Fungal Genome Initiative

White Paper developed
by the Fungal Research Community

February 8, 2002

Corresponding authors: Bruce Birren, Gerry Fink, and Eric Lander, Whitehead Institute Center for Genome Research, 320 Charles Street, Cambridge, MA 02141 USA.
Phone: 617-258-0900. E-mail: bwb@genome.wi.mit.edu.

1. Overview

Over the past 16 months, the fungal and genomics community has been working together to define a comprehensive Fungal Genome Initiative — an effort to jumpstart research on the fungal kingdom by prioritizing a set of fungi for genome sequencing. An initial high-priority set of 15 fungi has been selected, including fungi that present serious threats to human health, serve as important models for biomedical research, and provide a wide range of evolutionary comparisons at key branch points in the 1 billion years of fungal evolution. (Community discussion to identify additional high-priority targets is ongoing.)

The Fungal Genome Initiative was developed with broad participation of the fungal research community through a Steering Committee of fungal scientists and a series of meetings and presentations. The sequence information from these organisms will provide a wealth of data important to work in medicine, agriculture, and industry, and will spur progress in computational studies of eukaryotic biology and evolution.

In this white paper, we describe the rationale for a comprehensive Fungal Genome Initiative.

2. History and Promise of Fungal Genomics

The sequence of the genome of *Saccharomyces cerevisiae* was a landmark in genomics (Goffeau et al. 1996). The sequence had an immediate impact on all labs working with yeast by eliminating the years of effort previously associated with gene discovery. But the availability of the yeast genome sequence was even more profound because it made possible the first global studies of eukaryotic gene expression and gene function (e.g., Giaever et al., 1999; Winzeler et al., 1999; Birrell et al., 2001; Ooi et al., 2001).

The importance of the yeast sequence reached far beyond the yeast labs. It provided scientists working on human genes access to a deep body of knowledge about how those genes functioned, and allowed use of the exquisite tools of yeast genetics to further interrogate protein functions and interactions (Dodt et al., 1996; Foury et al., 1997; Primig et al., 2000; Bennett et al., 2001). The usage statistics of the Saccharomyces Genome Database (SGD) demonstrate the broad reliance on these data. The site currently averages nearly 120,000 accesses each week from over 10,000 unique sites (http://genome-www.stanford.edu/usage/sgd/). These data are clearly supporting research outside traditional yeast labs.

Since the sequencing of *S. cerevisiae*, progress on other fungal genomes has been remarkably limited. No additional finished fungal genomes have been published to date, although the fission yeast *Schizosaccharomyces pombe* and the filamentous fungi *Neurospora crassa* and *Ashbya gossypii* are nearing completion. A handful of other fungal projects are underway, although many have yielded only fragmentary sequence and many of these have not been made public. By contrast, over 50 microbial genomes have been completed and more than 120 bacterial genome sequence projects are underway.

This represents a serious imbalance considering the exceptional contributions that fungal genome sequence can provide to the study of eukaryotic biology and human medicine. Although fungal genomes are somewhat larger than bacterial genomes, many of the most important ones are relatively modest in size (7–40 Mb) and contain few repeats. They are thus ideal targets for whole-genome shotgun (WGS) sequencing. (Ten-fold coverage of a 10-Mb genome corresponds to only 200,000 passing reads.)

The high gene density of fungi makes them extremely cost-effective in terms of eukaryotic gene discovery. For example, *S. cerevisiae* contains a gene approximately every 2 kb (Goffeau et al., 1996), while the larger Neurospora genome averages a gene every 4 kb (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/).

Within fungal genomes lies the evolutionary history of the origins of many important biological processes found in higher eukaryotes. In addition, their experimental tractability make fungi among the most useful model systems in cell biology. Fungal cellular physiology and genetics share key components with animal cells, including multicellularity, cytoskeletal structures, development and differentiation, sexual reproduction, cell cycle, intercellular signaling, circadian rhythms, DNA methylation and

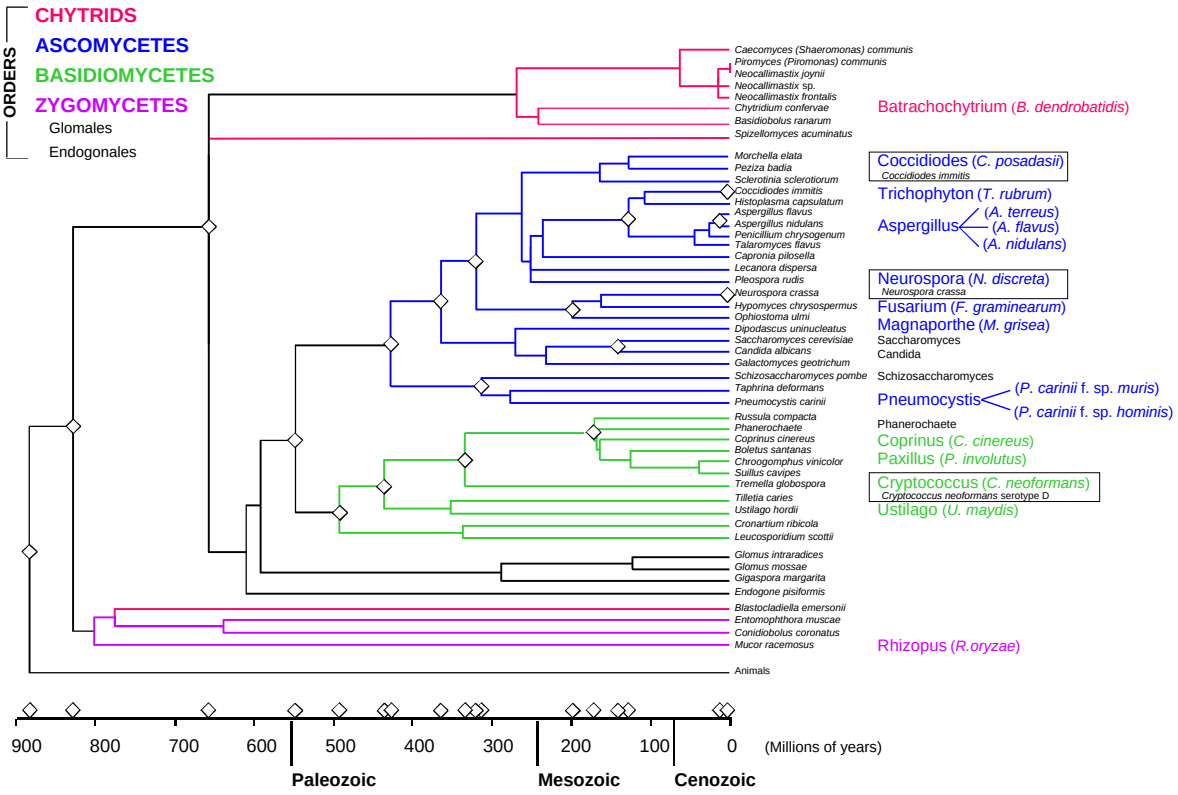
regulation of gene expression through modifications to chromatin structure, and programmed cell death. The shared origins of the genes responsible for these fundamental biological functions between humans and fungi continue to make the history and function of these fungal genes of vital interest to human biology.

3. Rationale for a Fungal Sequencing Program

We propose that fungal genomics should be approached in a kingdom-wide manner — that is, by selecting a balanced collection of fungi (rather than choosing individual fungi in isolation) that maximizes the overall value for comparative genomics, evolutionary studies, eukaryotic biology, and medical studies. The sequence of each organism should not only enable research in a specific research community, but should enhance the value of each additional sequence for comparative studies of the evolution of genes, chromosomes, and regulatory and biochemical pathways.

In this white paper, we describe a coherent collection of fungi that individually have great importance and together complement each other to provide a much larger whole (Figure 1). We begin by outlining the general considerations that justify a program for fungal sequencing.

FIGURE 1



Organisms in black type have sequence projects completed or underway.

◇ = branch points, and their approximate dates, across which evolutionary comparisons will be made.

▭ = boxed pairs show sequence comparisons for species that are approximately 5–10 million years apart.

3.1 Impact on human health. Fungal pathogens are devastating to human health. Fungal infections have lethal consequences for the growing population of patients immunocompromised with AIDS or therapeutically immunosuppressed after cancer chemotherapy or transplantation surgery. Emerging fungal infections represent an equally serious threat to healthy human populations. Identifying

effective therapies against these eukaryotes has been more difficult than for bacterial pathogens, and, as a result, few effective antifungals are currently available. Most of the existing drugs have serious side effects, and resistance to these compounds is an increasing problem. Genome sequence from pathogenic fungi will be the most efficient step in identifying potential targets for therapeutic intervention and vaccination among the largely unknown set of fungal proteins. In addition to their role as pathogens, the importance of fungi to human health includes their role in the development and production of critical pharmaceuticals, including antibiotics such as penicillin, cephalosporins, and cyclosporin, which have combined annual sales over $5 billion. Further, new recommendations are expected to triple the number of Americans taking cholesterol-lowering statins, produced by fungi, from the current 12 million up to 36 million in just a few years.

