

## Fungal Genomes and Beyond

Nemat O. Keyhani\*

Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611

The sequencing of fungal genomes is advancing at breakneck-speed, producing voluminous amounts of data. Within the next five years, it is possible that over a couple thousand genomes, representing every major fungal family will be completed and available to the scientific community. In order for this data to have a truly transformative effect on mycological and other research, however, several factors need to be addressed. These include; (1) the establishment of user-friendly platforms for examining, sorting, and sifting through the genomes, (2) integration, or at least cross-communication, between the various databases that house the genomic data, and (3) investment in community resources that can act as repositories for and provide materials to researchers, i.e. strains, clones, plasmids, etc. The frameworks for some these needs, e.g. the materials available from the Fungal Genetics Stock Center (FGSC, University of Missouri), are already established and should be reinforced, whereas for others, e.g. data accessibility, the sooner that a plan can be implemented the better.

The Fungal Kingdom is considered to contribute greater than 15% of the species richness found in the major groups of organisms [1]. Displaying a wide range of facultative to obligate saprophytic, symbiotic, and pathogenic lifestyles, fungi have colonized all terrestrial ecosystems. Often noticeable in soils only in their fruiting body form, a largely hidden under- or inaccessible world is populated by fungal organisms performing ecological functions that have significant impacts on plant, animal, and even other microbial communities. Fungal contenders exist for our planet's largest and amongst the oldest organisms, i.e. *Armillaria solidipes* also known as the honey mushroom found in the Malheur National Forest in Oregon, estimated at being between 2000-9,000 years old, spans over several thousand acres. In all, the current total number of fungal species is estimated at over 1.5 million with a little over 100,000 described to date. The described fungi have been placed into approximately 140 orders and over 550 families.

Due to this diversity and their biological and biotechnological importance the sequencing of fungal genomes is well underway [2]. Including yeasts, the total number of fungal genomes sequenced or near completion to date is close to 500 (Genomes Online Database, GOLD, <http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi>). Although the available fungal genomes are perhaps somewhat biased towards fungi of medical importance, several new whole-genome sequencing initiatives are likely to provide robust and more taxonomically diverse datasets. The Fungal Genome Initiative at the Broad Institute currently houses over 50 fungal genomes, mainly from the Ascomycota, and includes model organisms such as *Neurospora* and *Aspergillus* species, but also extends to the genomes of several basidiomycetes, chytrids, and at least one mucormycete (<http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative>). The Mycorrhizal Genome Initiative, designed to sequence and analyze the genomes of ascomycetes and basidiomycetes that function within the context of symbiosis in woody shrubs and trees, has selected 28 species for whole genome sequencing (<http://mycor.nancy.inra.fr/IMG/Mycogenomes/index.html>). These data, coupled to the sequencing of a "suite of fungal decayers", including brown and white rot fungi will provide the genomic framework for

understanding the biology of these organisms and potentially exploiting them in applications ranging from reforestation and maintenance of ecological biodiversity to bioenergy production.

Perhaps the most significant near-term development in fungal genomics is the recent DOE-JGI funding of the 1000 Fungal Genomes Project, which seeks to sequence and assemble a thousand fungal genomes within the next five years (<http://1000.fungalgenomes.org/home/>). The objectives of this initiative are to sequence at least two representatives from every family or family-level clade of Fungi, thus providing an exceptionally diverse sampling of fungi. These data would benefit scientists across a wide spectrum of research and allow for a level of comparative genomics far deeper than that possible today. Additional international as well as single fungal genome projects exist. At least 23 fungal genomes have been completed or are in progress in various Chinese laboratories and several of these have already been published [3]. It is unclear, however, whether these data (including a number of individual fungal genomes sequenced within the US and throughout the world) have been integrated into any of the databases listed above, a critical omission that needs to be addressed.

