

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

Johanna Rhodes¹, Mathew A. Beale², Mathieu Vanhove¹, Joseph N. Jarvis^{3,4,5}, Shichina
Kannambath⁶, John A. Simpson⁷, Anthea Ryan⁷, Graeme Meintjes⁸, Thomas S. Harrison⁶,
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 *Cryptococcus neoformans* and

represent a novel pathway for rapid within-host adaptation and evolution of resistance to firstline 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* (*Cn*) (1). This mycosis poses a considerable public health problem in sub-Saharan Africa, which has the highest estimated annual incidence of cryptococcal meningitis (CM) globally (2), with the majority of infections caused by *Cryptococcus neoformans sensu stricto* (previously referred to as *Cryptococcus neoformans* var. *grubii*) (3).

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 (4). Microevolution occurs in response to drug pressure, leading to resistance, a phenomenon previously described in *Cn* (5, 6). 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 (7).

Various methods of within-host evolution are available to eukaryotic pathogens, most notably sexual and parasexual reproduction, although these are difficult to observe in fungi due to being cryptic. Aneuploidy, recombination in the telomeres, and mutator states (8) also provide means of rapid within-host evolution, with other mechanisms still likely to be discovered. The accumulation of SNPs alongside copy number variation and aneuploidy has been witnessed during infection in different fungal pathogen species, enabling rapid adaptive evolution (9) and conferring resistance to antifungal drugs (10). Candidiasis is caused by numerous *Candida* species, yet within-host evolution to these species differs. Infection by *Candida albicans* is usually susceptible to azole antifungal drugs, but resistance can evolve via the acquisition of drug-resistant aneuploid isolates, which contain an isochromosome of the left arm of chromosome 5 (11). 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, *Candida glabrata* are intrinsically poorly susceptible to azoles, and have more recently evolved multi-drug resistance to both azoles and echinocandins (12-14).

Within-host diversity and recombination has been witnessed in eukaryotic pathogens, notably *C. albicans*: mutation and recombination rates can be increased under stressful conditions, such as drug treatment (15, 16), resulting in loss-of-heterozygosity (LOH) and aneuploidy (15). These genetic alterations contribute to the maintenance of a population of *C. albicans* within the host environment (15), and drug pressure can result in diverging levels of fitness (17).

Similar responses to antifungal drugs have been observed in *Cn*; Point mutations in the *Cn* ortholog *ERG11* were also shown to confer fluconazole resistance, by causing the amino acid substitution G484S (8). Sionov *et al.* (6) demonstrated large scale chromosomal duplications (primarily chromosome 1) are fundamental to overcoming fluconazole (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 (6), although other genes are also thought to be involved in FLC resistance (18, 19). Previous studies of serially collected *Cn* isolates have confirmed in-host microevolution, including the occurrence of large-scale genomic rearrangements (20-22). Like *C. albicans*, the *Cn* genome is capable of undergoing chromosomal duplication and loss under stresses such as drug pressure or invasion of the human host (23). These chromosomal duplications are often lost when the selective pressure is removed (6).

The development of mutator states via hypermutability is a rapidly expanding area of study in bacteria, particularly *Pseudomonas aeruginosa* in cystic fibrosis (CF) patients. Here, hypermutability has been shown to have an association with antimicrobial resistance (24, 25), causing significant implications in the early treatment of cystic fibrosis patients to prevent chronic infection (25, 26). 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 (27). Frameshift mutations in an ortholog of the mismatch repair gene *MSH2*

have also been shown to contribute to microevolution in the sister species of *Cn*,
Cryptococcus gattii (28).

Here we describe a comparative genome-sequencing based approach to investigate
microevolution in serially collected isolates of *Cn*. 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 (29-33). Ethical approval was obtained from the Wandsworth
Research Ethics Committee covering St. George's University of London (29-32). 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/ μ L. 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 subset of the study isolates, according to the methods outlines in Meyer *et al.* (34), with modifications as outlined in Beale *et al.* (35).

Molecular methods

Cn was isolated from HIV infected individuals on location by plating CSF onto Sabourand Dextrose (SD) agar (Oxoid, Fisher Scientific), and growing at 30°C for 48 hours. A representative sample of the *Cn* 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°C until further testing. This approach ensures all genetic

diversity is maintained through the process, and single colony picking only occurs at the final stage of liquid culture and DNA extraction.

Cn isolates from patient CSF were stored in 20% glycerol at -80°C until ready for testing. Frozen stocks were plated onto SD agar and cultured for 72 hours. A single colony was inoculated into 6ml Yeast Peptone Digest broth (Oxoid) supplemented with 0.5M NaCl and cultured at 37°C with agitation (165 rpm) for 40 hours, followed by genomic DNA extraction using the Masterpure Yeast DNA purification kit (Epicentre) modified by addition of two cycles of rapid bead beating (45 seconds at 4.5 m/second) using a FastPrep 24 homogeniser (MP Bio). Genomic DNA libraries were prepared using the TruSeq DNA v2 or TruSeq Nano DNA kit (Illumina), and whole genome sequencing was performed on an Illumina HiSeq 2500 at Medical Research Council Clinical Genomics Centre (Imperial College London) as previously described (36). 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 *Cn* reference genome H99 (37) using the Burrows-Wheeler Aligner (BWA) v0.75a mem algorithm (38) with default parameters to obtain high depth alignments (average 104x). Samtools (39) 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 (40) version 3.4-46
‘RealignerTargetCreator’ and ‘IndelRealigner’, following best practice guidelines (41).
Single nucleotide polymorphisms (SNPs) and INDELs were called from all alignments
using GATK (40) 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 *Cn* 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, Cambridge, MA, USA) and the latest release
(CNA3) of the *Cn* reference genome H99 and gene ontology.
Some isolates were suspected of having non-haploid 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 (40) version
3.4-46 ‘DepthOfCoverage’ under default settings. The *Cn* H99 (37) was again used as
reference. In order to determine aneuploidy, whole-genome coverage data was
normalised and regions displaying normalised coverage equal to 2 were deemed diploid
events (likewise, normalised coverage equal to 3 were deemed triploid events, and so
on), whereas normalised coverage equal to zero was deemed a deletion event.
Susceptibility testing

The susceptibility testing of all relapse isolates were 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, fluconazole, voriconazole, posaconazole, itraconazole, micafungin, anidulafungin and caspofungin, and commercialises 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. 1:20 dilution was prepared 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 (42, 43).

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 (44) as described in Abdolrasouli *et al.* (45): 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 *Cn* 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 (46) pathway source for *Cn* H99, using a p -value cut-off of $p < 0.05$.

Identifying sites under selection

BayeScan 2.01 (47) 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) (48). 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/ μ L. 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 *Cn*

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 *Cn* H99 reference genome (37), 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).