Fungi also exert a heavy influence on agriculture and our ability to feed the world's population. Plant pathogens destroy vast amounts of crops in the field and after harvest each year. For example, in the United States, where over $600 million is spent annually on agricultural fungicides, one-third of the peanut crop is lost to fungal infection. Worldwide it is estimated that the annual loss of rice due strictly to the rice blast fungus would be sufficient to feed 60 million people. Overall, crop losses due to fungi exceed $200 billion annually. Access to genome sequence is paramount to advance our knowledge of fungal infection as well as the interaction of pathogen and host. Sequence data will also provide crucial information on how these organisms reproduce and persist in the environment. In addition to their role as pathogens, fungi have additional critical but poorly understood roles in agriculture. Namely, the mycorrhizal fungi that grow interdependently with plant roots are critical for nutrient uptake by plants.

Finally, the effectiveness of fungi in causing both human and plant diseases has lead DARPA and NIAID to recognize a number of these organisms as potential biological weapons. Our heightened need for preparedness against biological warfare and bioterrorist attacks requires rapid development of the kind of diagnostic tools, treatments, and vaccines that will come from fungal genome sequencing.

3.2 Impact on human biology. Although *S. cerevisiae* provides key insights into the function of many human proteins, the small size of its genome limits its value as a tool for this purpose. A deeper sampling of fungal genes will rapidly increase the number of human proteins for which we can access homologues in model organisms. A recent review indicates that homologues exist for approximately 30% of all human proteins within the limited available fungal sequence data. This is more than twice the number of similar proteins found by examining the genome of *S. cerevisiae* alone (Zeng et al., 2001). Genome sequence from many diverse fungi, coupled with comparative and functional genomics, will advance our understanding of the eukaryotic proteome, which in turn will help define what it means to be a eukaryote. In doing so, we will learn not only how to manipulate fungi for our own purposes, but how to therapeutically intervene to manipulate human physiology for the treatment of metabolic and infectious diseases.

3.3 Impact on comparative genomics and evolutionary science. Comparative genomics and evolutionary genomic studies hold great promise, but these fields are still in their infancy. For example, most comparative analyses still consist of pair-wise sequence alignment of genes or short regions. In principle, more power can be obtained through the much wider comparison of many related organisms. In practice, however, many new methods will need to be developed to extract this power.

A fungal sequencing program represents an ideal system for establishing the power and limits of comparative studies because:

- The small genome sizes allow for the comparison of many complete eukaryotic genomes for small cost and effort.
- The fungal kingdom, with more than 1 million different species, displays extraordinary diversity.
- Genomes can be selected representing a wide variety of evolutionary distances, ranging from as little as 5 million years to approximately 1 billion years.

- Genomes can be selected representing specific branch points in the phylogenetic tree to illuminate the molecular basis for key biological innovations.
- Fungi offer outstanding opportunities to study natural populations and evolution. For example, there are over 4000 well-characterized natural isolates of Neurospora deposited at the Fungal Genetics Stock Center, taken from widespread and ecologically diverse regions. These represent five heterothallic species, four that must out-cross and one that can effectively self-cross.

Analyses of representative genomes distributed across the fungal tree are expected to provide the molecular basis for understanding the extraordinary diversity that has arisen over the estimated 1 billion years since divergence from a common ancestor. Further, the fungi in this study have been selected to allow comparisons of genomes with a wide range in times of divergence, including pairings of sibling species to permit comparison of very recent evolutionary events. The ability to compare genomes that represent a variety of evolutionary distances and relationships is exactly what is needed to build effective tools for comparative analysis and to define what are the appropriate evolutionary distances for answering different questions. Such comprehensive studies are difficult to accomplish with large mammalian genomes. Our ability to perform them rapidly in fungi will establish the foundation for comparative genomics and inform optimal choices when sequencing capacity exists for broad, deep sampling of these larger genomes.

3.4 User community. A fungal sequencing program will generate data of great interest to a number of user communities, including those engaged in comparative genomics, evolutionary studies, fungal biology, infectious disease, and computational biology.

4. Community Discussions Leading to This Proposal

4.1 Initial meeting. In November 2000, Dr. Gerry Fink, Director of the Whitehead Institute, invited a small group of fungal geneticists and biologists to discuss ways to accelerate the slow pace of fungal genome sequencing. Participants included academic and industrial fungal scientists as well as those with experience in genome sequencing and analysis. Attendees included: Gerry Fink, Whitehead Inst.; John Taylor, University Calif., Berkeley; Joe Heitman, Duke University; Ron Morris, UMDNJ-Robert Wood Johnson Medical School; Peter Hecht, Microbia, Inc.; and Eric Lander, Bruce Birren, Chad Nusbaum, and other members of the Whitehead Center for Genome Research.

The group unanimously concluded that the dearth of publicly available fungal genome sequence was a major barrier to biomedical research. Moreover, the group noted that the entire *S. cerevisiae* genome could be sequenced using just a few days' capacity at a large sequencing center. A broad initiative was conceived in which organisms would not be selected one at a time, but would be considered as part of a cohesive strategy. The primary selection criteria endorsed were:

- Importance of the organism in human health and commercial activities.
- Value of the organism as a tool for comparative genomics.
- Presence of genetic resources and an established research community.

4.2 Broadening the discussion. Subsequently, John Taylor and Bruce Birren each made presentations at the International Fungal Genetics meeting in Asilomar, CA, in March 2001 to describe the concept and solicit feedback about the desirability of such an initiative and about the organisms that should be considered.

To further develop the idea of a broad fungal sequencing program, Dr. Fink organized and chaired an Advisory Committee.

Advisory Committee:

Janis Antonivics, University of Virginia, Charlottesville
Bruce Birren, Whitehead Institute for Biomedical Research
Brendan Cormack, Johns Hopkins University School of Medicine
Melanie Cushion, University of Cincinnati Medical School
Ralph Dean, North Carolina State University
Peter Hecht, Microbia, Inc.
Joe Heitman, Duke University
John Kinsey, University Kansas Medical School; Director, Fungal Genetics Stock Center
H. Corby Kistler, USDA-ARS Cereal Disease Lab, University of Minnesota
James Kronstad, University of British Columbia
Eric Lander, Whitehead Institute for Biomedical Research
Lene Lange, Novozymes, A/S
Joyce Longcore, University of Maine
Ron Morris, UMDNJ-Robert Wood Johnson Medical School
Peter Phillipsen, University of Basel
Patricia Pukkila, University of North Carolina, Chapel Hill
Matthew Sachs, Oregon Health and Science University; Chair, Fungal Genetics Policy Comm.
Chuck Staben, University of Kentucky, Lexington
James Sweigard, E.I. Dupont de Nemours and Co.
John Taylor, University of California, Berkeley
Olen Yoder, Syngenta.

This group prepared a draft white paper that described what the community had begun to refer to as the Fungal Genome Initiative. The white paper laid out the goals and rationale for the initiative, as well as suggested a structure for such a broad, collaborative process. This white paper was circulated during the summer of 2001 among federal agencies with an interest in promoting fungal research. In fact, the white paper served as the direct inspiration for NHGRI's process for prioritizing organisms for sequencing.

4.3 Public meeting. To solicit the widest range of opinions, the Advisory Committee worked with NHGRI and NSF to sponsor a workshop on Fungal Genomics in Alexandria, VA, in November 2001. The agenda for the workshop is attached as Appendix A.

The Steering Committee for this workshop included Gerry Fink, Whitehead Institute, Chair; John Taylor, University of California, Berkeley; Ron Morris, UMDNJ-Robert Wood Johnson Medical School; Joe Heitman, Duke University; Ralph Dean, North Carolina State University; and Mary Anne Nelson, University of New Mexico.

Over 60 attendees heard 19 speakers representing the academic and industrial interests in medical, agricultural, industrial, evolutionary, basic biological fungal research, and informatics give their summary of what genome resources were most needed to spur research and development in their areas of interest. In addition, representatives were present from various federal agencies, including NHGRI, NIAID, NIGMS, NSF, USDA, DTRA, and NCBI. The invitee list and agenda is attached as Appendix B.

The workshop produced strong endorsement of an initiative in fungal genomics that would include rapid sequencing and public release of many fungal genomes chosen with a single, coherent plan. The attendees elaborated a plan and organizational structure to provide ongoing oversight, to ensure appropriate channels of interaction with the many independent research groups working on individual organisms, and to advise on the informatics needs of the different user communities and the best practices needed to meet these needs.

5. Choice of Organisms

The Advisory Committee, with input from the fungal biology community, has identified a coherent and complementary collection of 15 fungi for initial sequencing. These fungi are:

MEDICAL

Cryptococcus neoformans, Serotype A
Coccidioides posadasii
Pneumocystis carinii (human and mouse)
Trichophyton rubrum
Rhizopus oryzae

COMMERCIAL

Magnaporthe grisea
Aspergillus flavus
Aspergillus nidulans
Aspergillus terreus
Fusarium graminearum

EVOLUTION AND FUNGAL DIVERSITY

Neurospora discreta
Coprinus cinereus
Batrachochytrium dendrobatidis
Ustilago maydis
Paxillus involutus

In the final portion of this white paper, we discuss the specific rationale for the choice of fungi in terms of biological interest, genome size, availability of haploid strains, community, and evolutionary diversity.

