G3: Genes|Genomes|Genetics Early Online, published on February 10, 2017 as doi:10.1534/g3.116.037499

A population genomics approach to assessing the genetic basis of within-host microevolution underlying recurrent cryptococcal meningitis infection

Johanna Rhodes[1], Mathew A. Beale[2], Mathieu Vanhove[1], Joseph N. Jarvis[3,4,5], Shichina Kannambath[6], John A. Simpson[7], Anthea Ryan[7], Graeme Meintjes[8], Thomas S. Harrison[6], Matthew C. Fisher[1*], Tihana Bicanic[6*]

1 - Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

2 - Division of Infection and Immunity, Faculty of Medical Sciences, University College London, London, United Kingdom

3 - Botswana-UPenn Partnership, Gaborone, Botswana

4 - Division of Infectious Diseases, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

5 - Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

6 - Institute of Infection and Immunity, St. George's University London, London, United Kingdom

7 - National Health Laboratory Service, Greenpoint, Cape Town, South Africa

8 - Clinical Infectious Diseases Research Initiative, Institute of Infectious Disease and Molecular Medicine and Department of Medicine, University of Cape Town, Cape Town, South Africa

*Corresponding authors: matthew.fisher@imperial.ac.uk; tbicanic@sgul.ac.uk

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 to assess the genetic basis of infection in 17 HIV-infected individuals with recurrent cryptococcal meningitis. 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, whilst 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 non-synonymous 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.

Introduction

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 cryptococcal meningitis (CM) globally (Park et al. 2009), with the majority of infections caused by *Cryptococcus neoformans sensu stricto* (previously referred to as *Cryptococcus neoformans* var. *grubii)* (Jarvis & Harrison 2007).

Standard treatment for HIV-associated CM includes the long-term use of azole drugs such as fluconazole, following initial 1-2 week 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 sterilisation of cerebral spinal fluid (CSF)) can relapse due to persisting infections, which in some cases appear to have evolved resistance to firstline 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).

59 Various methods of within-host evolution are available to eukaryotic pathogens, most
60 notably sexual and parasexual reproduction, although these are difficult to observe due
61 to the often cryptic nature recombination of fungi. Aneuploidy, recombination in the
62 telomeres, and mutator states (Rodero et al. 2003) also provide means of rapid within-
63 host evolution, with other mechanisms still likely to be discovered. The accumulation of
64 SNPs alongside copy number variation and aneuploidy has been witnessed during
65 infection in different fungal pathogen species, enabling rapid adaptive evolution (Calo et
66 al. 2013) and conferring resistance to antifungal dugs (Hickman et al. 2015). Candidiasis
67 is caused by numerous *Candida* species, yet within-host evolution amongst these
68 species differs. Infectious strains of *Candida albicans* are usually susceptible to azole
69 antifungal drugs, but resistance can evolve via the evolution of drug-resistant aneuploid
70 isolates, which contain an isochromosome of the left arm of chromosome 5 (Selmecki et
71 al. 2009). The left arm of chromosome 5 contains two important genes involved in
72 resistance to antifungals: *ERG11*, a target of azoles, and *TAC1*, a transcription factor that
73 activates drug efflux pump expression. Conversely, *Candida glabrata* are intrinsically
74 poorly susceptible to azoles, and have more recently evolved multi-drug resistance to
75 both azoles and echinocandins (Pfaller 2012; Panackal et al. 2006; Alexander et al. 2013).

76 The occurrence of within-host diversity and recombination has been witnessed in
77 eukaryotic pathogens, notably *C. albicans*: mutation and recombination rates can be
78 increased under stressful conditions, such as drug treatment (Forche et al. 2009; Ford et
79 al. 2015), resulting in loss-of-heterozygosity (LOH) and aneuploidy (Forche et al. 2009).
80 These genetic alterations contribute to the maintenance of a population of *C. albicans*

81 within the host environment (Forche et al. 2009), and drug pressure can result in
82 diverging levels of fitness (Cowen et al. 2001).

83 Similar responses to antifungal drugs have been observed in *C. neoformans*; Point
84 mutations in the ortholog *ERG11* were also shown to confer fluconazole resistance, by
85 causing the amino acid substitution G484S (Rodero et al. 2003). Sionov *et al.* (Sionov et
86 al. 2010) demonstrated large scale chromosomal duplications (primarily chromosome 1)
87 are fundamental to overcoming fluconazole (FLC) drug pressure in a mouse model,
88 contributing to failure of FLC therapy. The duplication of chromosome 1 included
89 increased copy number of genes *ERG11*, the target of FLC, and *AFR1*, a transporter of
90 azoles (Sionov et al. 2010), although other genes are also thought to be involved in FLC
91 resistance (Sionov et al. 2013; Paul et al. 2015). Previous studies of serially collected *C.*
92 *neoformans* isolates have confirmed in-host microevolution, including the occurrence of
93 large-scale genomic rearrangements (Fraser et al. 2005; Blasi et al. 2001; Illnait-Zaragozi
94 et al. 2010). Like *C. albicans*, the *C. neoformans* genome is capable of undergoing
95 chromosomal duplication and loss under stresses such as drug pressure or invasion of
96 the human host (Fries et al. 1996). These chromosomal duplications are often lost when
97 the selective pressure is removed (Sionov et al. 2010).

98 The development of mutator states via hypermutability is a rapidly expanding area of
99 study in bacteria, particularly *Pseudomonas aeruginosa* in cystic fibrosis (CF) patients.
100 Here, hypermutability has been shown to have an association with antimicrobial
101 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- to 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*, *Cryptococcus 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 fluconazole. We used whole-genome sequencing 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 cryptococcal meningitis were studied. All patients were either part of observational studies or clinical trial (Bicanic et al. 2007; Bicanic et al. 2008; Jarvis et al. 2012; Longley et al. 2008; Jarvis et al. 2010). 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; Bicanic et al. 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 cryptococcal meningitis and were treated using induction therapy with 7-14 days' 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 fluconazole consolidation at 400 mg/d for 8 weeks 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/μL. The single patient in Uganda received induction therapy with fluconazole 1600 mg/d for 2 weeks followed by fluconazole consolidation and maintenance and ART, as above (*n* = 1 pair). As part of 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

144 subset of the study isolates, according to the methods outlines in Meyer *et al.* (Meyer et
145 al. 2009), with modifications as outlined in Beale *et al.* (Beale et al. 2015).

146 Molecular methods

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

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

163 Data policy

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.

Whole genome sequence 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 104x). 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).

Single nucleotide polymorphisms (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 are 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 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,

184 Cambridge, MA, USA) and the latest release (CNA3) of the *C. neoformans* reference
185 genome H99 and gene ontology.

186 Some isolates were suspected of having non-haploid genomes due to the high number
187 of low confidence variants. For these isolates, 'HaplotypeCaller' was repeated in diploid
188 mode.

189 The average (mean) coverage for each isolate were determined using GATK (McKenna
190 et al. 2010) version 3.4-46 'DepthOfCoverage' under default settings. The *C.*
191 *neoformans* H99 (Loftus et al. 2005) was again used as reference. In order to determine
192 aneuploidy, whole-genome coverage data was normalised and regions displaying
193 normalised coverage equal to 2 were deemed diploid events (likewise, normalised
194 coverage equal to 3 were deemed triploid events, and so on), whereas normalised
195 coverage equal to zero was deemed a deletion event.

