Literature-Based Gene Curation and Proposed Genetic Nomenclature for Cryptococcus

Citation	Inglis, D. O., Skrzypek, M. S., Liaw, E., Moktali, V., Sherlock, G., & Stajich, J. E. (2014). Literature-based gene curation and a proposed genetic nomenclature for Cryptococcus. Eukaryotic Cell, 13(7), 878-883. doi:10.1128/EC.00083-14
DOI	10.1128/EC.00083-14
Publisher	American Society for Microbiology
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

EC
Journals.ASM.org

Literature-Based Gene Curation and Proposed Genetic Nomenclature for *Cryptococcus*

Diane O. Inglis,[a] Marek S. Skrzypek,[a] Edward Liaw,[b] Venkatesh Moktali,[b,c] Gavin Sherlock,[a] Jason E. Stajich[b]

Department of Genetics, Stanford University, Stanford, California, USA[a]; Department of Plant Pathology and Microbiology and Institute for Integrative Genome Biology, University of California—Riverside, Riverside, California, USA[b]; Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, USA[c]

***Cryptococcus*, a major cause of disseminated infections in immunocompromised patients, kills over 600,000 people per year worldwide. Genes involved in the virulence of the meningitis-causing fungus are being characterized at an increasing rate, and to date, at least 648 *Cryptococcus* gene names have been published. However, these data are scattered throughout the literature and are challenging to find. Furthermore, conflicts in locus identification exist, so that named genes have been subsequently published under new names or names associated with one locus have been used for another locus. To avoid these conflicts and to provide a central source of *Cryptococcus* gene information, we have collected all published *Cryptococcus* gene names from the scientific literature and associated them with standard *Cryptococcus* locus identifiers and have incorporated them into FungiDB (www.fungidb.org). FungiDB is a panfungal genome database that collects gene information and functional data and provides search tools for 61 species of fungi and oomycetes. We applied these published names to a manually curated ortholog set of all *Cryptococcus* species currently in FungiDB, including *Cryptococcus neoformans* var. *neoformans* strains JEC21 and B-3501A, *C. neoformans* var. *grubii* strain H99, and *Cryptococcus gattii* strains R265 and WM276, and have written brief descriptions of their functions. We also compiled a protocol for gene naming that summarizes guidelines proposed by members of the *Cryptococcus* research community. The centralization of genomic and literature-based information for *Cryptococcus* at FungiDB will help researchers communicate about genes of interest, such as those related to virulence, and will further facilitate research on the pathogen.**

Cryptococcus neoformans and *Cryptococcus gattii* are major causes of disseminated disease in immunocompromised patients, which can result in severe, often fatal, meningitis (1). *C. gattii* is a global pathogen, and its prevalence in the United States has risen; it is responsible for an increasing number of infections in the U.S. Pacific Northwest (2). Furthermore, recent reports suggest that cryptococcosis due to *C. gattii* is even more common in immunocompetent and AIDS patients than previously estimated (3, 4). *C. neoformans* also occurs worldwide and is a common complication in HIV-positive individuals and in AIDS patients (4). *C. neoformans* has emerged as a model organism for the study of pathogenesis, and molecular tools for transformation and gene deletion, protein localization with green fluorescent protein (GFP) (5), and conditional knockdowns with RNA interference (RNAi) (6) have enabled detailed analyses of links between gene function and molecular mechanisms of pathogenesis. The construction of a gene knockout collection is under way for *C. neoformans* var. *grubii* strain H99, and to date, more than 1,200 knockout strains have been deposited at the Fungal Genetics Stock Center (7). Both a community-developed microarray platform and transcriptome-sequencing (RNA-seq) strategies have been used in gene expression studies that have identified major virulence determinants and essential pathways, such as capsule biogenesis (8), stress (9) and temperature response (10), and sexual development (11). Genome sequencing and transcript profiling have revealed the molecular inventory of the organism and facilitated comparative genomics studies among the species and strains (12). For example, genomic comparisons of different *Cryptococcus* genomes revealed a unique chromosomal translocation in strains of *C. neoformans* var. *grubii* that influence virulence attributes (13).

The genomes of *C. neoformans* var. *neoformans* strain JEC21 and the related strain B-3501A were sequenced and published almost a decade ago (12). The genome of *C. neoformans* var. *grubii* strain H99 was sequenced by Duke University and the Broad Institute (www.broadinstitute.org/annotation/genome/cryptococcus_neoformans/), and a chromosome level assembly and improved annotation for H99 has recently been completed (14). The genome sequencing of two strains of *C. gattii*, R265 and WM276, has also been completed, and analysis of the strains revealed considerable genomic variation in clinical and environmental isolates and changes in chromosome copy numbers in isolates with fluconazole heteroresistance (15, 16). The sequencing of additional *Cryptococcus* genomes is under way (3). The analysis of these additional genomes will provide insights into *Cryptococcus* biology and will likely direct further research studies on *Cryptococcus*.

