# Fungal Genome Initiative

## A White Paper for Fungal Comparative Genomics

**June 10, 2003**

**Submitted by The Fungal Genome Initiative Steering Committee**
**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

The goal of the Fungal Genome Initiative is to provide the sequence of key organisms across the fungal kingdom and thereby lay the foundation for work in medicine, agriculture, and industry. The fungal and genomics communities have worked together for over 2 years to choose the most informative organisms to sequence from the more than 1.5 million species that comprise this kingdom. The February 2002 white paper identified an initial group of 15 fungi. These fungi 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 spanned by the fungal evolutionary tree.

The Fungal Genome Initiative (FGI) has garnered attention from a broad group of scientists through presentations at meetings, publications, and the release of its first genome sequences. The biological community's interest in the project has grown steadily, resulting in nearly 100 nominations of organisms to be sequenced. Simultaneously, the methods and strategies for effective comparative studies have been clarified by recent whole-genome comparisons of yeasts. Recognizing the power of these comparative approaches, the FGI Steering Committee has identified a coherent set of 44 new fungi as immediate targets for sequencing with an emphasis on clusters of related species.

In this white paper, we propose to sequence additional fungi that include well-studied models important to human health and welfare from poorly understood regions of the fungal kingdom. The data obtained from these fungi will support comparative analyses of other fungi that are also critical to human health and that serve as research models. The FGI will thus propel research in medicine, industry, and agriculture, and will spur progress in computational studies of eukaryotic biology and evolution though the rapid release of sequence from clusters of genomes that are closely related to the most important fungi in research and medicine.

## 2. History and Promise of Fungal Genomics

**2.1 Impact of the yeast sequence.** The sequence of the genome of *Saccharomyces cerevisiae* was a landmark in genomics (Goffeau et al. 1996). 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; Ideker et al. 2001; Ooi et al. 2001; Lee et al. 2002) and 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; Jorgensen et al. 2002; Segal et al. 2003).

**2.2 Other fungal sequencing.** Since the sequencing of *S. cerevisiae*, progress on other fungal genomes has been limited. The sequence for the fission yeast *Schizosaccharomyces pombe* was published only in 2002 (Wood et al. 2002) and the sequence for the first filamentous fungus, *Neurospora crassa*, was published this year (Galagan et al. 2003). Although several other fungal projects are underway outside of the FGI, to date most have yielded only fragmentary sequence and most of these genomes are not freely available. In fact, the freely available genome sequences generated through the FGI represent the largest and most complete set of fungal genome sequences yet produced.

The paucity of fungal genome sequencing is remarkable considering the exceptional contributions that fungal genome sequences can provide to the study of eukaryotic biology and human medicine. Fungal genomes are relatively modest in size (7–40 Mb) and contain few repeats. They are thus ideal targets for whole-genome shotgun (WGS) sequencing. Further, 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 3.7 kb (Galagan et al. 2003).

Within fungal genomes lies the evolutionary history of the origins of many important biological processes found in higher eukaryotes, and 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 make understanding the history and function of fungal genes and genomes of vital interest to human biology.

**2.3 Insights from the sequence of *N. crassa*.** The draft sequence of *N. crassa* underscores the potential of this project (Galagan et al. 2003). The 10,000 predicted genes in this 40-Mb genome correspond to more than twice the number found in the fission yeast *S. pombe* and only about 25% fewer than found in Drosophila. More than 4100 of these predicted proteins lack significant matches to known proteins in the public databases and more than 5800 — a number equal to the entire *S. cerevisiae* genome — lack significant matches to genes in either *S. cerevisiae* or *S. pombe*. These data confirm the early stage of genomic characterization of the filamentous fungi. In addition, when compared with other sequenced eukaryotes, a full 1421 predicted *N. crassa* proteins show their best match to proteins in either plants or animals. Of these, 584 lack high-scoring matches to genes in either sequenced yeast. These genes point to aspects of biology that are shared between filamentous fungi and higher eukaryotes that in many cases are not found within the yeasts.

Despite the fact that Neurospora is the most extensively studied filamentous fungus, the genome sequence revealed some important surprises. Among those revealed by comparative genomics was the presence of genes putatively involved in secondary metabolism. Secondary metabolism, the synthesis of small bioactive molecules, is a key aspect of the biology of filamentous fungi, producing well-known antibiotics and toxins. With the exception of carotenoid and melanin pigment synthesis, Neurospora was not known to possess secondary metabolism. Nonetheless, a number of non-ribosomal polypeptide synthetases and polyketide synthases were identified. The production of secondary metabolites may prove to be an important aspect of the toxicity of fungal pathogens, and represent a potential new diagnostic opportunity.

Although not a pathogen, the *N. crassa* genome revealed a number of genes similar to genes required for pathogenesis in fungal plant pathogens. In several cases, the only known functions for many of these genes within the pathogens was their requirement for pathogenicity. Hence, the value of obtaining genome sequences from model fungi for understanding pathogenesis is extremely high, both because of the intrinsic value the model fungi have as tractable experimental systems and because of the additional power obtained from comparative genome analyses.

## 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 set of fungi (rather than choosing individual fungi in isolation) that maximizes the overall value through a comparative approach. In this way the sequence of well-chosen organisms will not only enable ongoing research on that organism but will enhance the value of other sequences through comparative studies of the evolution of genes, chromosomes, regulatory and biochemical pathways and pathogenesis.

In this white paper, we describe a collection of fungi that have been selected to further expand our understanding of the fungal kingdom and to provide much deeper insight into several extremely important fungal groups. The comparative genomic approach we describe will accelerate the pace of discovery for sorely needed diagnostic tools and therapeutic agents. We begin by outlining the general considerations that justify a program for fungal sequencing.

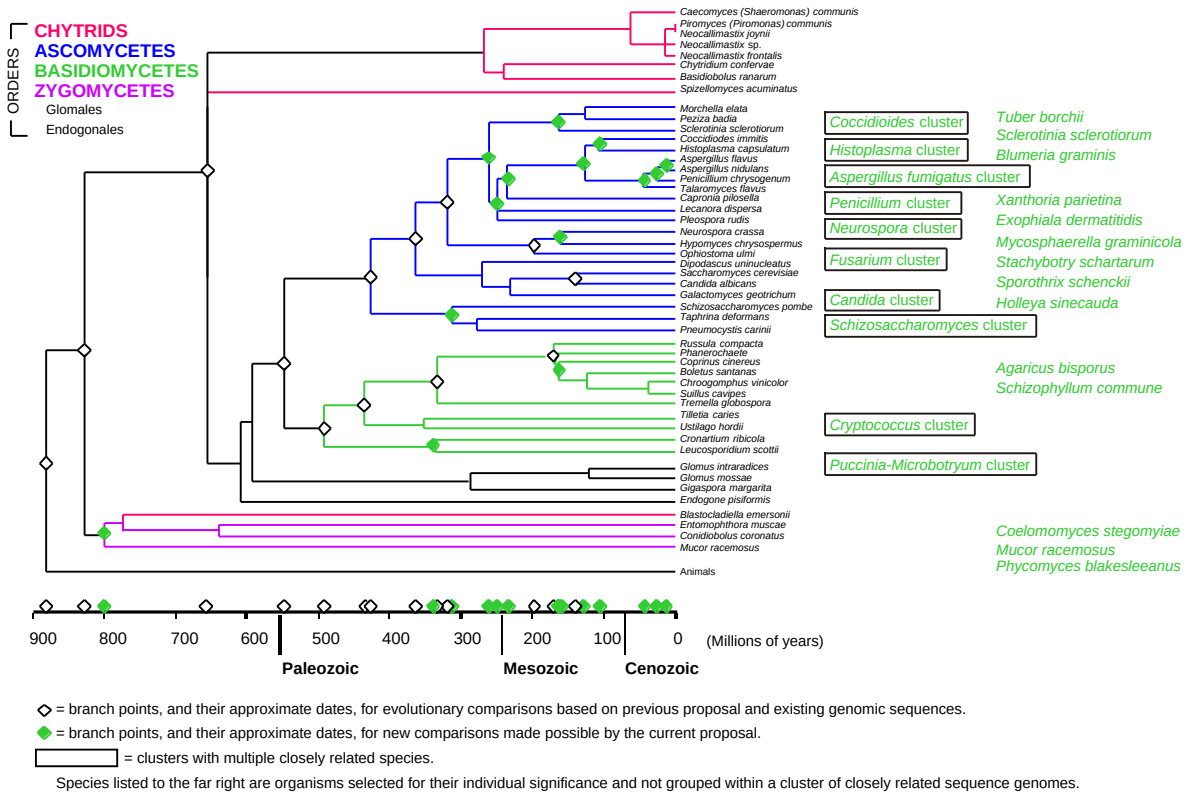


**FIGURE 1. Phylogenetic tree.**

**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. Fungal disease now represents as much as 15% of all hospital-acquired infections. Emerging fungal infections represent an equally serious threat to healthy human populations, including severe allergic reactions to fungal spores and molds. Identifying effective therapies against these eukaryotes has been more difficult than for bacteria, and, as a result, few effective antifungals are currently available. The worldwide market for antifungals is projected to reach $6.5 billion by 2008. 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.

One of the greatest needs clinically is the availability of diagnostics that can provide facile and accurate identification of particular fungal species. Genome sequences provide the opportunity for unique DNA probes that could be used for identification. Our work on Neurospora suggests that fungi may also have pathways yielding secondary metabolites whose presence in the urine or blood would be indelible signatures of the organism causing the infection.

In addition to their role as pathogens, the importance of fungi to human health includes their production of a vast array of secondary metabolites, including toxins and carcinogens that destroy human and animal foodstuffs. The ability to produce and secrete these metabolites underlies their role in the development and production of critical pharmaceuticals with billions of dollars in annual sales, including

antibiotics such as penicillin and the cephalosporins, as well as the cholesterol-lowering statins and cyclosporin.

Fungi also exert a heavy influence on agriculture and our ability to feed the world's population. Fungal 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, crop losses due to fungi exceed $200 billion annually. Genome sequence will be paramount for understanding fungal infection as well as host/pathogen interactions. Sequence data will also provide crucial information on how these organisms reproduce, persist in the environment and interact with their hosts. In addition to their role as pathogens, fungi have additional vital but poorly understood positive roles in agriculture. For example, the mycorrhizal fungi that grow interdependently with plant roots are critical for nutrient uptake by plants.

**3.2 Impact on human biology.** Although *S. cerevisiae* provides key insights into the function of many human proteins, analysis of filamentous fungi reveal a much larger set of proteins shared with humans (Zeng et al. 2001; Galagan et al. 2003). A deeper sampling of fungal genes will rapidly increase the number of human proteins for which we can access homologues in model organisms. Genome sequence from many diverse fungi, coupled with comparative and functional genomics, will advance our understanding of the eukaryotic proteome. In doing so, we will learn not only how to manipulate fungi, but also how 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 mammalian comparative analyses consist of pair-wise sequence alignment of regions. Only once has this involved complete genomes (Mouse Genome Sequencing Consortium 2002). Recently, alignment of a single region from multiple genomes has provoked great excitement about the ability to do this on a genome-wide basis (Boffelli et al. 2003). While of great value, these studies have been largely defined by the limited availability of mammalian genome sequence. A fungal sequencing program represents an ideal system for developing comparative methods for eukaryotic genomes 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 less than 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.

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.

**3.4 Lessons from yeast comparative studies.** Recent work using budding yeast illustrates the power of whole-genome comparative analysis for studying genome evolution, gene identification and gene regulation (Cliften et al. 2003; Kellis et al. 2003). For example, the alignment of high-quality assemblies of four closely related yeasts — *S. cerevisiae*, *S. paradoxus*, *S. bayanus*, and *S. mikatae* — revealed all large-scale chromosomal rearrangements between the species. In addition, the conservation of gene sequences during evolution permits real genes to be distinguished from random open reading frames (ORFs). By these means, ~500 previously annotated yeast ORFs were suggested to be deleted from the gene catalogue, an additional 188 small genes were detected, and the boundaries of more than 300 genes were revised. On the one hand, it is sobering that these many corrections were recognized for the heavily

studied *S. cerevisiae*. One the other hand, these data reveal the power of evolutionary comparisons to recognize and correct genes, a technique applicable to other fungi.

Finally, the yeast comparisons provided the ability to detect the small signals of genetic regulatory elements above the noise of surrounding non-conserved sequence, using computational methods alone. Without prior knowledge of the function of individual genes or factors, virtually all known regulatory motifs were discovered. These methods will be of great value to the study of filamentous fungi, for which there is presently very little information about the components of transcriptional and post-transcriptional regulation.

These studies suggest a clear strategy for further fungal sequencing, namely to sequence close relatives of the most important fungi and use multiple alignments of whole genome assemblies to identify functionally conserved elements and investigate genome evolution. Because the genomes selected for comparative sequencing are also used in laboratory studies, research on these organisms to will be dramatically accelerated. The prime motivation is to use comparative genomics to illuminate to the greatest degree possible the biology and genetics of organisms of proven importance.

The species of yeast used in the recent comparative analyses of yeast genomes (Kellis et al. 2003; Cliften et al. 2003) were specifically chosen based on their evolutionary distance as being maximally informative for alignment and identification of functional elements. The FGI Steering Committee has used the extensive resources of the community to identify fungi for genome sequencing that are appropriately close to important fungi for which there is also genome sequence. This has been necessary because the existing sequence data do not support effective comparisons. In fact, at present, the only filamentous fungi for which multiple genome alignments can be attempted are for *Aspergillus fumigatus*, *A. nidulans*, and *A. oryzae*. However, these species are more distantly related than would be ideal for comparative analysis.

## 4. History of the FGI

**4.1 Origins.** In November 2000, Dr. Gerry Fink 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.

The group concluded that the dearth of publicly available fungal genome sequence was a major barrier to biomedical research. 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.

These principles were laid out in a draft white paper that described a broad, collaborative Fungal Genome Initiative. This white paper was circulated during the summer of 2001 among federal agencies and served as the direct inspiration for NHGRI's process for prioritizing organisms for sequencing.

**4.2 Steering Committee.** To provide advice and oversight of a sustained effort, a Steering Committee was organized, consisting of:

Gerry Fink, Whitehead Institute for Biomedical Research, Steering Committee, Chair
Ralph Dean, North Carolina State University; Fungal Genetics Policy Committee, Chair
Peter Hecht, Microbia, Inc.
Joe Heitman, Duke University
Ron Morris, UMDNJ-Robert Wood Johnson Medical School
Matthew Sachs, Oregon Health and Science University
John Taylor, University of California, Berkeley
Mary Anne Nelson, University of New Mexico
Bruce Birren, Whitehead Institute for Biomedical Research

These fungal biologists represent a cross-section of interests in mycology and fungal genetics and are responsible for setting the direction of the project and reviewing progress. The Steering Committee meets annually, having met most recently in March 2003 at the International Fungal Genetics Conference at Asilomar, California. Throughout the year, the Steering Committee interacts through e-mail and telephone conference calls. In addition, regular meetings take place between the Committee Chair and the Whitehead Institute/MIT Center for Genome Research (WICGR) staff.

**4.3 Ongoing community input**. In November 2001, the Steering Committee convened an NHGRI and NSF sponsored workshop on Fungal Genomics in Alexandria, Virginia. Over 60 attendees representing academic, government and industrial interests in medical, agricultural, industrial, evolutionary, basic biological fungal research and informatics discussed the genome resources that were most needed to spur research and development in their areas of interest. The workshop produced strong endorsement of an initiative that would include rapid sequencing and public release of many fungal genomes chosen as part of a single, coherent plan.

One outcome of the workshop was formalization of the communication channels between the FGI Steering Committee and the broader research community. Presentations about the FGI at major fungal conferences keep the community apprised of progress and always include the solicitation of community input. Notice of new data releases are sent through our own mailing list and that of the Fungal Genetics Stock Center. New nominations of candidate fungi arrive weekly along with other e-mail correspondence (FGI_Info@genome.wi.mit.edu). To date, nearly 100 nominations of organisms for sequencing have been considered by the Steering Committee. Of particular importance in evaluating these nominations has been the FGI Steering Committee's connection with other groups interested in fungal phylogeny and pathogenesis. Specifically, the input from scientists associated with the NSF-funded Fungal Tree of Life project (Dr. J. Spatafora, PI) has been valuable in identifying fungi of the appropriate evolutionary distance for our purposes, and our contacts with the American Phytopathological Society have provided a great deal of useful information.

