

# A Population Genomics Approach to Assessing the Genetic Basis of Within-Host Microevolution Underlying Recurrent Cryptococcal Meningitis Infection

Johanna Rhodes,\* Mathew A. Beale,<sup>†</sup> Mathieu Vanhove,\* Joseph N. Jarvis,<sup>\*,§,\*\*\*</sup>

Shichina Kannambath,<sup>††</sup> John A. Simpson,<sup>\*\*</sup> Anthea Ryan,<sup>\*\*</sup> Graeme Meintjes,<sup>§§</sup> Thomas S. Harrison,<sup>††</sup>

Matthew C. Fisher,<sup>\*,1</sup> and Tihana Bicanic<sup>††,1</sup>

\*Department of Infectious Disease Epidemiology, Imperial College London, W2 1PG United Kingdom, <sup>†</sup>Division of Infection and Immunity, Faculty of Medical Sciences, University College London, WC1E 6BT United Kingdom,

<sup>‡</sup>Botswana-UPenn Partnership, 0022 Gaborone, Botswana, <sup>§</sup>Division of Infectious Diseases, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104,

<sup>\*\*</sup>Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, WC1E 7HT, United Kingdom, <sup>††</sup>Institute of Infection and Immunity, St. George's University London, SW17 0RE United Kingdom, <sup>\*\*</sup>National Health Laboratory Service, Greenpoint, Cape Town, 8005

South Africa, and <sup>§§</sup>Department of Medicine, University of Cape Town, 7925, South Africa

ORCID IDs: 0000-0002-1338-7860 (J.R.); 0000-0002-4740-3187 (M.A.B.); 0000-0002-4758-9161 (S.K.); 0000-0002-1862-6402 (M.C.F.)

**ABSTRACT** Recurrence of meningitis due to *Cryptococcus neoformans* after treatment causes substantial mortality in HIV/AIDS patients across sub-Saharan Africa. In order to determine whether recurrence occurred due to relapse of the original infecting isolate or reinfection with a different isolate weeks or months after initial treatment, we used whole-genome sequencing (WGS) to assess the genetic basis of infection in 17 HIV-infected individuals with recurrent cryptococcal meningitis (CM). Comparisons revealed a clonal relationship for 15 pairs of isolates recovered before and after recurrence showing relapse of the original infection. The two remaining pairs showed high levels of genetic heterogeneity; in one pair we found this to be a result of infection by mixed genotypes, while the second was a result of nonsense mutations in the gene encoding the DNA mismatch repair proteins *MSH2*, *MSH5*, and *RAD5*. These nonsense mutations led to a hypermutator state, leading to dramatically elevated rates of synonymous and nonsynonymous substitutions. Hypermutator phenotypes owing to nonsense mutations in these genes have not previously been reported in *C. neoformans*, and represent a novel pathway for rapid within-host adaptation and evolution of resistance to first-line antifungal drugs.

## KEYWORDS

cryptococcosis  
evolution  
genomics  
hypermutation  
resistance

The HIV/AIDS pandemic has led to a large population of profoundly immunocompromised individuals that are vulnerable to infection by the opportunistic fungus pathogen *Cryptococcus neoformans* (Hagen *et al.* 2015). This mycosis poses a considerable public health problem in sub-Saharan Africa, which has the highest estimated annual incidence of CM globally (Park *et al.* 2009), with the majority of infections caused by *C. neoformans sensu stricto* (previously referred to as *C. neoformans* var. *grubii*) (Jarvis and Harrison 2007).

Standard treatment for HIV-associated CM includes the long-term use of azole drugs such as fluconazole (FLC), following initial 1–2 wk induction treatment with amphotericin B, which is often not available (Brouwer *et al.* 2004). Microevolution occurs in response to

drug pressure, leading to resistance, a phenomenon previously described in *C. neoformans* (Ormerod *et al.* 2013; Sionov *et al.* 2010). Patients who appear successfully treated [evidenced by symptom resolution and sterilization of cerebral spinal fluid (CSF)] can relapse due to persisting infections, which in some cases appear to have evolved resistance to first-line antifungal drugs. In the absence of continued antifungal therapy and restoration of their immune system through antiretroviral therapy (ART), patients with HIV/AIDS also have a high probability of recurrence of CM (Bozzette *et al.* 1991).

Various methods of within-host evolution are available to eukaryotic pathogens, most notably sexual and parasexual reproduction, although

these are difficult to observe due to the often cryptic nature of recombination of fungi. Aneuploidy, recombination in the telomeres, and mutator states (Rodero *et al.* 2003) also provide means of rapid within-host evolution, with other mechanisms still likely to be discovered. The accumulation of single nucleotide polymorphisms (SNPs) alongside copy number variation (CNV) and aneuploidy has been witnessed during infection in different fungal pathogen species, enabling rapid adaptive evolution (Calo *et al.* 2013) and conferring resistance to antifungal drugs (Hickman *et al.* 2015). Candidiasis is caused by numerous *Candida* species, yet within-host evolution among these species differs. Infectious strains of *Candida albicans* are usually susceptible to azole antifungal drugs, but resistance can evolve via the evolution of drug-resistant aneuploid isolates, which contain an isochromosome of the left arm of chromosome 5 (Selmecki *et al.* 2009). The left arm of chromosome 5 contains two important genes involved in resistance to antifungals: *ERG11*, a target of azoles, and *TAC1*, a transcription factor that activates drug efflux pump expression. Conversely, *Ca. glabrata* strains are intrinsically poorly susceptible to azoles, and have more recently evolved multi-drug resistance to both azoles and echinocandins (Pfaller 2012; Panackal *et al.* 2006; Alexander *et al.* 2013).

The occurrence of within-host diversity and recombination has been witnessed in eukaryotic pathogens, notably *Ca. albicans*: mutation and recombination rates can be increased under stressful conditions, such as drug treatment (Forche *et al.* 2009; Ford *et al.* 2015), resulting in loss-of-heterozygosity (LOH) and aneuploidy (Forche *et al.* 2009). These genetic alterations contribute to the maintenance of a population of *Ca. albicans* within the host environment (Forche *et al.* 2009), and drug pressure can result in diverging levels of fitness (Cowen *et al.* 2001).

Similar responses to antifungal drugs have been observed in *C. neoformans*; point mutations in the ortholog *ERG11* were also shown to confer FLC resistance, by causing the amino acid substitution G484S (Rodero *et al.* 2003). Sionov *et al.* (2010) demonstrated that large-scale chromosomal duplications (primarily chromosome 1) are fundamental to overcoming FLC drug pressure in a mouse model, contributing to failure of FLC therapy. The duplication of chromosome 1 included increased copy number of genes *ERG11*, the target of FLC, and *AFR1*, a transporter of azoles (Sionov *et al.* 2010), although other genes are also thought to be involved in FLC resistance (Sionov *et al.* 2013; Paul *et al.* 2015). Previous studies of serially collected *C. neoformans* isolates have confirmed in-host microevolution, including the occurrence of large-scale genomic rearrangements (Fraser *et al.* 2005; Blasi *et al.* 2001; Illnait-Zaragozi *et al.* 2010). Like *Ca. albicans*, the *C. neoformans* genome is capable of undergoing chromosomal duplication and loss under stresses such as drug pressure or invasion of the human

host (Fries *et al.* 1996). These chromosomal duplications are often lost when the selective pressure is removed (Sionov *et al.* 2010).

The development of mutator states via hypermutability is a rapidly expanding area of study in bacteria, particularly *Pseudomonas aeruginosa* in cystic fibrosis patients. Here, hypermutability has been shown to have an association with antimicrobial resistance (Oliver *et al.* 2000; Maciá *et al.* 2005), causing significant implications in the early treatment of cystic fibrosis patients to prevent chronic infection (Burns *et al.* 2001; Maciá *et al.* 2005). Few studies have explored hypermutation in pathogenic fungi; however, mutations in the yeast *Saccharomyces cerevisiae* genes *PMS1*, *MLH1*, and *MSH2*, which are all involved in mismatch repair, have been shown to lead to 100–700-fold increases in mutations throughout the genome (Strand *et al.* 1993). Frameshift mutations in an ortholog of the mismatch repair gene *MSH2* have also been shown to contribute to microevolution in the sister species of *C. neoformans*, *C. gattii* (Billmyre *et al.* 2014).

Here, we describe a comparative genome sequencing-based approach to investigate microevolution in serially collected isolates of *C. neoformans*. These isolates were grown and stored from fresh CSF of patients with CM, prior to starting and during antifungal therapy using induction with amphotericin B-based regimens, followed by FLC. We used WGS to describe the nature of infection in 17 patients to gain insights into the dynamics of recurrent infections.

## MATERIALS AND METHODS

### Samples and patients

Sixteen South African patients and one Ugandan patient demonstrating clinical evidence of CM were studied. All patients were either part of observational studies or clinical trials (Bicanic *et al.* 2007, 2008; Jarvis *et al.* 2010, 2012; Longley *et al.* 2008). Ethical approval was obtained from the Wandsworth Research Ethics Committee covering St. George's University of London (Longley *et al.* 2008; Bicanic *et al.* 2007, 2008; Jarvis *et al.* 2012). In South Africa, additional ethical approval was obtained from the University of Cape Town Research Ethics Committee; in Uganda, from the Research Ethics Committee of Mbarara University of Science and Technology. All patients initially presented with CM and were treated using induction therapy with 7–14 d amphotericin B deoxycholate 0.7–1 mg/kg/d, with or without 100 mg/kg/d of flucytosine (with one patient, IFNR63, also receiving adjunctive interferon  $\gamma$ ), followed by FLC consolidation at 400 mg/d for 8 wk and maintenance therapy at 200 mg/d for 6–12 months ( $n = 16$  pairs), until immune restoration on ART with a CD4 count of  $> 200$  cells/ $\mu$ l. The single patient in Uganda received induction therapy with FLC 1600 mg/d for 2 wk followed by FLC consolidation and maintenance and ART, as above ( $n = 1$  pair). As part of the study procedure, patients enrolled in clinical trials had quantitative cryptococcal cultures performed on serial CSF samples. Patients with a recurrence of their cryptococcal disease following initial treatment and positive CSF culture for *Cryptococcus* at the time of disease recurrence were included in the study. We studied the clinical cryptococcal isolates taken on initial diagnosis (prior to initiation of treatment) and compared each with the *Cryptococcus* isolated from CSF on recurrence of disease in the same patient (Table 1).