FungiDB (www.fungidb.org) (17) is an expanding fungal genome database that provides sequence information, functional data, and intuitive search tools for 61 species spanning the fungal and oomycetal tree of life, with genomes from Ascomycota, Ba-

Received 7 April 2014 **Accepted** 7 May 2014
Published ahead of print 9 May 2014
Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/EC.00083-14.

doi:10.1128/EC.00083-14
The authors have paid a fee to allow immediate free access to this article.

sidiomycota, Mucormycotina, Chytridiomycota, and Oomycetes. FungiDB also contains functional genomics data, including cell cycle microarray data for *Saccharomyces cerevisiae*; RNA-seq data for *Coprinopsis cinerea*, *Neurospora crassa*, and *Rhizopus oryzae*; and yeast two-hybrid interaction data for *S. cerevisiae*. The genome sequences and predicted gene product information, as well as functional genomics data, are available at FungiDB for five strains representing two species and two varieties of *Cryptococcus*.

The amount of functional information for *Cryptococcus* genes is growing rapidly, and there is a pressing need for a central resource that collects, organizes, and disseminates these data. Currently, functional information about *Cryptococcus* is dispersed in the published literature, requiring potentially time-consuming literature reading for researchers to find and incorporate characterized gene function data into their research. During this process, important references can be overlooked, often due to locus misidentification and conflicts in gene naming. In numerous instances, characterized genes with already published names have been given new names in subsequent publications, or a name associated with one locus has been used to refer to another locus (see the supplemental material).

To minimize nomenclature-related confusion among researchers, it has become imperative to gather and organize data from genetic, molecular, and epidemiological studies of *Cryptococcus* into a single resource of collective knowledge. We have compiled an up-to-date and comprehensive gene set for *Cryptococcus* published gene names, along with their locus identifiers (IDs) and descriptions, and have used these names and descriptions to annotate the set of syntenic orthologs across five *Cryptococcus* genomes. In addition, we propose a set of gene-naming guidelines that has been developed based on community protocols and in consultation with the *Cryptococcus* community. We have also established a *Cryptococcus* Gene Registry where gene names can be reserved in advance of publication. All of these data and resources are freely accessible at FungiDB. The centralized location of genomic and curated information for *Cryptococcus* will enable researchers to communicate effectively about particular genes, such as those related to virulence, and will facilitate gene function-based research on this important fungal pathogen.

MATERIALS AND METHODS

Cryptococcus gene name collection. PubMed searches at NCBI (www.ncbi.nlm.nih.gov/pubmed) (18) were performed using the keywords "*Cryptococcus*" and "gene" or "*Cryptococcus*" and "mutant." Replacing the term "*Cryptococcus*" with "*Filobasidiella*" did not result in any additional papers with gene names. These publications were then screened for the presence of a *Cryptococcus* gene that included a name. Papers that contained names associated with standard locus identifiers for the H99 or JEC21 strain were entered in a FileMakerPro (FileMaker, Inc.) database. For gene names associated with a GenBank identifier, sequences were compared by BLAST to the *Cryptococcus* genomes in FungiDB, and that gene name was assigned to the matching locus identifier. Papers containing DNA or protein sequence of the described gene were compared by BLASTn or BLASTp, respectively, to the *Cryptococcus* genomes in FungiDB, and the appropriate locus identifier was associated with the gene name. In some instances, no sequence information at all was provided or the primer sequences listed were insufficient for unambiguous locus determination. In these cases, the corresponding author of the publication was contacted directly by e-mail for clarification of the sequence and/or H99 or JEC21 locus identifier.

Expert review of published gene information. Community review of the published gene information for *Cryptococcus* was performed by e-mail communication with an international group of principal investigators from 30 *Cryptococcus* research laboratories. We asked these researchers to review a preliminary list of JEC21 and H99 gene names and descriptions with the specific goal of identifying errors, missed gene names, or any other comments they wished to make about their genes of interest or on gene naming in *Cryptococcus*. The responses we received were analyzed, and appropriate changes were noted. Comments received by e-mail regarding gene name protocols from each laboratory were documented and compiled into a standard protocol that reflected the majority opinion of the responders.

Ortholog determination. Ortholog assignment of *Cryptococcus* genes was determined by FungiDB 3.0 (released May 2014) (17). FungiDB uses the EupathDB (www.eupathdb.org) (19) analysis pipelines to load and analyze genomic sequences. Briefly OrthoMCL (20) was used to identify orthologs across 88 eukaryotic genomes, 16 archaeal genomes, and 34 bacterial genomes to create clusters that are made available in OrthoMCL-DB (version 5) (21). Genes from all fungal genomes in FungiDB are mapped into OrthoMCL-DB ortholog clusters by protein similarity (22). Mercator (23) was used to construct a synteny map of the homologous regions. The ortholog and synteny groupings of genes hosted in FungiDB was used to determine the orthology of genes across the five *Cryptococcus* genomes. Syntenic orthologs for every curated gene (see the supplemental material) were determined by manual visual inspection of the syntenic orthology of the genomes of *C. neoformans* var. *grubii* strain H99, *C. neoformans* var. *neoformans* strains JEC21 and B-3501A, and *C. gattii* strains WM276 and RM265 (see the supplemental material) (Fig. 1).