6. Sequencing Approach

6.1 Deep-shotgun sequencing. For each fungus, we propose that deep-shotgun sequence with long-range links be obtained. Specifically, we propose the collection of 10X whole-genome shotgun (WGS) sequence from paired-end reads obtained from 4-kb plasmids (90%) and 40-kb fosmids (10%). In both cases the inserts will be prepared with randomly sheared genomic DNA. Such coverage of a genome corresponds to only 250,000 attempted reads per 10 Mb of genome, allowing for an 80% pass rate. For reference, this corresponds to about 2.5 days of capacity at the Whitehead Institute. Thus, a 40-Mb fungus could be sequenced with 10 days of capacity.

The proposed sequence will provide a total of approximately 100X physical coverage that should suffice to provide a high-quality assembly, although not yet a finished sequence. This expectation is borne out by experience with WGS assembly at Whitehead:

- With ~5X coverage of mouse, we have been able to assemble the mouse genome into contigs with an N50 length of 16 kb and supercontigs with an N50 length of 1.2 Mb.
- With ~10X coverage of *N. crassa*, we have been able to assemble the genome into contigs with an N50 length of 88 kb and supercontigs with an N50 length of 613 kb.

(Note: The Neurospora project has somewhat shorter supercontigs than the mouse, despite similar total physical coverage. This is because the Neurospora project employed (multi-copy) cosmids with inserts prepared by partial digestion, whereas the more recent mouse project employed (single-copy) fosmids with inserts prepared by random shear. The latter provides better representation and will be used in the projects discussed here.)

6.2 Finished sequence. We believe that producing a high-quality, deep-shotgun assembly in a rapid fashion should be the highest priority for this initiative, and that it will be extremely valuable to produce finished reference sequences for some of these genomes. We propose that all clones required for finishing be retained and that the decision to carry out finishing be prioritized by NHGRI staff on the basis of evolving assessment of cost and capacity.

6.3 Community distribution of large insert clones. The fosmid libraries for each genome should be made available for distribution through the Fungal Genetics Stock Center (FGSC). FGSC has been continuously funded by NSF for over 40 years to maintain and distribute resources for fungal biology, and it has agreed to participate in the Fungal Genome Initiative to distribute fosmid clones. (Indeed, Whitehead has an ongoing collaboration with the FGSC for this purpose in the context of the Neurospora project.)

7. Detailed Description of Organisms

The organisms, the rationale for sequencing and relevant genome information are summarized in Table 1 and are described in more detail on the following pages.

TABLE 1

	NAME	SIGNIFICANCE	EST. SIZE (Mb)
MEDICINE	***Cryptococcus neoformans***, Serotype A	Encapsulated basidomycete; cause of fatal meningitis in humans	24
	Coccidiodes posadasii	Soil ascomycete endemic in southwest, leads to fatal infection; bioterrorism threat	29
	Pneumocystis carinii (human and mouse)	Leading opportunistic infection associated with AIDS; drug-resistance emerging	7.5 and 6.5
	Trichophyton rubrum	Most common fungal infection in the world, uniquely adapted for growth on human skin	12
	Rhizopus oryzae	Cause of mucormycosis	35
COMMERCE	***Magnaporthe grisea***	Cause of rice blast; leading model for fungal–host interactions	40
	Aspergillus flavus	Source of aflatoxin in food stuffs and cause of human aspergillosis	40
	Aspergillus nidulans	Key model system for study of genetics and cell biology	31
	Aspergillus terreus	Major source of lovastatin	30
	Fusarium graminearum	Cause of head blight (scab) on wheat and barley crops; also produces toxins that destroy grain	40
EVOLUTION/ FUNGAL DIVERSITY	***Neurospora discreta***	Fungal model for study of population genetics	40
	Coprinus cinereus	Well-developed laboratory model for multicellular fungus showing coordinated development	37.5
	Batrachochytrium dendrobatidis	Recently described chytrid; cause of widespread amphibian population decline	30
	Ustilago maydis	Well-studied model for pathogen–host interactions	20
	Paxillus involutus	Mycorhizzal fungus that easily establishes mutualistic growth with tree roots under laboratory conditions	40

7.1. *Cryptococcus neoformans*, Serotype A

SIGNIFICANCE: *Cryptococcus neoformans* is an encapsulated fungal pathogen causing fatal meningitis in humans. The infection, initiated by inhalation into the lungs, occurs mainly in immunocompromised individuals, but can also occur in healthy individuals. *Cryptococcus neoformans* is usually found in tissues in the yeast form. Infection of the brain and meninges is the most common clinical manifestation. In immunocompetent individuals, the initial infection is usually controlled and asymptomatic and the organism remains dormant in a lymph node complex, much like tuberculosis. Reactivation occurs in immunocompromised hosts where the fungus can spread via the blood to infect the central nervous system. Once *C. neoformans* reaches this stage it can cause meningitis that is uniformly fatal if untreated. Few antifungal agents exist and drug-resistant strains are emerging.

There are four serotypes of *C. neoformans*. The serotype D MATa strain JEC21 was the first strain chosen for sequence studies because of its advanced genetic tools; however, more than 90% of clinical isolates and more than 99% of isolates from AIDS patients are of the more divergent serotype A strains. Therefore, understanding this disease requires sequence of a serotype A strain for comparison with the serotype D genome.

GENERAL DESCRIPTION: *Cryptococcus neoformans* is unique among the most common human fungal pathogens in that it is a basidiomycete, thus it is evolutionarily divergent from the more common pathogenic ascomycetes (e.g., *Candida albicans*) and more closely related to wood rotting fungi (e.g., *Phanerochaete chrysosporium*), mushrooms (e.g., *Coprinus cinereus*), and plant pathogens (e.g., *Ustilago maydis*).

Cryptococcus neoformans elaborates two specialized virulence factors, the polysaccharide capsule, which inhibits phagocytosis, and melanin, which serves as an antioxidant. The typical vegetative form of *C. neoformans* is the yeast form. The organism can also undergo sexual reproduction and form basidiospores. Sexual reproduction appears to occur much less frequently in nature than asexual or vegetative reproduction.

GENOME FACTS: Most isolates of *C. neoformans* are haploid. The size of the genome is approximately 24 Mb with 12 chromosomes. *Cryptococcus neoformans* has a defined sexual cycle involving mating between cells of the MATa and MATα types. Thus, classical genetic approaches can be applied to study this organism.

Molecular biology approaches have been developed, including transformation systems and gene disruption by homologous recombination. In addition, episomal plasmids, cDNA expression libraries, two-hybrid libraries, and several robust animal models (including the rabbit meningitis model, murine tail-vein injection, and murine and rat inhalation models) are available. It is now possible to identify a gene of interest, disrupt the gene, and test the impact on biology and virulence.

COMMUNITY: There is a large network of cooperating laboratories working on *C. neoformans* that have organized efforts to share genome maps and sequence information. The medical and fungal research communities have a broad interest in sequencing *C. neoformans* due to its importance as a human pathogen and its evolutionary divergence from other human fungal pathogens. Joe Heitman (Duke University) will supply the haploid DNA for this organism. In addition, the comparison of the genomes of the different *C. neoformans* serotypes will be of interest to a wide range of genome and evolutionary scientists outside those working directly on this organism.

7.2. *Coccidiodes posadasii* (= *Coccidiodes immitis* non-California species)

GENERAL: *Coccidioides posadasii* is an ascomycete soil fungus found in the desert regions of North America, Mexico, and scattered areas in South America, particularly Argentina. It is one of a pair of species (the other being *C. immitis*) found in the Central Valley of California, San Diego, and Baja, California, that causes coccidioidomycosis. NIH has allocated funding for sequencing an isolate of *C. immitis*. Our inclusion of *C. posadasii* in this proposal is to take advantage of that sequence for comparative genomics among close relatives.

In the soil, *C. posadasii* produces hyphae and mitospores. *C. posadasii* is specifically adapted to live in mammals, including humans. When mitospores are inhaled, they enlarge to make spherules, which divide internally to make endospores. One spherule can release thousands of endospores, which, in turn, mature into spherules. No sex is known in *C. posadasii*.

SIGNIFICANCE: *C. posadasii* is medically significant because it can cause serious and sometimes fatal disease in otherwise healthy people. The rate of infection varies from year to year, but an average of 100,000 persons per year are infected in the United States alone. Approximately 60% of those infected are asymptomatic, 35% are symptomatic, and some illnesses last months before resolving. Approximately 5% of infections require medical therapy and some of these infections are fatal. The percentage of serious infections is much higher in people with poor immune systems (e.g., people on chemotherapy, with transplants, or AIDS). Infections in older people are also more likely to require hospitalization or be fatal. When *C. posadasii* spreads beyond the lungs, it frequently goes to the bones and joints, skin, and brain. Brain infection (meningitis) is lethal if not treated. Bone and skin infection are less dangerous but can persist for life. Further information is available at: http://www.cdc.gov/ncidod/eid/vol2no3/kirkland.htm and http://www.arl.arizona.edu/vfce/.

As a bioterrorism threat, the U.S. government classified *C. immitis* as an organism with bioterrorist potential. At the time of that legislation, *C. posadasii* was thought to be identical to *C. immitis* and an equal threat.

On the evolutionary front, *C. posadasii* and *C. immitis* have become models for studying the evolutionary biology of pathogenic fungi. Over 200 strains of the two species from across the Western Hemisphere have been typed at nine microsatellite loci, more than for any other fungal species. Most of the strains are from *C. posadasii* due to its greater geographic range.

GENOME FACTS: The *C. posadasii* genome is approximately 29 Mb and consists of four chromosomes. It is a haploid organism, so sequence assembly and analysis will not be complicated by heterozygosity.