**4.4 FGI website, data release and user community**. Since the February 2001 release of the *N. crassa* genome sequence, WICGR has maintained a growing set of resources for fungal genomics on the web. The FGI website (www-genome.wi.mit.edu/annotation/fungi/fgi/) lists the status of all FGI projects and provides forms for submitting sequencing candidates. Currently, genome assemblies are available through the FGI website for five filamentous fungi and three yeasts. These fungal databases have received over 2.9 million hits since their launch and currently average 440,000 hits/month. This level of use demonstrates the wide appeal of these data, including scientists engaged in comparative genomics, evolutionary studies, fungal biology, infectious disease and computational biology.

**4.5 WICGR fungal collaborations.** WICGR has forged many strong collaborative relationships involving fungal genome sequencing and analysis. In fact, each fungal sequencing project represents a successful collaboration between WICGR and a research community of varying size. In the case of *N. crassa*, WICGR organized the community analysis project that engaged over 70 scientists in analysis of the genome sequence. In other cases, WICGR is helping to coordinate comparative analyses for multiple genome sequences, such as with *Aspergillus fumigatus*, *A. oryzae*, and *A. nidulans*. A similar collaboration exists with The Institute for Genomic Research (TIGR) and a consortium of university-based labs to jointly analyze genome sequence from two different Cryptococcus species. In several instances, WICGR has built alliances directly responsible for public release of data previously held in private hands, such as with Monsanto, Bayer, and Exelixis.

## 5. Sequencing Progress

The WICGR has fully released 8 fungal genome assemblies, including 5 from the 15 on our first white paper (Table 1). Funds to sequence 2 of the original 15 genomes, *Magnaporthe grisea* and *Fusarium graminearum*, were provided by awards outside of NHGRI. Two more assemblies are in "pre-release", currently available on our website to collaborators for quality review pending full public release later this month. Sequence data for all FGI genomes are made available in advance of assembly according to NHGRI policy on rapid data release by regular deposition of traces at the NCBI trace repository. We are on schedule to release seven high-priority fungi as per our plan provided to NHGRI. Difficulty obtaining DNA samples of sufficient quality for Fosmid cloning required us to revise the exact order of genomes to be sequenced, highlighting our need and ability to remain flexible when working from a list of targets.

**Table 1. WICGR Fungal Genome Releases**

| Species | Status | Predicted genome size (Mb) | Assembled bases (Mb) | N50 scaffold size (Mb) | Accession # |
|---|---|---|---|---|---|
| ***Neurospora crassa*** | annotated assembly released | 40 | 38,044,343 | 0.61 | AABX01000000 |
| ***Magnaporthe grisea*** | annotated assembly released | 40 | 37,878,070 | 1.6 | trace repository |
| ***Aspergillus nidulans*** | assembly released | 31 | 30,068,514 | 2.44 | AACD01000000 |
| ***Fusarium graminearum*** | assembly released | 40 | 36,093,143 | 5.36 | AACM01000000 |
| ***Cryptococcus neoformans***, serotype A | assembly released | 20 | 19,223,796 | 1.3 | AACO01000000 |
| ***Ustilago maydis*** | assembly in pre-release | 20 | 19,762,689 | 0.82 | trace repository |
| ***Coprinus cinereus*** | assembly in pre-release | 38 | 36,259,524 | 2.06 | trace repository |
| ***Coccidiodes immitis*** | in sequencing | 29 | — | — | — |
| ***Rhizopus arrhizus*** | in sequencing | 35 | — | — | trace repository |
| ***Saccharomyces paradoxus*** | assembly released | 12 | 11,570,000 | 0.509 | AABY00000000 |
| ***Saccharomyces mikatae*** | annotated assembly released | 12 | 11,220,000 | 0.334 | AABZ00000001 |
| ***Saccharomyces bayanus*** | annotated assembly released | 12 | 11,320,000 | 0.234 | AACA00000002 |

—, not applicable.

## 6. Choice of Organisms

The Steering Committee, with considerable input from the fungal research community, has identified a coherent and complementary collection of 44 additional fungi for sequencing (the complete list of nominations received to date are provided in the Appendix). At present, no applications to support sequencing any of these 44 new fungi are pending or contemplated. In the final portion of this white paper, we discuss the specific rationale for the choice of individual fungi in terms of biological interest, genome size, availability of haploid strains, community, and evolutionary diversity.

In response to our initial white paper, the NHGRI review committee designated seven fungi as high priority from the list of 15 candidates. Once again we look to this committee and NHGRI staff for guidance concerning the appropriate time line for sequencing the complete list.

Several of these fungi have been selected as well-studied laboratory models from poorly characterized regions of the fungal tree. However, given the proven impact of comparative analysis, the majority of these new fungal candidates have been selected to create clusters of sequenced genomes for species closely related to key fungal pathogens and research models. The points of nucleation for these comparative studies were chosen by the community and represent not only the most important fungi with respect to human health, but also those fungi that account for the vast majority of government and private spending on basic research and development of diagnostics and antifungals. The fungi we will develop clusters of genome sequences around are: *Candida, Aspergillus, Cryptococcus, Coccidioides, Penicillium, Fusarium, Histoplasma, Neurospora, and Schizosaccharomyces*. The single biggest change between the organisms described in the previous white paper and this one is the new emphasis on *Candida*. This reflects both the overwhelming impact of *Candida* on the U.S. healthcare system (roughly 75% of all superficial and systemic fungal infections and more than 85% of hospital-acquired fungal infections are caused by *Candida*) and the lagging development of suitable genomic resources. Overall, the value of these comparative data in enabling research on these organisms that are already central to fungal genetics and medical mycology will be tremendous.

Figure 2 displays the clusters of fungal species proposed by the FGI. The tree topology only represents the classification relationships among the selected taxa. The tree branches do not reflect the evolutionary divergence.

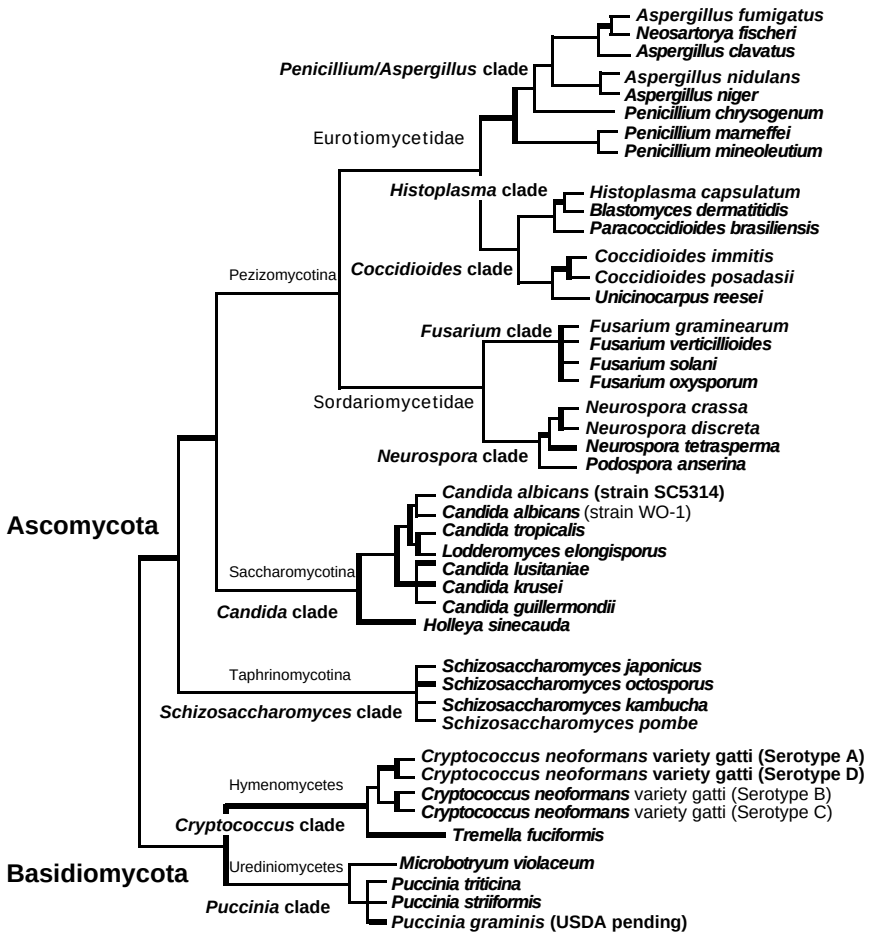


**FIGURE 2. Cluster tree.**

## 7. Sequencing Approach

**7.1 Deep-shotgun sequencing**. For each fungus, we propose to generate a high-quality draft sequence. Specifically, we will produce assemblies representing 8X whole-genome shotgun (WGS) sequence from paired-end reads obtained from 4-kb plasmids (80%), 10-kb plasmids (10%) and 40-kb Fosmids (10%). All libraries will be prepared from randomly sheared genomic DNA. Our experience with shotgun sequencing of fungal genomes indicates that 8X sequence coverage produces a high-quality draft assembly with the vast majority of each genome (well over 96%) present in the assembly. Further, these assemblies achieve long-range continuity as a result of the links provided by the Fosmid end sequences and the combined physical coverage of all libraries, which is approximately 50X. Although fungal genomes vary considerably in structure and nucleotide composition, a typical 8X assembly yields N50 contig sizes of 30–110 kb, and the N50 scaffold size of ~2 Mb. The high quality of this draft sequence is sufficient for most of the comparative studies this project will support. For example, our yeast comparative studies employed roughly this coverage (Kellis et al. 2003).

**7.2 Polishing and finishing.** For genomes that serve as particularly important references, it may be valuable to further improve the quality of the assembly. We propose that all clones required for automated polishing and/or finishing be retained and that the decision to carry out subsequent polishing and/or targeted finishing work be prioritized by NHGRI staff on the basis of evolving assessment of cost and capacity. One important feature of these high-quality draft assemblies is that almost all gaps and low-quality regions are small and are spanned by Fosmid subclones. This provides the opportunity for rapid and efficient improvement in the quality of the assembly. Fosmids that span gaps or low-quality regions can be automatically identified and used as templates for highly automated directed sequencing.

## 8. Detailed Description of Organisms

The organisms, the rationale for sequencing, and relevant genome information are summarized in Table 2 and are described in more detail on the following pages.

**Table 2. Organism Summaries**

| | Name | Significance | Est. size (Mb) |
|---|---|---|---|
| **MEDICINE** | | | |
| ***Candida*** cluster | ***Candida albicans*** (strain WO-1) | Most common human pathogen, related to laboratory strain being sequenced (SC5314) | 16 |
| | ***Candida tropicalis*** | The second most pathogenic of the ***Candida*** species; extremely close relative to ***C. albicans*** to be used for genomic comparison | 16 |
| | ***Lodderomyces elongisporus*** | Closest sexual relative to ***C. albicans*** and source of haploid genome for comparison to ***C. albicans*** | 16 |
| | ***Candida lusitaniae*** | Haploid relative of ***C. albicans*** with different codon usage and known complete sexual/meiotic cycles | 16 |
| | ***Candida krusei*** | Haploid relative of ***C. albicans*** with different codon usage and known complete sexual/meiotic cycles | 16 |
| | ***Candida guillermondii*** | Haploid relative of ***C. albicans*** with different codon usage and known complete sexual/meiotic cycles | 16 |
| ***Aspergillus fumigatus*** cluster | ***Neosartorya fischeri*** | Sexual ***Aspergillus*** species and closest relative to ***A. fumigatus***, the #2 U.S. health problem. Also a causative agent of aspergillosis | 25–30 |
| | ***Aspergillus clavatus*** | Close relative to ***A. fumigatus***, the second most common fungal pathogen in the U.S. | 25–30 |
| ***Fusarium*** cluster | ***Fusarium verticillioides*** | Pathogenic fungus that infects immunosuppressed patients | 46 |
| | ***Fusarium solani*** | Deadly threat to immunosuppressed, especially neutropenic and transplant patients | 40 |
| | ***Fusarium oxysporum*** | Pathogenic fungus that infects immunosuppressed patients | 33 |
| ***Histoplasma*** cluster | ***Paracoccidioides brasiliensis*** | Most prevalent systemic mycose in Latin America | 25 |
| | ***Blastomyces dermatitidis*** | Causative agent of blastomycosis, the principle systemic mycoses | 28 |
| ***Coccidioides*** cluster | ***Unicinocarpus reesei*** | Closest known relative of ***Coccidioides*** species, most severe of the U.S. systemic mycoses and select agents. | 30 |
| ***Penicillium*** cluster | ***Penicillium marneffei*** | Only dimorphic ***Penicillium***; cause of grave pneumonia in AIDS patients | 22–33 |
| | ***Penicillium mineoleutium*** | Close relative ***Penicillium*** marneffei | 30 |
| | ***Stachybotrys chartarum*** | Black mold, major indoor environmental threat | 40 |
| | ***Sporothrix schenckii*** | Pathogenic dimorphic fungus with worldwide distribution | 40 |
| | ***Exophiala (Wangiella) dermatitidis*** | Causative agent of human dermatomycoses; excellent model for other dematiaceous fungal pathogens | 19 |
| ***Cryptococcus*** cluster | ***Cryptococcus neoformans variety gattii*** (Serotype B) | Encapsulated basidiomycete; leading cause of infectious meningitis | 20 |
| | ***Cryptococcus neoformans variety gattii*** (Serotype C) | Encapsulated basidiomycete; leading cause of infectious meningitis | 20 |
| | ***Tremella fuciformis*** | Close relative of ***Cryptococcus***; well-developed sexual fruiting body to compare with ***Cryptococcus***; obligate mycoparasite on wood rot fungi | 20 |

| | | | |
|---|---|---|---|
| **COMMERCE** | | | |
| | ***Penicillium chrysogenum*** | Primary penicillin source; reveals genomic effects of mutagenesis and selection | 34 |
| | ***Aspergillus niger*** | Widely used for the industrial production of enzymes and metabolites | 36 |
| | ***Agaricus bisporus*** | Most widely cultivated mushroom, annual worldwide production valued at $4.5 billion | 40 |
| | ***Tuber borchii*** | Edible truffle; ectomycorrhizal fungus, with experimental plantation | 34 |
| **EVOLUTION / FUNGAL DIVERSITY** | | | |
| | ***Saccharomyces cerevisiae*** RM11-1a | Natural isolate, now used in laboratory studies | 12 |
| ***Neurospora*** cluster | ***Neurospora tetrasperma*** | Pseudohomothallic species, diverged from ***N. crassa*** about 2.5 million years ago | 40 |
| | ***Podospora anserina*** | Model filamentous fungus, supports comparative studies of ***Neurospora*** | 34 |
| ***Schizosaccharo-myces*** cluster | ***Schizosaccharomyces japonicus*** | Supports comparative analysis of model yeast, ***S. pombe*** | 14 |
| | ***Schizosaccharomyces octosporus*** | Supports comparative analysis of model yeast, ***S. pombe*** | 14 |
| | ***Schizosaccharomyces kambucha*** | Supports comparative analysis of model yeast, ***S. pombe*** | 14 |
| | ***Schizophyllum commune*** | Major model for mushroom-forming fungi | 38 |
| | ***Phycomyces blakesleeanus*** | Model filamentous zygomycete | 30 |
| | ***Mucor racemosus*** | Model for dimorphic growth among poorly understood zygomycetes | 39 |
| | ***Xanthoria parietina*** | Lichen; classic example of a mutualistic symbiotic relationship | 30–40 |
| | ***Coelomomyces stegomyiae*** | Haploid chytrids, the main eukaryotic pathogen of mosquito larvae | 50 |
| **AGRICULTURAL** | | | |
| | ***Mycosphaerella graminicola*** | Cause of septoria tritici leaf blotch, the second most important wheat disease in the U.S. | 32–40 |
| | ***Blumeria graminis*** | Cause of most important leaf disease of barley; most intensively studied powdery mildew species at the molecular level | 35–45 |
| | ***Sclerotinia sclerotiorum*** | Broadest known host-range of any plant pathogen; infects more than 408 species | 26–44 |
| | ***Holleya sinecauda*** | Pathogen on mustard seeds, close relative to ***Saccharomyces*** and ***Candida*** | 8.5 |
| ***Puccinia-Microbotryum*** cluster | ***Microbotryum violaceum*** | Best studied pathogen in natural plant populations; closely related with ***Puccinia*** species | 25–30 |
| | ***Puccinia triticina*** | Wheat leaf rust, the most widespread and common wheat pathogen in the U.S. and worldwide | 90 |
| | ***Puccinia striiformis*** | Basidiomycete fungal pathogen of crops and grasses | 90 |

### 8.1 *Candida* cluster

**SIGNIFICANCE:** *Candida albicans* is the most common human fungal pathogen, likely because it is generally a benign commensal that resides on the mucosal surfaces of most if not all organisms. *C. albicans* is capable of causing superficial infections (vaginitis, thrush) in normal hosts, and severe systemic infections in immunocompromised hosts. *C. albicans* is diploid, but recent discoveries herald a new era in understanding the sexual cycle and its role in virulence. These advances stemmed directly from the genome project and the discovery of the MAT locus.

