Title
Toward genome-enabled mycology.

Permalink
https://escholarship.org/uc/item/5f1159jc

Journal
Mycologia, 105(6)

ISSN
0027-5514

Authors
Hibbett, DS
Stajich, JE
Spatafora, JW

Publication Date
2013

DOI
10.3852/13-196

License
CC BY-NC-ND 4.0

Peer reviewed
Toward genome-enabled mycology

David S. Hibbett
Biology Department, Clark University, Worcester, Massachusetts 01610

Jason E. Stajich
Department of Plant Pathology and Microbiology, University of California, Riverside, California 92521

Joseph W. Spatafora
Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331

Abstract: Genome-enabled mycology is a rapidly expanding field that is characterized by the pervasive use of genome-scale data and associated computational tools in all aspects of fungal biology. Genome-enabled mycology is integrative and often requires teams of researchers with diverse skills in organismal mycology, bioinformatics and molecular biology. This issue of Mycologia presents the first complete fungal genomes in the history of the journal, reflecting the ongoing transformation of mycology into a genome-enabled science. Here we consider the prospects for genome-enabled mycology and the technical and social challenges that will need to be overcome to grow the database of complete fungal genomes and enable all fungal biologists to make use of the new data.

Key words: bioinformatics, fungi, genomics, phylogeny, Polyporales

INTRODUCTION

To understand the functional biology and evolutionary relationships of Fungi, mycologists have readily embraced new technologies. In the 1960s, for example, fungal biologists enthusiastically adopted electron microscopy, which allowed unprecedented resolution of subcellular characters that informed both cell biology and systematics (Bracker 1967). Publications on fungal ultrastructure appeared frequently throughout the 1970s and 1980s (Kimborough 1981, Mims 1991), but in the 1990s they were largely replaced by a flood of phylogenetic analyses based on the revolutionary new technique of PCR (Fig. 1). Today another mycological revolution is taking place, one based on fungal genomes. This issue of Mycologia includes the first publication of new genomes in the history of the journal (Binder et al. 2013), reflecting the ongoing transformation of mycology into a genome-enabled science. Here we consider the effect of genomics on mycology and the adjustments that will be necessary to let all fungal biologists benefit from the new data.

Fungal genomes accumulated slowly at first but now are being produced at an exponential rate. The first genome of a fungus (or any eukaryote), Saccharomyces cerevisiae, was published in 1996 (Goffeau et al. 1996), followed by Schizosaccharomyces pombe (Wood et al. 2002) and Neurospora crassa (Galagan et al. 2003), while the first basidiomycete genome was that of the model white-rot fungus Phanerochaete chrysosporium (Martinez et al. 2004). These early species were targeted because of their importance as experimental organisms, but genomics is no longer reserved for model systems. In 2012 alone the Joint Genome Institute (JGI) of the US Department of Energy completed whole-genome sequences of more than 50 fungal species (http://genome.jgi.doe.gov/programs/fungi/index.jsf) (Fig. 1) and GenBank (http://www.ncbi.nlm.nih.gov/genome/brow/) currently houses 417 fungal genomes representing 267 species, but this is just the beginning. The 1000 Fungal Genomes (1KFG) project, which was initiated in 2011 (http://1000.fungalgenomes.org), aims to sequence representatives of approximately two genera from each of the roughly 656 recognized families of Fungi (Kirk et al. 2008), and the Genomes Online Database (GOLD) reports 2364 fungal genome projects completed or in progress, representing more than 40% of all eukaryotic genome projects (http://www.genomesonline.org/cgi-bin/GOLD/phylogenetic_distribution.cgi).

The current issue of Mycologia marks a small (but we think significant) step toward genome-enabled mycology. Binder et al. (Binder et al. 2013) report whole genomes of three species of Polyporales (Agaricomycetes): Bjerkandera adusta, Phlebia brevispora and Ganoderma lucidum, which were sequenced and annotated by the JGI. Binder et al. is focused primarily on phylogenetics, while a set of companion papers illustrates the potential for combining genomics and phylogenetics with functional biology of Fungi. A major motivation for genomic studies of Polyporales is their effect on the carbon cycle through wood decay (Grigoriev et al. 2011). Accordingly, the companion papers emphasize gene families encoding enzymes that are important in decay, including carbohydrate-active enzymes (Hori et al. 2013),

Submitted 21 Jun 2013; accepted for publication 4 Jul 2013.

1 Corresponding author. E-mail: dhibbett@clarku.edu
lignin-degrading peroxidases (Ruiz-Duenas et al. 2013) and cytochrome P450s (Yadav et al. 2013).