COMMUNITY: As discussed above, the sequence will be of broad interest to fungal researchers and to eukaryotic biologists in general. The haploid DNA for *C. posadasii* will be supplied by John Taylor at the University of California, Berkeley. Currently, cDNA libraries from the spherule (invasive form) have been produced. Two *C. immitis* genomic libraries were made by The Institute for Genomics Research (http://www.tigr.org), where the NIH-sponsored sequence of *C. immitis* is being obtained. TIGR has sequenced 2000 expressed sequence tags (cDNA) and more than 2000 genomic clones. This effort has already resulted in the discovery of more than 350 new genes in *C. immitis*. Laboratories currently studying *C. posadasii* with the intention of producing a vaccine include Theo Kirkland, UCSD; Demo Pappagianis, UCD; John Galgiani, UAZ; Rebecca Cox, Texas Chest Hospital; and Garry Cole, U. Toledo.

7.3. *Pneumocystis carinii* f. sp. *hominis* / *Pneumocystis carinii* f. sp. *muris*

SIGNIFICANCE: Pneumonia caused by *Pneumocystis carinii* remains the leading opportunistic infection associated with AIDS patients. Currently, less than a dozen genes of human *P. carinii* have been identified. In addition, human *P. carinii* remains resistant to most common antifungal and antiprotozoal agents. It will be the goal of this project to provide the full genome sequence of human *P. carinii* to aid in rational drug development for treatment and prophylaxis of human pneumonia. Genomic data from the ongoing rat *P. carinii* sequencing project combined with the proposed sequencing of the mouse and human *P. carinii* organisms will provide an opportunity to compare synteny and genetic composition of the lower branches of the fungal kingdom with those occupying higher levels.

GENERAL DESCRIPTION: Organisms known collectively as *P. carinii* are found in the lungs of almost every mammalian species evaluated for their presence. The members of the Pneumocystis group are mammalian-species specific. Thus, the human *P. carinii* cannot infect rats and vice versa. In fact, rat *P. carinii* cannot infect a more closely related species, such as the mouse. In a movement toward recognition of these organism populations as distinct species, the Pneumocystis scientific community instituted nomenclature recommended by the Botanical Code of Nomenclature for physiological fungal variants. It is the human and mouse forms of *P. carinii* that we propose for sequencing: *P. carinii* f. sp. *hominis* and *P. carinii* f. sp. *muris*.

The lack of a culture system has hindered progress in understanding the *P. carinii* life cycle, mode of transmission, and metabolic functions. Unlike any extant fungus, *P. carinii* possesses only a single copy of the nuclear ribosomal RNA locus and has little to no detectable ergosterol. It is resistant to standard antifungals, including amphotericin B, and standard azoles such as fluconazole.

The placement of *P. carinii* in the fungal kingdom is a recent phenomenon; until just a few years ago it was mistakenly thought to be a protozoan. Phylogenetic trees based on nuclear 16S-like RNA or mitochondrial 26S-like RNA sequences have not identified any closely related family members, though the fission yeast *Schizosaccharomyces pombe* appears to have the closest affinity. It has been suggested that the Pneumocystis species represent an early divergent line in the fungal kingdom that may have branched coincident with the branching of the basidiomycete and ascomycete lineages.

GENOME FACTS: The genome of human *P. carinii* is estimated to be 7.5 Mb and mouse *P. carinii* to be 6.5 Mb. The genomes are assumed to be haploid, but recent evidence suggests the rat *P. carinii* genome may be aneuploid. Shotgun libraries of chromosomes and chromosome groups are being used to sequence the rat *P. carinii* genome. Sufficient material can be obtained from human and mouse *P. carinii* to create the libraries and mapping reagents as has been accomplished in rat.

COMMUNITY: The medical and fungal research communities have a broad interest in sequencing this organism due to its importance as a human pathogen and because of its unique placement within the fungal kingdom. Approximately 35 laboratories worldwide work directly with Pneumocystis. Some laboratories interested in aspects of the biology of human *P. carinii* (e.g., emerging drug resistance) are impeded in such studies due to the technical challenge of culturing these organisms. Other laboratories investigate the host response to *P. carinii* using the mouse and rat as the primary models. Two websites host Pneumocystis genome information: http://biology.uky.edu/Pc and http://gene.genetics.uga.edu. An additional website is under construction (http://www.pneumocystis.org) that will compile genomic tools such as EST libraries and software for imaging contigs and chromosomes. Melanie Cushion at the University of Cincinnati will serve as the contact for obtaining *P. carinii* haploid DNA. The ability to perform comparative analyses of the genome sequences for the rat, mouse, and human *P. carinii* will be of broad interest to genome scientists as well as those interested in the evolution of pathogen–host interactions.

7.4. *Trichophyton rubrum* Strain D12

SIGNIFICANCE: Thirty to seventy percent of adults are asymptomatic carriers of dermatophytes — fungi that cause skin and nail infections — of which athlete's foot (*Trichophyton rubrum*) is the most common. Transmission of this fungal infection is common in public facilities, including swimming pools, gyms, camps, prisons, and military bases. In addition to widespread chronic disease in all populations, dermatophytes can cause acute and severe problems in specific geographic locations, for example, the 1968 epidemic in the Mekong valley during the Vietnam war. (The strain that will be sequenced was involved in the Mekong epidemic.)

While not fatal, these infections cause tremendous pain and account for significant costs to society. Infection is not limited to the feet, but can occur on the torso, groin, head, hair follicles, and arms. Available drugs are effective against common dermatophyte infections. However, many of these drugs can be obtained over the counter, creating a potential for over or improper use that contributes to the development of drug resistance.

GENERAL DESCRIPTION: Of the 42 known dermatophyte species, 31 are established to be pathogenic to humans. The species are distinguished by morphology of hyphae and conidia (spores) as well as mating criteria. Many of these species have a sexual cycle. The dermatophytes are divided into geophilic (soil dwelling), zoophilic (animal-specific), and arthropophilic (human-specific) species.

T. rubrum is a cosmopolitan arthropophilic dermatophyte found throughout the world. It is the most frequent etiologic agent of skin and nail infections. Once an infection with *T. rubrum* has been controlled by therapy, the patient is a life-long carrier of *T. rubrum*. Similar to other arthropophilic species, *T. rubrum* appears to have lost its mating ability. Fifty-eight percent of the dermatophyte species isolated at clinical laboratories are *T. rubrum*, 27% are *T. mentagrophytes*, 7% are *T. verrucosum*, and 3% are *T. tonsurans*. Based on these frequencies, *T. rubrum* is the best choice for the genome sequencing of a dermatophyte.

GENOME FACTS: Knowledge of the dermatophyte genome is extremely limited. There are no published reports on the size, complexity, or AT content of the genome. The genomic DNA can be separated on a pulsed field gel electrophoresis (PFGE) apparatus into four chromosomes. The genome is thought to be 30–40 Mb in size, its AT content is average at 50% (based on the few sequenced genes), and 5–10% of the genome is in repetitive DNA. Gene analysis in *T. rubrum* has been very limited. Partial coding regions of eight nuclear-encoded genes have been analyzed, including actin, heat shock protein 70, and two major surface antigens. The eight genes have been analyzed from different strains of *T. rubrum.*

COMMUNITY: At this time there are few basic research laboratories working on dermatophyte infections. Descriptive studies, case reports, and discussions of therapeutic options are common. Clinical microbiology laboratories report on the isolation and identification of various dermatophyte species. Susceptibility levels to many of the current drugs are determined for many species. Some evolutionary biologists analyze how the dermatophytes are related to other organisms using ribosomal RNA sequences and a select number of other genes. In terms of basic investigative biological research, however, there is little to report for a group of organisms that cause the most common fungal infection in humans.

Grant support is similarly lacking. In the last 10 years, support from the NIH for research on dermatophytes has been limited to a grant to study immunology, a few grants to study current and novel drug therapies, and a grant to fund an Allergic Disease Center that includes the study of dermatophyte infections. A genome sequence for *T. rubrum* would allow current technologies and techniques to be applied to the dermatophytes for the development of new diagnostics, therapies, and vaccines.

Ted White (University of Washington), in addition to supplying the haploid DNA for this organism, assembled a group of researchers who have worked on dermatophytes and who support the sequencing effort. These include clinical researchers and epidemiologists.

7.5. ***Rhizopus oryzae* (= *R. arrhizus*)**

SIGNIFICANCE: The genus *Rhizopus* is classified under the family Mucoraceae in the order Mucorales of the phylum Zygomycota. *R. oryzae* is the most important and representative agent of mucormycosis. *R. oryzae* and the other members of Mucorales are recovered in profusion from decaying vegetables, fruits and their seeds, grains, compost piles, soil, animal excreta, and molding bread. Since it is ubiquitous in nature, mucormycosis cases are reported worldwide. The majority of mucormycosis patients have a serious underlying condition, such as diabetes mellitus, immunosuppression, starvation, burns, or other major trauma. Pathologically, mucormycosis is characterized by vascular invasion with hyphae, infarction and necrosis of tissue, and by an acute or subacute course of infection. Of the many forms of clinical manifestation, the most common form caused by Rhizopus species is the rhinocerebral and craniofacial mucormycosis. This originates in the paranasal sinus but presents in contiguous structures of orbit, palate, face, nose, or brain. Ordinarily death occurs in untreated cases within 4 weeks of onset.

