Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*

Christopher A. Desjardins[1], Charles Giamberardino[2], Sean M. Sykes[1], Chen-Hsin Yu[2], Jennifer L. Tenor[2], Yuan Chen[2], Timothy Yang[2], Alexander M. Jones[2], Sheng Sun[3], Miriam R. Haverkamp[4], Joseph Heitman[3], Anastasia P. Litvintseva[5], John R. Perfect[2,*], Christina A. Cuomo[1,*]

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

[2]Division of Infectious Diseases, Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA

[3]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

[4]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[5]Mycotic Diseases Branch, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30329 USA

*Corresponding authors: Christina A. Cuomo (cuomo@broadinstitute.org), John R. Perfect (john.perfect@duke.edu)

Running title: Evolution of virulence in *Cryptococcus neoformans*

Keywords: *Cryptococcus*, evolution, phylogeny, coalescent analysis, population structure, recombination, introgression, GWAS, virulence, melanin, mating-type locus

Abstract

Cryptococcus neoformans is an opportunistic fungal pathogen that causes approximately 625,000 deaths per year from nervous system infections. Here, we leveraged a unique, genetically diverse population of *C. neoformans* from sub-Saharan Africa, commonly isolated from mopane trees, to determine how selective pressures in the environment coincidentally adapted *C. neoformans* for human virulence. Genome sequencing and phylogenetic analysis of 387 isolates, representing the global VNI and African VNB lineages, highlighted a deep, non-recombining split in VNB (herein VNBI and VNBII). VNBII was enriched for clinical samples relative to VNBI, while phenotypic profiling of 183 isolates demonstrated that VNBI isolates were significantly more resistant to oxidative stress and more heavily melanized than VNBII isolates. Lack of melanization in both lineages was associated with loss-of-function mutations in the *BZP4* transcription factor. A genome-wide association study across all VNB isolates revealed sequence differences between clinical and environmental isolates in virulence factors and stress response genes. Inositol transporters and catabolism genes, which process sugars present in plants and the human nervous system, were identified as targets of selection in all three lineages. Further phylogenetic and population genomic analyses revealed extensive loss of genetic diversity in VNBI, suggestive of a history of population bottlenecks, along with unique evolutionary trajectories for mating type loci. These data highlight the complex evolutionary interplay between adaptation to natural environments and opportunistic infections, and that selection on specific pathways may predispose isolates to human virulence.

Introduction

Cryptococcus neoformans is an opportunistic fungal pathogen of enormous clinical importance. An estimated one million new *Cryptococcus* infections occur each year resulting in approximately 625,000 deaths (Park et al. 2009). In Africa, cryptococcosis is the third most common cause of death in patients with HIV/AIDS, and mortality from cryptococcal meningitis exceeds the death rate from tuberculosis (Park et al. 2009). Of the two varieties recognized within the *C. neoformans* species complex, over 95% of cryptococcosis cases are caused by *C. neoformans* var. *grubii* (Chayakulkeeree and Perfect 2006) (serotype A lineage).

Genetic analyses have subdivided *C. neoformans* var. *grubii* into three distinct lineages. The VNI and VNII lineages are highly clonal and globally distributed, while the third lineage, VNB, is genetically diverse yet largely restricted to sub-Saharan Africa (Litvintseva et al. 2006). Environmental reservoirs also differ between these groups; VNB is associated with collection from tree hollows and soil at the base of trees, particularly the mopane tree (*Colophospermum mopane*) which predominates in the African savannah (Litvintseva et al. 2011). VNI is associated with both trees and pigeon excreta (Litvintseva et al. 2011), whereas the environmental reservoir for VNII is not well defined, as the vast majority of known VNII isolates have clinical origin.

The three lineages also differ in the prevalence of each of the two mating types, however, the extent to which this impacts recombination is not well understood. In *Cryptococcus*, the two mating types, a and α, are encoded by a ~100 kb mating type locus (*MAT*) (Lengeler et al. 2002). Sexual reproduction and meiotic recombination can result both from bisexual mating of the opposite mating types as well as from unisexual mating between *MAT*α isolates (Lin et al. 2005). VNI and VNII are predominantly of a single mating type (*MAT*α) while VNB contains a

mix of both mating types (*MAT*α and *MAT***a**); the presence of both mating types in VNB has led to the suggestion of more extensive recombination in this group compared to VNI and VNII (Litvintseva et al. 2003, 2006). Previous genetic analyses of *C. neoformans* suggested recombination in both single and multiple mating type populations (Litvintseva et al. 2003, 2006; Hiremath et al. 2008), although these studies were based on AFLP or MLST data and therefore could only provide limited resolution.

C. neoformans is an environmental opportunistic pathogen, incapable of human to human transmission or dissemination back to the environment from infected hosts. Therefore direct selection of traits to confer virulence or transmission advantages in humans cannot occur. Rather, traits with increased virulence in humans are the result of coincidental selection (Brown et al. 2012), resulting from interactions of *C. neoformans* with other eukaryotes such as predatory amoeba and nematodes in natural environments (Steenbergen et al. 2001). Isolates collected from the environment have been shown to be less virulent in animal models than those recovered from clinical sources (Fromtling et al. 1989; Litvintseva and Mitchell 2009), suggesting that there are advantageous genotypes for growth in a human host. Understanding the selective forces operating on *C. neoformans* in nature is critical to understanding genetic diversity as it relates to virulence in humans.

Three primary virulence factors for *C. neoformans* include a polysaccharide capsule that inhibits phagocytosis, melanin that protects from environmental stresses, and the ability to grow at human body temperature, 37°C (Bulmer et al. 1967; Kwon-Chung and Rhodes 1986). Numerous additional virulence factors have been identified in *C. neoformans*, such as phospolipase B (Cox et al. 2001), urease (Cox et al. 2000), and a number of signaling cascades (Kozubowski et al. 2009). While the importance of these genes has been confirmed by gene deletion, natural genetic variation encoded within these and other factors could also affect

virulence in a quantitative fashion, as genetically similar isolates have been identified that have differential ability to cause disease in animals (Litvintseva and Mitchell 2009).

Previous genomic work has provided reference assemblies of *C. neoformans* (Loftus et al. 2005; Janbon et al. 2014) and the sister species *C. gattii* (D'Souza et al. 2011; Farrer et al. 2015). Here, we leverage the reference assembly of *C. neoformans* strain H99 to conduct large-scale sequencing and variant calling of 387 isolates of *C. neoformans* var. *grubii,* covering the phylogenetic, geographic and ecological breath of the species. We trace the complex evolutionary history of the group using phylogenetic and population genomic analyses, and dissect the genetic differences between clinical and environmental isolates through tests for genome-wide association and selection. For the latter analyses we focus on VNB, as the high genetic diversity and suggested frequent recombination in the lineage, along with ease of isolation from both clinical and environmental sources, make it particularly suited to these tests. Furthermore, we conduct a high-throughput phenotypic screen of 183 VNB isolates to explore how genetic differences relate to responses to environmental stresses. These analyses provide insight into how genetic variation relates to clinical prevalence and how selection in the environment has adapted *C. neoformans* for infection of human hosts.

Results

Phylogenetic analysis reveals deep, non-recombining split in VNB lineage

To better understand the population structure and genetic basis of virulence across *C. neoformans* var. *grubii* lineages, we sequenced and identified variants for 387 isolates, including 185 from VNI, 186 from VNB, and 16 from VNII (Methods, **Supplemental Table S1**). A phylogeny estimated from ~ 1 million variable positions showed clear separation of the three

monophyletic lineages VNI, VNII, and VNB (**Fig. 1**). However, inspection of the phylogeny also revealed a strongly supported (100% bootstrap support), deep split within the VNB lineage, that matched a division seen in previous analyses of AFLP data (Litvintseva et al. 2003), MLST data (Litvintseva et al. 2006; Chen et al. 2015), four nuclear loci (Ngamskulrungroj et al. 2009), and 23 whole genome sequences (Vanhove et al. 2016). This split was further supported by analysis of population ancestry using STRUCTURE (k=4) and of sequence similarity using principle components analysis (PCA), both of which produced four distinct groups representing VNI, VNII, and two VNB lineages with little to no intermixing (**Fig. 2A,B**). We therefore propose that VNB be subdivided into two distinct lineages, VNBI and VNBII, following the nomenclature of (Chen et al. 2015). VNBI encompassed 122 isolates from this study, including previously studied *MAT***a** isolates Bt63 and Bt204, while VNBII included 64 isolates from this study, including previously studied *MAT***a** isolates Bt65 and Bt206 (Litvintseva et al. 2003, 2006; Morrow et al. 2012).

To determine if these groups represented reproductively isolated, non-recombining lineages, we calculated decay of linkage disequilibrium (LD) for VNI, VNBI, VNBII, all VNB, and VNB combined with VNI. We excluded VNII from the calculations due to small sample size. Neither the combined set of VNB and VNI, nor VNB alone, demonstrated evidence of 50% linkage disequilibrium (LD) decay within the 250 kb range of the analysis (**Fig. 2C**), suggesting limited recombination between members of these populations. However, when split into VNI, VNBI, and VNBII, all lineages demonstrated 50% LD decay in 5 to 7.5 kb (**Fig. 2C, Table 1**). This level of recombination is similar to that previously reported for wild populations of *Saccharomyces cerevisiae* and *S. paradoxus* (Liti et al. 2009).

In the phylogenetic analysis, VNI was further subdivided into three distinct clades, two of which were globally distributed (VNIa and VNIb) and one of which was restricted to sub-Saharan Africa (VNIc) (**Supplemental Fig. S1**). These three clades were highly supported (100%

bootstrap support) in this phylogeny; these subdivisions are consistent with one prior MLST analysis (Litvintseva et al. 2006), but not with two major subdivisions identified in a more recent MLST analysis (Ferreira-Paim et al. 2017) (Methods). The VNIa clade contains the highest diversity (pi=0.00178) compared to VNIc (0.00125) and VNIb (0.00084). STRUCTURE and PCA analyses showed the three clades to be largely distinct (**Supplemental Fig. S1B**), with a small subset of isolates showing evidence of mixed ancestry of all three clades. However, faster LD decay was observed in the full VNI set of isolates compared to that of the individual subclades (**Supplemental Fig. S1C**), suggesting there is active recombination between the VNI subclades and that VNI should be considered a single lineage.