196 Susceptibility testing

197 The susceptibility testing of all relapse isolates were performed with the MICRONAUT-
198 AM susceptibility testing system for yeast (Merlin) as recommended by the
199 manufacturer. MICRONAUT-AM allows the determination of MICs of amphotericin B,
200 flucytosine, fluconazole, voriconazole, posaconazole, itraconazole, micafungin,
201 anidulafungin and caspofungin, and commercialises the well-established, but laborious,
202 CLSI broth microdilution technique. Briefly, for each isolate five colonies were used to
203 prepare a 0.5-McFarland-standard suspension in 0.9% NaCl. 1:20 dilution was prepared
204 in 0.9% NaCl and 1:5 dilution was prepared in 11ml RPMI broth provided with the kit.

100µl AST indicator and 50µl Methylene blue solutions were mixed with the broth for manual susceptibility testing. The broth was then inoculated onto Merlin MICRONAUT 96-well testing plates (100ul/well) and incubated at 30°C for 72 hours. 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 *C. albicans* EUCAST (Vers. 7.0 / 12-08-2014) values (Rodríguez-Tudela et al. 2010; Alastruey-Izquierdo & 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.* (Abdolrasouli et al. 2015): 35 isolates from this study in addition to 27 isolates ('non-study') were included to show the phylogenetic context of true relapse infections. These non-study isolates, whilst from a clinical source, were not recurrent isolates and were not isolated as part of the clinical trials described in the earlier Methods section. Resulting phylogenies were visualised 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.

Gene Ontology and KEGG pathway analysis

Non-synonymous (nsSNPs) mutations unique to each timepoint for each pair were assessed for significantly overrepresented gene ontology (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 & 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 programme 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, 100 to 1,000) for the neutral model.

Results

Clinical and demographic information

The study included paired isolates from 17 patients, with a median age of 32 years (IQR 26-36) and median CD4 count at CM diagnosis of 22 (IQR 9-71) cells/μL. Six patients were male, 9 were female, with the gender of two patients unrecorded. The median time between initial and recurrence isolates was 115 days (minimum 55 days, maximum 409 days). 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 7 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 fluconazole for 2 weeks prior to recurrence. In four patients (CCTP27, CCTP50, RCT24, IFNR63) who were adherent to both ART and fluconazole at recurrence were assessed as having CM immune reconstitution inflammatory syndrome (CM-IRIS).

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 analysed 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 (AT) 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 whole-genome sequencing (WGS) to characterise these Pairs. STs for Pairs 1 and 6 were inconclusive, and suggestive of a mixed infection 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 +/-31.2 (standard deviation). To enable comparative studies and detect micro-evolutionary 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 Supplementary Materials Table 5.

Due to a high number of low-confidence SNPs filtered out in some isolates, which is suggestive of heterozygous SNPs, variant calling was re-run in diploid mode (see Methods) for all isolates in Pairs 3, 4, 5 and 17 (results in Supplementary Materials Table 6).

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

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

304 All pairs, with the exception of Pair 7, were isolated from patients in South Africa; Pair 7
305 was isolated from a patient in Uganda. We classified the second isolate as a relapse of

306 the original infection if more than 97% of SNPs were in common between original and
307 recurrent isolates. The majority of pairs had >99% SNP similarity (Table 2) between
308 original and recurrent isolates, with Pairs 6, 7 and 14 displaying 97%, 98% and 97%
309 similarity respectively. Therefore, all pairs, with the exception of Pairs 3 and 17 (SNP
310 similarity 44% and 56%) could be classified as relapsed infections on this basis. This
311 confirms previous results obtained by MLST, and that the original and recurrent isolates
312 sequenced of Pairs 1 and 7 (which had previously demonstrated a potential mixed
313 infection) were indeed true relapse infections.

314 Within the VNB pairs (3, 12 and 17), the accumulation of SNPs between original and
315 recurrent infection varied widely. We observed 178 and 304 SNPs/day for CCTP50-d257
316 and CCTP50-d409 respectively (Pair 3), 8 SNPs/day for Pair 12, and 968 SNPs/day for Pair
317 17. Due to the variation in SNP accumulation between Pair 12 and Pairs 3 and 17, we
318 hypothesised that Pair 12 was a true relapse of the original infection, whilst Pairs 3 and
319 17 are showing inflated SNP numbers due to reinfection or an anomalous rate of
320 evolution.

321 Antifungal susceptibility testing

322 Fluconazole susceptibility testing (Table 1) using the Etest® (bioMerieux) was carried out
323 for 12 isolates (including three paired isolates) in this study by the accredited central
324 Microbiology laboratory in Cape Town at the time of the clinical episode; five of these
325 (CCTP27-d121 in Pair 1; CCTP50 and CCTP50-d257 in Pair 3; RCT24-d154 in Pair6;
326 IFNR11-d203 in Pair 15) had MICs above the established epidemiological cut-off value (≥

8µg/ml) for fluconazole (FLC). All pairs were retested following 5-10 years frozen storage in glycerol using the MICRONAUT-AM system for yeast susceptibility (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 ug/ml, whilst the recurrent isolate was resistant at an MIC of 64 ug/ml; in Pair 3 (CCTP50), the initial isolate MIC was 16 ug/ml (intermediate), whilst a highly resistant MIC of 256 ug/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* isolated from the same patient were relapse infection of the original isolate, or infection 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 relapse 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 relapse of the original infections. However, only 46% and 56% of SNPs were found to be in common between initial and relapse infection 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 were repeated for each of the 14 *C. neoformans* chromosomes individually (Supplementary Figure 1).

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 (Supplementary Figure 1). All three genotypes were phylogenetically similar for three other chromosomes, however long branches and clustering with additional non-study 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 (6 chromosomes), or were phylogenetically distinct from one another, and were more phylogenetically related with other study or additional isolates (6 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 non-synonymous SNPs (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.

388 Five SNPs, all intergenic, were found to be common at Day 1, along with five different
389 SNPs, also intergenic, within VNB pairs (Pairs 3, 12 and 17). Three intergenic SNPs were
390 common to all VNII pairs (Pairs 2, 4, 5 and 9) at timepoint Day 1, whilst 14 SNPs were
391 common to all VNII pairs at the point of recurrent infection, five of which were
392 intergenic. The remaining 9 SNPs were located in the 5' untranslated region (UTR) gene
393 *SMF1* (CNAG_05640), a metal ion transporter with a natural resistance-associated
394 macrophage protein. Selection analysis indicated that this gene was not under selection
395 pressure, however.

396 To evaluate the genetic divergence, Wright's fixation indexes (F_{ST}) were calculated to
397 identify SNPs under selection in VNI original and recurrent infection populations
398 investigating 96,856 loci (see Methods). No putative loci under either diversifying or
399 balancing selection could be detected using a false discovery rate (FDR) of 0.05. F_{ST}
400 values were limited to not exceed 3.47^{x10-5}.