The most prominent predisposing factor for the facial cranial mucormycosis is diabetes mellitus, and the diabetic population is on the rise worldwide. The second most common form of mucormycosis is pneumonia, which occurs most frequently among patients with hematologic disorders, lymphoma, severe nutropenia, or history of deferoxamine therapy. Infection in these patients has been observed to be fatal in a very short period of time in all cases reported to date. Although the number of cases has been small, children appear overrepresented among the mucormycosis patients without a known underlying disease. Rhizopus species also can cause skin and soft tissue infection in the setting of local trauma or by the hematogenous route.

The mucormycotic agents are the least-studied pathogenic fungi at the molecular level, and understanding their pathobiology will be helped considerably by obtaining the genomic sequence of a representative species.

GENERAL DESCRIPTION: The zygomycetes are evolutionarily divergent from ascomycetes, which harbor the most common human pathogenic fungi. *R. oryzae*, like other members of Mucorales, is a rapidly growing mold that propagates by hydrophobic sporangiospores that readily disperse after maturation. It is one of the common laboratory contaminants due to its ubiquity in soil and decomposing organic material. It is different from ascomycetes or basidiomycetes in that its hyphae are tube-like, without septation, and the cell wall contains chitosan and chitin instead of glucans, mannans, and chitin. Asexual spores are produced within sporangia and are released upon maturation. The fungus can also undergo sexual reproduction and produce zygospores upon mating between the positive and negative strain on an appropriate mycological medium.

GENOME FACTS: *R. oryzae* occurs as a haploid; its genome size is unknown. Although *R. oryzae* has a defined sexual cycle involving mating between the positive and the negative strains, classical genetic approaches have rarely been applied to study this organism. That is because germination of zygospores takes months and the rate of germination is erratic. Molecular biological approaches have been developed, including transformation systems and gene disruption by homologous integration. In addition, episomal plasmids, cDNA expression libraries, and animal models are available. It is now possible to identify the genes of interest and study their impact on pathobiology.

COMMUNITY: As discussed above, the sequence will be of broad interest to fungal researchers and to eukaryotic biologists in general. The community studying mucormycosis is much smaller than communities studying mycosis of fungi that cause cryptococcosis, candidiasis, and aspergillosis. However, the availability of genomic sequence from *R. oryzae*, will certainly induce an influx of researchers into the field. Ashraf Ibrahim at the UCLA School of Medicine will supply haploid DNA for this organism.

7.6. ***Magnaporthe grisea***

SIGNIFICANCE: Rice blast disease, caused by *Magnaporthe grisea*, is the major fungal threat to the world food supply. Loss of rice due to this disease represents sufficient food to feed 60 million people per year. *Magnaporthe grisea* is the premier organism for the study of fungal–host interactions because it grows on defined media, unlike many phytopathogenic fungi such as the mildews and rusts. The process of infection is similar to that of other fungal pathogens but is mediated through a specialized infection cell that produces an appressorium, which pierces the host tissue. *Magnaporthe grisea* is especially useful in dissecting the host–pathogen interaction because the early steps in infection — including germination, appressorium formation, and penetration — can be studied independently of the host. Moreover, numerous genes controlling host recognition, endogenous signaling pathways, infection-related morphogenesis, conidiation, and pathogenicity have been cloned and characterized. In addition, the host range of relatives is broad — races of the fungus attack specific cultivars of rice or other grasses, including turfgrass and major cereals.

GENERAL DESCRIPTION: *Magnaporthe grisea* is a haploid filamentous ascomycete. The fungus is heterothallic with mating controlled by the alternative mating loci, Mat1-1 and Mat1-2. Mating occurs readily when opposite mating types are mixed on oatmeal agar. Several positively selectable drug resistance markers are available that facilitate genetic analysis. *M. grisea*, like many foliar pathogens, is well adapted to attack and penetrate its host. All aerial parts of the plant are subject to invasion, but losses are most devastating when the panicle or node at the base of the panicle is infected and killed, resulting in loss of grain set.

GENOME FACTS: The genome of *M. grisea* is estimated to be approximately 40 Mb contained in 7 chromosomes. There are several published genetic maps for *M. grisea* based on molecular (mainly RFLP) markers. These maps have been integrated, providing markers every 125 kb across the genome.

Other resources for genome analysis include a number of different DNA libraries, represented by cDNA, cosmid, and BAC libraries. At least 2400 sequenced cDNAs, or ESTs, have been determined and deposited in GenBank. Three BAC libraries have been constructed for *M. grisea*.

COMMUNITY: There are more than a dozen major laboratories working in rice blast in the United States. The initiative to sequence *M. grisea* is widely endorsed by applied and basic researchers as well as by scientists focused on rice research. The availability of genome sequence for both the host (rice) and *M. grisea* make this sequence especially valuable to those interested in the genomic basis for pathogen–host interactions. In addition, the observed synteny between *M. grisea* and *N. crassa* make the comparative analysis of these genomes especially appealing to genome scientists. The overall purpose of the initiative is to build on available resources and to provide a comprehensive understanding of the pathogenic process and related fungal biology. Ralph Dean of North Carolina State University will supply the haploid DNA.

7.7. ***Aspergillus flavus***

SIGNIFICANCE: *Aspergillus flavus* causes aspergillosis, a life-threatening human disease, particularly in patients who are immunosuppressed or have chronic lung disease. *Aspergillus flavus* is responsible for about 30% of the cases of aspergillosis. *Aspergillus flavus* also produces aflatoxin, the most important of the known mycotoxins. The fungus infests and produces aflatoxin on a variety of stored grains, including corn and other agricultural commodities. Aflatoxin, one of the most carcinogenic compounds known, causes liver cancer in humans. Annual economic losses for crops contaminated with aflatoxin is hundreds of millions of dollars.

Because it produces aflatoxin, *A. flavus* is mutagenic, teratogenic, and acutely toxic to most animals and man. *Aspergillus oryzae*, a variant of *A. flavus*, is used to make soy sauce. Apparently, commercial strains of *A. oryzae* do not produce aflatoxin.

GENERAL DESCRIPTION: *Aspergillus flavus* is a member of the ascomycetes. Like almost all of its medically and commercially important relatives, *A. flavus* is an imperfect ascomycete that does not produce ascospores. It grows rapidly as a haploid filamentous fungus on solid or liquid media under a variety of nutritional conditions. It makes asexual spores but has no meiotic spore products. It is usually differentiated from other ascomycetes by its pattern of asexual spore production. *A. flavus* and *A. oryzae* appear to be variants of the same species.

GENOME FACTS: *Aspergillus flavus* has 8 chromosomes with at least one marker on each chromosome. The genome is estimated to be about 40 Mb, similar to other Aspergilli. There is a 60-kb gene cluster encoding the enzymatic activities, and regulatory proteins necessary for the biosynthesis of aflatoxin are syntenic with a similar cluster involved with the biosynthesis of sterigmatocystin, an intermediate in aflatoxin biosynthesis that is produced by *A. nidulans*. Aflatoxin mutants, developmental mutants, and GFP-tagged strains are available. Plasmid and cosmid libraries are available, and an EST library is currently being sequenced, which will also be publicly available.

COMMUNITY: Many laboratories worldwide are interested in *A. flavus* from a medical and agronomic mycology point of view, but only about 10 laboratories do molecular work with the organism. A website that reflects the medical and agronomic interests can be found at http://www.aspergillus.man.ac.uk/. Haploid DNA for *A. flavus* is available from Ron Morris (UMDNJ–The Robert Wood Johnson Medical Center). The ability to perform comparative analysis of the *A. flavus* genome with that of the other Aspergillus species make this an especially useful data set for computational scientists interested in evolution and comparative genomics.

7.8. *Aspergillus nidulans*

SIGNIFICANCE: *Aspergillus nidulans* is one of the critical fungal systems in genetics and cell biology. *Aspergillus nidulans* is important because it is closely related to a large number of other Aspergillus species of industrial and medical significance — e.g., *A. niger*, *A. oryzae*, *A. flavus*, and *A. fumigatus* — and serves as a model for their biology. Unlike these other Aspergilli, which are asexual, *A. nidulans* has a well-characterized, conventional genetic system. It undergoes DNA-mediated transformation, and genes from other Aspergilli as well as some mammalian genes function in *A. nidulans*.

Aspergillus nidulans is a particularly useful model organism for studies of cell biology and gene regulation. The initial work on the genetics of tubulin and microtubules was done in *A. nidulans*. Similarly, *A. nidulans* contributes to our understanding of mitosis and the intracellular functions of the mitotic motors kinesin and cytoplasmic dynein. Carbon regulation and nitrogen regulation are also well studied. One useful consequence of these regulatory studies is the characterization and development of the alcA alcohol dehydrogenase regulatable promoter (which is induced by alcohol and repressed by glucose) as a useful tool to control gene expression.

GENERAL DESCRIPTION: *Aspergillus nidulans* is an ascomycete. It grows rapidly as a filamentous fungus on solid or in liquid media under a variety of nutritional conditions. *Aspergillus nidulans* is homothallic, which means that any two strains can be mated directly. It is normally haploid, but can also be induced to grow as a heterokaryon or a vegetative diploid. It produces both asexual spores (conidia) and sexual spores (ascospores). It undergoes development to produce at least nine different cell types.