A high level of diversity was observed within the VNB lineage, resulting in long branch lengths (Figure 1). Although all VNB isolates were mapped to the *Cn* H99 reference (37), which is a VNI lineage isolate, we do not believe SNP numbers observed in the VNB lineage are a consequence of phylogenetic distance to the reference genome. This is because SNP determination revealed only 21.6% SNPs were common to the three VNB Pairs included in this study (Pairs 3, 12 and 17), highlighting the large amounts of genetic diversity seen within this lineage.

Phylogenetic analysis showed that of the 17 pairs of relapse isolates, three pairs were lineage VNB, whilst four and ten belonged to lineages VNII and VNI respectively (Figure 1). The average pairwise SNP diversity was far higher amongst 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 more than 97% of SNPs were in common between original and recurrent isolates. The majority of pairs had >99% SNP similarity (Table 3) 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 of Pairs 1 and 7 (which had previously demonstrated a potential 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/day for CCTP50-d257 and CCTP50-d409 respectively (Pair 3), 8 SNPs/day for Pair 12, and 968 SNPs/day for Pair 17. Due to the variation in SNP accumulation between Pair 12 and Pairs 3 and 17, we hypothesised that Pair 12 was a true relapse of the original infection, whilst Pairs 3 and 17 are showing inflated SNP numbers due to reinfection or an anomalous rate of evolution.

Antifungal susceptibility testing

Fluconazole susceptibility testing (Table 1) using the Etest® (bioMerieux) 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 Pair6; IFNR11-d203 in Pair 15) had MICs above the established epidemiological cut-off value ($\geq 8\mu\text{g/ml}$) 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 $\mu\text{g/ml}$, whilst the recurrent isolate was resistant at an MIC of 64 $\mu\text{g/ml}$; in Pair 3 (CCTP50), the initial isolate MIC was 16 $\mu\text{g/ml}$ (intermediate), whilst 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 *Cn* 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 3). 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 3). 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 *Cn* 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 *Cn* var. *grubii* 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.

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, whilst 14 SNPs were common to all VNII pairs at the point of recurrent infection, five of which were intergenic. The remaining 9 SNPs were located in the 5' untranslated region (UTR) gene *SMF1* (CNAG_05640), a metal ion transporter with a natural resistance-associated macrophage protein. Selection analysis indicated that this gene was not under selection pressure, however.

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 Methods). No putative loci under either diversifying or balancing selection could be detected using a false discovery rate (FDR) of 0.05. F_{ST} values were limited to not exceed 3.47×10^{-5} .

Aneuploidy as a generator of diversity in recurrent infection

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

Ormerod *et al.* (5) 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 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.* (5). 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 *Cn* 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 (SREBP) (49). An alcohol dehydrogenase (*GNO1* – CNAG_06168) was also present, which is thought to be involved in the defence against host response (50). 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 aneuploidies 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* (51). 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 *Cn* is usually due to the persistence and recurrence of the original infecting isolate (52), 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 (53). 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 relapses of infection of the same genotype, which differ by only a few SNPs, whilst 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 *Cn* incurs numerous unique small and large-scale changes during infection, and that a subset of these may have adaptive

value. Whilst this study is concerned with the genomics of recurrent infections by identifying SNP changes and ploidy potentially involved in the persistence of *Cn* infection, future work should investigate the potential role of gene expression changes and gene networks involved in changes in fitness amongst 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 (54) and differentiating between dormancy and reactivation (55) respectively. Such approaches are likely to increase our understanding of clinical cases of recurrent *Cn* infections through identifying the genetic basis of phenotypic switching (56) 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. 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 (57). 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.

The occurrence of aneuploidy 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 (6, 58). Sionov *et al.* (6) 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 *Cn* chromosome 1 included *ERG11* and *AFR1*, which are both transporters of azole drugs. Whilst 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.* suggested that *ERG11* contributed to the duplication of chromosome 1 (6); we observed only one isolate (IFNR11 of Pair 15) displaying a duplication of chromosome 1,

but a single copy of *ERG11*, suggesting ploidy was not complete throughout the chromosome. Since this isolate was the initial infection, we can assume that the duplication of chromosome 1 was not solely due to 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.* (5) showed an aneuploidy (duplication) in chromosome 12 between serially collected isolates. Four pairs included in this study (Pair 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 (6), one can assume that this aneuploidy was maintained due to previous drug exposure potentially not reported by the patient, or that aneuploidy helps *Cn* adapt to the host environment (59). Chromosome 12 experienced triploidy in the Pair 1 recurrent isolate (CCTP27-d121); this pair also demonstrated drug resistance to fluconazole, with a FLC MIC of 4 at initial infection, and a FLC MIC of 64 at recurrent infection. Ormerod *et al.* (5) hypothesise that the large number of genes affected by the increased copy number of chromosome 12 contributes to metabolome differences; however, we hypothesise copy number variation 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 *Cn* virulence.

Antimicrobial drugs impose strong selection pressure on pathogens, which may lead to the evolution of drug resistance (60, 61); there are, however, fitness costs associated with the evolution of resistance to antifungal drugs that may impact fitness (17).

Genome-wide scans for sites under selection leads to the identification of possible sites of drug resistance. We did not identify any significant sites when comparing VNI original infection versus recurrent infection, and the number of VNB and VNII isolates were too low for analysis. Whilst 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 amongst individuals due to stochasticity and clonal interference (61). It is also likely that as there is little recombination in VNI isolates compared to VNB and VNII isolates (62-64) and therefore linkage is complete across the genome, further hampering selection analysis. We therefore 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. It is therefore important for clinicians to request susceptibility testing in real time, at the very least in all cases or recurrent CM.

Whilst Pair 17 did not exhibit a high percentage of common SNPs between the original and recurrent isolates indicative of a relapse infection, the elevated rate of SNPs observed in all chromosomes of both isolates suggested this was not a re-infection 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*. Whilst previous studies have shown hypermutator phenotypes aid adaptation to stress (65), and we hypothesise that hypermutation may lead to adaptation of drug resistance under the stress of antifungal treatment. These results are the first to the authors' knowledge to report on nonsense mutations in *MSH2*, *RAD5* and *MSH5* in *Cn*. 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 occur 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 *Cn* 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 *Cn*, 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 *Cn* by imposing a high genetic load.

Acknowledgements

Special thanks to Mr. Ian Wright for his financial contribution to this project, and Dr. Winnie Wu and StarLab UK for supplying filter pipette tips free of charge. This work was funded by a Medical Research Council grant (MRC MR/K000373/1) awarded to MCF and a grant from Imperial College London (WPIA/F24085) awarded to JR. JJ was supported by the Wellcome Trust (training fellowship, WT081794) and is supported by a grant from the Penn Center for AIDS Research (CFAR), an NIH-funded program (P30 AI 045008). Thankyou to Sean Lobo and Ahsan Awais Bokhair for initial work on MICs and MLST, respectively.