Polyporales exhibit both bipolar (single locus) and tetrapolar (two-locus) mating systems, which Gilbertson suggested might be correlated with decay mode (Gilbertson 1980, 1981). It is fitting, therefore, that one of the companion papers in this issue (James et al. 2013) reconstructs mechanisms of transformations between bipolar and tetrapolar mating systems in Polyporales and other Agaricomycetes and evaluates their possible relationship to decay type. Other functionally significant gene families treated in the companion articles include ABC transporters (Kovalchuk et al. 2013b), cerato-platanins (Chen et al. 2013), hydrophobins (Mgbeahuruieke et al. 2013) and proteins associated with RNA interference (RNAi) (Yang et al. 2013). Finally, a phylogenetic study by (Ortiz-Santana et al. 2013) analyzes evolutionary relationships in the “Antrodia clade” of brown-rot Polyporales, using ribosomal RNA genes representing 123 species in 26 genera.

The common element uniting all papers in this issue is the application of phylogenetic methods. Still, it would be premature to say that phylogenetics and

**Fig. 1.** Numbers of papers in Mycologia focused on electron microscopy (top) or phylogenetics (middle; based on a survey of the title field in searches of ISI-cited article for Mycologia 1945–2013) and fungal genomes deposited in GenBank as of Jun 2013, beginning with *S. cerevisiae* in 1996 (arrow, bottom panel). After the publication of the PCR (arrow, middle panel) molecular phylogenetics replaced ultrastructural analyses. The asterisk corresponds to the publication of the “Deep Hypha” issue of Mycologia (Blackwell et al. 2006).
genomics have been fully integrated in fungal biology. To be sure, there have been advances in resolving the backbone of the fungal phylogeny using genomic data (Fitzpatrick et al. 2006, Robbertse et al. 2006, Liu et al. 2008, Liu et al. 2009, Medina et al. 2011, Robbertse et al. 2011, Ebersberger et al. 2012, Floudas et al. 2012, Padamsee et al. 2012, Binder et al. 2013), but detailed knowledge about the fine branches of the fungal tree of life are still based almost entirely on phylogenetic analyses of one or a few genes sampled in many species, as in the analysis of Ortiz-Santana et al. (2013). Similarly, most of the phylogenies in the companion articles illustrate the relationships within gene families, but for the most part they have not been reconciled with the organismal phylogeny—the studies of Yang et al. (2013) and Ruiz-Duenas et al. (2013) are exceptions—and therefore they do not explicitly reconstruct the patterns of gene duplication and loss that could be associated with functional diversification. Below, we consider the prospects for a new genome-enabled mycology, highlighting selected recent studies that illustrate how genomic, phylogenetic and functional perspectives can be integrated with applications to all branches of fungal biology.

ENVISIONING GENOME-ENABLED MYCOLOGY

Genome-enabled mycology will be characterized by the pervasive use of genome-scale data and associated computational tools in descriptive and hypothesis-driven investigations in all aspects of fungal biology, including genetics, cell and molecular biology, physiology, ecology, conservation biology, systematics, symbiosis and pathology. The transformation of mycology into a genome-enabled science is being accelerated by advances in DNA sequencing technology and by more accessible and distributed high-end computing infrastructures. Genome sequencing currently is restricted largely to Fungi in pure culture, but single-cell genomics (Stepanauskas 2012) and bioinformatics approaches to filtering mixtures of genomes are enabling analyses of obligate symbionts (Vogel and Moran 2013) as well as complex environmental samples. It is not unreasonable to think that in the near future mycologists will routinely sequence the complete genomes of isolates that are the subjects of their research, much as ITS sequences are now used to confirm identifications.

The traditional “organismal” branches of mycology, including ecology, systematics and population biology, already have been profoundly influenced by genomics. The transformation of fungal ecology into a molecular discipline actually began in the 1990s, when Sanger sequencing of PCR amplicons started to be used to identify environmental samples (Horton and Bruns 2001). Much more detailed profiles of fungal communities are now routinely obtained with next-generation sequencing, yielding surprising reports of diverse Fungi from unusual habitats, such as corals and deep-sea hydrothermal vents (Le Calvez et al. 2009, Jones and Richards 2011, Amend et al. 2012). A rich taxonomic catalog alone, however, does not provide an understanding of the functions of Fungi in their habitats. As more robustly annotated genomes become available, backed by functional and biochemical analyses, it will become easier to incorporate metatranscriptomic and metaproteomic approaches in environmental surveys (Damon et al. 2012, Seifert et al. 2013), which will provide insights into the ecological services provided by individual species as well as entire communities (Damon et al. 2011, Aylward et al. 2013, Takasaki et al. 2013).