To determine whether there was some level of genetic exchange between the groups below the level of detection of genome-wide LD and STRUCTURE analyses, we repeated the STRUCTURE analysis at 500 kb intervals and identified 46 isolates with evidence of recent introgressions of at least 50 kb from different VN groups (**Fig. 3**, Methods). These introgressions were composed of segments ranging in size from 5 to 260 kb (0.02–1.4% of the genome) and occurred in all pairwise combinations of VNI, VNBI, and VNBII as recipient and donor; VNII donors but not recipients were detected. Thirty-two VNBI isolates contained introgressions, which was significantly more than the 14 VNI isolates and 5 VNBII isolates that contained introgressions ($p < 1.32\times10^{-5}$ and $p < 0.0033$, respectively, Fisher's exact test). Isolates with introgressions included both mating types and were dispersed across the VNB phylogeny, but within VNI only occurred within the African-specific subclade (**Fig. 1**). Overall, introgressed regions were not functionally enriched for any specific gene functions; however, the most common introgression in VNBI (present in 14 clinical and environmental isolates, from VNI) encompasses 40 kb and encodes a glycolipid mannosyltransferase involved in capsule synthesis (CNAG_00926) (O'Meara and Alspaugh 2012) and secretome component carboxypeptidase D (Geddes et al. 2015). Conversely, the most common introgression in VNI

(present in 8 clinical and environmental isolates, from VNBI) encompasses 60 kb and contains virulence factor *VCX1* (Kmetzsch et al. 2010) and azole drug target *ERG11*. This suggests a low level of mixing of genomic regions with clinically relevant phenotypes may be occurring between lineages in sub-Saharan Africa.

VNBII isolates are enriched for clinical isolation source relative to VNBI

While the VNB lineage has previously been associated with a high frequency of environmental samples and the **a** mating type, we found differences between the relative frequencies of these properties between VNBI and VNBII. While VNBI contained large numbers of both clinical and environmental samples (52 vs. 70, respectively), VNBII contained almost exclusively clinical isolates (60 vs. 3, respectively), which was a significant difference ($p < 0.0001$, Fisher's exact test). Environmental isolates in both VNBI and VNBII were scattered across the phylogenetic tree and showed no evidence of clustering, although all three environmental VNBII isolates were relatively basal in the lineage (**Fig. 1**). SNP calling and phylogenetic analysis of 23 recently published Zambian isolates (Vanhove et al. 2016) showed a similar bias in environmental and clinical isolates in the two lineages; the Zambian mixed environmental and clinical VNB-A clade matched our VNBII lineage, while the VNB-B clade, which was entirely environmental samples, matched our VNBI lineage (**Supplemental Fig. S2**). Furthermore, VNBI had a greater proportion of *MAT***a** isolates than VNBII (33/122 vs. 9/64, respectively), although the difference was not significant.

Analysis of annotated whole-genome assemblies of representatives of the four lineages revealed very few gene gains or losses specific to VNBI and VNBII (**Supplemental Table S2**). Most gains and losses were hypothetical proteins, with the exceptions of a gain of an α-β hydrolase and loss of a phosphopyruvate hydratase and an alginate lyase in VNBI, and a gain

of an L-iditol 2-dehydrogenase in VNBII. However, a large number of fixed polymorphisms were present between the two lineages, in addition to private unfixed alleles in each lineage and unfixed alleles shared between the two lineages (**Supplemental Fig. S3A**). Therefore small polymorphisms may underlie much of the phenotypic variation within and between these lineages.

To identify genomic features that might contribute to increased prevalence in clinical populations, we conducted a genome-wide association analysis (GWAS) of small polymorphisms and loss-of-function mutations across VNBI and VNBII. Correction for population stratification was included so that features that evolved at the base of VNBI would not overwhelm the analysis. The top 10 loci associated with the clinical phenotype included a number of known virulence factors and genes involved in oxidative stress response (**Table 2**). The top scoring locus was the intergenic region downstream of CNAG_07989 ($p < 3.0 \times 10^{-7}$, score test), which encodes a velvet family protein, and a deletion of that gene prevents filamentation (Chacko et al. 2015). Also identified were the intergenic region downstream of *LMP1*, a known virulence factor identified through laboratory passage of H99 (Janbon et al. 2014), *CCK2*, which encodes a casein kinase related to *CCK1* (essential for cell integrity and virulence (Wang et al. 2011)), and three genes, *PDX3*, *FOX21*, and *FHB1*, previously noted to respond to oxidative and nitrosative stress (de Jesús-Berríos et al. 2003; Upadhya et al. 2013).

We also examined geographical associations of the VNBI and VNBII lineages. VNBII isolates were isolated almost exclusively from patients treated at the Princess Marina Hospital in southern Botswana, while VNBI isolates were common in all four heavily sampled locations, including patients from both hospitals and the environment in south, northeast, and north-central Botswana (**Supplemental Table S3**). Further evaluation of the VNBI population revealed little evidence for genetic structure separating or connecting specific locations (**Supplemental Table**

S4). We also examined the combined phylogeny (**Supplemental Fig. S2**) of the Botswana isolates with the previously sequenced Zambian isolates (Vanhove et al. 2016) to determine if these patterns extended outside of Botswana. Surprisingly, both VNBI and VNBII Zambian isolates largely clustered near the base of each lineage, suggesting some genetic separation between Botswanan and Zambian VNB lineages and that some geographic isolation may be occurring at the scale of country. The Zambian VNI isolates, on the other hand, were dispersed throughout our global VNIa and VNIb clades, though not the African VNIc clade, suggesting greater gene flow between VNI populations in the region.

VNBI and VNBII differ in their ability to melanize and respond to oxidative stress

To determine if the distinct genetic backgrounds of VNBI and VNBII, or variants identified in the association analysis, correlated with any relevant phenotypes, we assayed 183 of the 186 sequenced VNBI and VNBII isolates for growth under a variety of stress conditions, along with a set of control isolates including the VNI reference strain H99. We specifically tested response to oxidative stressors H_2O_2 and paraquat, melanization capacity with L-DOPA, and resistance to the antifungal drug fluconazole (results in **Supplemental Table S5**). We did not identify any isolates with mutations in stress response genes *PDX3*, *FOX21*, or *FHB1* with extreme response phenotypes. However, we did identify significant phenotypic differences between VNBI and VNBII isolates and between clinical and environmental isolates. To control for the interaction of lineage and isolation source, we analyzed only clinical isolates for lineage comparisons and only VNBI isolates for isolation source comparisons. We also excluded all isolates that were completely unable to melanize in the comparison of melanization between environmental and clinical isolates to prevent any effect of sampling bias (Methods).

VNBI clinical isolates were significantly more melanized than VNBII clinical isolates ($p < 0.00017$, Mann-Whitney *U* test, **Fig. 4A**), and significantly more resistant to both H_2O_2 ($p < 0.0049$, Mann-Whitney *U* test, **Supplemental Fig. S4A**) and paraquat ($p < 0.042$, Mann-Whitney *U* test, **Supplemental Fig. S4B**), but not fluconazole ($p < 0.78$, Mann-Whitney *U* test, **Supplemental Fig. S4C**). Within VNBI, environmental isolates were more heavily melanized ($p < 0.00047$, Mann-Whitney *U* test) than clinical isolates and more resistant to both H_2O_2 ($p < 0.00098$, Mann-Whitney *U* test) and paraquat ($p < 4.1\times10^{-5}$, Mann-Whitney *U* test). The groups did not significantly differ in their resistance to fluconazole ($p < 0.071$, Mann-Whitney *U* test).

We then conducted a second GWAS analysis to identify genotypic variants associated with increased or decreased responses in the phenotypic assays (**Supplemental Tables S6-S9**). These analyses revealed that loss-of-function mutations in *BZP4* (CNAG_03346) were highly associated with reduced melanization ability ($p < 3.43\times10^{-9}$, score test). Four clinical isolates, two from VNBI and two from VNBII, each had a different loss-of-function mutation in *BZP4*. Three of these isolates had a reduced melanization ability similar to that of a *lac1* deletion mutant, the primary gene required for melanization (Salas et al. 1996), while the fourth isolate showed a slightly reduced melanization potential (**Fig. 4A-C**). *BZP4* was previously associated with a reduced melanization capacity in a deletion screen of *C. neoformans* transcription factors(Jung et al. 2015). We then further screened genes previously categorized as having melanization defects in another deletion screen (Liu et al. 2008) and identified a second gene, *CHO2* (CNAG_03139) at lower rank in the association analysis ($p < 0.0083$, score test). Loss-of-function of this gene is present in all VNBII isolates and could be related to the reduced melanization capability of VNBII. While GWAS analysis did not produce any promising candidates for fluconazole resistance, copy number analysis revealed a duplication of drug target *ERG11* in the four most resistant isolates (Bt65, Bt89, MW-RSA3179, and MW-RSA2967).

283

284 **Population genomic analysis identifies extensive loss of diversity in VNBI**

285

286 To better understand genetic diversity at the lineage level, we calculated a number of population
287 genetic statistics for VNI, VNBI, and VNBII, including π, Tajima's D, and F_{ST} (**Table 1**). Genome-
288 wide calculations of the diversity metric π revealed that VNBI and VNBII had nearly twice the
289 genetic diversity of VNI (**Table 1**). However, when diversity was plotted along each
290 chromosome, VNBI and VNBII appeared quite different. As shown in **Fig. 5A** for chromosome 5,
291 the VNI group contained long tracts of low diversity while VNBII contained predominantly high
292 diversity regions, as expected. By contrast, VNBI showed a unique pattern with high diversity
293 regions interspersed with long tracts of low diversity. These regions of low diversity correspond
294 with tracts of low Tajima's D, which suggests an excess of rare alleles resulting from a
295 population bottleneck or recent selective sweep. Confirming this, both VNI and VNBI had
296 significantly more rare alleles across the entire genome (<2% frequency) than VNBII ($p<2.2\times10^{-16}$
297 for both comparisons, Fisher's exact test, **Supplemental Fig. S3B**). To determine if this
298 pattern appeared in all chromosomes, we plotted the distribution of Tajima's D values for each
299 chromosome individually (**Fig. 5B**). While VNI and VNBII have unimodal distributions for most
300 chromosomes, VNBI has a bimodal distribution of Tajima's D across nearly all chromosomes,
301 with the possible exception of chromosome 6. However, when we evaluated genes located in
302 the regions of reduced Tajima's D we found no significant functional enrichment in any of the
303 lineages.

