MycoCosm portal: gearing up for 1000 fungal genomes

Igor V. Grigoriev*, Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik Nordberg, Inna Dubchak and Igor Shabalov

DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598

December 2013

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Nucleic Acids Research Advance Access published December 1, 2013

Nucleic Acids Research, 2013, 1–6
doi:10.1093/nar/gkt1183

MycoCosm portal: gearing up for 1000 fungal genomes

Igor V. Grigoriev*, Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik Nordberg, Inna Dubchak and Igor Shabalov

US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Received October 9, 2013; Revised October 31, 2013; Accepted November 1, 2013

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013

ABSTRACT

MycoCosm is a fungal genomics portal (http://jgi.doe.gov/fungi), developed by the US Department of Energy Joint Genome Institute to support integration, analysis and dissemination of fungal genome sequences and other 'omics' data by providing interactive web-based tools. MycoCosm also promotes and facilitates user community participation through the nomination of new species of fungi for sequencing, and the annotation and analysis of resulting data. By efficiently filling gaps in the Fungal Tree of Life, MycoCosm will help address important problems associated with energy and the environment, taking advantage of growing fungal genomics resources.

INTRODUCTION

Genomics is a powerful tool for understanding the genomic basis of biological processes in individual organisms and their communities. The Fungal Genomics Program (http://jgi.doe.gov/fungi) of the US Department of Energy (DOE) Joint Genome Institute (JGI) has partnered with the international scientific community to support genomic exploration of the kingdom Fungi and to help address problems related to energy and the environment. The exponential growth of data resulting from several large-scale genomics initiatives like the 1000 Fungal Genomes Project requires new approaches for data integration and analysis (1). We designed MycoCosm, an integrated fungal genomics resource, to allow researchers to easily access large amounts of genomics data and analyze them using web-based analytical tools. The MycoCosm portal enables in-depth multidimensional analysis of individual genomes and efficient comparative genomics of fungi, which may be applied to phylogenetically related fungi, or to those sharing the same lifestyle, ecological niche or phenotypic trait.

The first JGI fungal portal was created for the white rot fungus *Phanerochaete chrysosporium* in 2004 (2). To enable comparative genome analysis within the same analytical framework, in 2008 JGI integrated data from six genomes of Dothideomycete plant pathogens (three sequenced by JGI and three others provided by the Dothideomycetes Genome Consortium). This served as a prototype for MycoCosm, the first official version of which was released in March 2010 in response to a call for a pan-fungal genomic resource from the community of fungal biologists brought together by the Burroughs Wellcome Fund (Alexandria, VA). By 2011, MycoCosm included over 100 fungal genomes (3), and continues to grow. Here we provide an update on the latest developments of MycoCosm, which includes flexible tools for the analysis of exponentially growing genomics and other 'omics' data sets, and extends user community participation to the entire genome project cycle, from nomination of species for sequencing through genome annotation and comparative genomics.

Significantly larger and better-integrated genomics data

Since our last report (3), MycoCosm has grown to host >250 publicly available fungal genomes (Figure 1), 80% of which were sequenced by JGI and 20% by other groups added for comparative purposes. Individual genomes, or biologically relevant groups of them, may be searched using keywords or sequence, investigated using MycoCosm analytical tools, and downloaded for custom analyses.

Also in the past two years, all MycoCosm nodes have been populated with newly sequenced and annotated genomes including the recently activated nodes for Glomeromycota, Entomophthoromycotina, Kickxellomycotina, Blastocladiomycota, Neocallimastigomycota and others. Several new branches such as Xylonomycetes, Zoopagomycotina and Cryptomycota have been added for a more complete picture of the Fungal Tree of Life (Figure 2). Several model fungi have been added to MycoCosm,

*To whom correspondence should be addressed. Tel: +1 925 296 5860; Email: ivgrigoriev@lbl.gov

Published by Oxford University Press 2013. This work is written by US Government employees and is in the public domain in the US.

for which genes are linked to the corresponding records in curated databases such as the Saccharomyces Genome Database (4), the Aspergillus Genome Database (5) and others. GenBank records for JGI-sequenced fungal genomes are linked to MycoCosm using "JGIDB" qualifiers (http://www.ncbi.nlm.nih.gov/genbank/collab/db_xref/).