Author contributions

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

References

1. **Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, Falk R, Parnmen S, Lumbsch HT, Boekhout T.** 2015. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. *Fungal Genetics and Biology* **78**:16–48.
2. **Park BJ, Wannemuehler KA, Marston BJ, Govender N, Pappas PG, Chiller TM.** 2009. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS* **23**:525–530.
3. **Jarvis JN, Harrison TS.** 2007. HIV-associated cryptococcal meningitis. *AIDS* **21**:2119–2129.
4. **Brouwer AE, Rajanuwong A, Chierakul W, Griffin GE, Larsen RA, White NJ, Harrison TS.** 2004. Combination antifungal therapies for HIV-associated cryptococcal meningitis: a randomised trial. *Lancet* **363**:1764–1767.
5. **Ormerod KL, Morrow CA, Chow EW, Lee IR, Arras SD, Schirra HJ, Cox GM, Fries BC, Fraser JA.** 2013. Comparative Genomics of Serial Isolates of *Cryptococcus neoformans* Reveals Gene Associated with Carbon Utilization and Virulence. *G3 (Bethesda)* 1–12.
6. **Sionov E, Lee H, Chang YC, Kwon-Chung KJ.** 2010. *Cryptococcus neoformans* Overcomes Stress of Azole Drugs by Formation of Disomy in Specific Multiple Chromosomes. *PLoS Pathog* **6**:e1000848.
7. **Bozzette SA, Larsen RA, Chiu J, Leal MA, Jacobsen J, Rothman P, Robinson P, Gilbert G, McCutchan JA, Tilles J.** 1991. A placebo-controlled trial of maintenance therapy with fluconazole after treatment of cryptococcal meningitis in the acquired immunodeficiency syndrome. California Collaborative Treatment Group. *N Engl J Med* **324**:580–584.
8. **Rodero L, Mellado E, Rodriguez AC, Salve A, Guelfand L, Cahn P, Cuenca-Estrella M, Davel G, Rodríguez-Tudela JL.** 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**:3653–3656.
9. **Calo S, Billmyre RB, Heitman J.** 2013. Generators of Phenotypic Diversity in the Evolution of Pathogenic Microorganisms. *PLoS Pathog* **9**:e1003181–4.
10. **Hickman MA, Paulson C, Dudley A, Berman J.** 2015. Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in *Candida albicans*. *Genetics* **200**:781–794.

- 665 11. **Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J.** 2009. Acquisition of
666 aneuploidy provides increased fitness during the evolution of antifungal drug
667 resistance. *PLoS Genet* **5**:e1000705.
- 668 12. **Pfaller MA.** 2012. Antifungal drug resistance: mechanisms, epidemiology, and
669 consequences for treatment. *Am J Med* **125**:S3–13.
- 670 13. **Panackal AA, Gribskov JL, Staab JF, Kirby KA, Rinaldi M, Marr KA.** 2006. Clinical
671 significance of azole antifungal drug cross-resistance in *Candida glabrata*. *J Clin*
672 *Microbiol* **44**:1740–1743.
- 673 14. **Alexander BD, Johnson MD, Pfeiffer CD, Jimenez-Ortigosa C, Catania J, Booker R,**
674 **Castanheira M, Messer SA, Perlin DS, Pfaller MA.** 2013. Increasing echinocandin
675 resistance in *Candida glabrata*: clinical failure correlates with presence of FKS
676 mutations and elevated minimum inhibitory concentrations. *CLIN INFECT DIS*
677 **56**:1724–1732.
- 678 15. **Forche A, Magee PT, Selmecki A, Berman J, May G.** 2009. Evolution in *Candida*
679 *albicans* Populations During a Single Passage Through a Mouse Host. *Genetics*
680 **182**:799–811.
- 681 16. **Ford CB, Funt JM, Abbey D, Issi L, Guiducci C, Martinez DA, Delorey T, Li BY,**
682 **White TC, Cuomo C, Rao RP, Berman J, Thompson DA, Regev A.** 2015. The
683 evolution of drug resistance in clinical isolates of *Candida albicans*. *eLife* **4**:e00662.
- 684 17. **Cowen LE, Kohn LM, Anderson JB.** 2001. Divergence in fitness and evolution of
685 drug resistance in experimental populations of *Candida albicans*. *J Bacteriol*
686 **183**:2971–2978.
- 687 18. **Sionov E, Chang YC, Kwon-Chung KJ.** 2013. Azole Heteroresistance in
688 *Cryptococcus neoformans*: Emergence of Resistant Clones with Chromosomal
689 Disomy in the Mouse Brain during Fluconazole Treatment. *Antimicrob Agents*
690 *Chemother* **57**:5127–5130.
- 691 19. **Paul S, Doering TL, Moye-Rowley WS.** 2015. *Cryptococcus neoformans* Yap1 is
692 required for normal fluconazole and oxidative stress resistance. *Fungal Genetics*
693 *and Biology* **74**:1–9.
- 694 20. **Fraser JA, Huang JC, Pukkila-Worley R, Alspaugh JA, Mitchell TG, Heitman J.**
695 2005. Chromosomal translocation and segmental duplication in *Cryptococcus*
696 *neoformans*. *Eukaryotic Cell* **4**:401–406.
- 697 21. **Blasi E, Brozzetti A, Francisci D, Neglia R, Cardinali G, Bistoni F, Vidotto V,**
698 **Baldelli F.** 2001. Evidence of microevolution in a clinical case of recurrent
699 *Cryptococcus neoformans* meningoencephalitis. *Eur J Clin Microbiol Infect Dis*
700 **20**:535–543.

