AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

A New Lineage of *Cryptococcus gattii* (VGV) Discovered in the Central Zambezian Miombo Woodlands

Rhys A. Farrer,[a,b,c,d] Miwha Chang,[e] Michael J. Davis,[e] Lucy van Dorp,[c] Dong-Hoon Yang,[e] Terrance Shea,[d] Thomas R. Sewell,[f] Wieland Meyer,[g,h] Francois Balloux,[c] Hannah M. Edwards,[f] Duncan Chanda,[i] Geoffrey Kwenda,[j] Mathieu Vanhove,[f] Yun C. Chang,[e] Christina A. Cuomo,[d] Matthew C. Fisher,[f] Kyung J. Kwon-Chung[e]

aMedical Research Council Centre for Medical Mycology, University of Exeter, Exeter, United Kingdom
bMedical Research Council Centre for Medical Mycology, University of Aberdeen, Aberdeen, United Kingdom
cUCL Genetics Institute, University College London, London, United Kingdom
dBroad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
eMolecular Microbiology Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA
fMRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom
gMolecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney, NSW, Australia
hWestmead Hospital (Research and Education Network), Westmead Institute for Medical Research, Sydney, NSW, Australia
iAdult Centre of Excellence, University Teaching Hospital, Lusaka, Zambia
jDepartment of Biomedical Sciences, School of Health Sciences, University of Zambia, Lusaka, Zambia

ABSTRACT We discovered a new lineage of the globally important fungal pathogen *Cryptococcus gattii* on the basis of analysis of six isolates collected from three locations spanning the Central Miombo Woodlands of Zambia, Africa. All isolates were from environments (middens and tree holes) that are associated with a small mammal, the African hyrax. Phylogenetic and population genetic analyses confirmed that these isolates form a distinct, deeply divergent lineage, which we name VGV. VGV comprises two subclades (A and B) that are capable of causing mild lung infection with negligible neurotropism in mice. Comparing the VGV genome to previously identified lineages of *C. gattii* revealed a unique suite of genes together with gene loss and inversion events. However, standard *URA5* restriction fragment length polymorphism (RFLP) analysis could not distinguish between VGV and VGIV isolates. We therefore developed a new *URA5* RFLP method that can reliably identify the newly described lineage. Our work highlights how sampling understudied ecological regions alongside genomic and functional characterization can broaden our understanding of the evolution and ecology of major global pathogens.

IMPORTANCE *Cryptococcus gattii* is an environmental pathogen that causes severe systemic infection in immunocompetent individuals more often than in immunocompromised humans. Over the past 2 decades, researchers have shown that *C. gattii* falls within four genetically distinct major lineages. By combining field work from an understudied ecological region (the Central Miombo Woodlands of Zambia, Africa), genome sequencing and assemblies, phylogenetic and population genetic analyses, and phenotypic characterization (morphology, histopathological, drug-sensitivity, survival experiments), we discovered a hitherto unknown lineage, which we name VGV (variety *gattii* five). The discovery of a new lineage from an understudied ecological region has far-reaching implications for the study and understanding of fungal pathogens and diseases they cause.

Cryptococcosis is a severe fungal infection responsible for high levels of mortality and morbidity worldwide (1). The etiological agents are two species complexes of the genus *Cryptococcus*: *Cryptococcus neoformans* and *C. gattii*. While the first described cases of clinical cryptococcosis due to these two distinct species complexes were reported in the mid-1890s under the names *Saccharomyces hominis* (2) and *S. subcutaneous tumefaciens* (3), respectively, clinical *Cryptococcus* isolates have been taxonomically treated as a single species (*C. neoformans*) for more than 100 years (4). Heterogeneity among cryptococcosis-causing yeast isolates became increasingly apparent from the middle of the 20th century onward and led to the recognition of four serotypes (A, B, C, and D) on the basis of the antigenic determinant of capsular polysaccharide (5, 6). Subsequent discovery of two distinct sexual cycles produced by the isolates of serotypes A and D versus serotypes B and C (7, 8) and phylogenetic analysis using various gene sequences (9–11) confirmed these complexes to be genetically divergent enough to be considered separate species. Thus, in 2002, the isolates of serotypes B and C were formally classified as *C. gattii* (12) while *C. neoformans* includes all serotype A and D strains (13).

Over the past 2 decades, population structure analysis of the two species using molecular typing methods such as PCR fingerprinting (14), amplified fragment length polymorphism (AFLP) analysis (15), and multilocus sequencing (16) has demonstrated that both species contain genetically diverse lineages that qualify them to be considered two species complexes, and those species complexes have been further subdivided into numerous molecular types (14, 17). To date, four major lineages, denoted VGI/AFLP4, VGII/AFLP6, VGIII/AFLP5, and VGIV/AFLP7, are recognized for *C. gattii*. Recently, a fifth genotype was described on the basis of a single strain but with several different designations, including clade B (based on multilocus sequence typing [MLST]), VGIIIc/VGIV, and *C. decagattii* (17, 18). Elevation of these five lineages to separate species has been proposed previously (17). However, such taxonomic treatment is currently controversial, mainly due to the lack of clear biological differences between the lineages and of a clear consensus with respect to the limits and numbers of the putative species boundaries. As such, the various *C. gattii* lineages are collectively considered the "*C. gattii* species complex" (18).

In this paper, we describe the discovery of a new lineage/molecular type within the *C. gattii* species complex, which we designate VGV (variety *gattii* five). The six VGV isolates were identified among 32 *C. gattii* isolates recovered from soil, animal dung, and tree bark samples collected in Zambia by Vanhove et al. in 2013 (19). In this paper, we characterize genomic and phenotypic features of the VGV molecular type. Additionally, we present a new, improved genome assembly and gene sets for *C. decagattii* (17) which we confirmed for the first time to represent a separate lineage and which we therefore name "VGVI" for consistency with the other lineages.

RESULTS

Comparative and population genomics for the six lineages of C. gattii. We discovered a new lineage of *Cryptococcus gattii* from a panel of 32 (of 55) genome-sequenced isolates recovered from Southern tree hyrax (*Dendrohyrax arboreus*) middens, midden soil, or tree holes from the Central Zambezian Miombo Woodland ecoregion, a densely forested ecoregion that covers much of Central and East Africa (Fig. 1) (Table 1; see also Table S1 in the supplemental material). Isolates from the new lineage, which we have named VGV, were collected from a 430-km span of northern Zambia, including the Mupata Hills (Luanshya, Copperbelt Province), Mutinondo Wilderness area, and Kapishya (Mpika, Northern Province), suggesting that the lineage has a broad regional distribution across this ecoregion (Fig. 1A). All VGV isolates were identified as serotype B, which also encompasses strains from VGI, VGII, the VGIIIa subgroup, and rare isolates among VGIV strains.

Phylogenetic analyses demonstrated that VGV and VGVI (*C. decagattii*) and the four previously described lineages are genetically distinct from each other (Fig. 2). Indeed, none of the *C. gattii* lineages appear to be the result of hybridization based on the
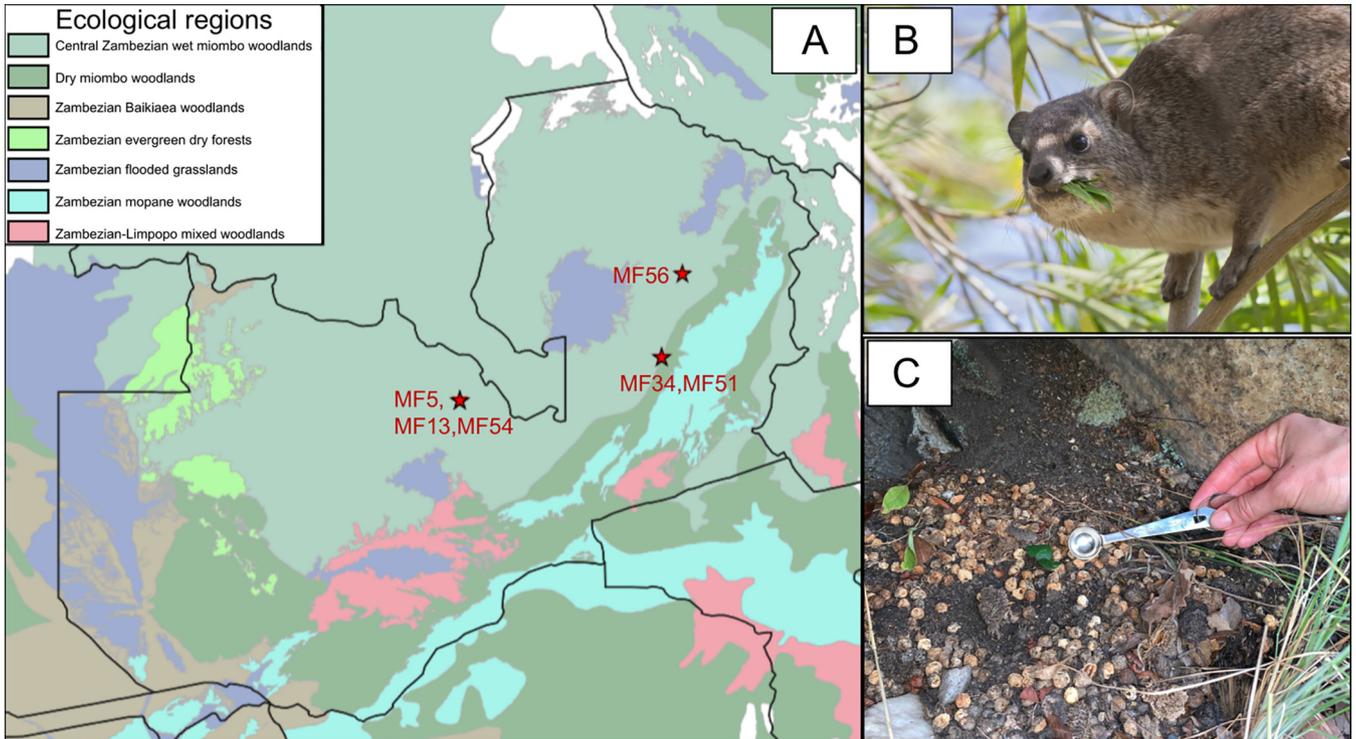
FIG 1 Environmental sampling of *C. gattii* VGV in Zambia. (A) Location of *C. gattii* VGV isolates across Central Zambian Miombo Woodlands. Isolates MF5, MF13, and 54 (Mupata Hills, Copperbelt province) and MF56 (Kapishya, Northern Province) were found in or near Hyrax middens created by Southern tree hyrax (*Dendrohyrax arboreus*). MF34 and MF51 were found from sampling Miombo tree holes in the Mutinondo wilderness area, Northern Province. (B) A tree hyrax feeding on leaves. (C) Sampling from hyrax middens at Kapishya from which MF56 was isolated.

distribution of private alleles (i.e., alleles not shared between lineages) (Fig. 3a and b), maximum likelihood phylogenetic reconstruction (Fig. 2a), Wright's fixation index (F_{ST}) (see Fig. S4 in the supplemental material) or NeighborNet Network data (Fig. 2b). Additional population genetics analyses confirmed low levels of genetic exchange between the six well-resolved *C. gattii* lineages. For example, principal-component analysis (PCA) resolved distinct grouping for the lineages, with the first component (PC1) separating VGII from all other lineages, forming distinct clusters for VGIII and VGVI on PC2 (Fig. 2c). The projection of PC3 and PC4 further allowed identification of distinct tight clusters for the VGI, VGIV, and VGV lineages (Fig. 2d).