GENOME FACTS: The size of the *A. nidulans* genome is approximately 31 Mb. It has 8 well-marked chromosomes containing an estimated 11,000–12,000 genes. Approximately 900 genes have been identified in *A. nidulans* by conventional matings; 432 have been mapped to locus, and 254 are cloned and sequenced.

A partial sequence (2–3-fold coverage) of the *A. nidulans* genome was produced by Monsanto. This has been released, but only to academic laboratories under certain restricted conditions that limit genomic investigation. Because of the low coverage, the sequence itself has only relatively small contigs. From this sequence, 29% of the genes have a known function, 23% have a putative function, and 48% are unknown. In addition, although 61% of *S. cerevisiae* ORFs have homologues in *A. nidulans*, 31% of *A. nidulans* ORFs are unrepresented in *S. cerevisiae*.

Plasmid, cosmid, EST, and BAC genomic libraries are publicly available. A physical map is available at http://gene.genetics.uga.edu/index.html. Chromosome linkage maps are available at http://www.gla.ac.uk/Acad/IBLS/molgen/aspergillus/.

COMMUNITY: The well-characterized genetic system of *A. nidulans* and its close relation to medically and industrially significant organisms lend strong support to the sequencing of its genome. Currently, approximately 40 laboratories worldwide focus on the genetics and molecular biology of *A. nidulans*. Nine of these laboratories — representing Australia, England, France, Germany, and the United States — have joined to form a steering group to promote sequencing and annotation of the *A. nidulans* genome. Haploid DNA for *A. nidulans* is available from Ron Morris (UMDNJ–The Robert Wood Johnson Medical Center). The ability to perform comparative analysis of the *A. nidulans* genome with that of the other Aspergillus species make this an especially useful data set for computational scientists interested in evolution and comparative genomics.

7.9. ***Aspergillus terreus***

SIGNIFICANCE: *Aspergillus terreus* is the major source of lovastatin. The worldwide market for statins is more than $12 billion annually. Lovastatin was the first drug of the "statin" class to be approved by the FDA for the treatment of hypercholesterolemia in humans. Lovastatin reduces both normal and elevated LDL cholesterol concentrations. This includes reduction of serum cholesterol. In addition, lovastatin is a polyketide whose biosynthesis and regulation is likely relevant to that of other structurally related and commercially important molecules.

In addition to lovastatin, *A. terreus* produces numerous other secondary metabolites (e.g., patulin, citrinin, isoterrin, asterriquinone) and commercially important enzymes (e.g., xylanase).

Aspergillus terreus is also related to important human pathogens (e.g., *A. fumigatus*), strains that produce dangerous toxins on agricultural products (e.g., *A. flavus*), as well as strains used for industrial enzyme production (e.g., *A. oryzae*) and, therefore, will be useful in comparative genome analysis studies.

GENERAL DESCRIPTION: *Aspergillus terreus* is a filamentous ascomycete and is commonly found in soil. It is occasionally reported as an opportunistic pathogen of humans and animals. It grows well on either solid or liquid media and readily produces characteristic globe-shaped structures bearing asexual spores.

GENOME FACTS: *Aspergillus terreus* has a haploid genome that is expected to be 25–35 Mb in size, and has a G+C content of 50–60%. More than 30 genes have been sequenced from *A. terreus*, including 18 genes required for lovastatin biosynthesis. Most genes contain multiple, small introns.

This species is highly amenable to molecular genetic manipulation with a variety of available tools, including gene libraries, transformation methods, dominant selectable markers, and defined gene expression systems.

COMMUNITY: *Aspergillus terreus* is actively studied by a number of laboratories in academia, biotechnology, and the pharmaceutical industry. Both genomic and cDNA libraries have been constructed from this organism. "Early" lovastatin production strains as well as wild isolates are publicly available from fungal stock centers (e.g., ATCC, NRRL). Todd Milne (Microbia, Inc.) will provide haploid DNA. Comparative analysis of the *A. terreus* genome with that of the other Aspergillus species make this an especially useful data set for computational scientists interested in evolution and comparative genomics.

7.10. *Fusarium graminearum*

SIGNIFICANCE: Fungi in the genus Fusarium cause a variety of blights, seedling disease, root rots, or wilts on nearly every species of cultivated plant. Fusarium head blight (scab) of wheat and barley has emerged as the plant disease with the greatest impact on U.S. agriculture and society. The disease, caused by *Fusarium graminearum*, is increasingly becoming a threat to the world's food supply due to recent head blight outbreaks in Asia, Canada, Europe, and South America. The fungus also causes disease on corn and rice. For many Fusarium diseases, effective fungicides and highly resistant plant cultivars are not available. The pathogen poses a two-fold threat: first, infested cereals show significant reduction in seed quality and yield; second, scabby grain is contaminated with trichothecene and estrogenic mycotoxins, making it unsuitable for food or feed.

As a food safety issue, trichothecene toxins such as "vomitoxin" (deoxynivalenol) pose a serious hazard to human and animal health because these sesquiterpenoids are potent inhibitors of eukaryotic protein biosynthesis. Vomitoxin causes weight loss and feeding refusal in nonruminant livestock. Human ingestion of grain contaminated with *F. graminearum* is associated with alimentary toxic aleukia as well as illness characterized by nausea, vomiting, anorexia, and convulsions. Trichothecenes are also powerful modulators of human immune function. Fusarium infections are also a significant risk factor for immunocompromised individuals, and do not respond to currently available antifungal compounds. Without effective therapy, disseminated Fusarium infections are almost always fatal.

GENERAL DESCRIPTION: *Fusarium graminearum* is a filamentous ascomycete. The fungus grows well on defined medium in pure culture as a haploid mycelial thallus and can complete a sexual cycle in 4–6 weeks; abundant asexual sporulation occurs in less than 1 week. Although it is homothallic, outcrossing of strains is possible and can be promoted by manipulation of the mating type locus. A repeatable pathogenicity assay is adapted for both plants and animals, allowing for the study of the genetics of pathogenicity.

GENOME FACTS: The genome size of *F. graminearum* is approximately 40 Mb. A genetic map containing 444 AFLP-defined loci and with 9 linkage groups has been constructed. The G+C content of characterized portions of the genome, such as the mating type locus, is approximately 49%. Cosmid and BAC libraries with 10-fold coverage are available. Approximately 12,000 ESTs from various growth conditions and sources are available and funding has been provided to sequence at least 18,000 additional ESTs. Several selectable markers and reliable transformation procedures are available for efficient transformation. Low-level genomic sequence coverage (1–3-fold) has been obtained by private companies, including Syngenta. These data are not publicly available.

COMMUNITY: Because of its importance as a plant pathogen, national and international communities are studying Fusarium and broad support exists for sequencing *F. graminearum* in particular. The U.S. Wheat and Barley Scab Initiative meets annually and has a well-maintained website (http://www.scabusa.org). It also has its own USDA-funded grants program that distributes over $5 million per year. While most funded scientists are involved in plant improvement programs, a growing number are focused on characterizing the genome of the fungus. Furthermore, scientists working on related species of Fusarium are interested in having a genome for a closely related and pathogenic fungus. *F. graminearum* haploid DNA will be made available by H. Corby Kistler at the University of Minnesota. The International Society of Plant Pathology maintains a standing committee on Fusarium and an international Fusarium workshop is held on a regular basis. As the first Fusarium to be sequenced the *F. graminearum* genome sequence will immediately be valuable to those working with other species, including important plant and human pathogens.

7.11. *Neurospora discreta*

SIGNIFICANCE: Medical: Neurospora species are the preeminent model filamentous fungi. Research on filamentous ascomycete opportunistic fungi will be enhanced by the ability to compare genetic and genomic results to those obtained with Neurospora microarrays. Evolution: Comparative genomics is in its infancy. Our ability to compare the *N. discreta* genome to the already sequenced *N. crassa* genome will teach us what is possible in close comparisons of eukaryotic genomes — lessons that will be essential to the many comparisons of larger eukaryotic genomes that are certain to come. In terms of fungal evolution, there is no fungal counterpart to Drosophila, an organism that is both a genetic and population genetic model. Neurospora will become that organism due to the fact that it already is a genetic model and the fact that *N. discreta* has just been recognized to colonize recently burned vegetation throughout western North America. Importantly, sufficient cell mass can be found on vegetation to enable functional-genomics studies on gene expression under natural conditions. Genomic comparison between the existing *N. crassa* sequence and that of *N. discreta* will provide the information needed to exploit the *N. discreta* collections. Many basic questions in evolution and ecology will be addressed using the *N. discreta* genomic sequence and the collections of individuals; this includes the molecular basis for adaptation, one of the "holy grails" of evolutionary biology.

GENERAL DESCRIPTION: *Neurospora discreta* is a filamentous ascomycete that is ubiquitous worldwide. As is typical for other Neurospora species, it is associated with burned vegetation. In particular, *N. discreta* is found in recently burned forest trees and large shrubs, including pine, fir, juniper, aspen, alder, and willow. In the past 18 months a collection of over 400 *N. discreta* isolates from New Mexico to Alaska was built, with many sites providing populations of over 30 individuals. The NSF provided funding for the Whitehead Institute to obtain a complete genome sequence for *N. crassa*. The sequence for *N. discreta* will enable researchers to take advantage of the *N. crassa* sequence for comparative genomics among close relatives.