401 Aneuploidy as a generator of diversity in recurrent infection

402 Normalised whole-genome coverage was plotted to observe possible aneuploidy
403 (increase or decrease in copies of chromosomes) and copy number variation (CNV)
404 events. Aneuploidy events were observed in 7 genome pairs, suggesting either
405 interspersed or tandem duplications of large segments of the genome.

406 Ormerod *et al.* (Ormerod et al. 2013) previously published a study showing relapse
407 isolates exhibiting aneuploidies of chromosome 12. We observed aneuploidy of
408 chromosome 12 in four pairs (Pairs 1, 5, 10 and 14) included in this study. The

409 aneuploidy spanned different regions of chromosome 12 in all pairs, but all aneuploidies
410 were present in the right arm of the chromosome: in Pair 5 the aneuploidy was
411 restricted to 392 kbp of one chromosome arm; in Pair 1, the aneuploidy spanned an
412 entire arm of chromosome 12 (603 kbp). Pair 14 displayed this aneuploidy in both the
413 initial and recurrent infection, and the aneuploidy spanned the whole chromosome; the
414 recurrent infection isolate of Pair 10 also displayed aneuploidy along the whole
415 chromosome. Evaluation of the read depth along chromosome 12 revealed triplication
416 of the chromosome 12 arm in the Pair 1 recurrent infection isolate, a phenomenon also
417 seen in Ormerod *et al.* (Ormerod et al. 2013). Further analysis of read depth in the Pair
418 5 recurrent isolate revealed a diploid genome, and chromosome 12 was also
419 experiencing triploidy. Current annotation of the *C. neoformans* H99 genome reveals
420 the presence of 327 genes in chromosome 12; the right arm of chromosome 12 has 260
421 genes present. We scanned the genes present in this arm of chromosome 12 for genes
422 potentially involved in virulence, which might prove advantageous to the progression of
423 infection or drug resistance. One such gene was *SFB2* (CNAG_06093), which is involved
424 in the conservation of the sterol regulatory element-binding protein pathway (SREBP)
425 (Chang et al. 2009). An alcohol dehydrogenase (*GNO1* – CNAG_06168) was also present,
426 which is thought to be involved in the defence against host response (de Jesús-Berríos
427 et al. 2003). Analysis for enrichment of metabolic pathways also revealed that the
428 genes present in this chromosome arm are significantly involved in the metabolism of
429 drugs (corrected *p*-value $p < 3.81e^{-2}$).

We searched for copy number variation 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 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 loss of heterozygosity (LoH), whilst 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 to 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). Whilst chromosome 1 was duplicated in the initial infection isolate of Pair 15 (IFNR11), *ERG11* was found to be haploid; the ploidy of 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 nsSNPs was present in *ERG11* in the recurrent infection (CCTP52-d55). Clinical notes show that the patient from which Pair 4 was isolated was given fluconazole (400 mg/d) on initial infection, did not attend follow up or receive ART or further fluconazole, and was then re-admitted 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 4 chromosomes, whilst 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 (Supplementary Figure 1), suggesting that on this basis the recurrent infection was distinct enough as to be defined as a non-relapse infection in this Pair. However, on further investigation this was found not to be the case.

We analysed the coverage profiles and synonymous/non-synonymous 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). We therefore 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 harbour 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 harbour 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 suggest 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 (Supplementary Table 2). 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*, 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 whole-genome sequencing 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

513 relapses of infection of the same genotype, which differ by only a few SNPs, whilst initial
514 infection by a population of dissimilar genotypes will see a difference of many SNPs
515 between initial and relapse infection as genetic drift occurs. Our results show that *C.*
516 *neoformans* incurs numerous unique small and large-scale changes during infection, and
517 that a subset of these may have adaptive value. Whilst this study is concerned with the
518 genomics of recurrent infections by identifying SNP changes and ploidy potentially
519 involved in the persistence of *C. neoformans* infection, future work should investigate
520 the potential role of gene expression changes and gene networks involved in changes in
521 fitness amongst the populations of infecting genotypes that underpin the recurrence of
522 infection. Previous studies in *Brucella* infection and TB have highlighted the merits of
523 using transcriptomics to identify patients requiring more intensive treatment (Dufort et
524 al. 2016) and differentiating between dormancy and reactivation (Kondratieva et al.
525 2014) respectively. Given the higher genetic variation observed in the VNB lineage, the
526 genomic data could also be exploited to investigate genome content variation, as this is
527 known to be a major determinant in yeast phenotypic variation (56). Such approaches
528 are likely to increase our understanding of clinical cases of recurrent *C. neoformans*
529 infections through identifying the genetic basis of phenotypic switching (D'Souza &
530 Heitman 2001) and the gene regulatory networks involved in latency, virulence and
531 resistance to antifungal therapies.

532 One pair (Pair 3) included in this study did not display a relapse of the initial infecting
533 isolate. Analysis of this pair showed that only 46% of SNPs were in common between
534 the initial and recurrent infection, suggesting that relapse was caused by a new, albeit

similar, genotype. These co-infection events are rarely reported in literature; however, Hagen *et al*. did find evidence of co-infection in a single patient using AFLP (58). The extensive chromosomal copy number variations, 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 ug/ml, whilst the recurrent isolate at day 257 had a highly resistant FLC MIC of 256 ug/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.

557 The occurrence of aneuploidy, where an abnormal number of chromosomes is observed,
558 is seen as an evolutionary process that rapidly alters fitness, and has been described in
559 multiple human fungal pathogens as a means of generating drug resistance (Selmecki et
560 al. 2006; Sionov et al. 2010). Sionov *et al.* (Sionov et al. 2010) reported the duplication
561 of multiple chromosomes in response to high concentrations of FLC, which resulted in
562 genotypes developing FLC drug resistance. Associated gene duplications in *C.*
563 *neoformans* chromosome 1 included *ERG11* and *AFR1*, which are both transporters of
564 azole drugs. Whilst duplications of *ERG11* were seen in seven pairs (2, 3, 4, 5, 9, 12 and
565 17), these were not necessarily associated with an entire duplication of chromosome 1.
566 Sionov *et al.* suggested that *ERG11* contributed to the duplication of chromosome 1
567 (Sionov et al. 2010); we observed only one isolate (IFNR11 of Pair 15) displaying a
568 duplication of chromosome 1, but a single copy of *ERG11*, suggesting ploidy was not
569 complete throughout the chromosome. Since this isolate was the initial infection, we
570 can assume that the duplication of chromosome 1 was not solely due to stress of azole
571 drug treatment, suggesting that ploidy can be activated under different conditions, such
572 as the stress associated with adaptation to the host. A possible limitation is that the
573 observed duplication may be due to prolonged frozen storage.