The new VGV lineage is represented by six isolates falling within two distinct subclades (A and B). Clade A comprises three VGV isolates (MF5, MF13, and MF54) that were recovered from soil and animal dung sampled in hyrax middens, from which we also identified VGI and VGII isolates (Fig. 1a). Clade B comprises a further three VGV isolates: two (MF34 and MF51) that were recovered approximately 345 km away from clade A and a third (MF56) that was recovered approximately 430 km away from the other clade B isolates. Clade B isolates were recovered from both a tree hole and hyrax

TABLE 1 Environmental isolates of VGIV and VGV from the Central Zambezian Miombo Woodlands

Isolate	Source	Location	Latitude	Longitude	VG type	Serotype
MF5	Hyrax midden	Mupata Hill	−13.12	28.18	VGV	B
MF13	Hyrax midden	Mupata Hill	−13.12	28.18	VGV	B
MF34	Tree hole	Mutinondo	−12.45	31.29	VGV	B
MF51	Tree hole	Mutinondo	−12.45	31.29	VGV	B
MF54	Hyrax midden	Mupata Hill	−13.12	28.18	VGV	B
MF56	Hyrax midden	Kapishya	−11.16	31.60	VGV	B
MF46	Hyrax midden	Mutinondo	−12.48	31.32	VGIV	C
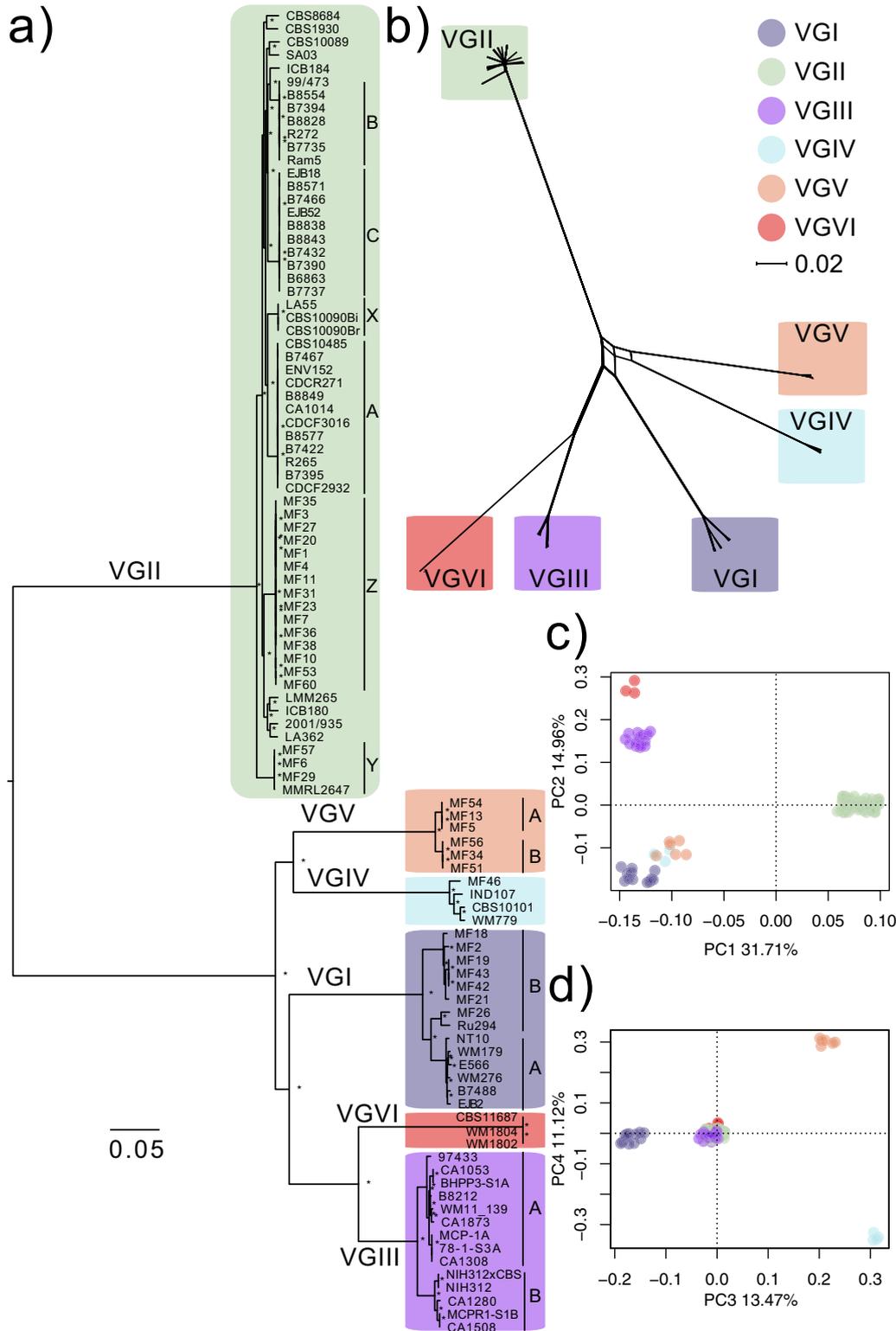
FIG 2 Whole-genome analysis supports classification of VGV as a distinct lineage. (a) Maximum likelihood (RAxML) phylogeny of 101 *C. gattii* genomes generated over all nonambiguous sites with at least one SNP in at least one isolate (1,518,323 sites, or 8.7% of the total genome). Isolate names are colored according to lineage (dark blue = VGI, green = VGII, purple = VGIII, light blue = VGIV, orange = VGV, and red = VGVI). Asterisks indicate 100% bootstrap support at each node corresponding to 1,000 tree-building replicates. (b) SplitsTree NeighborNet Network. (c and d) Principal-component analysis (PCA) of genomic variant sites, showing separation of isolates into lineages (isolates are plotted with 2% random noise [jitter] for clarity of individual points). The first four principal components (PC1 and PC2 [c] and PC3 and PC4 [d]) account for 71.26% of the total genetic variation.
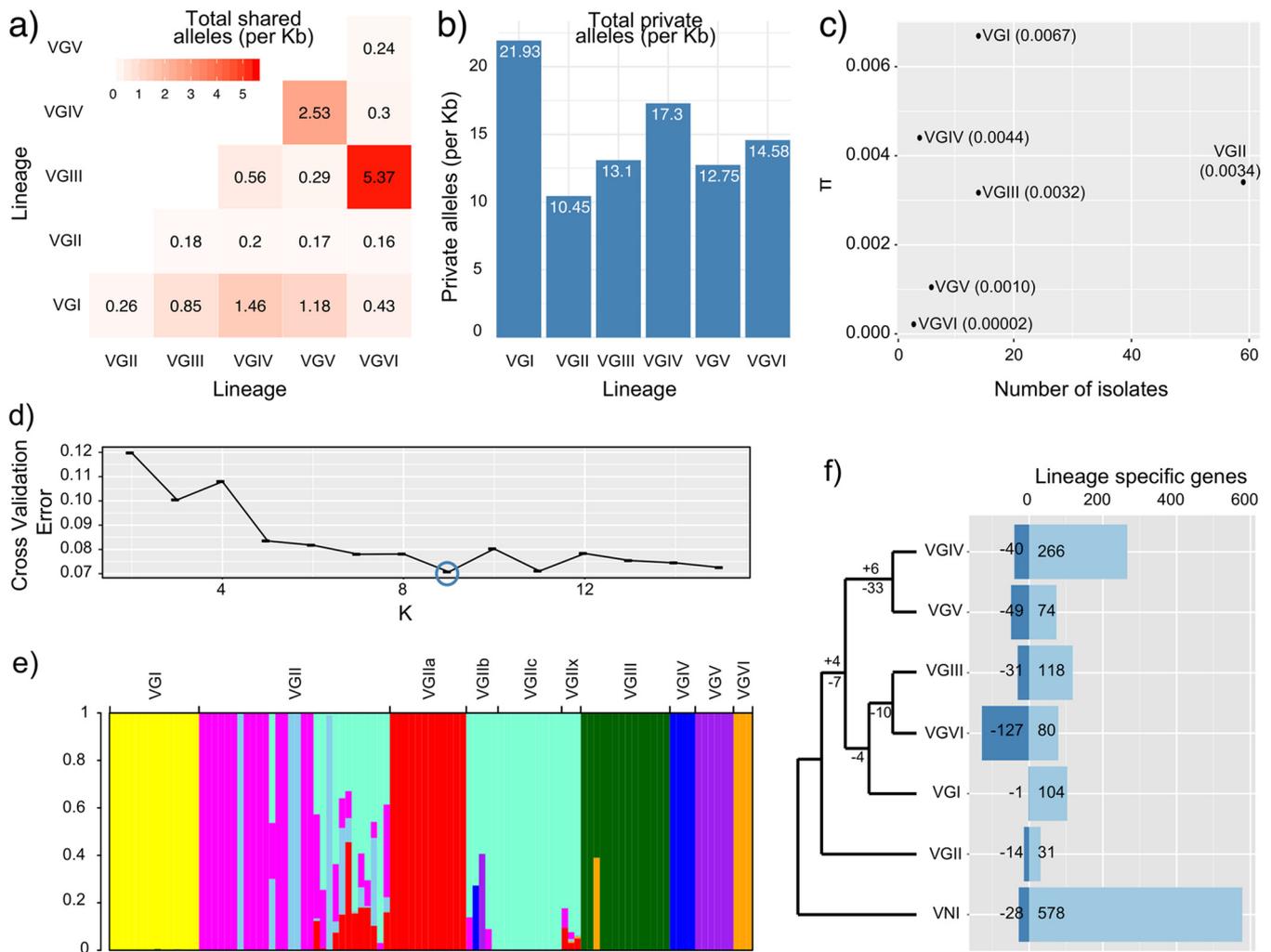
FIG 3 Population genetic analyses for each of the *C. gattii* lineages (based on 101 genomes). (a) Alleles shared between *C. gattii* lineages (SNPs per kilobase). (b) Alleles not shared between *C. gattii* lineages (i.e., private alleles) (SNPs per kilobase). (c) Nucleotide diversity (π) within each lineage against the number of isolates representing each lineage. (d) Admixture *K* optimization based on cross-validation error. (e) Unsupervised ADMIXTURE clustering analysis of all isolates at *K* = 9. (f) Lineage-specific gene counts and lineage-specific gene loss counts. The tree topology is based on the core ortholog RAxML tree, setting equal branch lengths, and the numbers of multilineage-specific gene gains and losses are shown above internal nodes.

middens, showing that the lineage can occupy both tree and dung, environments that are both associated with hyrax activity. The fact that clades A and B were found in different geographic locations might reflect the degree of spatial genetic structure within VGV. All the VGV isolates were located in regions of the granite and acidic kopjes/hills that are found occurring patchily across this ecoregion.

C. gattii VGV is highly divergent from all previously recognized *C. gattii* lineages. VGV isolates differ from VGII (reference isolate R265) by ~0.75 million single nucleotide polymorphisms (SNPs) on average (44 SNPs/kb); thus, those isolates and the other lineages are similarly distant from VGII (Table S1). Analysis of the relative proportions of shared private alleles for the *C. gattii* lineages (Fig. 3a and b) indicates that VGII shared the fewest alleles (<4.6 kb total; <0.2 SNPs per kb) (Fig. 2 and Fig. 3a) with any of the other lineages, reflecting its greater divergence. The newly discovered VGV lineage shared fewer alleles with VGVI (0.24 per kb) and VGIII (0.29 per kb) than with VGI (1.18 per kb) and VGIV (2.53 per kb). The lineages that shared the most private alleles were VGVI and its closest relative, VGIII (92 kb total; 5.37 per kb), which account for an average of 12% of all SNPs (based on alignments to VGII) found in isolates from those lineages.