The genome of one particular strain, SC5314 is nearing completion at Stanford, however assembling the shotgun sequence has proven difficult due to polymorphisms in the diploid genome. We propose to sequence a second *C. albicans* isolate, clinical isolate WO-1, to provide important information about the pathogenic form of *C. albicans*. This isolate differs considerably from the SC5314 strain. It is the most carefully characterized MTL-homozygous strain for both white-opaque switching, a phenotypic change that correlates with host cell specificity and mating. We also propose to sequence two species closely related to *C. albicans*: *C. tropicalis* and *Lodderomyces elongisporus.* The asexual diploid yeast *C. tropicalis* is the second most pathogenic of the *Candida* species. Unlike *C. albicans*, which is a normal commensal on human mucous membranes, the detection of *C. tropicalis* is more often associated with the development of deep fungal infections. *L. elongisporus* is the only known ascosporogenous species in the *C. albicans* clade. It usually produces a single ascospore.

To overcome the difficulty in assembling these polymorphic diploid genomes, we propose to sequence three additional haploid *Candida* species, which have defined sexual cycles and yet lie within the clade of *Candida* species that underwent the CTG capture event, and that have the same as genetic code as *C. albicans*. These species are *C. lusitaniae*, *C. guilliermondii*, and *C. krusei*. In particular, reagents to study *C. lusitaniae* (aka *Clavispora lusitaniae*) are well developed and include molecular genetic techniques to characterize the sexual cycle, a transformation system, gene disruption by homologous recombination, and congenic a and α strains constructed for genetic studies. Also, *C. albicans* has the ability to mate yet appears to lack genes required for meiosis and sporulation in *S. cerevisiae*. We would like to identify genes that are conserved among these three *Candida* species that can mate, undergo meiosis, and sporulate in order to identify the common gene set involved in these processes. Our results may reveal that *C. albicans* posesses homologs of genes required for these processes, which would stimulate efforts to study meiosis and sporulation. Alternatively, it may reveal that a limited number of key genes are missing; these genes could then be introduced to endow *C. albicans* with these key developmental pathways and allow standard genetic methods to be conducted in this system. The comparisons among these *Candida* species, including three haploids, will improve genome assembly and annotation, and will provide insight to genome evolution.

**GENERAL DESCRIPTION:** *Candida* species belong to Ascomycota, Saccharomycotina subphyla and all species can be cultured under laboratory conditions. The species we propose to sequence, with the possible exception of *C. krusei*, are descended from an ancestral strain in which the CTG codon was reconfigured ~150 million years ago. In fact, sequencing the genome of *C. krusei* will enable us to determine if CTG codons encode leucine as in most organisms or serines as in other *Candida* species.

**GENOME FACTS:** The genome size of *Candida* species is around 20 Mb. All the sequences can be readily aligned to the *C. albicans* strain SC5314 assembly (available at www-sequence.stanford.edu/), which will expedite assembly and annotation.

**COMMUNITY:** There is a very large research community, over 200 investigators, working on basic research using *C. albicans*. The genomic data of these closely related species will greatly improve the genome assembly and annotation of *C. albicans*, which will accelerate the understanding of both the pathogenesis of this fungal pathogen and the genome divergence among these species. The genomic DNA of *C. albicans* (strain WO-1) and three haploid *Candida* species will be provided by Dr. David Soll at University of Iowa. *C. tropicalis* and *L. elongisporus* DNA will be provided by Dr. Clete Kurtzman at the USDA in Peoroa.

### 8.2 *Aspergillus fumigatus* cluster

**SIGNIFICANCE:** *Aspergillus fumigatus* is the most frequently encountered opportunistic *Aspergillus* pathogen. *A. fumigatus* has been reported to be the major organism isolated from air samples, presumably because its spores are very small. The human infection caused by *A. fumigatus* is rated as the #2 health problem in the United States. Because of the importance to public health, the *A. fumigatus*genome has been sequenced through joint effort by TIGR, the Sanger Centre, and the Institut Pasteur. Here we propose to sequence two close relatives of *A. fumigatus* — *Neosartorya fischeri* and *A. clavatus* — to facilitate comparative study and improve the genome annotation. *Neosartorya fischeri* is the sexual stage of an *Aspergillus* species, so comparison with the asexual *A. fumigatus* may identify genes required for sexual reproduction. *N. fischeri* is also a causative agent of aspergillosis. It causes pulmonary aspergillosis in transplant patients. *A. clavatus* is found in soils and animal manure. Although this species is only occasionally pathogenic, it can produce petulin, which may be associated with disease in humans and other animals.

**GENERAL DESCRIPTION:** Both *A. clavatus* and *N. fischeri* belong to the family Trichocomaceae. Experience with batch and fed-batch fermentation under controlled small- and large-scale conditions is long standing, whereas continuous culturing has been developed and applied for various physiological and genomics studies. Both organisms produce different cell types during their development.

**GENOME FACTS:** The sizes of the two genomes are ~35 Mb. There is no long-link library available so far.

**COMMUNITY**: Currently, more than 70 laboratories worldwide investigate a number of genetic, molecular-biological, biochemical, physiological and process-engineering aspects of *Aspergillus*. The public sequencing of these genomes, in addition to the already available genomes of *A. fumigatus* and the model fungus *A. nidulans*, will considerably enrich the possibility for comparative and evolutionary genomics. This activity will also stimulate the growing collaboration between the global *Aspergillus* community.

### 8.3 *Fusarium* cluster

**SIGNIFICANCE:** The three *Fusarium* species we propose to sequence have significance for medicine, agriculture, and food safety. Disseminated *Fusarium* infections are emerging as a significant risk factor for immunosuppressed humans, especially neutropenic and transplant patients. Currently available antifungal therapies are largely ineffective against disseminated *Fusarium* infections, which are often fatal. *Fusarium* species also are the leading and most serious cause of the important opthalmologic disease, mycotic keratitis. The three species most commonly associated with human infections are, in order of prevalence, *F. solani*, *F. oxysporum* and *F. verticillioides* (formerly known as *F. moniliforme*). Each fungal species also is an important plant pathogen. *F. oxysporum* causes wilt disease and root rots on over 100 species of cultivated plants worldwide and is perhaps the single most destructive fungal plant pathogen. *F. solani* is an important pathogen of leguminous crops and other hosts, and *F. verticillioides* causes *Fusarium* ear rot of corn throughout the world. *Fusarium* species also may contaminate agricultural products with mycotoxins harmful to both humans and livestock. Corn infected with *F. verticillioides* contains fumonisins, which have been associated with human esophageal cancer and have been demonstrated to cause equine leukoencephalomalacia and pulmonary edema in swine. Fumonisin FB1 is hepatotoxic to all animal species tested and has been shown to be hepatocarcinogenic in mouse. The FDA recently recommended maximum levels for fumonisins in both animal feed and human food products.

A whole-genome sequence assembly was recently obtained for *F. graminearum*. Having comparative *Fusarium* genomes would aid in gene identification for these relatively poorly characterized organisms by providing regions of synteny with orthologous ORFs conserved through evolutionary time. In addition, the assemblies would aid in the understanding of whole-genome evolution and comparative genome organization of *Fusarium* species possessing four chromosomes (the *F. graminearum* complex and related species), 12 chromosomes (the *G. fujikuroi* complex including *F. verticillioides*), and those with variable chromosome numbers (*F solani* and *F. oxysporum*). The assemblies would also would provide information on the origin of supernumerary chromosomes.

**GENERAL DESCRIPTION:** *Fusarium* species are the asexual form of certain ascomycetous fungi in the Hypocreales. *F. oxysporum* is known solely by its asexual stage, but is aligned with fungi having a *Gibberella* sexual state; it is closely related to *F. graminearum* (*G. zeae*). *F. verticillioides,* having a sexual form known as *Gibberella moniliformis*, is less closely related to *G. zeae*, followed by the more distant *Fusarium solani* that has a *Nectria haematococca* sexual state. All are easily grown in culture and readily produce asexual and (except for *F. oxysporum*) sexual spores.

Model host systems for *Fusarium* pathogenicity studies are available for plants (*Arabidopsis*, tomato) and animals (mouse, greater wax moth caterpillar). Forward genetic screens are identifying factors important for *Fusarium* pathogenicity. Because the mutagen is typically an insertional tag (either a transposon or introduced plasmid) a genome sequence would facilitate more rapid identification of the disrupted gene. *Fusarium* genes can be efficiently targeted for disruption. A genome sequence would facilitate reverse genetics — phenotypic analysis of mutated candidate genes.

**GENOME FACTS:** Members of the *F. solani* species complex have both variable karyotypes and variable genome sizes due to the presence of supernumerary chromosomes that were first discovered for filamentous fungi in this organism. Typical strains have a genome size of 40 Mb and 15 chromosomes. *F. oxysporum* strains have highly variable genomes ranging in size from 33 Mb to >50 Mb. Chromosome number varies in *F. oxysporum* due to aneuploidy. Mitotic linkage maps are available. *F. verticillioides* has a genome size of approximately 46 Mb and a well-developed genetic map. It and closely related strains of the *G. fujikuroi* species complex each have 12 chromosomes.

**COMMUNITY:** Owing to its importance as a human and plant pathogen, there is clear interest in *Fusarium* genomics among plant pathologists and medical mycologists. The International Society of Plant Pathology maintains a standing committee on *Fusarium* and an International Fusarium Workshop is held on a regular basis. The first Fusarium Genomics Workshop was held recently at Asilomar, and a second is scheduled for Copenhagen in March 2004. Currently 142 scientists from 28 countries are members of the Fusarium Genomics Consortium.

### 8.4 *Paracoccidioides brasiliensis*

**SIGNIFICANCE:** *Paracoccidioides brasiliensis* is the most prevalent systemic mycose in Latin America. Most infections with *Paracoccidioides* occur in Central and South America, particularly Brazil, Venezuela, and Colombia. It is a granulomatous disease that produces a primary pulmonary infection, often asymptomatic. Inhalation of the conidia followed by infection of the lungs is the presumable mode of infection. The disease has a long latent period and may become apparent only years after exposure. In addition to the primary pulmonary form of the disease, disseminated forms may also be observed. Reticuloendothelial system, skin, mucous membranes and lymph nodes are frequently affected in cases of disseminated disease. A similarity of paracoccidioidomycosis with blastomycosis suggests a relationship of the two etiological agents.

The high positive skin test reactivity in some areas (70%) indicates that around 10 million people are infected in South America. The genome sequence will greatly help our understanding of the fungus pathobiology and provide insights to possible ways of treatment.

**GENERAL DESCRIPTION:** Although originally classified as an imperfect fungus (phylum Deteromycota), studies based on rDNA sequences have placed *P. brasiliensis* in the phylum Ascomycota. *P. brasiliensis* is a thermally dimorphic fungus. After entering the host, the fungus converts to the yeast form. It is a very interesting model in which to study dimorphism since there is a clear relationship between morphogenesis and pathogenicity. The habitat of *P. brasiliensis* is presumably the soil, where it has been isolated as conidia. It has also been isolated from the digestive tract of some animals. The main antigenic component described in *P. brasiliensis* is gp43, an exocellular glycoprotein that is a receptor for murine laminin and may be a virulence factor.

**GENOME FACTS:** *P. brasiliensis* lab strains are diploid, although there is a paper suggesting that some *P. brasiliensis* individuals are haploid; its genome size is ~25 Mb. Chromosomes can be separated by pulsed field gel electrophoresis in five distinct bands. No sexual phase is known. Interestingly, there are extensive similarities between *P. brasiliensis* and *C. albicans* genes. Several groups have extensively worked in the biology and immunolgy of the fungus and two laboratories have constructed cDNA libraries and have begun sequencing ESTs. These labs have also constructed cosmid libraries. Molecular biology tools are being developed, such as transformation systems and gene disruption protocols.

**COMMUNITY:** There is a large research community working with *P. brasiliensis* in Brazil. There are active collaborations betweeen the Latin American researchers and North American labs to study this pathogen. They would greatly benefit from its genome sequence. Paulo S. R. Coelho (Universidade de Sao Paulo at Ribeirao Preto, Brazil) will supply diploid DNA.

### 8.5 *Blastomyces dermatitidis*

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

*B. dermatitidis* is an Ascomycete in a class with four other systemic dimorphic fungi, including *Histoplasma capsulatum*, *Paracoccidioides brasiliensis*, and *Sporothrix schenckii*. *B. dermatitidis* and *H. capsulatum* are the most closely related genetically. In addition, it was reported that yeast-phase specific virulence genes in these fungi are regulated by conserved mechanisms. There is fundamental biological interest and medical significance to elucidate the genomic basis of biological similarities and differences between human fungal pathogens. In a study of the molecular evolution of 10 fungal pathogens, *B. dermatitidis* and *H. capsulatum* were the two closest and differed at only 12 bases within 1713 bases of 18S rDNA sequenced. Washington University is sequencing (3-fold coverage) a pathogenic human isolate of *B. dermatitidis* — ATCC 26199 — used in many published studies, to compare its sequence with that of *H. capsulatum*. Here we propose to produce whole-genome shotgun assemblies for a different *B. dermatitidis* strain and two of its close relatives: *P. brasiliensis* and *S. schenckii* from this important class. These genomic sequences, together with available genomic sequence of *H. capsulatum*, will provide power for the comparative study to better understand pathogenicity and evolution.

**GENERAL DESCRIPTION:** As with *H. capsulatum* (and other systemic fungi such as *Coccidioides*), *B. dermatitidis* is a dimorphic fungus. It grows as a saprophytic mold form in the soil and a pathogenic yeast form in host tissue.

**GENOME FACTS:** *B. dermatitidis* is a haploid fungus with two mating types. Scientists at University of Wisconsin-Madison have developed the tools to genetically manipulate the fungus using molecular approaches. DNA-mediated gene transfer can be done reliably and with high efficiency by electroporation or using Agrobacterium. Multiple dominant selectable markers, marked auxotrophic strains, LacZ and Gus reporters for promoter analysis, and reverse genetic and forward genetic approaches have all been published and are available to the research community.

**COMMUNITY:** Multiple *B. dermatitidis* strains are available at Dr. Bruce Klein's lab at University of Wisconsin-Madison. These strains have well-characterized *in vitro* and *in vivo* models of infection, published and used by many researchers in the fungal community. Dr. Klein's lab will provide the haploid DNA sample for the sequencing project.

### 8.6 *Unicinocarpus reesei* (*Coccidioides* cluster)

**SIGNIFICANCE:** *Unicinocarpus reesii* is the closest known relative of *Coccidioides*, a species that causes serious and sometimes fatal disease in otherwise healthy people. Morphologically, *U. reesei* is similar to *Coccidioides*. It differs from *Coccidioides* in two key aspects: it is not a human pathogen and it mates by out-breeding to produce meiotic progeny. A recent study of the phylogenetics of *U. reesei* showed that it harbors cryptic species that can be recognized by DNA polymorphism. Although not a human pathogen, *U. reesii* is keratinophilic and can be found in the lungs of rodents.

*Coccidioides posadasii* has been sequenced by The Institute for Genomic Research, and *Coccidioides immitis* is being sequencing at the Whitehead Institute/MIT Center for Genome Research. Here, we propose to sequence *U. reesei* in order to provide a comparison to the two *Coccidioides* species as well as for annotating the genomes of all three species. In addition, the comparison of these genomes will identify genes found in both pathogenic *Coccidioides* species but not in *U. reesii;* these are good candidates for genes involved in human pathogenicity.

