

UC Riverside

UC Riverside Previously Published Works

Title

Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts.

Permalink

<https://escholarship.org/uc/item/9000j11g>

Journal

mBio, 2(1)

ISSN

2150-7511

Authors

D'Souza, CA
Kronstad, JW
Taylor, G
et al.

Publication Date

2011

DOI

10.1128/mbio.00342-10

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

Genome Variation in *Cryptococcus gattii*, an Emerging Pathogen of Immunocompetent Hosts

C. A. D'Souza,^a J. W. Kronstad,^a G. Taylor,^b R. Warren,^b M. Yuen,^a G. Hu,^a W. H. Jung,^c A. Sham,^a S. E. Kidd,^{a,d} K. Tangen,^a N. Lee,^a T. Zeilmaker,^a J. Sawkins,^a G. McVicker,^a S. Shah,^a S. Gnerre,^e A. Griggs,^e Q. Zeng,^e K. Bartlett,^f W. Li,^g X. Wang,^g J. Heitman,^g J. E. Stajich,^h J. A. Fraser,ⁱ W. Meyer,^d D. Carter,^j J. Schein,^b M. Krzywinski,^b K. J. Kwon-Chung,^k A. Varma,^k J. Wang,^a R. Brunham,^l M. Fyfe,^m B. F. F. Ouellette,^{a,n} A. Siddiqui,^b M. Marra,^b S. Jones,^b R. Holt,^b B. W. Birren,^e J. E. Galagan,^e and C. A. Cuomo^e

The Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada^a; Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada^b; Department of Biotechnology, Chung-Ang University, Anseong-Si, Gyeonggi-Do, Republic of Korea^c; Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Westmead Millennium Institute, Sydney Emerging Disease and Biosecurity Institute, Sydney Medical School—Westmead Hospital, The University of Sydney, Westmead, New South Wales, Australia^d; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA^e; School of Occupational and Environmental Hygiene, University of British Columbia, Vancouver, British Columbia, Canada^f; Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA^g; Department of Plant Pathology and Microbiology, University of California, Riverside, California, USA^h; Centre for Infectious Disease Research, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australiaⁱ; School of Molecular Bioscience, University of Sydney, Sydney, New South Wales, Australia^j; Molecular Microbiology Section, Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA^k; British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada^l; Office of the Medical Health Officer, Vancouver Island Health Authority, Victoria, British Columbia, Canada^m; and Ontario Institute for Cancer Research, Toronto, Ontario, Canadaⁿ

C.A.D. and J.W.K. contributed equally to this article and should be considered co-first authors.

ABSTRACT *Cryptococcus gattii* recently emerged as the causative agent of cryptococcosis in healthy individuals in western North America, despite previous characterization of the fungus as a pathogen in tropical or subtropical regions. As a foundation to study the genetics of virulence in this pathogen, we sequenced the genomes of a strain (WM276) representing the predominant global molecular type (VGI) and a clinical strain (R265) of the major genotype (VGIIa) causing disease in North America. We compared these *C. gattii* genomes with each other and with the genomes of representative strains of the two varieties of *Cryptococcus neoformans* that generally cause disease in immunocompromised people. Our comparisons included chromosome alignments, analysis of gene content and gene family evolution, and comparative genome hybridization (CGH). These studies revealed that the genomes of the two representative *C. gattii* strains (genotypes VGI and VGIIa) are colinear for the majority of chromosomes, with some minor rearrangements. However, multiortholog phylogenetic analysis and an evaluation of gene/sequence conservation support the existence of speciation within the *C. gattii* complex. More extensive chromosome rearrangements were observed upon comparison of the *C. gattii* and the *C. neoformans* genomes. Finally, CGH revealed considerable variation in clinical and environmental isolates as well as changes in chromosome copy numbers in *C. gattii* isolates displaying fluconazole heteroresistance.

IMPORTANCE Isolates of *Cryptococcus gattii* are currently causing an outbreak of cryptococcosis in western North America, and most of the cases occurred in the absence of coinfection with HIV. This pattern is therefore in stark contrast to the current global burden of one million annual cases of cryptococcosis, caused by the related species *Cryptococcus neoformans*, in the HIV/AIDS population. The genome sequences of two outbreak-associated major genotypes of *C. gattii* reported here provide insights into genome variation within and between cryptococcal species. These sequences also provide a resource to further evaluate the epidemiology of cryptococcal disease and to evaluate the role of pathogen genes in the differential interactions of *C. gattii* and *C. neoformans* with immunocompromised and immunocompetent hosts.

Received 22 December 2010 Accepted 10 January 2011 Published 8 February 2011

Citation D'Souza, C. A., J. W. Kronstad, G. Taylor, R. Warren, M. Yuen, et al. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. mBio 2(1):e00342-10. doi:10.1128/mBio.00342-10.

Editor Françoise Dromer, Institut Pasteur

Copyright © 2011 D'Souza et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to J. W. Kronstad, kronstad@interchange.ubc.ca.

Fungal pathogens of humans mainly cause disease in people with underlying immune defects due to AIDS, cancer, or immunosuppressive therapy. For example, strains of *Cryptococcus neoformans* emerged as life-threatening agents of fungal meningitis, with an increase in the number of global cases coincidentally

occurring with the HIV/AIDS epidemic (1, 2). These pathogens are responsible for an estimated one million cases of cryptococcal meningitis globally per year in AIDS patients, leading to approximately 625,000 deaths (3). In contrast, the closely related species *Cryptococcus gattii* has the distinct ability to cause disease in oth-

erwise healthy people, although *C. gattii* infections have also been reported in immunocompromised patients, including HIV/AIDS and organ transplant patients (4–6). *C. gattii* was initially classified as a variety of *C. neoformans* but has recently been recognized as a separate species (7). The two species share all of the major recognized virulence factors such as production of a polysaccharide capsule, melanin deposition in the cell wall, and robust growth at 37°C, but they differ in a number of traits (reviewed in reference 8). For example, unlike those of *C. neoformans*, *C. gattii* strains can assimilate D-proline, D-tryptophan, and L-malic acid, use glycine as both carbon and nitrogen sources, and resist growth inhibition by cycloheximide and canavanine. Clinically, *C. gattii* infections result in a higher incidence of lung and brain granulomas and neurological complications, and these infections often require prolonged treatment with antifungal drugs compared with those caused by *C. neoformans* (8).

The ability of *C. gattii* to infect both immunocompromised and immunocompetent individuals has been dramatically demonstrated by the emergence of cryptococcosis on Vancouver Island in British Columbia (BC), Canada, over the last 10 years. Specifically, there have been >240 reported human cases (mainly in immunocompetent people) and hundreds of animal cases of cryptococcosis since 1999 (9–11). With an average annual incidence of 5.8 per million persons, this is a much higher rate of infection than that found in the rest of the world, and 19 of the human cases were fatal despite aggressive antifungal therapy (10). The outbreak observed on Vancouver Island has now spread to the British Columbia mainland as well as the Pacific Northwest region of the United States (9, 12–16). The outbreaks and increased incidence of cryptococcosis in a temperate climatic zone suggest a fundamental shift in the ecological adaptation of this pathogen.

Extensive environmental sampling revealed that *C. gattii* is present in the air and soil and on multiple tree species in British Columbia (9, 16–18). Previous characterization of *C. gattii* isolates suggested that the species was primarily a tropical or subtropical pathogen and revealed that the natural habitat of *C. gattii* in Australia was in tree species such as the eucalypts (19–21). Molecular typing of clinical and environmental isolates by PCR fingerprinting, amplified fragment length polymorphism analysis, and multilocus sequence typing identified four molecular types within the *C. gattii* complex (VGI, VGII, VGIII, and VGIV), with VGI being the most commonly isolated genotype worldwide (22–25). The VGII molecular type is predominant among clinical and environmental isolates from Vancouver Island, with further identification of VGIIa and VGIIb subtypes among the isolates. The VGIIa subtype is predominant in the environment and responsible for the majority of cases of cryptococcosis, although fatal disease has also been caused by VGI and VGIIb strains (10, 26). In a disquieting trend of continual expansion in the United States, a virulent and novel subtype, VGIIc, has emerged in Oregon and is now contributing to illness in the region along with the VGIIa subtype (13).

In this report, we present the genome sequences of strains representing VGI and VGIIa and compare these genomes with each other and with the sequenced genome of a serotype D strain (JEC21) of the opportunistic species *C. neoformans* (27). We also describe comparative genome hybridization (CGH) experiments that extend the analysis of the genomes to include *C. gattii* outbreak strains with different levels of virulence as well as strains

resistant to high levels of the antifungal drug fluconazole. Together, these studies reveal extensive rearrangements between the representative *C. gattii* and *C. neoformans* genomes that may have contributed to sexual isolation and speciation. Additionally, considerable variation was observed between and within molecular types of *C. gattii*, and this finding is consistent with the existence of separate species within the *C. gattii* complex. Overall, the *C. gattii* sequences provide a reference platform for studying virulence and for further detailed epidemiological characterization of isolates causing the unusual emergence of this pathogen in North America.

RESULTS AND DISCUSSION

Genome sequences of the *C. gattii* strains WM276 and R265.