Nearly 1 in 10 nucleotides in the *C. gattii* genome has an alternative allele across the six lineages (1.55×10^6 sites; 9.01% of the *C. gattii* genome). Indeed, >180 kb of these unique/private alleles were identified for each lineage, including for VGV, which had 220-kb private alleles (12.75 per kb) (Fig. 3b). VGI is the most distinct in terms of both the highest count of private alleles (378 kb/21.93 SNPs per kb) and its nucleotide diversity (π) (Fig. 3c), which is reflected in the three distinct subclades of VGI isolates in the whole-genome phylogeny (Fig. 2a and b). Conversely, the three VGVI isolates are thought to be derived from a single isolate recovered from a patient in Mexico and subsequently distributed to different laboratories where they have been renamed and sequenced (14, 17, 20, 21). Its few clonal differences are illustrated by its low nucleotide diversity (π) (Fig. 3c).

Unsupervised model-based clustering identified highly structured ancestry components enriched in each of the lineages. The clustering solution with the lowest cross-validation error value ($K = 9$) grouped the VGV isolates into a single genetically homogenous group (Fig. 3d and e; see also Fig. S1) while identifying four unique components within the VGII lineage. Of these, subclades VGIIx and VGIIb share small proportions of ancestry with other defined VGII subclades. For example, VGIIb is inferred to share ancestry with other VGII subclades (isolates Ram5 and B8554) and other lineages (B7394 has alleles from VGIV, and B7735 has alleles from VGV). Conversely, none of the isolates in VGIIa and VGIIc have a demonstrable admixture with other subclades or lineages, both being formed by single unique ancestry components. VGIII isolate B8212 (a clinical isolate collected in Oregon, USA, in 2007 [22]) is also modelled as sharing ancestry with VGVI.

Finer-scale clustering was performed by considering patterns of genome-wide haplotype sharing in fineSTRUCTURE (23). Here, VGV isolates forming a separate cluster with greater haplotype similarity to isolates from VGI, VGIII, and VGIV than VGII (Fig. S2). While the haplotype sharing patterns overwhelmingly accorded with each lineage being genetically distinct, a notable exception was VGIII isolate B8212, which shares haplotypes with VGIV and VGVI (also in accordance with model-based clustering), perhaps owing to a small amount of genetic exchange with one or both of those lineages. As also observed using ADMIXTURE-based clustering (Fig. 3e), two isolates from VGII, B7394 and B7735, were also genetically distinct and were assigned to their own cluster which was most closely related to isolates from subclade VGIIb.

All six VGV isolates were haploid, with no evidence for aneuploidy based on allele frequencies and depth of coverage (Fig. S3). However, we did observe copy number variation (CNV) between the three VGVI isolates derived from a single clinical isolate from Mexico (14, 17, 20, 21). Specifically, isolate CBS11687 acquired an ~200-kb duplication of supercontig 5 (sc5) (position 1040000 through the end of the supercontig). Separately, isolate WM1804 had a smaller, 40-kb duplication on sc21 (positions 150,000 to 190,000). Isolate WM1802 had neither CNV. In terms of base changes, the three VGVI isolates (WM1802, WM1804, and CBS11687) differed by only 419 SNPs, with the lowest number found between WM1804 and CBS11687 ($n = 126$) and the highest number found between WM1802 and CBS11687 ($n = 315$). These genetic differences may have occurred as a result of microevolution during or following passaging or cryopreserving, although large CNVs are common in *C. gattii* (24, 25). All of the newly isolated VGI ($n = 7$) and VGVA ($n = 3$) samples from Zambia had a small, <10-kb duplication within supercontig 6 of the R265 genome (kb position 400 to position 410). This genomic region encodes a single 87-amino-acid (aa) protein that is conserved in *C. neoformans* and *C. gattii* but has no functional annotation (by PFAM, GO terms, KEGG-EC, TMHMM, or SigP4).

The results from our phylogenetic and population genetic analyses are in line with previous work (26), indicating that lineages within the *C. gattii* species complex have remained largely genetically isolated since their divergence. Pairwise lineage calculations of θ, representing Weir's formulation of Wright's fixation index (F_{ST}), suggest very low levels of genetic exchange between the lineages (Fig. S4), which is also reflected in analyses of genetic structure (Fig. 2 and 3; see also Fig. S1 and S2). Both depth-of-

TABLE 2 The genome assemblies used for phylogenetic analysis and orthology detection[a]

Species	Lineage	Isolate	Genome ID	Genome size (Mb)	No. of contigs/scaffolds	No. of protein-coding genes
C. gattii	VGI	WM276	CNB_WM276_v2	18.36	14	6,565
C. gattii	VGI	EJB2	Cryp_gatt_EJB2_V1	17.53	282	6,763
C. gattii	VGII	R265	Cryp_gatt_R265	17.55	14	6,192
C. gattii	VGII	Ram5	Cryp_gatt_Ram5_V1	17.32	24	6,463
C. gattii	VGIII	CA1280	Cryp_gatt_CA1280_V1	17.49	89	6,683
C. gattii	VGIII	CA1873	Cryp_gatt_CA1873_V1	17.39	33	6,634
C. gattii	VGIV	IND107	Cryp_gatt_IND107_V2	17.56	34	6,696
C. gattii	VGV	MF34	Cryp_gatt_MF34*	17.89	15	6,322
C. gattii	VGVI	WM1802	Cryp_deca_WM1802*	17.42	863	6,092
C. neoformans	VNI	H99	CNA2	18.87	14	6,962

[a]Asterisks (*) indicate genome assemblies that are newly described in this paper. All others have been described previously (26).

coverage plots and F_{ST} nonoverlapping sliding 10-kb window plots across the mating type locus (*MAT*) at the start of supercontig 18 demonstrate that all VGIV and VGV isolates included in this study were *MATα* (the reference genome of R265 is *MATα*; high depth of coverage and θ values of >0.98 across the *MAT* loci). In contrast, for VGI, VGII, VGIII, and VGVI, *MAT***a** isolates were included in our panel.

Genome assembly and analysis of VGV reveals *C. gattii* lineage-specific differences. We assembled and annotated a nearly complete genome for the newly discovered *C. gattii* VGV lineage (isolate MF34) using both Oxford Nanopore and Illumina sequencing reads. The resultant assembly consisted of 15 contigs corresponding to the 14 chromosomes; the single break in one chromosome corresponds to the ribosomal DNA (rDNA) region. Other than underrepresenting rDNA genes, this assembly provides a complete representation of the genome, with telomeric repeats (TTAGGG) present at 28 contig ends. Gene annotation revealed 6,322 predicted protein-coding genes, which is similar to the data from the seven other representative *C. gattii* isolates with publicly available complete genomes (26, 27) representing the four previously known major lineages (ranging from 6,092 to 6,763), as well as *C. neoformans* H99 (28) ($n = 6,962$) (Fig. S5 and S6).

To establish the evolution of protein-coding genes in *C. gattii*, we compared the gene content data for two representative annotated genomes per lineage where possible (no second annotated reference genomes were available for VGIV, VGV, and VGVI), identifying 4,565 single-copy core orthologs that are shared among the five lineages of *C. gattii* and *C. neoformans* (~74% of *Cryptococcus* genes) (Table 2). For VGVI, we sequenced and assembled the WM1802 isolate, obtaining a similar genome length (17.42 Mb) and a similar count of protein-coding genes ($n = 6,092$). For VGII, we included the updated VGII R265 PacBio assembly in our panel of genomes (29) (Table 2). Detection orthology analysis performed between just the two R265 assemblies identified 91% of genes in 1:1 orthology ($n = 5,642$), ~4% of genes unique to the new assembly ($n = 252$), and ~6% of genes in paralogous clusters ($n = 364$). The previous VGII R265 assembly had 635 genes that were not called in the new assembly, likely representing a difference in the annotation protocol. Analysis of core eukaryotic genes (CEGs) and BUSCO revealed a high level of completeness of gene sets and increased completeness in the new annotation (Fig. S6). Furthermore, all assemblies generated using long-read sequencing technology assembled into 14 scaffolds/supercontigs, suggesting that all *Cryptococcus* lineages/species have conserved numbers of chromosomes.

Ortholog amino acid differences within and between lineages were consistent with results from our phylogenetic and population genetic analyses. VGV MF34 had the highest amino acid sequence similarity to VGIV IND107 (53,000 amino acid differences = 97.55% similarity), which was observed in both alignment-based and ortholog-based phylogenies (Fig. 2 and Fig. 4). The most similar interlineage orthologs were those of VGIII and VGVI (49,500 predicted amino acid differences = 97.71% similarity) (Table S1), while the most distinct pairwise comparisons were those between *C. gattii*
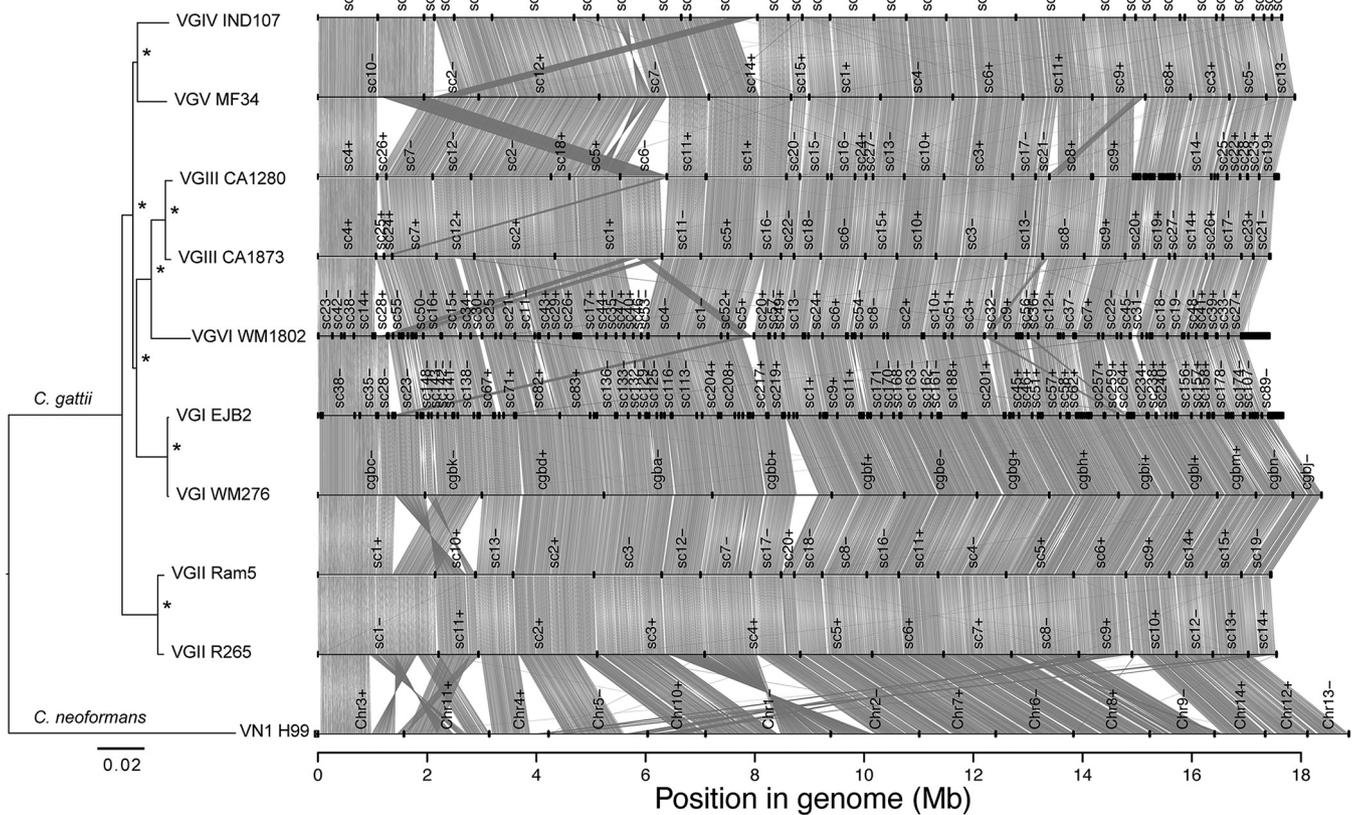
FIG 4 A phylogenetic tree for 10 *Cryptococcus* genomes belonging to the six *C. gattii* lineages and 1 *C. neoformans* lineage outgroup alongside their genome synteny. The phylogenetic tree was constructed in RAxML, with branch lengths indicating the mean number of nucleotide substitutions per site. To the right is a synteny plot, visualizing regions that span three or more orthologs between any two species as a connected gray line. Supercontig numbers are shown above each genome axis if longer than 400 kb, where a plus sign (+) represents the forward orientation and a minus sign (−) represents the negative orientation.

and *C. neoformans* (between 205,000 and 218,000 amino acid changes; ~90% protein similarity).