**GENERAL:** *Unicinocarpus reesii* is an ascomycete soil fungus (Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; Onygenaceae; Uncinocarpus) and is the closest known relative of the *Coccidioides* species. It can be grown readily in the lab.

**GENOME FACTS:** Information on the genome of this fungus is very limited. The type strain is available from the ATCC stock center (ATCC 34533). To our knowledge, no plasmid, cosmid, or BAC genomic libraries are available.

**COMMUNITY:** The sequence will be of broad interest to fungal researchers and eukaryotic biologists in general. DNA for *U. reesii* will be supplied by John Taylor at the University of California, Berkeley.

### 8.7 *Penicillium* cluster

**SIGNIFICANCE:** *Penicillium marneffei* is endemic to the Southeast Asian region where it is the second most frequent opportunistic fungal infection in AIDS patients after *Cryptococcus*. Approximately 21% of all AIDS patients in these regions are infected with *P. marneffei*. Although *P. marneffei* has been primarily associated with HIV-positive individuals, a number of cases of acquired infection in non-HIV patients have been reported. In a subset of these cases, other immunodeficiencies were identified. In all cases, untreated *P. marneffei* infections are fatal.

*P. marneffei* causes both systemic and superficial infections. It is the only known *Penicillium* species that exhibits dimorphic growth, converting to the yeast form at 37°C. Similar to many other fungal pathogens, infection is likely to occur through the inhalation of the asexual spores produced by the filamentous form. In infected individuals, pulmonary alveolar macrophages and peripheral blood mononuclear cells contain multiple *P. marneffei* cells in the yeast form. Cell-mediated immune responses via pulmonary alveolar macrophages appear to be the primary defense mechanism against *P. marneffei*. But *P. marneffei* yeast cells are capable of dividing and killing macrophages in immunocompromised hosts, and this strategy may provide a mode of dissemination throughout the host.

*P. marneffei* infections respond to treatment with antimycotics, such as amphotericin B and a number of the azole drugs. These treatments are rarely curative and continued treatment is required to avoid relapse, especially in immunocompromised individuals. This dormancy and reactivation poses a serious threat to patient health and continued prophylactic administration of antimycotics can lead to incidences of drug resistance.

**GENERAL DESCRIPTION**: *P. marneffei* is most closely related to the *Talaromyces* species, with biverticilliate conidiophore morphology. In addition to its importance as a pathogen, *P. marneffei's* highly developed molecular genetic systems and similarity with many other less amenable fungal pathogens such as *Histoplasma* and *Coccidioides* represents an excellent genetic model system. Comparisons between the *P. marneffei* genome sequence with those of *Histoplasma* and *Coccidiodes*, as well as monomorphic species such as *A. nidulans* and *A. fumigatus*, will be highly informative. To accurately annotate the *P. marneffei* genome, we propose to also sequence a closely related species, *Penicillium mineoleutium*.

**GENOME FACTS:** The estimated genome size for *P. marneffei*, based on pulse-field gel electrophoretic studies, is between 22–33 Mb. The genome size for *P. mineoleutium* is not known. The organisms are haploid and abundant repeated sequences have not been identified. The GC content of over 2000 random *P. marneffei* sequences does not significantly differ from that of other related fungi such as *A. fumigatus* and *A. nidulans*. *P. marneffei* phase-specific cDNA libraries are available. A genomic DNA library of short (<500 bp) fragments is being used to generate glass-based microarray slides for expression studies and is also available. Clones of interest based on expression data will be sequenced and all clones are available.

**COMMUNITY:** Many groups in Southeast Asia, Europe, and the United States are involved in identifying genes implicated in infection, diagnosis, and disease. Matthew Fisher (The Institute of Zoology, U.K.) is currently studying the evolution and population genetics of *P. marneffei* using both microsatellites and single-copy genes. Alex Andrianopoulos and Michael J. Hynes (University of Melbourne, Australia) are involved in understanding the molecular mechanisms controlling development and pathogenicity. K.Y. Yuen (University of Hong Kong), in collaboration with the Pasteur Institute, has sequenced over 2000 random sequences from *P. marneffei*, and in collaboration with the Beijing Genomics Institute plans to sequence the genome of a highly pathogenic variant of *P. marneffei*. However, these groups do not intend to release this sequence to the public domain.

Genomic DNA from the *P. marneffei* type strain (FRR 2161, ATCC 18224) will be supplied by Alex Andrianopoulos (University of Melbourne, Australia). Genomîc DNA from *P. mineoleutium* will be supplied by Sybren de Hoog at Fungal Biodiversity Center, Utrecht, The Netherlands.

### 8.8 *Stachybotrys chartarum* (Ehrenb. ex Link) Hughes (synonyms = *S. atra, S. alternans*)

**SIGNIFICANCE**: *Stachybotrys chartarum* (Ehrenb. ex Link) Hughes (synonyms = *S. atra, S. alternans*) is the fungus most publicized in the popular press over the past several years due to its growth in water-damaged homes and its presumed association with sick-building syndrome. It is the etiologic agent of stachybotyrotoxicosis, a severe disease of large domestic animals, especially horses, that was first reported in the 1930s in Eastern Europe. The disease was found to be due to fungal contamination of straw and grain and is manifested by pancytopenia, hemorrhage, and neurological disturbances, often resulting in death. When fed to horses, toxins purified from the fungus were shown to be capable of causing illness and death. In the mid-1990s, *S. chartarum* emerged as a major health concern with a cluster of cases of pulmonary hemorrhage and hemosiderosis in infants that may have resulted from exposure to a hemolysin, produced by this organism. This remains controversial, with the CDC first reporting and later questioning a causal relationship between fungal contamination and disease. In addition, the fungus has been implicated in sick-building syndrome. *S. chartarum* grows on water-saturated cellulose materials, such as wallpaper, gypsum and fiberboard, and produces conidia that may contain trichothecene mycotoxins, notably satratoxins F, G and H, with the latter being the most prevalent. These trichothecene mycotoxins are acutely toxic to mammals, inducing cytotoxicity, inflammation, and severe pulmonary irritation.

G**ENERAL DESCRIPTION**: *S. chartarum* is a member of the Deuteromycota in the form-order Moniliales and family Dematiaceae based on its production of simple determinate conidiophores with whorls of phialides containing dark-brown to black ridged conidia (8–11 × 3–6 μm) encased in slime. Its capability of growing on moist material rich in cellulose and low in nitrogen has led to it being found on the walls of buildings where water problems, such as plumbing or roof leaks, excess humidity, or flooding damage have occurred. The concern over indoor fungal contamination is due to its production of several mycotoxin secondary compounds. The secondary compounds include the trichothecenes and related trichoverroids that are potent translation inhibitors; the immunosuppressants phenylspirodrimanes and cyclosporin; the stachybocins, which are endothelin-receptor antagonists; and the atranones, a recently described class of compounds for which biological activity has not yet been described. It is the variety of secondary compounds produced by *S. chartarum* and their potential for human health problems that make this organism a prime candidate for genomic analysis. Comparative genomic analysis between this fungus and the *Fusarium* species that produce trichothecenes are also of interest.

**GENOME FACTS**: There has been little work the on genomic or molecular analyses of *S. chartarum*. In fact, although its closest relatives phylogenetically are haploid ascomycetes, it is not known whether the fungus is haploid or diploid. Molecular work has primarily focused on identifying species-specific sequences that can be used for detection of *S. chartarum* by PCR from air and environmental samples. Detailed analyses of trichothecene biosynthesis has been performed in the *Fusarium* species, including identification of biosynthetic and regulatory genes. It will be of great interest to do comparative studies with the parallel pathways in *Stachybotrys* to see whether similar regulatory controls are used.

**COMMUNITY**: Interest has accelerated on the possible role of *S. chartarum* as an etiologic agent of sick-building syndrome and infant pulmonary hemorrhage/hemosiderosis, as well as the use of mycotoxins, in particular trichothecenes, in biological warfare. A symposium at the 15th Congress of the International Society for Human and Animal Mycology focused on mycotoxins and the sick-building syndrome, with discussions on the health effects of the fungus in home and work environments and the chemistry and toxicity of *S. chartarum.* The presenters were from the United States, Canada, and Australia, and the attendees included numerous mycologists from globally diverse countries. Elucidation of the genomic sequence of *S. chartarum* would greatly enhance studies directed toward establishing the role of this fungus in human and animal disease and provide the information required for development of diagnostic tools, treatments, and a vaccine against mycotoxins.

### 8.9 *Sporothrix schenckii*

**SIGNIFICANCE:** *Sporothrix schenckii* is a pathogenic dimorphic fungus with worldwide distribution and is the causative agent of sporotrichosis (rose picker's disease). In its saprophytic form it develops hyaline, regularly septated hyphae and pyriform conidia, which can be found single or in groups in a characteristic daisy-like arrangement. The yeast or parasitic form shows ovoid cells with single or multiple budding. Information obtained from the phylogenetic analysis of various *S. schenckii* genes groups this human pathogen closer to the filamentous fungi and plant pathogens than to other human pathogens, such as *Candida albicans*, *Cryptococcus neoformans*, and *Histoplasma capsulatum.* Like other human pathogens, however, *S. schenckii* is dimorphic, and switches from a hyphal form to a yeast form at 37°C. At present, the information available on this emerging pathogen is inadequate when compared with that available for other pathogenic fungi, and would be greatly enhanced by the genome sequence.

Also, *S. schenckii* is an excellent model system for the study of proliferation during a developmental switch, i.e., the proliferation accompanying morphogenetic yeast to mycelium transition. *S. schenckii* yeast cells can easily be synchronized and induced to undergo transformation to the mycelium form of the fungus. Dimorphism in *S. schenckii* depends on transmembrane signaling pathways that respond to cell density, external pH, cyclic nucleotides, extracellular calcium concentration and protein kinase C (PKC) effectors such as phorbol esters. *S. schenckii* is a particularly useful model for the study of the role of calcium in the regulation of morphogenesis. Two calcium uptake peaks were observed to occur after the induction of the yeast to mycelium transition, and the inhibition of these uptakes resulted in the inhibition of this transition.

**GENERAL DESCRIPTION:** *Sporothrix schenckii* is an ascomycete (Ascomycota; Ophiostomales; Ophiostomataceae; mitosporic Ophiostomataceae). It can be grown readily as yeast in a rich, defined medium at pH 7.2 or as the mycelium form at pH 4.0. It also grows in Sabouraud dextrose agar where both phases are observed to be present.

**GENOME FACTS:** The information available regarding the genome of this fungus is very limited. The DNA content per cell was calculated for strain ATCC 58251 to be ~$1.87 \times 10^{-8}$ µg/cell. Only six genes have been cloned and sequenced in both the genomic and cDNA forms in this strain. These genes have been found to contain introns that vary in size from approximately 60–100 bp.

No plasmid, cosmid or BAC genomic libraires are currently publicly available.

**COMMUNITY:** The community of scientists interested in *S. schenckii* is relatively small. DNA for the genome sequence will be supplied by Dr. Nuri Rodríguez-del Valle at the University of Puerto Rico.

### 8.10 *Exophiala (Wangiella) dermatitidis*

**SIGNIFICANCE:** *Exophiala/Wangiella dermatitidis* is a model human fungal pathogen. This fungus has recently become better known as a paradigm for the causative agents of phaeohyphomycosis and other emerging dermatomycoses of humans, because of its increasing detection as a systemic pathogen in both immunocompetent and immunocompromised patients.

*Exophiala (Wangiella) dermatitidis* is one of many form-species of the Fungi Imperfecti, which contain many poorly studied human pathogens that cause a variety of diseases, including phaeohyphomycosis, subcutaneous and cerebral chomoblastomycosis, cladosporiosis, dematiaceous eumycotic mycetoma, tinea nigra and black piedra. Many of these diseases can require long-term antifungal intervention and surgery and can sometimes lead to the death of patients. Although no single species causes a great number of infections each year, collectively they cause much morbidity or mortality because of their natural resistance to most antifungal drugs, which is partly attributable to their cell wall melanin. Studies of *E./W. dermatitidis* have demonstrated that this particular agent of phaeohyphomycosis is an exquisitely suitable model for discovering attributes about the other dematiaceous agents of human disease. In part, this is because this fungus produces a greater variety of disease manifestations in humans than most of the other species. It is also a good model for learning about the mechanisms regulating polarized and nonpolarized growth and cellular differentiation in fungi, because *E./W. dermatitidis* can be grown as a budding yeast, a filamentous mold or as a multicellular mass. In fact, most of what is known about the growth of fungi in the multicellular (sclerotic) phenotype, which is characteristic of the tissue phases of all the agents of chromoblastomycosis, comes from studies of *E./W. dermatitidis*. Because of the link between cellular morphogenesis and virulence, many recent studies have focused on the cell wall related attributes of this fungus, which might identify new targets for the development of antifungal drugs.

**GENERAL DESCRIPTION:** The zoopathogenic fungus *Exophiala (Wangiella) dermatitidis* is one of many form-species of the Fungi Imperfecti, which are darkly pigmented (dematiaceous) owing to the deposition of 1,8-dihydroxynaphthalene (1,8-DHN) melanin in their cell walls. Among the fungi comprising the so-called black yeast clade, *E./W. dermatitidis* is the most extensively studied physiologically, biochemically, and molecularly, and possibly also morphologically and cytologically. Members of this fungal clade comprise a unique group of species distinct from the other major Ascomycota clades, which include pathogens such as *Candida albicans* (Class Hemiascomycetes) *Histoplasma capsulatum*, *Blastomyces dermatitidis*, *Coccidioides immitis* and many others.

**GENOME FACTS:** The genome of *E./W. dermatitidis* strain 8656 has been estimated preliminarily to be about 19 Mb. The chromosome number is 4 for the uninucleate, haploid yeast cells of this strain. Many color mutants, auxotrophic mutants, temperature-sensitive morphological and cell cycle mutants, and a variety of molecularly targeted and random disruption mutants have been identified.

**COMMUNITY:** The community of scholars involved with any single dematiaceous human pathogenic fungus is relatively small. However, there is little dispute that genome sequencing of any one of these pathogens would significantly focus this community toward that species, and also attract other new participants. The significantly advanced study already aimed at *E./W. dermatitidis*, together with the advanced nature of its molecular tractability, suggests to most that if the genome of any one of these unique species is sequenced, then it should be that of this established model. Dr. Paul J. Szaniszlo at University of Texas at Austin, in addition to supplying the haploid DNA for this organism, assembled a group of researchers who worked on dematiaceous human pathogenic fungi and who support the sequencing effort.

### 8.11 *Cryptococcus* cluster

**SIGNIFICANCE:** *Cryptococcus neoformans* is a human fungal pathogen that is distributed throughout the world. *C. neoformans* is now the leading cause of infectious meningitis. The organism is a basidiomycete, and thus is quite divergent from most other human fungal pathogens, which are predominantly ascomycetes. Existing antifungal therapies are limited to amphotericin B, azoles, and flucytosine, and drug resistance is emerging. The organism has been diverging over the past 40 million years into four distinct serotypes that are significantly diverged from each other and now represent three varieties or even possibly distinct species.

Ongoing genome projects at TIGR, Whitehead/Duke, and Stanford address three serotype A (strain H99) and serotype D (B3501, JEC21) strains. Here we propose to sequence the genome of two divergent serotype B and C strains, which represent the divergent gattii variety. Strains of this variety are mainly localized in tropical and subtropical regions of the world, including Australia, South America, and Africa; there are also isolates from California. In contrast to serotype A and D strains, which infect immunocompromised hosts, the gattii strains infect immunocompetent hosts. Strikingly, strains of the gattii variety have recently caused an outbreak on Vancouver Island in Canada involving more than 50 infections in otherwise healthy human patients, and many cases in animals, including dogs, cats, porpoises and others. Recent studies have recapitulated the sexual cycle of the gattii variety and led to the development of transformation and gene disruption approaches.