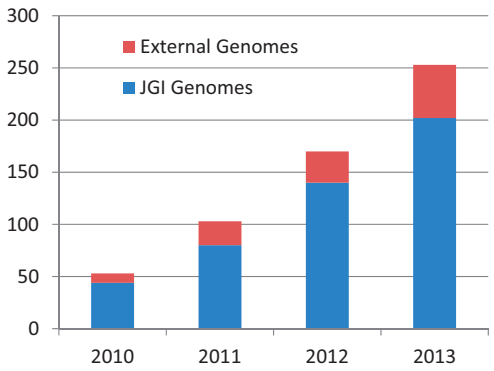

Figure 1. Growth of annotated genomes in MycoCosm. The genomes sequenced by JGI are shown in blue and those sequenced by others are shown in red.

Each genome in MycoCosm has its own portal where genomics and other 'omics' data can be explored in depth, in a single-genome context (Figure 3). Genomes in MycoCosm are also grouped for comparative genomics analyses (Figure 4), into two kinds of groups: phylogenetic and ecological. Phylogenetic groups are represented (and linked to) by the nodes of the MycoCosm tree, and facilitate comparative analyses of organisms with a shared evolutionary history. Ecological groups represent a curation effort in which fungi of similar lifestyles (e.g. endophytes, mycorrhizal symbionts, plant pathgens, etc.) have been assembled into groups that are linked to the chapters of the Genomics Encyclopedia of Fungi (6).

To encourage the research community to participate in the further growth of MycoCosm, users can now nominate diverse fungal species for sequencing (Figure 5) in the context of the 1000 Fungal Genomes Project (http://1000.fungalgenomes.org). Nominated species will be peer reviewed, and, if approved, will be sequenced, annotated and fully integrated into MycoCosm's comparative genomics framework.