Overall, synteny is conserved within *C. gattii* (26), though with notable differences between some lineages. For example, VGV has a single 171-kb inversion on supercontig 7 (positions 544,906 to 716,249) compared with the middle of VGIV IND107 supercontig 7 and the middle of VGIII CA1280 supercontig 5 (Fig. 4). VGVI also has some syntenic differences from its closest relative, VGIII (Fig. 2) (see also Fig. 3 and Fig. 4). For example, approximately half of VGVI supercontig 5 is syntenic for the start of VGIII (CA1873) supercontig 16, while the second half of VGVI supercontig 5 is syntenic for a middle region of VGIII supercontig 1, indicative of a chromosomal translocation. Further improvements and additional genome assemblies should establish the full number and genetic impact of lineage-specific genomic rearrangements.

Lineage-specific genes and multilineage-specific genes (found in two or more lineages) were identified in each of the lineages (Fig. 3f; see also Fig. S7). Many of these lineage-specific genes (223/605; 37%) were previously identified from a panel of genome assemblies without the addition of VGV and VGVI (26). A further 53/605 (9%) of the newly detected lineage-specific genes were previously categorized as multilineage-specific genes. The lineage-specific genes in newly sequenced lineages (VGV and VGVI) include 74 genes that were unique to VGV and 49 genes that were uniquely absent in VGV. Genes unique to VGV include those encoding two sugar transporters (D1P53_002216 and D1P53_002944), an alcohol dehydrogenase (D1P53_004471), and an aldehyde dehydrogenase (D1P53_006242). Conversely, eight transmembrane proteins and a single uncharacterized secreted protein were uniquely missing in VGV. All of the genes involved in the ergosterol biosynthesis pathway were present in single-copy
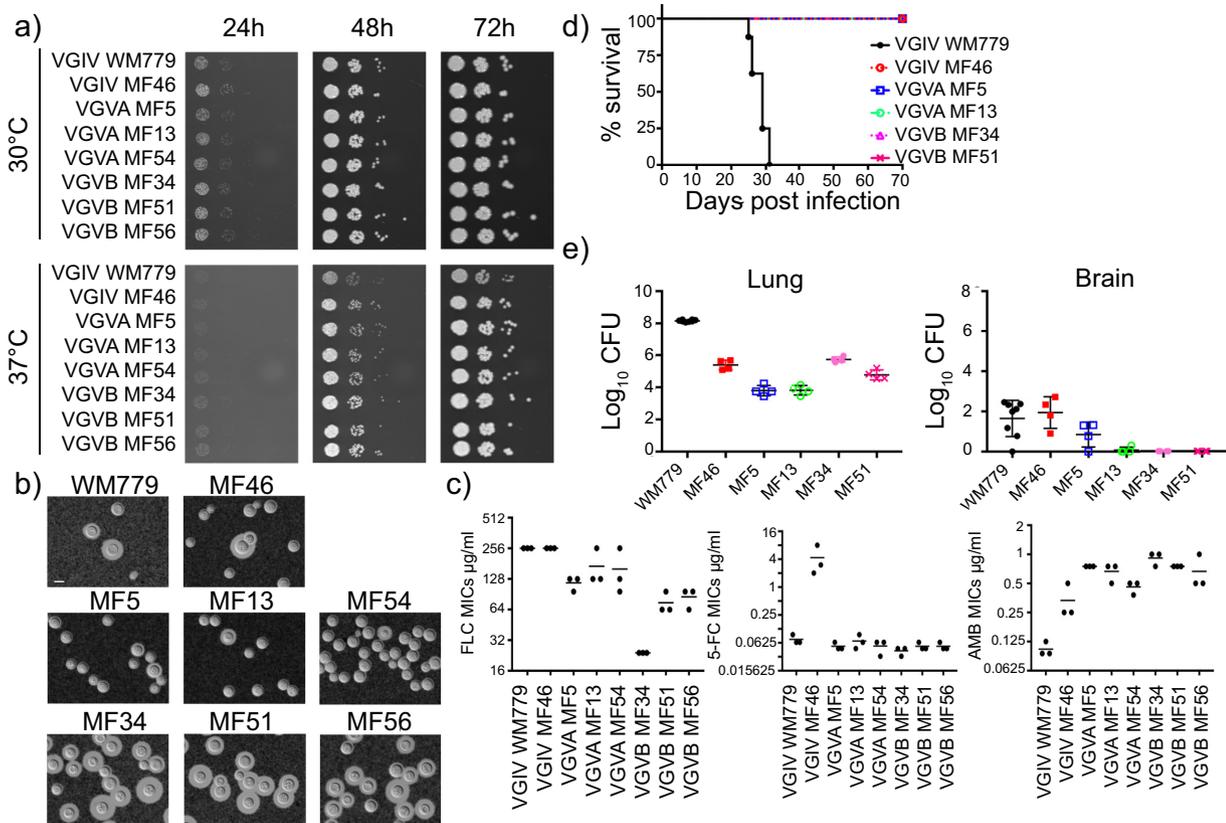
FIG 5 Phenotypic characteristics. (a) Growth of six VGV isolates and two VGIV control isolates at 30°C and 37°C on YEPD agar. (b) India ink staining of VGV cells grown in YEPD broth for 24 h at 30°C. The isolates of clade A (MF5, MF13, and MF53) produced thinner capsules than the isolates of clade B cells, which produced capsules similar in size to those produced by the two VGIV control isolates. Bar = 5 μm. (c) MIC of VGV isolates for FLC, 5-FC, and amphotericin B. All tested isolates showed high FLC MICs ranging from 20 to 256 μg/ml. The MIC for 5-FC was low, except for MF46, an environmental isolate of VGV, for which the MIC for 5-FC averaged 4 μg/ml. All VGV isolates showed higher MICs for amphotericin B than for the VGIV controls. (d) Survival curve of mice infected by four VGV isolates (intrapharyngeal aspiration of 5,000 cell/mouse) and two VGIV control isolates. Only the mice infected with VGIV isolate WM779 succumbed to infection. (e) Lung (left) and brain (right) fungal loads. VGV isolates grew moderately well in lungs, but the CFU levels in the brains were negligible.

form in VGV. The VGVI WM1802 genome includes 80 genes that are unique and lacks 127 genes that are uniquely absent. Among the unique genes in WM1802, 14 are predicted to be involved in transport and include genes encoding three monosaccharide transporters, one hexose transporter, one cadmium ion transporter, and one monocarboxylic acid transporter.

Predictably, *C. neoformans* VNI H99 had the greatest number of lineage-specific/absent genes, with 578 unique genes and 28 absent genes. These included 47 genes predicted to encode transmembrane proteins (including four sugar transporters, five major facilitator superfamily [MFS] transporters, and a caffeine resistance transporter), and 34 secreted proteins. Fewer genes were uniquely absent in *C. neoformans*, which included genes encoding eukaryotic translation initiation factor 3 subunit B, an ACC oxidase, a copper amine, an allantoate permease of the major facilitator, and a 3-hydroxyacyl-dehydrogenase with oxidoreductase activity. Full details of all lineage-specific genes are provided in Table S1.

Phenotypic characteristics of VGV. All six isolates that belonged to the new VGV lineage based on whole-genome sequencing (Table 1) were first incorrectly identified as VGIV based on the *URA5* restriction fragment length polymorphism (RFLP) banding pattern. Unlike most of VGIV, all six VGV isolates were serotype B. These were further characterized as *MATα* and as melanin and urease positive (Fig. S8) and grew well at 37°C (Fig. 5A). However, the VGV strains grew slightly more slowly on canavanine glycine bromothymol blue (CGB) agar (Fig. S8) than the serotype C VGIV strain (MF46)

isolated from the same environment in the Central Miombo Woodland (Table 1). Hence, the positive blue/green color development on CGB agar took longer in VGV than in the control strains (Fig. S8). Since VGV is genetically closest to VGIV (Fig. 2a and b; see also Fig. S2), two serotype C VGIV strains were used as control isolates for further phenotypic comparisons (WM779 isolated from a cheetah in South Africa [16] and MF46 isolated from Miombo tree bark in Zambia near Hyrax middens) (Table 1).

The size and morphology of VGV yeast cells were typical for *Cryptococcus* and indistinguishable from those of cells of the control strains (Fig. 5B). Two distinct patterns of capsule formation were found among the six VGV isolates grown in yeast extract-peptone-dextrose (YEPD) broth (Fig. 5B). The isolates recovered from clade A (MF5, MF13, and MF54) produced thinner (≤1-μm-thick) capsules than those recovered from clade B (MF34, MF51, and MF56), which produced thick (2-to-4-μm-thick) capsules similar to those seen with the VGIV control strains.

The VGV isolates and the control strains of VGIV manifested unusually high resistance toward fluconazole (FLC), particularly given that they were sampled from an environmental niche. The three isolates of VGV clade A were more resistant to FLC, with drug MICs of ≥128 μg/ml, than the three isolates in clade B, which showed drug MICs of 24 to 64 μg/ml. All six VGV isolates showed a drug MIC of 0.0625 μg/ml for 5-fluorocytocine (5-FC), a value similar to that shown by WM779. The drug MIC of MF46 for 5-FC, 4 μg/ml, was unusually high. The VGV amphotericin B MIC ranged between 0.5 and 1 μg/ml, higher than the levels measured for the control strains, which showed drug MICs below 0.5 μg/ml (Fig. 5C).

To explore the relative pathogenicities among VGV subclades we selected two isolates from clade A (MF5 and MF13) and two isolates from clade B (MF34 and MF51) for inoculation in mouse models. Mice infected by all four isolates survived for 70 days, while all the mice infected by WM779, a virulent serotype C control isolate, succumbed to infection within 30 days postinfection (Fig. 5D). VGIV environmental isolate MF46 (serotype C) caused no death in the mouse model. Fungal loads in the lungs of VGV-infected mice were substantially lower than those seen with the mice infected with WM779 and slightly lower than those seen with the mice infected with MF46. Brain fungal loads of mice infected with the VGV strains were also low to negligible. The control isolates of VGIV showed little neurotropism (Fig. 5E). Histopathological analysis of the lungs demonstrated significant pathology in WM779-infected mice, with yeast found throughout the lung together with notable disruption of lung tissue. In many locations, extensive leukocyte recruitment was evident in areas of concentrated infection (Fig. 6).