304

305 In most cases, when two lineages had regions of low diversity and Tajima's D, they also had
306 high F_{ST}. However, there was one notable exception to this on chromosome 5, where a region
307 of reduced diversity in both VNBI and VNBII included a region of high F_{ST} abruptly adjacent to a
308 region of low F_{ST} (**Supplemental Fig. S5**). The low F_{ST} region includes CNAG_07442, which

encodes a phosphatidylglycerol transfer protein that was the most abundant antigen recovered from an experimental mouse vaccine using strain cap59 (Specht et al. 2015), suggesting a similar antigen would be found in both lineages. The high F_{ST} region included CNAG_01058, a gene involved in capsule maintenance (Brown et al. 2014) and CNAG_01060, a potential substrate of the *FBP1* gene required for proliferation in macrophages and dissemination (Liu and Xue 2014).

Inositol and other sugar transporters are targets of selection in all VN lineages

As Tajima's D can identify effects of both selection and demography (Tajima 1989), we used the composite likelihood ratio test (CLR) (Nielsen et al. 2005), which is specific to selection, in a genomic scan for regions under selection in each lineage (**Supplemental Fig. S6**). Surprisingly, there was very little overlap between the genes in regions under positive selection and the genes in regions of reduced diversity in any of the lineages, suggesting that selective sweeps were not fixing genes under selection. To further evaluate genes under positive selection, we identified genes in regions with the top 5% of CLR values for each of VNI, VNBI and VNBII. A functional enrichment analysis of these genes revealed that major facilitator superfamily (MFS) transporters, and in particular the sugar transporter subset, were under selection in all lineages (**Supplemental Table S10**). Overlaying these genes onto a phylogeny of all sugar transporters in the reference strain H99 showed the selection was concentrated in transporters of inositol, xylose, maltose, α-glucosides, lactose, and galactose (**Supplemental Fig. S7**). Six inositol transporters were found to be under selection in at least one lineage, including *ITR1A* (CNAG_04552), which was under selection in all three lineages (**Table 3**). Furthermore, genes involved in inositol catabolism, including four copies of inositol 2-dehydrogenase and one copy of inositol oxygenase were also under selection, notably in VNI where all five genes were under

334 selection. Other groups of genes under selection with enriched functions include
335 monoxygenases in VNI and β-glucosidases in VNBII.
336
337 Subtelomeric regions of *Cryptococcus* have been previously highlighted as regions of high
338 diversity heavily enriched for transporters (Chow et al. 2012). To determine if subtelomeric
339 regions showed signatures of selection, we searched for enrichment of windows under selection
340 in the outer 40 kb of each chromosome for VNI, VNBI, and VNBII. Chromosomes 1, 4, 5, and
341 14 showed subtelomeric enrichment of selection in all three examined lineages, and all
342 chromosomes except 2 and 6 showed enrichment in at least one lineage (**Supplemental Table**
343 **S11**). This suggests subtelomeric regions encode numerous targets of selection across the VNI,
344 VNBI, and VNBII lineages. Enrichment analysis comparing genes under selection in the
345 subtelomeric regions to the remainder of the genome showed the same patterns as the
346 genome-wide selection analysis. Additionally, an enrichment of fungal transcription factors was
347 identified (**Supplemental Table S12**), including *FZC20*, *FZC41*, *SIP402*, and *FZC22*, the last of
348 which has been shown to have reduced virulence in mice when deleted (Jung et al. 2015).
349
350 **Mating type loci *MAT*α and *MAT*a have distinct evolutionary trajectories**
351
352 The mating type locus also showed strong signatures of selection. To better understand the
353 evolution of the *MAT* locus, we re-called SNPs against complete assemblies of each of the two
354 highly diverged mating types (*MAT*α and *MAT***a**) (Lengeler et al. 2002) and inferred phylogenies
355 of each mating type separately (**Fig. 6**, Methods). The *MAT*α phylogeny largely matched the
356 whole-genome phylogeny, with one major exception: VNBI *MAT*α appeared paraphyletic with
357 respect to VNBII (**Fig. 6A, B**), with two distinct alleles—one comprised of some VNBI and the
358 other clade comprised of the remaining VNBI plus all VNBII (bootstrap support: 100%). These

359 distinct clades did not map to specific sub-clades of VNBI in the whole-genome phylogeny but
360 rather were intermixed.
361
362 The *MAT***a** phylogeny also differed from the whole-genome phylogeny (**Fig. 6D, E**) in that VNI
363 and VNBI were closely related sister groups and separated by a large distance from VNBII
364 (bootstrap support: 100%). Bootstrap support for the monophyly of VNBI with respect to VNI
365 was limited (68%), indicating the possibility that the VNI *MAT***a** allele was derived from a basal
366 VNBI allele. Additionally, alleles co-segregating VNBI *MAT***a** isolates with VNI and VNBII *MAT***a**
367 isolates extend about 5.3 and 3.5 kb, respectively, upstream of the 5' end of the *MAT* locus
368 (**Supplemental Fig. S8**). This suggests a local introgression of the *MAT***a** locus from VNBI into
369 VNI, consistent with previous reports of hotspots that flank the *MAT* region (Hsueh et al. 2006;
370 Sun et al. 2012).
371
372 The *MAT*α and *MAT***a** alleles also showed distinct patterns of genetic diversity. Both alleles
373 showed spikes of high genetic variation in a small number of hotspots (**Fig. 6C, F**: π); all of the
374 regions corresponded to long terminal repeats in the annotations of the mating type loci,
375 suggesting only limited variation has accumulated in functional genes. Within *MAT*α, VNI alleles
376 had significantly fewer singleton polymorphisms than either VNBI or VNBII, resulting in shorter
377 terminal branch lengths (**Fig. 6B**; Wilcoxon rank sum test, $p < 3.8\times10^{-5}$ and $p < 5.2\times10^{-6}$,
378 respectively). While this could be the result of mixing within VNI but not VNBI or VNBII, the
379 same pattern was observed in the whole-genome data (**Fig. 1**; Wilcoxon rank sum test, $p <$
380 2.2×10^{-16} and $p < 1.7\times10^{-8}$), suggesting an overall increase in genetic variation in VNBI or VNBII
381 relative to VNI may be responsible for the difference. Between *MAT*α and *MAT***a**, *MAT***a** alleles
382 had significantly more singleton polymorphisms than *MAT*α, even when VNI isolates were
383 excluded (**Fig. 6B, E**; Wilcoxon rank sum test, $p < 1.3\times10^{-6}$), which could indicate less mixing
384 between *MAT*α alleles. Extensive linkage disequilibrium was seen in both *MAT* alleles (**Fig. 6C,**

385 **F**: R^2), suggesting that despite recombination occurring at the whole-genome level within VNI,
386 VNBI, and VNBII, limited or no recombination was occurring within the mating type loci. In total,
387 these analyses suggest that each mating type locus in *C. neoformans* has followed a distinct
388 evolutionary trajectory from the remainder of the nuclear genome.
389
390 **Discussion**
391
392 Phylogenomic analysis presented here provided strong support for a deep split separating the
393 VNB lineage in VNBI and VNBII, and further population genomic analysis revealed a lack of
394 recombination between the two. These lineages also differed in their ability to be recovered from
395 the environment, respond to oxidative stress, and melanize, and to a lesser extent the presence
396 of both mating types (*MAT***a** was less common in VNBII). Based on these genomic and
397 phenotypic differences, we propose that VNBI and VNBII should be considered two separate
398 lineages. The lower frequency of isolates from environmental sources in VNBII could be due to
399 an increased ability for VNBII isolates to colonize humans; there also may be geographic or
400 niche differences in VNBII environmental populations. Despite the fact that melanization is used
401 to identify isolates from the environment (*i.e.*, light brown isolates are detected but pure white
402 isolates are not), it is unlikely that decreased melanization of VNBII caused the group to be
403 under-sampled as most VNBII isolates melanized to at least some degree. Further genotypic
404 and phenotypic evaluation will be required to better delineate and understand the clinical
405 ramifications of the differences between VNBI and VNBII.
406
407 Both “into” and “out of” Africa hypotheses have been proposed for VNI (Litvintseva et al. 2011).
408 However, given the existence of the sub-Saharan Africa endemic VNIc and the presence of
409 African isolates spread across VNIa and VNIb, the “into” hypothesis seems unlikely. More likely,
410 the three VNI sub-lineages diversified and expanded within Africa and then two were globally

411 disseminated at a later date, possibly by migration of birds and humans. Why VNI but not VNBI
412 or VNBII would be globally dispersed is unclear. While the data presented here support an
413 African origin for VNB, a small number of South America VNB isolates have been described,
414 including a previous analysis of nuclear loci that placed two of these isolates close to Bt22
415 (Ngamskulrungroj et al. 2009), identified here as a VNBI isolate. Comprehensive sequencing of
416 South American VNB isolates will be required to confidently assess the origins of VNB diversity.
417
418 Given the lack of overlap between the tracts of low Tajima's D and the regions identified as
419 under selection, we hypothesize that the loss of diversity in VNI and VNBI are the result of
420 previous demographic events, such as population bottlenecks, rather than selective sweeps
421 fixing genes under selection. The global VNI lineage, with low overall diversity, likely emerged
422 from a small founder population that may have undergone repeated bottlenecks while spreading
423 geographically. Why the VNBI lineage, endemic to southeastern Africa, would have undergone
424 a major demographic event, is unclear, as it appears to be the most well-represented lineage in
425 current environmental sampling.
426
427 Previous analyses have highlighted the presence of recombination between isolates of different
428 lineages based on MLST gene analysis (Litvintseva et al. 2006; Chen et al. 2015). The
429 resolution possible with whole genome data demonstrates that many of these isolates predicted
430 to be recombinant by analysis of single genes such as *SOD1* are predominantly a single
431 ancestry based on whole genome analyses (*e.g.*, Bt65, Bt109). The detection of small
432 introgressions in VNBI, VNBII, and African VNI isolates suggests that a limited amount of
433 recombination is occurring in Africa where the three groups overlap in distribution. Given the
434 clinical significance of genes in these regions such as *ERG11*, better understanding of the
435 phylogenetic and geographic distribution of introgressions may be of clinical importance.
436