- 701 22. **Illnait-Zaragozi MT, Martínez-Machín GF, Fernández-Andreu CM, Hagen F,**
702 **Boekhout T, Klaassen CH, Meis JF.** 2010. Microsatellite typing and susceptibilities
703 of serial *Cryptococcus neoformans* isolates from Cuban patients with recurrent
704 cryptococcal meningitis. BMC Infectious Diseases **10**:289.
- 705 23. **Fries BC, Chen F, Currie BP, Casadevall A.** 1996. Karyotype instability in
706 *Cryptococcus neoformans* infection. J Clin Microbiol **34**:1531–1534.
- 707 24. **Oliver A, Cantón R, Campo P, Baquero F, Blázquez J.** 2000. High frequency of
708 hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science
709 **288**:1251–1254.
- 710 25. **Maciá MD, Blanquer D, Togores B, Sauleda J, Pérez JL, Oliver A.** 2005.
711 Hypermutation is a key factor in development of multiple-antimicrobial resistance
712 in *Pseudomonas aeruginosa* strains causing chronic lung infections. Antimicrob
713 Agents Chemother **49**:3382–3386.
- 714 26. **Burns JL, Gibson RL, McNamara S, Yim D, Emerson J, Rosenfeld M, Hiatt P,**
715 **McCoy K, Castile R, Smith AL, Ramsey BW.** 2001. Longitudinal assessment of
716 *Pseudomonas aeruginosa* in young children with cystic fibrosis. J Infect Dis
717 **183**:444–452.
- 718 27. **Strand M, Prolla TA, Liskay RM, Petes TD.** 1993. Destabilization of tracts of
719 simple repetitive DNA in yeast by mutations affecting DNA mismatch repair.
720 Nature **365**:274–276.
- 721 28. **Billmyre RB, Croll D, Li W, Mieczkowski P, Carter DA, Cuomo CA, Kronstad JW,**
722 **Heitman J.** 2014. Highly Recombinant VGII *Cryptococcus gattii* Population
723 Develops Clonal Outbreak Clusters through both Sexual Macroevolution and
724 Asexual Microevolution. mBio **5**:e01494–14–e01494–14.
- 725 29. **Bicanic T, Meintjes G, Wood R, Hayes M, Rebe K, Bekker LG, Harrison T.** 2007.
726 Fungal burden, early fungicidal activity, and outcome in cryptococcal meningitis in
727 antiretroviral-naïve or antiretroviral-experienced patients treated with
728 amphotericin B or fluconazole. CLIN INFECT DIS **45**:76–80.
- 729 30. **Bicanic T, Wood R, Meintjes G, Rebe K, Brouwer A, Loyse A, Bekker LG, Jaffar S,**
730 **Harrison T.** 2008. High-Dose Amphotericin B with Flucytosine for the Treatment
731 of Cryptococcal Meningitis in HIV-Infected Patients: A Randomized Trial. CLIN
732 INFECT DIS **47**:123–130.
- 733 31. **Jarvis JN, Meintjes G, Rebe K, Williams GN, Bicanic T, Williams A, Schutz C,**
734 **Bekker LG, Wood R, Harrison TS.** 2012. Adjunctive interferon-γ immunotherapy
735 for the treatment of HIV-associated cryptococcal meningitis. AIDS **26**:1105–1113.
- 736 32. **Longley N, Muzoora C, Taseera K, Mwesigye J, Rwebembera J, Chakera A, Wall E,**

- 737 **Andia I, Jaffar S, Harrison TS.** 2008. Dose Response Effect of High-Dose
738 Fluconazole for HIV-Associated Cryptococcal Meningitis in Southwestern Uganda.
739 CLIN INFECT DIS **47**:1556–1561.
- 740 33. **Jarvis JN, Meintjes G, Williams A, Brown Y, Crede T, Harrison TS.** 2010. Adult
741 meningitis in a setting of high HIV and TB prevalence: findings from 4961
742 suspected cases. BMC Infectious Diseases **10**:67.
- 743 34. **Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, Esposto MC, Fisher**
744 **M, Gilgado F, Hagen F, Kaocharoen S, Litvintseva AP, Mitchell TG, Simwami SP,**
745 **Trilles L, Viviani MA, Kwon-Chung J.** 2009. Consensus multi-locus sequence
746 typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med Mycol
747 **47**:561–570.
- 748 35. **Beale MA, Sabiiti W, Robertson EJ, Fuentes-Cabrejo KM, O’Hanlon SJ, Jarvis JN,**
749 **Loyse A, Meintjes G, Harrison TS, May RC, Fisher MC, Bicanic T.** 2015. Genotypic
750 Diversity Is Associated with Clinical Outcome and Phenotype in *Cryptococcal*
751 *Meningitis* across Southern Africa. PLoS Negl Trop Dis **9**:e0003847.
- 752 36. **Rhodes J, Beale MA, Fisher MC.** 2014. Illuminating Choices for Library Prep: A
753 Comparison of Library Preparation Methods for Whole Genome Sequencing of
754 *Cryptococcus neoformans* Using Illumina HiSeq. PLoS ONE **9**:e113501.
- 755 37. **Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J,**
756 **Miranda M, Anderson IJ, Fraser JA, Allen JE, Bosdet IE, Brent MR, Chiu R,**
757 **Doering TL, Donlin MJ, D’Souza CA, Fox DS, Grinberg V, Fu J, Fukushima M, Haas**
758 **BJ, Huang JC, Janbon G, Jones SJM, Koo HL, Krzywinski MI, Kwon-Chung JK,**
759 **Lengeler KB, Maiti R, Marra MA, Marra RE, Mathewson CA, Mitchell TG, Pertea**
760 **M, Riggs FR, Salzberg SL, Schein JE, Shvartsbeyn A, Shin H, Shumway M, Specht**
761 **CA, Suh BB, Tenney A, Utterback TR, Wickes BL, Wortman JR, Wye NH, Kronstad**
762 **JW, Lodge JK, Heitman J, Davis RW, Fraser CM, Hyman RW.** 2005. The genome of
763 the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*.
764 Science **307**:1321–1324.
- 765 38. **Li H.** 2013. Aligning sequence reads, clone sequences and assembly contigs with
766 BWA-MEM.
- 767 39. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,**
768 **Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence
769 Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.
- 770 40. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A,**
771 **Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome
772 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
773 sequencing data. Genome Research **20**:1297–1303.

- 774 41. **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy**
775 **Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV,**
776 **Altshuler D, Gabriel S, DePristo MA.** 2013. From FastQ data to high confidence
777 variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc*
778 *Bioinformatics* **11**:11.10.1–11.10.33.

- 779 42. **Rodríguez-Tudela JL, Arendrup MC, Cuenca-Estrella M, Donnelly JP, Lass-Flörl C.**
780 2010. EUCAST breakpoints for antifungals. *Drug News Perspect* **23**:93–97.

- 781 43. **Alastruey-Izquierdo A, Cuenca-Estrella M.** 2012. EUCAST and CLSI: How to Assess
782 in Vitro Susceptibility and Clinical Resistance. *Curr Fungal Infect Rep* **6**:229–234.

- 783 44. **Stamatakis A.** 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic
784 analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–2690.

- 785 45. **Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, Chowdhary A, Meis JF,**
786 **Armstrong-James D, Fisher MC.** 2015. Genomic Context of Azole Resistance
787 Mutations in *Aspergillus fumigatus* Determined Using Whole-Genome Sequencing.
788 *mBio* **6**:e00536–15.

- 789 46. **Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M.** 2016. KEGG as a
790 reference resource for gene and protein annotation. *Nucleic Acids Res* **44**:D457–
791 62.