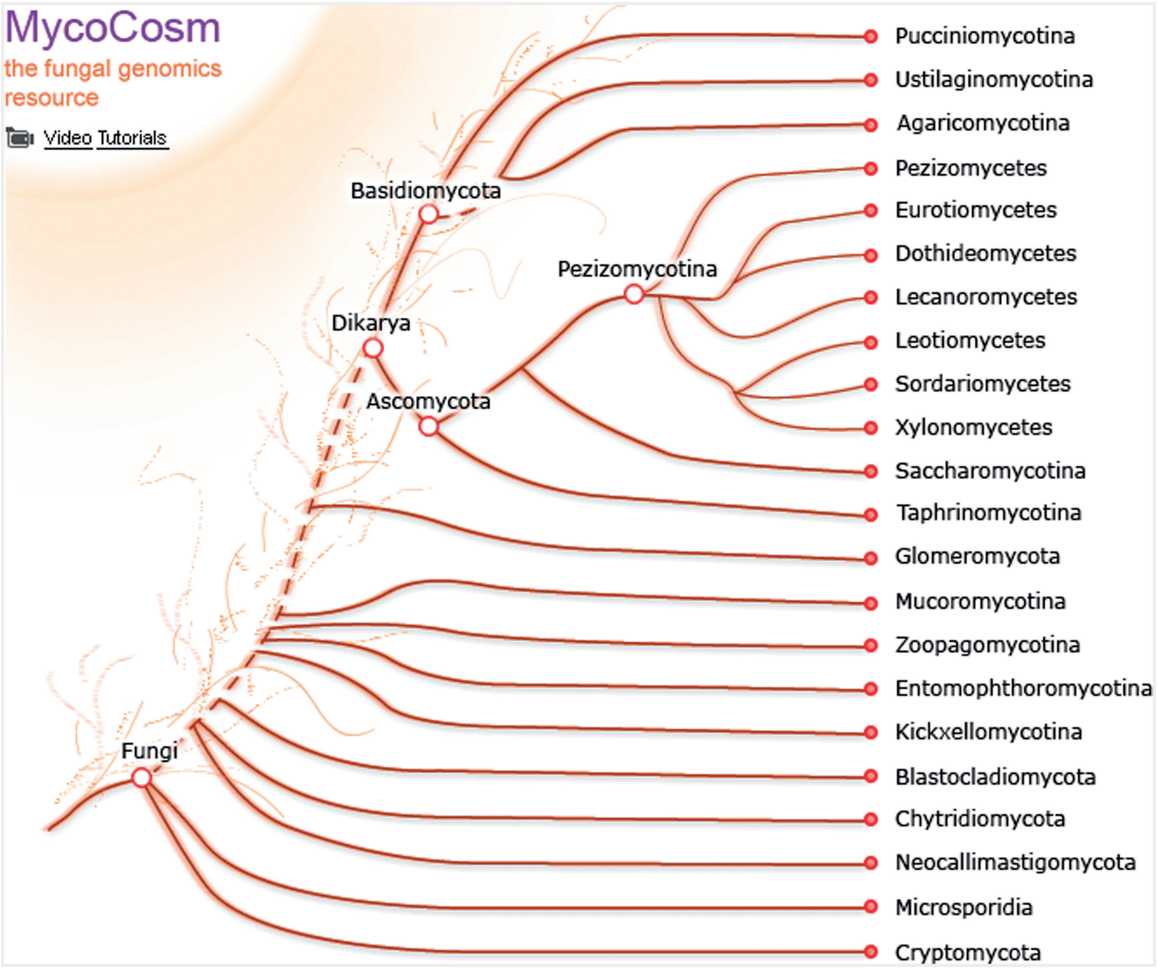

Figure 2. MycoCosm front page. The MycoCosm tree illustrates relationships between major groups of fungi. Each node of the tree is linked to the corresponding group page and portals for individual genomes within the group.