437 VNI showed a level of intragroup recombination comparable to VNBI and VNBII. This was
438 unexpected as both mating types are common in the VNB lineages, suggesting sexual
439 reproduction would dominate in VNB, while the **a** mating type is rare in VNI, suggesting
440 predominance of asexual reproduction in VNI. However, the high level of recombination seen in
441 VNI may indicate unisexual mating in VNI is nearly as common as bisexual mating in VNBI and
442 VNBII. It has been shown in *C. neoformans* var. *neoformans* (serotype D) isolates that
443 recombination is similar in unisexual and bisexual mating (Sun et al. 2014; Fu et al. 2015).
444
445 These population genomic analyses can also be used to evaluate lineage and species
446 boundaries that are currently under debate for *Cryptococcus* (Hagen et al. 2015; Kwon-Chung
447 et al. 2017). A multi-locus viewpoint is critical for a comprehensive understanding of
448 recombination and incomplete lineage sorting within a species, as recently shown for *Alternaria*
449 (Stewart et al. 2014). Despite high bootstrap support for the monophyly of the four population
450 subdivisions (VNI, VNII, VNBI and VNBII) in the whole-genome phylogenetic tree, the detection
451 of recombination between the lineages, including small introgressions, suggests that isolates
452 from the different lineages mate and form recombinant progeny; further analysis including of
453 hybrid strains (Chen et al. 2015) will help clarify if these lineages represent varieties or sibling
454 species. While whole-genome analyses may largely recapitulate the phylogenetic relationships
455 revealed from MLST studies, they can also reveal complex genetic exchange within a species at
456 a level of detail not possible in smaller scale studies.
457
458 Phylogenetic analyses of the *MAT* alleles suggest that these alleles have followed distinct
459 evolutionary trajectories from the remainder of the nuclear genome, involving incomplete sorting
460 or genetic mixing following divergence of the groups, as seen in the paraphyly of the VNBI
461 *MAT*α allele and close relationship of the VNI *MAT***a** allele to VNBI. These observations suggest
462 an outline of the evolution of mating in these groups: the presence of both mating types is

ancestral in VNBI and VNBII, and sexual reproduction in these groups involves opposite as well as potentially same sex mating. However, VNI may have diverged with only the *MAT*α mating type, possibly due to a small founding population size. This would have provided evolutionary pressure for unisexual reproduction, as is suggested by the evidence of extensive recombination within the nuclear genomes of VNI isolates. Following the divergence of VNI from VNBI and VNBII, we speculate that VNI then re-acquired the *MAT***a** allele from a VNBI-like ancestor.

The selection and association analyses, along with the phenotypic assays, provided clues to both selection pressures experienced by *C. neoformans* in the environment and which of these traits may translate to pathogenicity in humans. Selection on both inositol transport and utilization could affect the lineage-specific host range of *C. neoformans* and easily translate to altered virulence levels in humans, as inositol is abundant in the human brain (Fisher et al. 2002) and is required by *C. neoformans* for virulence (Shea et al. 2006) and mating (Xue et al. 2010). Gene expansions of inositol transporters in *C. neoformans* have been previously proposed to have evolved for adaptation growth on trees yet preadapted *C. neoformans* for growth in the human brain (Xue et al. 2010; Xue 2012). Selection on xylose facilitators suggests potential adaptation to different tree hosts with different chemical compositions. GWAS analysis also revealed a gene known to be involved in the yeast-hyphal transition. As the inability to maintain yeast phase is known to correlate with decreased virulence (Wang et al. 2012), it is easy to hypothesize that natural adaptations to environments where increased time in the yeast state would be selected for could result in isolates with increased virulence in humans. Future genomic and phenotypic analyses focused on VNI may help better understand the specific properties that enabled the global dispersal of this lineage.

488 It was initially surprising that within VNBI, environmental isolates were more resistant to
489 oxidative stress and more heavily melanized than clinical isolates, particularly given that
490 melanization is a known virulence factor in humans. However, the diversity of stressors in the
491 environment is likely greater than within the human host, and it has already been hypothesized
492 that the ability of *C. neoformans* to survive in macrophages was derived from an ability to
493 defend against predatory protozoa in the environment (Steenbergen et al. 2001). Further
494 experiments are needed to disentangle the stresses faced in these two contexts and the
495 differential ability of environmental and clinical isolates to respond to these stresses. Here,
496 GWAS analysis linked loss-of-function mutations in the transcription factor *BZP4* to significantly
497 reduced melanization in a number of unrelated clinical isolates, highlighting the potential clinical
498 impact of this complex interplay. In sum, these data provide insight into how selection pressure
499 over the evolutionary history of *C. neoformans* may have pre-adapted to successful invasion of
500 the human nervous system, and suggests many pathways to explore for differential virulence
501 within standing genetic diversity or outbreaks.
502
503 **Methods**
504
505 **Sample Selection**
506 We selected 387 isolates for analysis: 185 from VNI, 186 from VNB, and 16 from VNII. We
507 sought to represent isolation sources and mating types as evenly as possible for each of VNI
508 and VNB, and therefore preferentially chose isolates containing *MAT***a** and those isolated from
509 environmental sources, as both are less frequently collected, particularly for VNI. We also
510 included a limited number of VNII isolates, but did not focus on this lineage as it is rarely
511 isolated. The majority of isolates, particularly VNB isolates, were collected in Botswana, and
512 more specific collection details of these isolates are given in **Supplemental Material** and a
513 previous MLST analysis (Chen et al. 2015).

514

515 **Sample Preparation, Sequencing, and Variant Identification**

516 Samples were prepared for DNA extraction as described (**Supplemental Material**). Genomic
517 DNA was sheared to ~250 bp using a Covaris LE instrument and adapted for Illumina
518 sequencing as previously described (Fisher et al. 2011). Libraries were sequenced on an
519 Illumina HiSeq to generate 101 base reads. Reads were aligned to the *C. neoformans* var.
520 *grubii* H99 assembly (Genbank accession GCA_000149245.2) using BWA-MEM version 0.7.12
521 (Li 2013). Variants were then identified using GATK version 3.4 (McKenna et al. 2010) as
522 described (**Supplemental Material**), and functionally annotated with SnpEff version 4.2
523 (Cingolani et al. 2012). Read data from 46 previously sequenced Zambian isolates (Vanhove et
524 al. 2016) was downloaded from NCBI (BioProject PRJEB13814) and variants were called for all
525 samples as described above.

526

527 **Phylogenetic Analysis**

528 For phylogenetic analysis, the 1,069,080 sites with an unambiguous SNP in at least one isolate
529 and with ambiguity in at most 10% of isolates were concatenated; insertions or deletions at
530 these sites were treated as ambiguous to maintain the alignment. Phylogenetic trees were
531 estimated using RAxML version 8.2.4 (Stamatakis 2014) under the GTRCAT model in rapid
532 bootstrapping mode. We then placed the 46 previously sequenced Zambian isolates (Vanhove
533 et al. 2016) in phylogenetic context with the isolates in this study using FastTree version 2.1.8
534 compiled for double precision (Price et al. 2010).

535

536 **Population Genomic Analysis**

537 Population structure was investigated using multiple approaches. Principal components analysis
538 was run on all variants using SMARTPCA (Patterson et al. 2006). Major ancestry subdivisions
539 were delineated using STRUCTURE (Pritchard et al. 2000) v2.3.4 in the site-by-site mode and

k=4, with 15% of the positions randomly subsampled from those containing at least 2 variants and missing in fewer than 5% of isolates. Identification of VNI subdivisions and smaller inter-lineage introgressions using STRUCTURE were performed as described (**Supplemental Material**). Population genomic metrics were calculated using vcftools version 1.14 (Danecek et al. 2011) and PopGenome (Pfeifer et al. 2014); further details are given in **Supplemental Material**. The CLR selection metric was also calculated using PopGenome , and enrichment tests were conducted using Fisher's exact test corrected with the Benjamini-Hochberg method for multiple comparisons (**Supplemental Material**).

Association Analysis

For each phenotype, two matrices of these features were constructed from the variant calls of all isolates. In the first matrix, rare variants (those at ≤ 5% frequency) were collapsed by gene or intergenic region, while common variants (those at > 5% frequency) were considered independently. In the second matrix, we assessed whether each gene had a loss-of-function mutation (defined as s "HIGH" impact mutation by SnpEff), regardless of the frequency of the mutation. Each variant matrix was then analyzed with GEMMA (Zhou and Stephens 2012) using a univariate linear mixed model and a relatedness matrix to account for population stratification. Score test results of both analyses were then combined and examined. *ERG11* and *AFR1* copy number of drug resistant isolates were estimated as the normalized read depth of these genes from whole genome BWA-MEM alignments, where the per base depth calculated using samtools (Li et al. 2009) v1.3 mpileup was summed across each gene and normalized to the average genome depth.