RESULTS AND DISCUSSION

Collection of published gene information for *Cryptococcus*. The corpus of published literature for an organism or species takes many years to generate and holds great potential as a source of information about the biology of an organism and its genes and proteins. Because these data are produced over time and vary in availability, searching for biological information about a gene or protein can be challenging. However, when the data about gene function from published research are collected and organized into a central Web-accessible location, researchers gain immediate access to all available information about the genes they study. The manual curation of data from the published literature provides a rich and valuable resource that combines sequences with gene names and functional information about proteins, such as phenotypes of mutant strains, Gene Ontology (24) annotations, and expression profiles, and connects them to powerful analysis tools. The analysis of such manually curated data can further stimulate hypothesis-driven research, either on a small scale or on a genome-wide scale. Despite the tremendous value this type of resource provides to the research community, databases with literature-based curation are available for only a limited number of fungal species.

In an effort to promote research and effective communication among researchers about the gene identity and function of the pathogen, we have collected and organized all available published data about *Cryptococcus* gene function and have made these data available at FungiDB. By searching PubMed, we found over 200 published *Cryptococcus* papers, dating from 1994 to 2013, with gene-related information, from which we collected 678 published names or aliases (synonyms) for 648 genes and wrote brief descriptions about the function of each gene (see the supplemental material). We also made use of a gene description prediction algorithm, Automated Assignment of Human Readable Descriptions (AHRD), to provide additional descriptions of gene function as a comparison to the curated descriptions (see the

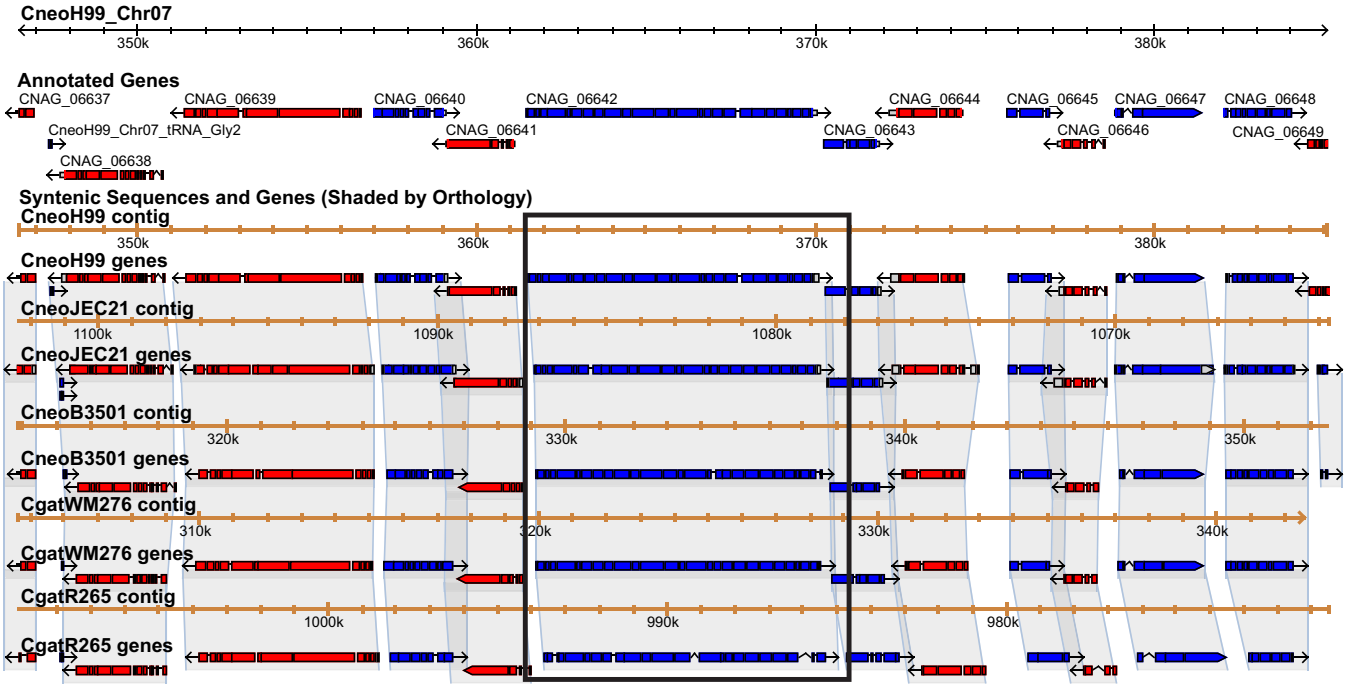

FIG 1 View of syntenic orthologs on the FungiDB locus summary and Genome Browser page. A view centered on CNAG_06642/*TOR1* is shown as an example.

supplemental material). Although the medical impact of *C. gattii* is high, functional gene analysis has been limited in strains of the species. Thus, the majority of our annotations came from publications describing research in *C. neoformans* var. *grubii* strain H99 and *C. neoformans* strain JEC21 with a small number of annotations coming from strain B-3501A.