Histopathological analysis of the VGV isolates displayed substantially lower pulmonary yeast levels. That said, mice infected by MF46, MF34, and MF51 had higher levels of yeast than those infected by MF5 and MF13 as shown by both CFU counts (Fig. 5E) and the results of histopathological analysis (Fig. 6), in which MF51 was shown to represent clade B. The lung histopathology of the mice infected by MF34 was similar to that of the mice infected by MF51 (data not shown). Notable for its absence, leukocyte infiltration was mostly low or absent from sites of VGV infection. MF13 showed some leukocyte infiltration to a subset of infectious foci (Fig. 6).

Identification of VGV by *URA5* RFLP. The patterns of *URA5* RFLP, obtained by double digestion with Sau96I and HhaI, have been widely used to identify the lineage/molecular type in both *C. neoformans* and *C. gattii* species complexes (14). The *URA5* RFLP of clade A isolates obtained by Sau96I/HhaI digest showed a pattern identical with that of VGIV (Fig. S9A). Those of clade B, however, produced an additional 1.3-kb amplicon which was absent from clade A and from all of the other VG molecular type reference isolates. This 1.3-kb amplicon was present even in uncut DNA of clade B isolates (Fig. S9A and B), but its nature is not known at this juncture. Since the *URA5* RFLP patterns of VGIV and VGV were not conclusively different, we compared the *URA5* gene sequence of VGV isolate MF34 to that of VGIV reference strain WM779 to identify possible restriction enzymes that can clearly distinguish the two lineages. This led to

Alcian Blue, 2.5X	Alcian Blue,10X	H&E, 10X

VGIV
WM779

VGIV
MF46

VGV
MF5

VGV
MF13

VGV
MF51

VGV D20 LUNGS
Bar, 50μm (X10 magnification) and 200μm (X2.5 magnification)
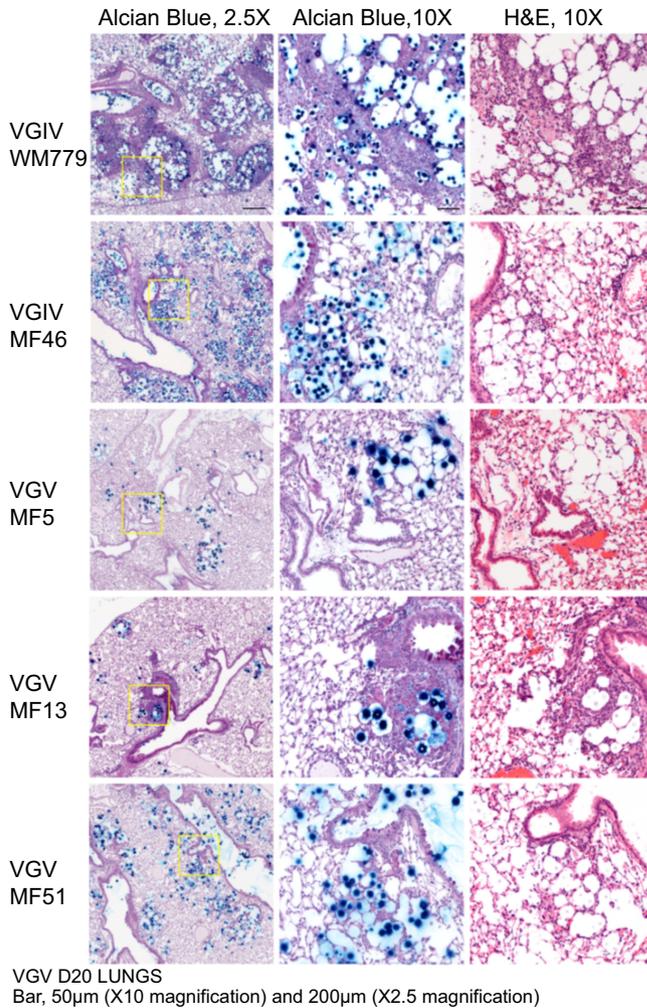
FIG 6 Histopathology of lung infected by VGV isolates. Sections of the mouse lungs infected by three different VGV isolates and two VGIV control strains were stained with alcian blue and periodic acid-Schiff stain and counterstained with hematoxylin (left and middle columns) or with standard hematoxylin and eosin (right column). Note that alcian blue stains cryptococcal cells blue. Images in the left columns were acquired using a 2.5× objective. Images in the middle and right columns are higher-magnification (10×) images of the areas indicated by the yellow boxes in the 2.5× images.

our identifying two highly discriminatory restriction enzymes: StuI and EarI. The expected sizes (measured in base pairs) of the *URA5* gene fragments resulting from StuI digestion are as follows: 221 bp, 237 bp, and 322 bp in VGIV and 237 bp and 543 bp in VGV. The EarI digestions produced 247-bp and 501-bp fragments in VGIV and 247-bp and 300-bp fragments in VGV. We compared *URA5* RFLP data from 17 VGIV isolates (Table S1) with the data from 6 VGV isolates obtained by StuI or EarI digestion, and the results are shown in Fig. S9C and D.

Type strain of VGV. We have designated MF34 (clade B isolate) as the type strain of VGV. MF34 was isolated from a tree hole located in Mutinondo (latitude −12.45, longitude 31.29), Central Zambezian Miombo Woodlands (Table 1). Its genome has been assembled and annotated to near-completion (15 scaffolds with N_{50} = 1.3 Mb and telomeric repeats at 28 contig ends). MF34 is serotype B and *MATα* and causes mild pneumonia in C57BL/6 mice with negligible neurotropism.

DISCUSSION

Over the past decade, increased sampling worldwide and whole-genome sequencing (WGS) methods have uncovered greater genetic diversity of important pathogens,

including the *C. neoformans* and *C. gattii* species complexes. For example, sampling from Botswana revealed the existence of the *C. neoformans* VNB lineage (30), which itself has recently been shown to be deeply split into two genetically isolated lineages, VNBI and VNBII (31). Thus far, VGVI is the only lineage that exists as a single genotype since the three isolates previously designated *C. decagattii* appear to have been isolated from the same patient (21). The previously identified lineages of *Cryptococcus* have recently been designated separate taxonomic species based on phylogenetic species recognition criteria (17). While we agree that *Cryptococcus* contains a number of genetically diverse and monophylectic clades that can be viewed as species under an evolutionary species concept (32), we have previously argued that it is premature to give each clade a separate taxonomic name at this juncture (18, 33). One notable concern raised by Kwon-Chung et al. (18) was that the proposed seven-species taxonomy (33) was likely to be unstable due to incomplete knowledge of the true extent of *Cryptococcus* diversity worldwide. Our discovery of *C. gattii* VGV from the Miombo woodlands of Zambia clearly shows that we have not yet achieved a full understanding of the global biodiversity of *Cryptococcus* and that further exploration will likely yield additional phylogenetic species. Until we have a more accurate consensus on the true numbers of *Cryptococcus* lineages, we propose that the names "VN" and "VG" serve as practical "Zip code" identifiers within *C. neoformans* and *C. gattii*, offering a convenient way to describe newly discovered lineages or recombinants without introducing unwanted nomenclatural instability and confusion.

Our discovery of *C. gattii* VGV in samples from hyrax-associated environments suggests an association with these mammals. Hyrax are small herbivores that are most closely related to elephants (Proboscidea) and sea cows (Sirenia) and are characterized by the behavior of defecating in communal latrines, usually located in crevices in rocky kopjes, over many generations (34). These locations are often sheltered in rocky caves and droppings are likely to accumulate for upwards of 50,000 years, in some cases forming a stable paleoenvironmental hot spot of urea-rich nitrogenous material (35). *Cryptococcus* has a pronounced trophism for urea as a nutritive substrate, and pigeon guano is known to support prolific growth of *C. neoformans* and (to a lesser extent) *C. gattii* (36). Our finding that hyrax middens are hotspots of *Cryptococcus* diversity suggests that their ecological stability in landscapes that are low in nitrogen availability may lead to them being important arenas for the evolution of *Cryptococcus* and that they will likely be fertile ground for further discovery of diversity within this genus.

Fungal association with small mammals may suggest adaptations that confer pathogenicity, known as the "endozoan, small-mammal reservoir" hypothesis (37), and deserves to be explored further following our findings of an association of *Cryptococcus* with hyrax. Accordingly, alongside further study of potential mammalian reservoirs, the search for VGV clinical isolates is also needed in order to understand the true virulence potential of VGV and whether it can spill over into humans. Murine models have shown that environmental isolates are less virulent than clinical isolates of the same molecular type in both the *C. gattii* and *C. neoformans* species complexes, suggesting that polymorphic virulence factors exist (38, 39). However, despite its great genetic distance from all other lineages, the new VGV lineage is not clearly distinguishable from others by existing methods such as serotyping or the routinely used approach of RFLP analysis employing Sau96I and HhaI *URA5* digestion (14). Thus, it is possible that previously analyzed isolates belonging to VGV may have been misidentified using non-WGS methods. The most likely candidates for the search of clinical VGV are VGIV serotype B isolates recovered from patients. Geographically, the most likely place to find the VGV clinical isolates appear to be in sub-Saharan Africa since the current panel of isolates were found in the Zambian environment within an ecoregion that includes Tanzania, Burundi, Democratic Republic of the Congo, Angola, and Malawi.

Previous work has shown that most isolates of the *C. gattii* species complex cause pulmonary infection in a murine model with low neurotropism (20, 40, 41). The four VGV isolates tested here were less neurotropic than the MF46 VGIV isolate that was collected from the same Zambian environment, and all the examined Zambian envi-

ronmental isolates were significantly less virulent than a VGIV control strain, WM779. It remains to be shown whether the differences in neurotropism are due to lineage-specific genes or to alleles in VGV. As previous work has shown in *C. neoformans* (42), capsule size difference manifested by clades A and B *in vitro* was unrelated to virulence in mice.

Although serotypes have not yet been conclusively linked to virulence in *Cryptococcus*, they remain important for strain identification. The majority of *C. gattii* isolates tested to date have been of serotype B, except for a subset of VGIII and the majority of VGIV isolates, which are serotype C. The six VGV isolates are also all serotype B—but, due to the lower growth rate on CGB agar, the CGB reaction was weaker than that seen with other isolates of serotype B or serotype C. It took 24 h longer for VGV than for the other VG isolates (VGI to VGIV) to turn the medium dark blue. As the six VGV isolates were all serotype B whereas the majority of VGIV isolates (the most closely related lineage) reported thus far have been serotype C, it is possible that VGV may also occur in serotype C. Additional environmental sampling of VGV is therefore necessary to establish the dominant serotype, since the current sample size of six is insufficient for definitive conclusions.

Surprisingly, five of the six VGV isolates and the two control VGIV isolates were highly resistant to fluconazole (MIC of >64 μg/ml), a commonly used antifungal drug. The three isolates of VGV clade A were more resistant to FLC than the isolates of VGV clade B. Although the *C. gattii* species complex was previously known to be more resistant to FLC on average than *C. neoformans* (43), such high resistance to FLC in environmental isolates is notable and was not previously reported (44). All of the VGV isolates had identical nucleotide sequences for *ERG11* and *AFR1*, demonstrating that the resistant isolates did represent not a result of genetic differences in the target or transporter of FLC. However, innate fungal resistance to FLC can be due to multiple factors besides the *ERG11* gene or efflux pumps and the mechanism(s) of FLC resistance in VGV remains a subject for future investigation. Why environmental VGV isolates should have such high resistance to azoles is unclear, as, given the relatively pristine environments from which they were recovered, it is unlikely that they had come into contact with agrichemicals. More likely, fluconazole resistance is a pleiotrophic effect that has evolved as a consequence of exposure to xenobiotics other than azoles. Further investigations into the evolution of FLC resistance in VGV may take on additional importance as clinical cases due to VGV are a distinct possibility in the Sub-Saharan regions where 12% of the members of the Zambian population are living with the HIV (https://www.unaids.org/en/regionscountries/countries/zambia).