N. discreta is a quintessential ascomycete fungus, with sexual spores, asexual spores, two mating types, mating regulated by two idiomorphs, and mating partners attracted by pheromones. What sets *N. discreta* apart is its abundance throughout western North America, and the fact that many natural colonies are large enough to provide RNA for microarray experiments in natural, field conditions.

GENOME FACTS: The *N. discreta* genome is approximately 40 Mb with 7 chromosomes. It is a haploid organism, so sequence assembly and analysis will not be complicated by heterozygosity.

COMMUNITY: A large community of Neurospora biologists exist overseas and in the United States, and an International Neurospora Conference is held biannually at Asilomar (next meeting March 14–17, 2002). The existence of an *N. crassa* genome is already attracting developmental biologists to the fungal genetics community; and it is certain the existence of an *N. discreta* sequence will do the same for evolutionary biologists and bioinformaticists. The comparison of the genomes will also be of interest to a broad range of genome and evolutionary scientists outside those working directly on this organism.

7.12. *Coprinus cinereus* (Schaeff. ex Fr.)

SIGNIFICANCE: Genomic analysis will reveal whether the genes responsible for the multicellularity of *Coprinus cinereus* are the progenitors of those in other multicellular organisms (e.g., nuclear receptors) or are innovations restricted to the fungal kingdom. *Coprinus cinereus* is a higher fungus (Agaricales) comprised of many cell types and, therefore, provides a window on the development of multicellularity within a single kingdom. The cells of this organism undergo coordinated developmental events not observed in unicellular fungi, including secondary and tertiary gill formation, cap expansion and autodigestion, active spore discharge, and stalk elongation, phototropism, and gravitropism.

GENERAL DESCRIPTION: *Coprinus cinereus* is a multicellular basidiomycete with a typical mushroom form that undergoes a complete sexual cycle. Unlike most mushrooms, *C. cinereus* can complete its entire life cycle (2 weeks) in the laboratory. Its easy cultivation on simple defined media permits extensive genetic and molecular analysis.

Spores of each mating type produce a monokaryotic mycelium. The fusion of compatible hyphae establishes a dikaryotic mycelium that forms a fruiting body — a miniature mushroom with three distinct tissues (gill, stalk, and cap). The primordium grows and the basidial cells of the hymenial layers in the gills initiate and complete nuclear fusion, meiosis, and haploid basidiospore formation. Spore maturation and active discharge from the cap are accompanied by elongation of the stalk and autodigestion of the remaining gill tissue.

GENOME FACTS: The haploid genome size of *C. cinereus* is estimated at 37.5 Mb. The 13 chromosomes range in size from 1–5 Mb. Over 100 markers affecting mating type, hyphal growth, fruit body morphogenesis, and DNA repair have been assigned to 10 linkage groups. *Coprinus cinereus* is easily transformed by DNA, and targeted gene disruption is achieved in some instances. Both RFLP and RAPD markers are widely utilized. Duplicated genes frequently become methylated at CpG dinucleotides. Chromosome-specific cosmid libraries have been prepared for 8 of the 13 chromosomes. Both commercially prepared cDNA libraries and a cDNA library suitable for yeast two-hybrid screens are available. A partial contig map of the smallest chromosome (~1 Mb) has been assembled and its sequence is currently being determined.

COMMUNITY: Currently, 12 laboratories primarily utilize *C. cinereus*, and several dozen more utilize *C. cinereus* for comparative studies. Approximately 100 sequences have been deposited in GenBank, and a limited number of ESTs (300) have been analyzed. The first sequence of a mushroom will be of particular interest to evolutionary biologists as well as to those interested in the large commercial markets for many mushroom species. Patricia J. Pukkila at the University of North Carolina will provide haploid DNA.

7.13. ***Batrachochytrium dendrobatidis***

SIGNIFICANCE: As a representative of the chytrids, the sequence of *Batrachochytrium dendrobatidis* will be our first window into this largely uncharacterized class of fungi. *Batrachochytrium dendrobatidis* is a monocentric chytrid firmly within the order Chytridiales, based on ultrastructural characteristics and molecular sequence data. Chytridiales is the largest (about 80 genera and 500 species) of the five chytrid orders.

As members of terrestrial and aquatic microbial communities, chytrids are parasites and saprobes of many microscopic organisms (e.g., pollen, algae, and invertebrates) and play an important ecological role in the degradation of recalcitrant materials. Some species are parasites of higher plants, the most notable of which is *Synchytrium endobioticum*. The finding of this potato pathogen in Nova Scotia recently caused a trade dispute between the United States and Canada.

Batrachochytrium dendrobatidis is an amphibian parasite that is thought to be a primary causative agent of the global amphibian decline. Researchers have found that this fungus invades the top layers of amphibian skin cells and causes thickening of the keratinized layer. Because these species drink and breathe through their skin, the fungus may kill them by disrupting these mechanisms. Alternatively, the fungus may be secreting a toxin.

GENERAL DESCRIPTION: *Batrachochytrium dendrobatidis* is readily grown in axenic culture. No resting spore or sexual stage is known. There are dozens of cultures of different strains of *B. dendrobatidis*, including the type strain, available worldwide.

GENOME FACTS: Relatively little is known about the genome of *B. dendrobatidis*. The size is estimated to be between 35–40 Mb from pulsed field separation of chromosomes. It is likely that the fungus is diploid and that the genome size is one diploid equivalent. It is estimated that there are 20 chromosomes. The G+C content is typical for fungi and estimated to be approximately 40%, based on the sequences of 34 randomly cloned DNA regions. Initial sample sequencing will be undertaken to confirm the genome size.

COMMUNITY: Researchers associated with the NSF-sponsored IRCEB project concerning the role of disease in amphibian decline are attempting to determine the population genetics of *B. dendrobatidis* and will be eager to work with the sequence. Others seeking the basis of pathogenicity may benefit as well. Researchers, including trainees of the NSF-sponsored PEET project concerning the phylogeny and systematics of the Chytridiales, will benefit from this project as they attempt to find genes with various mutation rates to use in determining relationships in this earliest lineage of fungi. Haploid DNA for study will be available from Joyce Longcore (University of Maine).

7.14. *Ustilago maydis*

SIGNIFICANCE: *Ustilago maydis* is an important model system for studying pathogen–host interactions. *Ustilago maydis* has been studied for more than 100 years by plant pathologists. Molecular genetic research with *U. maydis* focuses on recombination, the role of mating in pathogenesis, and signaling pathways that influence virulence. Recently, the fungus has emerged as an excellent experimental model for the molecular genetic analysis of phytopathogenesis, particularly in the characterization of infection-specific morphogenesis in response to signals from host plants. *Ustilago maydis* also serves as an important model for other basidiomycete plant pathogens that are more difficult to work with in the laboratory, such as the rust and bunt fungi.

Ustilago maydis is a good representative of the basidiomycetes from an evolutionary perspective, and the genomic sequence would be useful for comparison with fungi from other groups. In this context, genomic analyses would provide an important avenue to understanding the host range of fungal phytopathogens. The analysis of *U. maydis* would provide a framework for studying the hundreds of other Ustilago species that attack important crops, such as barley, wheat, sorghum, and sugarcane. Comparisons would also be possible with other basidiomycete fungi, such as the important human pathogen *C. neoformans*.

Commercially, *U. maydis* is an excellent model for the discovery of antifungal drugs. In addition, maize tumors caused by *U. maydis* are prized in Hispanic cuisine and there is interest in improving commercial production.

GENERAL DESCRIPTION: *Ustilago maydis* is a basidiomycete fungal pathogen of maize and teosinte. The fungus induces tumors on host plants and forms masses of diploid teliospores. These spores germinate and form haploid meiotic products that can be propagated in culture as yeast-like cells. Haploid strains of opposite mating type fuse and form a filamentous, dikaryotic cell type that invades plant tissue to reinitiate infection.

GENOME FACTS: The genome size of *U. maydis* is approximately 20 Mb and many haploid and diploid strains are available. The technology is available to exploit genomic sequence information for genomewide functional studies (e.g., large-scale gene disruption). Cosmid, cDNA, and BAC libraries have been constructed by various groups. The British Columbia Genome Sequence Centre in Vancouver fingerprinted 2000 BAC clones of the related species *U. hordei* as part of a comparative program for Ustilago species. This project also includes plans to fingerprint BAC clones from *U. maydis*, to construct microarrays of random DNA fragments, and to perform serial analysis of gene expression (SAGE). A complete sequence of the *U. maydis* genome is needed, however, for the study of this fungus.

COMMUNITY: *U. maydis* is significant as a model system and the sequencing of its genome is robustly supported by the fungal research community. Jim Kronstand at the University of British Columbia will supply the haploid DNA for sequencing. Approximately 20 research groups (excluding plant pathologists) around the world work on *U. maydis* or related species. Strong collaborative connections exist for some of these groups, and representatives of the Ustilago community met at the most recent Fungal Genetics Conference at Asilomar to discuss coordinated efforts to make use of the genomic sequence. An international conference dedicated to Ustilago research will take place in August 2002. Several private companies have produced partial sequences of this genome. There are no plans to make these data publicly available. The ability to analyze *U. maydis* and maize sequences will be used by a wide scientific community studying pathogen–host interactions.