Systematic locus identifiers for *C. neoformans* JEC21 genes were often included in papers published after 2005, but at least 50 of the papers we examined were published prior to the release of the JEC21 systematically named gene set. For this older set of papers, we used BLASTp (25) to match sequences labeled with GenBank identifiers and/or sequences presented in figures or tables in the papers to those in the current genome release at FungiDB. A small number of papers required verification with the authors to unambiguously identify the locus that was linked to the gene name. We encourage researchers to include a reference to the *Cryptococcus* systematic locus ID in future publications to facilitate the linking of gene annotations to their respective genes.

We found a surprisingly small number of gene name conflicts—when the same gene has been published under more than one name or when more than one gene has been called by the same name—although we uncovered future potential conflicts with the nomenclature for unpublished genes that are under investigation. We resolved the gene name conflicts that exist in the literature by using the rule that the first published name becomes the unique standard name for that gene and any additional names used in subsequent publications become aliases, or synonyms. This has been standard practice in the *S. cerevisiae* and *Candida albicans* communities for many years, so there is ample precedent for the approach. This manually curated and community-reviewed (described below) collection of published gene names and descriptions will aid researchers in their characterization of gene function and factors responsible for the pathogenicity of *Cryptococcus*.

***Cryptococcus* community gene annotation review.** To leverage the collective knowledge of the *Cryptococcus* research community and to garner support for the development and use of a freely available manually curated genomic resource for *Cryptococcus*, an initial set of 593 genes with names and descriptions (with associated PubMed identifiers) were sent by email to the heads of 30 *Cryptococcus* research laboratories for a 5-week period of review (December 2013 to January 2014). The file contained data from the H99 strain and/or the JEC21 strain that were curated from published, gene-centered literature. BLASTp results for H99 genes compared with *S. cerevisiae* proteins and specific questions or notes about particular genes were also supplied. Of these requests, we received timely and detailed responses from 26 of the research groups. All responses were analyzed for new gene name assignments associated with a literature reference, for leads on the first instance of publication of a gene name, and for the existence of conflicts in gene names. We found that the researchers who responded were highly meticulous about every published gene being represented on the list and adamant that the first name used in publication should become the standard name used in every subsequent publication. We also found that *Cryptococcus* researchers universally want a consistent gene-naming system based on both the identities of the orthologs and the gene nomenclature in use by the *S. cerevisiae* research community (described in detail below). Altogether, the efforts of the community during the gene review period supplied us with comments for over 200 genes and alerted us to additional references that resulted in an additional 42 published gene names that were included in the final curated gene set.

A community-based genetic nomenclature for *Cryptococcus*. The use of a community-supported genetic nomenclature is an essential first step in the development of a rational system for communicating information about genes and mutants of a species. In the absence of a unified gene-naming system, gene name