574 Ormerod *et al.* (Ormerod et al. 2013) showed an aneuploidy (duplication) in
575 chromosome 12 between serially collected isolates. Four pairs included in this study
576 (Pair 1, 5, 10 and 14) all showed aneuploidy in chromosome 12; however, Pair 14 (ID
577 IFNR19) displayed this aneuploidy in both the initial and relapse infections. Since
578 aneuploidies are typically lost upon removal of drug pressure (Sionov et al. 2010), one

579 can assume that this aneuploidy was maintained due to previous drug exposure
580 potentially not reported by the patient, or that aneuploidy helps *C. neoformans* adapt to
581 the host environment (Morrow & Fraser 2013). Chromosome 12 experienced triploidy
582 in the Pair 1 recurrent isolate (CCTP27-d121); this pair also demonstrated drug
583 resistance to fluconazole, with a FLC MIC of 4 at initial infection, and a FLC MIC of 64 at
584 recurrent infection. Ormerod *et al.* (Ormerod et al. 2013) hypothesise that the large
585 number of genes affected by the increased copy number of chromosome 12 contributes
586 to metabolome differences; however, we hypothesise copy number variation of
587 chromosome 12 is a response to FLC stress, resulting in increased MIC, and that some
588 genes present on chromosome 12, such as *ERG8* and *CAP6*, may be targets of azole
589 drugs or involved in *C. neoformans* virulence.

590 Antimicrobial drugs impose strong selection pressure on pathogens, with may lead to
591 the evolution of drug resistance (Mu et al. 2010; Didelot et al. 2016); there are, however,
592 fitness costs associated with the evolution of resistance to antifungal drugs that may
593 impact fitness (Cowen et al. 2001). Genome-wide scans for sites under selection leads
594 to the identification of possible sites of drug resistance. We did not identify any
595 significant sites when comparing VNI original infection versus recurrent infection, and
596 the number of VNB and VNII isolates were too low for analysis. Whilst these results
597 could be interpreted as there being no sites under selection in the VNI isolates sampled
598 in this study, it is more likely that similar patterns would not be seen amongst
599 individuals due to stochasticity and clonal interference (Didelot et al. 2016). It is also
600 likely that as there is little recombination in VNI isolates compared to VNB and VNII

601 isolates (Khayhan et al. 2013; Litvintseva et al. 2011; Litvintseva 2005) and therefore
602 linkage is complete across the genome, further hampering selection analysis. We
603 therefore found no evidence for genetically determined alterations in drug resistance in
604 the study isolates.

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

610 Whilst Pair 17 did not exhibit a high percentage of common SNPs between the original
611 and recurrent isolates indicative of a relapse infection, the elevated rate of SNPs
612 observed in all chromosomes of both isolates suggested this was not a re-infection as
613 seen in Pair 3 (Figure 3c). Rather, our results suggest that the isolates in this pair were
614 exhibiting a hypermutator phenotype, as a result of two nonsense mutations in the DNA
615 mismatch repair gene *MSH2*, and one nonsense mutation in each of the DNA mismatch
616 repair genes *RAD5* and *MSH5*. Whilst previous studies have shown hypermutator
617 phenotypes aid adaptation to stress (Magditch et al. 2012), and we hypothesise that
618 hypermutation may lead to adaptation of drug resistance under the stress of antifungal
619 treatment. These results are the first to the authors' knowledge to report on nonsense
620 mutations in *MSH2*, *RAD5* and *MSH5* in *C. neoformans*. Further investigation is required
621 to determine whether these nonsense mutations have a role in drug resistance

622 phenotypes using transcriptomic approaches and creating single-gene knockout
623 mutants of *MSH2, RAD5* and *MSH5*. It is also necessary to test the virulence of
624 hypermutator isolates in the mouse model and to describe the impact of the increase in
625 mutation rate that occur a result of this hypermutation. Our study only includes one
626 pair of hypermutator genotypes, so further sampling is required to identify whether this
627 phenomenon is specific to the VNB lineage, whether hypermutators occur in the VNII
628 and VNI lineages, and whether they are clinically relevant.

629 This work represents the most extensive comparative genome-sequencing based study
630 to investigate microevolution in serially collected isolates of *C. neoformans* to date. The
631 observation of an infection of a single patient with a population of VNB isolates is
632 clinically relevant, as widely used drug regimens with azole monotherapy may not be
633 effective against such a genetically diverse infection. It is also likely that the extensive
634 genetic diversity seen in clinically isolated VNB isolates may be due to mixed infection.
635 Hypermutation due to nonsense mutations in the DNA mismatch repair genes *MSH2,*
636 *RAD5* and *MSH5* cause an increased mutation and rate of aneuploidy in *C. neoformans*,
637 which may confer an increased ability to adapt to drug pressure. Further sampling is
638 required to identify whether hypermutation is a phenomenon only observed in the VNB
639 lineage, and how these mutations impact the fitness of *C. neoformans* by imposing a
640 high genetic load.

641 **Acknowledgements**

642 Special thanks to Mr. Ian Wright for his financial contribution to this project, and Dr.

643 Winnie Wu and StarLab UK for supplying filter pipette tips free of charge. This work was
644 funded by a Medical Research Council grant (MRC MR/K000373/1) awarded to MCF and
645 a grant from Imperial College London (WPIA/F24085) awarded to JR. JJ was supported
646 by the Wellcome Trust (training fellowship, WT081794) and is supported by a grant from
647 the Penn Center for AIDS Research (CFAR), an NIH-funded program (P30 AI 045008).
648 Thankyou to Sean Lobo and Ahsan Awais Bokhari for initial work on MICs and MLST,
649 respectively.

650 **Author contributions**

651 TB, JJ, JS, AR, GM and TSH conducted the clinical studies and provided isolates included
652 in this study. TB, JJ and MCF conceived the research question. JR and MB conceived the
653 experiments; MB performed library preparations for sequencing, with assistance from
654 MV. JR performed alignments, variant calling and downstream analysis, and wrote the
655 paper. MV performed BayeScan analysis. All authors read and contributed to this paper.

656 **References**

657 Abdolrasouli, A. et al., 2015. Genomic Context of Azole Resistance Mutations in
658 Aspergillus fumigatus Determined Using Whole-Genome Sequencing. *mBio*, 6(3),
659 pp.e00536–15.

660 Alastruey-Izquierdo, A. & Cuenca-Estrella, M., 2012. EUCAST and CLSI: How to Assess in
661 Vitro Susceptibility and Clinical Resistance. *Current Fungal Infection Reports*, 6(3),
662 pp.229–234.

663 Alexander, B.D. et al., 2013. Increasing echinocandin resistance in Candida glabrata:
664 clinical failure correlates with presence of FKS mutations and elevated minimum
665 inhibitory concentrations. *Clinical Infectious Diseases*, 56(12), pp.1724–1732.

666 Beale, M.A. et al., 2015. Genotypic Diversity Is Associated with Clinical Outcome and

667 Phenotype in Cryptococcal Meningitis across Southern Africa J. M. Vinetz, ed. *PLoS*
668 *Neglected Tropical Diseases*, 9(6), p.e0003847.

669 Bicanic, T. et al., 2007. Fungal burden, early fungicidal activity, and outcome in
670 cryptococcal meningitis in antiretroviral-naive or antiretroviral-experienced patients
671 treated with amphotericin B or fluconazole. *Clinical Infectious Diseases*, 45(1),
672 pp.76–80.

673 Bicanic, T. et al., 2008. High-Dose Amphotericin B with Flucytosine for the Treatment of
674 Cryptococcal Meningitis in HIV-Infected Patients: A Randomized Trial. *Clinical*
675 *Infectious Diseases*, 47(1), pp.123–130.