7.15. ***Paxillus involutus***

SIGNIFICANCE: *Paxillus involutus* is a common ectomycorrhizal fungus that is a mutualist with many northern temperate forest trees. The physiological ecology of *P. involutus* is the best studied among ectomycorrhizal taxa, because it grows fairly rapidly in culture and its mycorrhizae are easily established with tree roots under laboratory conditions. It is commonly used in microcosm experiments in dual culture with birch or pine seedlings. These systems have been used to study carbon metabolism, nitrogen and phosphorous acquisition and transport, and the ability of the organism to scavenge nutrients from complex organic substrates.

GENERAL DESCRIPTION: *Paxillus involutus* is a member of the Boletales, a large order of ectomycorrhizal and saprobic basidiomycetes. The order is evolutionarily significant because switches between saprobic and ectomycorrhizal lifestyles have occurred at least three times within the order, and because it is related to the Euagarics yet is independent from it. *Boletus edulis*, a highly prized edible species, and *Serpula lacrimans*, the dry-rot fungus that destroys European buildings, are other species in the order. *P. involutus* is the only member of the Boletales that is targeted for genomic sequencing.

P. involutus in the broad sense is a complex of closely related cryptic species that include *P. filamentosus, P. vernalis, P. validus, P. obscurosporus, P. albidulus,* and *P. rubicundulus*. Some of these species have specific host or habitat associations. The group thus provides a useful system for studying the evolution of host and ecological specificity.

In nature and in the laboratory, *P. involutus* usually grows as a heterokaryon composed of two different nuclei, each of a different mating type, but axenic haploid strains can be isolated from spores. Completion of the sexual cycle in the laboratory is not possible.

GENOME FACTS: The genome size (40 Mb) is in the range of other basidiomycetes, such as *Schizophyllum commune*, and *C. cinereus*. Currently there is sequence from roughly 100 genes and random genomic fragments deposited in GenBank. Libraries of ESTs, and genomic DNA have been prepared in two labs.

COMMUNITY: As discussed, the sequence will be of broad interest to fungal researchers and to eukaryotic biologists in general. Groups in England, Sweden, France, Germany, and the United States work on the *P. involutus* species. Most past work has been focused on its physiology, but more recently there has been a shift toward molecular genetic work, particularly in England and Sweden. Tom Bruns (University of California, Berkeley) will supply the haploid DNA for this organism.

References

Birrell GW, Giaever G, Chu AM, Davis RW, Brown JM (2001) A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl. Acad. Sci. USA* 98: 12608–12613.

Bennett CB, Lewis LK, Karthikeyan G, Lobachev KS, Jin YH, Sterling JF, Snipe JR, Resnick MA (2001) Genes required for ionizing radiation resistance in yeast. *Nat. Genet.* 29: 426–434.

Dodt G, Braverman N, Valle D, Gould SJ (1996) From expressed sequence tags to peroxisome biogenesis disorder genes. *Ann. NY Acad. Sci.* 804: 516–523.

Foury F (1997) Human genetic diseases: A cross-talk between man and yeast. *Gene* 195: 1–10.

Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 21: 278–283.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. *Science* 274: 563–567.

Ooi SL, Shoemaker DD, Boeke JD (2001) A DNA microarray-based genetic screen for nonhomologous end-joining mutants in *Saccharomyces cerevisiae*. *Science* 294: 2552–2556.

Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, Davis RW, Esposito RE (2000) The core meiotic transcriptome in budding yeasts. *Nat. Genet.* 26: 415–423.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Davis RW, et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901–906.

Zeng Q, Morales AJ, Cottarel G (2001) Fungi and humans: Closer than you think. *Trends. Genet.* 17: 682–684.

Appendix A – Agenda for Fungal Genomics Workshop Nov 6, 2001

Time	Current and Desired Resources	Speaker	Subject
8:15	Resource development for the fungal kingdom	Gerry Fink	Introduction and lessons learned from S. cerevisiae
8:30	Resource development for the fungal kingdom	John Taylor	Evolution
8:45	Resource development for the fungal kingdom	Ron Morris	Aspergillus
	Current Resources/ Needs	----------	----------
9:00	1) Medical	Melanie Cushion	P. carinii
9:15	2) Medical	John Perfect	Human pathogens
9:30	3) Medical	Joe Heitman	C. neoformans
9:45	4) Agriculture	Ralph Dean	Resources for agriculturally important fungi
10:00	5) Industry	Todd Milne	Industries' needs
10:15	6) Evolution	Rytas Vilgalys	What we can learn from fungi
10:30	**Break**	----------	----------
10:45	Recent genome experiences	Fred Dietrich	A. gossypii
11:00	Recent genome experiences	Matt Sachs	N. crassa
11:15	Functional genomics: What's possible	Jef Boeke	Yeast
11:30	Functional genomics: What's possible	Anita Sil	Clinical
11:45	Bioinformatics	Laura Robertson	Industries' needs
12:00	Bioinformatics:	Mike Cherry	Whole-genome informatics
12:30	**Lunch**	----------	----------
	Discussion	**Moderator**	----------
1:30	Fungal priorities	John Taylor	----------
2:00	Sequencing: Needs and how to accomplish them	Ron Morris	----------
2:30	Functional genomics: Needs and how to accomplish them	Jeff Boeke	----------
3:00	**Break**	----------	----------
3:20	Bioinformatics: Needs and how to accomplish them	Mary Anne Nelson	----------
3:50	Outreach	Patricia Pukkila	----------
4:00	Open forum discussion	Gerry Fink Eric Lander	----------
4:50	**Closing**	Ron Morris	----------

Appendix B – Invitee List: Fungal Genomics Workshop Nov 6, 2001

Steering Committee	Affiliation
Dean, Ralph	North Carolina State Univ.
Fink, Gerry	Whitehead Institute
Heitman, Joseph	Duke Univ. Medical Center
Morris, Ron	UMDNJ-Robert Wood Johnson Medical School
Nelson, Mary Anne	Univ. of New Mexico
Taylor, John	Univ. of California, Berkeley

Invitees	Affiliation	Invitees	Affiliation
Barrett, Rob	Whitehead Institute	Kane, Matt	NSF
Birren, Bruce	Whitehead Institute	Kinsey, Jak	FGSC, KU Med. Ctr.
Boeke, Jef	Johns Hopkins Univ. Sch. of Med.	Kistler, Corby	Univ. of Minnesota
Calvo, Sarah	Whitehead Institute	Klausner, Richard	NCI
Cherry, Mike	Stanford Univ. Sch. of Med.	Kwon-Chung, June	NIAID
Church, Deanna	NCBI	Lander, Eric	Whitehead Institute
Cole, Gary	Medical College of Ohio	Lappartient, Anne	AventisCropScience
Cooke Jr., Charles L.	DTRA	Latgé, Jean-Paul	Institut Pasteur
Cormack, Brendan	Johns Hopkins Medical School	Levis, Caroline	INRA Versailles
Cushion, Melanie	Univ. of Cincinnati Coll. of Med.	Lodge, Jennifer	Saint Louis Univ.
Delgado, Nelson	Medical College of Ohio	Loftus, Brendan J.	TIGR
d'Enfert, Christophe	Institut Pasteur	Milne, Todd	Microbia, Inc.
Dietrich, Fred S	Duke Univ. Med. Ctr.	Natvig, Donald	Univ. of New Mexico
Dixon, Dennis M.	NIAID	Nierman, William	TIGR
Drell, Daniel	DOE	Nusbaum, Chad	Whitehead Institute
Dujon, Bernard	Institut Pasteur	Orbach, Marc	Univ. of Arizona
Duncan, Rory	NIAID	Perfect, John R.	Duke Univ. Med. Ctr.
Dunlap, Jay C.	Dartmouth College	Pukkila, Patricia J.	Univ. of North Carolina
Ebbole, Daniel	Texas A&M Univ.	Ralph Dean	N. Carolina State Univ.
Eisen, Jonathan A.	TIGR	Reese, David	EPA
Ellis, Leland	USDA	Ridley, Susan Porter	NSF
Eversole, Kellye	Eversole Associates	Robertson, Laura	Proteome, Div. of Incyte Genomics
Farman, Mark	Univ. of Kentucky	Sachs, Matthew	Oregon Health & Science Univ.
Feingold, Elise	NHGRI	Sherman, David	Lab. Bordelais de Rech. en Informatique
Feldblyum, Tamara	TIGR	Shuster, Jeffrey	Paradigm Genetics, Inc.
Fitzhugh, Will	Whitehead Institute	Sil, Anita	Univ. of California, SF
Galagan, James	Whitehead Institute	Smulian, George	Univ. of Cincinnati Coll. of Med.
Gerry Fink	Whitehead Institute	Strausberg, Bob	NCI
Gill, Steven	TIGR	Sweigard, Jim	Dupont
Giovanni, Maria	NIAID	Tompkins, Laurie	NIGMS
Gold, Scott	Univ. of Georgia	Tornow, Joanne	NSF
Greenberg, Judith	NIGMS	Turgeon, Gillian	Syngenta
Guyer, Mark	NHGRI	Vilgalys, Rytas	Duke Univ. Med. Ctr.
Hecht, Peter	Microbia, Inc.	White, Ted	Univ. of Washington
Henkart, Maryanna	NSF	Xu, Jin-Rong	Purdue Univ.
Hung, Chiung-yu	Medical College of Ohio		