The *C. gattii* strain WM276, which represents the VGI molecular type, which is commonly isolated worldwide, was sequenced at 6.5× coverage and assembled into 14 chromosomes with a combined size of 18.4 Mb (see Table S1a and b in the supplemental material). The assembly of the genome was supported by a physical map constructed by fingerprinting bacterial artificial chromosome (BAC) clones and by the inclusion of sequence reads from the ends of the BAC clones. This strategy, together with several rounds of finishing and gap closure, improved the quality of the genome assembly and resulted in a high level of completion. As a result, only eight internal gaps remain in the sequence, and telomeric sequences comprised of repeats of the motif (TTAGGGG)_n were identified at both ends of eight chromosomes; telomeric sequences were identified at only one end of four additional chromosomes, and telomeric sequences have yet to be identified for the remaining two chromosomes (see Table S1b in the supplemental material). With the assembled genome sequence, we identified and annotated 6,565 potential open reading frames (ORFs; excluding pseudogenes). These ORFs were identified and annotated using Pegasys (28), which includes the Atlas database (29) and Apollo (30, 31) for the curators to interpret gene models and make decisions based on the evidence presented to them. The evidence was built from gene prediction models, BLASTp (32) against nonredundant (nr) protein resources, and tBLASTx against fungal dbEST (33) sequences. This annotation processing pipeline also included the identification and the location of retroelements (see annotations in Data Set S1, tab A, in the supplemental material). With regard to the latter elements, the centromeres in the WM276 chromosomes can be defined by the clustering of the retrotransposons TCN1 and TCN6 at these locations, as previously described for the genomes of the *C. neoformans* strains JEC21 and B3501A (27). We also identified the rRNA gene cluster on chromosome 2 (positions 1516460 to 1924811; ~0.4 Mb) but excluded rRNA and other RNA-carrying genes from the annotations, in keeping with a focus on protein-encoding genes.

The WM276 genome and the detailed manual annotations have been employed as the reference genome for *C. gattii* and used in the annotation of the *C. gattii* R265 genome. The genome of this strain, which represents the VGIIa subtype, responsible for the majority of cases of cryptococcosis in British Columbia, was sequenced at 6.5× coverage (see Materials and Methods in Text S1 and sequencing statistics in Table S1c in the supplemental material). Automated annotation of the assembled R265 sequence was performed, and the annotation information obtained from the WM276 genome was employed to compare gene models. The

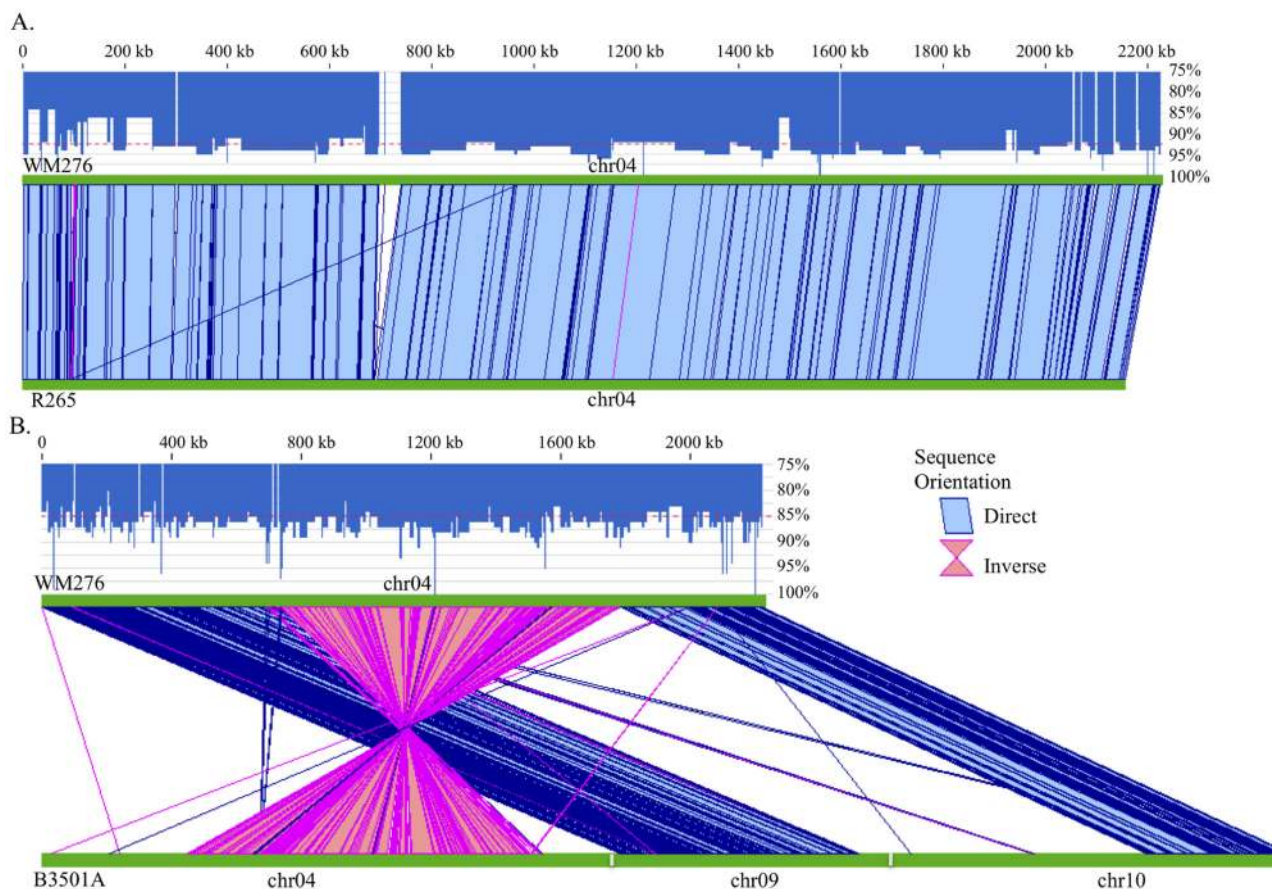


FIG 1 Alignments of colinear and rearranged chromosomes. (A) Alignments of the sequence of chromosome 4 from the two *C. gattii* strains, WM276 and R265. (B) Alignment of chromosome 4 from *C. gattii* strain WM276 with the corresponding sequences from the *C. neoformans* strain B3501A, which are present on chromosomes 4, 9, and 10. Aligned chromosomes are indicated by green lines. The blue bar above each chromosome pair shows the percent sequence identity (panel A, ~92.4%; panel B, ~85%; overall percentage indicated by the red dashed line). The blue and pink segments between the chromosomes indicate whether the alignments are direct or inverted, respectively (blocks are bounded by a darker color and filled with a lighter color).

R265 genome assembly was also supported by a BAC physical map and BAC end sequences.

Comparison of the *C. gattii* genome sequences. To obtain an overview of the organizations of the *C. gattii* genomes for strains WM276 and R265, we aligned their chromosomes using the nucleic acid sequence comparison tool Cross-Match (<http://www.phrap.org>), a Smith-Waterman local alignment algorithm (34, 35); alignments were visualized using the tool XMatchView (<http://www.bcgsc.ca/platform/bioinfo/software/xmatchview>). Cross-Match analysis of these two genomes revealed that the majority of chromosomes were colinear, as illustrated for chromosome 4 in Fig. 1A. However, some rearranged chromosomes and regions of inversion were also present between the two genomes (see Text S1, p. 19 in the supplemental material). This is in striking contrast to the more substantial genome rearrangements observed in the comparisons with the serotype D genome, as shown in Fig. 1B and described below. The Cross-Match analysis also revealed a higher-than-expected overall nucleotide sequence divergence of 7.6% between the VGI and VGII genomes. This observation suggests potential speciation within the *C. gattii* complex, although we note that the percent divergence is 10 to 15% between the *neoformans* and *grubii* varieties of *C. neoformans* (36). Perhaps these observa-

tions indicate that these *C. neoformans* varieties should be separate species.

We compared the gene contents of the *C. gattii* strains by mapping syntenic orthologs between the two strains. First, the assemblies were aligned using NUCmer (37), and alignments separated by less than 200 bases were merged into syntenic regions. Next, gene coordinates were transferred from WM276 to R265 using the base-to-base correspondence of the alignments. Upon mapping 6,120 genes from WM276 to 6,102 genes in R265, we found some 2:1 and 1:2 mappings that appeared mostly to be splits or merges of genes (see Data Set S1, tab B, in the supplemental material). A total of 445 WM276 genes did not map to the R265 genome, but closer inspection allowed the alignment of 291 loci to R265 regions that did not contain good gene structures. This group had a small average polypeptide size (205 amino acids [aa]) compared to the overall average (522 aa), and the small size and lack of conservation suggest that these are dubious genes. The remaining 154 genes did not fall into alignment with regions of the R265 genome. The full set of 6,210 genes that were predicted in R265 also included 108 that were annotated uniquely for this genome.

The majority of genes specific to one genome or the other encoded hypothetical proteins. However, some specific examples of

differential gene content between the two *C. gattii* strains included the genes encoding the predicted Argonaute proteins Ago1 and Ago2 (CGB_D9320W and CGB_D9160C, respectively), which were present in WM276 but not R265; this gene deficiency in R265 was not associated with gaps in the R265 genome assembly (see Text S1, p. 20 and Table S2 in the supplemental material). The highly conserved Argonaute proteins are part of the RNA-induced silencing complex (RISC) (38) and function in RNA interference and related phenomena. Another example of a difference in gene content is an 880-bp deletion that removes upstream sequence and the coding region for 209 N-terminal amino acids of the pheromone receptor-like protein Cpr2 (CNBG_5530) in R265; this gene is intact in WM276 (CGB_A1720W) (see Text S1, p. 21 and Table S2 in the supplemental material) (39). The WM276 ortholog CGB_A1730W that lies downstream of this locus is also absent from R265. Aside from these differences, the *CPR2* flanking region was conserved in both of the *C. gattii* strains, suggesting that the Cpr2 gene may be under selective pressure. In this regard, it is possible that selective loss of some gene functions may contribute to the higher virulence of the R265 strain relative to that of WM276 (24, 40). It has been proposed, for example, that pathogens may become adapted to the selective pressure of a host niche through inactivation of so-called antivirulence genes (41). Identification of the compendium of antivirulence genes for a particular pathogen could lead to a better understanding of the emergence of novel virulence traits, perhaps including factors relevant to the emergence of cryptococcosis in western North America in the case of *C. gattii*. Overall, the observed gene differences between the two *C. gattii* strains will provide opportunities for functional examination of their contributions to virulence differences.

