Genomics of the fungal kingdom: Insights into eukaryotic biology

James E. Galagan, Matthew R. Henn, Li-Jun Ma, et al.

Genome Res. 2005 15: 1620-1631
Access the most recent version at doi:10.1101/gr.3767105

Perspective

Genomics of the fungal kingdom: Insights into eukaryotic biology

James E. Galagan,[1] Matthew R. Henn, Li-Jun Ma, Christina A. Cuomo, and Bruce Birren

The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02141, USA

The last decade has witnessed a revolution in the genomics of the fungal kingdom. Since the sequencing of the first fungus in 1996, the number of available fungal genome sequences has increased by an order of magnitude. Over 40 complete fungal genomes have been publicly released with an equal number currently being sequenced—representing the widest sampling of genomes from any eukaryotic kingdom. Moreover, many of these sequenced species form clusters of related organisms designed to enable comparative studies. These data provide an unparalleled opportunity to study the biology and evolution of this medically, industrially, and environmentally important kingdom. In addition, fungi also serve as model organisms for all eukaryotes. The available fungal genomic resource, coupled with the experimental tractability of the fungi, is accelerating research into the fundamental aspects of eukaryotic biology. We provide here an overview of available fungal genomes and highlight some of the biological insights that have been derived through their analysis. We also discuss insights into the fundamental cellular biology shared between fungi and other eukaryotic organisms.

[Supplemental material is available online at www.genome.org. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: J. Stajich.]

The over 1.5 million members of the Fungal Kingdom (Hawksworth 1991) impact nearly all other forms of life as either friend or foe. Fungi play a critical role in the environment through the decomposition of organic material and through symbiotic relationships with prokaryotes, plants (including algae), and animals. In particular, fungi share a long history with human civilization. References in Greek literature, mushroom stones from Mesoamerica dating to 1000–300 BC (Lowy 1971), and dried mushrooms of *Piptoporus betulinus* found in a pouch around a Stone Age man's neck in the Alps (Rensberger 1992) all attest to this long relationship. The relationship can be beneficial, as in the case of biotransformations such as fermentation and the production of antibiotics or extremely detrimental, as demonstrated by the devastating impacts of mycoses, plant diseases, and mycotoxins (Moss 1987).

Found within the 900 million years (Myr) of evolutionary history of the fungi is an enormous biological diversity (Fig. 1). This diversity encompasses four major groups of fungal organism, i.e., ascomycetes, basidiomycetes, zygomycetes, and chytrids. Fungal cellular physiology and genetics share key components with animal and plant cells, including multicellularity, cytoskeletal structures, development and differentiation, sexual reproduction, cell cycle, intercellular signaling, circadian rhythms, DNA methylation, and chromatin modification. The shared origins of the genes responsible for these fundamental biological functions between humans and fungi continue to make the understanding of these fungal genes of vital interest to human biology. In addition, their genomes are more easily sequenced and annotated relative to most metazoans and their experimental tractability makes fungi among the most useful model systems in cell biology.

Despite the importance and utility of fungi, until quite recently what was known about their genomes was primarily derived from the sequence of the yeast *Saccharomyces cerevisiae*. But in the last 5 yr, however, there has been an explosion in fungal genomics that has greatly expanded our view of the genetic and physiological diversity of these organisms. We provide here an overview of available fungal genomes and highlight some of the biological insights that have been derived through their analysis. We also discuss insights into the fundamental cellular biology shared between fungi and other eukaryotic organisms. These highlights are not intended to be comprehensive. Specifically, we focus on results derived from whole-genome analysis of fungi other than yeasts, as the genomics of *S. cerevisiae* and related organisms is covered elsewhere in this issue.

Fungal genomics' history and resources

The era of fungal genomics—and indeed eukaryotic genomics—was ushered in by the sequencing of the complete genome of the yeast *S. cereviseae*, reported in 1996 (Goffeau et al. 1996). This milestone revolutionized work in yeast and enabled the first global studies of eukaryotic gene function and expression. However, the yeast genome sequence provided only a limited glimpse of the biological diversity of the fungal kingdom. The subsequent completion of *Schizosaccharomyces pombe* (Wood et al. 2002) and *Neurospora crassa* (Galagan et al. 2003) revealed the limits of yeast as a proxy for all other fungi. In particular, the genome of *N. crassa*—the first filamentous fungus to be sequenced—possessed nearly twice as many genes as *S. cerevisiae* and *S. pombe* and lacked homologs to known proteins for over 40% of these genes.

Despite evident need, progress in sequencing fungal genomes was initially slow. To accelerate the pace of fungal genom-

[1]Corresponding author.
E-mail jgalag@mit.edu; fax (617) 258-0903.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.3767105.

15:1620–1631 ©2005 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/05; www.genome.org

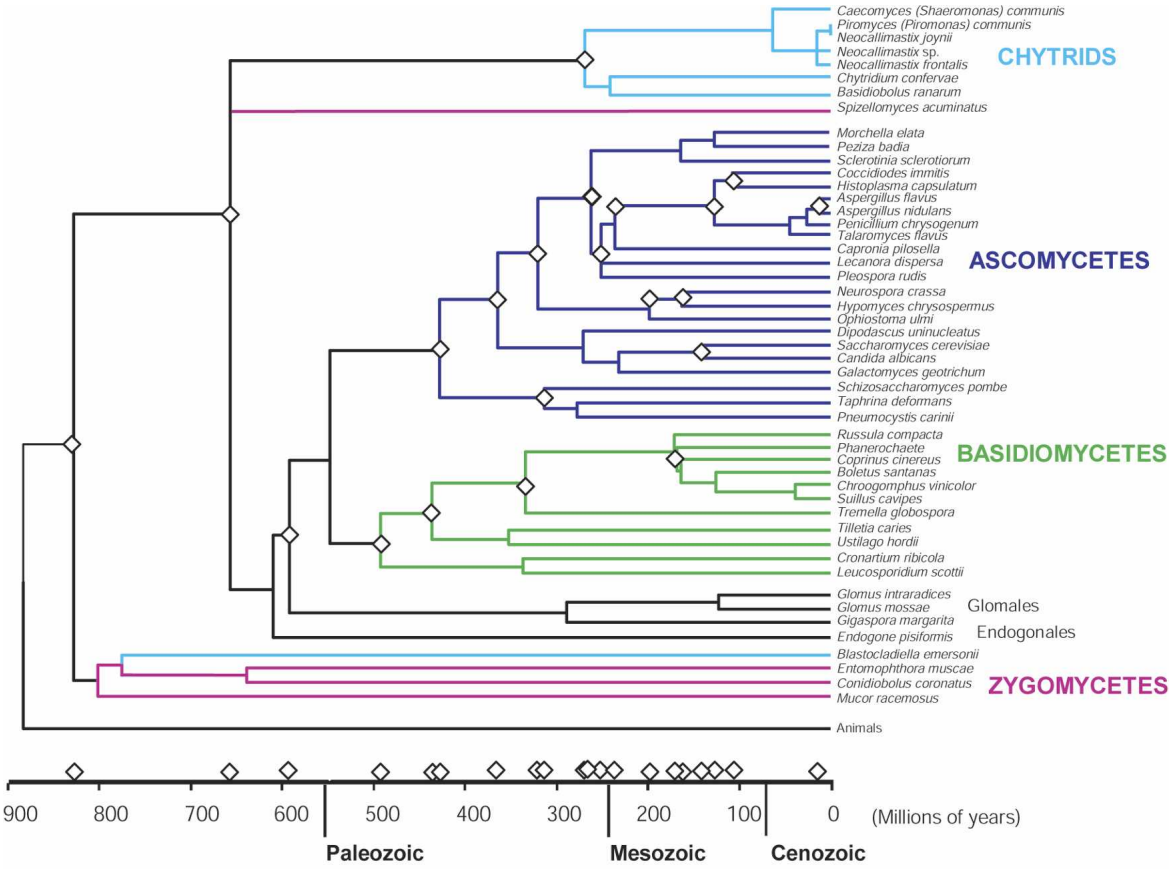

Figure 1. Phylogeny of the fungal kingdom. Major fungal groups colored as indicated by text to *right*. Diamonds indicate evolutionary branch points, and their approximate dating (*bottom*), captured by fungal genomes sequenced in or in progress.

ics, in 2000, a consortium of mycologists in collaboration with scientists from the Whitehead Institute/MIT Center for Genome Research—now the Broad Institute—launched the Fungal Genome Initiative (FGI—http://www.broad.mit.edu/annotation/fgi/). The goal of the FGI is to sequence the genomes of fungi from throughout the kingdom. Importantly, the fungi to be sequenced are not selected one at a time without consideration of each other. Rather, they form groups of organisms that maximize their combined value for comparative genomics, evolutionary studies, eukaryotic biology, and medical studies. When the FGI was launched, two fungal genomes were available. Since that time, 23 different fungal genomes have been released (Table 1) through the FGI. These genomes have been matched by a roughly equal number from other centers and projects including the Joint Genome Institute (JGI), the Washington University Genome Sequencing Center, Génolevures, TIGR, the Sanger Institute, the Marine Biological Laboratories (MBL), the Stanford Genome Technology Center, the Duke Center for Genome Research, and the University of British Columbia. These data have been generated through the support of numerous funding agencies, including the National Human Genome Research Institute, the National Science Foundation, the National Institute of Allergy and Infectious Disease, and the US Department of Agriculture and the Department of Energy. Of particular note is the growing partnership between academia and industry, which has resulted in the release of several privately held fungal genome sequences from companies including Monsanto, Syngenta, Biozentrum, Bayer CropScience AG, and Exelixis.