**GENERAL DESCRIPTION:** *Cryptococcus* and *Tremella* species belong to basidiomycetes, Hymenomycetes lineage, Tremellales order. Population genetic studies reveal that the serotype B/C strains are ~40 million years diverged from the serotype A and D strains. And the B and C serotype strains are ~10 million years diverged from each other but can still interbreed and may constitute a single species that is currently called *C. neoformans variety gattii* but has been proposed could be renamed *Cryptococcus gattii*. A detailed comparison of the serotype A, B, C, and D strains will reveal general principles about the evolution of this species complex relevant to all fungal evolution;, in particular, about the evolution of a primary human fungal pathogen that infects normal hosts and an opportunistic human pathogen that infects immunocompromised hosts.

We also propose to sequence *Tremella mesenterica*, one of the tremella wood-rotting jelly fungi. *Cryptococcus* species are nested within a single group of *Tremella* spp. including *T. fuciformis* and *T. mesenterica*. The biology of *Tremella* species might be especially relevant to understanding the *Cryptococcus* group. Most *Tremella* are weak-to-strong mycoparasites—the cultivated species are good wood rotters but grow and fruit best when co-inoculated onto logs with other fungi (esp. *Hypoxylon*). Their ecology might provide insights into the wood-rotter origins of *C. neoformans*.

**GENOME FACTS:** Two BAC libraries have been constructed for the serotype B strains WM276 (Australian environmental isolate) and R265 (Vancouver Outbreak Isolate), and 0.75X genome coverage has been achieved for WM276 at the Vancouver Genome Center under the direction of Jim Kronstad. These BACS have also been end sequenced and fingerprinted and there is a BAC scaffold map for WM276. The Heitman lab has identified candidate serotype C strains for sequencing and is prepared to send high-quality genomic DNA for these efforts.

**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. Dr. Joe Heitman (Duke University) will supply the haploid DNA for two *Cryptococcus species*. Dr. Rytas Vilgalys at Duke University, department of Biology will provide *Tremella fuciformis* DNA for genome sequencing.

### 8.12 *Penicillium chrysogenum*

**SIGNIFICANCE**: *Penicillium chrysogenum* strains are the primary commercial sources of penicillin. Penicillin continues to play a critical role in modern antimicrobial treatment as the precursor for many β-lactam antibiotics (representing one-third of the antibiotic market and $8 billion in annual sales). Worldwide production of penicillin is in excess of 40,000 metric tons annually. The lineage of *P. chrysogenum* production strains (and their genomes) represents the most intensive industrial strain improvement effort in history, resulting in more than a 3000-fold increase in yield over 50 years. Top production strains generate 65–80 grams of penicillin per liter. This collection of strains provides the best opportunity to understand the effects of repeated rounds of mutagenesis and selection by humans on genome structure, gene expression, and microbial physiology. Lessons gleaned from *P. chrysogenum* will accelerate the engineering of other important drug-producing strains, such as *Aspergillus terreus*.

**GENERAL DESCRIPTION**: *Penicillium chrysogenum* is a ubiquitous filamentous fungus associated with the ascomycete class. It grows readily on liquid or solid media. On solid media, it conidiates readily to produce aerial brush-like structures (usually green) tipped with spores.

**GENOME FACTS:** The genome of *P. chrysogenum* is haploid and approximately 34 Mb. More than 50 genes have been sequenced and most contain multiple, short introns and have a GC content of 48–55%. There is at least one relatively well-characterized dsRNA mycovirus (3.5-kb genome) found in some strains.

*Penicillium chrysogenum* is particularly amenable to molecular genetic analysis with a variety of available tools, including gene libraries, transformation methods, dominant selectable markers, defined promoters for heterologous or regulated gene expression, and proficient homologous recombination for the generation of precise genome deletions or alterations. Limited genomic analysis of the "Wisconsin Series" of the strain improvement lineage reveals that more "developed" strains contain duplications of the penicillin biosynthetic cluster.

**COMMUNITY:** Due to the tremendous industrial significance of this species, there has been significant and long-standing interest in understanding the physiology and regulation of penicillin production. As described above, there are numerous strains and reagents available. There is also a significant community of *P. chrysogenum* researchers that could benefit from a public sequencing effort.

### 8.13 *Aspergillus niger*

**SIGNIFICANCE:** *Aspergillus niger* is an important industrial fungus that is widely used for the production of enzymes and metabolites, such as citric acid, but also as a host to produce heterologous proteins. *Aspergillus niger* has been used to study fungal protein secretion, proteolysis, cell wall biosynthesis, cell morphology and polarity, degradation of plant (cell wall) polysaccharides, central carbon metabolism and nutrient transport, both genetically and biochemically. Moreover, ecological and evolution studies have been performed with this fungus and related Aspergilli. In the last decade tools have been developed and applied covering the genetics (parasexual cycle), molecular-genetics (transformation, gene-disruption and [over]expression), cell-biology (*in situ* hybridization, GFP-analysis, FRET) and mutant-induction (transposons, *A. tumefaciens* transformation). Although initial transcriptomic and proteomic experiments are carried out within the academic community, the *A. niger* research is seriously hampered by the lack of a publicly available genome sequence. In the private sector, two such sequences have been determined but not freely accessible.

**GENERAL DESCRIPTION:** *Aspergillus niger* belongs to the family Trichocomaceae and is an imperfect fungus. Experience with batch and fed-batch fermentation under controlled small- and large-scale conditions is long standing, whereas continuous culturing has been developed and applied for various physiological and genomics studies. *A. niger* is multinucleate and normally haploid, but can also grow as a heterokaryon or diploid. Conidia are normally mononucleate. Different cell types are formed as it undergoes development.

**GENOME FACTS:** The size of the *A. niger* genome is 35.9 Mb. It has eight well-marked chromosomes containing ~15,000 genes. Over 110 genes have been assigned genetically to linkage groups and over 200 genes are cloned and characterized. Two times the sequence (4–7.5 fold coverage) of different *A. niger* strains was produced by Integrated Genomics and DSM, respectively, but these sequences are not freely available.

There is a BAC library in the private domain (DSM) based on a different *A. niger* strain, a descendent of NRRL 3122. An ordered cosmid library of *A. niger* N402 of 6000 clones is available from the group of van den Hondel/Ram, University of Leiden, The Netherlands. EST sequences produced by A. Tsang, Dept. of Biology, Concordia University (tsang@vax2.concordia.ca) are present at NCBI (no. BE758761-BE760957). Information about the genetic map and chromosomes are available.

**COMMUNITY:** The industrially important fungus *A. niger* is genetically the best characterized organism to date among the asexual filamentous fungi. This strongly justifies public sequencing of its genome in addition to the already available genomes of the model fungus *A. nidulans* and the medically important fungus *A. fumigatus*. Moreover, *A. niger* is the most suitable fungus to study processes like protein and metabolite secretion and other fermentation-related studies.

Currently, more than 70 laboratories worldwide investigate a number of genetic, molecular-biological, biochemical, physiological and process-engineering aspects of this fungus. Within the EUROFUNG consortium (www.EUROFUNG.net), in which *A. niger* research is a major activity, a task force was formed headed by Dr. Steve Oliver (Manchester University) and Dr. Cees van den Hondel (Leiden University) to promote *A. niger* sequencing and annotation. This activity will join well with the current activities for *A. nidulans* and *A. fumigatus*. It will considerably enrich the possibility for comparative and evolutionary genomics. This activity will also stimulate the growing collaboration between the global *Aspergillus* community, especially between the laboratories within Europe and the United States.

### 8.14 *Agaricus bisporus*

**SIGNIFICANCE:** *Agaricus bisporus* is the most widely cultivated edible mushroom, with a world annual production of more than 2 million tons, valued at $4.5 billion or € 4.5 billion. It is cultivated in all regions of the world with the largest production in Europe, North America and China. It is a lignin degrading saprotroph found growing in a variety of habitats, and a major component of the litter-decomposing mycota under some *Cupressus*, *Prosopis*, and *Picea* species. *Agaricus bisporus* provides a model for developmental biology of gill-bearing, fleshy, fruitbody-forming fungi, ectomycorrhizae and homobasidiomycetous plant pathogens for leaf-litter/lignocellulose-degrading fungi that are vital to carbon cycling. Lignocellulose constitutes 95% of the terrestrial biomass for nitrogen and heavy-metal sequestration and movement within forest floor ecosystem, and for versatile control of meiotic self-reproduction. It also has the potential for molecular farming for production of high-value heterologous proteins.

*Agaricus bisporus* exhibits a number of distinct and clearly observable developmental phases and therefore is a good model to study fungal morphogenesis. *Agaricus bisporus* development involves sporulation, spore germination, vegetative hyphal morphologies (which can be influenced by local nutritional and water potential environments), cell aggregation, primordial formation, sporophore growth, tissue differentiation, and sporophore senescence. Identifying the genes/proteins involved in these processes is likely to have relevance and inform the study of fungal morphogenesis and differentiation as a whole. Many of these changes are due to interactions with the chemical, physical and biological environments. The mushroom is subject to attack by a variety of pathogens and therefore the sporophore is an ideal tissue to study fungal–fungal, fungal–bacterial and fungal–viral interactions. It is in this organism in which fungal viruses were first detected.

**GENERAL DESCRIPTION:** *Agaricus bisporus* is white-button mushroom, a well-studied basidiomycete with respect to agronomy, genetics, physiology, morphogenesis, substrate utilization, strain instability, pests and diseases. Unlike well-studied model homobasidiomycetes (e.g., *Coprinus* and *Schizophyllum*), the fertile mycelium lacks clamp connections and there is no evidence of regulation of nuclear number (6–35 nuclei per cell). It can be cultivated on a range of media allowing crossings to be made. The commercial crop varieties are subject to breeding programs within commercial companies and research laboratories.

**GENOME FACTS:** The genome size of *Agaricus bisporus* is 34 Mb. The haploid chromosome number is 13. Auxotrophic, resistance and developmental mutants are available. Specific markers are available for mating type, cap color, basidiospore number and quantitative trait loci. Extensive culture collections are available of cultivated strains and the wild types collected from a geographically diverse variety of habitats. Both homokaryotic and heterokarotic cultures are available. To date there are more than 150 cDNA/genes and 477 ESTs sequenced. The cosmid, cDNA and chromosome libraries are available. Genetic maps are also available with a wide range of markers including isozymes, RFLPs, RAPDs, SCARs and QTLs. Two class I retrotransposons and one class II DNA transposon have been described from the species. Other repeated elements are being characterized.

**COMMUNITY:** The *Agaricus bisporus* community consists of growers, spawn producers and scientists. The growers are widely dispersed in the host countries serving local and international markets. The research community is largely based in Europe, North America and China. Sequencing of the *A. bisporus* genome will benefit and inform the entire mushroom community as well as the wider homobasidiomycete community. There are more than nine laboratories and over 20 investigators working on genetics and sequencing projects of *A. bisporus*.

**8.15 *Tuber borchii Vittad.***

**SIGNIFICANCE:** *Tuber borchii*, an edible hypogeous fungus, is the only ectomycorrhizal ascomycete whose biodiversity, molecular ecology and basic biology have started to be addressed by molecular tools. As a multicellular fungus forming strongly flavored fruitbodies, *T. borchii* is highly prized as a natural food complement (or flavoring additive) in various parts of the world. There is a fairly strong interest in improving its production as well as that of other commercially valuable truffle species such as *T. magnatum* and *T. melanosporum*. Fruiting body development only takes place after the vegetative mycelium has established a productive, symbiotic interaction with a compatible host plant. This association (ectomycorrhiza; ECM) greatly improves plant nitrogen nutrition as well as plant survival under a variety of adverse conditions and is thought to have contributed decisively to the evolution of woody plants. In fact, about 80% of present-day woody plants are found in association with ectomycorrhizal (ascomycete or basidiomycete) fungi.

It belongs to the Peziales, a large and heterogeneous order of fungi presently devoid of any genomic sequence coverage. This order is ecologically significant because it includes both hypogeous and epigeous fungi, with either a saprotrophic or symbiotic lifestyle that colonize a variety of forest and semidesertic ecosystems. The edible mycorrhizal mushroom *Morchella* and the saprobe *Ascobolus* are other well-known members of this order. *T. borchii* colonizes a variety of trees and shrubs in temperate forests. It has an interesting geographical distribution, being present in Central/Southern Europe and China, but not in other continents. In contrast, closely related *Tuber* species (a total of 60 have been described worldwide so far) are restricted to one continent or even to a single area. Laboratory or field specimens of most species are available. Therefore, the genome sequence of *T. borchii* would provide not only a first window into pezizalean fungi, but also a valuable framework for studying the evolution of ecological specificity and the underlying molecular mechanisms of environmental adaptation.

**GENERAL DESCRIPTION:** *Tuber borchii* is a filamentous ascomycete. Protocols for the *Agrobacterium*-mediated genetic transformation of *T. borchii* are being developed. *T. borchii* can be grown as a pure mycelial culture on a variety of chemically defined (solid or liquid) media, and such mycelia can be employed successfully for *in vitro* mycorrhization. Several *T. borchii* strains with distinctively different mycelial morphologies and symbiotic capacities, including the type strain ATCC 95640 are available.

**GENOME FACTS:** The genome of *T. borchii* is thought to be haploid, with an estimated GC content of 50%. More than 300 DNA sequences from *Tuber* are available in public databases; two-thirds are either full-length or EST sequences from *T. borchii*. An expression profiling study comparing *T. borchii* mycelia and fruitbodies has recently been conducted. Additional EST analyses and large-scale expression studies are in progress. Genomic and cDNA libraries for *T. borchii*, and for the related, highly prized species *T. magnatum*, are available. Yeast expression and two-hybrid libraries for heterologous complementation and protein–protein interaction screens are being constructed.

**COMMUNITY:** The position of *T. borchii* within the Pezizales, one of the most basal groups of filamentous ascomycetes, makes this organism particularly attractive to fungal researchers and eukaryotic biologists interested in understanding ascomycete phylogeny and the evolution of morphological characters and ecological strategies. An active scientific community that can provide specific expertise in the cultivation, molecular typization, and molecular biology of truffles is present in Italy and France. A website, created and maintained by the Italian *Tuber* scientific network (http://www.truffle.org) hosts information on various *Tuber* species along with molecular tools for their identification.

### 8.16 *Saccharomyces cerevisiae* — RM11-1a a natural isolate

**SIGNIFICANCE:** *Saccharomyces cerevisiae* is arguably the most important model organism for studies of genetics and eukaryotic biology. As the first eukaryote to have its genome sequenced, it has also become the model of choice for functional and comparative genomics. To date, the sequence of only a single laboratory strain of *S. cerevisiae*, S288C, is available, and a sequence of an independent natural isolate would greatly enhance biological, genomic, and evolutionary studies. High-quality draft sequence of three related *Saccharomyces* species was recently generated (Kellis et al. 2003). The sequence divergence between *S. paradoxus* (the closest of these species) and *S. cerevisiae* is 20%, considerably greater than that between human and rhesus macaque. In contrast, the sequence divergence between RM11 and S288C is estimated to be 0.5–1%, approaching that between human and chimp. RM11 sequence would thus fill an important gap in one of the key objectives of the Fungal Genome Initiative — to compare genomes that represent a variety of evolutionary distances and relationships. The comparison of *S. cerevisiae* with other *Saccharomyces* species found a number of differences in *S. cerevisiae*, and the authors pointed out that "sequencing of additional strains will be required to determine whether [these changes] represent differences in the S288C strain alone or in *S. cerevisiae* in general." Sequence of RM11, an independent natural isolate that shares no recent history with S288C, will provide this information. Indeed, we have found a number of polymorphisms between S288C and RM11 that represent new mutations in S288C relative to natural isolates of *S. cerevisiae* and other yeast species and that alter the function of the protein products in S288C, perhaps reflecting adaptation to the new selective regime in the laboratory environment. Thus some conclusions drawn from studies in laboratory strains may not accurately represent the biology of *S. cerevisiae*. Sequence of a natural isolate will allow identification of such mutations genome-wide and enable studies of the true wild-type alleles, as well as shedding light on natural genetic variation within *S. cerevisiae*. Sequence of RM11 will also *identify S. cerevisiae* genes deleted in S288C and quantify the rate of intraspecific telomeric "genome churning." RM11 has been used as a model for mapping loci that affect gene expression and other complex phenotypes, and the sequence will greatly facilitate positional cloning of the genes involved.