676 Billmyre, R.B. et al., 2014. Highly Recombinant VGII Cryptococcus gattii Population
677 Develops Clonal Outbreak Clusters through both Sexual Macroevolution and Asexual
678 Microevolution. *mBio*, 5(4), pp.e01494–14–e01494–14.

679 Blasi, E. et al., 2001. Evidence of microevolution in a clinical case of recurrent
680 Cryptococcus neoformans meningoencephalitis. *European journal of clinical*
681 *microbiology & infectious diseases : official publication of the European Society of*
682 *Clinical Microbiology*, 20(8), pp.535–543.

683 Bozzette, S.A. et al., 1991. A placebo-controlled trial of maintenance therapy with
684 fluconazole after treatment of cryptococcal meningitis in the acquired
685 immunodeficiency syndrome. California Collaborative Treatment Group. *The New*
686 *England journal of medicine*, 324(9), pp.580–584.

687 Brouwer, A.E. et al., 2004. Combination antifungal therapies for HIV-associated
688 cryptococcal meningitis: a randomised trial. *Lancet*, 363(9423), pp.1764–1767.

689 Burns, J.L. et al., 2001. Longitudinal assessment of Pseudomonas aeruginosa in young
690 children with cystic fibrosis. *The Journal of infectious diseases*, 183(3), pp.444–452.

691 Calo, S., Billmyre, R.B. & Heitman, J., 2013. Generators of Phenotypic Diversity in the
692 Evolution of Pathogenic Microorganisms H. True-Krob, ed. *PLoS pathogens*, 9(3),
693 pp.e1003181–4.

694 Chang, Y.C. et al., 2009. Conservation of the sterol regulatory element-binding protein
695 pathway and its pathobiological importance in Cryptococcus neoformans.
696 *Eukaryotic Cell*, 8(11), pp.1770–1779.

697 Cowen, L.E., Kohn, L.M. & Anderson, J.B., 2001. Divergence in fitness and evolution of
698 drug resistance in experimental populations of Candida albicans. *Journal of*
699 *bacteriology*, 183(10), pp.2971–2978.

700 de Jesús-Berríos, M. et al., 2003. Enzymes that counteract nitrosative stress promote
701 fungal virulence. *Current Biology*, 13(22), pp.1963–1968.

Desnos-Ollivier, M. et al., 2010. Mixed Infections and In Vivo Evolution in the Human Fungal Pathogen Cryptococcus neoformans. *mBio*, 1(1), pp.e00091–10–e00091–10.

Didelot, X. et al., 2016. Within-host evolution of bacterial pathogens. *Nature reviews Microbiology*, 14, pp.150–162.

Drotschmann, K., Clark, A.B. & Kunkel, T.A., 1999. Mutator phenotypes of common polymorphisms and missense mutations in MSH2. *Current Biology*, 9(16), pp.907–910.

Dufort, M.J. et al., 2016. Gene expression profiles predict treatment outcomes in Brucellosis. *The Journal of Immunology*, 196(1 Supplement), pp.66.21–66.21.

D'Souza, C.A. & Heitman, J., 2001. It infects me, it infects me not: phenotypic switching in the fungal pathogen Cryptococcus neoformans. *Journal of Clinical Investigation*, 108(11), pp.1577–1578.

Excoffier, L., Hofer, T. & Foll, M., 2009. Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), pp.285–298.

Foll, M. & Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2), pp.977–993.

Forche, A. et al., 2009. Evolution in Candida albicans Populations During a Single Passage Through a Mouse Host. *Genetics*, 182(3), pp.799–811.

Ford, C.B. et al., 2015. The evolution of drug resistance in clinical isolates of Candida albicans. *eLife*, 4, p.e00662.

Fraser, J.A. et al., 2005. Chromosomal translocation and segmental duplication in Cryptococcus neoformans. *Eukaryotic Cell*, 4(2), pp.401–406.

Fries, B.C. et al., 1996. Karyotype instability in Cryptococcus neoformans infection. *Journal of clinical microbiology*, 34(6), pp.1531–1534.

Hagen, F. et al., 2015. Recognition of seven species in the Cryptococcus gattii/Cryptococcus neoformans species complex. *Fungal Genetics and Biology*, 78(C), pp.16–48.

Hickman, M.A. et al., 2015. Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in Candida albicans. *Genetics*, 200, pp.781–794.

Illnait-Zaragozi, M.T. et al., 2010. Microsatellite typing and susceptibilities of serial Cryptococcus neoformans isolates from Cuban patients with recurrent cryptococcal

735 meningitis. *BMC Infectious Diseases*, 10(1), p.289.

736 Jarvis, J.N. & Harrison, T.S., 2007. HIV-associated cryptococcal meningitis. *AIDS*, 21(16),
737 pp.2119–2129.

738 Jarvis, J.N. et al., 2012. Adjunctive interferon-γ immunotherapy for the treatment of
739 HIV-associated cryptococcal meningitis. *AIDS*, 26(9), pp.1105–1113.

740 Jarvis, J.N. et al., 2010. Adult meningitis in a setting of high HIV and TB prevalence:
741 findings from 4961 suspected cases. *BMC Infectious Diseases*, 10(1), p.67.

742 Kanehisa, M. et al., 2016. KEGG as a reference resource for gene and protein annotation.
743 *Nucleic acids research*, 44(D1), pp.D457–62.

744 Khayhan, K. et al., 2013. Geographically Structured Populations of Cryptococcus
745 neoformans Variety grubii in Asia Correlate with HIV Status and Show a Clonal
746 Population Structure O. Zaragoza, ed. *PLoS ONE*, 8(9), p.e72222.

747 Kondratieva, T. et al., 2014. Latent tuberculosis infection: What we know about its
748 genetic control? *Tuberculosis*, 94(5), pp.462–468.

749 Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
750 MEM.

751 Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*,
752 25(16), pp.2078–2079.

753 Litvintseva, A.P., 2005. Multilocus Sequence Typing Reveals Three Genetic
754 Subpopulations of Cryptococcus neoformans var. grubii (Serotype A), Including a
755 Unique Population in Botswana. *Genetics*, 172(4), pp.2223–2238.

756 Litvintseva, A.P. et al., 2011. Evidence that the Human Pathogenic Fungus Cryptococcus
757 neoformans var. grubii May Have Evolved in Africa K. Nielsen, ed. *PLoS ONE*, 6(5),
758 p.e19688.

759 Loftus, B.J. et al., 2005. The genome of the basidiomycetous yeast and human pathogen
760 Cryptococcus neoformans. *Science*, 307(5713), pp.1321–1324.

761 Longley, N. et al., 2008. Dose Response Effect of High-Dose Fluconazole for HIV-
762 Associated Cryptococcal Meningitis in Southwestern Uganda. *Clinical Infectious*
763 *Diseases*, 47(12), pp.1556–1561.

764 Maciá, M.D. et al., 2005. Hypermutation is a key factor in development of multiple-
765 antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung
766 infections. *Antimicrobial agents and chemotherapy*, 49(8), pp.3382–3386.

767 Magditch, D.A. et al., 2012. DNA Mutations Mediate Microevolution between Host-
768 Adapted Forms of the Pathogenic Fungus Cryptococcus neoformans R. C. May, ed.
769 *PLoS pathogens*, 8(10), p.e1002936.