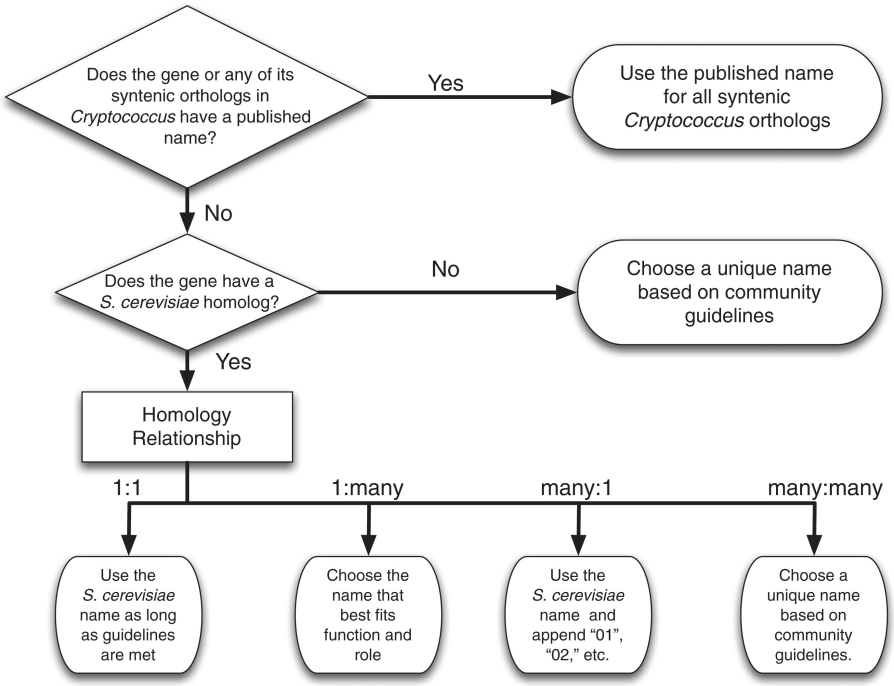

FIG 2 Flowchart for assigning names to *Cryptococcus* genes. A gene name is assigned based on prior publication in any of the *Cryptococcus* species or by homology to a gene in the model yeast *S. cerevisiae*. In cases of ambiguous orthology relationships with many *Cryptococcus* genes and 1 copy in *S. cerevisiae*, names are based on the yeast name but with a two-digit number appended. Where there are multiple *S. cerevisiae* homologs for a *C. neoformans* gene, the name will be chosen from that which best fits the role or function inferred from the *Cryptococcus* experiment. Where there are many paralogous copies in both *C. neoformans* and *S. cerevisiae*, a unique name not directly tied to the *S. cerevisiae* name is chosen for the gene.

conflicts occur and lead to confusion and errors in research when either a single name is used to refer to more than one gene, a single gene has been referred to by more than one name, or the name assigned to a gene is also used in related fungi for a gene with an entirely different function. Inconsistent formatting of gene names can also lead to confusion, as genes, proteins, or mutants of a species are often distinguished by uppercase and lowercase letters and/or italicization of the names. Standard names for published *Cryptococcus* genes from the H99 and JEC21 genomes were assigned by the first appearance in the main body of a peer-reviewed publication. Extended lists of gene names in publication supplements that lack characterization were not incorporated as standard names at FungiDB. We now consider this set of published names to be the standard names across both *Cryptococcus* species (five reference genomes) and encourage their use in all subsequent publications that refer to these genes or their syntenic orthologs across new *Cryptococcus* genomes.

Comments and protocols for naming *Cryptococcus* genes were collected by e-mail correspondence with laboratory heads during the gene list review period. While most researchers agree that choosing a name is up to the individual researcher, all were in favor of setting guidelines to ensure consistency and prevent conflicts. Here, we summarize the community comments and, based on these comments, propose a set of guidelines for *Cryptococcus* (Fig. 2) adapted from the model yeast *S. cerevisiae* that may also be applicable to other fungi. (i) Use the same name for *C. neoformans*, *C. neoformans* var. *grubii*, and *C. gattii* syntenic orthologs in the reference genomes H99, JEC21, B-3501A, WM267, and R265. Ideally, syntenic orthologs of new *Cryptococcus* genomes whose sequence is currently being determined (3) will also use these gene names in future publications. (ii) Avoid using "Cn" in the formal gene name unless for purposes of comparison in a discussion. The "Cn" ("Cg" for *C. gatii*) will be dropped from the standard gene name when it is recorded in FungiDB. (iii) Be certain that a published standard name does not already exist for the gene of interest. Check FungiDB for standard names and the community comments field on the gene for unpublished gene name reservations. (iv) Check to be sure that the name has not already been used to name a different *Cryptococcus* gene. A search of FungiDB should uncover names already in use. Also, a PubMed search with the gene name and "*Cryptococcus*" may pull up a publication with the name if it has been published. (v) Check to be sure that the name being considered has not already been used to name homologs of different genes in *S. cerevisiae*. A search at the *Saccharomyces* Genome Database (SGD) (www.yeastgenome.org) will reveal whether the name has been associated with an *S. cerevisiae* gene with a different function. (vi) Use the format of 3 capital letters plus a number for the gene name, consistent with the *S. cerevisiae* nomenclature (26). Gene names should be capitalized and italicized (e.g., *ACT1*), while protein names have the first letter capitalized with the rest of the letters lowercase and are not italicized (e.g., Act1). (vii) The names of mutant genes should be designated by italicization in lowercase (e.g., *act1*). (viii) Avoid using names that begin with a "C" if it stands for *Cryptococcus* as the first word in the name. (ix) Use the reciprocal best match of an *S. cerevisiae* protein with a BLASTp E value cutoff of $<10e-5$ to determine likely orthology (27). (x) In a case where more than one *Cryptococcus* protein matches a single *S. cerevisiae* protein (e.g., Xyz1), multiply the digit by 100 and then use the last digit to represent the different genes (e.g., *XYZ101*, *XYZ102*, etc.). (xi) For genes that

TABLE 1 Gene model predictions for *Cryptococcus* species[a]