Comparison of the genome sequences of *C. gattii* and *C. neoformans*. The sequences of the WM276 chromosomes were each aligned with the sequences of chromosomes of the *C. neoformans* strain B3501A (27) using the Cross-Match approach. At the whole-genome level, the WM276 genome shows 87.0% identity to the B3501A genome, and the *C. gattii* R265 genome has 85.6% identity with that of B3501A. In contrast to the alignments observed for the *C. gattii* strains, which showed extensive conservation of synteny (described above and see Text S1, p. 19 in the supplemental material), many rearrangements were observed for the comparisons of the *C. gattii* and *C. neoformans* chromosomes (see Text S1, p. 22 and 23 in the supplemental material). In particular, a striking three-part chromosomal rearrangement was observed that involved chromosomes 4, 9, and 10 in the two strains; the rearrangements that resulted in chromosome 4 are illustrated in Fig. 1B. Within the chromosomes from each genome, two out of six breakpoints involved in the rearrangements were associated with the retrotransposons TCN1 and TCN6. For example, we found TCN1 and TCN6 elements at the junctions within WM276 chromosomes 4 and 10 at the locations where sequences aligned to B3501A chromosomes 4 and 9 were juxtaposed. These junctions are located within the centromeres of these chromosomes, a finding consistent with the highly repetitive nature of these regions. We hypothesize that three steps would be needed to rearrange the chromosomes in the manner observed in Fig. 1B starting from common ancestral sequences and that one of the steps may have involved the TCN1/TCN6 retroelements at the centromeres of the respective chromosomes. These rearrangements were also observed when the R265 genome was compared with that of B3501A,

indicating that the chromosomal rearrangements are ancient in the *C. gattii* lineage (data not shown).

The differences in chromosome arrangements between *C. gattii* and *C. neoformans* may have functional significance with regard to virulence, and it is possible that the observed rearrangements are indicative of dynamic genome changes that contributed to speciation. The mating type locus (*MAT*) has been associated with virulence in *C. neoformans* (42, 43), and this locus was found on rearranged chromosome 9 in both of the *C. gattii* genomes. In general, chromosomal rearrangements, segmental duplications and whole-chromosome copy number variations have been described for *C. neoformans* and are well documented in other fungi (44–51). Genomic rearrangements can serve as direct targets for natural selection and can accumulate in different lineages to contribute to the genotypic divergence and speciation through inhibition of proper pairing and recombination of rearranged chromosomes (52). Certainly, the presence of the rearrangements in both of the *C. gattii* genomes supports the idea that these chromosomal changes contributed to speciation within cryptococci.

To assess gene content differences, we also compared the *C. gattii* and *C. neoformans* genomes by reciprocal BLAST (32) of their gene sets and identified genes exclusively found in JEC21 (254 genes) or WM276 (565 genes) (see Data Set S1, tabs C and D, in the supplemental material). The JEC21 genome was used in this analysis because of the detailed gene annotations available for this genome and because of the nearly identical gene sets between JEC21 and B3501A (27). The majority of the genes encoded hypothetical proteins. However, examples of orthologs found in JEC21 but absent in WM276 included those encoding the 60S ribosomal protein L31, a haloacid dehalogenase, an inositol/phosphatidylinositol kinase, an alpha-L-arabinofuranosidase, and a sphingosine-1-phosphate phosphatase (see Data Set S1, tab C, in the supplemental material). Examples of orthologs found in WM276 but absent in JEC21 included those encoding the Rad51-like DNA repair protein and several enzymes such as phenylacrylic acid decarboxylase, arsenate reductase, 6-phosphogluconolactonase, haloalkanoic acid dehalogenase, and isochorismatase (see Data Set S1, tab D, in the supplemental material). The functional significance of these differences in gene contents warrants further investigation; in particular, the function of the isochorismatase is interesting because this enzyme has a predicted role in catecholic siderophore/secondary metabolite biosynthesis (53–55). In addition, the isochorismatase gene was part of a deletion associated with loss of virulence in a WM276 mutant, as described below.

Cognate isochorismatase orthologs of the WM276 protein were found in *C. gattii* strain R265 and *C. neoformans* var. *grubii* strain H99 but not in *C. neoformans* var. *neoformans*. Although the latter finding agrees with the inability of *C. neoformans* var. *neoformans* to produce siderophores (56, 57), divergent paralogs with weaker similarity to the isochorismatase domain were present; this property has not been investigated in *C. gattii* to our knowledge. Moreover, catecholic siderophore biosynthesis has not been well characterized in fungi. While a BLAST query of the non-redundant database with the WM276 ortholog using default criteria did not find any significantly similar proteins in other basidiomycetes or *Saccharomyces cerevisiae*, we found significant hits to proteins in other ascomycete fungi such as *Aspergillus* spp. (E value = 5E–58) and *Botryotinia* (E value = 8E–68).

Evolutionary relationships among the *Cryptococcus* species and other basidiomycetes based on multiple single-copy orthologs. All of the predicted protein sequences from the five *Cryptococcus* genomes (WM276, R265, JEC21, B3501A, and H99) were clustered into orthologous groups using OrthoMCL (58), (see Materials and Methods in Text S1 in the supplemental material). OrthoMCL includes recent paralogs within ortholog groups as within-species BLAST hits that are reciprocally better than between-species hits. Single-copy orthologs were identified as the clusters with exactly one member per species. Following the identification of 5,171 groups of single-copy orthologs conserved among the five strains, alignments for each of these groups were concatenated (2,817,121 characters), and a phylogenetic tree was generated using the maximum likelihood analysis method implemented in PhyML3.0 (59). We calculated the time since divergence of *C. neoformans* var. *grubii* versus *C. gattii* to be ~34 million years (myr) based on the average branch length between *C. neoformans* var. *grubii* representative strain H99 and the *C. gattii* strains in the phylogenetic tree (0.1355) and assuming a commonly utilized neutral mutation rate of $2E-9$ per nucleotide per year for protein-coding genes (60, 61). The multiortholog-based ultrametric tree was then generated by the PATHd8 algorithm (62) using the age estimate of 34 myr for the most recent common ancestor of the *C. gattii* and *C. neoformans* var. *grubii* lineages (Fig. 2A; see also Materials and Methods in Text S1 in the supplemental material). The divergence between the VGI and VGII *C. gattii* strains WM276 and R265 was found to be 12.4 myr, and this result advocates for the existence of speciation between these molecular types. In fact, phylogenetic analysis of a selection of globally collected isolates indicates considerable genetic variation within the *Cryptococcus* species complex, warranting *C. gattii* molecular types to be considered individual varieties, if not species (24, 63).

To examine phylogenetic relationships in the broader context of other basidiomycete fungi, we also carried out phylogenetic analysis with the five *Cryptococcus* genomes and other basidiomycete genomes, including the human pathogen *Malassezia globosa* (64), the plant pathogen *Ustilago maydis* (65), the wood-rotting fungus *Phanerochaete chrysosporium* (66), and the mushroom *Coprinus cinereus* (67). The ascomycetous fungus *Saccharomyces cerevisiae* was included as a distantly related outgroup taxon. Following identification of 1,519 single-copy ortholog groups conserved among the 10 genomes, alignments for each of these groups were concatenated (837,857 characters), and a phylogenetic tree was generated by maximum likelihood analysis, as described above. The resultant phylogenetic tree was calibrated based on a recent estimate of ~500 million years of divergence between ascomycetous and basidiomycetous fungi (68) (Fig. 2B). According to this calibration, the two *C. gattii* molecular types would have diverged about 11 myr ago, an estimate similar to the one based solely on the *Cryptococcus* phylogeny described above. It appears that the cryptococci, representatives of the class Tremellomycetes, diverged about 291 myr from the common ancestor of *Phanerochaete chrysosporium* and *Coprinus cinereus* (class Agaricomycetes). We also carried out analysis of the evolution of gene families identified among the five cryptococci and the other fungi mentioned above, and we compared the mitochondrial gene contents of all of the sequenced *Cryptococcus* genomes (see Text S1 in the supplemental material).