- 792 47. **Foll M, Gaggiotti O.** 2008. A genome-scan method to identify selected loci
793 appropriate for both dominant and codominant markers: a Bayesian perspective.
794 *Genetics* **180**:977–993.

- 795 48. **Excoffier L, Hofer T, Foll M.** 2009. Detecting loci under selection in a hierarchically
796 structured population. *Heredity* **103**:285–298.

- 797 49. **Chang YC, Ingavale SS, Bien C, Espenshade P, Kwon-Chung KJ.** 2009.
798 Conservation of the sterol regulatory element-binding protein pathway and its
799 pathobiological importance in *Cryptococcus neoformans*. *Eukaryotic Cell* **8**:1770–
800 1779.

- 801 50. **de Jesús-Berríos M, Liu L, Nussbaum JC, Cox GM, Stamler JS, Heitman J.** 2003.
802 Enzymes that counteract nitrosative stress promote fungal virulence. *Current*
803 *Biology* **13**:1963–1968.

- 804 51. **Drotschmann K, Clark AB, Kunkel TA.** 1999. Mutator phenotypes of common
805 polymorphisms and missense mutations in MSH2. *Current Biology* **9**:907–910.

- 806 52. **Spitzer ED, Spitzer SG, Freundlich LF, Casadevall A.** 1993. Persistence of initial
807 infection in recurrent *Cryptococcus neoformans* meningitis. *Lancet* **341**:595–596.

- 808 53. **Desnos-Ollivier M, Patel S, Spaulding AR, Charlier C, Garcia-Hermoso D, Nielsen**
809 **K, Dromer F.** 2010. Mixed Infections and In Vivo Evolution in the Human Fungal
810 Pathogen *Cryptococcus neoformans*. *mBio* **1**:e00091–10–e00091–10.
- 811 54. **Dufort MJ, Hanniffy S, Álvarez RC, Gorvel VA, Altman MC, Khaenam P, Solera J,**
812 **Gersuk VH, Bosilkovski M, Moriyón I, Linsley PS, Gorvel J-P.** 2016. Gene
813 expression profiles predict treatment outcomes in Brucellosis. *J Immunol*
814 **196**:66.21–66.21.
- 815 55. **Kondratieva T, Azhikina T, Nikonenko B, Kaprelyants A, Apt A.** 2014. Latent
816 tuberculosis infection: What we know about its genetic control? *Tuberculosis*
817 **94**:462–468.
- 818 56. **D’Souza CA, Heitman J.** 2001. It infects me, it infects me not: phenotypic
819 switching in the fungal pathogen *Cryptococcus neoformans*. *J Clin Invest*
820 **108**:1577–1578.
- 821 57. **Vanhove M, Beale MA, Rhodes J, Chanda D, Lakhi S, Kwenda G, Molloy S,**
822 **Karunaharan N, Stone N, Harrison TS, Bicanic T, Fisher MC.** 2016. Genomic
823 epidemiology of *Cryptococcus* yeasts identifies adaptation to environmental
824 niches underpinning infection across an African HIV/AIDS cohort. *Molecular*
825 *Ecology* 1–34.
- 826 58. **Selmecki A, Forche A, Berman J.** 2006. Aneuploidy and isochromosome formation
827 in drug-resistant *Candida albicans*. *Science* **313**:367–370.
- 828 59. **Morrow CA, Fraser JA.** 2013. Ploidy variation as an adaptive mechanism in human
829 pathogenic fungi. *Seminars in Cell and Developmental Biology* **24**:339–346.
- 830 60. **Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, Chotivanich K,**
831 **Wilairatana P, Krudsood S, White NJ, Udomsangpetch R, Cui L, Ho M, Ou F, Li H,**
832 **Song J, Li G, Wang X, Seila S, Sokunthea S, Socheat D, Sturdevant DE, Porcella SF,**
833 **Fairhurst RM, Wellems TE, Awadalla P, Su X-Z.** 2010. *Plasmodium falciparum*
834 genome-wide scans for positive selection, recombination hot spots and resistance
835 to antimalarial drugs. *Nat Genet* **42**:268–271.
- 836 61. **Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ.** 2016. Within-host evolution
837 of bacterial pathogens. *Nat Rev Microbiol* **14**:150–162.
- 838 62. **Khayhan K, Hagen F, Pan W, Simwami S, Fisher MC, Wahyuningsih R,**
839 **Chakrabarti A, Chowdhary A, Ikeda R, Taj-Aldeen SJ, Khan Z, Ip M, Imran D, Sjam**
840 **R, Sriburee P, Liao W, Chaicumpar K, Vuddhakul V, Meyer W, Trilles L, van Iersel**
841 **LJJ, Meis JF, Klaassen CHW, Boekhout T.** 2013. Geographically Structured
842 Populations of *Cryptococcus neoformans* Variety *grubii* in Asia Correlate with HIV
843 Status and Show a Clonal Population Structure. *PLoS ONE* **8**:e72222.

63. **Litvintseva AP, Carbone I, Rossouw J, Thakur R, Govender NP, Mitchell TG.** 2011. Evidence that the Human Pathogenic Fungus *Cryptococcus neoformans* var. *grubii* May Have Evolved in Africa. PLoS ONE **6**:e19688.
64. **Litvintseva AP.** 2005. Multilocus Sequence Typing Reveals Three Genetic Subpopulations of *Cryptococcus neoformans* var. *grubii* (Serotype A), Including a Unique Population in Botswana. Genetics **172**:2223–2238.
65. **Magditch DA, Liu T-B, Xue C, Idnurm A.** 2012. DNA Mutations Mediate Microevolution between Host-Adapted Forms of the Pathogenic Fungus *Cryptococcus neoformans*. PLoS Pathog **8**:e1002936.

Figure legends

Figure 1 – Phylogenetic analysis of *Cn* 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.

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, normalised 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 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 Chromosomes 2, 4, 6 and 9 compared to the Day 1 isolate in Pair 17, whereas ploidy remains unchanged for Chromosomes 1 and 12.

Figure 3 – Hypotheses of routes of infection of the human host by *Cn* 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 sterile the CSF, resistance may develop on FLC maintenance therapy due to selection pressure, resulting in relapsed infection from proliferation of a drug-resistant isolate. **b)** VNB lineage *Cn* exists the environment as a population, which 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 *Cn* to infect a single immunocompromised individual. Some isolates may be susceptible to antifungal drugs and are thus becoming removed, whilst 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.