In total, over 40 fungal genomes sequences are currently publicly available with over 40 additional projects underway (Tables 1, 2). These genomes represent important human pathogens, plant pathogens, saprophytes, and model organisms. They also encompass fungi that grow as yeasts, form mycelium or pseudo-hyphae, or are capable of dimorphic (or polymorphic) growth. In addition, they include representatives of all four major fungal groups. i.e., ascomycetes, basidiomycetes, zygomycetes, and chytrids. Importantly, the majority of available fungal genomes fall into clusters of related genomes that enable comparative analysis across a range of evolutionary distances (Fig. 2). These clusters also include related organisms that differ in terms of specific physiological traits (i.e., pathogenicity), thus allowing these traits to be explored through comparison.

Access to these fungal genomic data is available through a growing number of online resources. These resources include the Broad Institute Fungal Genome Initiative Web site (http://www.broad.mit.edu/FGI/), the JGI Integrated Microbial Resource database (http://img.jgi.doe.gov/pub/main.cgi/), the TIGR fungal database (www.tigr.org.tbd/fungal), NCBI Entrez (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj), the Munich Information Center for Protein Sequences (MIPS—http://mips.gsf.de/projects/fungi/), MetaDB (http://www.neurotransmitter.net/metadb/), and the Genomes Online database (http://www.genomesonline.org/). Particularly useful

Table 1. Complete fungal and Oomycete genomes[a]

Genus/species	Taxonomy	Sequencing center(s)
Ashbya gossypii[c] (aka *Eremothecium*)	Saccharomycetes	Biozentrum an Snygenta AG
Aspergillus fumigatus[c]	Eurotiomycetes	Sanger Institute & TIGR
Aspergillus nidulans[c]	Eurotiomycetes	Broad Institute & Monsanto
Aspergillus terreus	Eurotiomycetes	Broad Institute
Botrytis cinerea	Leotiomycetes	Broad Institute & Syngenta AG; Genoscope
Candida albicans[c,d]	Saccharomycetes	Stanford Genome Technology Center, Sanger Institute, & Broad Institute
Candida glabrata[c]	Saccharomycetes	Génolevures
Candida guilliermondii (anamorph of *Pichia*)	Saccharomycetes	Broad Institute
Candida lusitaniae (aka *Clavispora*)	Saccharomycetes	Broad Institute
Candida tropicalis[b]	Saccharomycetes	Broad Institute & Génolevures
Chaetomium globosum	Sordariomycetes	Broad Institute
Coprinus cinereus	Homobasidiomycetes	Broad Institute
Coccidiodes immitis[b,c,d]	Eurotiomycetes	Broad Institute
Coccidioides posadasii[b,d]	Eurotiomycetes	TIGR & Broad Institute
Coprinus cinereus	Homobasidiomycetes	Broad Institute
Cryptococcus neoformans (anamorph of *Filobasidiella*)[b,c]	Homobasidiomycetes	Broad Institute, Genome Sciences Center Canada, Duke Center for Genome Research, Stanford Genome Technology Center, & TIGR
Debaryomyces hansenii[c]	Saccharomycetes	Génolevures
Encephalitozoon cuniculi	Microsporidia	Genescope
Fusarium graminearum[c] (aka *Gibberella zeae*)	Sordariomycetes	Broad Institute
Kluyveromyces lactis[c]	Saccharomycetes	Génolevures
Magnaporthe grisea[c]	Sordariomycetes	Broad Institute
Neurospora crassa[c]	Sordariomycetes	Broad Institute
Phanerochaete chrysosporium[c]	Homobasidiomycota	Joint Genome Institute
Phytophthora ramorum	Oomycete	Joint Genome Institute
Phytophthora sojae	Oomycete	Joint Genome Institute
Podospora anserina	Sordariomycete	CNRS & Genoscope
Rhizopus oryzae[c]	Zygomycota	Broad Institute
Saccharomyces bayanus[b,d]	Saccharomycetes	Washington University Genome Sequencing Center, Broad Institute, Génolevures
Saccharomyces castellii	Saccharomycetes	Washington University Genome Sequencing Center
Saccharomyces cerevisiae[b,c]	Saccharomycetes	Stanford Genome Technology Center, Sanger Institute, & Broad Institute
Saccharomyces kluyveri	Saccharomycetes	Washington University Genome Sequencing Center & Génolevures
Saccharomyces kudriazevii	Saccharomycetes	Washington University Genome Sequencing Center
Saccharomyces mikatae[c]	Saccharomycetes	Broad Institute & Washington University Genome Sequencing Center
Saccharomyces paradoxus[c]	Saccharomycetes	Broad Institute
Schizosaccharomyces pombe	Schizosaccharomycetes	Sanger Institute
Sclerotinia sclerotiorum	Leotiomycetes	Broad Institute
Stagonospora nodorumm[c] (anamorph of *Phaeosphaeria*)	Dothideomycetes	Broad Institute & International Stagonospora nodorum Genomics Consortium
Uncinocarpus reesei	Eurotiomycetes	Broad Institute
Ustilago maydis[c]	Ustilaginomycota	Broad Institute, Bayer CropScience AG, & Exelixis
Yarrowia lipolytica[c]	Saccharomycetes	Génolevures
Zygosaccharomyces rouxii	Saccharomycetes	Génolevures

[a]Status as of September 10, 2005.
[b]Multiple Strains.
[c]Annotated
[d]Some strains/species still in progress.
A table including URLs is included in Supplemental material.

species-specific fungal databases include the Saccharomyces Genome Database (SGD—http://www.yeastgenome.org/) and CadidaDB (http://genolist.pasteur.fr/CandidaDB/). Of note is the Fungal Genetics Stock Center (FGSC http://www.fgsc.net/), which provides access to clones and other experimental resources in conjunction with several fungal genome projects. A more complete list of online resources is presented in the Supplemental material.

Fungal genome sequencing

The revolution in fungal genomics has been driven by the evolution of genome sequencing technology. Current whole genome shotgun (WGS) sequencing and assembly technologies produce fungal genome sequences with unparalleled accuracy and long-range contiguity at ever-reduced cost. These methods represent an advance over the clone-by-clone approaches used to sequence the first eukaryotic genomes. The clone-by-clone approach relied on labor-intensive clone-restriction mapping to pick sequencing templates, and required separate shotgun libraries for each clone to be prepared, tested, sequenced, and assembled. Ultimately, these maps were not sufficient to protect against both unnecessary overlap and errors originating both with the maps and sequencing. The adoption of more efficient high-throughput sequencing methods coupled with the simplic-

Table 2. Fungal and Oomycete genome projects in progress[a]