### Multi-locus sequence typing

To discern whether mixed or single genotype infections were extracted from CSF, multi-locus sequence typing (MLST) was performed on three independent colonies for a subset of the study isolates, according to the methods outlined in Meyer *et al.* (2009), with modifications as outlined in Beale *et al.* (2015).

Copyright © 2017 Rhodes *et al.*

doi: <https://doi.org/10.1534/g3.116.037499>

Manuscript received November 15, 2016; accepted for publication February 3, 2017; published Early Online February 10, 2017.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at [www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.037499/-/DC1](http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.037499/-/DC1).

<sup>1</sup>Corresponding authors: Department of Infectious Disease Epidemiology, Imperial College London, VC3, Norfolk Place, London W2 1PG, United Kingdom. E-mail: [matthew.fisher@imperial.ac.uk](mailto:matthew.fisher@imperial.ac.uk); and Institute of Infection and Immunity, St. George's University London, Cranmer Terrace, London SW17 0RE, United Kingdom. E-mail: [tbicanic@sgul.ac.uk](mailto:tbicanic@sgul.ac.uk)

■ **Table 1** Details of *C. neoformans* isolates and MICs (if available) at time of isolation from South Africa and Uganda (pair 7 only) used in this study

Pair #	Isolate ID	Day of Isolation	FLC MIC at Isolation	Treatment of CM Episode (if Known)
1	CCTP27	1	4	AmB 3 d
1	CCTP27-d121	121	64	VOR until CD4 > 200 cells/ $\mu$ l
2	CCTP32	1	4	AmB 7 d
2	CCTP32-d132	132	6	AmB 7 d
3	CCTP50	1	16	AmB 14 d
3	CCTP50-d257	257	256	AmB 14 d
3	CCTP50-d409	409	N/A	AmB 7 d
4	CCTP52	1	N/A	FLC 400 mg/d
4	CCTP52-d55	55	N/A	FLC 400 mg/d
5	RCT9	1	2	AmB 0.7 mg/kg/d plus 5FC for 14 d
5	RCT9-d99	99	N/A	AmB until death
6	RCT24	1	4	AmB plus 5FC for 14 d
6	RCT24-d154	154	12	FLC 800 mg/d
7	1600-1	1	N/A	FLC 16,000 mg/d for 14 d
7	1600-1-d106	106	N/A	
8	IFNR63	1	N/A	AmB plus 5FC
8	IFNR63-d128	128	N/A	
9	IFNR24	1	N/A	AmB 7 d
9	IFNR24-d101	101	N/A	
10	IFNR18	1	N/A	AmB 7 d
10	IFNR18-d134	134	N/A	
11	IFNR14	1	N/A	AmB 7 d
11	IFNR14-d97	97	N/A	
12	IFNR13	1	N/A	AmB 7 d
12	IFNR13-d95	95	1	
13	IFNR6	1	N/A	AmB 7 d
13	IFNR6-d73	73	1	
14	IFNR19	1	N/A	AmB 7 d
14	IFNR19-d111	111	N/A	
15	IFNR11	1	N/A	AmB 7 d
15	IFNR11-d203	203	12	
16	IFNR27	1	N/A	AmB 7 d
16	IFNR27-d204	204	N/A	
17	IFNR23	1	N/A	AmB 7 d
17	IFNR23-d179	179	N/A	

ID, identifier; FLC, fluconazole; MIC, minimum inhibitory concentration; CM, cryptococcal meningitis; AmB, amphotericin B 1 mg/kg/d (as per hospital guidelines at that time, unless otherwise stated); VOR, voriconazole 300 mg/d; N/A, not applicable; 5FC, flucytosine.

## Molecular methods

*C. neoformans* was isolated from HIV-infected individuals on location by plating CSF onto Sabourand Dextrose (SD) agar (Oxoid, Fisher Scientific), and growing at 30° for 48 hr. A representative sample of the *C. neoformans* population was taken by selecting a broad “sweep” of all colonies on the SD agar plate, which was stored in cryopreservative medium (80% SD broth, 20% glycerol) at –80° until further testing. This approach ensures that all genetic diversity is maintained through the process, and single colony picking only occurs at the final stage of liquid culture and DNA extraction.

Frozen stocks were plated onto SD agar and cultured for 72 hr. A single colony was inoculated into 6 ml Yeast Peptone Digest broth (Oxoid) supplemented with 0.5 M NaCl and cultured at 37° with agitation (165 rpm) for 40 hr, followed by genomic DNA extraction using the Masterpure Yeast DNA purification kit (Epicentre) modified by addition of two cycles of rapid bead beating (45 sec at 4.5 m/sec) using a FastPrep 24 homogenizer (MP Bio). Genomic DNA libraries were prepared using the TruSeq DNA v2 or TruSeq Nano DNA kit (Illumina), and WGS was performed on an Illumina HiSequation 2500 at the Medical Research Council Clinical Genomics Centre (Imperial College London) as previously described (Rhodes *et al.* 2014).

## WGS analysis

Raw Illumina reads were aligned to the *C. neoformans* reference genome H99 (Loftus *et al.* 2005) using the Burrows–Wheeler Aligner (BWA) v0.75a mem algorithm (Li 2013) with default parameters to obtain high depth alignments (average 104 ×). Samtools (Li *et al.* 2009) version 1.2 was used to sort and index resulting BAM files, and generate statistics regarding the quality of alignment. Picard version 1.72 was used to identify duplicate reads and assign correct read groups to BAM files. Furthermore, BAM files were locally realigned around insertions and deletions (INDELs) using GATK (McKenna *et al.* 2010) version 3.4-46 “RealignerTargetCreator” and “IndelRealigner,” following best practice guidelines (Van der Auwera *et al.* 2013).

SNPs and INDELs were called from all alignments using GATK (McKenna *et al.* 2010) version 3.4-46 “HaplotypeCaller” in haploid mode, with a requirement that all variants called and emitted were above a phred-scale confidence threshold of 30. Both SNPs and INDELs were hard filtered due to a lack of training sets available for *C. neoformans* by running VariantFiltration with parameters “DP < 5 || MQ < 40.0 || QD < 2.0 || FS > 60.0”; this expression ensured that low-confidence variants were filtered out if they met just one of the filter expression criteria. Resulting high-confidence variants were mapped to

genes using VCF-annotator (Broad Institute, Cambridge, MA) and the latest release (CNA3) of the *C. neoformans* reference genome H99 and gene ontology (GO).

Some isolates were suspected of having nonhaploid genomes due to the high number of low-confidence variants. For these isolates, “HaplotypeCaller” was repeated in diploid mode.

The average (mean) coverage for each isolate were determined using GATK (McKenna *et al.* 2010) version 3.4-46 “DepthOfCoverage” under default settings. The *C. neoformans* H99 (Loftus *et al.* 2005) was again used as reference. In order to determine aneuploidy, whole-genome coverage data were normalized and regions displaying normalized coverage equal to two were deemed diploid events (likewise, normalized coverage equal to three was deemed a triploid event, and so on), whereas normalized coverage equal to zero was deemed a deletion event.

### Susceptibility testing

The susceptibility testing of all relapse isolates was performed with the MICRONAUT-AM susceptibility testing system for yeast (Merlin) as recommended by the manufacturer. MICRONAUT-AM allows the determination of MICs of amphotericin B, flucytosine, FLC, voriconazole, posaconazole, itraconazole, micafungin, anidulafungin, and caspofungin, and commercializes the well-established, but laborious, CLSI broth microdilution technique. Briefly, for each isolate, five colonies were used to prepare a 0.5 McFarland-standard suspension in 0.9% NaCl. A 1:20 dilution was prepared in 0.9% NaCl and a 1:5 dilution was prepared in 11 ml RPMI broth provided with the kit. Next, 100  $\mu$ l AST indicator and 50  $\mu$ l methylene blue solutions were mixed with the broth for manual susceptibility testing. The broth was then inoculated onto Merlin MICRONAUT 96-well testing plates (100  $\mu$ l/well) and incubated at 30° for 72 hr. The lowest concentration of an antifungal agent with no detectable growth (MIC) was determined for each isolate based on fungal growth (pink) or no growth (blue). Obtained MICs were interpreted according to *Ca. albicans* EUCAST (Vers. 7.0/12-08-2014) values (Rodríguez-Tudela *et al.* 2010; Alastruey-Izquierdo and Cuenca-Estrella 2012).

### Phylogenetic analysis

Whole-genome SNPs were converted into relaxed interleaved Phylip format. Rapid bootstrap phylogenetic analysis using 500 bootstrap replicates was carried out on 62 isolates in total (Table 1) using RAxML-HPC version 7.3.0 (Stamatakis 2006) as described in Abdolrasouli *et al.* (2015); 35 isolates from this study in addition to 27 isolates (“nonstudy”) were included to show the phylogenetic context of true relapse infections. These nonstudy isolates, while from a clinical source, were not recurrent isolates and were not isolated as part of the clinical trials described in the earlier *Materials and Methods* section. Resulting phylogenies were visualized in FigTree version 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). The same process was completed for each chromosome individually for all 62 isolates, using 250 replicates in the rapid bootstrap analysis.

### GO and KEGG pathway analysis

Nonsynonymous SNP (nsSNP) mutations unique to each timepoint for each pair were assessed for significantly overrepresented GO annotations and metabolic pathways. Briefly, genes found to contain a nsSNP mutation were interrogated for overrepresented Biological Process Ontology in the *C. neoformans* H99 database. GO terms that were found to be associated with genes mapping to the InterPro domain database were transferred to GO associations, using a *P*-value cut-off of

$P < 0.05$ . For metabolic pathway enrichment in genes containing nsSNPs, genes were interrogated against the KEGG (Kanehisa *et al.* 2016) pathway source for *C. neoformans* H99, using a *P*-value cut-off of  $P < 0.05$ .