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013

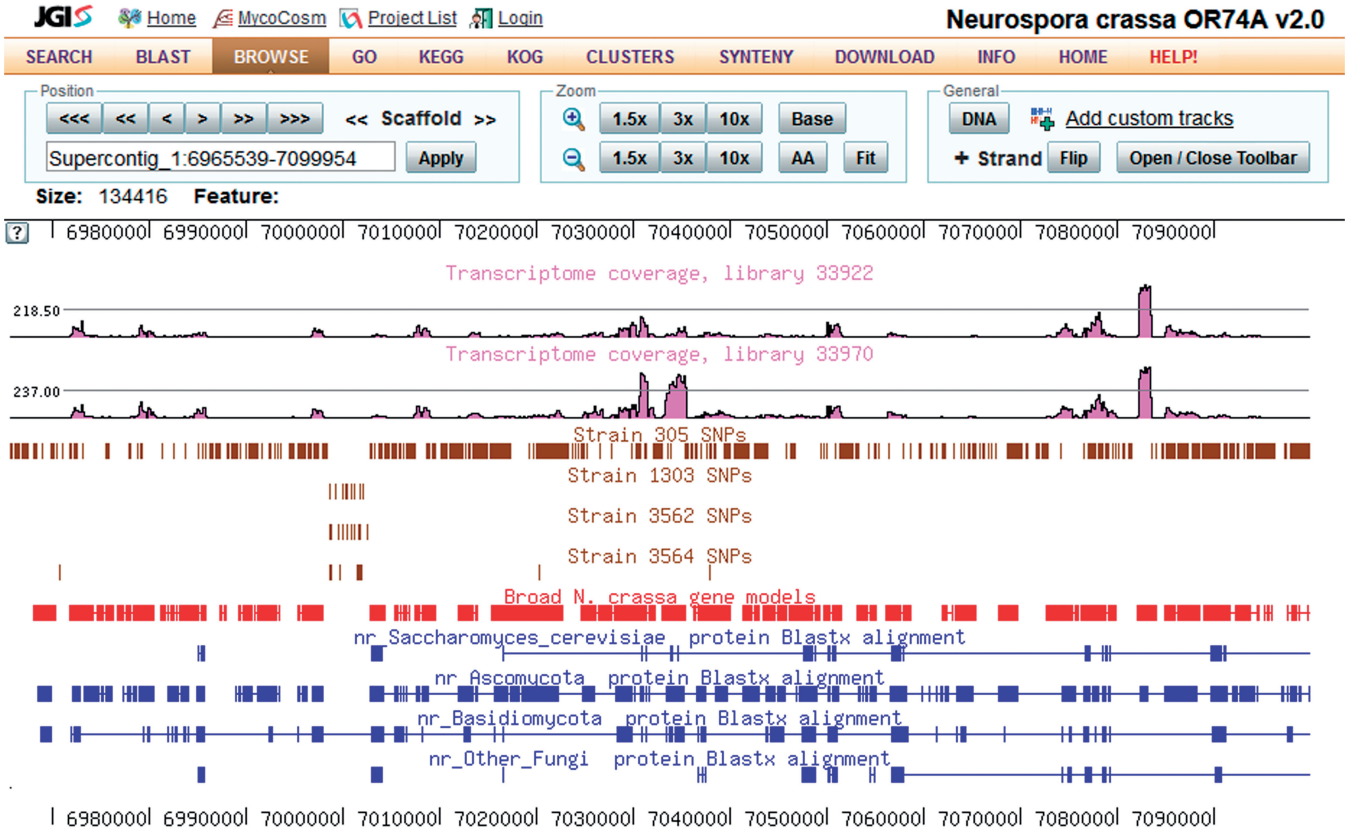

Figure 3. Genome Browser for *Neurospora crassa*. Horizontal tracks represent genomic features of a given assembly scaffold. Red bars indicate *N. crassa* genes predicted by the Broad Institute and uploaded into MycoCosm. The four tracks with brown bars illustrate SNPs derived from resequencing of mutant strains of *N. crassa* (7). The two pink curves represent transcriptome profiling using RNA-Seq during *N. crassa* growth under two different conditions. Additional tracks can be displayed and configured using the toolbar.

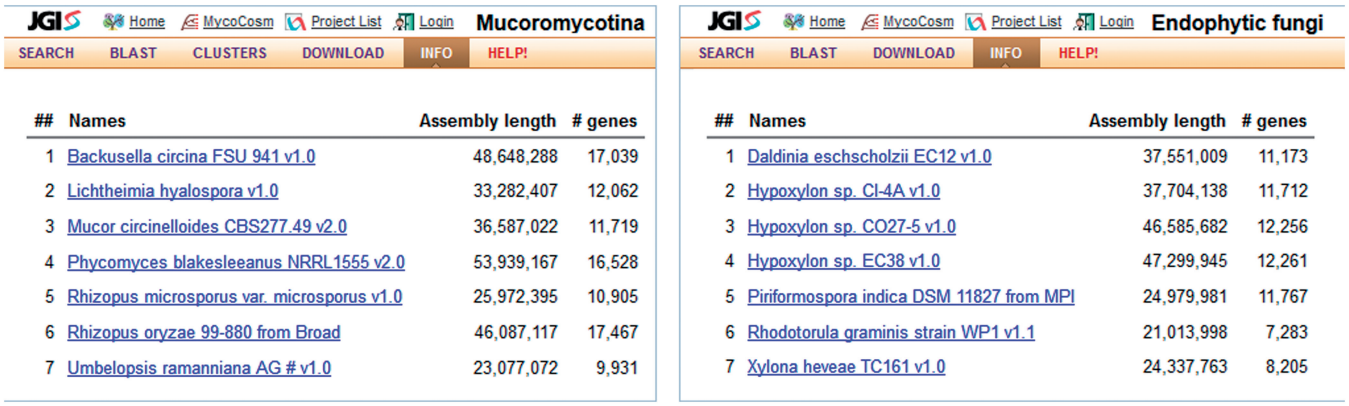

Figure 4. Examples of MycoCosm group pages. A PhyloGroup containing seven annotated genomes of fungi from Mucoromycotina is shown on the left. An EcoGroup with seven annotated genomes of phylogenetically diverse endophytes is shown on the right.

Uniform genome annotation

For genome comparison to produce meaningful results, application of the same set of functional annotation databases to all genomes is required. For genomes sequenced and annotated outside of JGI, the existing gene models are fed into the functional annotation part of the pipeline, using the same databases used for annotation of JGI genomes, to ensure consistent references for genome comparisons.