Genus/species	Taxonomy	Sequencing Center(s)
Alternaria brassicicola	Dothideomycetes	Washington University Genome Sequencing Center
Aspergillus clavatus	Eurotiomycetes	TIGR
Aspergillus fischerianus	Eurotiomycetes	TIGR
Aspergillus flavus	Eurotiomycetes	TIGR
Aspergillus niger	Eurotiomycetes	Joint Genome Institute
Aspergillus parasiticus	Eurotiomycetes	University of Oklahoma
Batrachochytrium dendrobatidis[b]	Chytridiomycete	Broad Institute & Joint Genome Institute
Blastomyces dermatitidis (anamorph of*Ajellomyces*)	Eurotiomycetes	Washington University Genome Sequencing Center
Candida dubliniensis	Saccharomycetes	Sanger Institute
Candida parapsilosis	Saccharomycetes	Sanger Institute
Fusarium oxysporum	Sordariomycetes	Broad Institute
Fusarium verticillioides	Sordariomycetes	Broad Institute & Syngenta AG
Glomus intraradices	Glomeromycetes	Joint Genome Institute
Histoplasma capsulatum (anamorph of *Ajellomyces*)[b]	Eurotiomycetes	Broad Institute & Washington University Genome Sequencing Center
Kluyveromyces marxianus	Saccharomycetes	Génolevures
Kluyveromyces thermotolarans	Saccharomycetes	Génolevures
Kluyveromyces waltii	Saccharomycetes	Broad Institute
Laccaria bicolor	Homobasidiomycetes	Joint Genome Institute
Lodderomyces elongisporus	Saccharomycetes	Broad Institute
Melampsora larici-populina	Urediniomycetes	Joint Genome Institute
Mycosphaerella fijiensis	Dothideomycetes	Joint Genome Institute
Nectria haematococca	Sordariomycetes	Joint Genome Institute
Nosema locustae[c] (aka *Anionospora*)	Microsporidia	Marine Biological Laboratory
Peronospora parasitica	Oomycetes	Washington University Genome Sequencing Center
Phakopsora meibomiae	Urediniomycetes	Joint Genome Institute
Phakopsora pachyrhizi	Urediniomycetes	Joint Genome Institute
Phytophthora capsici	Oomycete	Joint Genome Institute
Phytophthora infestans	Oomycete	Broad Institute & Sanger Institute
Pichia angusta	Saccharomycetes	Génolevures
Pichia farinose	Saccharomycetes	Génolevures
Piromyces sp.	Chytridiomycete	Joint Genome Institute
Pneumocystis carinii[b]	Pneumocystidomycetes	Broad Institute & University of Cincinnati
Puccinia graminis	Urediniomycetes	Broad Institute
Pyrenophora tritici-repentis	Dothidcemycetes	Broad Institute
Saccharomyces exiguus	Saccharomycetes	Génolevures
Saccharomyces servazzii	Saccharomycetes	Génolevures
Schizosaccharomyces japonicus	Schizosaccharomycetes	Broad Institute
Schizosaccharomyces octosporus	Schizosaccharomycetes	Broad Institute
Trichoderma reesei (anamorph of *Hypocrea jecorina*)	Sordariomycetes	Joint Genome Institute
Trichoderma virens	Sordariomycetes	Joint Genome Institute
Xanthoria parietina	Lecanoromycetes	Joint Genome Institute

[a]Status as of September 10, 2005.
[b]Multiple Strains.
[c]Annotated.
A table including URLs is included in Supplemental material.

ity of WGS strategies has greatly accelerated the pace of genome sequencing while dramatically reducing costs. Advances in assembly algorithms (Myers et al. 2000; Aparicio et al. 2002; Batzoglou et al. 2002; Jaffe et al. 2003; Mullikin and Ning 2003) and the inclusion of end sequences from large insert clones (e.g., Fosmids or BACs) routinely yield assemblies with high-sequence quality and continuity. For example, within the draft assembly of *Fusarium graminearum* an average base falls in a scaffold 5.4 Mb in length, while many scaffolds approach the length of intact chromosomes. Moreover, >99% of the individual bases in this assembly have consensus quality scores equivalent to that of a manually finished sequence.

Despite these advances, a number of challenges remain. Repetitive sequences present the single biggest difficulty in assembling WGS sequence data. The modest level of repetitive sequence ameliorates this problem in most fungi. However, the high identity repeats associated with telomeres, centromeres, and rDNA arrays remain difficult. Often these regions are not cloned in bacterial libraries, while in other cases these regions are cloned and sequenced but not correctly assembled. Although follow up analyses (Farman and Leong 1995; Li et al. 2005) can accurately reconstruct telomeres, more robust automated methods are needed, as are independent mapping methods for assessing the size and position of these difficult to sequence regions.

A special case of repeated sequences are diploid genomes. In diploids, the extent of heterozygosity can vary dramatically across chromosomal regions. Regions of low polymorphism will be incorrectly merged during assembly, while highly polymorphic regions are separated. Consequently, allelic differences are difficult to distinguish from distinct paralogs. When possible, these complications have been avoided by sequencing a haploid form of the organism, or minimized by sequencing a closely related haploid as an aid. But in many cases, such as with *Candida ablicans*, sequencing a diploid is unavoidable (Jones et al. 2004;

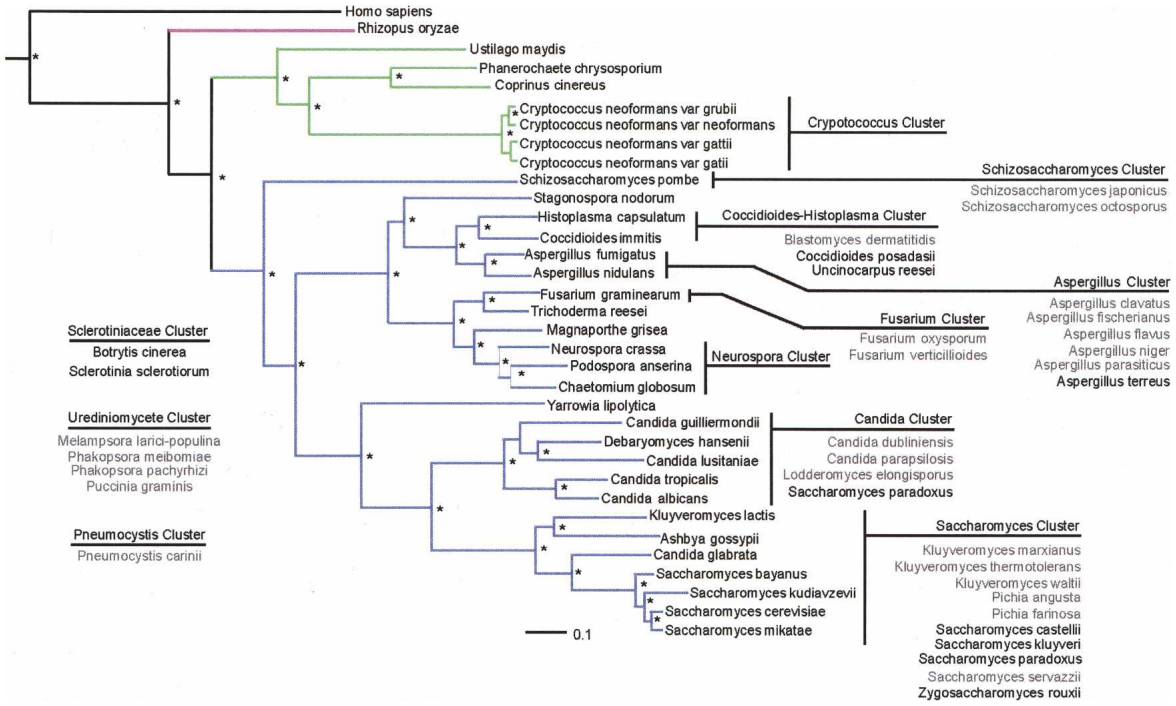

Figure 2. Phylogeny of sequenced fungal genomes and genome clusters. Maximum likelihood tree of 33 fungi with available genome sequence. Additional genome sequences are shown to *right* with genomes in progress shown in gray. Clusters of related fungi are indicated. Phylogeny is rooted with *Dictyostelium discoideum* (data not shown), and major fungal groups are colored as in Figure 1. Tree was generated based on protein sequence from 25 genes; asterisk indicates bootstrap values of >80 using PHYML with JTT model. Phylogeny was generated by Jason Stajich and modified with permission (http://fungal.genome.duke.edu/).

Braun et al. 2005). New assembly algorithms are being developed to more accurately assemble whole-genome sequence data from diploid data sources (Vinson et al. 2005).

The challenges facing fungal genome sequencing are being met by new mapping and sequencing technologies. At least two different mapping approaches offer independent validation of genome assemblies without cloning, i.e., HAPPY mapping and optical mapping. HAPPY mapping is an established technique that determines the proximity of DNA markers through PCR assays using multiple pools of diluted, randomly broken genomic DNA (Dear and Cook 1989, 1993). This methodology is technically simple and does not produce large clone libraries. Optical mapping is a technology that has been newly applied to genome assembly (Zhou et al. 2004). The method produces genome-wide restriction maps based on images of single DNA molecules of megabase length. Comparing the order and distance between restriction sites to in silico digests of genome assemblies provides an independent assembly validation. Both HAPPY and optical mapping do not involve cloning and thus provide access to regions not present in WGS libraries. Both also allow sequences to be assigned to chromosomal locations.

Ongoing advances in sequencing technology also promise to further revolutionize fungal genomics. Although much work is still needed to optimize and fully validate these new approaches, their value is already apparent. For example, pyrosequencing methods implemented by 454 Life Sciences, have successfully generated sequence from *N. crassa* that could not be acquired through conventional sequencing methods. These sequences were found to be AT rich, which likely precluded efficient cloning in bacterial libraries. New instruments also provide the ability to inexpensively produce amounts of data, albeit consisting of short reads—tens to hundreds of base pairs per read compared with 500–1000 for conventional Sanger. The potential cost reduction enables 5–100× more strains or species to be sequenced for the current cost of producing a single genome. While early efforts have focused on producing high-quality reference sequences for individual strains or species, these new technologies will propel us to more fully describe the molecular diversity within related strains.