### Identifying sites under selection

BayeScan 2.01 (Foll and Gaggiotti 2008) uses an outlier approach to identify candidate loci under natural selection. The method uses the allele frequencies that are characteristic of each population and estimates the posterior probabilities of a given locus under a model that includes selection and a neutral model. The program then determines whether the model that includes selection better fits the data. This approach allows the simultaneous assessment of the influence of both balancing and purifying selection. Loci under balancing selection will present low  $F_{ST}$  values whereas high  $F_{ST}$  values reflect patterns of local adaptation (purifying selection) (Excoffier *et al.* 2009). Analysis was not undertaken for the VNII and VNB lineages due to low numbers of isolates, which would be insufficient to overcome the strong population structure. VNI isolates at day 0 were assigned to a population and their associated relapse isolates constituted the second population. Analyses were conducted using the standard parameters including a 50,000 burn in period and 100,000 iterations. Several analyses were conducted varying the prior odds (from 10 or 100 to 1000) for the neutral model.

### Data availability

All raw reads and information on lineages of isolates in this study have been submitted to the European Nucleotide Archive under the project accession PRJEB11842.

## RESULTS

### Clinical and demographic information

The study included paired isolates from 17 patients, with a median age of 32 yr (IQR 26–36) and median CD4 count at CM diagnosis of 22 (IQR 9–71) cells/ $\mu$ l. Six patients were male, nine were female, with the gender of two patients unrecorded. The median time between initial and recurrence isolates was 115 d (minimum 55 d and maximum 409 d). In those for whom ART status was known, 2 of 16 (13%) patients were already on ART at the initial CM episode; 6 out of 15 (40%) patients had not started ART prior to CM recurrence.

Detailed clinical notes were available for the recurrent CM episode in seven patients: two (CCTP52 and RCT9) had not attended follow up and never started ART prior to admission with recurrence; both died of the recurrent CM. One patient (CCTP32) had not been taking FLC for 2 wk prior to recurrence. Four patients (CCTP27, CCTP50, RCT24, and IFNR63) who were adherent to both ART and FLC at recurrence were assessed as having CM immune reconstitution inflammatory syndrome.

### Sequencing of paired samples isolated from patients infected with *C. neoformans*

Prior to sequencing, multiple colonies from a subset of isolates included in this study were analyzed using MLST to investigate whether a mixed infection was present in the original CSF extract. The results show that mixed infections were not present in 12 out of 17 pairs included in this study. One pair (pair 7) was only tested once, and allele types were not sufficient to conclude whether sequence type (ST) 100 or 196 was present in both original and recurrent isolate. On two separate attempts, STs for pairs 3 and 17 could not be determined, reflecting a need for WGS to characterize these pairs. STs for pairs 1 and 6 were inconclusive, and suggestive of a mixed infection being present.



We recovered an average of 23.9 million reads from each isolate, with an average of 98.8% of reads mapped to the *C. neoformans* H99 reference genome (Loftus *et al.* 2005), and an average coverage of  $104 \pm 31.2$  SD. To enable comparative studies and detect microevolutionary changes, precise variant-calling was needed; variants were identified and false positive low-confidence variants were filtered out to provide a set of high-confidence SNPs (see *Materials and Methods*). Full alignment, coverage, and variant calling statistics are provided in Supplemental Material, Table S5 in File S1.

Due to a high number of low-confidence SNPs filtered out in some isolates, which is suggestive of heterozygous SNPs, variant calling was rerun in diploid mode (see *Materials and Methods*) for all isolates in pairs 3, 4, 5, and 17 (results in Table S6 in File S1).

A high level of diversity was observed within the VNB lineage, resulting in long branch lengths among isolates within this clade (Figure 1). Although all VNB isolates were mapped to the *C. neoformans* H99 reference (Loftus *et al.* 2005), which is a VNI lineage isolate, we do not believe that SNP numbers observed in the VNB lineage are inflated by the large phylogenetic distance to the reference genome. This is because SNP determination revealed only 21.6% of SNPs that we discovered were shared by the three VNB pairs included in this study (pairs 3, 12, and 17), highlighting the large amounts of genetic diversity seen within this lineage as we have previously noted (58).

Phylogenetic analysis showed that, of the 17 pairs of relapse isolates, three pairs were lineage VNB, while four and ten belonged to lineages VNII and VNI, respectively (Figure 1). The average pairwise SNP diversity was far higher among isolates from the VNB lineage (140,835 SNPs) compared to isolates in the VNI (17,808 SNPs) and VNII (938 SNPs) lineages, showing that the VNI and VNII lineages are less diverse than VNB across our cohort. On average, isolates of the VNB, VNII, and VNI lineages accumulated 365, 12, and 3 unique SNPs per day between the time of the original isolation and the recurrence of infection. Isolates in the VNB lineage were more likely to experience a ploidy event, with an average of 1.6 changes in ploidy per isolate. Less than one isolate in the VNII and VNI lineages would, on average, experience ploidy events (0.375 and 0.26, respectively).

All pairs, with the exception of pair 7, were isolated from patients in South Africa; pair 7 was isolated from a patient in Uganda. We classified the second isolate as a relapse of the original infection if  $> 97\%$  of SNPs were in common between original and recurrent isolates. The majority of pairs had  $> 99\%$  SNP similarity (Table 2) between original and recurrent isolates, with Pairs 6, 7, and 14 displaying 97, 98, and 97% similarity, respectively. Therefore, all pairs, with the exception of pairs 3 and 17 (SNP similarity 44 and 56%), could be classified as relapsed infections on this basis. This confirms previous results obtained by MLST, and that the original and recurrent isolates sequenced from pairs 1 and 7 (which had previously indicated a potentially mixed infection) were indeed true relapse infections.

Within the VNB pairs (3, 12, and 17), the accumulation of SNPs between original and recurrent infection varied widely. We observed 178 and 304 SNPs/d for CCTP50-d257 and CCTP50-d409, respectively (pair 3), 8 SNPs/d for pair 12, and 968 SNPs/d for pair 17. Due to the variation in SNP accumulation between pair 12 and pairs 3 and 17, we hypothesized that pair 12 was a true relapse of the original infection, while pairs 3 and 17 were showing inflated SNP numbers due to reinfection or an anomalous rate of evolution.

### Antifungal susceptibility testing

FLC susceptibility testing (Table 1) using the Etest (bioMérieux) was carried out for 12 isolates (including three paired isolates) in this study

by the accredited central Microbiology laboratory in Cape Town at the time of the clinical episode; five of these (CCTP27-d121 in pair 1; CCTP50 and CCTP50-d257 in pair 3; RCT24-d154 in pair 6; and IFNR11-d203 in pair 15) had MICs above the established epidemiological cut-off value ( $\geq 8$   $\mu\text{g/ml}$ ) for FLC. All pairs were retested following 5–10 yr frozen storage in glycerol using the MICRONAUT-AM system for yeast susceptibility (*Materials and Methods*): all were found to be sensitive to FLC.

The fourfold increase in FLC MIC observed in pairs 1 and 3 initial and recurrent infections provide a sound basis for relapse of infection due to drug resistance: in pair 1 (patient CCTP27), the initial isolate had a susceptible FLC MIC of 4  $\mu\text{g/ml}$ , while the recurrent isolate was resistant at a MIC of 64  $\mu\text{g/ml}$ ; in pair 3 (CCTP50), the initial isolate MIC was 16  $\mu\text{g/ml}$  (intermediate), while a highly resistant MIC of 256  $\mu\text{g/ml}$  was found on recurrence at day 257.

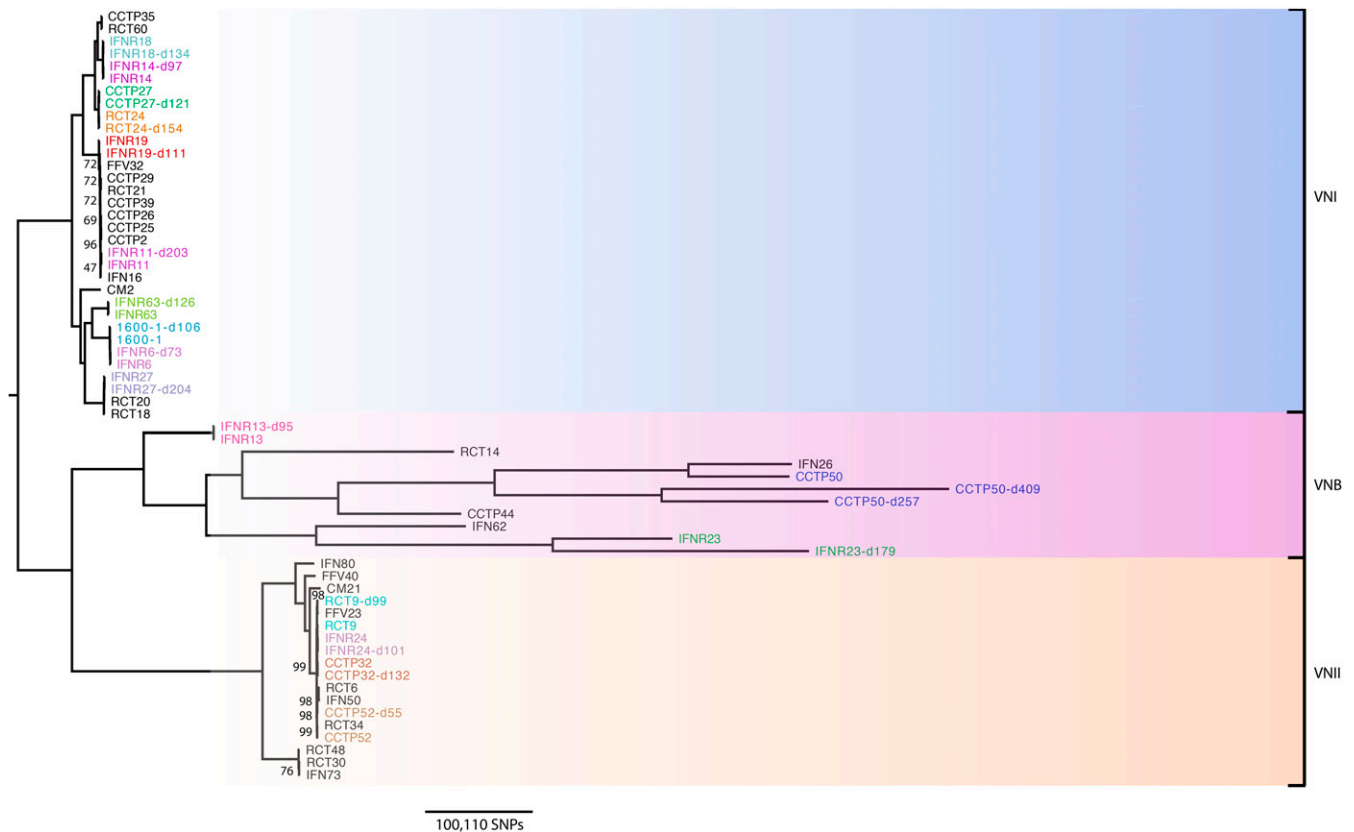
### Serial isolates share a recent common ancestor, suggesting relapse of infection