One area of mycology that has readily adopted genome-scale data is phylogenetics (Robbertse et al. 2006, Liu et al. 2008, Liu et al. 2009, Robbertse et al. 2011, Floudas et al. 2012, Padamsee et al. 2012). Systematics and genomics are complementary because they share the application of phylogenetic methods and principles. Moreover phylogenetic hypotheses are used to guide selection of species for sequencing, as in the 1KFG project. Thanks to the application of taxonomy-based sampling strategies, it is all but certain that in the near future there will be richly sampled, genome-based phylogenetic trees for virtually all groups of Fungi, with most clades supported with high confidence. At the same time, it has become clear that even with genome-scale datasets there are certain nodes in the fungal tree for which alternate resolutions cannot be rejected (Medina et al. 2011). Indeed, with the availability of genome-scale data, considerations of statistical consistency in phylogenetic methods (Felsenstein 1988) have become even more important, in that model misspecification or the inclusion of fast-evolving sites and genes may result in erroneous nodes being resolved with high confidence (Lartillot et al. 2007, Rodríguez-Ezepeleta et al. 2007, Philippe et al. 2011). The discussion so far has emphasized analyses of sequence data, but whole genomes also are making it possible to discover new kinds of characters to test or develop phylogenetic hypotheses, such as gene fusions or rearrangements (Lee et al. 2010, Slot and Bruns 2001). Probabilistic models for these kinds of characters are now needed to analyze such data using maximum likelihood or Bayesian approaches.

The availability of genome-scale data has raised awareness of conflicts between gene trees and species trees, due to incomplete lineage sorting, horizontal
gene transfer and patterns of gene duplication and loss (Rannala and Yang 2008, Salichos and Rokas 2013). Standard molecular systematics practice seeks to identify single-copy genes that presumably track the organismal phylogeny. In a genome-enabled framework, we envision that fungal systematists will not make a priori determinations of which genes are appropriate for phylogenetic inference but instead will screen for gene tree/species tree conflicts across diverse gene families as a routine part of a phylogenomic analysis. The output of such analyses will include reconciliations of patterns of gene duplication and loss (Durand et al. 2006, Vernot et al. 2008, Górecki et al. 2011), as well as automated identification of genes and chromosomes that might have been transferred horizontally (Richards et al. 2011). Visualization of complex histories, such as the simultaneous tracking of large numbers of loci through speciation, will be a challenge for implementing these approaches.

In genome-enabled mycology, functional biology will be seamlessly integrated with ecology, systematics and population biology. This integration has been driven for some time by the movement of fungal organismal biologists toward molecular data, which have been widely used in the post-PCR era to address classical questions in ecology and evolution. To complete the integration, however, it also will be necessary for mechanistically oriented fungal biologists to adopt a comparative perspective. Much of the progress in biology since the modern synthesis has been achieved by reducing biological phenomena to their most basic and fundamental units, studied in model systems. Tremendous gains have been and will continue to be made through studies focused on model systems, such as S. cerevisiae (Hinnebusch and Johnston 2011), N. crassa (Lambrechts et al. 2009), Aspergillus spp. (Galagan et al. 2005, Pel et al. 2007), Coprinopsis cinerea (Stajich et al. 2010) and Schizophyllum commune (Ohm et al. 2010). Nevertheless, the growing availability of complete genomes is driving mycology in a more integrative direction, in which comparative analyses become the primary sources of functional hypotheses that can be further evaluated in tractable model organisms.

Functional insights may be obtained from genome-based comparative analyses from clades to populations. At a macro-evolutionary scale, phylogenomic analyses will not reconstruct patterns of organismal relationships only but also will localize the origins of genetic diversity, such as that due to gene family expansion and horizontal transmission, that may be correlated with functional innovations and ecological shifts (Richards et al. 2006, Slot and Rokas 2011, Campbell et al. 2012, Floudas et al. 2012). Similarly comparative transcriptomic data, evaluated in a phylogenetic context, will provide insights into the functional bases of evolutionary transformations in diverse non-model systems, for which genetic manipulations may or may not be possible (Dunn et al. 2013). At the level of populations, genomewide association studies (McCarthy et al. 2008) based on single nucleotide polymorphisms (SNPs) will provide the ability to characterize variation at the nucleotide level across the genome, with much greater granularity than traditional QTL approaches. Such analyses are already revolutionizing population genetic studies of fungi, providing insight into mechanisms of adaptation and population divergence (Cowen et al. 2002, Tsai et al. 2008, Neafsey et al. 2010, Ellison et al. 2011, Louis 2011).