Category	Value				
	B-3501A	JEC21	H99	WM276	R265
Total no. of predicted genes	6,500	6,476	6,975	6,575	6,216
No. of named genes (% of total)	629 (9.7)	634 (9.8)	640 (9.2)	618 (9.2)	610 (9.8)
No. of genes remaining to be named	5,871	5,842	6,348	5,957	5,606

[a] The number of protein-coding gene models predicted in five *Cryptococcus* genomes; the number of named genes annotated from the literature, including those with names assigned by orthology; and the count of the remaining genes in each genome requiring naming and annotation.

lack a named *S. cerevisiae* ortholog, choose a unique name or use the name of the closest well-characterized ortholog in another species. The general rule, however, is that if there is functional characterization of a *Cryptococcus* protein and it is involved in a different process than in other species, the name should reflect its function in *Cryptococcus*.

In some cases, exceptions to the preferred standard nomenclature were made. For example, for genes within the mating-type locus, we used an "alpha" designation appended to the name for all strains that bear a *MAT*α mating-type locus: *SXI1alpha*, *STE3alpha*, etc. The word "alpha" must be spelled out for database purposes, and this exception to the gene nomenclature was also made for database reasons in *S. cerevisiae* for its *MAT*α locus. The "α" designation was used in the original identification of the genes in the *MAT*α loci (28) and is appropriate for use in publication. All of the *Cryptococcus* strains currently in FungiDB (H99, JEC21, B-3501A, WM276, and R265) are of the *MATalpha* mating type. When whole genome sequences of *MAT***a** strains become available through future sequencing efforts, the *MAT***a** version of these genes will have an "**a**" appended. The adoption of a genetic nomenclature with proven success, as in the system in *S. cerevisiae*, will serve to unify the community behind a set of agreed upon standards, and the use of these guidelines will improve communication among researchers engaged in gene-centered research in *Cryptococcus*.

Ortholog determination and annotation across three *Cryptococcus* species. A total of 6,975 protein-coding gene models are predicted in the recently updated genome of *C. neoformans* var. *grubii* strain H99 (14), 6,476 are predicted for the *C. neoformans* strain JEC21 (12), and 6,500 are predicted for B-3501A. For *C. gattii* strains R265 and WM276, 6,216 and 6,575 gene models are predicted, respectively (Table 1) (15) and are available for searching in FungiDB.

One of the suggestions repeated by *Cryptococcus* researchers is that gene names should be applied consistently across the *Cryptococcus* species and strains. To comply with this request requires an error-free ortholog set with one-to-one protein matches so that only those proteins with orthology supported by synteny data are aligned in an ortholog group. Therefore, we selected genes corresponding to the published gene set from the H99 and JEC21 strains and also the genes that were submitted for name reservation (described below) and determined the orthologs across the species (Table 1) by manually examining synteny across the five related genomes (Fig. 1) hosted at FungiDB. The manual determination of the syntenic ortholog groups of the remaining ~5,400 *Cryptococcus* genes is an ongoing effort. Published names were applied across all species, even if they were not in the preferred 3-letter plus a number format, (e.g., the histone chaperone *HIRA* [7] and the fungus-specific glucosylceramide *EGCrP1* [29]) (see the supplemental material). Adding an additional name to a gene to conform to a naming guideline only exacerbates the problem by associating more than one name with a single gene, a circumstance we hope to avoid.

By providing this set of manually curated orthologs between the H99, JEC21, B-3501A, WM276, and R265 genomes, we have enabled researchers to view at a glance the roughly 9% of genes (Table 1) that already have published names and those names that have already been assigned to a locus, thus increasing the chance for uniform and meaningful gene name assignments in future publications.

A gene name registry for *Cryptococcus*. A gene name registry is a system in which a new and unique gene name can be assigned to a locus prior to publication. It lets other researchers know that the gene is under investigation and publication of the name is imminent. It can also alert investigators to potential conflicts in gene names, which most researchers strongly seek to avoid. After polling, 14 laboratory heads stated that they would support and use a gene name registry for *Cryptococcus* at FungiDB, and 27 genes were voluntarily submitted for prepublication registry during the gene list preview period (see the supplemental material). The remaining respondents focused on gene name review and simply did not answer the question. These gene registry comments are available in the community comments field for the respective genes at FungiDB, and the complete list of community comments, including gene name reservations, can be downloaded at www.fungidb.org.

Conclusions. Published gene names and descriptions for three *Cryptococcus* species have been made available in a consistent format for searching and downloading to facilitate the dissemination of gene and protein information. A genetic nomenclature for *Cryptococcus* has been built by community feedback and consensus from dedicated researchers. The creation of a *Cryptococcus* gene registry at FungiDB has received community-wide support. The organization and annotation of these data in one centrally located database provides the *Cryptococcus* research community with a rich source of genome annotation data, a valuable place to consider gene name information, and a place for community comments. All of these resources are available for search and download at FungiDB. We hope that the collection of *Cryptococcus* gene annotations in a publically available, centrally located resource will promote further functional research on this medically important pathogen.