Fungal gene annotation

Gene prediction in fungi

Gene annotation in the fungi is aided by the comparatively streamlined gene structures in these organisms. Fungal genomes display coding densities ranging from 37% to 61% and, as with other sequenced eukaryotes, gene density is inversely correlated with genome size. Coding sequence lengths average between 1.3 and 1.9 kb. Relative to metazoans, fungal genes are interrupted by few introns, although the fungi display a striking diversity of gene structures. Intron densities in fungi range from 5–6 introns per gene in basidiomycetes such as *Cryptococcus neoformans* (Loftus et al. 2005), to one to two introns per gene on average for many recently sequenced ascomycetes (e.g., *Neurospora*, *Magnaporthe*) (Galagan et al. 2003; Borkovich et al. 2004; Dean et al. 2005) to <300 introns in total in the hemiascomycete yeast *S. cerevisae* (Goffeau et al. 1996). Introns are typically short in fungi, averaging between 80 and 150 bp in many ascomycetes. The basidiomycete *C. neoformans* is exceptional with regard to intron size, with an average intron length of 68 bp and possessing many introns as short as 35 bp. As described below, the structural di-

versity of introns in fungi provides a unique opportunity to study their evolution.

The relatively simple gene structures of most fungi facilitate accurate gene prediction. However, the majority of fungal species lack significant EST data. As a consequence, gene prediction in fungi relies heavily on de novo gene prediction. Given the significant differences in the characteristics of exons and introns between fungi, the training of gene prediction tools on organism-specific data is paramount. A growing number of de novo gene predictors provide this capability. These include GeneID (Guigo et al. 1992; Parra et al. 2000), FGenesh and FGenesh+ (Salamov and Solovyev 2000), SNAP (Korf 2004), Augustus (Stanke and Waack 2003; Stanke et al. 2004; Stanke and Morgenstern 2005), and GlimmerM (Salzberg et al. 1999). In addition to these tools, the programs GeneWise (http://www.ebi.ac.uk/Wise2/index.html) and Exonerate (Slater and Birney 2005) enable gene prediction based solely on the alignment of homologous protein or coding sequences.

Comparative gene prediction is a particularly attractive strategy for fungi, given clusters of related genomes. The utility of comparative gene prediction for fungi has been demonstrated through the comparative annotation of related *Saccharomyces* species (Kellis et al. 2003) and the adaptation of the program TWINSCAN for use with *C. neoformans* (Tenney et al. 2004). The latter program is a pairwise de novo gene prediction algorithm that utilizes homology from an informant genome to make predictions on a reference genome (Korf et al. 2001). TWINSCAN was trained on *C. neoformans* Serotype D (Loftus et al. 2005) and gene predictions were made using *C. neoformans* Serotype A (http://www.broad.mit.edu/) as the informant genome. Verification using known genes or RT–PCR indicated that 60%–72% of the predictions were exactly correct. Given the relative complexity of the *C. neoformans* genome as compared with other fungi, these results are encouraging. A challenge for the future will be to enable training of TWINSCAN and other comparative gene-prediction algorithms for the growing number of fungal genome clusters.

Alternative splicing in fungi

As with many other eukaryotes, a factor complicating gene annotation in fungi is the occurrence of alternative splicing. Examples of fungal genes with multiple alternative transcripts have been previously reported, but large-scale EST sequencing coupled with complete genome sequences is providing a more comprehensive view (Ebbole et al. 2004; Nelson 2004). Perhaps the most extensive genome-wide survey on alternative splicing in a fungus comes from the Basidiomycete *C. neoformans* Serotype D (Loftus et al. 2005). For this project, a high coverage of EST sequence was generated, resulting in alignments between at least one EST for the majority of predicted genes. These data revealed evidence of alternative splicing for 277 genes or 4.2% of the total. Although fungi appear to use alternative splicing less frequently than metazoans (estimates in human range from 40% of genes with alternative splicing to more than 80%) (Modrek and Lee 2002; Johnson et al. 2003; Kampa et al. 2004), these data represent the largest fraction of genes with alternative transcripts reported for any fungus so far, and likely represents a lower bound. Futhermore, the authors identified a variety of alternative splicing mechanisms including exon skipping and truncation, and extensions of both the 5′ and 3′ ends. The results for *C. neoformans* are noteworthy, as previous data from *S. cerevisiae* (Davis et al. 2000; Grate and Ares Jr. 2002; Barrass and Beggs 2003) and *S. pombe* (Romfo et al. 2000) suggested that alternative splicing might only be prevalent in multicellular eukaryotes (Ast 2004). In addition, the few examples of alternative splicing in ascomycete yeasts (Romfo et al. 2000), as well as from other fungi (Ebbole et al. 2004), primarily involve intron retention. The results from *C. neoformans*, a basidiomycetous yeast, indicate that alternative splicing is likely more prevalent and richer than expected, even in single-celled organisms.

Genome evolution

One noteworthy observation coming from the comparison of multiple genome sequences is how divergent fungi are at the genome level, despite apparent morphological and physiological similarities. For example, comparisons of the genomes of *Magnaporthe grisea* and *N. crassa*, related ascomycetes thought to have shared a common ancestor as recently as 200 million years ago (Mya) (Taylor et al. 1999; Berbee and Taylor 2000; Heckman et al. 2001), revealed an average amino acid identity of only 47% and virtually no conserved synteny (Dean et al. 2005). Only 113 regions were identified containing four or more genes in conserved colinearity. More generally, analyses of available complete fungal genomes reveal a rapid breakdown of conserved synteny over a relatively short evolutionary time span (data not shown). Even members of the same genus can display a remarkable divergence at the genomic level. A comparison of three species of *Aspergillus*—*A. nidulans*, *A. fumigatus*, and *A. oryzae*—revealed only 68% average amino acid identity between any pair of species (Galagan et al. 2005), an evolutionary distance comparable to that between human and fish (Dujon et al. 2004). At this distance, roughly 70% of *A. nidulans* could be mapped to a syntenic block with either *A. fumigatus* or *A. orzyae*, with roughly 50% of *A. nidulans* in conserved synteny across all three species (Galagan et al. 2005). Within these blocks of synteny, numerous microrearrangements were evident. These included many small inversions that have been shown to be a common pattern of rearrangement in eukaryotes (Seoighe et al. 2000; Aparicio et al. 2002; Kellis et al. 2003). But other patterns of breakage were equally prevalent, including translocations and segmental insertions, deletions, and duplications. Duplications and translocations in particular have been shown to be a common response of yeast undergoing experimental evolution (Dunham et al. 2002; Koszul et al. 2004) and thus are expected to contribute to the long-term evolution of fungal genomes. Moreover, a whole-genome duplication in yeast followed by massive gene loss, first predicted by Wolfe and colleagues (Wolfe and Shields 1997), has been confirmed by comparative analysis (Dujon et al. 2004; Kellis et al. 2004) and shown to have had a significant impact on yeast fermentation from carbon sources. Comparisons of fungi have also confirmed spatial patterns of rearrangement observed in other eukaryotes, namely, that rearrangements are far more common near telomeres and are frequently associated with repetitive sequence elements (Huynen et al. 2001; Carlton et al. 2002; Coghlan and Wolfe 2002; Kellis et al. 2003; Galagan et al. 2005; Lephart et al. 2005).

Together, these studies indicate that fungal genomes in particular, and eukaryotic genomes in general, are remarkably dynamic. In addition, studies in fungi are allowing us to investigate specific aspects of genome evolution in greater depth, and in some cases, connect genome evolutionary events with specific aspects of physiology. We review here two specific areas that

have received considerable recent attention and for which the availability of genome sequence has led to new insights, i.e., intron evolution and genome defense. Owing to space, many other topics could not be considered, although several have been reviewed elsewhere (Archer and Dyer 2004; Borkovich et al. 2004; Fraser and Heitman 2004; Odds et al. 2004; Ryan and Smith 2004; Bell-Pedersen et al. 2005; Veneault-Fourrey and Talbot 2005; Yu and Keller 2005).

Intron evolution

Although introns have been the object of intense study since their discovery over a quarter of a century ago (Sambrook 1977; Gilbert 1978), numerous questions remain concerning their origins, role, and evolutionary dynamics (for a review, see Lynch and Richardson 2002). Fungal genomes are particularly tractable for the study of introns for a number of reasons. First, the fundamental aspects of intron biology are shared between fungi and other eukaryotes, and thus, lessons learned from fungi are likely to have wide relevance. Second, as described above, fungal genomes are gene dense with relatively simple gene structures, facilitating the accurate prediction of intron boundaries. Finally, as described above, the fungi display a wider diversity of average intron density. This architectural diversity, coupled with the availability of fungal genomes spanning the kingdom, provides an opportunity to investigate intron dynamics. Several recent studies illustrate the utility of fungi for intron studies, and have provided new insight into the patterns and mechanisms of intron evolution.