**GENERAL DESCRIPTION:** RM11-1a is a haploid derivative of Bb32(3), a natural isolate collected by Robert Mortimer from a California vineyard. It has high spore viability (80–90%) when crossed with different lab strains. Strains of both mating types and with a number of auxotrophic markers are available. RM11 has been subject of extensive phenotypic characterization, including growth under a wide range of conditions and gene expression profiling. Measurements of gene expression in RM11, S288C, and segregants from a cross between them show that 1/3–1/2 of the genome is differentially expressed, and that these differences are due to at least 500–1000 separate loci, some of which affect expression in *cis* and others in *trans*. Sequence of RM11 will greatly facilitate rapid identification of these loci and comprehensive characterization of regulatory variation. RM11 also has significantly longer life span than laboratory yeast strains and accumulates age-associated abnormalities at a lower rate. It is being used in studies of aging.

**GENOME FACTS:** The *S. cerevisiae* genome is 12.16 Mb with 16 chromosomes. Availability of finished S288C sequence will allow assembly of the RM11 sequence from a low-coverage (3–4X) shotgun, analogous to assembly of the chimp genome with the human genome as the reference. Thus the RM11 sequence can be generated with only ~10% of the sequencing capacity required for 10X coverage of a new 40-Mb fungal genome. Sequencing a haploid strain will ensure that assembly and analysis will not be complicated by polymorphic sites. Sequence divergence from S288C is estimated at 1 in 100–200 bp, and this sequence variation is distributed throughout the genome, confirming that RM11 shares no recent history with S288C.

**COMMUNITY:** In addition to several labs that have been working with RM11 specifically, the sequence will be of interest to the large wine yeast community, as well as to the broader *S. cerevisiae* community. More generally, the usefulness of the sequence will extend far beyond the yeast community to scientists with interests in natural genetic variation, comparative genomics, and evolution. Leonid Kruglyak (FHCRC) will supply haploid genomic DNA and has assembled a team to analyze differences between the S288C and RM11 genomes.

### 8.17 *Neurospora* cluster

**SIGNIFICANCE:** *Neurospora* species are the preeminent model filamentous fungi. Comparative studies among closely related *Neurospora* species to the already sequenced *N. crassa* is the best approach to address the mechanisms and genetic consequences of variation in the reproductive systems of filamentous fungi, and will teach us what is possible in close comparisons of muticellular, high-eukaryotic genomes. We propose to sequence *Neurospora tetrasperma* and *Podospora anserina* in addition to *N. discreta*, which was listed in the previous FGI white paper. The divergent time among these four species is about 75 million years, which provides suitable evolutionary distance to identify functional conservation among them. *N. tetrasperma* is a pseudohomothallic species that diverged from the truly heterothallic species *N. crassa* about 2.5 million years ago. The evolutionary distance between *N. discreta* and *N. crassa* is over 5 million years. *Podospora anserina* is used to study various aspects of fundamental biology (including aging, prion, cell differentiation, peroxisome, translation, catalytic RNA) as well as biological questions more focused on basic fungal biology (including vegetative incompatibility, mating types, spore killer). *Podospora anserina* is related to *Neurospora* but has diverged substantially from this species (circa 75 million years ago), making it a valuable model in evolution comparative studies. In addition, it presents a lifestyle completely different from *Neurospora* species; it inhabits dung and thus relies solely on sexual reproduction for dispersion and displays a constitutive aging phenomenon. The differences also include the apparent lack of gene silencing mechanisms (RIP, MIP, or quelling) making it an attractive model for experiments that require transgenesis.

**GENERAL DESCRIPTION:** Both *N. tetrasperma* and *P. anserina* are Sordariales of ascomycete, Pezizomycotina. They both grow rapidly on solid or liquid media and have simple nutritional requirements. *Podospora anserina* is heterothallic and sexual reproduction takes less than a week. Culture of *N. tetrasperma* is also availabe.

**GENOME FACTS:** The genome size of *N. tetrasperma* is ~40 Mb with seven chromosomes. It is effectively diploid. The genome size of *P. anserina*, measured by pulse-field analysis, is estimated to be 34 Mb and has seven nuclear chromosomes and a mitochondrial chromosome. More than 100 genes have been identified by conventional mating and more than 60 genes have been analyzed up to the molecular level and deposited in GenBank for *P. anserina* species. Cosmid, BAC and cDNA libraries are available for *P. anserina*.

**COMMUNITY:** A large community of *Neurospora* biologists exist in the United States and overseas, and an International Neurospora Conference is held biannually at Asilomar (last meeting was March 14–17, 2002). There are over 10 laboratories around the world (three in the U.S., five in France, one in the Netherlands, and one in Germany) working on *P. anserina*, some for more than 50 years. The existence of the *N. crassa* genome is already attracting developmental biologists to the fungal genetics community; and it is certain that the additional *Neurospora* 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 apart from those working directly on this organism. Dr. Don Natvig at University of New Mexico will supply *N. tetrasperma* DNA, and a DNA sample of *P. anserina* is available from Dr. R. Debuchy at the Institut de Génétique et Microbiologie de l'Université d'Orsay at France.

### 8.18 *Schizosaccharomyces* cluster

**SIGNIFICANCE:** Much of what is known about human genome biology has been extrapolated from experimentally tractable model organisms, particularly budding yeast (*Saccharomyces cerevisiae*), fission yeast (*Schizosaccharomyces pombe*), flies (*Drosophila melanogaster*), and worms (*Caenorhabditis elegans*). Such an approach revealed a remarkable similarity in genome organization between *S. pombe* and metazoans. Like metazoans, *S. pombe* has relatively large chromosomes with large repetitive centromeres, large, ill-defined, inefficient origins, and intron-rich genes. In addition to similar genome structure on the sequence level, *S. pombe* and metazoans have similar genome packaging mechanisms; *S. pombe* telomeres and heterochromatin are composed of proteins similar to those found in metazoans.

The similarities between the genomes of *S. pombe* and metazoans make *S. pombe* an important model of this aspect of eukaryotic biology. The comparisons between *S. pombe* and closely related species will guide a better understanding of this organism, as was recently acheived in a *S. cerevisiae* comparative study (Kellis et al. 2003). We have selected three additional fission yeast — *S. japonicus, S. octosporus* and *S. kambucha* — that are well-positioned evolutionarily, to provide a clear view of conservation among them. *S. pombe* and *S. kambucha* are closely related, sharing about 98% nucleotide identity, while the other two show separation on the order of 100 million years. Comparisons among these four genomes would provide a powerful way to identify and study their organization, function and evolution. In particular, such an analysis should identify conserved non-coding sequences important in genome structure and metabolism, including understood features such as origins and centromeres and less well-defined features such as cohesion sites, recombinational enhancers, MARs/SARs, heterochromatin boundary elements, and conserved features of unknown function. These features provide precendent to help understand the nature of the surprisingly large amount of non-coding sequence conserved between humans and mice.

As well as being important models for eukaryotic biology, the schizosaccharomycetes are a crucial group for evolutionary studies. They represent a basal lineage in Ascomycota; in several studies they are placed as the deepest branch. That alone makes them an interesting group within Ascomycota. Furthermore, this placement, along with the ancestral traits mentioned above, suggest that the schizosaccharomycetes represent an ancestral eukaryote, at least in terms of genome structure. However, the relationship of their genome structures and other traits to those of the basidiomycetes, which branched off before the schizosaccharomycetes, is unclear. The comparison of the basal ascomycetes (such the schizosaccharomycetes, neolecta, and pnuemocystis), the basidiomycetes, and other basal fungi will be a interesting area of evolutionary research.

The sequences of these three genomes would advance fungal and eukaryotic biology in at least three ways. First, they would allow for the recognition of conserved non-coning sequences involved in genome structure, genome function and other processes. Second, they would allow for a more accurate annotation of the *S. pombe* genome. Third, they would allow for a greater understanding of schizosaccharomycete evolution and the evolution of the ascomycetes as a whole.

**GENOME FACTS:** The fact that the *S. pombe* genome is finished and actively annotated means that the sequence of *S. japonicus, S. octosporus* and *S. kambucha* will be easy to assemble and will be immediately and widely useful. Finally, the genomes are small, about 15 Mb each. The suitability of the organism for experimentation allows for rigorous testing of hypotheses suggested by genomic comparison. Schizosaccharomycetes are amenable to well-developed genetics, cell biology and biochemistry.

**COMMUNITY:** The sequence data will be widely used in the *S. pombe* field, a large, active, genome-savvy group, and will be widely used for finding human sequence homologs that are not present in *Saccharomyces*. It will be widely used for studying evolutionary biology, such as the origins of eukaryotes. DNA samples for the three species are readily available from our collaborator Nick Rhind at University of Massachusetts Medical School.

### 8.19 *Schizophyllum commune*

**SIGNIFICANCE:** *Schizophyllum commune* is a major model system for mating and distribution of mating-type genes versus house-keeping genes in nature, evolution of tetrapolar mating types, mushroom formation, and general basidiomycete biology, including nuclear migration after mating and signal transduction in the mating-type pathways. *S. commune* is also medically important both as a cause of disease and potential source of pharmaceutical compounds. *Schizophyllum commune* completes its life cycle on defined minimal media within 2 weeks and is easy to grow, sporulate, germinate, cross, and protoplast. It is one of two model systems for tetrapolar mating (*Coprinus* is the other). There is great interest in the genome sequence for evolution as there are not yet true mushroom-forming homobasidiomycetes added to sequencing programs. Comparison of *Agaricales* and *Stereales* would be very important for genomic studies such as synteny, genome evolution, and mating-type gene evolution. The genome sequence would also facilitate the application of results for mushroom growers (Pleurotus, Lentinus, Agaricus) and in the pharmaceutical industry (schizophyllan). The *S. commune* genome will provide information for basic research on fungal growth (hydrophobins, nuclear migration, pheromone recognition) and development (mushroom induction, signaling pathways).

**GENERAL DESCRIPTION:** This wood-rotting, white rot bracket basidiomycete is found worldwide, with a high incidence on deciduous and conifer trees that has created losses to the timber industry of millions of dollars per year, particularly in North America. There is also an increasing incidence of pathogenic cases in humans. *Schizophyllum commune* can also be used for production of compounds for anti-cancer drugs.

*Schizophyllum commune* has been the major model system for mushroom-forming fungi during the last century. Mating-type genes are well defined. Fruiting-body formation occurs on standard media within 2 weeks and spores germinate readily. Many molecular tools, co-isogenic strains and markers, as well as genetic maps are available.

**GENOME FACTS:** The genome size of *S. commune* is 38 Mb. The organism has served as a model for mating-type analyses with full genetics applicable. Eleven chromosomes were visualized by electron microscopy and 14 chromosomes can be distinguished by CHEF gel blotting and hybridization of available marker genes. Although there is a genome sequencing project ongoing at The Greenomics section of the Plant Research International, The Netherlands, the data is not publicly available. Molecular genetic methods are fully developed, such as transformation, a gene knock-out system, and gene targeting for promotor studies. Cosmid libraries (Ullrich and Raper) and multiple cDNA libraries for co-isogenic strains with different developmental backgrounds (Raudaskoski) are available. There are also EST sequences being created at Southeast Missouri State University by Dr. Walt Lilly and Dr. Alan Gathman. About 500 EST sequences are available at GenBank.

**COMMUNITY:** A large research community exists in North America, Europe and Asia. Over 30 laboratories study this model organism. Prof. Erika Kothe at Friedrich-Schiller-Universität will supply haploid DNA sample for the genomic sequencing project.

### 8.20 *Phycomyces blakesleeanus*

**SIGNIFICANCE:** *Phycomyces blakesleeanus* is a filamentous zygomycete used in laboratories for 140 years. In the mid-1950s, Max Delbrück pioneered the use of *Phycomyces* as a model system for sensory physiology in his laboratory at CalTech and at several summer courses held at Cold Spring Harbor Laboratory. Researchers have been attracted by the sensitive and precise responses of *Phycomyces* to light and other stimuli, coupled with easy manipulations and good adaptation to laboratory life. It is a simple prototype of the many organisms that use light as a source of information but not as a significant source of energy. Development, genetics, and carotene production have been other subjects of pioneering research. *Phycomyces* is the second organism, after humans, known to require a vitamin. It was one of the first organisms in the research on spontaneous mutants and the second, after *Drosophila*, in which mutations were induced artificially. It was used to coin the concept and the name "heterokaryosis." The main areas of recent attention have been the various reactions to light, gravitropism, and some aspects of metabolism, particularly the production of carotene. Sensory physiology in *Phycomyces* has concentrated on the behavior of the sporangiophore. The *Phycomyces* sporangiophore is a specialized hyphae that grows away from the mycelium and can reach several centimeter in length. The tip of the sporangiophore contains a dark ball, the sporangia, filled with vegetative spores. *Phycomyces* can sense several stimuli, including light, gravity, wind, touch, and nearby objects, changing the sporangiophore direction and its speed of growth. Light can also modulate sporangiophore development and the synthesis of carotene.

**GENERAL DESCRIPTION:** *Phycomyces blakesleeanus* is a filamentous zygomycete belonging to the family Phycomycetaceae in the order Mucorales. As a member of the class zygomycetes, *Phycomyces* is about as distant from the ascomycetes, the plants, and the animals as these groups are from each other. Research on genetics and sensory physiology has been carried out with the standard wild type NRRL155. *Phycomyces* grows rapidly in solid and liquid media at 22°C. *Phycomyces* is naturally haploid and forced heterokaryons can be obtained after hyphal fusion. Vegetative spores are produced at the tip of the sporangiophores. *Phycomyces* mycelium is a coenocyte, since hyphae do not contain septa. Two standard mating types are available and sexual crosses can be performed. The vegetative cycle takes about 3–4 days to complete and the sexual cycle usually takes about 4–6 weeks.

**GENOME FACTS:** The *Phycomyces* genome is ~30 Mb with a G+C content of 35%. The ORFs sequenced so far have a G+C of ~50%. A short repetitive sequence, 31 bp, has been identified comprising about 5% of the genome. No active transponsons have been described, but several sequences resembling segments of transposable elements have been isolated. Genomic and cDNA libraries are available. Only 12 sequenced genes are available in public databases, which also contain several *Phycomyces* gene fragments. A project to sequence a large collection of ESTs isolated from mycelia grown under different growth conditions is currently being organized. A genetic map containing about 30 mapped genes in 11 linkage groups is available. A large collection of auxotroph and drug resistant mutants, as well as mutants altered in carotene biosynthesis, light sensitivity, sporangiophore growth, and fungal development, are available. Genome manipulation is currently limited to mutant isolation and genetic complementation. All the efforts to stably transform *Phycomyces* with exogenous DNA have been hampered by the low efficiency of transformation and by the rapid degradation of foreign DNA. Transient transformation, as it is performed in plants, should be possible. Future optimization of transformation protocols and selective markers should improve this technique.

**COMMUNITY:** About 15 research groups worldwide are currently using *Phycomyces* as a research organism in studies of photophysiology, gravitropism, cell metabolism, including the biosynthesis of carotene, cell wall dynamics, and cell differentiation. Other research groups occasionally resort to isolating and analyzing *Phycomyces* genes for evolutionary purposes. The availability of the *Phycomyces* genome will serve as a catalyst for the expansion of the *Phycomyces* community and will provide valuable sequence information about zygomycete fungi. Our knowledge of the basic processes investigated in *Phycomyces* will be enhanced by the sequence information provided by the *Phycomyces* genome project.

### 8.21 *Mucor racemosus*

**SIGNIFICANCE:** *Mucor racemosus* will be the first zygomycete to be sequenced. Most of the well-studied species of zygomycetes belong to the family Mucoraceae within the order Mucorales (e.g., *Rhizomucor* sp., *Mucor* sp., *Rhizopus* sp.). A handful of species belonging to *Mucor* and *Rhizomucor* display a repertoire of enzymes that are among the widest in the fungal kingdom, including proteases, amylases, cellulases, polygalacturonases, and lipases.

An advantage to sequencing *M. racemosus* is its dimorphic growth, which can be manipulated based on a change in oxygen. A novel expression system that combines the advantages of two different growth morphologies (yeast-like, homogenous growth facilitating high biomass production, and filamentous growth facilitating protein secretion) represents an alternative for the efficient production of recombinant proteins in microorganisms.