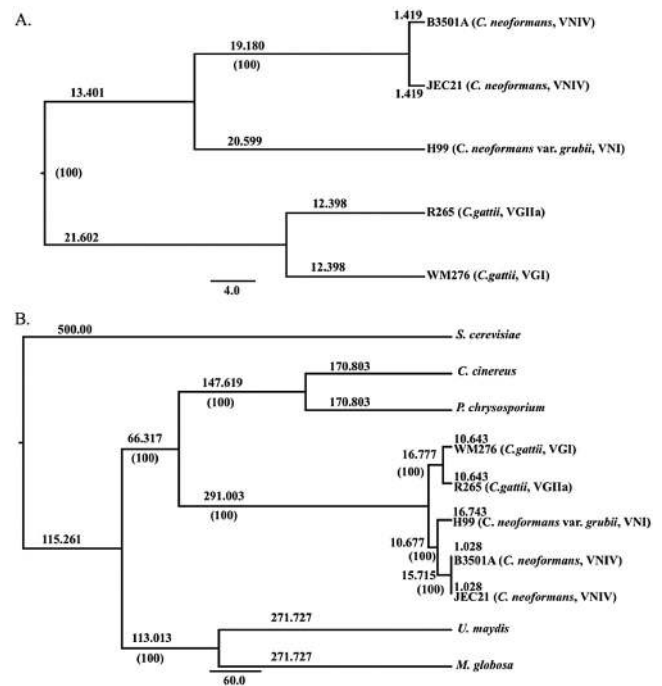


FIG 2 Multiortholog phylogeny of *Cryptococcus* strains and other basidiomycete fungi. Branch lengths of the indicated phylogenetic trees represent the number of nucleotide substitutions per site calibrated in million year divergence times (shown above the branches), and the numbers in brackets are values for bootstrap support of the branch. (A) A calibrated phylogenetic tree based on 5,171 single-copy orthologs conserved between the five indicated *Cryptococcus* strains was generated using the age estimate of the most recent common ancestor of the *C. gattii*-*C. neoformans* var. *grubii* lineages. Genome alignment results and the extent of divergence (12.4 myr) between the VGI and VGII *C. gattii* strains WM276 and R265 was found to be 12.4 myr, and this result advocates for the existence of speciation between these molecular types. (B) A phylogenetic tree based on 1,519 single-copy orthologs conserved among the 10 indicated fungal species was calibrated based on a recent estimate of ~500 myr of divergence between ascomycetous and basidiomycetous fungi. According to this calibration, the two *C. gattii* molecular types would have diverged about 11 myr ago. It appears that the cryptococci, representatives of the class Tremellomycetes, diverged about 291 myr from the common ancestors of *Phanerochaete chrysosporium* and *Coprinus cinereus*.

Genome variation within the VGI molecular type of *C. gattii*.

Comparative genome hybridization (CGH) studies were performed with the WM276 genome to begin an analysis of genome variation within the VGI molecular type of *C. gattii*. A whole-genome tiling array was initially employed to characterize a transformant of strain WM276 (WM276gfp2) that displayed a change in chromosome 11, as discovered by electrophoretic karyotyping (see Text S1, p. 24 in the supplemental material). This strain, which was generated by transformation with a gene encoding the green fluorescent protein, was also found to be avirulent in a mouse inhalation model of cryptococcosis (data not shown). CGH analysis revealed that a telomeric region of ~75 kb was missing on chromosome 11 in strain WM276gfp2 (see Text S1, p. 25 and 26 in the supplemental material). This region contained 24 genes, including a number of putative sugar transporters and glycosyl hydrolases (see Data Set S1, tab I, in the supplemental material). The potential for phenotypic consequences of the deletion was demonstrated by confirming that the mutant had a growth defect on raffinose, as expected from the deletion of the invertase gene CGB_K4300C (data not shown). The deleted region also

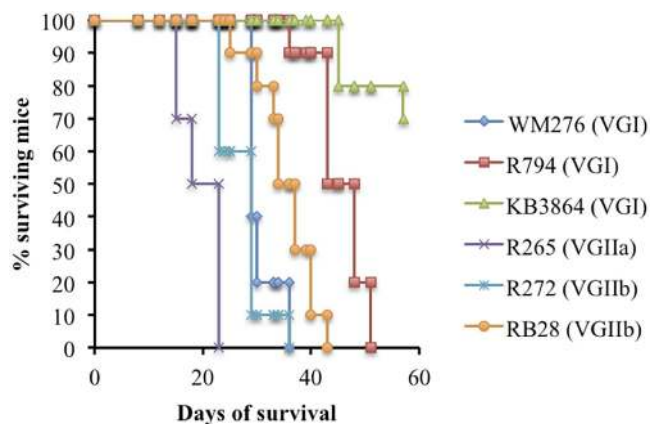


FIG 3 Virulence of Vancouver Island isolates of *C. gattii*. Ten A/JCr female mice were inoculated intranasally with 5×10^4 cells of each of the strains indicated and monitored for illness over 2 months. These assays were performed twice for all of the strains, with similar results. The analyses of virulence for strains WM276 (VGI), R265 (VGIIa), and R272 (VGIIb) (shown here only for comparison) were performed in the same experiment and previously published (40). The differences in virulence for WM276 versus those in the VGI strains R794 ($P < 0.0001$) and KB3864 ($P < 0.0001$) were statistically significant, and the mice infected with WM276 reached the endpoint of the experiment at day 36. The virulence of VGIIb strain RB28 was attenuated compared with that of VGIIb strain R272 (endpoint at day 36) and relative to that of the more virulent VGIIa strain R265 (endpoint at day 23) (40). The lower virulence of R272 relative to that of R265 has also been described in other studies (24, 40, 75, 76)

encoded an arsenite transporter, a peptidyl-prolyl *cis-trans*-isomerase, a Ras GTPase, an isochorismatase, and a copper-exporting ATPase. As mentioned above, the cognate WM276 isochorismatase ortholog was not found in the *C. neoformans* var. *neoformans* genomes. Overall, this analysis with the WM276 genome illustrated the utility of CGH to characterize genome variability, although additional work is needed to determine whether the loss of specific genes and/or mutations elsewhere in the WM276gfp2 genome account for the virulence defect. The number of variant genes in each *C. gattii* strain analyzed by CGH (including the clinical, environmental, and fluconazole-resistant strains) is indicated in Table S3 in the supplemental material.

An extensive collection of clinical and environmental isolates has been obtained as part of the analysis of the *C. gattii* outbreak on Vancouver Island (26). A survey of selected isolates of the VGI and VGII molecular types revealed differences in virulence in the mouse inhalation model of cryptococcosis (Fig. 3). For example, the virulence of strains R794 and KB3864 in the VGI set of strains was attenuated in comparison to WM276, and we therefore analyzed the genomes of these isolates by CGH with the WM276 tiling array. Examples of regions of difference (insertions/deletions/sequence divergence) between WM276 and the clinical isolate R794 are shown in Text S1, p. 27 in the supplemental material (all chromosomes are shown in Text S1, p. 26 in the supplemental material), and variant regions are described in Data Set S1, tab J, in the supplemental material. A substantial number of variant regions were found on different chromosomes, and many of these were at telomeric and subtelomeric regions. The genes that were deleted or highly diverged encoded, for example, an alpha-glucosidase (CGB_E0010C), an inositol oxygenase (CGB_G2390W), a hexose transport-related protein (CGB_H0010C), and myo-inositol

transporters (encoded on three different chromosomes: chr7, CGB_G2420C; chr10, CGB_J2530W; and chr12, CGB_L0070C). We also found a substantial number of genome differences in the comparison of the environmental isolate KB3864 with WM276 (see Text S1, p. 26 and Data Set S1, tab K, in the supplemental material). Some of these differences were the same as those found in R794, including variation in the regions containing genes encoding the alpha-glucosidase, the hexose transport-related protein, and a myo-inositol transporter.

Variations that potentially impact inositol metabolism are interesting, considering that this metabolite is found in high concentrations in the brain. Given the predilection of *Cryptococcus* for the central nervous system, the pathogen could potentially utilize Myo-inositol (a stereoisomer of inositol) as a sole carbon source through conversion to glucuronic acid by the action of Myo-inositol oxygenase (MIOX) (69). Variations in myo-inositol transporters in the VGI isolates are also noteworthy because myo-inositol transport has been implicated in mating and virulence (70). Additionally, transcriptome studies revealed that the transcript for myo-inositol phosphate synthase (*MYO1*) is abundant *in vivo* and that an inositol/phosphatidyl inositol phosphatase is upregulated upon phagocytosis of *C. gattii* by rat peritoneal macrophages (71–73). Recently, it was demonstrated that phosphatidylinositol 4-kinase is required for survival *ex vivo* in the hostile cerebrospinal fluid environment and within macrophages and for full virulence (74).

Genome variation within the VGII molecular type of *C. gattii*. Attenuated virulence of the VGIIb environmental strain RB28 was observed relative to VGIIa strain R265 in the mouse model (Fig. 3). In addition, attenuated virulence in the VGIIb clinical strain R272 had been demonstrated by other research groups and in our previous studies of the immune response to *C. gattii* (24, 40, 75, 76). The tiling array for the VGIIa genome of strain R265 was therefore used in CGH experiments to examine the genomes of these VGIIb subtype strains. Examples of regions of difference are shown in Text S1, p. 28 in the supplemental material (all chromosomes are shown in Text S1, p. 29 in the supplemental material). A smaller number of differences were identified in the tiling array than in the VGI analysis, and the variant regions in strains R272 and RB28 are listed in Data Set S1, tabs L and M, respectively, in the supplemental material. Genes encoding a putative oxidoreductase and a hexose carrier protein were either deleted or highly diverged in both strains. Genes specifically deleted or diverged in strain R272 encoded a putative endoribonuclease L-PSP, a TPR domain-containing protein, and a tartarate transporter. Genes specifically amplified in strain RB28 encoded a putative 2,4-dichlorophenoxyacetate alpha-ketoglutarate dioxygenase, a deoxyribose-phosphate aldolase, and a beta-1,4-glucosidase. Overall, this analysis revealed extensive variation between strains, thus precluding simple explanations of differences in virulence based on genome content.