In a study by Nielsen and colleagues (Nielsen et al. 2004), patterns of intron evolution were investigated in four Euascomycete fungal genomes (*A. nidulans*, *F. graminearum*, *M. grisea*, and *N. crassa*) spanning roughly 330 Myr. The conservation of predicted orthologous intron positions was determined in three of the fungi (using the fourth as an outgroup), and a probabilistic model was used to estimate the most likely rate of intron gain and loss giving rise to these observed conservation patterns. One immediate conclusion stemming from this work was the clear importance of intron gain within this group of Euascomycetes. All three non-outgroup lineages displayed significant numbers of predicted intron gains. In addition, even within this small set of organisms, differences in the pattern of intron dynamics were apparent, with the numbers of gained and lost introns approximately balanced in *M. grisea* and *F. graminearum*, but with roughly twice as many losses as gains in *N. crassa*. Rates of intron gain also varied substantially between gene families.

The subsequent sequencing of additional fungal genomes provided the opportunity to study intron dynamics over a wider evolutionary distance. With these additional data, Stajich and colleagues (J.E. Stajich, S.W. Roy, and F.S. Dietrich, in prep.) studied the patterns of intron gain and loss across 24 fungi spanning nearly the fungal kingdom. With *Homo sapiens* and *Arabidopsis thaliana* as outgroups, the authors developed a maximum likelihood approach to estimate intron loss and gain events and thereby calculate intron densities at various nodes in the fungal tree. Based on a set of more than 700 orthologous protein coding genes, the authors found numerous intron positions shared among plants, animal, and fungi, and they concluded both that these introns were common and present at the origin of the eukaryotic crown. Since the fungal last common ancestor, nearly all lineages were predicted to have suffered substantial intron loss, with particularly significant loss occurring at deeper branches and at the outset of the Hemiascomycete lineage. Interestingly, these authors also find intron gain to be as significant as intron loss in several recent lineages including the Euascomycetes (consistent with Nielsen et al. 2004) and the lineage leading to *C. neoformans*.

One intriguing characteristic of introns is the correlation between intron density and positional bias that has been observed in all eukaryotes sequenced to date (Mourier and Jeffares 2003). According to this 5′ positional bias, introns are evenly distributed within the coding sequence of genes in intron-rich organisms, but are biased toward the 5′ ends of genes in intron-poor organisms. It has been proposed that this bias may have arisen by intron loss through a mechanism of homologous recombination of spliced messages reverse transcribed from the 3′ polyadenylated tail (Fink 1987; Mourier and Jeffares 2003; Roy and Gilbert 2005b). The plausibility of such a recombination-based mechanism has been demonstrated in experiments with intron-containing Ty elements in yeast (Boeke et al. 1985). However, this hypothesis predicts that loss will be biased to introns in the 3′ portions of coding sequences. The pattern of intron loss reported by Nielsen et al. (2004) did not reveal such a 3′ bias. Instead, the rate of intron loss was lowest at the 3′ end, while the highest rates of intron loss occurred in the middles of genes.

A similar pattern of positional intron loss was revealed by more recent work by Stajich and colleagues (J. Stajich and F.S. Dietrich, in prep.). In this analysis, recent intron loss was investigated in the genomes of four closely related *Cryptococcus* species (which diverged <37 Mya [Xu et al. 2000]). The authors identified several loci where multiple intron losses appear to have occurred from a single event. As these loci lacked close paralogs, these losses cannot be explained by gene conversion and suggest intron loss through recombination with RNA. Interestingly, these events all occur in the middle of genes, leaving introns intact at the 3′ end. As with the results of Nielsen et al. (2004), these data suggest that intron loss alone, at least within the recent evolutionary history of the Cryptococci and the Euascomycetes, is not sufficient to explain the observed intron 5′ positional bias in these species.

These and other data (Bon et al. 2003) have revealed fungal introns to be remarkably dynamic. Current gene architectures appear to reflect an interplay between intron gain and loss, with the balance between the two processes varying over evolutionary time. Based on existing data, intron loss appears to dominate in certain fungal clades. This is consistent with the results of an analysis of eight genomes—spanning plants, animals, protests, and including two fungi, *S. cerevisiae* and *S. pombe* (Roy and Gilbert 2005a)—that reported massive and net intron loss in six of the lineages examined. The role of loss appears diminished in more recent fungal evolution with certain lineages gaining nearly as many introns as lost. The mechanism by which introns are gained in any organism remains to be conclusively established, although several theories have been proposed (Logsdon Jr. et al. 1998; Coghlan and Wolfe 2002; Lynch and Richardson 2002; Fedorov et al. 2003). As additional fungal genomes are sequenced, these and other mysteries surrounding intron evolution may eventually be solved.

Genome defense: Repeats, RIP, and RNAi

Repeat sequences are ubiquitous components of fungal genomes. In most genomes analyzed to date the majority of repeat sequences are associated with mobile genetic elements. Copies or remnants of both Class I (retroposons) or Class II (DNA trans-

posons) have been identified, and the number of distinct families of mobile elements continues to grow (for review, see Kempken and Kuck 1998; see also Dean et al. 2005). Microsatellite repeats (Toth et al. 2000) and low-complexity sequence, as well as centromere and telomere associated repeats (Schechtman 1990) are also common. Fungal genomes contain varying amounts of repeat sequence, with "typical" repeat content ranging from between 3% (e.g., *A. nidulans*, *A. fumgatus*, and *A. oryzae*) to 10% (e.g., *Neurospora*, *Magnaporthe*). However, species outside of this range at either extreme have also been identified. Although repeat sequences may play a beneficial role in generating genetic diversity, their presence can also be detrimental, particularly in terms of genome stability. As a result, many organisms have developed "genome defense" systems that repress the activity of transposable elements. Two different genome defense mechanisms in fungi have received particular recent attention as a consequence of genomics—Repeat Induced Point Mutation and RNA silencing.

The first eukaryotic genome defense system described, discovered in the fungus *N. crassa*, is a process called Repeat Induced Point Mutation (RIP) (Selker 1990, 2002; Davis et al. 2000; Galagan and Selker 2004). RIP is a homology-based process that mutates repetitive DNA and frequently leads to epigenetic silencing through DNA methylation. Importantly, RIP has been shown to act on all duplicated sequence, including long segmental duplications, mobile element duplications, and gene duplications. The properties of RIP immediately suggest an impact on genome evolution, and the completion of the *N. crassa* genome sequence allowed the full extent of this impact to be determined (Galagan et al. 2003). Consistent with a role as a defense against mobile elements, the analysis of the *N. crassa* genome revealed a complete absence of intact transposons. However, this defense was shown to come at a price: Essentially, all paralogs in *N. crassa* appear to predate RIP, and since the emergence of RIP gene evolution through gene duplication has been arrested. Gene duplication is widely considered to be essential for the generation of new function. RIP thus illustrates the extent to which genomes can go to defend against mobile elements, and the impact this defense can have on genome structure and evolution. This impact has wide relevance in the fungi, RIP—albeit in a less severe form—has been observed in a growing number of other fungi (Galagan and Selker 2004).

The genome defenses of *N. crassa* are additionally fortified by two different RNA silencing mechanisms, quelling and meiotic silencing by unpaired DNA (MSUD). RNA silencing is a term that encompasses a range of phenomena found in many eukaryotic organisms. Fundamentally, these phenomena involve the repression of sequences with similarity to short RNA molecules. RNA silencing was originally described as "quelling" in the fungus *N. crassa* (Cogoni et al. 1996) and "cosuppression" in plants (Napoli et al. 1990). RNA silencing was subsequently described in *Caenorhabditis elegans* and other metazoans, where it is called RNA interference (RNAi) (Fire et al. 1998; Ketting and Plasterk 2000). The core machinery for RNA silencing—Argonautes, Dicers, and helicases—appears conserved across species (Hutvagner and Zamore 2002) and was first studied as part of the quelling pathway in Neurospora. More recently, it was shown that *N. crassa* possesses a second RNA silencing pathway called MSUD (Aramayo and Metzenberg 1996; Shiu et al. 2001; Shiu and Metzenberg 2002). The analysis of the genome sequence revealed that quelling and MSUD appear to be paralogous pathways derived through the duplication of a core set of genes (Galagan et al. 2003; Borkovich et al. 2004). These two pathways have evolved to operate during different parts of the Neurospora life cycle. Quelling acts during vegetative growth while MSUD acts during the meiosis phases of sexual reproduction. Together with RIP, which acts during sexual reproduction but premeiotically, these pathways effectively silence genes in abberant copy number—and thus protect against genome instability—throughout the entire life cycle of *N. crassa* (Borkovich et al. 2004).