**GENERAL DESCRIPTION:** A number of major biological processes have been studied in detail in this fungus; these include apoptosis, fungal dimorphism, sexual cycle, carotene biosynthesis, lipid metabolism, and nutrition and heterlogous protein production and secretion. The dimorphic character of *M. racemosus* has been studied from a physiological approach, but the genetic and regulatory understanding still needs clarification. The change from hyphal to yeast-like growth is important for the understanding of many of the human fungal pathogens. *Mucor racemosus* could be used as a nonpathogenic model for such studies and in the future provide the basis for many new discoveries of both primary and secondary metabolites.

The Zygomycota are true members of the Fungal Kingdom, but still quite different from the other three phyla: Ascomycota, Basidiomycota, and Chytridiomycota. Characteristics unique to this class of fungi include a nonseptate hyphae, the formation of extremely resistant resting propagules (zygospores) based on fusion of gametangia, a cell wall composed of chitosan and chitis, extremely rapid growth on both liquid and solid substrates that allows the zygomycetes to efficiently compete for available nutrient materials, and the production of a wide spectrum of enzymes and secondary metabolites (e.g., strong pigments, fragrance, and biocides). This enzymatic production allows for growth of the fungi on a wide variety of substrates and within a wide pH range (pH 3–10).

**GENOME FACTS:** *Mucor racemosus* is a nonpathogenic, dimorphic fungus with a small genome. The genome includes 8 chromosomes with an estimated size of 39 Mb. Genetic knowledge is still scarce, although the fungus serves as a model organism in a number of areas of biological research.

Currently, databases include only a few dozen sequences from this species. The phylogeny has been studied using ribosomal sequence data.

**COMMUNITY:** Several European and American research groups use molecular biology to investigate the cellular processes of zygomycetes. This is carried out exclusively through nationally funded projects. The development of zygomycetes as alternative models for the study of fundamental and applied aspects of biotechnology will serve as a platform to design new processes and applications.

### 8.22 *Xanthoria parietina*

**SIGNIFICANCE:** Lichens are regarded as a classic example of a mutualistic symbiotic relationship. They are formed from a fungal "mycobiont" and one or more algae and/or cyanobacteria "photobiont" partners, with the fungal partner normally constituting at least 80% of the lichen thalli. They are able to grow in environmentally extreme habitats and lichen-dominated vegetation covers around 8% of the world's terrestrial surface. Lichen-forming fungi constitute more than 20% of all known fungal species and therefore represent a major component of fungal biodiversity. It has recently been suggested that they are the ancestors of many extant free-living fungi. Lichens are also important bioindicators, used to detect and map pollution deposition and other forms of environmental disturbance. They also form unusual secondary metabolites, some of which may have pharmacological activity. However, many fundamental aspects of lichen biology remain unknown, hindered in part by their slow rates of growth and difficulty to culture *in vitro*. Recent developments in molecular biology and genomics now provide a unique opportunity to investigate the biology of lichens, investigating topics such as the molecular nature of the symbiotic relationship, how they may be adapted for growth in extreme environments, and aspects of lichen population biology.

**GENERAL DESCRIPTION:** *Xanthoria parietina*(L.) Th. Fr. (1860) has been chosen as a model organism to represent lichen-forming fungi. It is a common lichen distributed worldwide, and is found in both urban and rural areas on a variety of nutrient rich and enriched substrata. It is an ascomycete and a member of the order Lecanorales, which encompasses the majority of lichen-forming fungi. The growth habit of *X. parietina* is typical of many lichens; forming stratified, foliose thalli with a Trebauxioid algal symbiotic partner. It can be readily bought in axenic culture from discharged ascospores, and then maintained on solid or liquid medium. It has a high growth rate in axenic culture compared with many of the other lichen-forming fungi.

**GENOME FACTS:** From previous studies with ascomycete fungi, it is estimated that the genome size of *X. parietina* will be between 30–40 Mb. It has already been possible to construct a genome library of *X. parietina* in a standard lambda-based vector system.

**COMMUNITY:** *Xanthoria parietina* is one of the most frequently studied lichen species in academia and biotechnology worldwide. Investigations have been made in the contexts of physiology, ecology and reproduction. *Xanthoria parietina* has also been used as a model species in pioneering molecular studies involving cloning of lichen genes involved in hydrophobin synthesis. The existence of a sequenced lichen genome would be a significant scientific resource for the worldwide lichen community. It would allow molecular-genetic and functional analysis of many aspects of lichen biology, e.g., investigations of mycobiont/photobiont interaction and development and production of lichen-specific secondary metabolites, genes involved with growth regulation, and nutrient recycling. Comparative genomics with other fungal genomes would also make this a fundamentally useful data set for other fungal biologists, together with a broad range of evolutionary biologists and bioinformaticists.

Haploid DNA for genome studies will be supplied from the Nottingham, U.K. group of Prof. David Archer and Drs. Peter Crittenden and Paul Dyer.

### 8.23 *Coelomomyces* species

**SIGNIFICANCE:** *Coelomomyces* is a water mold that is recognized as the main eukaryotic pathogen of mosquito larvae. The genus has approximately 60 different species and varieties that have been reported from all the major taxa of mosquitoes, including the major vectors of disease. Their ubiquitous occurrence and killing effect on these insects attracted early workers interested in biological control. Recorded epizootics in the field were not, however, duplicated by results in the laboratory. Efforts to infect larvae with field-collected fungal inoculum routinely failed. The answer to this paradox was discovered in 1974, when H. Whisler found that *Coelomomyces* required an alternate crustacean host for completion of the life cycle. By incorporating the copepod alternative host into the biocontrol scheme, *Coelomomyces* can be used to kill mosquitoes and could become a significant participant in the biological control of these medically significant dipterans. The candidate species selected for genomic sequencing is *C. stegomyiae*. It is the parasite of mosquitoes that causes yellow fever in humans.

**GENERAL DESCRIPTION:** *Coelomomyces* are classified in Blastocladiales Chytridiomycota. Species of *Coelomomyces* are obligate parasites of mosquitoes and chironomids, commonly called midges. Because of the high degree of host specificity and the lethal nature of these parasites, they are prime candidates for possible biological control agents for mosquitoes. Mosquito larvae are directly infected by motile zygotes, which, after injecting themselves into the host, produce a mycelium in the blood cavity (hemocoel) of the larva. The mature fungus differentiates into thick-walled sporangia and the larva dies. These resistant sporangia the resting phase of the life cycle. In the appropriate environmental conditions, they will produce meiospores of opposite mating type. These meiospores recognize and infect the alternate crustacean host, giving rise to the haploid or sexual phase of the cycle. Eleven different host-parasite combinations have now been studied and the alternate hosts include cyclopoid and harpacticoid copepods, and ostracods, taken from tropical, temperate, arctic, fresh water, and marine environments.

**GENOME FACTS:** Chytrids are the only fungi that have a developmental stage with flagellae. Based on this and other morphological criteria they were considered to be a distinct subkingdom, the basal one in fungi. Molecular knowledge of chytrids in general and *Coelomomyces* in particular is still limited. Efforts to carry out genetic investigations in the chytrid *Allomyces* were initiated in the 1970s, but they were unsuccessful largely because of the polyploidy. Sequences from three genes encoding RNA Polymerase II subunits has shown that *Coelomomyces*, unlike the majority of chytrids analyzed, contains only a single allelic form of each gene. The obligatory requirement of a sexual phase for *Coelomomyces* reproduction, not shared by the other chytrid fungi may explain these differences in the degree of polymorphism. The haploid isolates of *Coelomomyces* can be collected from the copepod host. *Coelomomyces* has a defined sexual cycle with two mating types. Thus, classical genetic approaches can be applied to study these organisms.

**COMMUNITY:** Within the Deep Hypha group, organized by J. Spatafora, M. Blackwell and J. Taylor, and supported by an NSF grant for bi-yearly meetings, there is a tightly-knit subgroup concerned with chytrids. Mainly it consists of David Porter (University of Georgia), Martha Powell (University of Alabama), Joyce Longcore (Bar Harbor Laboratory) and the graduate students who work with them. This core group is supported by a Pell Research Grant from NSF. In addition, Timothy James, a former Porter student now at Duke University has done some experiments on *Batrachochytrium dendrobatidis*, a pathogen of amphibians, and has links to the core chytrid group. Insofar as *Coelomomyces*, the principal worker has been and will continue to be Professor Howard Whisler, Professor Emeritus in the Department of Biology at the University of Washington. Recently, Whisler established a collaboration with Dr. Yajuan Liu, currently in the laboratory of Prof. Benjamin Hall. Whisler and Liu are committed to growing several *Coelomomyces* species and providing DNA for genomic analyses.

### 8.24 *Mycosphaerella graminicola* (anamorph: *Septoria tritici*)

**SIGNIFICANCE:** *Mycosphaerella graminicola* is the cause of *Septoria tritici* leaf blotch, which in most years is the second most important disease of wheat in the United States. Because of the importance of this disease, the fungus has been the object of intense study and is one of the first molecular genetic model systems among the filamentous fungi. *Mycosphaerella graminicola* is especially important in population studies of natural fungi. In the United States *Septoria tritici* leaf blotch occurs in most wheat-growing areas every year. The annual economic loss to U.S. wheat growers caused by *Septoria tritici* leaf blotch is more than $275 million. As expected for a pathogen of the world's most widely planted food crop, this disease is an important fungicide target, especially in Europe, which is the most important market for cereal fungicides for the U.S. agrochemical industry. The estimated volume of fungicide input in this area is ~$400 million, which is ~70% of the total fungicide input in cereals. *Septoria tritici* leaf blotch occurs throughout the world wherever wheat is grown. Although losses in the developing world are difficult to estimate, *M. graminicola* infection is very severe in Mediterranean and East African countries, such as the highlands in Ethiopia.

**GENERAL DESCRIPTION:** *Mycosphaerella graminicola* is a haploid, hemibiotrophic ascomycete with both filamentous and yeast-like growth phases. The fungus is easy to culture on both liquid and solid media. It has a bipolar, heterothallic mating system and crosses can be made easily on wheat leaves. Sexual recombination occurs frequently — several times per annum — in nature, and the population genetics of *M. graminicola* is better understood than that of any other fungus. Lab culture collections contain more than 5000 field isolates and several mapping populations are available. Other laboratory strains are available, including knock-out strains and experimental evolution populations that are expressing high levels of resistance to two classes of fungicides through accumulation of mutations during asexual reproduction. These strains will allow an analysis of the number of mutations involved in fungicide resistance. This is a first for a plant pathogen (as well as for a filamentous fungus) and puts *M. graminicola* in the same league as *C. albicans* and *S. cerevisiae* as a model system for *in vitro* experimental evolution studies in fungi.

**GENOME FACTS:** The genome size of *M. graminicola* has been estimated at 32–40 Mb, contained in ~20 small chromosomes, including the smallest chromosomes yet detected in filamentous fungi. A genetic linkage map with 23 linkage groups exists with markers spaced approximately every 5.3 cM, and a new high-density map (~1500 markers) is being constructed. Transformation protocols are available for developing targeted gene knock-outs for functional analysis, as well as several BAC libraries with estimated 8–10X coverage. The mating-type genes have been cloned, as well as a putative virulence factor and numerous ABC transporters that may function in pathogenicity. More than 1200 nucleotide sequences have been deposited in GenBank. In addition, a 0.1X shotgun sequence, ~200 kb of BAC sequence and a database of over 30,000 EST sequences containing more than 10,000 unique genes have been made and are available to academic researchers. A project to sequence the entire mitochondrial genome is scheduled to be completed by July 2003.

**COMMUNITY:** At least 12 laboratories worldwide are devoted to working on *M. graminicola*, as well as numerous international wheat-breeding companies and stations. The community is well organized and cooperative, as exemplified by the International Septoria and Stagonospora workshops, which occur every 5 years (the next is scheduled for December 2003). Furthermore, mutual visits between labs foster scientific exchange and cooperation, leading to a range of co-authored publications. In addition to *M. graminicola*, a large community exists for the banana sigatoka pathogens, *M. fijiensis* and *M. musicola*. If laboratories studying asexual fungi related to *Mycosphaerella* are included, the worldwide community probably numbers more than 40, although not all are involved in molecular genetics. In addition, a large community of plant pathologists studies the diseases caused by these organisms from the host side of the interaction, providing a huge impact for genomic information derived from the *M. graminicola*–wheat interaction.

### 8.25 *Blumeria graminis* f.sp. *hordei*

**SIGNIFICANCE:** *Blumeria graminis* f.sp. *hordei* (*Bgh*) is the preferred model fungus in studies of powdery mildew diseases on plants. Powdery mildew diseases are serious problems on many important crop plants (e.g., barley, wheat, grape, melon, cucumber, and pea). Therefore, huge resources are spent on chemical control, on development of new fungicides and on resistance breeding as the fungal populations constantly change to circumvent these control means. About 100 major resistance genes to *Bgh* are known in barley. However, introduced resistances are usually overcome within a couple of years, most likely due to the airborne spread of these fungi over long distances. So far, no powdery mildew fungal avirulence gene sequences have been published.

**GENERAL DESCRIPTION:** *Blumeria graminis* f.sp. *hordei* (*Bgh*) is a filamentous ascomycete. It was previously named *Erysiphe graminis* f.sp. *hordei*, and is one of many very closely related *formae speciales* attacking different species of the grass family. It is an example of the powdery mildews, a major form of plant pathogens. Powdery mildew fungi are obligate biotrophs living symbiotically with the epidermal cells on all green parts of the host. Thus, growth depends on the availability of a suitable host. *Bgh* is heterothallic with two mating types. Mating occurs readily when opposite mating types are mixed on plants. To take up nutrients, the fungus places a haustorium in the host cells, surrounded by a plant-derived membrane. The interaction between *Bgh* and plants is highly specialized and presents a fascinating biological phenomenon. Transcriptome and proteome studies of this symbiosis are emerging. Many *Bgh* genes are expected to have no homology to sequences in saprophytic and non-obligate plant pathogens. Such genes will be possible candidates for genes responsible for the obligate biotropic lifestyle. Comparative genomics — where identification of *Bgh* genes as well as genes "missing" in *Bgh* – will be a major goal of the genome sequencing project. This approach has significant potential to help define an artificial medium for these fungi, which in turn may lead to a useful transformation method.

The accessibility of these fungi on the leaf surface makes them favored for microscopic studies of the pathogen–host interaction. During the last 30 years numerous microscopic studies of the *Bgh*–barley interaction have dissected several specific developmental stages, where different types of plant resistance are manifested. For this and other reasons, research on powdery mildew resistance has had significant impact on plant pathology. Most recently, *Bgh* has been used on *Arabidopsis* as an interesting model for studies of "non-host" interactions.

**GENOME FACTS:** The sexual stage of *Bgh* has allowed construction of a genetic map of the fungus using seven avirulence genes and 352 DNA markers. The DNA markers include 182 AFLP-markers and 168 RFLP-markers, including 99 ESTs and 42 LTR-retrotransposon loci. The markers are distributed over 34 linkage groups covering 2114 cM, indicating a high level of recombination. The genome contains a significant amount of transponson-based repetitive DNA. Approximately 2000 unique ESTs are available form *Bgh*, in addition to BAC and YAC libraries. The genome size is estimated to be 30–50 Mb.

**COMMUNITY:** Several research groups in the United States, Europe and Japan are engaged in research on *Bgh* and other plant–powdery mildew interactions. The community is well integrated and regular international scientific meetings devoted to these fungi and their interactions with plants are being held. The 2nd International Powdery Mildew Conference will be held October 9–14, 2004, in Monterey, CA and is organized by Doug Gubler.