Phenotypic Assays

Freezer stocks were grown, plates, and pinned in 1536 array format as described (**Supplemental Material**); each single colony selected from the grown freezer stock became a 4

566 x 4 block of 16 colonies in two 1536 arrays. Four conditions were assayed: 5.6 mM H_2O_2, 1.0
567 mM paraquat, 10 µg/mL fluconazole, and 0.1 mg/mL L-DOPA. Arrays were incubated at 30°C
568 and images were acquired one-, two-, and three-days post pinning. Day two images were
569 analyzed using gitter (Wagih and Parts 2014) to assess either colony size in the case of H_2O_2,
570 paraquat of fluconazole, or brightness in the case of L-DOPA. Each condition was tested in
571 either one (H_2O_2), two (L-DOPA and fluconazole) or three (paraquat) independent experiments.
572 All independent experiments replicated well (R^2, L-DOPA = 0.51, fluconazole = 0.59, paraquat =
573 0.47-0.63). We then identified differences between VNBI and VNBII clinical isolates, and VNBI
574 environmental and clinical isolates, using the Mann-Whitney *U* test. As *C. neoformans* is
575 identified in the environment (but not in the clinic) by the ability to melanize to some degree (*i.e.*,
576 light brown isolates are detected but pure white isolates are not), in the comparison of
577 melanization of environmental and clinical VNBI isolates, we excluded isolates that were unable
578 to melanize (brightness >= 0.6) because the sampling of those isolates would be biased
579 towards clinical samples.
580
581 **Mating Type Analysis**
582 Analyses of the mating type loci were done with the same methodology as the whole genome
583 analyses with the following two exceptions. Reads were mapped to the appropriate high quality
584 locus-specific reference, either the H99 *MAT*α allele (Genbank accession AF542529) or the
585 125.91 *MAT***a** allele (Genbank accession AF542528). Second, linkage disequilibrium was
586 calculated on the site level rather than in windows, and visualized using the R package
587 LDheatmap (Shin et al. 2006).
588
589 **Data Access**
590 Genbank accession numbers of all sequencing data generated for this study are listed in
591 **Supplemental Table S1**.

592

593 **Acknowledgments**

594

595 We would like to thank Jose Munoz for providing helpful comments on the manuscript. This
596 project has been funded in whole or in part with Federal funds from the National Institute of
597 Allergy and Infectious Diseases, National Institutes of Health, Department of Health and
598 Human Services, under Grant Number U19AI110818 to the Broad Institute.
599 Support to J.R.P. came from Public Health Service Grants AI73896, AI93257. The use of
600 product names in this manuscript does not imply their endorsement by the US Department of
601 Health and Human Services. The finding and conclusions in this article are those of the authors
602 and do not necessarily represent the views of the CDC.

603

604 **Author Contributions**

605

606 C.A.C., A.L., and J.P. conceived and designed the project. M.H. provided the clinical isolates.
607 C.G., T.Y., Y.C., and A.M.J. performed the laboratory experiments. C.A.D., S.M.S., C.-H.Y.,
608 C.A.C., S.S., J.H., and C.G. analyzed the data. C.A.D., C.A.C., and A.L. wrote the paper.
609 C.A.C., J.P., and J.L.T. supervised and coordinated the project.

610

611 **Ethics Statement**

612

613 The clinical isolates were approved under Pro00029982 from the Duke Institutional Review
614 Board in USA and the Review Boards of Nyangabgwe and Princess Marina
615 Hospitals of Botswana.

616

Table 1. Genomic and population characteristics of global sampling of the four lineages of *C. neoformans* var. *grubii*. Listed statistics include sample size, physical distance needed for linkage disequilibrium to decay 50% (LD50), diversity (π), percent of the genome with reduced Tajima's D (TD), and number and percent of isolates isolated from clinical and environment sources, and of each mating type (*MAT*α or *MAT***a**). Given the limited sampling of VNII, many statistics were not determined (ND) or not applicable (NA).

Statistic	**VNI**	**VNBI**	**VNBII**	**VNII**
Sample Size	185	122	64	16
LD50 (kb)	7.5	6	5	ND
π	0.00209	0.00360	0.00358	ND
% with reduced TD	1.8	12	2.4	ND
Metadata				
# Clinical	156	51	60	13
# Environmental	24	71	3	1
# *MAT*α	182	89	55	16
# *MAT***a**	3	33	9	0
% Clinical/Environmental	NA	43/57	94/6	NA
% *MAT***a**/α	NA	27/73	14/86	NA

Table 2. GWAS analysis reveals loci associated with the clinical phenotype in VNBI and VNBII. Two GWAS analyses were conducted. In the first, variants under 5% frequency were combined by gene or intergenic region (rare) while variants over 5% frequency were treated independently (common). In the second analysis, loss-of-function mutations were identified and combined by gene (LOF). Both analyses were conducted using GEMMA corrected for population stratification with a relatedness matrix (Zhou and Stephens 2012). The 10 most significant loci according to the score test across both analyses are shown.

P value	Hit Type	Locus	Genes(s)
2.02×10^{-7}	LOF	CNAG_01348	cyanate hydratase
3.00×10^{-7}	rare	intergenic: CNAG_07989-CNAG_07661	velvet family protein, deletion prevents filamentation (Chacko et al. 2015); hypothetical protein
3.22×10^{-7}	common	intergenic: CNAG_07972-CNAG_06879	chromosome 3 centromeric region
1.19×10^{-6}	rare	CNAG_05440	pyridoxamine 5'-phosphate oxidase *PDX3*, downregulated under oxidative stress (Upadhya et al. 2013)
2.44×10^{-6}	rare	CNAG_03374	DNA replication complex GINS protein *PSF1*
2.91×10^{-6}	common	intergenic: CNAG_06765-CNAG_06764	low mating performance *LMP1*, deletion is avirulent (Janbon et al. 2014); short chain dehydrogenase *FOX21*, downregulated under oxidative stress (Upadhya et al. 2013)
3.99×10^{-6}	common	CNAG_06921	hypothetical protein
5.07×10^{-6}	common	CNAG_05037	hypothetical protein
5.62×10^{-6}	rare	CNAG_07427	Casein Kinase *CCK2*, related to *CCK1* which is essential for virulence (Wang et al. 2011)
6.00×10^{-6}	rare	CNAG_01464	Flavohemoglobin *FHB1*, nitrosative stress response (de Jesús-Berríos et al. 2003)

Table 3. Inositol transport and utilization genes under selection in individual lineages. Genes were present in the top 5% of windows evaluated with the CLR test for each lineage.

H99 Locus	Gene	VNI	VNBI	VNBII
CNAG_00097	inositol transporter *ITR1*	+	-	+
CNAG_04552	inositol transporter *ITR1A*	+	+	+
CNAG_00864	inositol transporter *ITR2*	-	+	-
CNAG_05377	inositol transporter *ITR3*	+	-	-
CNAG_00867	inositol transporter *ITR3A*	-	+	-
CNAG_05667	inositol transporter *ITR13B*	-	+	-
CNAG_05329	*myo*-inositol 2-dehydrogenase	+	-	+
CNAG_05320	inositol 2-dehydrogenase	+	+	-
CNAG_05321	inositol 2-dehydrogenase	+	+	-
CNAG_06530	inositol 2-dehydrogenase	+	-	-
CNAG_05316	inositol oxygenase	+	-	-

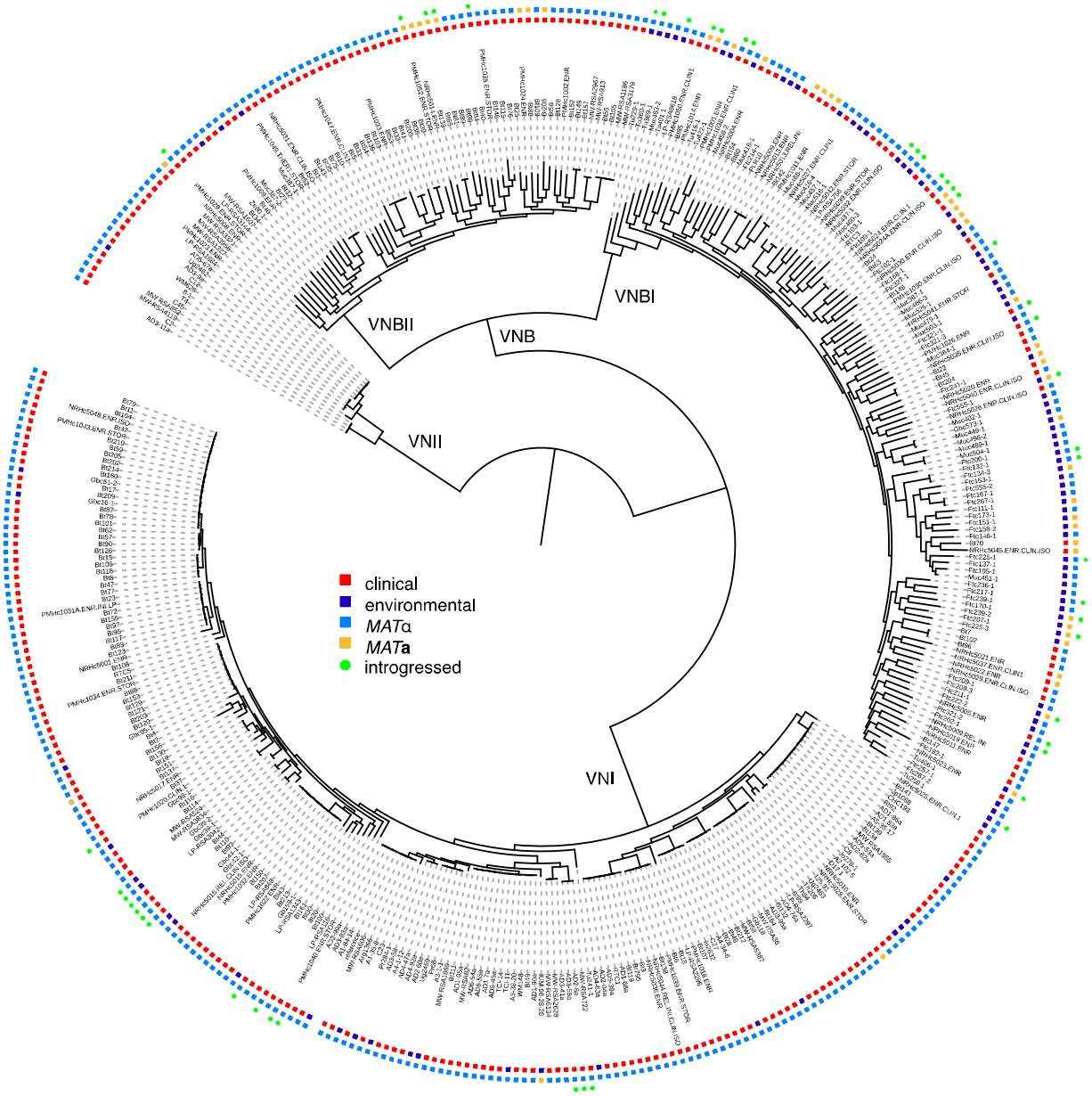

Figure 1. Phylogenomic analysis reveals a deep split within the VNB lineage. We propose that VNB be divided into two lineages: VNBI and VNBII. The phylogeny was estimated from 1,069,080 segregating sites using RAxML (Stamatakis 2014), and the tree was rooted with VNII as the outgroup. All lineages (VNI, VNII, VNBI and VNBII) had 100% bootstrap support. Isolation source (clinical vs. environmental), mating type (*MAT*α and *MAT***a**), and presence of >= 50 kb of introgressions from other lineages are also shown. VNBI samples were enriched for environmental isolation sources relative to VNBII ($p < 0.0001$, Fisher's exact test).