Tables

Table 1: Details of *Cn* isolates and MICs (if available) at time of isolation from South Africa and Uganda (Pair 7 only) used in this study. AmB = amphotericin B 1 mg/kg/d, as per hospital guidelines at that time, unless otherwise stated; VOR = 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-	409	n/a	AmB 7 days

	d409			
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	

899

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

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

902 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 (99.22)	1169 (0.40)	145 (58)	1096 (0.38)	145 (56)		
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)		

903

904 **Supplementary material**

905 Supplementary Table 1 - MLST results of independently testing three colonies per study
906 isolate.

907 Supplementary Table 2 - Number of SNPs in *MSH2* in each isolate, categorised by type
908 or location. SYN = synonymous SNP, NSY = non-synonymous SNP, p5UTR = 5' UTR,
909 p3UTR = 3' UTR, NON = nonsense mutation.

910 Supplementary Table 3 - Non-synonymous SNPs in Pair 3 isolates, per chromosome

911 Supplementary Table 4 - Number of non-synonymous and synonymous SNPs per isolate

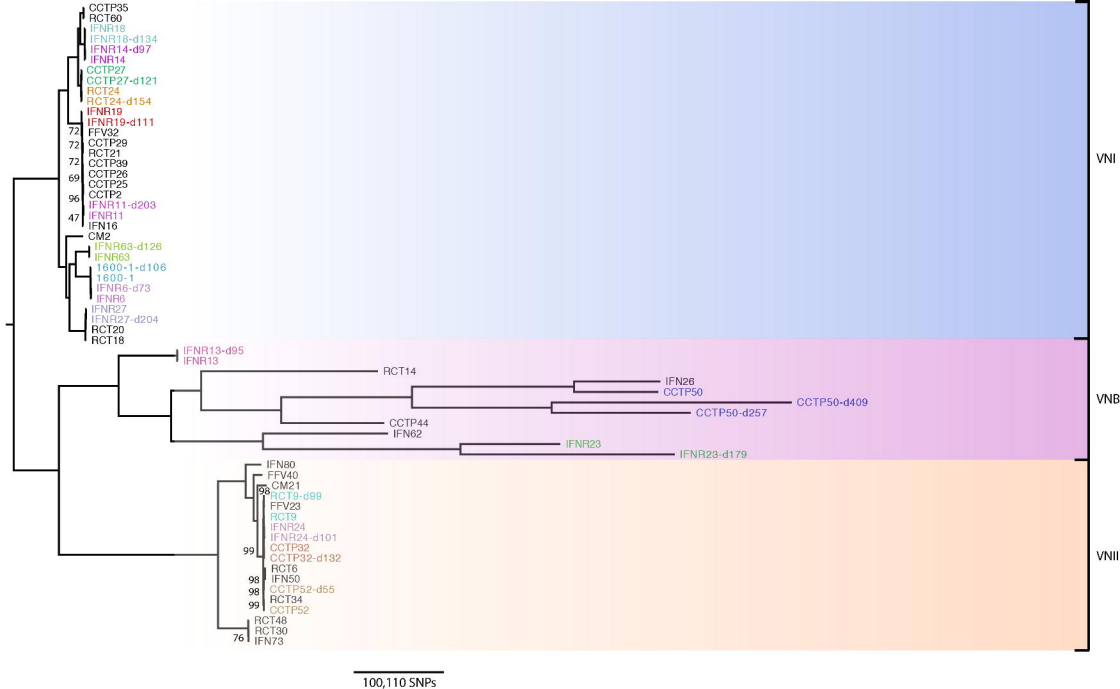
912 Supplementary Table 5 - Details of alignment and variant calling for all paired isolates
913 included in this study

914 Supplementary Table 6 - Number of homozygous and heterozygous SNPs in isolates of
915 Pairs suspected of diploidy.

916 Supplementary Figure 1 – Phylogenetic analysis each chromosome of *Cn* var. *grubii*
917 isolates in this study (coloured) with additional isolates (shown in black), to identify

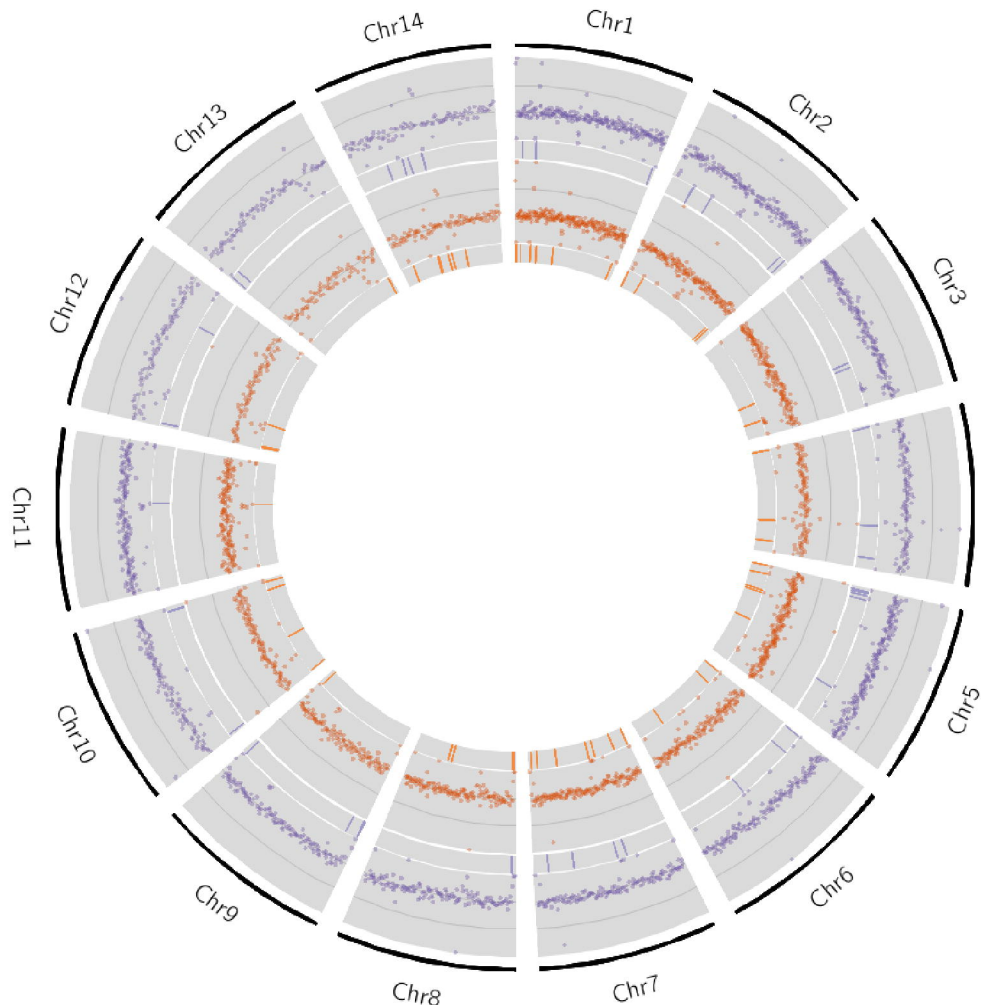
whether Pair 3 consists of different isolates or is a relapsed infection. We hypothesised that if Pair 3 was a relapsed infection, all isolates would share the same phylogenetic relationship in all 14 chromosomes. Rapid bootstrap analysis over 250 replicates was performed on SNP data from 62 isolates to generate an unrooted maximum-likelihood phylogeny (only branches not supported to 100% are indicated above branches). Branch lengths represent the number of SNPs between taxa.