770 McKenna, A. et al., 2010. The Genome Analysis Toolkit: A MapReduce framework for
771 analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), pp.1297–
772 1303.

773 Meyer, W. et al., 2009. Consensus multi-locus sequence typing scheme for Cryptococcus
774 neoformans and Cryptococcus gattii. *Medical Mycology*, 47(6), pp.561–570.

775 Morrow, C.A. & Fraser, J.A., 2013. Ploidy variation as an adaptive mechanism in human
776 pathogenic fungi. *Seminars in Cell and Developmental Biology*, 24(4), pp.339–346.

777 Mu, J. et al., 2010. Plasmodium falciparum genome-wide scans for positive selection,
778 recombination hot spots and resistance to antimalarial drugs. *Nature genetics*, 42(3),
779 pp.268–271.

780 Oliver, A. et al., 2000. High frequency of hypermutable Pseudomonas aeruginosa in
781 cystic fibrosis lung infection. *Science*, 288(5469), pp.1251–1254.

782 Ormerod, K.L. et al., 2013. Comparative Genomics of Serial Isolates of Cryptococcus
783 neoformans Reveals Gene Associated with Carbon Utilization and Virulence. *G3:*
784 *Genes| Genomes| Genetics*, pp.1–12.

785 Panackal, A.A. et al., 2006. Clinical significance of azole antifungal drug cross-resistance
786 in Candida glabrata. *Journal of clinical microbiology*, 44(5), pp.1740–1743.

787 Park, B.J. et al., 2009. Estimation of the current global burden of cryptococcal meningitis
788 among persons living with HIV/AIDS. *AIDS*, 23(4), pp.525–530.

789 Paul, S., Doering, T.L. & Moye-Rowley, W.S., 2015. Cryptococcus neoformans Yap1 is
790 required for normal fluconazole and oxidative stress resistance. *Fungal Genetics and*
791 *Biology*, 74, pp.1–9.

792 Pfaller, M.A., 2012. Antifungal drug resistance: mechanisms, epidemiology, and
793 consequences for treatment. *The American journal of medicine*, 125(1 Suppl),
794 pp.S3–13.

795 Rhodes, J., Beale, M.A. & Fisher, M.C., 2014. Illuminating Choices for Library Prep: A
796 Comparison of Library Preparation Methods for Whole Genome Sequencing of
797 Cryptococcus neoformans Using Illumina HiSeq K. Nielsen, ed. *PLoS ONE*, 9(11),
798 p.e113501.

799 Rodero, L. et al., 2003. G484S amino acid substitution in lanosterol 14-alpha
800 demethylase (ERG11) is related to fluconazole resistance in a recurrent

Cryptococcus neoformans clinical isolate. *Antimicrobial agents and chemotherapy*, 47(11), pp.3653–3656.

Rodríguez-Tudela, J.L. et al., 2010. EUCAST breakpoints for antifungals. *Drug news & perspectives*, 23(2), pp.93–97.

Selmecki, A., Forche, A. & Berman, J., 2006. Aneuploidy and isochromosome formation in drug-resistant Candida albicans. *Science*, 313(5785), pp.367–370.

Selmecki, A.M. et al., 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. H. D. Madhani, ed. *PLoS Genetics*, 5(10), p.e1000705.

Sionov, E. et al., 2010. Cryptococcus neoformans Overcomes Stress of Azole Drugs by Formation of Disomy in Specific Multiple Chromosomes S. G. Filler, ed. *PLoS pathogens*, 6(4), p.e1000848.

Sionov, E., Chang, Y.C. & Kwon-Chung, K.J., 2013. Azole Heteroresistance in Cryptococcus neoformans: Emergence of Resistant Clones with Chromosomal Disomy in the Mouse Brain during Fluconazole Treatment. *Antimicrobial agents and chemotherapy*, 57(10), pp.5127–5130.

Spitzer, E.D. et al., 1993. Persistence of initial infection in recurrent Cryptococcus neoformans meningitis. *Lancet*, 341(8845), pp.595–596.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), pp.2688–2690.

Strand, M. et al., 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *NATURE-LONDON-*, 365(6443), pp.274–276.

Van der Auwera, G.A. et al., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, 11(1110), pp.11.10.1–11.10.33.

Vanhove, M. et al., 2016. Genomic epidemiology of Cryptococcusyeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. *Molecular Ecology*, pp.1–34.

Figure legends

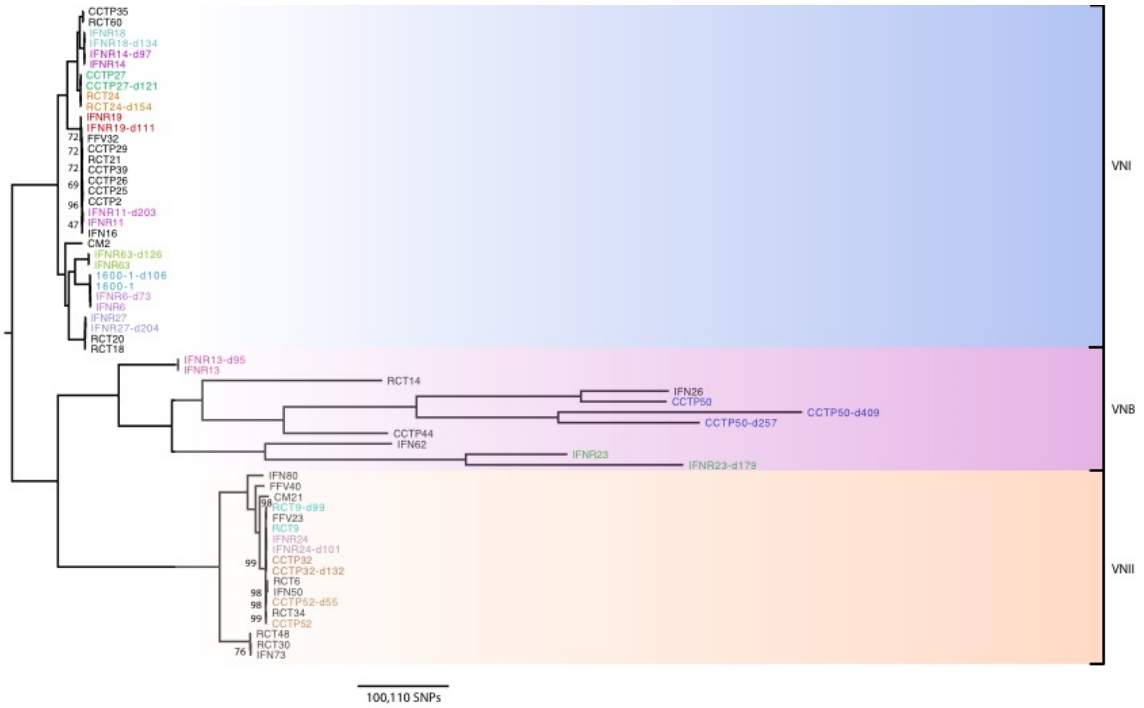
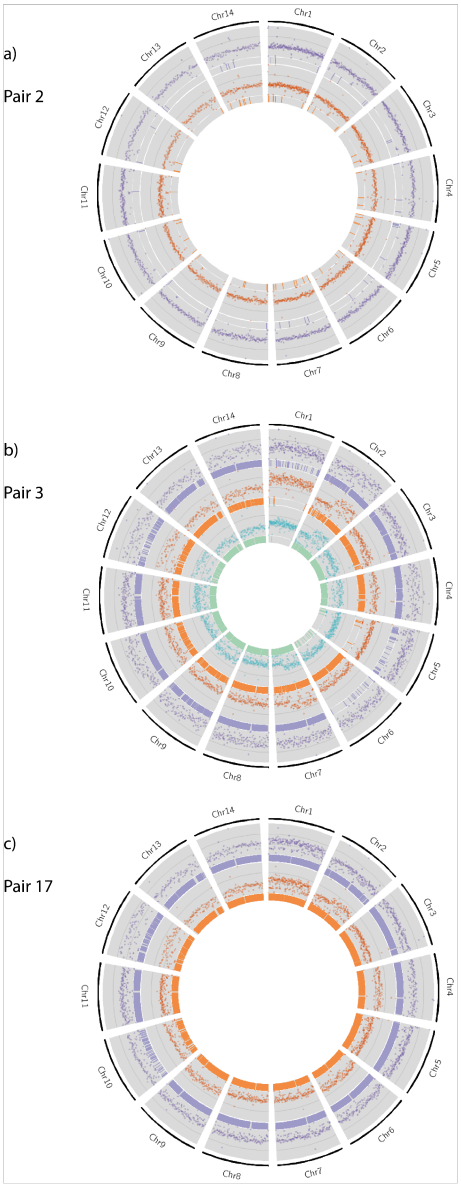