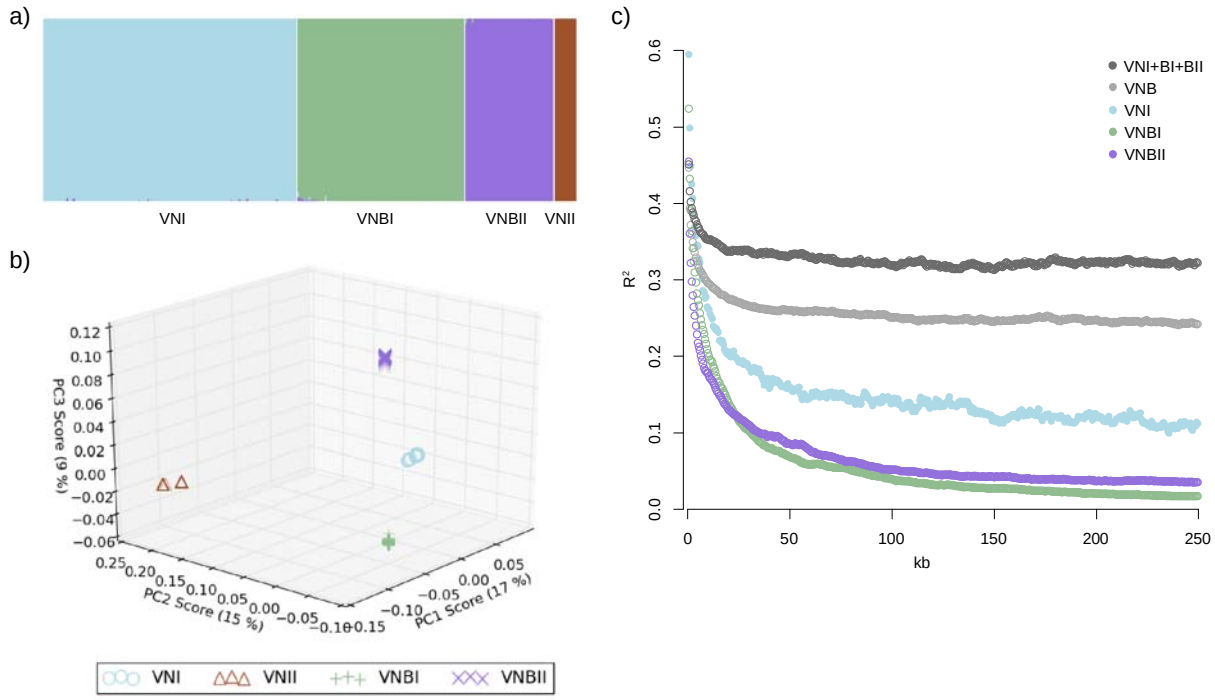

Figure 2. *C. neoformans* var. *grubii* is separated into four distinct non-recombining lineages: VNI, VNII, VNBI and VNBII. A) Results of STRUCTURE analysis (k=4) shows separation of all four lineages; B) Top three principle components from PCA analysis. The third principle component clearly separates VNBI from VNBII; C) Decay of linkage disequilibrium (LD) shows similar rates of recombination within groups and lack of recombination between groups. LD (R^2) was calculated for all pairs of SNPs separated by 0–250 kb and then averaged for every 500 bp. LD values for each window were then calculated by averaging over all pairwise calculations in the window. The chromosome with the mating type locus was excluded from the calculations.

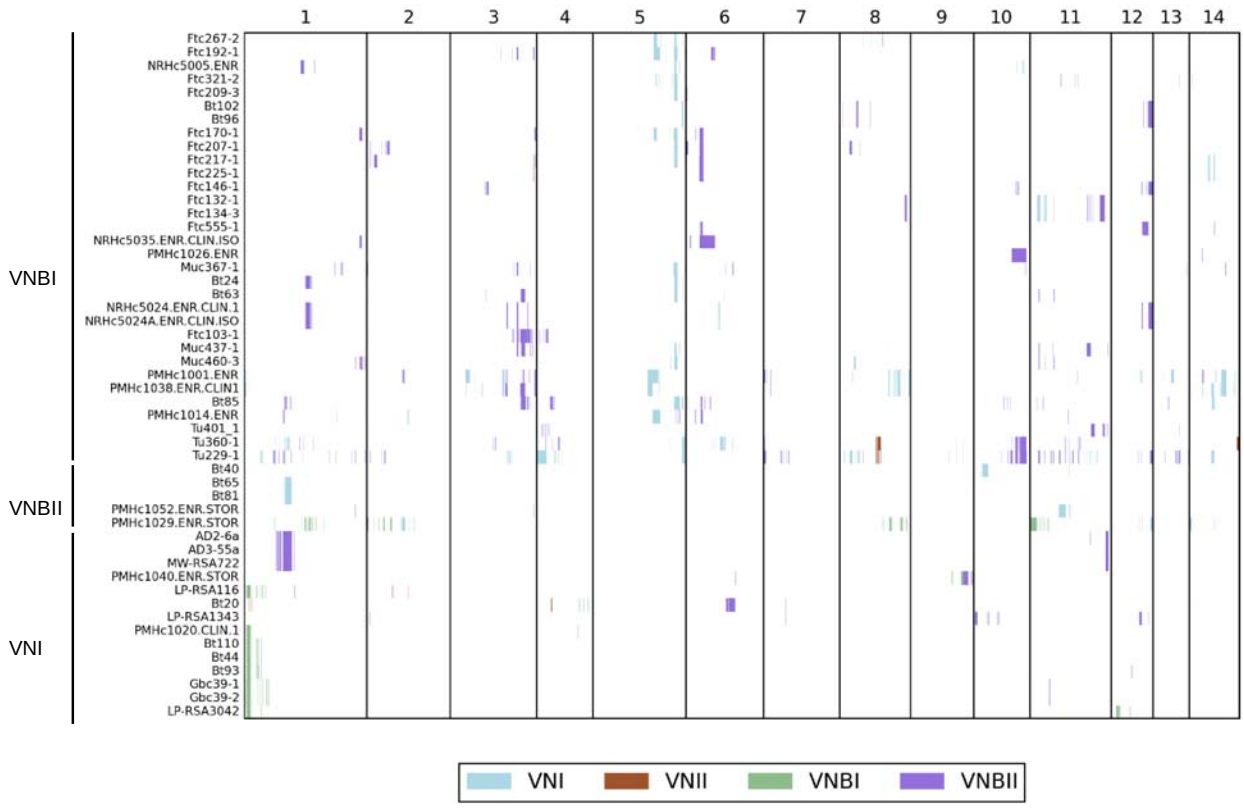

Figure 3. Small introgressions between VNI, VNII, VNBI, and VNBII are dispersed throughout the genome and phylogeny. Introgressions were detected by running STRUCTURE (k=4) on 500 kb blocks, excluding the mating type locus, and the group ancestry of each 5 kb within each block was identified. Recipient genomes are shown on the y-axis while genomic position is shown on the x-axis. White indicates non-introgressed regions while colored blocks indicate introgressed regions. The color key shows which color corresponds to which donor group.