Analysis of VGI and VGII strains showing heteroresistance to the antifungal drug fluconazole. The WM276 and R265 tiling arrays were also used in CGH experiments to examine genome changes in isolates of *C. gattii* from Canada, Australia, and India that showed heteroresistance to 64 $\mu\text{g/ml}$ of fluconazole. Heteroresistance is defined as the phenotypic manifestation of both drug resistance and susceptibility in mixed populations of a single clinical isolate (77). The two resistant VGI strains R1413F and R1412F each showed a different pattern of genome change compared with

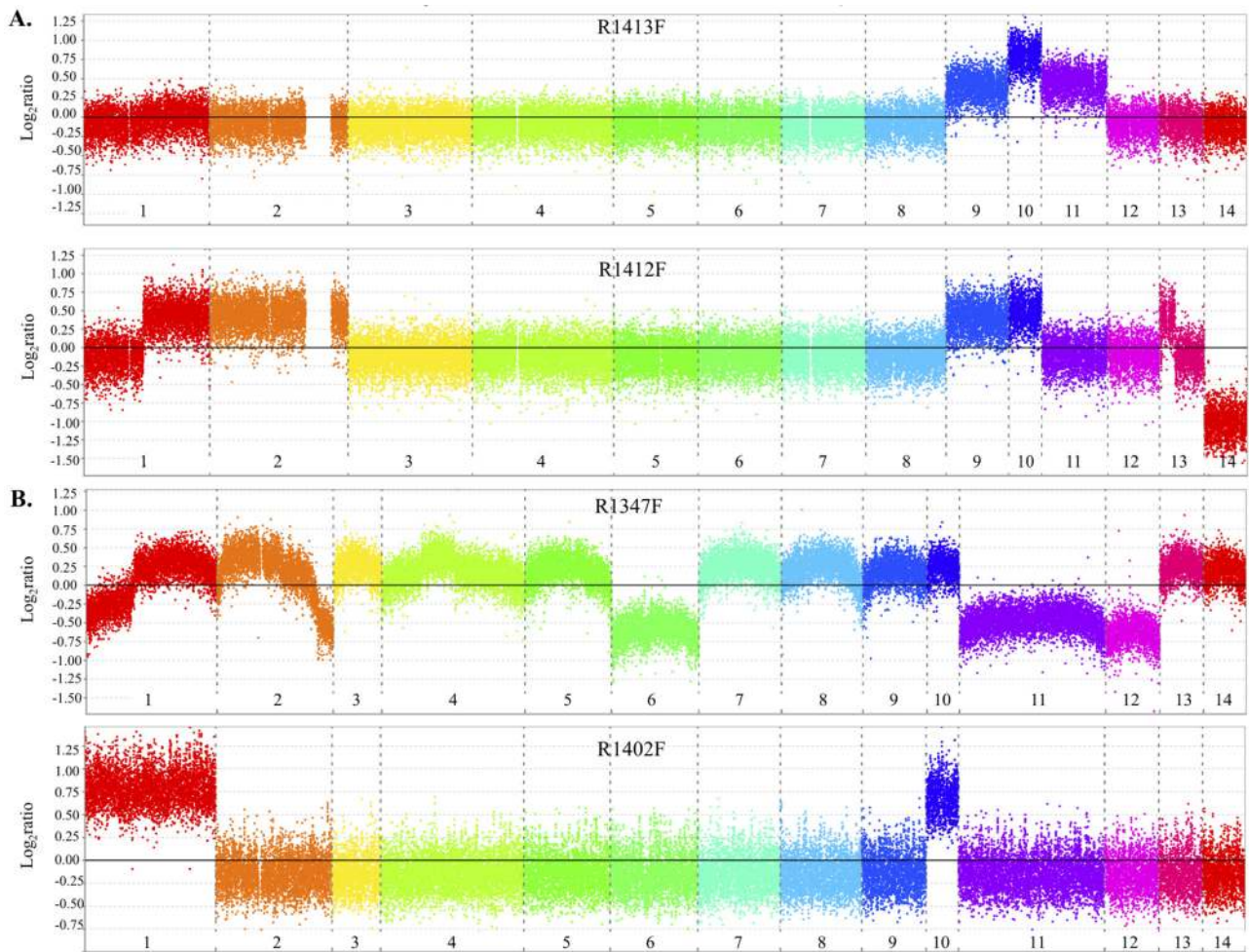


FIG 4 Comparative hybridization of fluconazole-resistant VGI isolates of *C. gattii* strains from the VGI and VGII molecular types. DNA from all resistant strains and their cognate parental strains was differentially labeled with fluorescent dyes prior to competitive hybridization to the VGI reference genome (WM276) or the VGII reference genome (R265) array. Log₂ ratios for assessing relative hybridization were averaged in windows of 400 bp. (A) VGI strains R1412F and R1413F (supported by FACS analysis) (see Text S1, p. 31 in the supplemental material). (B) VGII strains R1347F and R1402F (supported by FACS analysis) (see Text S1, p. 32 and 33 in the supplemental material). CGH plots for R1401F and R1346F can be found in Text S1, p. 30 in the supplemental material (see FACS analysis in Text S1, p. 32 and 33 in the supplemental material).

those of their parental strains. For strain R1413F, chromosomes 9 and 11 appear to have an elevated copy number of ~ 1.4 (based on a log₂ ratio of ~ 0.5), and chromosome 10 appeared to be disomic (log₂ ratio of ~ 1.0) (Fig. 4). These results suggest that the strain may contain a mixed population of cells, with different copy numbers present for the indicated chromosomes. Similarly, entire chromosomes 2, 9, and 10 have elevated copy numbers in the R1412F strain, and chromosomes 1 and 13 have elevated copy numbers for specific chromosomal segments. The latter chromosomes may have segmental duplications or more complicated rearrangements (e.g., translocations or the formation of isochromosomes). Surprisingly, this strain also had a reduced copy number for chromosome 14. One explanation for this observation is that the baseline ploidy of the strain may be diploid, and some chromosomes (e.g., 2, 9, and 10) may have copy numbers above $2N$ while chromosome 14 is present in a single copy. This conclusion is supported by fluorescence-activated cell sorting (FACS) experiments (78) that indicate that R1412F has a diploid character (see Text S1, p. 31 in the supplemental material). Variant regions in

VGI fluconazole-resistant isolates are listed in Data Set S1, tabs N and O, in the supplemental material).

The four VGII isolates that showed heteroresistance to $64 \mu\text{g/ml}$ of fluconazole also each had a different pattern of chromosome changes relative to those of the parental strains. Three of the strains showed relatively simple changes, with R1401F displaying an elevated copy number for a portion of chromosome 1 (see Text S1, p. 30 in the supplemental material), R1346F showing an increased copy number for all of chromosome 3 (see Text S1, p. 30), and R1402F having elevated copy numbers for chromosomes 1 and 10 (Fig. 4). Variant regions in fluconazole-resistant VGII isolates are listed in Data Set S1, tabs P to S, in the supplemental material.

FACS analysis supported the conclusion that R1346F, R1401F, and R1402F were primarily haploid (see Text S1, p. 32 and 33 in the supplemental material). Strain R1347F had a more complicated hybridization pattern, with log₂ ratios above 0 for all of chromosomes 3, 4, 5, 7, 8, 9, 10, and 14 and for segments of chromosomes 1 and 2 (Fig. 4). In contrast, log₂ ratios below 0 were

observed for chromosomes 6, 11, and 12, indicating that changes in ploidy had occurred relative to that of the parental strain R1347. FACS analysis of R1347F supported this conclusion because a mixed population of cells with ploidies above $2N$ was observed (see Text S1, p. 32 in the supplemental material). Overall, these results indicate that heteroresistance to fluconazole in strains of both the VGI and VGII molecular types of *C. gattii* is associated with changes in chromosome copy number and ploidy.

Disomic chromosomes have previously been identified in clinical isolates of *C. neoformans* var. *grubii*, and disomy has been recently associated with fluconazole heteroresistance in strains of both varieties of *C. neoformans* (49, 51). In particular, fluconazole resistance was attributed to an elevated copy number of chromosome 1 carrying the genes *AFR1* (ATP binding cassette [ABC] transporter; major exporter of azoles) and *ERG11* (cytochrome P450 lanosterol 14 α -demethylase; target of fluconazole). We found an elevated copy number for chromosome 1 in four of the six strains that we examined, a result consistent with amplification of the azole-transporter gene *AFR1* on this chromosome. We hypothesize that this may contribute in part to the observed fluconazole resistance, but amplification of chromosomes other than chromosomes 1 and 2 (carrying *AFR1* and *ERG11*, respectively) suggests that other mechanisms of fluconazole resistance that are independent of either *AFR1* or *ERG11* may occur in the strains. We should also note that changes in chromosome copy number may be a more general response to selective pressure because increased ploidy has recently been described in *C. neoformans* during the process of giant cell formation in infected animals (79, 80).

Summary. The genome sequences of the VGI and VGIIa genotypes of *C. gattii* revealed that the majority of the 14 chromosomes are colinear, with some minor rearrangements, but that the strains show considerable variation in gene content and overall sequence identity. In addition, multiortholog phylogenetic analysis supports the existence of speciation within the *C. gattii* complex. CGH analysis also revealed considerable variation in clinical and environmental isolates as well as changes in chromosome copy numbers and ploidy in *C. gattii* isolates displaying fluconazole heteroresistance. The genome sequences and the comparative studies reported here provide an opportunity to further define virulence functions that are distinct from or similar to those of the better-characterized sibling species *C. neoformans*. In particular, the genome sequences support detailed examinations of cryptococcal traits that might eventually explain the predilection of *C. gattii* for immunocompetent hosts, in contrast to the predilection of *C. neoformans* for immunocompromised people. For example, the genomes will facilitate the analysis of the high intracellular proliferation rate observed in *C. gattii* strains from the outbreak (13, 75) as well as the finding that *C. gattii* strains induce less protective inflammation in mice than *C. neoformans* strains (40).