To investigate whether the *C. neoformans* isolates from the same patients were relapse infections of the original isolates or infections with a new isolate, we undertook phylogenetic analyses to determine their relationships.

As described above, the high level of common SNPs and subsequent low level of unique SNPs between recurrent isolates indicated that all pairs, with the exception of pairs 3 and 17, were relapses of the original infections (Table 2). Phylogenetic analysis (Figure 1) confirmed that all pairs (excepting pairs 3 and 17) clustered together with short branch lengths, confirming the low level of divergence between original and recurrent isolates, thus confirming that they were relapses of the original infections. However, only 46 and 56% of SNPs were found to be in common between initial and relapse infections in pairs 3 and 17, respectively (Figure 1 and Table 2). These VNB pairs (pair 3: CCTP50, CCTP50-d257, and CCTP50-d409 and pair 17: IFNR23 and IFNR23-d179) showed markedly longer branch lengths, suggesting either reinfection or elevated rates of within-host evolution. Further analysis was undertaken to confirm or refute that reinfection by a different isolate was responsible for pairs 3 and 17. Phylogenetic analysis for all isolates included in Figure 1 was repeated for each of the 14 *C. neoformans* chromosomes individually (Figure S1).

Phylogenetic analysis of pair 3 showed that the original infecting genotype of CCTP50 was highly related to the isolate IFN26 (not included in this study but included in the phylogeny to assist with defining lineages, see *Materials and Methods*). All three genotypes from pair 3 were found to be phylogenetically clustered together, but with long branches (Figure 1). Chromosome-by-chromosome analysis indicated that pair 3 serially isolated genotypes displayed differing relationships for each chromosome, and all three serial genotypes were clustered together only in the phylogeny for chromosome 1 (Figure S1). All three genotypes were phylogenetically similar for three other chromosomes, however long branches and clustering with additional nonstudy isolates suggested differing evolutionary relationships. The three serially isolated genotypes of pair 3 were completely phylogenetically dissimilar in three chromosomes; the remaining chromosomes saw either the day 1 isolate (CCTP50) and day 409 isolate (CCTP50-d409), or the day 257 isolate (CCTP50-d257) and day 409 isolate (CCTP50-d409) phylogenetically more related.

Pair 17 isolates (ID IFNR23) clustered together in only two of the 14 chromosomal phylogenies explored (chromosomes 10 and 12); in the remaining chromosomal phylogenies, the pair 17 isolates either displayed a close phylogenetic relationship but with long branches (six



**Figure 1** Phylogenetic analysis of *C. neoformans* var. *grubii* isolates in this study (colored), with additional isolates (shown in black) added to distinguish true relapse infections, or recurrent infections, and associated lineages. We hypothesize that isolates resulting from true relapse infections would be closely related phylogenetically. Bootstrap analysis over 500 replicates was performed on WGS SNP data from 62 isolates, including the 35 isolates included in this study, to generate an unrooted maximum-likelihood phylogeny, with all branches supported to 69% or higher (with the exception of a particularly clonal VNI clade, including pair 15 only, which only had 47% branch support). Branch lengths represent the number of SNPs between taxa. SNP, single nucleotide polymorphism; WGS, whole-genome sequencing.

chromosomes), or were phylogenetically distinct from one another and were more phylogenetically related with other study or additional isolates (six chromosomes).

### Microevolution within the human host

Our data present a unique opportunity to observe microevolution of all three lineages of *C. neoformans* in the human host. Although multiple factors determine evolutionary rates, identifying nsSNPs that cause amino acid change is a standard method for inferring genetic diversity and observing natural selection on codons.

Less than 3% of nsSNPs were unique to recurrent isolates in all pairs, further suggesting that all pairs are relapse of the original infection, with the exception of the VNB pairs 3 and 17 (15.2 and 59.8% of all nsSNPs are unique, respectively).

SNPs unique to each timepoint for each pair were identified. All SNPs at day 1 in all pairs, and all SNPs at time of isolate of recurrent infection in all pairs, were compared. No SNPs were found to be common to all 17 pairs at either day 1 or at point of recurrent infection; however, there were VNII and VNB lineage-specific and timepoint-specific, common SNPs.

Five SNPs, all intergenic, were found to be common at day 1, along with five different SNPs, also intergenic, within VNB pairs (pairs 3, 12, and 17). Three intergenic SNPs were common to all VNII pairs (pairs 2, 4, 5, and 9) at timepoint day 1, while 14 SNPs were common to all VNII pairs at the point of recurrent infection, five of which were intergenic.

The remaining nine SNPs were located in the 5'-untranslated region (UTR) gene *SMF1* (CNAG\_05640), a metal ion transporter with a natural resistance-associated macrophage protein. However, selection analysis indicated that this gene was not under selection pressure.

To evaluate the genetic divergence, Wright's fixation indexes ( $F_{ST}$ ) were calculated to identify SNPs under selection in VNI original and recurrent infection populations investigating 96,856 loci (see *Materials and Methods*). No putative loci under either diversifying or balancing selection could be detected using a false discovery rate of 0.05.  $F_{ST}$  values were limited to not exceed  $3.47 \times 10^{-5}$ .

### Aneuploidy as a generator of diversity in recurrent infection

Normalized whole-genome coverage was plotted to observe possible aneuploidy (increase or decrease in copies of chromosomes Figures 2 and S2) and CNV events. Aneuploidy events were observed in seven genome pairs, suggesting either interspersed or tandem duplications of large segments of the genome.

Ormerod *et al.* (2013) previously published a study showing relapse isolates exhibiting aneuploidies of chromosome 12. We observed aneuploidy of chromosome 12 in four pairs (pairs 1, 5, 10, and 14) included in this study. The aneuploidy spanned different regions of chromosome 12 in all pairs, but all aneuploidies were present in the right arm of the chromosome: in pair 5 the aneuploidy was restricted to 392 kbp of one chromosome arm; in pair 1, the aneuploidy spanned an

■ **Table 2 A high number of shared SNPs in most pairs indicate a shared common ancestor**

Pair	Common SNPs (% Total)	Day 1 SNPs (% Total)	No. Day 1 nsSNPs (Genes Mapped)	Relapse SNPs (% Total)	No. Relapse nsSNPs (Genes Mapped)	Relapse #2 SNPs (% Total)	No. Relapse #2 nsSNPs (Genes Mapped)
1	13,490 (98.49)	97 (0.71)	9 (8)	110 (0.81)	19 (8)		
2	289,557 (99.29)	1,080 (0.37)	110 (55)	991 (0.34)	108 (45)		
3	261,718 (46.77)	127,631 (32.78)	25,331 (2,741)	45,769 (14.88)	8,342 (2,099)	124,498 (32.24)	23,957 (2,541)
4	289,574 (99.22)	1,169 (0.40)	145 (58)	1,096 (0.38)	145 (56)		
5	289,367 (99.20)	1,261 (0.43)	127 (54)	1,080 (0.37)	127 (59)		
6	13,122 (95.78)	133 (1.00)	5 (2)	445 (3.28)	82 (56)		
7	47,833 (97.83)	522 (1.08)	81 (65)	537 (1.11)	94 (72)		
8	46,883 (98.76)	294 (0.62)	29 (9)	296 (0.63)	26 (16)		
9	289,458 (99.27)	1,033 (0.36)	141 (58)	1,109 (0.38)	105 (55)		
10	12,687 (98.29)	104 (0.81)	4 (3)	117 (0.91)	11 (5)		
11	12,732 (98.38)	110 (0.86)	10 (2)	99 (0.77)	6 (5)		
12	221,973 (99.37)	636 (0.29)	55 (27)	778 (0.35)	61 (30)		
13	48,062 (98.58)	396 (0.82)	31 (11)	294 (0.61)	28 (15)		
14	28,960 (95.66)	288 (0.98)	44 (19)	1,026 (3.42)	105 (57)		
15	29,736 (98.40)	254 (0.85)	21 (14)	228 (0.76)	17 (13)		
16	45,437 (98.56)	325 (0.71)	30 (18)	341 (0.74)	30 (14)		
17	376,568 (56.43)	117,436 (23.77)	22,567 (4,153)	173,359 (31.52)	33,543 (4,427)		

Number of SNPs common to both initial and recurrent infection, along with number of SNPs and nonsynonymous SNPs unique to each timepoint. Percentages given to 2 d.p. SNP, single nucleotide polymorphism; No., number; nsSNP, nonsynonymous SNP.

entire arm of chromosome 12 (603 kbp). Pair 14 displayed this aneuploidy in both the initial and recurrent infection, and the aneuploidy spanned the whole chromosome; the recurrent infection isolate of pair 10 also displayed aneuploidy along the whole chromosome. Evaluation of the read depth along chromosome 12 revealed triplication of the chromosome 12 arm in the Pair 1 recurrent infection isolate, a phenomenon also seen in Ormerod *et al.* (2013). Further analysis of read depth in the pair 5 recurrent isolate revealed a diploid genome, and chromosome 12 was also experiencing triploidy. Current annotation of the *C. neoformans* H99 genome reveals the presence of 327 genes in chromosome 12; the right arm of chromosome 12 has 260 genes present. We scanned the genes present in this arm of chromosome 12 for genes potentially involved in virulence, which might prove advantageous to the progression of infection or drug resistance. One such gene was *SFB2* (CNAG\_06093), which is involved in the conservation of the sterol regulatory element-binding protein pathway (Chang *et al.* 2009). An alcohol dehydrogenase (*GNO1* – CNAG\_06168) was also present, which is thought to be involved in the defense against host response (de Jesús-Berrios *et al.* 2003). Analysis for enrichment of metabolic pathways also revealed that the genes present in this chromosome arm are significantly involved in the metabolism of drugs (corrected  $P$ -value  $P < 3.81e^{-2}$ ).