As in Metazoans, in fungi RNA silencing has emerged as a powerful experimental tool for manipulating gene expression. As described above, RNA silencing has been utilized in *C. albicans* to identify essential genes, and RNA-silencing experimental protocols have been developed for a host of other non-yeast fungi. Curiously, no endogenous microRNAs have yet been reported in fungi, although there is some evidence for antisense transcripts (Loftus et al. 2005). It has been suggested that these antisense transcripts may regulate gene expression through the RNAi pathway. But the extent, if at all, to which RNA silencing plays a gene regulatory role in fungi remains unknown.

Plant pathogenesis and environment

Fungi are central to the health of terrestrial ecosystems, and they have played a foundational role in the evolution of life on land. The colonization of land by eukaryotes is thought to have been established through the symbiosis of a fungus and a photosynthesizing organism (Gehrig et al. 1996; Heckman et al. 2001). The symbiosis between fungi and plants plays a crucial role in protecting plants from disease and facilitating nutrient uptake; 95% of all plant families have associated mycorrhizal fungi (Trappe 1987). Fungi also play a central role in degrading organic material. They are the dominant organisms in aerated soils (Frey et al. 1999), typically accounting for 10%–60% of the biomass in forest litter (Newell 1992; Metting 1993). In contrast to these beneficial roles, fungal plant pathogens have a devastating impact on agriculture. Fungi infect all major crop plants (Strange and Scott 2005) and lead to food contamination through the production of mycotoxins. In the United States alone, each year they are estimated to cause $33 billion dollars in damages (Madden and Wheelis 2003) and invoke expenditures of over $600 million on fungicides. Fungal pathogens have had a significant impact on human history. The dominance of tea over coffee in the British Empire can be traced to the failure of coffee in British Ceylon in the 1870s due to infection with leaf blight, caused by the fungus *Hemileia vastatrix*. These fields were planted instead with tea. Access to genome sequence promises to advance our knowledge of the underlying biology of fungal infection and the interaction of pathogen and host, as well as of the mechanisms by which fungi reproduce and persist in the environment. The reports on the genomes for *Phanerochaete chrysosporium* (Martinez et al. 2004) and *M. grisea* (Dean et al. 2005) illustrate this potential.

Lignin is a major component of plant cell walls and the second most abundant natural polymer (Martinez et al. 2004). Only a small group of fungi—termed white rot fungi—are able to degrade lignin, and as a consequence, these fungi play an important role in the global carbon cycle. The genome sequence of one white rot fungus, *P. chrysosporium*, has been generated and a preliminary analysis published (Martinez et al. 2004). The genome contains an extensive and highly redundant array of genes predicted to be involved in lignin degradation. Consistent with the ecosystem role of white rot, enzymes for carbohydrate catabolism outnumbered those for anabolism, the opposite of the

pattern seen in other sequenced eukaryotes. The genome also revealed an extensive array of secondary metabolite gene clusters. The authors suggest that these genes may be attractive targets for bioprocess engineering. Functional studies that utilize the genome sequence may reveal the underlying cellular networks responsible for the important ecological role of white rot fungi.

M. grisea causes the most destructive disease of rice and has emerged as a central model organism for the study of fungal plant diseases. Rice blast, the disease caused by *M. grisea* is estimated to destroy enough rice annually to feed 60 million people (Zeigler et al. 1994). The generation and preliminary analysis of the *M. grisea* genome sequence has provided insight into the molecular basis of fungal plant pathogenicity (Dean et al. 2005). In particular, the genome revealed an expanded family of G-protein-coupled receptors (GPCRs), including a subfamily, one member of which had been previously shown to be required for pathogenesis. The other novel members of this subfamily were shown to be expressed during infection, with two genes specifically up-regulated during the development of a specialized infection structure called the appressorium. The genome also suggested significant diversity among different *M. grisea* strains. Of the seven known *M. grisea* avirulence genes, only four were found in the strain sequenced. With additional sequences, we will better understand this diversity and the role it may play in plant disease. Ultimately, these efforts may lead to improved methods for pathogen control and higher yields of staple foodstuffs worldwide.

Mycoses and medicine

Fungal infections are the third most common hospital-acquired infection, and have emerged as a growing threat to human health (Beck-Sague and Jarvis 1993; Swartz 1994). They have lethal consequences for the growing population of patients immunocompromised with AIDS and leukemias or therapeutically immunosuppressed. The two most common fungal pathogens are *Candida* and *Aspergillus* species: Candidiasis is the most common HIV-related fungal infection with mortality reaching 49% (Gudlaugsson et al. 2003) while Aspergillosis has caused up to 10,000 hospitalizations per year with mortality as high as 20% (Dasbach et al. 2000). Emerging fungal infections represent an equally serious threat to healthy human populations. For example, in 2002, an outbreak of *C. neoformans* occurred on the east coast of Vancouver Island, British Columbia affecting at least 59, mostly immunocompetent, individuals and causing at least two deaths (Hoang et al. 2004). The incidence of Valley Fever caused by the dimorphic *Coccidioides* is increasing with more than 100,000 cases occurring each year in the United States alone (Chiller et al. 2003). Developing effective therapies against fungi has been more difficult than for bacterial pathogens, given the eukaryotic biology they share with humans; as a result, few effective antifungals are currently available. Most of the existing drugs have serious side effects, and resistance to these compounds is an increasing problem (Georgopapadakou 1998).

The analysis of the genomes of medically important fungi holds the potential to address these clinical issues. In particular, given the complete gene set for a pathogenic fungus, it becomes possible to predict genes necessary for fungal growth that lack human homologs. These may represent targets for antifungal drugs with fewer toxic side effects. This approach has been utilized to identify potential drug targets for *C. albicans* (Jones et al. 2004). Based on the human-curated gene set (see above), 228 genes were identified in the *C. albicans* genome sequence that were conserved in five other fully sequenced fungal genomes but that lacked significant sequence similarity in the human or mouse genomes (Braun et al. 2005). The authors suggested that, based on their predicted functions and localizations, these genes represent potential targets for small molecule inhibition.

The availability of complete genome sequence also facilitates genome-wide functional screens for drug targets. For example, De Backer and colleagues (De Backer et al. 2001) developed a method combining antisense RNA inhibition (see below) and promoter interference to identify genes critical for the growth of *C. albicans*, and subsequently used these genes as targets to identify new antifungals in a drug screen. The availability of an annotated genome sequence enabled the rapid identification of inhibited genes. Intriguingly, a significant fraction of *C. albicans* essential genes lacked homologs in *S. cerevisiae*, again highlighting the diversity of the fungal kingdom and the need for sequenced fungi beyond just a few models.

The growing complement of fungal genome sequences enables other strategies for investigating fungal infection. Comparing genomes from nonpathogenic species to related pathogenic organisms can identify genetic differences that contribute to infection and disease, while the comparison between strains with different host specificity may help clarify the genomic basis for differences in virulence and host interactions. Comparative analyses of these sorts are an exciting possibility arising from the sequencing of clusters of related genomes (as described above), often centering on a pathogenic fungus, but including related nonpathogenic fungi as in the case of *Coccidioides* spp. and *Uncinocarpus reesii*.

In addition to their role as pathogens, fungi also play a critical beneficial role in the development and production of pharmaceuticals through the production of secondary metabolites including Lovestatin and antibiotics such as penicillin, cephalosporins, and cyclosporine. The genomes of filamentous fungi have revealed an extensive—and occasionally unexpected—repertoire of secondary metabolites (Galagan et al. 2003; Kroken et al. 2003; Borkovich et al. 2004; Dean et al. 2005; Yu and Keller 2005). The burgeoning genomic resource available for fungi promises many further insights and discoveries into the friend and foe relationship between fungi and man.

The future of fungal genomics

The growing number of complete fungal genomes provides an unprecedented opportunity to study the biology and evolution of an entire eukaryotic kingdom. However, sequence is only the tip of the iceberg for fungal genomics. The availability of genome sequence has catalyzed the development of genome-wide functional studies for a growing number of fungal species. In particular, microarrays—both public and commercial—are available for numerous fungi, enabling not only expression studies, but also cross-genome hybridization, the identification of transcription-factor binding sites and chromatin modifications, and population genotyping. High-throughput proteomic methods are also increasingly being applied, providing insight into the protein modification and translational control. In addition, as highlighted above, comprehensive gene knock-out or knock-down projects are underway for several species. Ultimately, these data will enable a true systems biological approach to understanding fungal biology and evolution, and in particular the biology un-

derlying the widespread medical, agricultural, and environmental impact of fungi.