MATERIALS AND METHODS

Genome sequencing and PCR analysis. The whole-genome shotgun sequence assembler ARACHNE (81) was used to assemble Sanger sequencing reads for the WM276 genome, and the assembly was improved using information from bacterial artificial chromosome (BAC)-based physical maps (82, 83). Automated annotation of the WM276 genome was performed with an in-house genome annotation algorithm called Pegasys (28), based on comparisons with annotated gene models and expressed sequence tags (ESTs) for *C. neoformans* strain JEC21, followed by manual curation with the genome browser and editor Apollo (30, 31). A summary

of the sequencing details is presented in Table S1, and see below for GenBank accession numbers.

The sequence of the R265 genome was obtained using Sanger sequencing and assembled with ARACHNE (81). The R265 genes were annotated primarily by transferring annotations from WM276 and also calling a small number of novel genes. R265 was aligned to WM276 using PatternHunter (84), and gene calls in aligned blocks were mapped from WM276 to R265 using an in-house mapping program. To call genes specific to R265, candidate gene structures were identified using GENEID (85), FGENESH (86), and GLEAN (87), and the resulting 108 genes were supported by predictions of a Pfam protein domain or alignment with an EST sequence.

PCR assays were employed to examine the *AGO1*, *AGO2*, *CPR2*, and *CGB_A1730W* genes in *C. gattii* strains (see Table S2 in the supplemental material). Additional details on DNA isolation, primer sequences, and PCR conditions, as well as sequence analysis, are provided in Text S1 in the supplemental material. Text S1 also contains a table of strains and lists of the software and data sources employed in the work.

Identification and alignment of orthologs and phylogenetic analysis of *Cryptococcus* strains and other fungal taxa. Orthology data sets were generated, including 10-way clusters among selected basidiomycetes and *S. cerevisiae* and 5-way clusters between the sequenced *Cryptococcus* spp. For each data set, all predicted protein sequences from the appropriate genomes were searched against each other with BLASTP (32) and clustered into orthologous groups using OrthoMCL (58) with the default criteria (E value < $1E-5$). Among the five *Cryptococcus* strains, 5,171 single-copy orthologs were identified as the clusters with exactly one member per species. Multiple sequence alignments were constructed with MUSCLE (88), and the alignments were trimmed using a heuristic method implemented in trimAl (89), with the automated option that selects optimal parameters to trim the input alignment. Alignments for all 5,171 clusters were concatenated into a single file containing 2,817,121 characters, converted to the Phylip format. Phylogenetic analysis of the five *Cryptococcus* strains was performed using maximum likelihood method PhyML3.0 (59) implemented in SeaView 4 (90). The JTT (Jones, Taylor, Thornton) amino acid substitution model (91) was used, along with the tree topology search operation that combines NNI (Nearest Neighbor Interchange) and SPR (Subtree Pruning and Regrafting) moves, the proportion of invariable sites and category of substitution rate were optimized by the program, and gaps were treated as unknown characters. The starting tree to be refined by the maximum likelihood algorithm was a distance-based BIONJ (Bio Neighbor Joining) tree estimated by the program (59). Statistical support for phylogenetic grouping was assessed by approximate likelihood ratio tests based on a Shimodaira-Hasegawa-like procedure (SH-aLRT) (92) and by bootstrap analysis (500 resamplings). An ultrametric tree was generated using PATHd8 (62), with the maximum likelihood tree as a starting point and fixing the age of the most recent common ancestor (MRCA) involved in the *C. neoformans* (H99)-*C. gattii* split at 34 myr. This age was derived using the neutral mutation rate of $2E-9$ per nucleotide per year for protein-coding genes. While PATHd8 does not assume a molecular clock exists, it runs a clock test, allowing for substitution rate variation along all lineages. Molecular clock tests indicated that three-fourth of nodes were rejected at a confidence level of 0.95. Pathd8 parameters are provided in Materials and Methods of Text S1 in the supplemental material. This file also provides details of the methods used for testing the phylogenetic relationships of the five *Cryptococcus* strains with *S. cerevisiae*, *U. maydis*, *C. cinereus*, *M. globosa*, and *P. chrysosporium*.

MCL clustering of annotated proteins into families and likelihood analysis of gene family expansion and contraction. The MCL (Markov Cluster) algorithm was used to globally identify gene families in the fungal genomes in our data set. MCL detects proteins with very similar domain architectures rather than attempting to detect each domain individually, thus accurately assigning proteins (even ones with different domain structures) into distinct multigene families (93). The algorithm CAFE (com-

putational analysis of gene family evolution) was used to detect significant gene family size changes between any two lineages (94, 95). For additional details, see Text S1 in the supplemental material.

Isolation and analysis of fluconazole-heteroresistant strains. Two representative strains each from Canada (R.B.-13, R.B.-14), Australia (RAM-15, VPB571-058), and India (B-5765, B-5788) were used to study heteroresistance to fluconazole and for CGH and FACS analysis. The methods for isolation of heteroresistant strains of *C. gattii* have been described (96), and additional details of strain processing, CHG analysis, and FACS are provided in Text S1 in the supplemental material (and in reference 49). Note that we have redesignated the previously isolated *C. gattii* strains (in parentheses) as follows: R-1346 (R.B.-13), R-1347 (R.B.-14), R-1401 (RAM-15), R-1402 (VPB571-058), R-1412 (B-5765), and R-1413 (B-5788). For FACS analysis, cells of fluconazole-resistant *C. gattii* strains were grown in yeast extract-peptone-dextrose (YPD) medium with fluconazole, harvested from liquid medium at log phase, and processed for flow cytometry as described previously (78, 97).

Virulence assays. Two virulence assays were performed using a mouse model of cryptococcosis. In the first, the virulence of the *C. neoformans* var. *grubii* and *C. gattii* strains expressing green fluorescent protein (GFP) fusions was tested using 10 C57BL/6 female mice per strain. The second assay analyzed the virulence of VGI and VGII strains of *C. gattii*. In this experiment, 10 A/JCr female mice were inoculated intranasally with 5×10^4 cells of each strain (RB28, R794, and KB3864) and monitored for illness over 2 months. The experiment also included the VGI strain WM276, the VGIIa strain R265, and the VGIIb strain R272. The virulence data for these three strains were previously reported (40), as described in the legend to Fig. 3. Survival data were analyzed using Kaplan-Meier curves, and the groups of mice infected with different strains were compared by using the log rank test to assess statistical significance and confidence of the virulence data. Additional details are provided in Text S1 in the supplemental material.

Ethics statement. The virulence assays employing mice were carried out in strict accordance with the guidelines of the Canadian Council on Animal Care. The protocol for the assays was approved by the University of British Columbia Committee on Animal Care (protocol A07-0117).

Nucleotide sequence accession numbers. Sequences were deposited in GenBank with the following accession numbers: chr1, CP000286; chr2, CP000287; chr3, CP000288; chr4, CP000289; chr5, CP000290; chr6, CP000291; chr7, CP000292; chr8, CP000293; chr9, CP000294; chr10, CP000295; chr11, CP000296; chr12, CP000297; chr13, CP000298; and chr14, CP000299. The assembled R265 genome and annotations were submitted to GenBank (project accession number AAFP01000000).

ACKNOWLEDGMENTS

We thank Thomas Sharpton (Gladstone Institute, CA) and Scott DiGiustini (Vancouver, British Columbia, Canada) for unconditional advice on the gene family evolution analysis. We also thank the Cryptococcal working group and the British Columbia Centre for Disease Control, especially Eleni Galanis, for advice, as well as Chris Walsh, Han Hao, and Brett Finlay for additional assistance.

The sequencing and annotation of the WM276 genome was funded by Genome Canada and Genome British Columbia and by grants from the National Institute of Allergy and Infectious Disease (R01 AI053721) and the Canadian Institutes of Health Research to J.W.K. Sequencing and annotation of the R265 genome at the Broad Institute was supported by the National Human Genome Research Institute (grant U54HG003067). Additional support was obtained from an R01 grant (AI50113) to J.H. and an intramural program grant to K.J.K.-C. and A.V. from the National Institute of Allergy and Infectious Diseases and the National Institutes of Health (NIH/NIAD, MD), the National Health and Medical Research Council Australia Research grant 990738 to W.M. for the construction of physical maps of the *C. gattii* genomes, and a grant to D.C. by the Howard Hughes Medical Institute under the International Scholars Program (55000640).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00342-10/-/DCSupplemental>.

Text S1, PDF file, 4.5 MB.
Text S2, RTF file, 0.152 MB.
Text S3, RTF file, 0.138 MB.
Table S1, PDF file, 0.086 MB.
Table S2, PDF file, 0.070 MB.
Table S3, PDF file, 0.067 MB.
Data Set S1, XLSX file, 0.788 MB.