All fungi sequenced at JGI are annotated using the JGI Annotation Pipeline, which integrates an array of tools for gene prediction, annotation and analysis (8). Using repeat-masked genome assemblies as input, protein-coding gene models are predicted using a combination of gene predictors of three general kinds: *ab initio*, homology-based and transcriptome-based. The *ab initio* Fgenesh (9) and GeneMark (10) are specifically trained for the given genome. Homology-based Fgenesh+ (9) and Genewise

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013

Phylum name	Subphylum name	Class name	Order name	Family name	GOLD Projects		
Ascomycota	Pezizomycotina	Arthoniomycetes	Arthoniales	Arthoniaceae	None sequenced	Nominate	Add to GOLD
Ascomycota	Pezizomycotina	Arthoniomycetes	Arthoniales	Chrysothricaceae	None sequenced	Nominate	Add to GOLD
Ascomycota	Pezizomycotina	Arthoniomycetes	Arthoniales	Roccellaceae	None sequenced	Nominate	Add to GOLD
Ascomycota	Pezizomycotina	Arthoniomycetes	Incertae sedis	Melaspileaceae	None sequenced	Nominate	Add to GOLD
Ascomycota	Pezizomycotina	Dothideomycetes	Acrospermales	Acrospermaceae	None sequenced * Acrospermum compressum * Oomyces carneoalbus	Nominate Manfred Binder Manfred Binder	Add to GOLD CBS KNAW CBS KNAW
Ascomycota	Pezizomycotina	Dothideomycetes	Botryosphaeriales	Bagnisiellaceae	None sequenced	Nominate	Add to GOLD
Ascomycota	Pezizomycotina	Dothideomycetes	Botryosphaeriales	Botryosphaeriaceae	Aplosporella prunicola Botryosphaeria dothidea	Whole Genome Whole Genome	targeted Complete
Ascomycota	Pezizomycotina	Dothideomycetes	Botryosphaeriales	Guignardiaceae	Phyllosticta citricarpa	Nominate	Add to GOLD

Figure 5. 1000 Fungal Genomes nomination page. Over 500 families of fungi and associated genome projects are displayed. Green and yellow rows highlight families with no, or only one, sequenced genome, where additional species can be nominated for sequencing via the use of the 'Nominate' button. Nomination requires user registration and commitment to provide JGI with DNA and RNA samples for sequencing.

(11) are seeded by BLASTx (12) alignments of fungal sequences in NCBI's nonredundant protein database. Transcriptome-based models are predicted from assembled RNA-sequencing data derived from the cultured fungus. For each genomic locus, the best representative gene model is selected based on a combination of protein homology and transcriptome support; these make up a 'gene catalog'. In addition to protein-coding genes, tRNAs are predicted using tRNAscan-SE (13).

All predicted proteins are then functionally annotated. SignalP (14) is used to detect the sequence motifs responsible for protein localization, TMHMM (15) identifies possible transmembrane domains and InterProScan (16) predicts functional domains from Pfam (17) and other databases. Protein alignments to the NCBI's nonredundant (http://www.ncbi.nlm.nih.gov), Swiss-Prot (18), KEGG (19) and KOG (20) databases additionally facilitate functional interpretation. Interpro, KEGG and Swiss-Prot hits are used to map gene ontology (GO) terms (21). Detailed annotation reports from these methods are provided on every coding gene's protein page. Additionally, summaries listing numbers of genes by category in the GO, KEGG and KOG classifications are accessible from the portal menu, and can be compared with other selected genomes to explore gene family expansions and contractions across genomes.

Predicted gene models, along with different lines of supporting evidence including transcriptomics and similarity to proteins in other genomes, are displayed on a genome browser (Figure 3). The browser, which is based on the UCSC Genome Browser (22), displays genomic features as horizontal tracks, which can be customized to the needs of each user. Individual elements from these tracks are linked to the corresponding pages for additional details. For example, predicted gene models are linked to annotation reports, while genome conservation tracks are linked to VISTA Point (23). Two new types of data, single-nucleotide polymorphisms (SNPs) and gene expression profiles, have been added to the browser, to enable the visualization of resequencing and transcriptomics data, respectively (Figure 3). These are important additions to the resources developed, for example, for model fungi such as the ascomycete *Neurospora crassa* (7) or the basidiomycete *Schizophyllum commune* (24).