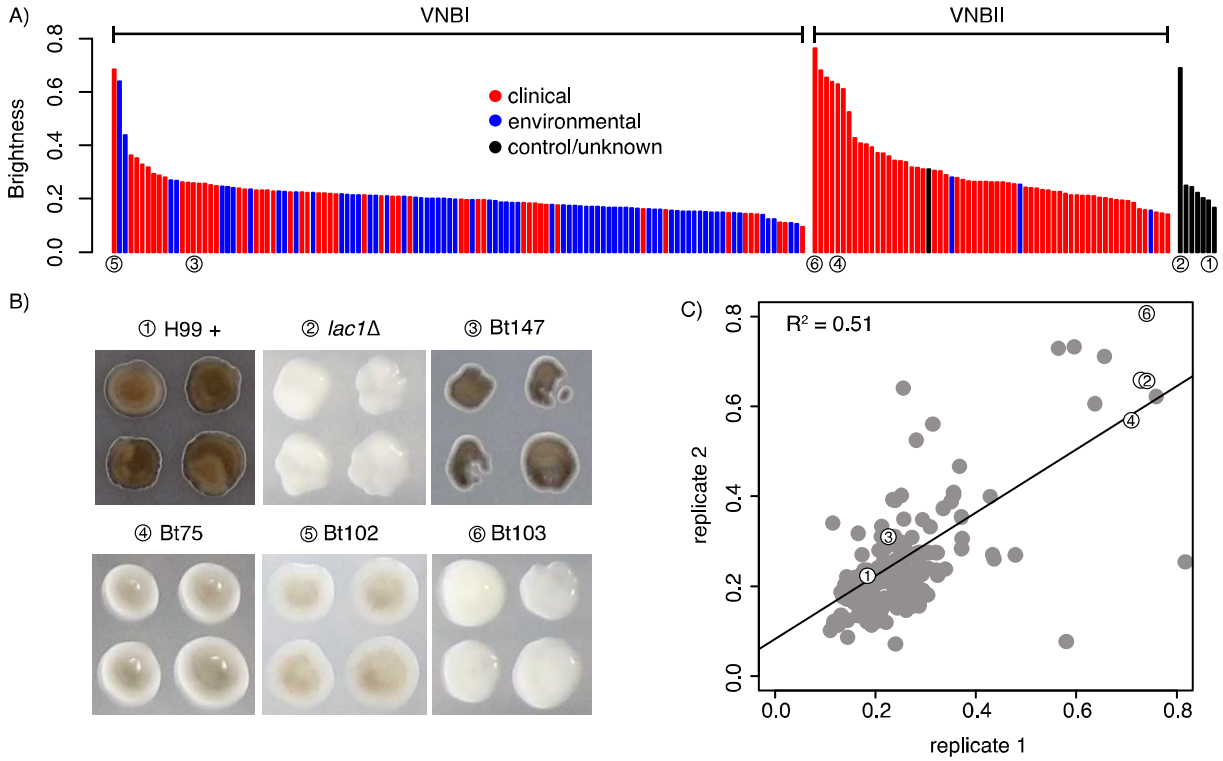

Figure 4. Phenotypic and GWAS analyses demonstrate that VNBI isolates have a greater ability to melanize than VNBII isolates and identify *BZP4*-deficient isolates as having reduced melanization capacity. A) Isolates were provided with L-DOPA and colony brightness was assayed; isolates with the lowest brightness are the most melanized. Clinical isolates are shown in red, environmental isolates are shown in blue, control/unknown isolates are shown in black, and isolates described in (B) are underscored with numbered circles. VNBI clinical isolates were significantly more melanized than VNBII clinical isolates ($p < 0.00017$) and VNBI environmental isolates were significantly more melanized than VNBI clinical isolates ($p < 0.00047$). For the latter comparison, the least melanized (brightness >= 0.6) samples were excluded to prevent an effect of sampling bias. B) GWAS analysis identified loss-of-function mutations in *BZP4* as being associated with a lack of melanization. The four isolates with *BZP4* loss-of-function mutations are shown here in the L-DOPA assay, along with the positive control H99 and the negative control *lac1*Δ. C) A strong correlation of melanization was present between replicates. Isolates shown in (B) are indicated with numbered circles.

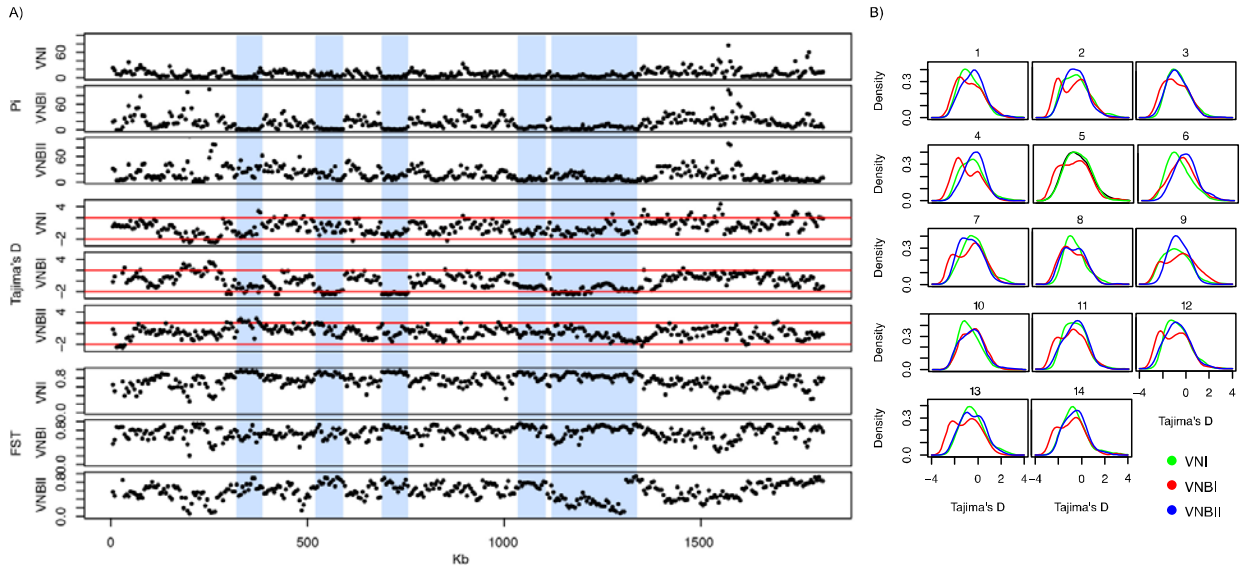

Figure 5. Population genomic analysis revealed long tracts of low genetic diversity and Tajima's D in VNI and VNBI. A) Low diversity tracts are shown here in chromosome 5. In VNBI, these tracts are interspersed between regions of high diversity. Vertical blue bars highlight these areas of reduced diversity and Tajima's D in VNBI, which are generally accompanied by high F_{ST} between populations. Statistics shown include π, Tajima's D, and F_{ST}. B) Density distribution of Tajima's D across all 14 chromosomes for VNI, VNBI, and VNBII. VNI and VNBII show predominantly unimodal distributions for most chromosomes, while VNBI shows a bimodal distribution for all chromosomes except 6.

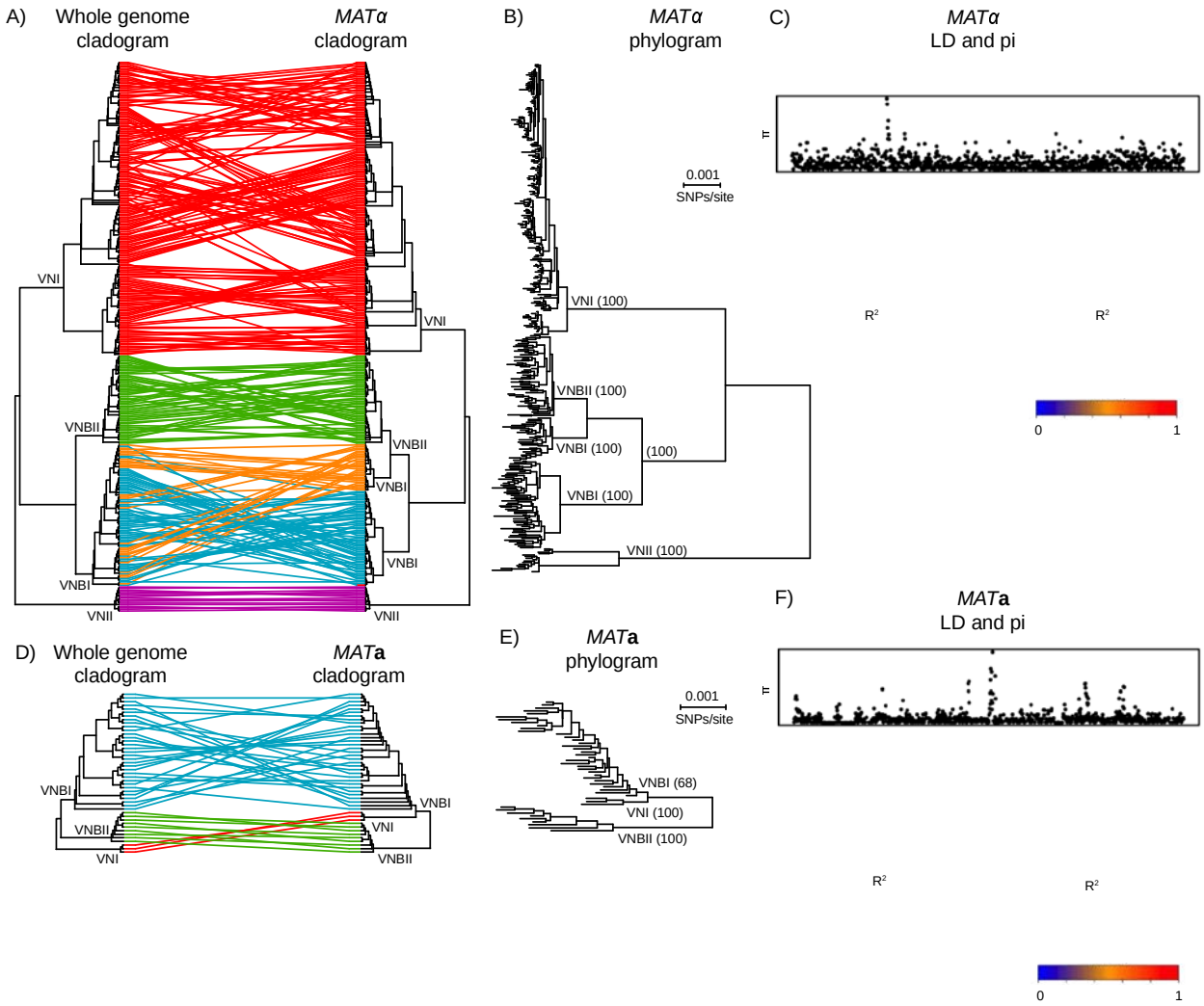

Figure 6. Phylogenetic and linkage analyses reveal distinct evolutionary trajectories of the *MAT* locus alleles. A) The topology of the whole-genome phylogeny and *MAT*α phylogeny are compared as cladograms. VNBI isolates contain two distinct α alleles; B) *MAT*α phylogram showing branch lengths and bootstrap support for major clades; C) Linkage disequilibrium (R^2) and diversity (π) across the *MAT*α locus; D) Topology of the whole-genome phylogeny and *MAT***a** phylogeny are compared as cladograms; E) *MAT***a** phylogram showing branch lengths and bootstrap support for major clades. VNI is close sister-group to VNBI, in contrast to the whole-genome phylogeny, and there is limited support for monophyly of the VNI *MAT***a** allele with respect to VNBI (bootstrap support: 68%); F) Linkage disequilibrium (R^2) and diversity (π) across the *MAT***a** locus.

References

Brown JCS, Nelson J, VanderSluis B, Deshpande R, Butts A, Kagan S, Polacheck I, Krysan DJ, Myers CL, Madhani HD. 2014. Unraveling the biology of a fungal meningitis pathogen using chemical genetics. *Cell* **159**: 1168–1187.

Brown SP, Cornforth DM, Mideo N. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends Microbiol* **20**: 336–342.

Bulmer GS, Sans MD, Gunn CM. 1967. *Cryptococcus neoformans* I. Nonencapsulated mutants. *J Bacteriol* **94**: 1475–1479.

Chacko N, Zhao Y, Yang E, Wang L, Cai JJ, Lin X. 2015. The lncRNA RZE1 controls cryptococcal morphological transition. *PLOS Genet* **11**: e1005692.