Comparative phylogenomic and population-genomic analyses can be used to address the genetic bases of any fungal phenotype. So far, much of the emphasis in fungal comparative genomics has concerned the diversity of fungal nutritional modes, metabolic capabilities, host ranges and mechanisms of pathogenesis (Martin et al. 2010, Spanu et al. 2010, Eastwood et al. 2011, Fernandez-Fueyo et al. 2012, Floudas et al. 2012, Morin et al. 2012, O’Connell et al. 2012, Vogel and Moran 2013). Such a focus is understandable, given the profound affect that Fungi have on natural and agricultural ecosystems and human health and their potential as industrial organisms. Another area that we expect will benefit in the near term is the study of developmental evolution (evo-devo). The stunning diversity of morphology across the Fungi (Hibbett 2007, Petersen 2012) begs to be explained by the development of novel genomic components and unique regulation and regulatory networks of pre-existing genes and gene families (Stajich et al. 2010, Taylor Ellison 2010, Jedd 2011).

**Achieving genome-enabled mycology**

The studies cited above show that the transformation toward genome-enabled mycology is well underway. However, to realize the full potential of genomics for fungal biology it will be necessary to overcome diverse obstacles, including technical barriers, problems associated with acquisition of biological material and social challenges, such as the need for training, networking and development of community standards for data sharing.

While the sequencing of fungal genomes has become routine, most available fungal genomes are not closed or sequences of complete chromosomes but are draft assemblies comprising scaffolds of varying lengths. The inability to close genomes is
partially due to the inability of “short-read” sequencing systems (e.g., those based on the Illumina platform) to assemble repetitive intergenic regions. The growing length of sequence reads in short-read systems, now reaching 250 bp, and improvements in assembly software have dramatically improved our ability to generate more complete genomes, however, and there are numerous emerging “long-read” technologies (e.g., the Pacific Biosciences RSII system) that are capable of producing sequences of several thousand bases that stand to significantly alleviate problems in genome assembly (Niedringhaus et al. 2011, Chin et al. 2013, Eisenstein 2013).

Even with longer sequence reads, genome assembly will remain a non-trivial task, requiring bioinformatics resources that are not available in most mycological laboratories. Computer speed and memory requirements are a limitation for assembly, although there are new ways to mitigate this requirement by reducing the redundancy of the data (Pell et al. 2012). Creation of high quality assemblies also requires expertise in assembly software and techniques. Additional general bioinformatics protocols using open source tools still need to be developed if the promise of high quality sequence assembly is to come to individual laboratories. Genome annotation is also a major challenge that requires computational infrastructure and specific bioinformatics expertise, but it is enabled by freely available software and protocols that have supported successful sequencing projects and teams (Haas et al. 2008, 2011; Holt and Yandell 2011; Yandell and Ence 2012).

Some of the hardest problems in fungal genomics are those associated with complex environmental samples. Longer sequence reads will help and bioinformatics approaches are being developed to deal with mixtures of genomes (Keeling et al. 2010, Vogel and Moran 2013), but the incomplete documentation of fungal diversity will remain a major constraint. Roughly 100,000 species of Fungi have been described (Kirk et al. 2008), but only about a quarter of these (26,540 species) are represented by a sequence of any locus, and only 267 species are represented by a genome sequence in GenBank (although this is increasing rapidly). These numbers pale in comparison to estimates of actual fungal diversity, which range from hundreds of thousands to many millions of species (Blackwell 2011, Hawksworth 2012). For the foreseeable future, fungal metagenomic and metatranscriptomic studies will have to contend with huge numbers of unidentified genes and organisms. The ongoing 1KFG project and fungal barcoding efforts (Schoch et al. 2012) are critical for fungal environmental genomics because they will provide annotated reference genomes from diverse branches of the fungal phylogeny, as well as taxonomic sequence databases for identification purposes.

One of the limitations of the 1KFG project and fungal genomics in general is the need to obtain nucleic acids for all the target species. Indeed, acquisition of material for sequencing has proven to be a major bottleneck for fungal genomics. Similarly, barcoding efforts are limited by the need to obtain high quality, reliably identified specimens from which DNA can be isolated. Ultimately, fungal genomics, including environmental studies, depends on taxonomic experts to collect and identify species in nature, deposit vouchers in herbaria and (ideally) bring the organisms into pure culture. Thus, the growth of fungal genomics should result in greater support for basic biodiversity surveys and collections infrastructure, coupled with the development of taxonomic and molecular databases (Hibbett et al. 2011, Voigt and Kirk 2011).