We searched for CNV in genes known to be involved in drug resistance and virulence. *CAP10* appeared to be haploid in all isolates, with the exception of pairs 4 and 5, where the initial infection (CCTP52) and recurrent infection (RCT9-d99) were found to be diploid, respectively. However, on closer inspection (see *Materials and Methods*), we believe that the isolates in pairs 4 and 5 (CCTP52 and RCT9-d99) have diploid genomes, implying that the *CAP10* gene is actually tetraploid. Whether the remaining isolates in these two pairs (CCTP52-d55 and RCT9) have diploid genomes could not be distinguished; however, it is clear that *CAP10* loses ploidy from initial infection to relapse for pair 4, with no evidence of LOH, while the reverse is true for pair 5. *CAP10* was also found to be tetraploid (as the genomes of these isolates were found to be diploid) in both initial and recurrent infections for pairs 3 and 17.

The *ERG11* gene on chromosome 1 was found to have increased copy number in numerous pairs (2, 3, 4, 5, 9, 12, and 17), and was not

found to be lineage-associated. However, this CNV was maintained throughout infection to recurrence in all pairs, with the exception of pair 4; since pair 4 initial infection (CCTP52) was found to have a diploid genome, *ERG11* was tetraploid, and lost this ploidy to be diploid with respect to the rest of the genome in the recurrent infection (CCTP52-d55). While chromosome 1 was duplicated in the initial infection isolate of pair 15 (IFNR11), *ERG11* was found to be haploid; the ploidy of the chromosome was subsequently lost in the recurrent infection isolate of pair 15 (IFNR11-d203).

*ERG11* in pair 4 (ID CCTP52) did not have any nsSNPs in the original infection (CCTP52), but one nsSNP was present in *ERG11* in the recurrent infection (CCTP52-d55). Clinical notes show that the patient from which pair 4 was isolated was given FLC (400 mg/d) on initial infection, did not attend follow up or receive ART or further FLC, and was then readmitted and died from CM recurrence at day 55. MIC values were unfortunately not available for either original or recurrent isolates.

### Nonsense mutations in DNA mismatch repair genes cause hypermutator states

Phylogenetic analysis on a chromosome-by-chromosome basis revealed that pair 3 isolates only clustered together in two of the fourteen chromosomes; the three isolates were phylogenetically dissimilar in four chromosomes, while day 257 and 409 isolates (pair 3 CCTP50-d257 and CCTP50-d409) were phylogenetically more similar to each other than to the day 1 isolate (CCTP50) in five chromosomes. Day 1 and day 409 isolates were more phylogenetically similar than to the day 257 isolate in three chromosomes. The lack of phylogenetic similarity (12 out of 14 chromosomes) shown in the three isolates of pair 3 indicated that these three isolates do not show a recent common ancestor, and provides evidence for reinfection with a new isolate, rather than relapse. In contrast, the two isolates in pair 17 were only phylogenetically related for five out of the 14 chromosomes (Figure S1), suggesting that on this basis the recurrent infection was distinct enough as to be defined as a nonrelapse infection in this pair. However, on further investigation this was found not to be the case.

We analyzed the coverage profiles and synonymous/nonsynonymous ratios of both isolates in pair 17 (Figure 2C). Although the

aneuploidies observed were extensive throughout the genome, the increases in ploidy appeared on similar chromosomes in both isolates. A similar observation was seen for the strikingly increased number of nsSNPs in both isolates in pair 17: 41,549 and 46,622 nsSNPs for IFNR23 and IFNR23-d179, respectively; an even larger increase in synonymous SNPs was also observed (68,094 and 82,172 synonymous SNPs for IFNR23 and IFNR23-d179, respectively). Therefore, we sought to identify a mechanism responsible for the high number of synonymous and nsSNPs, and ploidy.

Previous studies have reported that mutations in the DNA mismatch repair gene *MSH2* have resulted hypermutator effects in bacteria and the yeast *S. cerevisiae* (Drotschmann *et al.* 1999). Both pair 17 isolates were found to harbor two nonsense (*i.e.*, point mutations in the DNA sequence that result in a premature stop codon) mutations within the coding region of the gene encoding *MSH2*, the DNA mismatch repair protein. Nonsense mutations in *MSH2* were not observed in any other pairs included in this study. These mutations were in the same positions in both the original and recurrent isolates (Ser-888-STOP and Ser-86-STOP).

We then performed a genome-wide search in both pair 17 isolates to identify further nonsense mutations in DNA mismatch repair genes. Both pair 17 isolates were found to harbor a single nonsense mutation within the coding regions of genes encoding *MSH5* and *RAD5*. Again, nonsense mutations in these genes were not observed in any other pairs included in this study. These nonsense mutations caused Gln-1066-STOP in *RAD5* and Gln-709-STOP in *MSH5* in both original and recurrent isolates.

Since the likelihood of such mutations occurring by chance in independent genomes lacking a common ancestor is very small, this suggests that rather than being a reinfection, this was indeed a relapse of the original infection, and the phylogenetic dissimilarity between the two isolates was due to hypermutation. A total of 293 SNPs were located in *MSH2* in both pair 17 isolates, compared to a mean of 30 SNPs per isolate in the remaining pairs that we studied (Table S2 in File S1). More SNPs overall were observed in the recurrent isolate (IFNR23-d179) in both *RAD5* and *MSH5* (361 and 357, respectively) when compared to the original isolate (IFNR23 – 320 for *RAD5* and 305 for *MSH5*). These numbers are considerably higher than the average of 46 SNPs and 37 SNPs per isolate in the remaining pairs included in this study for *RAD5* and *MSH5*, respectively.

## DISCUSSION

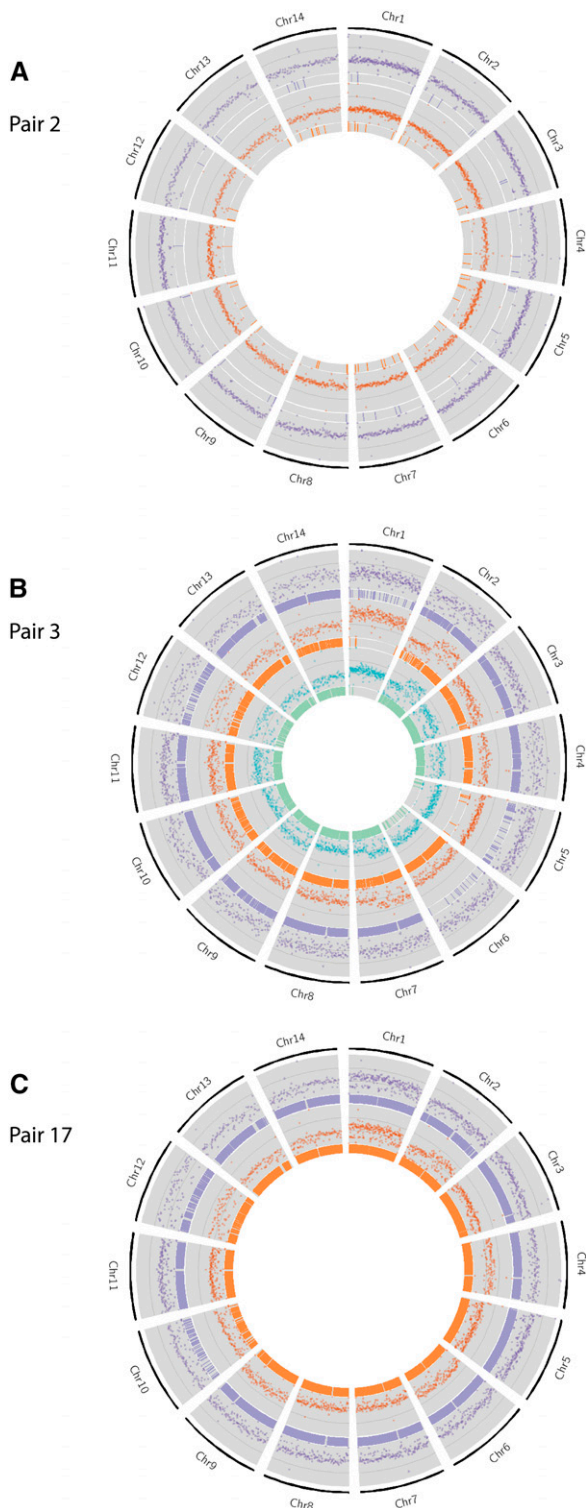
Relapse of CM caused by *C. neoformans* is usually due to the persistence and recurrence of the original infecting isolate (Spitzer *et al.* 1993), and studies often focus on rates of within-host microevolution between serially collected isolates. However, recent studies have shown that an infection of a population of dissimilar genotypes is responsible for 20% of relapse cases (Desnos-Ollivier *et al.* 2010). We used WGS to distinguish co-infections of a population of genotypes from a relapse of a single genotype owing to treatment failure (Figure 3). Using WGS we can distinguish between relapses of infection of the same genotype, which differ by only a few SNPs, while initial infection by a population of dissimilar genotypes will see a difference of many SNPs between initial and relapse infection as genetic drift occurs. Our results show that *C. neoformans* incurs numerous unique small and large-scale changes during infection, and that a subset of these may have adaptive value. While this study is concerned with the genomics of recurrent infections by identifying SNP changes and ploidy potentially involved in the persistence of *C. neoformans* infection, future work should investigate the potential role of gene expression changes and gene

networks involved in changes in fitness among the populations of infecting genotypes that underpin the recurrence of infection. Previous studies in *Brucella* infection and TB have highlighted the merits of using transcriptomics to identify patients requiring more intensive treatment (Dufort *et al.* 2016) and differentiating between dormancy and reactivation (Kondratieva *et al.* 2014), respectively. Given the higher genetic variation observed in the VNB lineage, the genomic data could also be exploited to investigate genome content variation, as this is known to be a major determinant in yeast phenotypic variation (56). Such approaches are likely to increase our understanding of clinical cases of recurrent *C. neoformans* infections through identifying the genetic basis of phenotypic switching (D'Souza and Heitman 2001) and the gene regulatory networks involved in latency, virulence, and resistance to antifungal therapies.