ACKNOWLEDGMENTS

We are especially grateful to Tamara Doering, Hiten Madhani, Christina Hull, and Michael Brent for productive discussions on gene-naming protocols in *Cryptococcus* and thank all members of the community who provided feedback for this work.

V.M. was supported by a grant from the Agriculture and Food Research Initiative of the USDA National Institute of Food and Agriculture (2011-68004-30104). Funding for this project was provided by a grant from the National Institutes of Health (R03 AI105636-02) to J.E.S. and G.S. FungiDB was supported by grants from the A. P. Sloan Foundation and the Burroughs Wellcome Foundation to J.E.S.

We have no competing interests to declare.

REFERENCES

1. **Perfect JR, Casadevall A.** 2002. Cryptococcosis. Infect. Dis. Clin. North Am. **16:**837–874, v-vi. http://dx.doi.org/10.1016/S0891-5520(02)00036-3.
2. **Harris JR, Lockhart SR, Debess E, Marsden-Haug N, Goldoft M, Wohrle R, Lee S, Smelser C, Park B, Chiller T.** 2011. *Cryptococcus gattii* in the United States: clinical aspects of infection with an emerging pathogen. Clin. Infect. Dis. **53:**1188–1195. http://dx.doi.org/10.1093/cid/cir723.
3. **Chaturvedi V, Nierman WC.** 2012. *Cryptococcus gattii* comparative genomics and transcriptomics: a NIH/NIAID White Paper. Mycopathologia **173:**367–373. http://dx.doi.org/10.1007/s11046-011-9512-9.
4. **Park BJ, Wannemuehler KA, Marston BJ, Govender N, Pappas PG, Chiller TM.** 2009. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. AIDS **23:**525–530. http://dx.doi.org/10.1097/QAD.0b013e328322ffac.
5. **del Poeta M, Toffaletti DL, Rude TH, Sparks SD, Heitman J, Perfect JR.** 1999. *Cryptococcus neoformans* differential gene expression detected in vitro and in vivo with green fluorescent protein. Infect. Immun. **67:**1812–1820.
6. **Skowyra ML, Doering TL.** 2012. RNA interference in *Cryptococcus neoformans*. Methods Mol. Biol. **845:**165–186. http://dx.doi.org/10.1007/978-1-61779-539-8_11.
7. **Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM.** 2008. Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. Cell **135:**174–188. http://dx.doi.org/10.1016/j.cell.2008.07.046.
8. **Haynes BC, Skowyra ML, Spencer SJ, Gish SR, Williams M, Held EP, Brent MR, Doering TL.** 2011. Toward an integrated model of capsule regulation in *Cryptococcus neoformans*. PLoS Pathog. **7:**e1002411. http://dx.doi.org/10.1371/journal.ppat.1002411.
9. **Maeng S, Ko YJ, Kim GB, Jung KW, Floyd A, Heitman J, Bahn YS.** 2010. Comparative transcriptome analysis reveals novel roles of the Ras and cyclic AMP signaling pathways in environmental stress response and antifungal drug sensitivity in *Cryptococcus neoformans*. Eukaryot. Cell **9:**360–378. http://dx.doi.org/10.1128/EC.00309-09.
10. **Kraus PR, Boily MJ, Giles SS, Stajich JE, Allen A, Cox GM, Dietrich FS, Perfect JR, Heitman J.** 2004. Identification of *Cryptococcus neoformans* temperature-regulated genes with a genomic-DNA microarray. Eukaryot. Cell **3:**1249–1260. http://dx.doi.org/10.1128/EC.3.5.1249-1260.2004.
11. **Kruzel EK, Giles SS, Hull CM.** 2012. Analysis of *Cryptococcus neoformans* sexual development reveals rewiring of the pheromone-response network by a change in transcription factor identity. Genetics **191:**435–449. http://dx.doi.org/10.1534/genetics.112.138958.
12. **Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA, Allen JE, Bosdet IE, Brent MR, Chiu R, Doering TL, Donlin MJ, D'Souza CA, Fox DS, Grinberg V, Fu J, Fukushima M, Haas BJ, Huang JC, Janbon G, Jones SJ, Koo HL, Krzywinski MI, Kwon-Chung JK, Lengeler KB, Maiti R, Marra MA, Marra RE, Mathewson CA, Mitchell TG, Pertea M, Riggs FR, Salzberg SL, Schein JE, Shvartsbeyn A, Shin H, Shumway M, Specht CA, Suh BB, Tenney A, Utterback TR, Wickes BL, Wortman JR, Wye NH, Kronstad JW, Lodge JK, Heitman J, Davis RW, Fraser CM, Hyman RW.** 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science **307:**1321–1324. http://dx.doi.org/10.1126/science.1103773.