### 8.26 *Sclerotinia sclerotiorum*

**SIGNIFICANCE:** *Sclerotinia sclerotiorum* (Lib.) de Bary is a pathogen with the broadest host range of any known fungus, infecting more than 408 plant species in 75 plant families. All dicotyledonous plants that have been tested are hosts for this fungus. Over the past 20 years the diseases caused by this fungus have continued to increase in incidence and severity: white mold of bean, head rot of sunflower, white mold of canola, and *Sclerotinia* stem rot of soybean. *Sclerotinia sclerotiorum* is also a natural pathogen of the experimental model plant *Arabidopsis thaliana*, providing an excellent laboratory model to study infection and host–pathogen interactions. Determinants of pathogenicity are complex; only partial or quantitative resistance has been identified in some hosts, although as yet there is little or no resistance available in commercial plant varieties. This makes *S. sclerotiorum* an ideal model for a multigenic, quantitative analysis of an economically important fungus. Because there is a completed genome for *Arabidopsis* and there are now substantial genome efforts for soybean, canola, and other crop plants, it is timely to exploit genomic resources for both pathogen and host. As pathogenesis in this fungus is predicated on early events involving pH signaling, down regulation of the host, oxidative burst, and fungal-induced alteration of host signaling pathways, genomic studies will lead not only to a more comprehensive understanding of the determinants of disease development but will facilitate the identification of unique fungal targets that are required for pathogenicity and that can be modulated to reduce the pathogenic machinery of this fungus. Of medical significance, polysaccharides produced by *S. sclerotiorum*, particularly (1-3)-B-D-glucans, have demonstrated antitumor properties attributed to immune system stimulation. *Sclerotinia sclerotiorum* also represents a number of other important plant pathogenic species in the same monophyletic lineage, including *Botrytis cinerea*, an important vineyard, horticultural, and greenhouse pathogen as well as the agent of the delightful "noble rot" in late harvest wines, and *Monilinia* spp., pathogens of many fruit trees and shrubs. Also related, within the class Discomycetes, is the edible morel.

**GENERAL DESCRIPTION:** *Sclerotinia sclerotiorum* is a haploid filamentous member of the monophyletic lineage Sclerotiniaceae of the Ascomycota. It produces multi-hyphal asexual propagules, sclerotia from which develop the sexual structures, and the apothecia from which discharges meiotically produced ascospores. The ascospores are easily cultured and sclerotia result; the meiospore-to-meiospore generation time is 1–3 months depending on methods and isolate phenotype. The discovery of a strain that does not need cold treatment to produce apothecia has greatly facilitated the study of this pathogen.

**GENOME FACTS:** Estimates of genome size based on electrophoretic karyotype analyses range from a conservative minimum of 25.7 to a possible maximum of 43.5 Mb. Unlike many phytopathogenic fungi, *S. sclerotiorum* exhibits relatively little karyotype variation among isolates and contains no minichromosomes. Several small-scale EST projects have begun that include ~3000 ESTs from cDNA libraries of sclerotial initials, vegetative mycelia grown in low-pH medium, and infected soybean plants. Two genomic libraries (one lambda and one cosmid) and a yeast two-hybrid library from the same wild-type isolate are available. Functional analysis of genes via gene replacement, overexpression, antisense, RNAi, and reporter gene strategies have been demonstrated.

**COMMUNITY:** Research interest in *S. sclerotiorum* has a long history dating back to the formal and detailed biological descriptions of Anton DeBary. At least 50 laboratories worldwide have ongoing projects related to the biology, pathogenicity and plant resistance to this fungus. Molecular tools are now well established and the community of molecular biologists working with *Sclerotinia* is growing. Genomic DNA for a *S. sclerotiorum* sequencing project is readily available from the laboratories of Martin Dickman (University of Nebraska), Linda Kohn (University of Toronto), and Jeffrey Rollins (University of Florida). As a result of industry, commodity group and grower concerns, the USDA-ARS is in the second year of funding (~$1 million) a multi-state, multi-institutional and multi-crop initiative to address questions regarding crop protection from diseases of *S. sclerotiorum.*

### 8.27 *Holleya sinecauda* aka *Nematospora sinecauda*

**SIGNIFICANCE:** *Holleya sinecauda* is a plant pathogen on mustard seeds that is closely related to *S. cerevisiae* and *Ashbya gossypii*. It has a small genome size and shows synteny with *A. gossypii*. The comparison of this genome sequence to these previously sequenced and well studied ascomycetes will allow the wealth of knowledge about yeast to inform our understanding of a plant pathogen's mode of growth and survival.

Sequence from the *H. sinecauda* will provide a valuable tool to help understand ascomycete genome evolution and initiate evolutionary developmental studies of fungal growth modes: *S. cerevisiae* grows as a yeast/pseudohyphal organism, *H. sinecauda* as a dimorphic fungus, and *A. gossypii* as a strictly filamentous fungus. The comparison of the different gene sets, particularly the sets of unique genes and absent genes, will suggest new areas of study in fungal developmental biology. *H. sinecauda* may be used in this respect as a model to analyze the dimorphic transition that is one of the major virulence determinants of *C. albicans*.

**GENERAL DESCRIPTION:** *Holleya sinecauda* is an ascomycete fungus. It is dimorphic and contains both yeast-like and hyphal stages. Yeast cells can undergo sporulation to produce asci with eight spores. During the life cycle, diploid stages occur as in *S. cerevisiae*, although at present no sexual cycle has been described. A transformation system has recently been developed at Hans-Knoell Institute, Germany, that allows integrative transformation. *H. sinecauda* is sensitive to the antibiotic G418 and therefore kanMX can be used as a dominant selectable marker. Plasmid biology has been established using *A. gossypii* centromeric DNA and an *A. gossypii* ARS-element. Vectors have been constructed to propagate at either high or low copy.

**GENOME FACTS:** *Holleya sinecauda* contains a small genome for a fungus; it is thought to be ~8.5 Mb. This is 500 kb to 1 Mb smaller than the genome of *A. gossypii*, reflecting a more compact organization of inter-ORF regions in *H. sinecauda* as compared with *A. gossypii*. The combination of small genome size and the syntenic gene arrangement with *A. gossypii* make *H. sinecauda* a very attractive organism to be sequenced. An *H. sinecauda* BAC library is currently being constructed.

**COMMUNITY:** *Holleya sinecauda* is well known primarily to researchers interested in fungal evolution and phylogeny. The *H. sinecauda* sequence, however, will be of great value to those studying molecular pathways and signaling routes in yeast. Dr. Juergen Wendland at Friedrich-Schiller University will continue tool development and work on the sexual cycle of *Holleya*. Dr. Wendland will also provide DNA samples for the genome sequencing.

### 8.28 *Microbotryum violaceum*

**SIGNIFICANCE:** *Microbotryum violaceum* is the best-studied pathogen in natural plant populations. It has become a model system for understanding the impact of disease on plant populations as well as the impact of sexually transmitted diseases on non-human populations. M*icrobotryum violaceum* causes anther-smut disease, affecting over 100 species in the Caryophyllaceae (carnation family) and related families. The disease is found commonly on native and introduced species in North America, Europe, and Asia, and it has close relatives in South America.

*Microbotryum* serves as a model for the study of sexually transmitted diseases in non-human populations and for understanding the process of host shifts and the emergence of new diseases. It is one of the few plant pathogens being studied as the focus of an NIH-funded project. Results from the study of this organism have increased our understanding of the effects of escape of resistance genes into natural populations, the coexistence of sexually transmitted pathogens and their hosts, the role of pathogens in regulating natural populations, and the emergence of new diseases via host shifts. *Microbotryum* induces a sex change in *Silene latifolia* and this is being used as a tool to probe the molecular mechanisms of sex-specific morphogenesis in plants. Under some conditions, *M. violaceum* undergoes intra-tetrad mating, which allows for the explicit study of meiotic parthenogenesis that is characteristic of many fungi, plants, and animals. It is the first fungus shown to have strongly dimorphic mating-type chromosomes, and therefore represents a tractable system for addressing the origin of sex chromosomes in haploid mating organisms.

**GENERAL DESCRIPTION:** *Microbotryum violaceum* is a basidiomycete fungus with a yeast-like saprophytic haploid stage that can be cultured on standard media, and a dikaryotic hyphal phase during which it is an obligate parasite. The diploid spores produced in the anthers undergo meiosis upon germination. There are two mating types, and conjugation is a prerequisite for infection. Development during germination and production of infection hyphae are well characterized and stimulated by host compounds.

**GENOME FACTS:** The genome size of *M. violaceum* is 25–30 Mb with 10–14 haploid chromosomes, depending on the isolate. Color mutants, auxotrophic mutants, and mating-type linked lethals have been identified and linkage maps are available. A large number of AFLP sites and microsatellites have been used as markers for population studies. A cosmid library has been developed and sequences are available for regions near the mating-type locus. Transformation protocols are published. Chromosome separation using pulsed-field gel electrophoresis is routine and individual chromosomes have been isolated. Collections of haploid cultures, including complete linear tetrads, are readily available through the web. There are an unusually large number of chromosomal rearrangements for an obligately sexual fungus. Retrotransposon-like elements are common and many of them are sequenced.

**COMMUNITY:** As a major model system in plant–pathogen interactions, *M. violaceum* is being studied by research groups in the United States, Sweden, France, The Netherlands, and Germany. Research on this fungus has received nearly $3 million in competitive grant funding from NSF and NIH. In addition, one of the common host plants of this fungus (*Silene latifolia*) is an important weed, and research on its pathogens and pests is being funded by the USDA. This host plant is also the object of studies supported by NSF on sex-chromosome evolution, floral development, metapopulation genetic structure, and sex-ratio distorter genes in plants.

### 8.29 *Puccinia* species

**SIGNIFICANCE:** *Puccinia* species are rust fungi that are classified in the phylum Basidiomycota and of the order Uredinales. The current estimate is that there are ~5000 species of rust fungi belonging to about 140 genera. Economically speaking, rust fungi are among the most important of all the Basidiomycota species and all are parasitic on plants. *Puccinia triticina* is the most widespread and common pathogen of wheat in the United States and worldwide. In the United States, winter and spring wheat crops are annually attacked by the leaf rust fungus. Losses can vary up to 10–15% in years that are favorable for development of rust infections and spread of infectious air-borne spores. Stripe rust, caused by *Puccinia striiformis*, is a major disease of wheat and barley in areas of the world where temperatures remain cool and moderate during the growing season. Wheat stripe rust is an important disease in northwest Europe, China, parts of central Asia, the Middle East, and the western United States and Canada. In recent years, stripe rust has become increasingly important in many countries as well as the southern Great Plains and midwestern United States. In the United States, the pathogen has caused losses of multi-million dollars every year since 1999. Both fungi were developed as bio-warfare weapons in the former USSR. *Puccinia striiformis* f. sp. *tritici* was listed as a potential bioterrorism threat to U.S. agriculture in the Public Health Security and Bioterrorism Preparedness and Response Act of 2002.

**GENERAL DESCRIPTION:** *Puccinia* are basidiomycete rust fungi. Most rust fungi have a complex life cycle that includes five spore stages and two hosts. Often the mechanism of infection of the two hosts is different. For example, the asexual urediniospores of *Puccinia graminis* (black stem rust of wheat) infects wheat by entering through the stomata. Infection of the alternate host (Barberry a dicot) occurs through direct penetration of the epidermis. All rust fungi are obligate biotrophs and have developed complex mechanisms for parasitizing their hosts.

*Puccinia* has phylogenetic significance, as it is the largest genus in one of the three major branches within the basidiomycetes. *Puccinia graminis*, another rust fungus, has a NSF/USDA proposal pending.

**GENOME FACTS:** The genome of *P. triticina* is estimated to be ~90 Mb and is comprised of 19 chromosomes. A genetic map is being developed using AFLPs. Several cDNA libraries have been developed from different stages of the fungus. Approximately 20,000 ETS have been sequenced generating ~4500 unique sequences. A fosmid library is also being constructed. At this point very little is known regarding the genome of *P. striiformis*. From work on other rust fungi, however, we predict that the haploid genome will consist of 17–21 chromosomes. Currently, a BAC library for *P. striiformis* f. sp. *tritici* is being constructed and from this an estimate of the genome size will be determined. Given that the sexual stage of this fungus is unknown, no genetic information is available. Molecular markers are being developed for population genetic studies.

**COMMUNITY:** Research on *P. triticina* is conducted at major cereal research centers in Canada, Australia, and at various laboratories in Europe. Sequencing the genomes of *P. triticina* and *P. striiformis* are widely supported by applied and basic researchers interested in rusts, as well as scientists focused on host–parasite interactions. The availability of the genome sequence of *P. triticina* will be especially valuable for comparative genomics with other basidiomycete plant pathogens, such as *Phakopsora pachyrhizi* (soybean rust) and *Ustilago maydis* (corn smut), as well as the filamentous ascoycete plant pathogens *Magnaporthe grisea* (rice blast) and *Fusariium graminearum*. Les J. Szabo (USDA ARS Cereal Disease Lab) will supply the DNA samples of both *P. triticina* and *P. striiformis.*

## 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.

Boffelli D, McAuliffe J, Ovcharenko D, Pachter L, Rubin EM (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 299: 1391–1394.

Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* Published online May 29, 2003.

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.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868.

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.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science.* 292: 929–934.

Jorgensen P, Nishikawa JL, Breitkreutz BJ, Tyers M (2002) Systematic identification of pathways that couple cell growth and division in yeast. *Science* 297: 395-400.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423: 241–254.

Lee TL, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298: 799–804.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.

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.

Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003) Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34: 166–176.

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.

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Dusterhoft A, Fritzc C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, Mottier S, Galibert F, Aves SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, del Rey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG, Nurse P, Cerrutti L (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415: 871–880.

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

## Appendix — Fungi nominated by community for sequencing (as of June 2003)

**ASCOMYCOTA**

***Mycosphaerella graminicola***
***Cenococcum geophilum***
***Aspergillus niger***
***Aspergillus versicolor***
***Neosartorya fischeri***
***Aspergillus clavatus***
***Penicillium chrysogenum***
***Penicillium roquefortii***
***Penicillium marneffei***
***Penicillium mineoleutium***
***Xanthoria parietina***
***Ramalina menziesii***
***Fusarium oxysporum***
***Fusarium verticillioides***
***Fusarium solani***
***Fusarium proliferatum***
***Stachybotrys chartarum***
***Sporothrix schenckii***
***Ophiostoma umli***
***Cryphonectria parasitica***
***Blastomyces dermatitidis***
***Paracoccidioides brasiliensis***
***Uncinocarpus reesei***
***Blumeria graminis***
***Sclerotinia sclerotiorum***
***Neurospora tetrasperma***
***Podospora anserina***
***Septoria lycopersici***
***Tuber borchii***
***Tuber melanosporum***
***Exophiala (Wangiella) dermatitidis***
***Saccharomyces cerevisiae*** (RM11-1a)
***Candida albicans*** (WO-1)
***Candida tropicalis***
***Lodderomyces elongisporus***
***Candida lusitaniae***
***Candida krusei***
***Candida guillermondii***
***Holleya sinecauda***
***Eremothecium gossypii***
***Schizosaccharomyces japonicus***
***Schizosaccharomyces octosporus***
***Schizosaccharomyces kambucha***
***Tolypocladium inflatum***
***Cordyceps militaris***
***Epichloe typhina***
***Ceratocystis fimbriata***
***Colletotrichum***
***Corollospora maritima***
***Xylaria hypoxylon***
***Leotia lubrica***
***Botrytis cinerea***
***Pyrenophora trici-repentis***
***Morchella esculenta***
***Taphrina deformans***

**BASIDIOMYCOTA**

***Schizophyllum commune***
***Agaricus bisporus***
***Microbotryum violaceum***
***Trichodoma***
***Tremella fuciformis***
***Tremella mesenterica***
***Cryptococcus neoformans variety gattii (***Serotype B***)***
***Cryptococcus neoformans variety gattii (***Serotype C***)***
***Tsuchiyaea wingfieldii,***
***Filobasidiella depauperata***
***Filobasidiella flava***
***Filobasidiella xianghuijun.***
***Puccinia triticina***
***Puccinia striiformis***
***Amanita phalloides***
***Flammulina velutipes***
***Armillaria***
***Cantharellus cibarius***
***Phallus impudicus***
***Phellinus pinii***
***Leptosphaeria maculans***
***Stagonospora nodorum***

**ZYGOMYCOTA**

***Phycomyces blakesleeanus***
***Mucor racemosus***

**CHYTRIDIOMYCOTA**

***Coelomomyces stegomyiae***
***Coelomomyces utahensis***
***Allomyces macrogynus***
***Blastocladiella emersonii***

**GENUSES NOMINATED BY RESEARCH COMMUNITY**

***Caldina*** (compare to lichen)
***Letharia*** (compare to lichen)
***Trichoderma (industry, biocontrol)***
***Rhytisma***
***Cladonia***
***Arthonia***
***Orbilia***
***Septobasidium***
***Sporidium***
***Tilletia***
***Polyporus***
***Glomus***
***Smittium***
***Chytridium***