In this paper, we present a nearly complete genome assembly for VGV type strain MF34. The MF34 genome allowed us to conclusively establish that VGV is a lineage of *C. gattii* that is separate and distinct from any previously identified and that it is not the result of hybridization, as has been seen for other divergent isolates (31). Indeed, while both *Cryptococcus* species complexes appear to have a conserved chromosome number of 14 based on the current panels of assembled and annotated genomes available, intrachromosomal and interchromosomal rearrangements as well as large CNVs appear to be common. This chromosomal variation may provide the genetic basis for phenotypic variation and may act as a genetic barrier to recombination between more highly divergent isolates such as those from separate lineages. At the within-lineage level, there are also a number of unique and uniquely lost "lineage-specific" genes which may contribute to phenotypic differences between lineages. However, it should be noted that many of the main phenotypes routinely measured, including virulence in animal models, growth rates, and ability to cause pulmonary versus central nervous system (CNS) infections, appear to differ as much within as between lineages.

One line of future inquiry that may help to explain this phenotypic diversity may come from the characterization of further transcriptional differences. For example, VGII has been shown previously to upregulate many of the ergosterol genes during coincubation with bone-marrow derived macrophages (45) and it will be important to determine whether other traits exist which differentiate the lineages of *Cryptococcus*.

Further, it will be important to examine whether similar lineage-specific differences underpin VGV's increased FLC resistance and whether clinically relevant traits such as drug resistance are linked to the environment within which these isolates have evolved. Ultimately, our report testifies to the deep reservoir of diversity that exists within *Cryptococcus*, which, despite decades of research into this genus, still harbors abundant surprises.

MATERIALS AND METHODS

Library preparation and sequencing of Zambian isolates. Environmental sampling took place in January and September of 2013. Samples were collected using Transwab Amies swabs (MWE; code MW170) and sterilized 30-ml screw-cap glass bottles. Amies liquid transport swabs were taken from tree bark ($n = 20$), soil ($n = 19$), and cracks in granite kopjes or droppings from rock Hyrax middens ($n = 16$). Samples were collected and processed according to previously established protocols (46, 47), and the samples were kept at 4°C before being processed on niger seed agar. All samples were collected under license from the Zambian Wildlife Authority (ZAWA).

Single colonies purified from the original isolation media were maintained under cryopreservation conditions at −80°C at Imperial College in London since 2013. The isolates were has been changed to revived on YPD agar (yeast extract 1%, peptone 1%, dextrose 2%) and incubated at 30°C before use. Genomic DNA was isolated with the cetyltrimethylammonium bromide (CTAB) extraction method as described previously (48) with modifications. Paired-end libraries (150 bp) were prepared and sequenced using an Illumina HiSeq 4000 platform by Novogene (Davis, CA). Two Oxford Nanopore libraries of isolate MF34 were constructed from genomic DNA using a one-dimensional (1D) library construction kit (model no. SQK-LSK109). A total of 243,660 reads with an N_{50} of 9,827 were generated on a FLO-MIN106 flow cell using a Minion. Reads were base called using Albacore v2.3.1. This resulted in 923,997,900 total bases (~46× coverage).

Genome assembly and annotation. For isolate MF34, a hybrid assembly of Oxford Nanopore long reads and Illumina short reads was generated. An initial assembly of the Oxford reads was generated using Canu v1.5 (49) with parameter genomeSize = 20,000,000. The assembly was inspected for the presence of telomeric repeat (TTAGGG) at contig ends; for two contig ends missing a telomeric repeat, contigs were extended by aligning unassembled Canu contigs to these ends using NUCmer v3.1 (50). Base-called reads were then aligned to the contigs using the Burrows-Wheeler transform algorithm (BWA mem) (51) with flag "-x ont2d" and the alignments used for polishing with nanopolish (52). Two rounds of Pilon v1.13 (53) correction were performed using Illumina BWA read alignments (51). Paired Illumina sequences of *C. decagattii* (VGVI) were assembled and scaffolded using SPAdes v3.1.1 (54) with *k*-mer lengths of 21, 33, 55, and 77. A summary of statistics for the assembly is provided in Table 2. Reads were aligned to the assembly with BWA v0.7.4-r385 mem (51), and Pilon v1.12 (53) was further used to improve the assembly. Scaffolds that were less than 1 kb were removed. The genome assembly has been submitted to NCBI (see below for project accession number).

The *C. gattii* VGV MF34 and VGVI WM1802 genomes were annotated using Genemark (55), BLASTx against Swiss-Prot (56) and KEGG (57), and HMMER hmmscan (58) against PFAM (59). We ran tRNAscan (60) and RNAmmer (61) to identify non-protein-coding genes. Gene predictions were checked for a variety of issues, including overlapping of noncoding genes, overlapping of coding genes, and the presence of in-frame stops. Genes were named according to evidence from BLASTx and HMMER in the following order of precedence: (i) Swiss-Prot (56), (ii) TIGRfam (62), and (iii) KEGG (57) (where BLASTx hits must meet the 70% identity and 70% overlap criteria to be considered a good hit and for the name to be applied). Otherwise, genes were classified as hypothetical proteins.

Genes were functionally annotated by assigning PFAM domains (59), GO terms, KEGG assignment, and ortholog mapping to genes of known function. HMMER3 (58) was used to identify PFAM (release 27) domains, and BLASTx was used against the KEGG v65 database (57) (E value, $<1 \times 10^{-10}$). GO terms were assigned using Blast2GO version2.3.5 (63), with a minimum E value of 1×10^{-10}. SignalP 4.0 (64) and TMHMM (65) were used to identify secreted proteins and transmembrane proteins, respectively (see Fig. S5 in the supplemental material). Gene sets were aligned to the 248 core eukaryotic genes (CEGs) and BUSCO basidiomycota_odb9 set to evaluate completeness (Fig. S6).

Read alignment and variant identification. The 36 newly sequenced isolates from this study were compared to an additional 65 isolates that had been sequenced and described in previous studies (20, 26, 38, 66, 67). These additional isolates were obtained from the NCBI Sequence read archive (SRA) and converted from SRA format to FASTQ using SRA Toolkit version 2.3.3-4. Illumina reads were aligned to the *C. gattii* VGII R265 reference genome assembly using Burrows-Wheeler Aligner (BWA) v0.7.4-r385 mem (51) with default parameters and were converted to sorted BAM format using SAMtools v0.1.9 (r783) (68).

Genome Analysis Toolkit (GATK) v2.7-4-g6f46d11 (69) was used to call both variant and reference nucleotides from the 101 alignments (as previously described [24]). Briefly, the Picard tools AddOrReplaceReadGroups, MarkDuplicates, CreateSequenceDictionary, and ReorderSam were used to preprocess the alignments (http://broadinstitute.github.io/picard/). GATK RealignerTarget-Creator and IndelRealigner were then used to resolve misaligned reads close to indels. Next, GATK Unified Genotyper (with the haploid Genotyper ploidy setting) was run with both SNP and indel genotype likelihood models (GLM). We also ran Base Recalibrator and PrintReads for base quality score recalibration on those initial sites for GLM SNP. We then recalled variants with Unified Genotyper with the parameter "—output-_mode EMIT_ALL_SITES." We merged and sorted all of the calls and then ran Variant Filtration with the

parameters "QD < 2.0, FS > 60.0, MQ < 40.0." Next, we removed any base that had a minimum genotype quality of below 50, a minimum proportion of alternative alleles (AD) of 80%, or a minimum depth of 10. Finally, we removed any positions that were called by both GLMs (i.e., incompatible indels and SNPs), any marked "LowQual" by GATK, any nested indels, and any sites that did not include a PASS flag.

Phylogenetic and population genetic analysis. The variants identified from the 101 alignments were filtered for positions that were homozygous (reference or SNP) and polymorphic in at least one isolate (Fig. 2), resulting in alignment of 1,517,353 nuclear sites and 970 mitochondrial sites. A FASTA file of these positions was created and converted into PHYLIP format, and a phylogenetic tree was generated using RAxML v7.7.8 (70) with 1,000 bootstrap replications. RAxML was run with the generalized time-reversible (GTR) and category (CAT) rate approximation, with final evaluation of the tree performed using GTR plus gamma-distributed rates. The same sites were analyzed using the NeighborNet Network of SplitsTree v4.14.6 (71).

A multisample variant call format (VCF) corresponding to all 101 genomes was made with VCFtools (72) and converted to ped and map file formats for use in PLINK v1.90 (73). Unsupervised ADMIXTURE (74) was run on a moderately linkage disequilibrium (LD)-pruned alignment for values of K between 1 and 15. A value of $K = 9$ provided the lowest cross-validation error (Fig. 3d and e; see also Fig. S1). To explore finer patterns of population structure among our sampled lineages, we applied a technique designed to characterize patterns of haplotype sharing between panels of "donor" and "recipient" haplotypes within a recombining population. We ran Chromopainter v2 (23) to infer, at each position in a recipient isolate's genome, the donor to which it is most closely ancestrally related relative to all others in the data set. To do this, we assumed a uniform recombination rate of 1.5 M (morgans)/megabase based on the genome-wide recombination rate previously estimated in *C. neoformans* (75) and with Chromopainter's switch and mutation rate parameters estimated using 10 runs of expectation-maximization (−n 190.29; −M 0.0011). We then ran Chromopainter in linked mode using the haploid switch (-j) under an "all-versus-all" framework, painting all samples using all others to produce a pairwise coancestry matrix describing the amount of DNA matches of each isolate to every other under the copying model.

Haplotype-based clustering was then implemented in fineSTRUCTURE (23) with an estimated normalization parameter of $c = 0.51$, sampling cluster assignments every 10,000 iterations for 1×10^6 Markov chain Monte Carlo (MCMC) iterations after 1×10^6 initial burn-in steps. We then performed an additional 1×10^5 hill-climbing iterations beginning with the MCMC sample with the highest posterior probability. This classified our data into 34 clusters (Fig. S2).

For the *C. neoformans* VNI H99-rooted *C. gattii* tree, we identified 1:1 orthologs for each of the nine isolates with Orthofinder v2.1.2 (76) using the Synima pipeline (77). We aligned orthologs with MUSCLE v3.8.31 (78), extracted the coding DNA sequences (CDSs) in a codon context, trimmed to the smallest contiguous sequence, and then concatenated alignments. In total, we aligned 2.16 Mb of transcripts for each genome. Prottest v3.4 (79) was used to determine the best-fitting amino acid transition model (JTT) according to Bayesian information criterion. The final tree was produced using RAxML v7.7.8 (70) and the CAT rate approximation and WAG amino acid replacement matrix with 1,000 bootstrap replicates. Synima (77) was used to visualize synteny between the genomes. The same pipeline was used to compare the previous and updated R265 genomes.

Phenotypic analysis. To determine the growth rate of *C. gattii* VGV, cells of all six VGV isolates were inoculated in YEPD broth and incubated at 30°C on a shaker (200 rpm) for 18 h. Cells were washed with sterile phosphate-buffered saline (PBS), and 2×10^5 cells/ml were resuspended in PBS. Three-microliter aliquots of 10-fold serial dilutions were spotted onto YEPD agar and incubated at 30°C and 37°C. For biological confirmation of the species, isolates were inoculated on canavanine glycine bromothymol blue (CGB) agar (80) for species-specific CGB reactions and on Christensen's urea agar (Sigma) and norepinephrine agar (81) for urease and melanin production reactions, respectively, and were incubated at 30°C for 48 h. India ink staining of the cells grown on YEPD broth for 24 h at 30°C was used for microscopic observation of the cell and polysaccharide capsule size. The reference strains used were WM148 or H99 (serotype A, VNI), WM626 (serotype A, VNII), WM179 (serotype B, VGI), WM178, R265 and R272 (serotype B, VGII), WM161 (serotype B, VGIII), and WM779 (serotype C, VGIV) (14). The mating type of each isolate was determined by PCR using primers specific to *STE12α* and *STE20*a (82).