Predicted gene models, and their annotations, can be improved through the process of manual curation, which is open to all authorized users and coordinated by the principal investigator of each project. Such collective efforts were critical for the analysis of several fungal genomes (24–31). Additionally, users can upload custom tracks to the browser and share them, either with specified collaborators or globally. This capability to upload data for individual genes or entire genomes, makes MycoCosm a powerful and customizable data exchange resource for the community.

Grouping genomes for comparative analyses

The wealth of genomes in MycoCosm creates an opportunity to compare groups of genomes, and the need for efficient tools to enable these comparisons. Mycocosm therefore includes tools that integrate single genomes into a comparative context, such as the ability to visualize variation in gene counts in different GO, KEGG and KOG categories across a user-selected assortment of genomes. Additionally, MycoCosm includes predefined groups of fungi to enable users to ask questions for fungal genomes in a group context. Two types of genome groups are available in MycoCosm and shown in Figure 4: (i) PhyloGroups, consisting of phylogenetically related species (corresponding to the nodes of the MycoCosm tree), and (ii) EcoGroups, containing fungi of similar lifestyle or ecology (but which may be phylogenetically distant).

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013

Both kinds of group offer consistent interfaces for comparative analyses. Genome sizes and gene counts may be compared on the Info page (Figure 4), giving a coarse-grained overview. Group-specific BLAST, search and download of genomes and annotations are available. Genes shared between genomes within a group can be explored using precomputed Markov clustering (32) of the full set of protein sequences in the group. Derived from all-versus-all BLAST alignments of the proteins, these gene clusters can be thought of as protein 'families'. The clusters can be filtered to show genes shared between all, or some, organisms in the group, or to find genes unique to them as described in details in help pages. Each cluster is linked to a page listing all the genes of the cluster with graphical display of protein domains, intron–exon gene structure and local synteny of cluster members. A larger genome-scale synteny calculated by VISTA (23) is available from the Synteny menu of individual genome portals, where dot-plots, for example, can illustrate the full spectrum of micro-, meso- and macro-synteny between different pairs of Dothideomycete species (31). This phenomenon is limited to this class of fungi and illustrates the utility of the group-centric approach for comparative genome analysis.

User nomination of new species for sequencing

Even though all major nodes of the MycoCosm tree are represented by at least one sequenced genome, many fungal clades remain unrepresented. The 250+ annotated genomes in MycoCosm are a small fraction of the total number of fungal species, estimates of which range from 1.5 million to as high as 5.1 million (33,34). Thus, sampling of the Fungal Tree of Life by genomics is sparse, and likely has left many biologically interesting clades unexplored. This undersampling will surely limit our attempts to understand functional diversity and evolution through fungal genomes, and will hamper metagenomics studies of fungal communities, where a broad array of reference genomes is required to interpret complex sequence data.

To encourage the scientific community to participate in determining what gaps in the Fungal Tree of Life will be filled in next, every node of MycoCosm is now equipped with a 'Nominate' link to enable users to recommend new genomes for sequencing via the 1000 Fungal Genomes Project (http://1000.fungalgenomes.org). Following this link from a given node, families of fungi with no, or only one, sequenced genomes (Figure 5, green and yellow rows) display a 'Nominate' button, which invites users to describe the species, access JGI User Guidelines and provide DNA and RNA samples to JGI for sequencing (if approved by peer review). The 1000 Fungal Genomes Project aims to have at least two sequenced genomes for every family of fungi. Through this and other JGI Community Sequencing Projects (35), we aim to fill the remaining gaps in the Fungal Tree of Life.

To enable visibility of ongoing JGI genome projects, the Projects menu of MycoCosm (http://jgi.doe.gov/projects) lists each project's status, principal investigator and other pertinent information. Combined with other genome projects listed in the GOLD database (36) and other sources, they form a 'master' list of fungal genome sequencing projects worldwide (http://1.usa.gov/JGI-1000-Fungi). This 'master' list helps us better focus our sequencing efforts on unexplored regions of the Fungal Tree of Life. Thus, we integrate a full genome project cycle (from sequencing target selection through assembly and annotation, to biology-driven data analysis) into MycoCosm enabling users to be active contributors to this process.