Supplementary Figure 2 – Chromosomal copy number variation can be observed in some isolates in Pairs 1, 4, 5, 10, 14 and 15, when compared to the *Cn* reference genome, H99. Normalised whole-genome depth of coverage is illustrated here, averaged over 10,000-bp bins and represented as a scatter plot. Bar plots represent the position of nsSNPs, where the purple track represents the original isolate, and orange represents the recurrent isolate. Note that Pairs 2, 3 and 17 are not present in this Figure, as they are represented in Figure 2.



a)

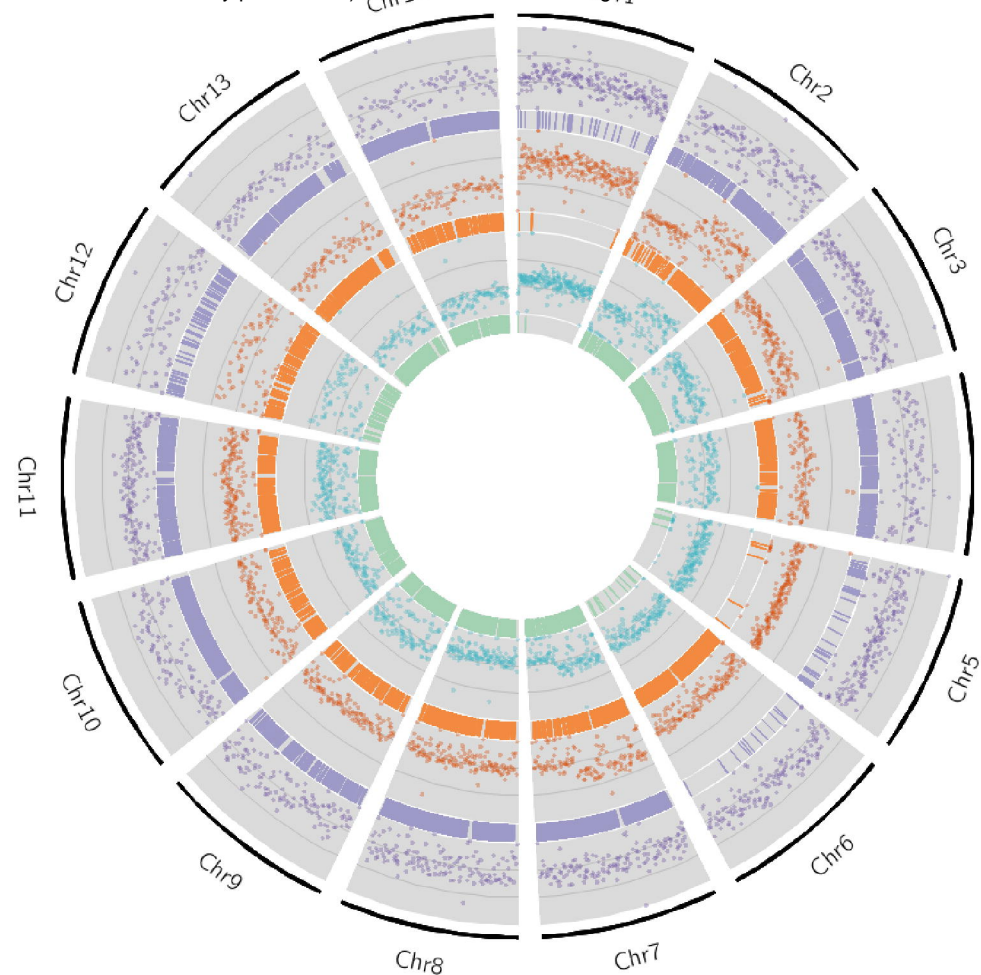
Pair 2



bioRxiv preprint doi: <https://doi.org/10.1101/083469>; this version posted October 26, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

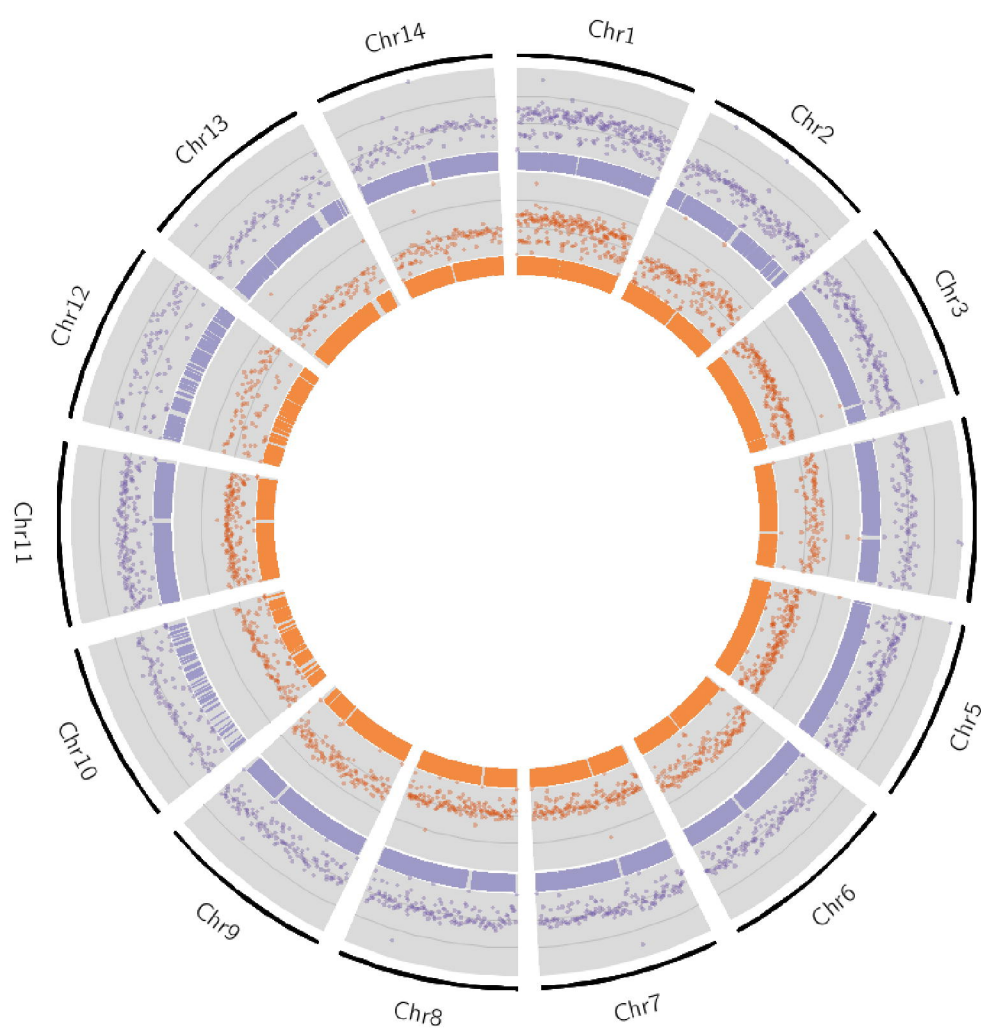
b)