REFERENCES

- Dismukes, W. E. 1988. Cryptococcal meningitis in patients with AIDS. *J. Infect. Dis.* 157:624–628.
- Levitz, S. M. 1991. The ecology of *Cryptococcus neoformans* and the epidemiology of cryptococcosis. *Rev. Infect. Dis.* 13:1163–1169.
- Park, B. J., K. A. Wannemuehler, B. J. Marston, N. G. Gøvdner, P. G. Pappas, and T. M. Chiller. 2009. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS* 23:525–530.
- Chaturvedi, S., M. Dyavaiah, R. A. Larsen, and V. Chaturvedi. 2005. *Cryptococcus gattii* in AIDS patients, southern California. *Emerg. Infect. Dis.* 11:1686–1692.
- Litvintseva, A. P., R. Thakur, L. B. Reller, and T. G. Mitchell. 2005. Prevalence of clinical isolates of *Cryptococcus gattii* serotype C among patients with AIDS in sub-Saharan Africa. *J. Infect. Dis.* 192:888–892.
- Blankenship, J. R., N. Singh, B. D. Alexander, and J. Heitman. 2005. *Cryptococcus neoformans* isolates from transplant recipients are not selected for resistance to calcineurin inhibitors by current immunosuppressive regimens. *J. Clin. Microbiol.* 43:464–467.
- Kwon-Chung, K. J., T. Boekhout, J. W. Fell, and M. Diaz. 2002. (1557) Proposal to conserve the name *Cryptococcus gattii* against *C. hondurians* and *C. bacillisporus* (Basidiomycota, Hymenomycetes, Tremellomycetidae). *Taxon* 51:804–806.
- Sorrell, T. C. 2001. *Cryptococcus neoformans* variety *gattii*. *Med. Mycol.* 39:155–168.
- Bartlett, K. H., S. E. Kidd, and J. W. Kronstad. 2008. The emergence of *Cryptococcus gattii* in British Columbia and the Pacific Northwest. *Curr. Infect. Dis. Rep.* 10:58–65.
- Galanis, E., and L. Macdougall. 2010. Epidemiology of *Cryptococcus gattii*, British Columbia, Canada, 1999–2007. *Emerg. Infect. Dis.* 16:251–257.
- Mak, S., B. Klinkenberg, K. Bartlett, and M. Fyfe. 2010. Ecological niche modeling of *Cryptococcus gattii* in British Columbia, Canada. *Environ. Health Perspect.* 118:653–658.
- Byrnes, E. J., III, R. J. Bildfell, S. A. Frank, T. G. Mitchell, K. A. Marr, and J. Heitman. 2009. Molecular evidence that the range of the Vancouver Island outbreak of *Cryptococcus gattii* infection has expanded into the Pacific Northwest in the United States. *J. Infect. Dis.* 199:1081–1086.
- Byrnes, E. J., III, W. Li, Y. Lewit, H. Ma, K. Voelz, P. Ren, D. A. Carter, V. Chaturvedi, R. J. Bildfell, R. C. May, and J. Heitman. 2010. Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. *PLoS Pathog.* 6:e1000850.
- Byrnes, E. J., III, R. J. Bildfell, P. L. Dearing, B. A. Valentine, and J. Heitman. 2009. *Cryptococcus gattii* with bimorphic colony types in a dog in western Oregon: additional evidence for expansion of the Vancouver Island outbreak. *J. Vet. Diagn. Invest.* 21:133–136.
- Datta, K., K. H. Bartlett, R. Baer, E. Byrnes, E. Galanis, J. Heitman, L. Hoang, M. J. Leslie, L. MacDougall, S. S. Magill, M. G. Morshed, K. A. Marr, and the *Cryptococcus gattii* Working Group of the Pacific Northwest. 2009. Spread of *Cryptococcus gattii* into Pacific Northwest region of the United States. *Emerg. Infect. Dis.* 15:1185–1191.
- MacDougall, L., S. E. Kidd, E. Galanis, S. Mak, M. J. Leslie, P. R. Cieslak, J. W. Kronstad, M. G. Morshed, and K. H. Bartlett. 2007. Spread of *Cryptococcus gattii* in British Columbia, Canada, and detection in the Pacific Northwest, USA. *Emerg. Infect. Dis.* 13:42–50.
- Kidd, S. E., Y. Chow, S. Mak, P. J. Bach, H. Chen, A. O. Hingston, J. W. Kronstad, and K. H. Bartlett. 2007. Characterization of environmental sources of the human and animal pathogen *Cryptococcus gattii* in British Columbia, Canada, and the Pacific Northwest of the United States. *Appl. Environ. Microbiol.* 73:1433–1443.
- Dixit, A., S. F. Carroll, and S. T. Qureshi. 2009. *Cryptococcus gattii*: an

- emerging cause of fungal disease in North America. *Interdiscip. Perspect. Infect. Dis.* :840452.
19. Kwon-Chung, K. J., and J. E. Bennett. 1984. Epidemiologic differences between the two varieties of *Cryptococcus neoformans*. *Am. J. Epidemiol.* 120:123–130.
 20. Ellis, D. H., and T. J. Pfeiffer. 1990. Natural habitat of *Cryptococcus neoformans* var. *gattii*. *J. Clin. Microbiol.* 28:1642–1644.
 21. Pfeiffer, T. J., and D. H. Ellis. 1992. Environmental isolation of *Cryptococcus neoformans* var. *gattii* from *Eucalyptus tereticornis*. *J. Med. Vet. Mycol.* 30:407–408.
 22. Boekhout, T., B. Theelen, M. Diaz, J. W. Fell, W. C. Hop, E. C. Abeln, F. Dromer, and W. Meyer. 2001. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiology* 147:891–907.
 23. Meyer, W., A. Castaneda, S. Jackson, M. Huynh, E. Castaneda, and the IberoAmerican Cryptococcal Study Group. 2003. Molecular typing of IberoAmerican *Cryptococcus neoformans* isolates. *Emerg. Infect. Dis.* 9:189–195.
 24. Fraser, J. A., S. S. Giles, E. C. Wenink, S. G. Geunes-Boyer, J. R. Wright, S. Diezmann, A. Allen, J. E. Stajich, F. S. Dietrich, J. R. Perfect, and J. Heitman. 2005. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. *Nature* 437:1360–1364.
 25. Meyer, W., D. M. Aanensen, T. Boekhout, M. Cogliati, M. R. Diaz, M. C. Esposto, M. Fisher, F. Gilgado, F. Hagen, S. Kaocharoen, A. P. Litvintseva, T. G. Mitchell, S. P. Simwami, L. Trilles, M. A. Viviani, and J. Kwon-Chung. 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med. Mycol.* 47: 561–570.
 26. Kidd, S. E., F. Hagen, R. L. Tschärke, M. Huynh, K. H. Bartlett, M. Fyfe, L. Macdougall, T. Boekhout, K. J. Kwon-Chung, and W. Meyer. 2004. A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). *Proc. Natl. Acad. Sci. U. S. A.* 101:17258–17263.
 27. Loftus, B. J., et al. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307:1321–1324.
 28. Shah, S. P., D. Y. He, J. N. Sawkins, J. C. Druce, G. Quon, D. Lett, G. X. Zheng, T. Xu, and B. F. Ouellette. 2004. Pegasys: software for executing and integrating analyses of biological sequences. *BMC Bioinformatics* 5:40.
 29. Shah, S. P., Y. Huang, T. Xu, M. M. Yuen, J. Ling, and B. F. Ouellette. 2005. Atlas—a data warehouse for integrative bioinformatics. *BMC Bioinformatics* 6:34.
 30. Lewis, S. E., S. M. Searle, N. Harris, M. Gibson, V. Lyer, J. Richter, C. Wiel, L. Bayraktaroglu, E. Birney, M. A. Crosby, J. S. Kaminker, B. B. Matthews, S. E. Prochnik, C. D. Smithy, J. L. Tupy, G. M. Rubin, S. Misra, C. J. Mungall, and M. E. Clamp. 2002. Apollo: a sequence annotation editor. *Genome Biol.* 3:RESEARCH0082.
 31. Ed, L., H. Nomi, G. Mark, C. Raymond, and L. Suzanna. 2009. Apollo: a community resource for genome annotation editing. *Bioinformatics* 25: 1836–1837.
 32. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
 33. Boguski, M. S., T. M. Lowe, and C. M. Tolstoshev. 1993. dbEST—database for “expressed sequence tags.” *Nat. Genet.* 4:332–333.
 34. Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* 147:195–197.
 35. Gotoh, O. 1982. An improved algorithm for matching biological sequences. *J. Mol. Biol.* 162:705–708.
 36. Kavanaugh, L. A., J. A. Fraser, and F. S. Dietrich. 2006. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Mol. Biol. Evol.* 23:1879–1890.
 37. Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
 38. Nakayashiki, H., N. Kadotani, and S. Mayama. 2006. Evolution and diversification of RNA silencing proteins in fungi. *J. Mol. Evol.* 63: 127–135.
 39. Hsueh, Y. P., C. Xue, and J. Heitman. 2009. A constitutively active GPCR governs morphogenic transitions in *Cryptococcus neoformans*. *EMBO J.* 28:1220–1233.
 40. Cheng, P. Y., A. Sham, and J. W. Kronstad. 2009. *Cryptococcus gattii* isolates from the British Columbia cryptococcosis outbreak induce less protective inflammation in a murine model of infection than *Cryptococcus neoformans*. *Infect. Immun.* 77:4284–4294.
 41. Maurelli, A. T. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiol. Lett.* 267: 1–8.
 42. Hull, C. M., and J. Heitman. 2002. Genetics of *Cryptococcus neoformans*. *Annu. Rev. Genet.* 36:557–615.
 43. Wickes, B. L. 2002. The role of mating type and morphology in *Cryptococcus neoformans* pathogenesis. *Int. J. Med. Microbiol.* 292:313–329.
 44. Zolan, M. E. 1995. Chromosome-length polymorphism in fungi. *Microbiol. Rev.* 59:686–698.
 45. Fierro, F., and J. F. Martin. 1999. Molecular mechanisms of chromosomal rearrangement in fungi. *Crit. Rev. Microbiol.* 25:1–17.
 46. Fries, B. C., D. L. Goldman, R. Cherniak, R. Ju, and A. Casadevall. 1999. Phenotypic switching in *Cryptococcus neoformans* results in changes in cellular morphology and glucuronoxylomannan structure. *Infect. Immun.* 67:6076–6083.
 47. Fraser, J. A., J. C. Huang, R. Pukkila-Worley, J. A. Alspaugh, T. G. Mitchell, and J. Heitman. 2005. Chromosomal translocation and segmental duplication in *Cryptococcus neoformans*. *Eukaryot. Cell* 4:401–406.
 48. Rustchenko, E. 2007. Chromosome instability in *Candida albicans*. *FEMS Yeast Res.* 7:2–11.
 49. Hu, G., I. Liu, A. Sham, J. E. Stajich, F. S. Dietrich, and J. W. Kronstad. 2008. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen *Cryptococcus neoformans*. *Genome Biol.* 9:R41.
 50. Sun, S., and J. Xu. 2009. Chromosomal rearrangements between serotype A and D strains in *Cryptococcus neoformans*. *PLoS One* 4:e5524.
 51. Sionov, E., H. Lee, Y. C. Chang, and K. J. Kwon-Chung. 2010. *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS Pathog.* 6:e1000848.
 52. Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16:351–358.
 53. Young, I. G., L. Langman, R. K. Luke, and F. Gibson. 1971. Biosynthesis of the iron-transport compound enterochelin: mutants of *Escherichia coli* unable to synthesize 2,3-dihydroxybenzoate. *J. Bacteriol.* 106:51–57.
 54. Litwin, C. M., T. W. Rayback, and J. Skinner. 1996. Role of catechol siderophore synthesis in *Vibrio vulnificus* virulence. *Infect. Immun.* 64: 2834–2838.
 55. May, J. J., T. M. Wendrich, and M. A. Marahiel. 2001. The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J. Biol. Chem.* 276:7209–7217.
 56. Jacobson, E. S., and M. J. Petro. 1987. Extracellular iron chelation in *Cryptococcus neoformans*. *J. Med. Vet. Mycol.* 25:415–418.
 57. Jacobson, E. S., A. P. Goodner, and K. J. Nyhus. 1998. Ferrous iron uptake in *Cryptococcus neoformans*. *Infect. Immun.* 66:4169–4175.
 58. Li, L., C. J. Stoeckert, Jr., and D. S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13: 2178–2189.
 59. Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696–704.
 60. Li, W. H., M. Tanimura, and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25:330–342.
 61. Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York, NY.
 62. Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* 56:741–752.
 63. Ngamskulrungraj, P., F. Gilgado, J. Faganello, A. P. Litvintseva, A. L. Leal, K. M. Tsui, T. G. Mitchell, M. H. Vainstein, and W. Meyer. 2009. Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. *PLoS One* 4:e5862.
 64. Dawson, T. L., Jr. 2007. *Malassezia globosa* and *restricta*: breakthrough understanding of the etiology and treatment of dandruff and seborrheic dermatitis through whole-genome analysis. *J. Investig. Dermatol. Symp. Proc.* 12:15–19.
 65. Kamper, J., et al. 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101.
 66. Martinez, D., L. F. Larrondo, N. Putnam, M. D. Gelpke, K. Huang, J. Chapman, K. G. Helfenbein, P. Ramaiya, J. C. Detter, F. Larimer, P. M. Coutinho, B. Henrissat, R. Berka, D. Cullen, and D. Rokhsar. 2004.

- Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* 22:695–700.
67. Stajich, J. E., et al. 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl. Acad. Sci. U. S. A.* 107:11889–11894.
 68. Lucking, R., S. Huhndorf, D. H. Pfister, E. R. Plata, and H. T. Lumbsch. 2009. Fungi evolved right on track. *Mycologia* 101:810–822.
 69. Mackenzie, E. A., and L. S. Klig. 2008. Computational modeling and in silico analysis of differential regulation of myo-inositol catabolic enzymes in *Cryptococcus neoformans*. *BMC Mol. Biol.* 9:88.
 70. Xue, C., T. Liu, L. Chen, W. Li, I. Liu, J. W. Kronstad, A. Seyfang, and J. Heitman. 2010. Role of an expanded inositol transporter repertoire in *Cryptococcus neoformans* sexual reproduction and virulence. *mBio* 1(1): e00084–10.
 71. Steen, B. R., S. Zuyderduyn, D. L. Toffaletti, M. Marra, S. J. Jones, J. R. Perfect, and J. Kronstad. 2003. *Cryptococcus neoformans* gene expression during experimental cryptococcal meningitis. *Eukaryot. Cell* 2:1336–1349.
 72. Hu, G., B. R. Steen, T. Lian, A. P. Sham, N. Tam, K. L. Tangen, and J. W. Kronstad. 2007. Transcriptional regulation by protein kinase A in *Cryptococcus neoformans*. *PLoS Pathog.* 3:e42.
 73. Goulart, L., L. K. Silva, L. Chiapello, C. Silveira, J. Crestani, D. Masih, and M. H. Vainstein. 2010. *Cryptococcus neoformans* and *Cryptococcus gattii* genes preferentially expressed during rat macrophage infection. *Med. Mycol.* 48:932–941.
 74. Lee, A., D. L. Toffaletti, J. Tenor, E. J. Soderblom, J. W. Thompson, M. A. Moseley, M. Price, and J. R. Perfect. 2010. Survival defects of *Cryptococcus neoformans* mutants exposed to human cerebrospinal fluid result in attenuated virulence in an experimental model of meningitis. *Infect. Immun.* 78:4213–4225.
 75. Ma, H., F. Hagen, D. J. Stekel, S. A. Johnston, E. Sionov, R. Falk, I. Polacheck, T. Boekhout, and R. C. May. 2009. The fatal fungal outbreak on Vancouver Island is characterized by enhanced intracellular parasitism driven by mitochondrial regulation. *Proc. Natl. Acad. Sci. U. S. A.* 106: 12980–12985.
 76. Ngamskulrunroj, P., C. Serena, F. Gilgado, R. Malik, and W. Meyer. 2011. Global VGIIa isolates are of comparable virulence to the major fatal *Cryptococcus gattii* Vancouver Island outbreak genotype. *Clin. Microbiol. Infect.* 17:251–258.
 77. Rinder, H.. 2001. Hetero-resistance: an under-recognised confounder in diagnosis and therapy? *J. Med. Microbiol.* 50:1018–1020.
 78. Tanaka, R., H. Taguchi, K. Takeo, M. Miyaji, and K. Nishimura. 1996. Determination of ploidy in *Cryptococcus neoformans* by flow cytometry. *J. Med. Vet. Mycol.* 34:299–301.
 79. Okagaki, L. H., A. K. Strain, J. N. Nielsen, C. Charlier, N. J. Baltes, F. Chretien, J. Heitman, F. Dromer, and K. Nielsen. 2010. Cryptococcal cell morphology affects host cell interactions and pathogenicity. *PLoS Pathog.* 6:e1000953.
 80. Zaragoza, O., R. Garcia-Rodas, J. D. Nosanchuk, M. Cuenca-Estrella, J. L. Rodriguez-Tudela, and A. Casadevall. 2010. Fungal cell gigantism during mammalian infection. *PLoS Pathog.* 6:e1000945.
 81. Batzoglou, S., D. B. Jaffe, K. Stanley, J. Butler, S. Gnerre, E. Mauceli, B. Berger, J. P. Mesirov, and E. S. Lander. 2002. ARACHNE: a whole-genome shotgun assembler. *Genome Res.* 12:177–189.
 82. Schein, J. E., K. L. Tangen, R. Chiu, H. Shin, K. B. Lengeler, W. K. MacDonald, I. Bosdet, J. Heitman, S. J. Jones, M. A. Marra, and J. W. Kronstad. 2002. Physical maps for genome analysis of serotype A and D strains of the fungal pathogen *Cryptococcus neoformans*. *Genome Res.* 12: 1445–1453.
 83. Warren, R. L., D. Varabei, D. Platt, X. Huang, D. Messina, S. P. Yang, J. W. Kronstad, M. Krzywinski, W. C. Warren, J. W. Wallis, L. W. Hillier, A. T. Chinwalla, J. E. Schein, A. S. Siddiqui, M. A. Marra, R. K. Wilson, and S. J. Jones. 2006. Physical map-assisted whole-genome shotgun sequence assemblies. *Genome Res.* 16:768–775.
 84. Ma, B., J. Tromp, and M. Li. 2002. PatternHunter: faster and more sensitive homology search. *Bioinformatics* 18:440–445.
 85. Blanco, E., G. Parra, and R. Guigo. 2007. Using geneid to identify genes. *Curr. Protoc. Bioinformatics Chapter 4:Unit 4.3.* doi:10.1002/0471250953.bi0403s18.
 86. Salamov, A. A., and V. V. Solovyev. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* 10:516–522.
 87. Elsik, C. G., A. J. Mackey, J. T. Reese, N. V. Milshina, D. S. Roos, and G. M. Weinstock. 2007. Creating a honey bee consensus gene set. *Genome Biol.* 8:R13.
 88. Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
 89. Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
 90. Gouy, M., S. Guindon, and O. Gascuel. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–224.
 91. Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8:275–282.
 92. Anisimova, M., and O. Gascuel. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55: 539–552.
 93. Enright, A. J., S. Van Dongen, and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–1584.
 94. Hahn, M. W., T. De Bie, J. E. Stajich, C. Nguyen, and N. Cristianini. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 15:1153–1160.
 95. De Bie, T., N. Cristianini, J. P. Demuth, and M. W. Hahn. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22:1269–1271.
 96. Varma, A., and K. J. Kwon-Chung. 2010. Heteroresistance of *Cryptococcus gattii* to fluconazole. *Antimicrob. Agents Chemother.* 54:2303–2311.
 97. Sia, R. A., K. B. Lengeler, and J. Heitman. 2000. Diploid strains of the pathogenic basidiomycete *Cryptococcus neoformans* are thermally dimorphic. *Fungal Genet. Biol.* 29:153–163.