FUTURE PLANS

Since the release of its first fungal genome 10 years ago, the number of fungal genomes sequenced by JGI has jumped to >200. Collecting, integrating and visualizing these genomes, and those in sequencing pipeline, for efficient comparative analysis requires new approaches and tools. Especially, functional genomics will quickly reveal putative functions of a large number of genes identified only through genome sequencing (as opposed to smaller-scale biochemical and genetic studies). New interactive tools for transcriptomics, epigenomics, population genomics and metagenomics will therefore be developed and integrated into MycoCosm, in support of data analysis and utilization, and hypothesis development.

ACKNOWLEDGEMENTS

We thank MycoCosm users for valuable feedback, members of JGI portal group for development of software modules used in MycoCosm and Byoungnam Min for technical assistance.

FUNDING

Director, Office of Science, Office of Biological and Environmental Research, Life Sciences Division, US Department of Energy [DE-AC02-05CH11231]. Funding for open access charge: US Department of Energy Contract [No. DE-AC02-05CH11231].

Conflict of interest statement. None declared.

REFERENCES

1. Grigoriev,I.V. (2013) A changing landscape of fungal genomics. In: Martin,F. (ed.), *The Ecological Genomics of Fungi*. John Wiley & Sons, Inc, Hoboken, NJ, pp. 1–20.
2. 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.
3. Grigoriev,I.V., Nordberg,H., Shabalov,I., Aerts,A., Cantor,M., Goodstein,D., Kuo,A., Minovitsky,S., Nikitin,R., Ohm,R.A. *et al.* (2012) The genome portal of the department of energy joint genome institute. *Nucleic Acids Res.*, **40**, D26–D32.
4. Cherry,J.M., Hong,E.L., Amundsen,C., Balakrishnan,R., Binkley,G., Chan,E.T., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R. *et al.* (2012) *Saccharomyces* genome