One pair (pair 3) included in this study did not display a relapse of the initial infecting isolate. Analysis of this pair showed that only 46% of SNPs were in common between the initial and recurrent infection, suggesting that relapse was caused by a new, albeit similar, genotype. These co-infection events are rarely reported in the literature; however, Hagen *et al.* (2015) did find evidence of co-infection in a single patient using AFLP (58). The extensive chromosomal CNVs, or aneuploidies, observed in pair 3 (Figure 2) also show that different genotypes were isolated at subsequent timepoints (days 257 and 409). Phylogenetic analyses showed that this pair belongs to the VNB lineage; it is known that a population of VNB genotypes can be found in one location, such as on the same tree (Vanhove *et al.* 2016). Therefore, it is possible for a single immunocompromised individual to inhale a cluster of basidiospores from a single mating population, which would lead to a cluster of related, but recombined, genotypes that then come to dominate the infection at different timepoints. Although a population can reside in an environmental reservoir, recombination between genotypes can occur, generating closely related, yet distinct, genotypes (Figure 3B). This latter hypothesis supports our observations of differing numbers of nsSNPs between day 257 (CCTP50-d257) and the original (CCTP50) and day 409 (CCTP50-d409) isolates, as well as the ploidies and MIC values (Table 1) seen at the time of sample isolation: the day 1 isolate (CCTP50) initially had an intermediate FLC MIC of 16 µg/ml, while the recurrent isolate at day 257 had a highly resistant FLC MIC of 256 µg/ml. These MIC values are suggestive of drug-resistant genotypes being present and selected for within this patient by the prolonged maintenance on FLC monotherapy following induction therapy with amphotericin B. It is also likely that the population of VNB isolates circulating in the patient were not sufficiently sampled by sequencing only one colony at each timepoint, and that deeper sequencing would have uncovered greater genomic diversity.

The occurrence of aneuploidy, where an abnormal number of chromosomes is observed, is seen as an evolutionary process that rapidly alters fitness, and has been described in multiple human fungal pathogens as a means of generating drug resistance (Selmecki *et al.* 2006; Sionov *et al.* 2010). Sionov *et al.* (2010) reported the duplication of multiple chromosomes in response to high concentrations of FLC, which resulted in genotypes developing FLC drug resistance. Associated gene duplications in *C. neoformans* chromosome 1 included *ERG11* and *AFR1*, which are both transporters of azole drugs. While duplications of *ERG11* were seen in seven pairs (2, 3, 4, 5, 9, 12, and 17), these were not necessarily associated with an entire duplication of chromosome 1. Sionov *et al.* (2010) suggested that *ERG11* contributed to the duplication of chromosome 1; we observed only one isolate (IFNR11 of pair 15) displaying a duplication of chromosome 1, but a single copy of *ERG11*, suggesting that ploidy was not complete throughout the chromosome. Since this isolate was the initial infection,





**Figure 2** Extensive chromosomal copy number variation was observed in all isolates in pairs 3 and 17, when compared to H99. Pair 2 is included to illustrate isolates without ploidy and extensive nsSNPs. Here, normalized whole-genome depth of coverage is shown, averaged over 10,000 bp bins, in scatter plots. Bar plots represent the position of nsSNPs. The purple track represents the original isolate, orange the recurrent isolate, and green (in the case of pair 3) for the final recurrent isolate. (A) No increase in ploidy is observed in either the original or recurrent isolate of Pair 2, and a small number of nsSNPs

we can assume that the duplication of chromosome 1 was not solely due to the stress of azole drug treatment, suggesting that ploidy can be activated under different conditions, such as the stress associated with adaptation to the host. A possible limitation is that the observed duplication may be due to prolonged frozen storage.

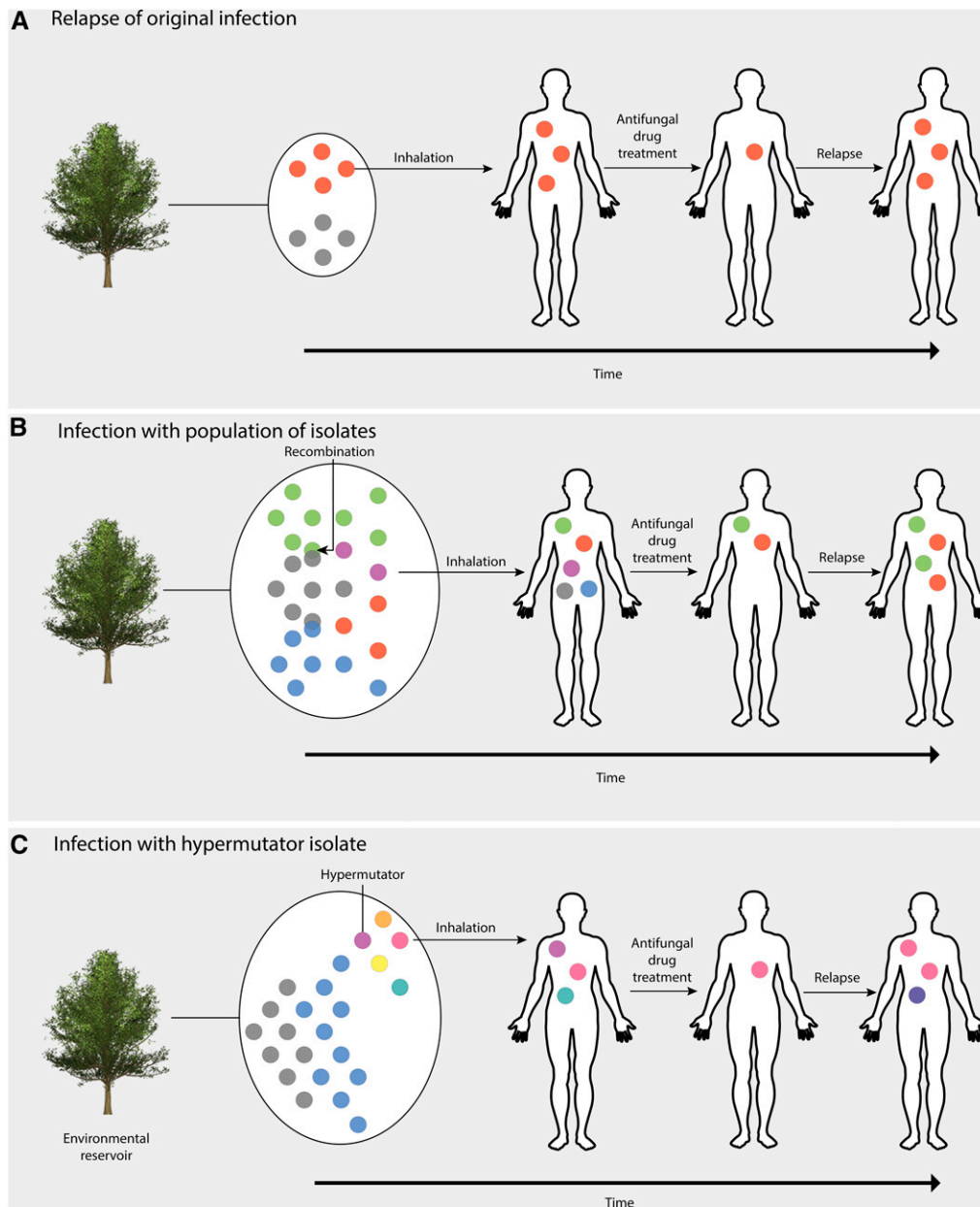
Ormerod *et al.* (2013) showed an aneuploidy (duplication) in chromosome 12 between serially collected isolates. Four pairs included in this study (pairs 1, 5, 10, and 14) all showed aneuploidy in chromosome 12; however, pair 14 (ID IFNR19) displayed this aneuploidy in both the initial and relapse infections. Since aneuploidies are typically lost upon removal of drug pressure (Sionov *et al.* 2010), one can assume that this aneuploidy was maintained due to previous drug exposure potentially not being reported by the patient, or that aneuploidy helps *C. neoformans* adapt to the host environment (Morrow and Fraser 2013). Chromosome 12 experienced triploidy in the pair 1 recurrent isolate (CCTP27-d121); this pair also demonstrated drug resistance to FLC, with a FLC MIC of four at initial infection, and a FLC MIC of 64 at recurrent infection. Ormerod *et al.* (2013) hypothesize that the large number of genes affected by the increased copy number of chromosome 12 contributes to metabolome differences; however, we hypothesize that CNV of chromosome 12 is a response to FLC stress, resulting in increased MIC, and that some genes present on chromosome 12, such as *ERG8* and *CAP6*, may be targets of azole drugs or involved in *C. neoformans* virulence.