The impacts of the availability of fungal genomes will depend on their accessibility to the larger scientific community, particularly to those that may not have in-depth expertise in bioinformatics. The National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) continues to be one of the major veins by which researchers can access fungal genomic data, and many of the tools available from NCBI address some of the desired functions listed below. However, not all fungal genomes appear to be equally accessible or available and no facile mechanism exists for corrections and annotations. Further investment in the development of platforms that have increased user-friendly interfaces is essential. This should include not only the ability to BLAST the genomes and/or locate a gene, but also to easily perform a range of other functions such as build (genomically) localized comparative analyses between organisms, find neighboring genes, identify potential regulatory regions, and link genes within biochemical and functional frameworks. An example of such a program, which should be incorporated especially as part of any large-scale fungal genome sequencing efforts, is the Comparative Fungal Genomics Platform (CFGP, <http://cfgp.riceblast.snu.ac.kr/main.php>). The tools available include easy to use links for analyzing, sorting, and storing genes of interest as well as links to DNA, RNA, and protein predictions and functions. A careful survey of the prospective needs of

\*Corresponding author: Nemat O. Keyhani, Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, E-mail: [keyhani@ufl.edu](mailto:keyhani@ufl.edu)

Received December 15, 2011; Accepted December 16, 2011; Published December 18, 2011

Citation: Keyhani NO (2011) Fungal Genomes and Beyond. Fungal Genom Biol 1:e101. doi:10.4172/2165-8056.1000e101

Copyright: © 2011 Keyhani NO. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

scientists before embarking on building upon these tools would seem to be prudent.

In addition, several specialized platforms already exist and can be used as templates for further expansion. The Saccharomyces Genome Database (<http://www.yeastgenome.org/>) is one of the most comprehensive of such efforts and includes a trove of tools for examining the data. These include features for querying and analyzing data along biochemical, developmental, and signaling pathways useful to a wide range of researchers without requiring detailed bioinformatic expertise. Extensive (literature) referencing for every gene that has been examined in terms of phenotype, expression, and/or biochemical characterization is invaluable. Additional databases particularly for model organisms also exist [4]. If the identified orthologous genes of the other extant fungal genomes could somehow be linked to these data, a significant contribution to facilitating annotation and research efforts would already be made. Another model to examine is The Arabidopsis Information Resource (TAIR, <http://www.arabidopsis.org/>), which provides a one-stop resource for the genome sequence, gene structure, product, and expression information, as well as DNA, seed stocks, plasmids, and references dealing with this model plant organism.

Despite their utility, significant efforts will be needed to apply these scaffoldsto such ambitious efforts as the 1000 Fungal Genome Project. Whether it would be better to design new interactive platforms essentially *de novo*, with smart features akin to gene-butlers with anticipatory algorithms and on-demand reference linking, or to build upon already established ones is a discussion that should be held given the vast amounts of data that are coming in the pipeline. For example, obtaining the sequence data itself can progress quite rapidly, however, high quality annotation of the genomes to a useful level, is likely to be a significant bottleneck. Regarding this latter step, the ongoing development of a system or technology (referred to as The SEED, [http://www.theseed.org/wiki/Home\\_of\\_the\\_SEED](http://www.theseed.org/wiki/Home_of_the_SEED)) for high-throughput,

high-quality annotation of genomes may be a model worth applying. The idea of the SEED project is to have researchers (not necessarily in bioinformatics) annotate single (or multiple) subsystems within their expertise over a collection of genomes rather than a single genome. This is in contrast to current approaches in which teams, which often have little expertise in the majority of the specific genes that they annotate, analyze whole genomes one at a time; a procedure that can result in significant levels of erroneous annotation. Unless a SEED-like project is coupled to the sequencing efforts, the annotation of the genomes may be cursory and hence limited in use to the wider scientific community.

Five years from now we may well have at our disposal the genomic sequences of several thousand fungi, across all the major families of the Fungal Kingdom. The extent to which these data can revolutionize the field will depend less upon the sequences themselves but more on the tools available for researchers to mine these genomes. The development of an integrated, learning-enabled, expert driven annotation and referenced web-based system, all with a user friendly, intuitive interface and linked to community available repository services, may seem to be a rather tall order. However, investment in building such an infrastructure would be critical for managing the potential data overload and stepping beyond the fungal genomes to yield resources capable of benefiting and amplifying both basic and applied research.

#### References

1. Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405: 212-219.
2. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, et al. (2011) Fueling the future with genomics. *Mycology* 2: 192-209.
3. An Z, Wang C, Liu X, Bennett JW (2010) China's fungal genomics initiative: a whitepaper. *Mycology* 1: 1-8.
4. Some additional databases include the *Candida* Genome database (<http://www.candidagenome.org/>), the *Aspergillus fumigatus* Genome database (<http://aspergillusgenome.org/>) the Wellcome-Trust Sanger Fungal Genomes database (<http://www.sanger.ac.uk/resources/downloads/fungi/>), as well as tools available at <http://fungalgenomes.org>.