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013

database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.
5. Arnaud,M.B., Cerqueira,G.C., Inglis,D.O., Skrzypek,M.S., Binkley,J., Chibucos,M.C., Crabtree,J., Howarth,C., Orvis,J., Shah,P. *et al.* (2012) The *Aspergillus* genome database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic Acids Res.*, **40**, D653–D659.
6. Grigoriev,I.V., Cullen,D., Goodwin,S.B., Hibbett,D., Jeffries,T.W., Kubicek,C.P., Kuske,C., Magnuson,J.K., Martin,F., Spatafora,J.W. *et al.* (2011) Fueling the future with fungal genomics. *Mycology*, **2**, 192–209.
7. McCluskey,K., Wiest,A.E., Grigoriev,I.V., Lipzen,A., Martin,J., Schackwitz,W. and Baker,S.E. (2011) Rediscovery by whole genome sequencing: classical mutations and genome polymorphisms in *Neurospora crassa*. *G3 (Bethesda)*, **1**, 303–316.
8. Grigoriev,I.V., Martinez,D.A. and Salamov,A.A. (2006) Fungal genomic annotation. In: Arora,D.K., Berke,R.A. and Singh,G.B. (eds), *Applied Mycology and Biotechnology*, Vol. 6. Elsevier, Amsterdam, pp. 123–142.
9. Salamov,A.A. and Solovyev,V.V. (2000) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.*, **10**, 516–522.
10. Ter-Hovhannisyan,V., Lomsadze,A., Chernoff,Y.O. and Borodovsky,M. (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.*, **18**, 1979–1990.
11. Birney,E., Clamp,M. and Durbin,R. (2004) GeneWise and genomewise. *Genome Res.*, **14**, 988–995.
12. Boratyn,G.M., Camacho,C., Cooper,P.S., Coulouris,G., Fong,A., Ma,N., Madden,T.L., Matten,W.T., McGinnis,S.D., Merezhuk,Y. *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.
13. Schattner,P., Brooks,A.N. and Lowe,T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*, **33**, W686–W689.
14. Petersen,T.N., Brunak,S., von Heijne,G. and Nielsen,H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
15. Krogh,A., Larsson,B., von Heijne,G. and Sonnhammer,E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
16. Hunter,S., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Das,U., Daugherty,L., Duquenne,L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **33**, W116–W120.
17. Punta,M., Coggill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Boursnell,C., Pang,N., Forslund,K., Ceric,G., Clements,J. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
18. The UniProt Consortium. (2013) Update on activities at the universal protein resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.
19. Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
20. Koonin,E.V., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Krylov,D.M., Makarova,K.S., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N., Rao,B.S. *et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.*, **5**, R7.
21. Blake,J.A., Dolan,M., Drabkin,H., Hill,D.P., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T., McCarthy,F. *et al.* (2013) Gene ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
22. Meyer,L.R., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Kuhn,R.M., Wong,M., Sloan,C.A., Rosenbloom,K.R., Roe,G., Rhead,B. *et al.* (2013) The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64–D69.
23. Frazer,K.A., Pachter,L., Poliakov,A., Rubin,E.M. and Dubchak,I. (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res.*, **32**, W273–W279.
24. Ohm,R.A., de Jong,J.F., Lugones,L.G., Aerts,A., Kothe,E., Stajich,J.E., de Vries,R.P., Record,E., Levasseur,A., Baker,S.E. *et al.* (2010) Genome sequence of the model mushroom *Schizophyllum commune*. *Nat. Biotechnol.*, **28**, 957–963.
25. Berka,R.M., Grigoriev,I.V., Otillar,R., Salamov,A., Grimwood,J., Reid,I., Ishmael,N., John,T., Darmond,C., Moisan,M.C. *et al.* (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat. Biotechnol.*, **29**, 922–927.
26. Duplessis,S., Cuomo,C.A., Lin,Y.C., Aerts,A., Tisserant,E., Veneault-Fourrey,C., Joly,D.L., Hacquard,S., Amselem,J., Cantarel,B.L. *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl Acad. Sci. USA*, **108**, 9166–9171.
27. Eastwood,D.C., Floudas,D., Binder,M., Majcherczyk,A., Schneider,P., Aerts,A., Asiegbu,F.O., Baker,S.E., Barry,K., Bendiksby,M. *et al.* (2011) The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science*, **333**, 762–765.
28. Floudas,D., Binder,M., Riley,R., Barry,K., Blanchette,R.A., Henrissat,B., Martinez,A.T., Otillar,R., Spatafora,J.W., Yadav,J.S. *et al.* (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, **336**, 1715–1719.
29. Goodwin,S.B., M'Barek,S.B., Dhillon,B., Wittenberg,A.H., Crane,C.F., Hane,J.K., Foster,A.J., Van der Lee,T.A., Grimwood,J., Aerts,A. *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.*, **7**, e1002070.
30. Martin,F., Aerts,A., Ahren,D., Brun,A., Danchin,E.G., Duchaussoy,F., Gibon,J., Kohler,A., Lindquist,E., Pereda,V. *et al.* (2008) The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, **452**, 88–92.
31. Ohm,R.A., Feau,N., Henrissat,B., Schoch,C.L., Horwitz,B.A., Barry,K.W., Condon,B.J., Copeland,A.C., Dhillon,B., Glaser,F. *et al.* (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes fungi*. *PLoS Pathog.*, **8**, e1003037.
32. Enright,A.J., Van Dongen,S. and Ouzounis,C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.
33. Hawksworth,D.L. (1991) The fungal dimension of biodiversity-magnitude, significance, and conservation. *Mycol. Res.*, **95**, 641–655.
34. Blackwell,M. (2011) The Fungi: 1, 2, 3. 5.1 million species? *Am. J. Bot.*, **98**, 426–438.
35. Martin,F., Cullen,D., Hibbett,D., Pisabarro,A., Spatafora,J.W., Baker,S.E. and Grigoriev,I.V. (2011) Sequencing the fungal tree of life. *New Phytologist*, **190**, 818–821.
36. Pagani,I., Liolios,K., Jansson,J., Chen,I.M.A., Smirnova,T., Nosrat,B., Markowitz,V.M. and Kyrpides,N.C. (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.*, **40**, D571–D579.

Downloaded from http://nar.oxfordjournals.org/ at University of California, Berkeley on December 27, 2013