The rise of genome-enabled mycology necessitates a new set of research skills. Just as fungal systematists of the late 20th century had to become proficient in basic techniques of molecular biology and phylogenetic analyses, today’s mycologists must become adept at techniques associated with information sciences, including management of large datasets and databases, use of platforms formerly restricted to computer science (e.g., UNIX) and the application of fundamental computer programming (e.g., Perl, Python, R) to solve context-dependent analytical issues. To put it another way, the keyboard has joined the pipet and the microscope among the essential resources in the mycologist’s toolkit.

Increased access to training in computational methods could help prevent the division of fungal biology into bioinformatics haves and have-nots, and some sequencing centers, such as the JGI, have been proactive in providing online tutorials, annotation jamborees and other training opportunities. Nevertheless, it is difficult for individual laboratories to gather all the expertise and resources necessary to execute all aspects of genome-enabled research projects, which often incorporate ancillary techniques such as light and electron microscopy, pathogenicity tests and biochemical assays. Genome-enabled mycology is by its nature a team-driven enterprise, with studies commonly having dozens of authors. Basic fungal biologists have important roles to play in these teams because they are uniquely qualified to propose compelling questions based on deep knowledge of fungal diversity, ecology, evolutionary relationships, development and functional biology.

A final class of challenges in genome-enabled mycology concerns data storage and access, including
strategies for distributing resources and policies for data use and publication. Most of the currently available fungal genomes have been produced by major sequencing centers. The benefits of such centralized resources are obvious, including the ability to provide diverse informatics services to users, such as FungiDB (Stajich et al. 2012), Mycocosm (Grigoriev et al. 2012), the Saccharomyces Genome Database (Cherry et al. 2012) and CFGP (Choi et al. 2013). However, as sequencing technology becomes more widespread and as focused communities of data users assemble, there also may be an increase in specialized genome collections, organized around particular taxonomic groups, diseases or habitats (e.g. marine Fungi or the human mycobiome). Already there is a partitioning of sequencing efforts, with the DOE-financed JGI focusing on species with relevance to energy and the carbon cycle and the Broad Institute focused principally on medically important species as a NIAID Genome Sequencing Center. As more data centers arise, it will be important to ensure that they are able to exchange information easily so that researchers can take advantage of all available genomic information.

One significant difference between genome-enabled mycology and traditional fungal biology concerns the timing of data release relative to publication. For example, fungal systematists typically deposit sequences into GenBank only after their analyses have been accepted for publication, but publicly financed genome centers have an obligation to make genome sequences available almost immediately upon completion (the waiting period for public genome release upon completion at genome centers such as JGI and the Broad Institute is 1 mo). Current community standards for data access are based on the Fort Lauderdale (Anonymous 2003) and Toronto (Birney 2009) agreements, which emphasized the shared responsibilities among data producers, researchers, supporting agencies, and editors and reviewers to ensure rapid release and fair use of data. A central message is that communication is crucial. Communities involved in data generation need to broadcast the goals of their projects, through Web pages, white papers or other means, and recruit interested parties to join in research collaborations and project planning. Data providers also have a responsibility to publish findings in a timely fashion so the data that have been made public can be more freely incorporated into other research projects. Conversely prospective data users should respect the legitimate interests of data providers to publish their findings and they should contact principal investigators to discuss appropriate use and potential conflicts as well as opportunities for collaboration. Reviewers and editors have a responsibility to confirm that community standards of data use are upheld.

As the number of fungal genomes skyrockets, there may be a tendency for some would-be data users to regard unpublished fungal genomes as cheap. The reality is that generation of material for sequencing is still difficult and time-consuming, work that is often performed by graduate students and postdoctoral fellows. This work will become even more challenging, as rare or recalcitrant species and exotic habitats are targeted. Faced with the threat of being scooped with data that they helped generate, some researchers might choose to negotiate project-specific memoranda of understanding or forego collaboration with publicly financed sequencing centers, so as to maintain control over data before publication. Unauthorized data use threatens to undermine what has been a remarkably successful model for generating a rich and expanding resource of fungal genomes and the collaborative spirit of mycological research.

ACKNOWLEDGMENTS

We thank all authors who contributed articles to the polypore phylogenomics section in this issue of Mycologia, and Editor-in-Chief Jeffrey K. Stone for his support of the project.

LITERATURE CITED