Pair 3

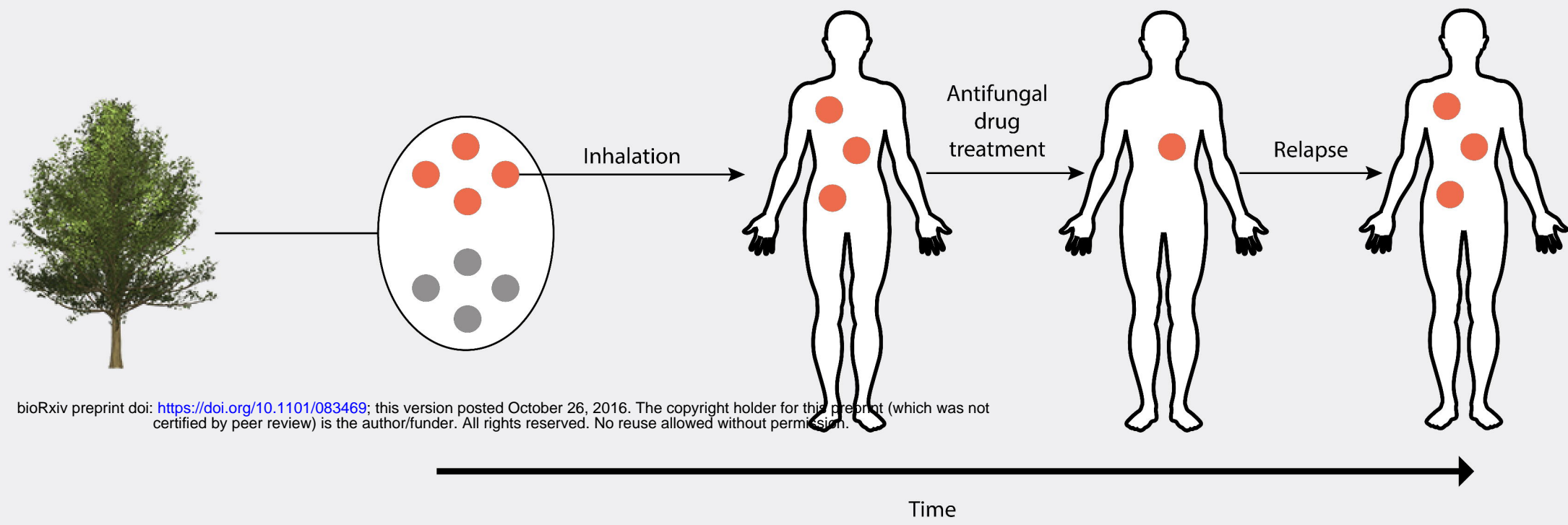


c)

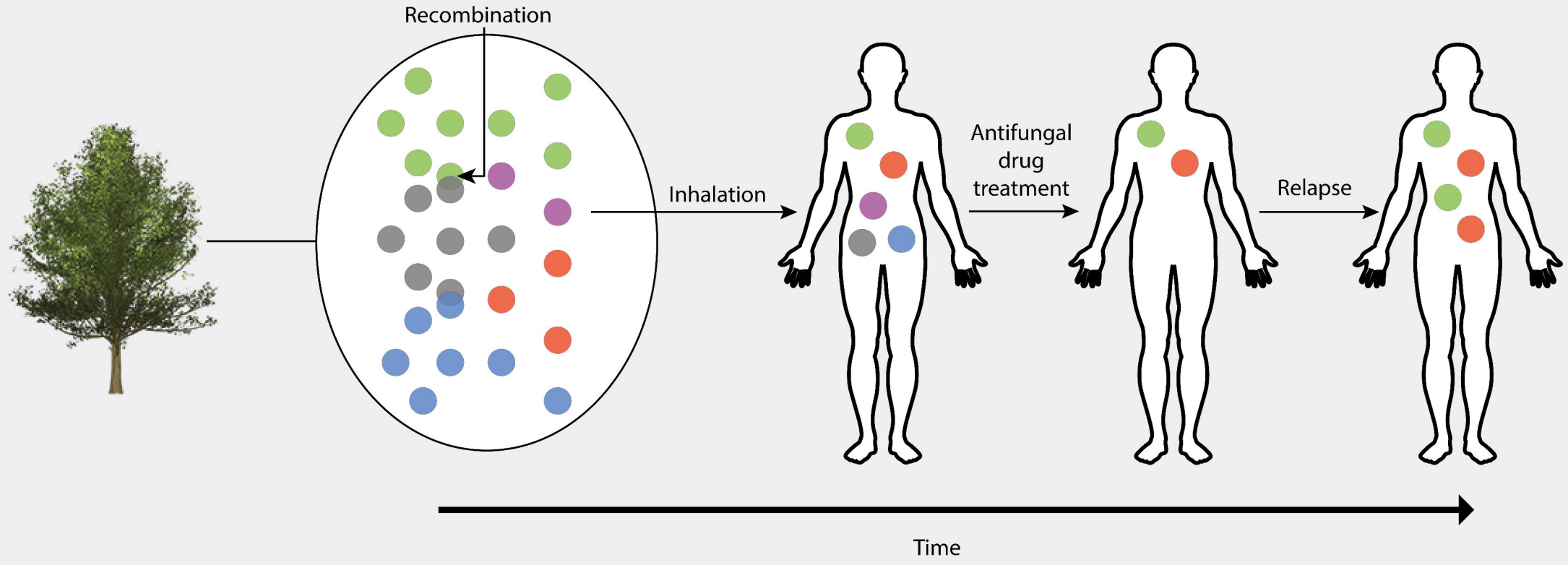
Pair 17



a) Relapse of original infection



b) Infection with population of isolates



c) Infection with hypermutator isolate