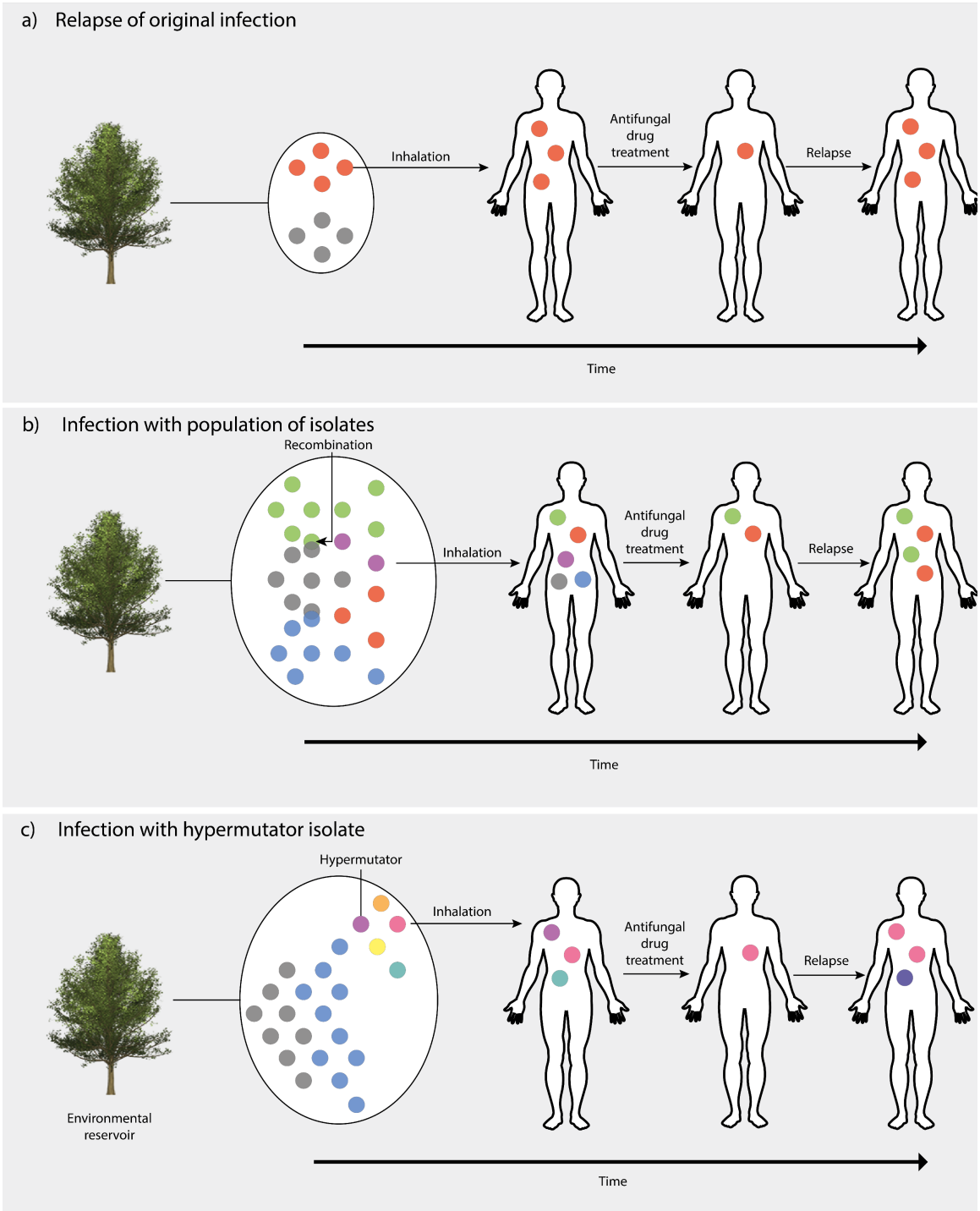
Figure 1 – Phylogenetic analysis of *C. neoformans* var. *grubii* isolates in this study (coloured), with additional isolates (shown in black) added to distinguish true relapse infections, or recurrent infections, and associated lineages. We hypothesise that isolates resulting from true relapse infections 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 to a particularly clonal VNI clade, including Pair 15 only, which only had 47% branch support). Branch lengths represent the number of SNPs between taxa.



844

845 Figure 2 – Extensive chromosomal copy number variation was observed in all isolates in
846 Pairs 3 and 17, when compared to H99. Pair 2 is included to illustrate isolates without
847 ploidy and extensive nsSNPs. Here, normalised whole-genome depth of coverage is
848 shown, averaged over 10,000-bp bins, in scatter plots. Bar plots represent the position
849 of nsSNPs. The purple track represents the original isolate, orange the recurrent isolate,
850 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 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 chromosome 6 being the exception: very few nsSNPs are located in chromosome 6 in CCTP50 and CCTP50-d409, whereas over 2000 nsSNPs are observed in chromosome 6 in CCTP50-d257. **c)** a gain in ploidy is observed for Chromsomes 2, 4, 6 and 9 compared to the Day 1 isolate in Pair 17, whereas ploidy remains unchanged for Chromosomes 1 and 12.


a) Relapse of original infection
Inhalation
Antifungal drug treatment
Relapse
Time
b) Infection with population of isolates
Recombination
Inhalation
Antifungal drug treatment
Relapse
Time
c) Infection with hypermutator isolate
Hypermutator
Inhalation
Antifungal drug treatment
Relapse
Environmental reservoir
Time