Antimicrobial drugs impose strong selection pressure on pathogens, which may lead to the evolution of drug resistance (Mu *et al.* 2010; Didelot *et al.* 2016); there are, however, fitness costs associated with the evolution of resistance to antifungal drugs that may impact fitness (Cowen *et al.* 2001). Genome-wide scans for sites under selection lead to the identification of possible sites of drug resistance. We did not identify any significant sites when comparing VNI original infection vs. recurrent infection, and the number of VNB and VNII isolates were too low for analysis. While these results could be interpreted as there being no sites under selection in the VNI isolates sampled in this study, it is more likely that similar patterns would not be seen among individuals due to stochasticity and clonal interference (Didelot *et al.* 2016). It is also likely that, as there is little recombination in VNI isolates compared to VNB and VNII isolates (Khayhan *et al.* 2013; Litvintseva *et al.* 2006, 2011), linkage is complete across the genome, further hampering selection analysis. Therefore, we found no evidence for genetically determined alterations in drug resistance in the study isolates.

MIC values were only obtained for 9 out of 35 isolates in this study at the time of sampling. Susceptibility testing at a later date revealed all the isolates to be susceptible to antifungal drugs including FLC, suggesting that any resistant phenotypes had been lost in the absence of drug selective pressure. Therefore, it is very important that clinicians request susceptibility testing in real time, at the very least in all cases of recurrent CM.

While pair 17 did not exhibit a high percentage of common SNPs between the original and recurrent isolates indicative of a relapse

are seen. (B) Increase-in-ploidy is observed in many chromosomes in the day 1 isolate for pair 3, some of which are lost over time. A large number of nsSNPs are observed in all chromosomes in isolates of pair 3, with Chr 6 being the exception: very few nsSNPs are located in Chr 6 in CCTP50 and CCTP50-d409, whereas over 2000 nsSNPs are observed in Chr 6 in CCTP50-d257. (C) A gain in ploidy is observed for Chrs 2, 4, 6, and 9 compared to the day 1 isolate in pair 17, whereas ploidy remains unchanged for Chr 1 and 12. Chr, chromosome; nsSNP, nonsynonymous single nucleotide polymorphism.



**Figure 3** Hypotheses of routes of infection of the human host by *C. neoformans* var. *grubii*. (A) Inhalation of a single population of basidiospores into a new host. Due to low within-host diversity and being drug naïve, there will be a bottleneck in population size due to antifungal drug treatment. However, if the initial drug regimen is insufficient to sterilize the cerebral spinal fluid, resistance may develop on fluconazole maintenance therapy due to selection pressure, resulting in relapsed infection from proliferation of a drug-resistant isolate. (B) VNB lineage *C. neoformans* exists in the environment as a population that can undergo recombination to produce genetically similar isolates, but with significantly diversity. Due to transmission bottlenecks, only a sample of the pathogen diversity will be transferred to the host, in this case by inhalation, but it is possible for a population of *C. neoformans* to infect a single immunocompromised individual. Some isolates may be susceptible to antifungal drugs and are thus removed, while other isolates may be inherently resistant and hence cause a relapse infection. (C) Mutations in the DNA mismatch repair gene *MSH2* cause an isolate to become a hypermutator. Some genotypes may be susceptible to antifungal drugs, but the high mutation rate allows the infection to adapt rapidly to the host and evolve drug resistance. These genotype proliferate in the host, thus causing relapse infection.

infection, the elevated rate of SNPs observed in all chromosomes of both isolates suggested that this was not a reinfection as seen in pair 3 (Figure 3C). Rather, our results suggest that the isolates in this pair were exhibiting a hypermutator phenotype, as a result of two nonsense mutations in the DNA mismatch repair gene *MSH2*, and one nonsense mutation in each of the DNA mismatch repair genes *RAD5* and *MSH5*. While previous studies have shown that hypermutator phenotypes aid adaptation to stress (Magditch *et al.* 2012), we hypothesize that hypermutation may lead to adaptation of drug resistance under the stress of antifungal treatment. These results are, to the authors' knowledge, the first to report on nonsense mutations in *MSH2*, *RAD5*, and *MSH5* in *C. neoformans*. Further investigation is required to determine whether these nonsense mutations have a role in drug resistance phenotypes using transcriptomic approaches and creating single-gene knockout mutants of *MSH2*, *RAD5*, and *MSH5*. It is also necessary to test the

virulence of hypermutator isolates in the mouse model and to describe the impact of the increase in mutation rate that occurs as a result of this hypermutation. Our study only includes one pair of hypermutator genotypes, so further sampling is required to identify whether this phenomenon is specific to the VNB lineage, whether hypermutators occur in the VNII and VNI lineages, and whether they are clinically relevant.

This work represents the most extensive comparative genome sequencing-based study to investigate microevolution in serially collected isolates of *C. neoformans* to date. The observation of an infection of a single patient with a population of VNB isolates is clinically relevant, as widely used drug regimens with azole monotherapy may not be effective against such a genetically diverse infection. It is also likely that the extensive genetic diversity seen in clinically isolated VNB isolates may be due to mixed infection. Hypermutation due to nonsense

mutations in the DNA mismatch repair genes *MSH2*, *RAD5*, and *MSH5* cause an increased mutation and rate of aneuploidy in *C. neoformans*, which may confer an increased ability to adapt to drug pressure. Further sampling is required to identify whether hypermutation is a phenomenon only observed in the VNB lineage, and how these mutations impact the fitness of *C. neoformans* by imposing a high genetic load.

## ACKNOWLEDGMENTS

The authors extend special thanks to Ian Wright for his financial contribution to this project, and Winnie Wu and StarLab UK for supplying filter pipette tips free of charge. They also thank Sean Lobo and Ahsan Awais Bokhari for initial work on minimum inhibitory concentrations and multi-locus sequence typing, respectively. This work was funded by a Medical Research Council grant (MRC MR/K000373/1) awarded to M.C.F. and a grant from Imperial College London (WPIA/F24085) awarded to J.R. J.N.J. was supported by the Wellcome Trust (training fellowship, WT081794) and is supported by a grant from the Penn Center for AIDS Research, a National Institutes of Health-funded program (P30 AI 045008).

Author contributions: T.B., J.N.J., J.A.S., A.R., G.M., and T.S.H. conducted the clinical studies and provided isolates included in this study. T.B., J.N.J., and M.C.F. conceived the research question. J.R. and M.A.B. conceived the experiments. M.A.B. performed library preparations for sequencing, with assistance from M.V. J.R. performed alignments, variant calling and downstream analysis, and wrote the paper. M.V. performed BayeScan analysis. All authors read and contributed to this article.

## LITERATURE CITED

- Abdolrasouli, A., J. Rhodes, M. A. Beale, F. Hagen, T. R. Rogers *et al.*, 2015 Genomic context of azole resistance mutations in aspergillus fumigatus determined using whole-genome sequencing. *MBio* 6(3): e00536 (erratum: *MBio* 6(4): e00939).
- Alastruey-Izquierdo, A., and M. Cuenca-Estrella, 2012 EUCAST and CLSI: how to assess in vitro susceptibility and clinical resistance. *Curr. Fungal Infect. Rep.* 6(3): 229–234.
- Alexander, B. D., M. D. Johnson, C. D. Pfeiffer, C. Jiménez-Ortigosa, J. Catania *et al.*, 2013 Increasing echinocandin resistance in *Candida glabrata*: clinical failure correlates with presence of *FKS* mutations and elevated minimum inhibitory concentrations. *Clin. Infect. Dis.* 56(12): 1724–1732.
- Beale, M. A., W. Sabiti, E. J. Robertson, K. M. Fuentes-Cabrejo, S. J. O'Hanlon *et al.*, 2015 Genotypic diversity is associated with clinical outcome and phenotype in cryptococcal meningitis across Southern Africa. *PLoS Negl. Trop. Dis.* 9(6): e0003847.
- Bicanic, T., G. Meintjes, R. Wood, M. Hayes, K. Rebe *et al.*, 2007 Fungal burden, early fungicidal activity, and outcome in cryptococcal meningitis in antiretroviral-naïve or antiretroviral-experienced patients treated with amphotericin B or fluconazole. *Clin. Infect. Dis.* 45(1): 76–80.
- Bicanic, T., R. Wood, G. Meintjes, K. Rebe, A. Brouwer *et al.*, 2008 High-dose amphotericin B with flucytosine for the treatment of cryptococcal meningitis in HIV-infected patients: a randomized trial. *Clin. Infect. Dis.* 47(1): 123–130.
- Billmyre, R. B., D. Croll, W. Li, P. Mieczkowski, D. A. Carter *et al.*, 2014 Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. *MBio* 5(4): e01494–e01514.
- Blasi, E., A. Brozzetti, D. Francisci, R. Neglia, G. Cardinali *et al.*, 2001 Evidence of microevolution in a clinical case of recurrent *Cryptococcus neoformans* meningoencephalitis. *Eur. J. Clin. Microbiol. Infect. Dis.* 20(8): 535–543.
- Bozzette, S. A., R. A. Larsen, J. Chiu, M. A. Leal, J. Jacobsen *et al.*, 1991 A placebo-controlled trial of maintenance therapy with fluconazole after treatment of cryptococcal meningitis in the acquired immunodeficiency syndrome. *N. Engl. J. Med.* 324(9): 580–584.
- Brouwer, A. E., A. Rajanuwong, W. Chierakul, G. E. Griffin, R. A. Larsen *et al.*, 2004 Combination antifungal therapies for HIV-associated cryptococcal meningitis: a randomised trial. *Lancet* 363(9423): 1764–1767.
- Burns, J. L., R. L. Gibson, S. McNamara, D. Yim, J. Emerson *et al.*, 2001 Longitudinal assessment of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *J. Infect. Dis.* 183(3): 444–452.
- Calo, S., R. B. Billmyre, and J. Heitman, 2013 Generators of phenotypic diversity in the evolution of pathogenic microorganisms. *PLoS Pathog.* 9(3): e1003181.
- Chang, Y. C., S. S. Ingavale, C. Bien, P. Espenshade, and K. J. Kwon-Chung, 2009 Conservation of the sterol regulatory element-binding protein pathway and its pathobiological importance in *Cryptococcus neoformans*. *Eukaryot. Cell* 8(11): 1770–1779.
- Cowen, L. E., L. M. Kohn, and J. B. Anderson, 2001 Divergence in fitness and evolution of drug resistance in experimental populations of *Candida albicans*. *J. Bacteriol.* 183(10): 2971–2978.
- de Jesús-Berrios, M., L. Liu, J. C. Nussbaum, G. M. Cox, J. S. Stamler *et al.*, 2003 Enzymes that counteract nitrosative stress promote fungal virulence. *Curr. Biol.* 13(22): 1963–1968.
- Desnos-Ollivier, M., S. Patel, A. R. Spaulding, C. Charlier, D. Garcia-Hermoso *et al.*, 2010 Mixed infections and *in vivo* evolution in the human fungal pathogen *Cryptococcus neoformans*. *MBio* 1(1): e00091–e00110.
- Didelot, X., A. S. Walker, T. E. Peto, D. W. Crook, D. J. Wilson *et al.*, 2016 Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* 14: 150–162.
- Drotschmann, K., A. B. Clark, and T. A. Kunkel, 1999 Mutator phenotypes of common polymorphisms and missense mutations in *MSH2*. *Curr. Biol.* 9(16): 907–910.
- D'Souza, C. A., and J. Heitman, 2001 It infects me, it infects me not: phenotypic switching in the fungal pathogen *Cryptococcus neoformans*. *J. Clin. Invest.* 108(11): 1577–1578.
- Dufort, M. J., S. Hanniffy, R. C. Álvarez, V. A. Gorvel, M. C. Altman *et al.*, 2016 Gene expression profiles predict treatment outcomes in Brucellosis. *J. Immunol.* 196(Suppl. 1): 66.21.
- Excoffier, L., T. Hofer, and M. Foll, 2009 Detecting loci under selection in a hierarchically structured population. *Heredity* 103(4): 285–298.
- Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.
- Forche, A., P. T. Magee, A. Selmecki, J. Berman, and G. May, 2009 Evolution in *Candida albicans* populations during a single passage through a mouse host. *Genetics* 182: 799–811.
- Ford, C. B., J. M. Funt, D. Abbey, L. Issi, C. Guiducci *et al.*, 2015 The evolution of drug resistance in clinical isolates of *Candida albicans*. *ELife* 4: e00662.
- Fraser, J. A., J. C. Huang, R. Pukkila-Worley, J. A. Alspaugh, T. G. Mitchell *et al.*, 2005 Chromosomal translocation and segmental duplication in *Cryptococcus neoformans*. *Eukaryot. Cell* 4(2): 401–406.
- Fries, B. C., F. Chen, B. P. Currie, and A. Casadevall, 1996 Karyotype instability in *Cryptococcus neoformans* infection. *J. Clin. Microbiol.* 34(6): 1531–1534.
- Hagen, F., K. Khayhan, B. Theelen, A. Kolecka, I. Polacheck *et al.*, 2015 Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. *Fungal Genet. Biol.* 78: 16–48.
- Hickman, M. A., C. Paulson, A. Dudley, and J. Berman, 2015 Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in *Candida albicans*. *Genetics* 200: 781–794.
- Illnait-Zaragozi, M. T., G. F. Martínez-Machín, C. M. Fernández-Andreu, F. Hagen, T. Boekhout *et al.*, 2010 Microsatellite typing and susceptibilities of serial *Cryptococcus neoformans* isolates from Cuban patients with recurrent cryptococcal meningitis. *BMC Infect. Dis.* 10(1): 289.