Acknowledgments

We are grateful to Jason Stajich for sharing his pre-publication results on intron evolution and the multigene phylogeny of sequenced fungi. This work was supported by grants from the NIH.

References

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297:** 1301–1310.

Aramayo, R. and Metzenberg, R.L. 1996. Meiotic transvection in fungi. *Cell* **86:** 103–113.

Archer, D.B. and Dyer, P.S. 2004. From genomics to post-genomics in *Aspergillus*. *Curr. Opin. Microbiol.* **7:** 499–504.

Ast, G. 2004. How did alternative splicing evolve? *Nat. Rev. Genet.* **5:** 773–782.

Barrass, J.D. and Beggs, J.D. 2003. Splicing goes global. *Trends Genet.* **19:** 295–298.

Batzoglou, S., Jaffe, D.B., Stanley, K., Butler, J., Gnerre, S., Mauceli, E., Berger, B., Mesirov, J.P., and Lander, E.S. 2002. ARACHNE: A whole-genome shotgun assembler. *Genome Res.* **12:** 177–189.

Beck-Sague, C. and Jarvis, W.R. 1993. Secular trends in the epidemiology of nosocomial fungal infections in the United States, 1980–1990. National Nosocomial Infections Surveillance System. *J. Infect. Dis.* **167:** 1247–1251.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L., and Zoran, M.J. 2005. Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nat. Rev. Genet.* **6:** 544–556.

Berbee, M.L. and Taylor, J.W. 2000. *The Mycota, vol VIIB, systematics and evolution*. Springer-Verlag, New York.

Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. 1985. Ty elements transpose through an RNA intermediate. *Cell* **40:** 491–500.

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuveglise, C., Munsterkotter, M., Guldener, U., Mewes, H.W., Van Helden, J., Dujon, B., et al. 2003. Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31:** 1121–1135.

Borkovich, K.A., Alex, L.A., Yarden, O., Freitag, M., Turner, G.E., Read, N.D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., et al. 2004. Lessons from the genome sequence of *Neurospora crassa*: Tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* **68:** 1–108.

Braun, B.R., van Het Hoog, M., d'Enfert, C., Martchenko, M., Dungan, J., Kuo, A., Inglis, D.O., Uhl, M.A., Hogues, H., Berriman, M., et al. 2005. A human-curated annotation of the *Candida albicans* genome. *PLoS Genet.* **1:** E1.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., et al. 2002. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419:** 512–519.

Chiller, T.M., Galgiani, J.N., and Stevens, D.A. 2003. Coccidioidomycosis. *Infect. Dis. Clin. North Am.* **17:** 41–57, viii.

Coghlan, A. and Wolfe, K.H. 2002. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.* **12:** 857–867.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. 1996. Transgene silencing of the al-1 gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO J.* **15:** 3153–3163.

Dasbach, E.J., Davies, G.M., and Teutsch, S.M. 2000. Burden of aspergillosis-related hospitalizations in the United States. *Clin. Infect. Dis.* **31:** 1524–1528.

Davis, C.A., Grate, L., Spingola, M., and Ares Jr., M. 2000. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* **28:** 1700–1706.

De Backer, M.D., Nelissen, B., Logghe, M., Viaene, J., Loonen, I., Vandoninck, S., de Hoogt, R., Dewaele, S., Simons, F.A., Verhasselt, P., et al. 2001. An antisense-based functional genomics approach for identification of genes critical for growth of *Candida albicans*. *Nat. Biotechnol.* **19:** 235–241.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H., et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434:** 980–986.

Dear, P.H. and Cook, P.R. 1989. Happy mapping: A proposal for linkage mapping the human genome. *Nucleic Acids Res.* **17:** 6795–6807.

———. 1993. Happy mapping: Linkage mapping using a physical analogue of meiosis. *Nucleic Acids Res.* **21:** 13–20.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., et al. 2004. Genome evolution in yeasts. *Nature* **430:** 35–44.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **99:** 16144–16149.

Ebbole, D.J., Jin, Y., Thon, M., Pan, H., Bhattarai, E., Thomas, T., and Dean, R. 2004. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: Analysis of expressed sequence tags. *Mol. Plant Microbe Interact.* **17:** 1337–1347.

Farman, M.L. and Leong, S.A. 1995. Genetic and physical mapping of telomeres in the rice blast fungus, *Magnaporthe grisea*. *Genetics* **140:** 479–492.

Fedorov, A., Roy, S., Fedorova, L., and Gilbert, W. 2003. Mystery of intron gain. *Genome Res.* **13:** 2236–2241.

Fink, G.R. 1987. Pseudogenes in yeast? *Cell* **49:** 5–6.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391:** 806–811.

Fraser, J.A. and Heitman, J. 2004. Evolution of fungal sex chromosomes. *Mol. Microbiol.* **51:** 299–306.

Frey, S.D., Elliott, E.T., and Paustian, K. 1999. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. *Soil Biol. Biochem.* **31:** 573–585.

Galagan, J.E. and Selker, E.U. 2004. RIP: The evolutionary cost of genome defense. *Trends Genet.* **20:** 417–423.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422:** 859–868.

Galagan, J.E., Calvo, S., Cuomo, C., Ma, L.J., Wortman, J., Batzoglou, S., Lee, S.I., Basturkmen, M., Spevak, C., Clutterbuck, J., et al. 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*. (in press)

Gehrig, H., Schussler, A., and Kluge, M. 1996. Geosiphon pyriforme, a fungus forming endocytobiosis with Nostoc (cyanobacteria), is an ancestral member of the Glomales: Evidence by SSU rRNA analysis. *J. Mol. Evol.* **43:** 71–81.

Georgopapadakou, N.H. 1998. Antifungals: Mechanism of action and resistance, established and novel drugs. *Curr. Opin. Microbiol.* **1:** 547–557.

Gilbert, W. 1978. Why genes in pieces? *Nature* **271:** 501.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274:** 546, 563–567.

Grate, L. and Ares Jr., M. 2002. Searching yeast intron data at Ares lab Web site. *Methods Enzymol.* **350:** 380–392.

Gudlaugsson, O., Gillespie, S., Lee, K., Vande Berg, J., Hu, J., Messer, S., Herwaldt, L., Pfaller, M., and Diekema, D. 2003. Attributable mortality of nosocomial candidemia, revisited. *Clin. Infect. Dis.* **37:** 1172–1177.

Guigo, R., Knudsen, S., Drake, N., and Smith, T. 1992. Prediction of gene structure. *J. Mol. Biol.* **226:** 141–157.

Hawksworth, D.L. 1991. The fungal dimension of biodiversity—magnitude, significance, and conservation. *Mycol. Res.* **95:** 641.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* **293:** 1129–1133.

Hoang, L.M., Maguire, J.A., Doyle, P., Fyfe, M., and Roscoe, D.L. 2004. *Cryptococcus neoformans* infections at Vancouver Hospital and Health Sciences Centre (1997–2002): Epidemiology, microbiology and histopathology. *J. Med. Microbiol.* **53:** 935–940.

Hutvagner, G. and Zamore, P.D. 2002. RNAi: Nature abhors a double-strand. *Curr. Opin. Genet. Dev.* **12:** 225–232.

Huynen, M.A., Snel, B., and Bork, P. 2001. Inversions and the dynamics of eukaryotic gene order. *Trends Genet.* **17:** 304–306.