859

860 Figure 3 – Hypotheses of routes of infection of the human host by *C. neoformans* var.
861 *grubii*. **a)** Inhalation of a single population of basidiospores into a new host. Due to low
862 within-host diversity and being drug naïve, there will be a bottleneck in population size
863 due to antifungal drug treatment. However, if the initial drug regimen is insufficient to
864 sterile the CSF, resistance may develop on FLC maintenance therapy due to selection
865 pressure, resulting in relapsed infection from proliferation of a drug-resistant isolate. **b)**
866 VNB lineage *C. neoformans* exists the environment as a population, which can undergo
867 recombination to produce genetically similar isolates, but with significantly diversity.
868 Due to transmission bottlenecks, only a sample of the pathogen diversity will be
869 transferred to the host, in this case, by inhalation, but it is possible for a population of *C.*
870 *neoformans* to infect a single immunocompromised individual. Some isolates may be
871 susceptible to antifungal drugs and are thus becoming removed, whilst other isolates
872 may be inherently resistant and hence cause a relapse infection. **c)** Mutations in the
873 DNA mismatch repair gene *MSH2* cause an isolate to become a hypermutator. Some
874 genotypes may be susceptible to antifungal drugs, but the high mutation rate allows the
875 infection to adapt rapidly to the host and evolve drug resistance. These genotype
876 proliferate in the host, thus causing relapse infection.

877 **Tables**

878 Table 1: Details of *C. neoformans* isolates and MICs (if available) at time of isolation from
879 South Africa and Uganda (Pair 7 only) used in this study. AmB = amphotericin B 1

880 mg/kg/d, as per hospital guidelines at that time, unless otherwise stated; VOR =

881 voriconazole 300 mg/d; 5FC = flucytosine; FLC = fluconazole.

Pair #	Isolate ID	Day of isolation	Fluconazole MIC at isolation	Treatment of CM episode (if known)
1	CCTP27	1	4	AmB 3 days
1	CCTP27-d121	121	64	VOR until CD4 > 200 cells/uL
2	CCTP32	1	4	AmB 7 days
2	CCTP32-d132	132	6	AmB 7 days
3	CCTP50	1	16	AmB 14 days
3	CCTP50-d257	257	256	AmB 14 days
3	CCTP50-d409	409	n/a	AmB 7 days
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 days
5	RCT9-d99	99	n/a	AmB until death
6	RCT24	1	4	AmB plus 5FC for 14

				days
6	RCT24-d154	154	12	FLC 800 mg/d
7	1600-1	1	n/a	FLC 16000 mg/d for 14 days
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 days
9	IFNR24-d101	101	n/a	
10	IFNR18	1	n/a	AmB 7 days
10	IFNR18-d134	134	n/a	
11	IFNR14	1	n/a	AmB 7 days
11	IFNR14-d97	97	n/a	
12	IFNR13	1	n/a	AmB 7 days
12	IFNR13-d95	95	1	
13	IFNR6	1	n/a	AmB 7 days
13	IFNR6-d73	73	1	
14	IFNR19	1	n/a	AmB 7 days
14	IFNR19-d111	111	n/a	
15	IFNR11	1	n/a	AmB 7 days
15	IFNR11-d203	203	12	

16	IFNR27	1	n/a	AmB 7 days
16	IFNR27-d204	204	n/a	
17	IFNR23	1	n/a	AmB 7 days
17	IFNR23-d179	179	n/a	

882

883 Table 2: A high number of shared SNPs in most pairs indicate a shared common ancestor.

884 Number of SNPs common to both initial and recurrent infection, along with number of

885 SNPs and non-synonymous SNPs unique to each timepoint. Percentages given to 2 d.p.

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	13490 (98.49)	97 (0.71)	9 (8)	110 (0.81)	19 (8)		
2	289557 (99.29)	1080 (0.37)	110 (55)	991 (0.34)	108 (45)		
3	261718 (46.77)	127631 (32.78)	25331 (2741)	45769 (14.88)	8342 (2099)	124498 (32.24)	23957 (2541)
4	289574	1169 (0.40)	145 (58)	1096 (0.38)	145 (56)		

	(99.22)						
5	289367 (99.20)	1261 (0.43)	127 (54)	1080 (0.37)	127 (59)		
6	13122 (95.78)	133 (1.00)	5 (2)	445 (3.28)	82 (56)		
7	47833 (97.83)	522 (1.08)	81 (65)	537 (1.11)	94 (72)		
8	46883 (98.76)	294 (0.62)	29 (9)	296 (0.63)	26 (16)		
9	289458 (99.27)	1033 (0.36)	141 (58)	1109 (0.38)	105 (55)		
10	12687 (98.29)	104 (0.81)	4 (3)	117 (0.91)	11 (5)		
11	12732 (98.38)	110 (0.86)	10 (2)	99 (0.77)	6 (5)		
12	221973 (99.37)	636 (0.29)	55 (27)	778 (0.35)	61 (30)		
13	48062 (98.58)	396 (0.82)	31 (11)	294 (0.61)	28 (15)		
14	28960 (95.66)	288 (0.98)	44 (19)	1026 (3.42)	105 (57)		
15	29736 (98.40)	254 (0.85)	21 (14)	228 (0.76)	17 (13)		
16	45437 (98.56)	325 (0.71)	30 (18)	341 (0.74)	30 (14)		
17	376568 (56.43)	117436 (23.77)	22567 (4153)	173359 (31.52)	33543 (4427)		

886

887 **Supplementary material**

888 Table S1 - MLST results of independently testing three colonies per study isolate.

889 Table S2 - Number of SNPs in *MSH2* in each isolate, categorised by type or location. SYN
890 = synonymous SNP, NSY = non-synonymous SNP, p5UTR = 5' UTR, p3UTR = 3' UTR,
891 NON = nonsense mutation.

892 Table S3 - Non-synonymous SNPs in Pair 3 isolates, per chromosome

893 Table S4 - Number of non-synonymous and synonymous SNPs per isolate

894 Table S5 - Details of alignment and variant calling for all paired isolates included in this
895 study

896 Table S6 - Number of homozygous and heterozygous SNPs in isolates of Pairs suspected
897 of diploidy.

898 Figure S1 – Phylogenetic analysis each chromosome of *C. neoformansi* isolates in this
899 study (coloured) with additional isolates (shown in black), to identify whether Pair 3
900 consists of different isolates or is a relapsed infection. We hypothesised that if Pair 3
901 was a relapsed infection, all isolates would share the same phylogenetic relationship in
902 all 14 chromosomes. Rapid bootstrap analysis over 250 replicates was performed on
903 SNP data from 62 isolates to generate an unrooted maximum-likelihood phylogeny (only
904 branches not supported to 100% are indicated above branches). Branch lengths
905 represent the number of SNPs between taxa.

906 Figure S2 – Chromosomal copy number variation can be observed in some isolates in
907 Pairs 1, 4, 5, 10, 14 and 15, when compared to the *C. neoformans* reference genome,
908 H99. Normalised whole-genome depth of coverage is illustrated here, averaged over

909 10,000-bp bins and represented as a scatter plot. Bar plots represent the position of
910 nsSNPs, where the purple track represents the original isolate, and orange represents
911 the recurrent isolate. Note that Pairs 2, 3 and 17 are not present in this Figure, as they
912 are represented in Figure 2.