13. **Morrow CA, Lee IR, Chow EW, Ormerod KL, Goldinger A, Byrnes EJ, III, Nielsen K, Heitman J, Schirra HJ, Fraser JA.** 2012. A unique chromosomal rearrangement in the *Cryptococcus neoformans* var. *grubii* type strain enhances key phenotypes associated with virulence. mBio **3:**e00310–11. http://dx.doi.org/10.1128/mBio.00310-11.
14. **Janbon G, Ormerod K, Paulet D, Byrnes IE, Yadav V, Chatterjee G, Mullapudi N, Hon C, Billmyre R, Brunel F, Bahn Y, Chen W, Chen Y, Chow E, Coppée J, Floyd-Averette A, Gaillardin C, Gerik K, Goldberg J, Gonzalez-Hilarion S, Gujja S, Hamlin JL, Hsueh YP, Ianiri G, Jones S, Kodira CD, Kozubowski L, Lam W, Marra M, Mesner LD, Mieczkowski PA, Moyrand F, Nielsen K, Proux C, Rossignol T, Schein JE, Sun S, Wollschlaeger C, Wood IA, Zeng Q, Neuvéglise C, Newlon CS, Perfect JR, Lodge JK, Idnurm A, Stajich JE, Kronstad JW, Sanyal K, Heitman J, Fraser JA, Cuomo CA, Dietrich FS.** 2014. Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. **10:**e1004261. http://dx.doi.org/10.1371/journal.pgen.1004261.
15. **D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, Jung WH, Sham A, Kidd SE, Tangen K, Lee N, Zeilmaker T, Sawkins J, McVicker G, Shah S, Gnerre S, Griggs A, Zeng Q, Bartlett K, Li W, Wang X, Heitman J, Stajich JE, Fraser JA, Meyer W, Carter D, Schein J, Krzywinski M, Kwon-Chung KJ, Varma A, Wang J, Brunham R, Fyfe M, Ouellette BF, Siddiqui A, Marra M, Jones S, Holt R, Birren BW, Galagan JE, Cuomo CA.** 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. mBio **2:**e00342–00310. http://dx.doi.org/10.1128/mBio.00342-10.
16. **Gillece JD, Schupp JM, Balajee SA, Harris J, Pearson T, Yan Y, Keim P, DeBess E, Marsden-Haug N, Wohrle R, Engelthaler DM, Lockhart SR.** 2011. Whole genome sequence analysis of *Cryptococcus gattii* from the Pacific Northwest reveals unexpected diversity. PLoS One **6:**e28550. http://dx.doi.org/10.1371/journal.pone.0028550.
17. **Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, Stoeckert CJ, Jr, Roos DS.** 2012. FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res. **40:**D675–D681. http://dx.doi.org/10.1093/nar/gkr918.
18. **NCBI Resource Coordinators.** 2014. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **42:**D7–D17. http://dx.doi.org/10.1093/nar/gkt1146.
19. **Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissinger JC, Kraemer ET, Li W, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Jr, Wang H, Warrenfeltz S.** 2013. EuPathDB: the eukaryotic pathogen database. Nucleic Acids Res. **41:**D684–D691. http://dx.doi.org/10.1093/nar/gks1113.
20. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178–2189. http://dx.doi.org/10.1101/gr.1224503.
21. **Chen F, Mackey AJ, Stoeckert CJ, Jr, Roos DS.** 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. **34:**D363–D368. http://dx.doi.org/10.1093/nar/gkj123.
22. **Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ, Jr.** 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinformatics Chapter 6:Unit 6.12. http://dx.doi.org/10.1002/0471250953.bi0612s35.
23. **Dewey CN.** 2007. Aligning multiple whole genomes with Mercator and MAVID. Methods Mol. Biol. **395:**221–236. http://dx.doi.org/10.1007/978-1-59745-514-5_14.
24. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25:**25–29. http://dx.doi.org/10.1038/75556.
25. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
26. **Cherry JM.** 1995. Genetic nomenclature guide. *Saccharomyces cerevisiae*. Trends Genet. March:11–12.
27. **Salichos L, Rokas A.** 2011. Evaluating ortholog prediction algorithms in a yeast model clade. PLoS One **6:**e18755. http://dx.doi.org/10.1371/journal.pone.0018755.
28. **Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J.** 2002. Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. Eukaryot. Cell **1:**704–718. http://dx.doi.org/10.1128/EC.1.5.704-718.2002.
29. **Ishibashi Y, Ikeda K, Sakaguchi K, Okino N, Taguchi R, Ito M.** 2012. Quality control of fungus-specific glucosylceramide in *Cryptococcus neoformans* by endoglycoceramidase-related protein 1 (EGCrP1). J. Biol. Chem. **287:**368–381. http://dx.doi.org/10.1074/jbc.M111.311340.