- Jarvis, J. N., and T. S. Harrison, 2007 HIV-associated cryptococcal meningitis. *AIDS* 21(16): 2119–2129.
- Jarvis, J. N., G. Meintjes, A. Williams, Y. Brown, T. Crede *et al.*, 2010 Adult meningitis in a setting of high HIV and TB prevalence: findings from 4961 suspected cases. *BMC Infect. Dis.* 10(1): 67.
- Jarvis, J. N., G. Meintjes, K. Rebe, G. N. Williams, T. Bicanic *et al.*, 2012 Adjunctive interferon- $\gamma$  immunotherapy for the treatment of HIV-associated cryptococcal meningitis: a randomized controlled trial. *AIDS* 26(9): 1105–1113.
- Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, 2016 KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44(D1): D457–D462.
- Khayhan, K., F. Hagen, W. Pan, S. Simwami, M. C. Fisher *et al.*, 2013 Geographically structured populations of *Cryptococcus neoformans* variety *grubii* in Asia correlate with HIV status and show a clonal population structure. *PLoS One* 8(9): e72222.
- Kondratieva, T., T. Azhikina, B. Nikonenko, A. Kaprelyants, and A. Apt, 2014 Latent tuberculosis infection: what we know about its genetic control? *Tuberculosis (Edinb.)* 94(5): 462–468.
- Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997v2. Available at: <https://arxiv.org/pdf/1303.3997.pdf>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16): 2078–2079.
- Litvintseva, A. P., R. Thakur, R. Vilgalys, and T. G. Mitchell, 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.
- Litvintseva, A. P., I. Carbone, J. Rossouw, R. Thakur, N. P. Govender *et al.*, 2011 Evidence that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have evolved in Africa. *PLoS One* 6(5): e19688.
- Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo *et al.*, 2005 The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307(5713): 1321–1324.
- Longley, N., C. Muzoora, K. Taseera, J. Mwesigye, J. Rwebembera *et al.*, 2008 Dose response effect of high-dose fluconazole for HIV-associated cryptococcal meningitis in southwestern Uganda. *Clin. Infect. Dis.* 47(12): 1556–1561.
- Maciá, M. D., D. Blanquer, B. Togores, J. Sauleda, and J. L. Pérez, 2005 Hypermutation is a key factor in development of multiple-antimicrobial resistance in *Pseudomonas aeruginosa* strains causing chronic lung infections. *Antimicrob. Agents Chemother.* 49(8): 3382–3386.
- Magditch, D. A., T. B. Liu, C. Xue, and A. Idnurm, 2012 DNA mutations mediate microevolution between host-adapted forms of the pathogenic fungus *Cryptococcus neoformans*. *PLoS Pathog.* 8(10): e1002936.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9): 1297–1303.
- Meyer, W., D. M. Aanensen, T. Boekhout, M. Cogliati, M. R. Diaz *et al.*, 2009 Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med. Mycol.* 47(6): 561–570.
- Morrow, C. A., and J. A. Fraser, 2013 Ploidy variation as an adaptive mechanism in human pathogenic fungi. *Semin. Cell Dev. Biol.* 24(4): 339–346.
- Mu, J., R. A. Myers, H. Jiang, S. Liu, S. Ricklefs *et al.*, 2010 Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat. Genet.* 42(3): 268–271.
- Oliver, A., R. Cantón, P. Campo, F. Baquero, and J. Blázquez, 2000 High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* 288(5469): 1251–1254.
- Ormerod, K. L., C. A. Morrow, E. W. Chow, I. R. Lee, S. D. Arras *et al.*, 2013 Comparative genomics of serial isolates of *Cryptococcus neoformans* reveals gene associated with carbon utilization and virulence. *G3 (Bethesda)* 3: 675–686.
- Panackal, A. A., J. L. Gribkov, J. F. Staab, K. A. Kirby, M. Rinaldi *et al.*, 2006 Clinical significance of azole antifungal drug cross-resistance in *Candida glabrata*. *J. Clin. Microbiol.* 44(5): 1740–1743.
- Park, B. J., K. A. Wannemuehler, B. J. Marston, N. Govender, P. G. Pappas *et al.*, 2009 Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS* 23(4): 525–530.
- Paul, S., T. L. Doering, and W. S. Moye-Rowley, 2015 *Cryptococcus neoformans* Yap1 is required for normal fluconazole and oxidative stress resistance. *Fungal Genet. Biol.* 74: 1–9.
- Pfaller, M. A., 2012 Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. *Am. J. Med.* 125(Suppl. 1): S3–S13.
- Rhodes, J., M. A. Beale, and M. C. Fisher, 2014 Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. *PLoS One* 9(11): e113501.
- Rodero, L., E. Mellado, A. C. Rodriguez, A. Salve, L. Guelfand *et al.*, 2003 G484S amino acid substitution in lanosterol 14- $\alpha$  demethylase (*ERG11*) is related to fluconazole resistance in a recurrent *Cryptococcus neoformans* clinical isolate. *Antimicrob. Agents Chemother.* 47(11): 3653–3656.
- Rodríguez-Tudela, J. L., M. C. Arendrup, M. Cuenca-Estrella, J. P. Donnelly, and C. Lass-Flörl, 2010 EUCAST breakpoints for antifungals. *Drug News Perspect.* 23(2): 93–97.
- Selmecki, A., A. Forche, and J. Berman, 2006 Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* 313(5785): 367–370.
- Selmecki, A. M., K. Dulmage, L. E. Cowen, J. B. Anderson, and J. Berman, 2009 Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet.* 5(10): e1000705.
- 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(4): e1000848.
- Sionov, E., Y. C. Chang, and K. J. Kwon-Chung, 2013 Azole heteroresistance in *Cryptococcus neoformans*: emergence of resistant clones with chromosomal disomy in the mouse brain during fluconazole treatment. *Antimicrob. Agents Chemother.* 57(10): 5127–5130.
- Spitzer, E. D., S. G. Spitzer, L. F. Freundlich, and A. Casadevall, 1993 Persistence of initial infection in recurrent *Cryptococcus neoformans* meningitis. *Lancet* 341(8845): 595–596.
- Stamatakis, A., 2006 RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688–2690.
- Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365(6443): 274–276.
- Van der Auwera, G. A., M. O., C. Carneiro, R. Hartl, G. Poplin, Del Angel *et al.*, 2013 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43: 11.10.1–11.10.33.
- Vanhove, M., M. A. Beale, J. Rhodes, D. Chanda, S. Lakhi *et al.*, 2016 Genomic epidemiology of *Cryptococcus* yeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. *Mol. Ecol.* Available at: <http://onlinelibrary.wiley.com/doi/10.1111/mec.13891/full>.

Communicating editor: J. Berman