Determination of MIC for antifungal antibiotics. MICs for fluconazole (FLC), 5-fluorocytosine (5FC), and amphotericin B were determined using Etest strips according to the Etest technical guide (AB Biodisk, Solna, Sweden), with slight modification. Fungal cells were grown in 5 ml of YEPD at 30°C for 18 h. Harvested cells were diluted in sterile saline solution to an optical density at 600 nm (OD$_{600}$) of 0.05 and plated on yeast nitrogen base (YNB) agar plates. Etest strips were placed at the center of the plates and incubated at 30°C for 72 h. The susceptibility endpoint was read at the first growth inhibition ellipse. The concentration ranges tested were as follows: for FLC, 0.016 to 256 μg/ml; for both 5-FC and amphotericin B, 0.002 to 32 μg/ml.

URA5 gene RFLP. The *URA5* gene of each isolate was amplified from genomic DNA by PCR to identify the molecular type using 50 ng of two primers: URA5 (5'-ATGTCCTCCCAAGCCCTCGACTCCG-3') and SJ01 (5'-TTAAG ACCTCTGAACACCGTACTC-3'). Reactions were carried out in a total volume of 50 μl as previously described (14). PCR was performed for 40 cycles at 94°C for 2 min of initial denaturation, 30 s of denaturation at 94°C, 30 s of annealing at 55°C, and 2 min of extension at 72°C. The reactions were completed by performing a final extension step for 10 min at 72°C. PCR products were analyzed by 1% agarose gel electrophoresis, and 5 μl of PCR products was double digested with Sau96I (10 U/μl) and HhaI (20 U/μl) for 3 h at 37°C. Then, digested samples were separated by 3% agarose gel electrophoresis

at 80 V for 5 h. The RFLP patterns of *URA5* gene were analyzed using well-defined bands in the gel images by comparing them with the patterns obtained from the standard reference strains.

Restriction enzyme analysis of the *URA5* gene to distinguish VGV from VGIV. We found that the *URA5* RFLP banding patterns (14) of VGV and VGIV were not clearly distinguishable using Sau96I and HhaI (Fig. S9A). We compared the DNA sequences of the *URA5* gene from MF34 (VGV) and WM779 (VGIV) and found two restriction enzymes, StuI and EarI, that can be used to distinguish the two molecular types based on *URA5* RFLP. Three microliters of *URA5* PCR products was digested with StuI (10 U/μl) or EarI (20 U/μl) (New England BioLabs Inc.) at 37°C for 4 h, and restriction fragments were separated by electrophoresis in 3% agarose Tris-acetate-EDTA (TAE) gels at 80 V for 5 h. Standard reference strains for molecular typing were used as controls.

Virulence in mice. The virulence of four VGV isolates, two from clade A and two from clade B, was assessed using 7-to-8-week-old female C57BL/6 mice (Taconic Farms). Isolates to be tested in mice were inoculated in YEPD broth and incubated overnight, washed twice, and diluted in PBS to 2.5×10^5 cells/ml. Mice (14 per isolate) were inoculated with 20 μl of cell suspension (5×10^3 cells/mouse) by pharyngeal aspiration. Eight mice were used for each isolate for determination of the survival rate, and six mice each were used for the analyses of fungal burden and histopathology at the indicated time points. Mice were monitored twice per day, and differences in survival were determined using GraphPad Prism, version 7 (GraphPad Software, San Diego, CA).

To assess the organ fungal burden, lungs and brains of four mice from each infected group were inspected. The mice infected with WM779 started to die on day 25 postinfection, and the lungs and brains were harvested immediately from the dead mice on day 25. Mice infected with other isolates were euthanized on day 60, and organs were harvested. Harvested lungs and brains were homogenized in 7 ml and 2 ml of sterile water, respectively, and 5-μl aliquots of 10-fold serial dilutions were plated on YEPD agar and incubated at 30°C for 48 h. Fungal colonies were counted and the tissue fungal burden was analyzed using GraphPad Prism, version 7 (GraphPad Software, San Diego, CA).

Histopathological analysis. For histopathological analysis, organs of infected mice from each group were harvested at 10 and 20 days postinoculation and fixed in 3.7% buffered formalin and embedded in paraffin. Sections were stained with hematoxylin and eosin (H&E) or alcian blue/periodic acid-Schiff (AB/PAS) at Histoserv Inc.

Ethics statement. The Institutional Animal Care and Use Committee of the National Institute of Allergy and Infectious Diseases approved all animal studies (approval no. LCIM-5E). Studies were performed in accordance with recommendations of the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

Data availability. The raw sequence and genome assembly of VGV MF34 are available in NCBI under BioProject accession no. PRJNA487802, and the culture has been deposited at the American Type Culture Collection (accession number pending). Raw sequence data were submitted to the NCBI Sequence Read Archive under BioProject identifiers (ID) PRJNA476154 (all *C. gattii* non-VGV isolates) and PRJNA480403 (all *C. gattii* VGV isolates).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .02306-19.

FIG S1, PDF file, 0.5 MB.
FIG S2, PDF file, 0.5 MB.
FIG S3, PDF file, 1 MB.
FIG S4, PDF file, 0.4 MB.
FIG S5, PDF file, 0.01 MB.
FIG S6, PDF file, 0.2 MB.
FIG S7, PDF file, 0.2 MB.
FIG S8, TIF file, 0.6 MB.
FIG S9, PDF file, 2.7 MB.
TABLE S1, XLSX file, 0.1 MB.

REFERENCES

1. Rajasingham R, Smith RM, Park BJ, Jarvis JN, Govender NP, Chiller TM, Denning DW, Loyse A, Boulware DR. 2017. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. Lancet Infect Dis 17:873–881. https://doi.org/10.1016/S1473-3099(17)30243-8.

2. Busse O. 1894. Uber parasitare Zelleinschlusse und ihre Zuchtung. Cent Bakt Parasit 16:175–180.

3. Barnett JA. 2010. A history of research on yeasts 14: medical yeasts part 2, *Cryptococcus neoformans*. Yeast 27:875–904. https://doi.org/10.1002/yea.1786.

4. Kwon-Chung KJ, Fraser JA, Doering TL, Wang Z, Janbon G, Idnurm A, Bahn Y-S. 2014. *Cryptococcus neoformans* and *Cryptococcus gattii*, the etiologic agents of cryptococcosis. Cold Spring Harb Perspect Med 4:a019760. https://doi.org/10.1101/cshperspect.a019760.

5. Evans EE. 1950. The antigenic composition of *Cryptococcus neoformans*: I. A serologic classification by means of the capsular and agglutination reactions. J Immunol 64:423–430.

6. Wilson DE, Bennett JE, Bailey JW. 1968. Serologic grouping of *Cryptococcus neoformans*. Proc Soc Exp Biol Med 127:820–823. https://doi.org/10.3181/00379727-127-32812.

7. Kwon-Chung KJ. 1975. A new genus, filobasidiella, the perfect state of *Cryptococcus neoformans*. Mycologia 67:1197–1200. https://doi.org/10.1080/00275514.1975.12019866.

8. Kwon-Chung KJ. 1976. A new species of Filobasidiella, the sexual state of *Cryptococcus neoformans* B and C serotypes. Mycologia 68:943–946. https://doi.org/10.1080/00275514.1976.12019972.

9. Franzot SP, Salkin IF, Casadevall A. 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. J Clin Microbiol 37:838–840.

10. Xu J, Vilgalys R, Mitchell TG. 2000. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. Mol Ecol 9:1471–1481. https://doi.org/10.1046/j.1365-294x.2000.01021.x.

11. Diaz MR, Boekhout T, Kiesling T, Fell JW. 2005. Comparative analysis of the intergenic spacer regions and population structure of the species complex of the pathogenic yeast *Cryptococcus neoformans*. FEMS Yeast Res 5:1129–1140. https://doi.org/10.1016/j.femsyr.2005.05.005.

12. Kwon-Chung KJ, Boekhout T, Fell JW, Diaz M. 2002. (1557) Proposal to Conserve the Name *Cryptococcus gattii* against *C. hondurianus* and *C. bacillisporus* (Basidiomycota, Hymenomycetes, Tremellomycetidae). Taxon 51:804–806. https://doi.org/10.2307/1555045.

13. Kwon-Chung KJ, Varma A. 2006. Do major species concepts support one, two or more species within *Cryptococcus neoformans*? FEMS Yeast Res 6:574–587. https://doi.org/10.1111/j.1567-1364.2006.00088.x.

14. Meyer W, Castañeda A, Jackson S, Huynh M, Castañeda E, IberoAmerican Cryptococcal Study Group. 2003. Molecular typing of IberoAmerican *Cryptococcus neoformans* isolates. Emerg Infect Dis 9:189–195. https://doi.org/10.3201/eid0902.020246.

15. Boekhout T, Theelen B, Diaz M, Fell JW, Hop WC, Abeln EC, Dromer F, Meyer W. 2001. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. Microbiology 147:891–907. https://doi.org/10.1099/00221287-147-4-891.

16. Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, Esposto MC, Fisher M, Gilgado F, Hagen F, Kaocharoen S, Litvintseva AP, Mitchell TG, Simwami SP, Trilles L, Viviani MA, Kwon-Chung J. 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med Mycol 47:561–570. https://doi.org/10.1080/13693780902953886.

17. Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, Falk R, Parnmen S, Lumbsch HT, Boekhout T. 2015. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet Biol 78:16–48. https://doi.org/10.1016/j.fgb.2015.02.009.

18. Kwon-Chung KJ, Bennett JE, Wickes BL, Meyer W, Cuomo CA, Wollenburg KR, Bicanic TA, Castañeda E, Chang YC, Chen J, Cogliati M, Dromer F, Ellis D, Filler SG, Fisher MC, Harrison TS, Holland SM, Kohno S, Kronstad JW, Lazera M, Levitz SM, Lionakis MS, May RC, Ngamskulrongroj P, Pappas PG, Perfect JR, Rickerts V, Sorrell TC, Walsh TJ, Williamson PR, Xu J, Zelazny AM, Casadevall A. 2017. The case for adopting the "species complex" nomenclature for the etiologic agents of cryptococcosis. mSphere 2:e00357-16. https://doi.org/10.1128/mSphere.00357-16.

19. Vanhove M, Beale MA, Rhodes J, Chanda D, Lakhi S, Kwenda G, Molloy S, Karunaharan N, Stone N, Harrison TS, Bicanic T, Fisher MC. 2017. Genomic epidemiology of Cryptococcus yeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. Mol Ecol 26:1991–2005. https://doi.org/10.1111/mec.13891.

20. Firacative C, Roe CC, Malik R, Ferreira-Paim K, Escandón P, Sykes JE, Castañón-Olivares LR, Contreras-Peres C, Samayoa B, Sorrell TC, Castañeda E, Lockhart SR, Engelthaler DM, Meyer W. 2016. MLST and wholegenome-based population analysis of *Cryptococcus gattii* VGIII links clinical, veterinary and environmental strains, and reveals divergent serotype specific sub-populations and distant ancestors. PLoS Negl Trop Dis 10:e0004861. https://doi.org/10.1371/journal.pntd.0004861.