Jaffe, D.B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov,

J.P., Zody, M.C., and Lander, E.S. 2003. Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13:** 91–96.
Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302:** 2141–2144.
Jones, T., Federspiel, N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T., et al. 2004. The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci.* **101:** 7329–7334.
Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., et al. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14:** 331–342.
Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423:** 241–254.
Kellis, M., Birren, B.W., and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428:** 617–624.
Kempken, F. and Kuck, U. 1998. Transposons in filamentous fungi—facts and perspectives. *Bioessays* **20:** 652–659.
Ketting, R.F. and Plasterk, R.H. 2000. A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* **404:** 296–298.
Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinform.* **5:** 59.
Korf, I., Flicek, P., Duan, D., and Brent, M.R. 2001. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17:** S140–S148.
Koszul, R., Caburet, S., Dujon, B., and Fischer, G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23:** 234–243.
Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., and Turgeon, B.G. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci.* **100:** 15670–15675.
Lephart, P.R., Chibana, H., and Magee, P.T. 2005. Effect of the major repeat sequence on chromosome loss in *Candida albicans*. *Eukaryot. Cell* **4:** 733–741.
Li, W., Rehmeyer, C.J., Staben, C., and Farman, M.L. 2005. TERMINUS—Telomeric end-read mining in unassembled sequences. *Bioinformatics* **21:** 1695–1698.
Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A., et al. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307:** 1321–1324.
Logsdon Jr., J.M., Stoltzfus, A., and Doolittle, W.F. 1998. Molecular evolution: Recent cases of spliceosomal intron gain? *Curr. Biol.* **8:** R560–R563.
Lowy, B. 1971. New records of mushroom stones from Guatemala. *Mycologia* **63:** 983–993.
Lynch, M. and Richardson, A.O. 2002. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12:** 701–710.
Madden, L.V. and Wheelis, M. 2003. The threat of plant pathogens as weapons against U.S. crops. *Annu. Rev. Phytopathol.* **41:** 155–176.
Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., et al. 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* **22:** 695–700.
Metting, F.B. 1993. Structure and physiological ecology of soil microbial communities. In *Soil microbial ecology: Applications in agricultural and environmental management*, pp. 3–25. Dekker, New York.
Modrek, B. and Lee, C. 2002. A genomic view of alternative splicing. *Nat. Genet.* **30:** 13–19.
Moss, M.O. 1987. Fungal biotechnology roundup. *Mycologist* **21:** 55–58.
Mourier, T. and Jeffares, D.C. 2003. Eukaryotic intron loss. *Science* **300:** 1393.
Mullikin, J.C. and Ning, Z. 2003. The phusion assembler. *Genome Res.* **13:** 81–90.
Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., et al. 2000. A whole-genome assembly of *Drosophila*. *Science* **287:** 2196–2204.
Napoli, C., Lemieux, C., and Jorgensen, R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2:** 279–289.
Nelson, M.A. 2004. EST evidence for alternative splicing in *Neurospora crassa*. In *Annual Neurospora Convention*.
Newell, S.Y. 1992. Estimating fungal biomass and productivity in decomposing litter. In *The fungal community: Its organization and role in the ecosystem* (eds. G.C. Carroll and D.T. Wicklow), pp. 521–561. Dekker, New York.
Nielsen, C.B., Friedman, B., Birren, B., Burge, C.B., and Galagan, J.E. 2004. Patterns of intron gain and loss in fungi. *PLoS Biol.* **2:** E422.
Odds, F.C., Brown, A.J., and Gow, N.A. 2004. *Candida albicans* genome sequence: A platform for genomics in the absence of genetics. *Genome Biol.* **5:** 230.
Parra, G., Blanco, E., and Guigo, R. 2000. GeneID in *Drosophila*. *Genome Res.* **10:** 511–515.
Rensberger, B. 1992. The Iceman: Now the research is on ice. *J. NIIT Res.* **4:** 25–27.
Romfo, C.M., Alvarez, C.J., van Heeckeren, W.J., Webb, C.J., and Wise, J.A. 2000. Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **20:** 7955–7970.
Roy, S.W. and Gilbert, W. 2005a. Complex early genes. *Proc. Natl. Acad. Sci.* **102:** 1986–1991.
———. 2005b. The pattern of intron loss. *Proc. Natl. Acad. Sci.* **102:** 713–718.
Ryan, M.J. and Smith, D. 2004. Fungal genetic resource centres and the genomic challenge. *Mycol. Res.* **108:** 1351–1362.
Salamov, A.A. and Solovyev, V.V. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10:** 516–522.
Salzberg, S.L., Pertea, M., Delcher, A.L., Gardner, M.J., and Tettelin, H. 1999. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59:** 24–31.
Sambrook, J. 1977. Adenovirus amazes at Cold Spring Harbor. *Nature* **268:** 101–104.
Schechtman, M.G. 1990. Characterization of telomere DNA from *Neurospora crassa*. *Gene* **88:** 159–165.
Selker, E.U. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* **24:** 579–613.
———. 2002. Repeat-induced gene silencing in fungi. *Adv. Genet.* **46:** 439–450.
Seoighe, C., Federspiel, N., Jones, T., Hansen, N., Bivolarovic, V., Surzycki, R., Tamse, R., Komp, C., Huizar, L., Davis, R.W., et al. 2000. Prevalence of small inversions in yeast gene order evolution. *Proc. Natl. Acad. Sci.* **97:** 14433–14437.
Shiu, P.K. and Metzenberg, R.L. 2002. Meiotic silencing by unpaired DNA: Properties, regulation and suppression. *Genetics* **161:** 1483–1495.
Shiu, P.K., Raju, N.B., Zickler, D., and Metzenberg, R.L. 2001. Meiotic silencing by unpaired DNA. *Cell* **107:** 905–916.
Slater, G.S. and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **6:** 31.
Stanke, M. and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33:** W465–W467.
Stanke, M. and Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19:** II215–II225.
Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. 2004. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32:** W309–W312.
Strange, R.N. and Scott, P.R. 2005. PLANT DISEASE: A threat to global food security. *Annu. Rev. Phytopathol.* **43:** 83–116.
Swartz, M.N. 1994. Hospital-acquired infections: Diseases with increasingly limited therapies. *Proc. Natl. Acad. Sci.* **91:** 2420–2427.
Taylor, T.N., Hass, H., and Kerp, H. 1999. The oldest fossil ascomycetes. *Nature* **399:** 648.
Tenney, A.E., Brown, R.H., Vaske, C., Lodge, J.K., Doering, T.L., and Brent, M.R. 2004. Gene prediction and verification in a compact genome with numerous small introns. *Genome Res.* **14:** 2330–2335.
Toth, G., Gaspari, Z., and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **10:** 967–981.
Trappe, J.M. 1987. Phylogenetic and ecologic aspects of mycotrophy in the angiosperms from an evolutionary standpoint. In *Ecophysiology of VA mycrorrhizal plants* (ed. G.R. Safir), pp. 5–25. CRC Press, Boca Raton, FL.
Veneault-Fourrey, C. and Talbot, N.J. 2005. Moving toward a systems biology approach to the study of fungal pathogenesis in the rice blast fungus *Magnaporthe grisea*. *Adv. Appl. Microbiol.* **57:** 177–215.
Vinson, J.P., Jaffe, D.B., O'Neill, K., Karlsson, E.K., Stange-Thomann, N., Anderson, S., Mesirov, J.P., Satoh, N., Satou, Y., Nusbaum, C., et al. 2005. Assembly of polymorphic genomes: Algorithms and

application to *Ciona savignyi*. *Genome Res.* **15:** 1127–1135.
Wolfe, K.H. and Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387:** 708–713.
Wood, V.R., Gwilliam, M.A., Rajandream, M., Lyne, R., Lyne, A., Stewart, J., Sgouros, N., Peat, J., Hayles, S., Baker, D., et al. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415:** 871–880.
Xu, J., Vilgalys, R., and Mitchell, T.G. 2000. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol. Ecol.* **9:** 1471–1481.
Yu, J.H. and Keller, N. 2005. Regulation of secondary metabolism in filamentous fungi. *Annu. Rev. Phytopathol.* **43:** 437–458.
Zeigler, R.S., Leong, S.A., and Teeng, P.S. 1994. *Rice Blast Disease*. CAB International, Wallingford, UK.
Zhou, S., Kile, A., Kvikstad, E., Bechner, M., Severin, J., Forrest, D., Runnheim, R., Churas, C., Anantharaman, T.S., Myler, P., et al. 2004. Shotgun optical mapping of the entire Leishmania major Friedlin genome. *Mol. Biochem. Parasitol.* **138:** 97–106.

Web site references

http://fungal.genome.duke.edu/; Duke Fungal Genomics Resource.
http://genolist.pasteur.fr/CandidaDB/; CandidaDB.
http://img.jgi.doe.gov/pub/main.cgi; JGI Integrated Microbial Resource Database.
http://mips.gsf.de/projects/fungi/; the Munich Information Center for Protein Sequences.
http://www.broad.mit.edu/annotation/fgi/; The Broad Institute Fungal Genome Initiative web site.
http://www.ebi.ac.uk/Wise2/index.html; GeneWise program.
http://www.fgsc.net/; Fungal Genetics Stock Center.
http://www.genomesonline.org/; The Genomes Online database.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj; the TIGR fungal database (www.tigr.org/tdb/fungal), NCBI Entrez.
http://www.neurotransmitter.net/metadb/; the Munich Information Center for Protein Sequences.
http://www.yeastgenome.org/; Saccharomyces Genome Database.

Erratum

Genome Research 15: 1620–1631 (2005)

Genomics of the fungal kingdom: Insights into eukaryotic biology

James E. Galagan, Matthew R. Henn, Li-Jun Ma, Christina A. Cuomo, and Bruce Birren

The authors wish to acknowledge that the phylogeny of the fungal kingdom presented in Figure 1 makes use of data obtained from Berbee and Taylor (2001).

Berbee, M.L. and Taylor, J.W. 2001. Fungal molecular evolution: Gene trees and geologic time. In *The Mycota Vol. VIIB, systematics and evolution* (eds. D.J. McLaughlin et al.), pp. 229–246. Springer, Berlin.

16:304 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org