21. Hagen F, Illnait-Zaragozí M-T, Meis JF, Chew WHM, Curfs-Breuker I, Mouton JW, Hoepelman AIM, Spanjaard L, Verweij PE, Kampinga GA, Kuijper EJ, Boekhout T, Klaassen C. 2012. Extensive genetic diversity within the Dutch clinical *Cryptococcus neoformans* population. J Clin Microbiol 50:1918–1926. https://doi.org/10.1128/JCM.06750-11.

22. 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. https://doi.org/10.1371/journal.pone.0028550.

23. Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet 8:e1002453. https://doi.org/10.1371/journal.pgen.1002453.

24. Chen Y, Farrer RA, Giamberardino C, Sakthikumar S, Jones A, Yang T, Tenor JL, Wagih O, Wyk MV, Govender NP, Mitchell TG, Litvintseva AP, Cuomo CA, Perfect JR. 2017. Microevolution of serial clinical isolates of *Cryptococcus neoformans* var. *grubii* and *C. gattii*. mBio 8:e00166-17. https://doi.org/10.1128/mBio.00166-17.

25. Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. BMC Genomics 17:700. https://doi.org/10.1186/s12864-016-3044-0.

26. Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, Chen Y, Voelz K, Heitman J, May RC, Fisher MC, Cuomo CA. 2015. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio 6:e00868-15. https://doi.org/10.1128/mBio.00868-15.

27. 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 BFF, 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:e00342-10. https://doi.org/10.1128/mBio.00342-10.

28. Janbon G, Ormerod KL, Paulet D, Byrnes EJ, Yadav V, Chatterjee G, Mullapudi N, Hon C-C, Billmyre RB, Brunel F, Bahn Y-S, Chen W, Chen Y, Chow EWL, Coppée J-Y, Floyd-Averette A, Gaillardin C, Gerik KJ, Goldberg J, Gonzalez-Hilarion S, Gujja S, Hamlin JL, Hsueh Y-P, 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, et al. 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. https://doi.org/10.1371/journal.pgen.1004261.

29. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman J, Sanyal K. 2018. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci U S A 115:3108–3113. https://doi.org/10.1073/pnas.1713725115.

30. Litvintseva AP, Thakur R, Vilgalys R, Mitchell TG. 2006. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. Genetics 172:2223–2238. https://doi.org/10.1534/genetics.105.046672.

31. Desjardins CA, Giamberardino C, Sykes SM, Yu C-H, Tenor JL, Chen Y, Yang T, Jones AM, Sun S, Haverkamp MR, Heitman J, Litvintseva AP, Perfect JR, Cuomo CA. 2017. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res 27:1207–1219. https://doi.org/10.1101/gr.218727.116.

32. Wiley EO. 1978. The evolutionary species concept reconsidered. Syst Zool 27:17–26. https://doi.org/10.2307/2412809.

33. Hagen F, Lumbsch HT, Arsenijevic VA, Badali H, Bertout S, Billmyre RB, Bragulat MR, Cabañes FJ, Carbia M, Chakrabarti A, Chaturvedi S, Chaturvedi V, Chen M, Chowdhary A, Colom M-F, Cornely OA, Crous PW, Cuétara MS, Diaz MR, Espinel-Ingroff A, Fakhim H, Falk R, Fang W, Herkert PF, Rodríguez CF, Fraser JA, Gené J, Guarro J, Idnurm A, Illnait-Zaragozi M-T, Khan Z, Khayhan K, Kolecka A, Kurtzman CP, Lagrou K, Liao W, Linares C, Meis JF, Nielsen K, Nyazika TK, Pan W, Pekmezovic M, Polacheck I, Posteraro B, Telles F de Q, Romeo O, Sánchez M, Sampaio A, Sanguinetti M, et al. 2017. Importance of resolving fungal nomenclature: the case of multiple pathogenic species in the *Cryptococcus* genus. mSphere 2:e00238-17. https://doi.org/10.1128/mSphere.00238-17.

34. Scott L. 1990. Hyrax (Procaviidae) and dassie rat (Petromuridae) middens in palaeoenvironmental studies in Africa, p 408–427. *In* Packrat middens: the last 40,000 years of biotic change. University of Arizona Press, Tucson, AZ.

35. Chase BM, Scott L, Meadows ME, Gil-Romera G, Boom A, Carr AS, Reimer PJ, Truc L, Valsecchi V, Quick LJ. 2012. Rock hyrax middens: a palaeoenvironmental archive for southern African drylands. Quat Sci Rev 56:107–125. https://doi.org/10.1016/j.quascirev.2012.08.018.

36. Nielsen K, Obaldia ALD, Heitman J. 2007. *Cryptococcus neoformans* mates on Pigeon guano: implications for the realized ecological niche and globalization. Eukaryot Cell 6:949–959. https://doi.org/10.1128/EC.00097-07.

37. Taylor JW, Barker BM. 2019. The endozoan, small-mammal reservoir hypothesis and the life cycle of Coccidioides species. Med Mycol 57:S16–S20. https://doi.org/10.1093/mmy/myy039.

38. Springer DJ, Billmyre RB, Filler EE, Voelz K, Pursall R, Mieczkowski PA, Larsen RA, Dietrich FS, May RC, Filler SG, Heitman J. 2014. *Cryptococcus gattii* VGIII isolates causing infections in HIV/AIDS patients in Southern California: identification of the local environmental source as arboreal. PLoS Pathog 10:e1004285. https://doi.org/10.1371/journal.ppat.1004285.

39. Litvintseva AP, Mitchell TG. 2009. Most environmental isolates of *Cryptococcus neoformans* var. *grubii* (serotype A) are not lethal for mice. Infect Immun 77:3188–3195. https://doi.org/10.1128/IAI.00296-09.

40. Ngamskulrungroj P, Chang Y, Sionov E, Kwon-Chung KJ. 2012. The primary target organ of *Cryptococcus gattii* is different from that of *Cryptococcus neoformans* in a murine model. mBio 3:e00103-12. https://doi.org/10.1128/mBio.00103-12.

41. Davis MJ, Moyer S, Hoke ES, Sionov E, Mayer-Barber KD, Barber DL, Cai H, Jenkins L, Walter PJ, Chang YC, Kwon-Chung KJ. 2019. Pulmonary iron limitation induced by exogenous type I IFN protects mice from *Cryptococcus gattii* independently of T cells. mBio 10:e00799-19. https://doi.org/10.1128/mBio.00799-19.

42. Dykstra MA, Friedman L, Murphy JW. 1977. Capsule size of *Cryptococcus neoformans*: control and relationship to virulence. Infect Immun 16:129–135.

43. Gomez-Lopez A, Zaragoza O, Dos Anjos Martins M, Melhem MC, Rodriguez-Tudela JL, Cuenca-Estrella M. 2008. *In vitro* susceptibility of *Cryptococcus gattii* clinical isolates. Clin Microbiol Infect 14:727–730. https://doi.org/10.1111/j.1469-0691.2008.02021.x.

44. Khan ZU, Randhawa HS, Kowshik T, Chowdhary A, Chandy R. 2007. Antifungal susceptibility of *Cryptococcus neoformans* and *Cryptococcus gattii* isolates from decayed wood of trunk hollows of *Ficus religiosa* and

Syzygium cumini trees in north-western India. J Antimicrob Chemother 60:312–316. https://doi.org/10.1093/jac/dkm192.

45. Farrer RA, Ford CB, Rhodes J, Delorey T, May RC, Fisher MC, Cloutman-Green E, Balloux F, Cuomo CA. 2018. Transcriptional heterogeneity of Cryptococcus gattii VGII compared with non-VGII lineages underpins key pathogenicity pathways. mSphere 3:e00445-18. https://doi.org/10.1128/mSphere.00445-18.

46. Litvintseva AP, Carbone I, Rossouw J, Thakur R, Govender NP, Mitchell TG. 2011. Evidence that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have evolved in Africa. PLoS One 6:e19688. https://doi.org/10.1371/journal.pone.0019688.

47. Randhawa HS, Kowshik T, Khan ZU. 2005. Efficacy of swabbing versus a conventional technique for isolation of *Cryptococcus neoformans* from decayed wood in tree trunk hollows. Med Mycol 43:67–71. https://doi.org/10.1080/13693780410001712025.

48. Fujimura H, Sakuma Y. 1993. Simplified isolation of chromosomal and plasmid DNA from yeasts. Biotechniques 14:538–540.

49. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

50. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. 1999. Alignment of whole genomes. Nucleic Acids Res 27:2369–2376. https://doi.org/10.1093/nar/27.11.2369.

51. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv 13033997 [q-bio.GN]. https://arxiv.org/abs/1303.3997.

52. Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using only Nanopore sequencing data. Nat Methods 12:733–735. https://doi.org/10.1038/nmeth.3444.

53. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

54. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

55. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 26:1107–1115. https://doi.org/10.1093/nar/26.4.1107.

56. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48. https://doi.org/10.1093/nar/28.1.45.

57. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.

58. Finn RD, Clements J, Eddy SR. 2011. HMMER Web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–W37. https://doi.org/10.1093/nar/gkr367.

59. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.

60. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.

61. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.

62. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373. https://doi.org/10.1093/nar/gkg128.

63. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.

64. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.

65. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting

transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.

66. Farrer RA, Voelz K, Henk DA, Johnston SA, Fisher MC, May RC, Cuomo CA. 2016. Microevolutionary traits and comparative population genomics of the emerging pathogenic fungus *Cryptococcus gattii*. Philos Trans R Soc Lond B Biol Sci 371:20160021. https://doi.org/10.1098/rstb.2016.0021.

67. Engelthaler DM, Hicks ND, Gillece JD, Roe CC, Schupp JM, Driebe EM, Gilgado F, Carriconde F, Trilles L, Firacative C, Ngamskulrungroj P, Castañeda E, Lazera MDS, Melhem MSC, Pérez-Bercoff Å, Huttley G, Sorrell TC, Voelz K, May RC, Fisher MC, Thompson GR, Lockhart SR, Keim P, Meyer W. 2014. *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. mBio 5:e01464-14. https://doi.org/10.1128/mBio.01464-14.

68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

69. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.

70. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. https://doi.org/10.1093/bioinformatics/btl446.

71. Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73. https://doi.org/10.1093/bioinformatics/14.1.68.

72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

73. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795.

74. Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. https://doi.org/10.1101/gr.094052.109.

75. Roth C, Sun S, Billmyre RB, Heitman J, Magwene PM. 2018. A high-resolution map of meiotic recombination in *Cryptococcus deneoformans* demonstrates decreased recombination in unisexual reproduction. Genetics 209:567–578. https://doi.org/10.1534/genetics.118.300996.

76. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2.

77. Farrer RA. 2017. Synima: a synteny imaging tool for annotated genome assemblies. BMC Bioinformatics 18:507. https://doi.org/10.1186/s12859-017-1939-7.

78. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.

79. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165. https://doi.org/10.1093/bioinformatics/btr088.

80. Kwon-Chung KJ, Polacheck I, Bennett JE. 1982. Improved diagnostic medium for separation of *Cryptococcus neoformans* var. *neoformans* (serotypes A and D) and *Cryptococcus neoformans* var. *gattii* (serotypes B and C). J Clin Microbiol 15:535–537.

81. Kwon-Chung KJ, Polacheck I, Popkin TJ. 1982. Melanin-lacking mutants of *Cryptococcus neoformans* and their virulence for mice. J Bacteriol 150:1414–1421.

82. Halliday CL, Bui T, Krockenberger M, Malik R, Ellis DH, Carter DA. 1999. Presence of alpha and a mating types in environmental and clinical collections of *Cryptococcus neoformans* var. *gattii* strains from Australia. J Clin Microbiol 37:2920–2926.