Chayakulkeeree M, Perfect JR. 2006. Cryptococcosis. *Infect Dis Clin North Am* **20**: 507–544, v–vi.

Chen Y, Litvintseva AP, Frazzitta AE, Haverkamp MR, Wang L, Fang C, Muthoga C, Mitchell TG, Perfect JR. 2015. Comparative analyses of clinical and environmental populations of *Cryptococcus neoformans* in Botswana. *Mol Ecol* **24**: 3559–3571.

Chow EWL, Morrow CA, Djordjevic JT, Wood IA, Fraser JA. 2012. Microevolution of *Cryptococcus neoformans* driven by massive tandem gene amplification. *Mol Biol Evol* **29**: 1987–2000.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**: 80–92.

Cox GM, McDade HC, Chen SCA, Tucker SC, Gottfredsson M, Wright LC, Sorrell TC, Leidich SD, Casadevall A, Ghannoum MA, et al. 2001. Extracellular phospholipase activity is a virulence factor for *Cryptococcus neoformans*. *Mol Microbiol* **39**: 166–175.

Cox GM, Mukherjee J, Cole GT, Casadevall A, Perfect JR. 2000. Urease as a virulence factor in experimental cryptococcosis. *Infect Immun* **68**: 443–448.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

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

D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, Jung WH, Sham A, Kidd SE, Tangen K, et al. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *mBio* **2**: e00342-10.

740 Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, Chen Y, Voelz K, Heitman
741 J, May RC, et al. 2015. Genome evolution and innovation across the four major lineages
742 of *Cryptococcus gattii*. *mBio* **6**: e00868-15.

743 Ferreira-Paim K, Andrade-Silva L, Fonseca FM, Ferreira TB, Mora DJ, Andrade-Silva J, Khan A,
744 Dao A, Reis EC, Almeida MTG, et al. 2017. MLST-Based Population Genetic Analysis in
745 a Global Context Reveals Clonality amongst *Cryptococcus neoformans* var. *grubii* VNI
746 Isolates from HIV Patients in Southeastern Brazil. *PLoS Negl Trop Dis* **11**: e0005223.

747 Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A,
748 Ambrogio L, et al. 2011. A scalable, fully automated process for construction of
749 sequence-ready human exome targeted capture libraries. *Genome Biol* **12**: R1.

750 Fisher SK, Novak JE, Agranoff BW. 2002. Inositol and higher inositol phosphates in neural
751 tissues: homeostasis, metabolism and functional significance. *J Neurochem* **82**: 736–
752 754.

753 Fromtling RA, Abruzzo GK, Ruiz A. 1989. Virulence and antifungal susceptibility of
754 environmental and clinical isolates of *Cryptococcus neoformans* from Puerto Rico.
755 *Mycopathologia* **106**: 163–166.

756 Fu C, Sun S, Billmyre RB, Roach KC, Heitman J. 2015. Unisexual versus bisexual mating in
757 *Cryptococcus neoformans*: consequences and biological impacts. *Fungal Genet Biol* **78**:
758 65–75.

759 Geddes JMH, Croll D, Caza M, Stoynov N, Foster LJ, Kronstad JW. 2015. Secretome profiling
760 of *Cryptococcus neoformans* reveals regulation of a subset of virulence-associated
761 proteins and potential biomarkers by protein kinase A. *BMC Microbiol* **15**: 206.

762 Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, Falk R, Parnmen S,
763 Lumbsch HT, Boekhout T. 2015. Recognition of seven species in the *Cryptococcus*
764 *gattii*/*Cryptococcus neoformans* species complex. *Fungal Genet Biol* **78**: 16–48.

765 Hiremath SS, Chowdhary A, Kowshik T, Randhawa HS, Sun S, Xu J. 2008. Long-distance
766 dispersal and recombination in environmental populations of *Cryptococcus neoformans*
767 var. *grubii* from India. *Microbiology* **154**: 1513–1524.

768 Hsueh Y-P, Idnurm A, Heitman J. 2006. Recombination hotspots flank the *Cryptococcus*
769 mating-type locus: implications for the evolution of a fungal sex chromosome. *PLoS*
770 *Genet* **2**: e184.

771 Janbon G, Ormerod KL, Paulet D, Byrnes EJ III, Yadav V, Chatterjee G, Mullapudi N, Hon C-C,
772 Billmyre RB, Brunel F, et al. 2014. Analysis of the genome and transcriptome of
773 *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and
774 microevolution leading to virulence attenuation. *PLoS Genet* **10**: e1004261.

775 Jung K-W, Yang D-H, Maeng S, Lee K-T, So Y-S, Hong J, Choi J, Byun H-J, Kim H, Bang S, et
776 al. 2015. Systematic functional profiling of transcription factor networks in *Cryptococcus*
777 *neoformans*. *Nat Commun* **6**: 6757.

778 Kmetzsch L, Staats CC, Simon E, Fonseca FL, Oliveira DL de, Sobrino L, Rodrigues J, Leal AL,
779 Nimrichter L, Rodrigues ML, et al. 2010. The vacuolar Ca2+ exchanger Vcx1 Is involved
780 in calcineurin-dependent Ca2+ tolerance and virulence in *Cryptococcus neoformans*.
781 *Eukaryot Cell* **9**: 1798–1805.

782 Kozubowski L, Lee SC, Heitman J. 2009. Signalling pathways in the pathogenesis of
783 *Cryptococcus*. *Cell Microbiol* **11**: 370–380.

784 Kwon-Chung KJ, Bennett JE, Wickes BL, Meyer W, Cuomo CA, Wollenburg KR, Bicanic TA,
785 Castañeda E, Chang YC, Chen J, et al. 2017. The Case for Adopting the "Species
786 Complex" Nomenclature for the Etiologic Agents of Cryptococcosis. *mSphere* **2**: e00357-
787 16.

788 Kwon-Chung KJ, Rhodes JC. 1986. Encapsulation and melanin formation as indicators of
789 virulence in *Cryptococcus neoformans*. *Infect Immun* **51**: 218–223.

790 Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J. 2002. Mating-
791 type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes.
792 *Eukaryot Cell* **1**: 704–718.

793 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
794 *ArXiv13033997 Q-Bio*. http://arxiv.org/abs/1303.3997 (Accessed April 5, 2016).

795 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
796 Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map
797 format and SAMtools. *Bioinformatics* **25**: 2078–2079.

798 Lin X, Hull CM, Heitman J. 2005. Sexual reproduction between partners of the same mating
799 type in *Cryptococcus neoformans*. *Nature* **434**: 1017–1021.

800 Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A,
801 Koufopanou V, et al. 2009. Population genomics of domestic and wild yeasts. *Nature*
802 **458**: 337–341.

803 Litvintseva AP, Carbone I, Rossouw J, Thakur R, Govender NP, Mitchell TG. 2011. Evidence
804 that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have
805 evolved in Africa. *PLOS ONE* **6**: e19688.

806 Litvintseva AP, Marra RE, Nielsen K, Heitman J, Vilgalys R, Mitchell TG. 2003. Evidence of
807 sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-
808 Saharan Africa. *Eukaryot Cell* **2**: 1162–1168.

809 Litvintseva AP, Mitchell TG. 2009. Most environmental isolates of *Cryptococcus neoformans*
810 var. *grubii* (Serotype A) are not lethal for mice. *Infect Immun* **77**: 3188–3195.

811 Litvintseva AP, Thakur R, Vilgalys R, Mitchell TG. 2006. Multilocus sequence typing reveals
812 three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (Serotype A),
813 including a unique population in Botswana. *Genetics* **172**: 2223–2238.

814 Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. 2008. Systematic genetic
815 analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. *Cell* **135**:
816 174–188.

817 Liu T-B, Xue C. 2014. Fbp1-mediated ubiquitin-proteasome pathway controls *Cryptococcus*
818 *neoformans* virulence by regulating fungal intracellular growth in macrophages. *Infect*
819 *Immun* **82**: 557–568.

820 Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M,
821 Anderson IJ, Fraser JA, et al. 2005. The genome of the basidiomycetous yeast and
822 human pathogen *Cryptococcus neoformans*. *Science* **307**: 1321–1324.

823 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
824 D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce
825 framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–
826 1303.

827 Morrow CA, Lee IR, Chow EWL, Ormerod KL, Goldinger A, Byrnes EJ, Nielsen K, Heitman J,
828 Schirra HJ, Fraser JA. 2012. A Unique Chromosomal Rearrangement in the
829 *Cryptococcus neoformans* var. *grubii* Type Strain Enhances Key Phenotypes Associated
830 with Virulence. *mBio* **3**: e00310-11.

831 Ngamskulrungroj P, Gilgado F, Faganello J, Litvintseva AP, Leal AL, Tsui KM, Mitchell TG,
832 Vainstein MH, Meyer W. 2009. Genetic Diversity of the *Cryptococcus* Species Complex
833 Suggests that *Cryptococcus gattii* Deserves to Have Varieties. *PLOS ONE* **4**: e5862.

834 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for
835 selective sweeps using SNP data. *Genome Res* **15**: 1566–1575.

836 O'Meara TR, Alspaugh JA. 2012. The *Cryptococcus neoformans* capsule: a sword and a shield.
837 *Clin Microbiol Rev* **25**: 387–408.

838 Park BJ, Wannemuehler KA, Marston BJ, Govender N, Pappas PG, Chiller TM. 2009.
839 Estimation of the current global burden of cryptococcal meningitis among persons living
840 with HIV/AIDS. *AIDS Lond Engl* **23**: 525–530.

841 Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLOS Genet* **2**:
842 e190.

843 Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient
844 swiss army knife for population genomic analyses in R. *Mol Biol Evol* **31**: 1929–1936.

845 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees
846 for Large Alignments. *PLoS ONE* **5**: e9490.

847 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus
848 genotype data. *Genetics* **155**: 945–959.

849 Salas SD, Bennett JE, Kwon-Chung KJ, Perfect JR, Williamson PR. 1996. Effect of the laccase
850 gene CNLAC1, on virulence of *Cryptococcus neoformans*. *J Exp Med* **184**: 377–386.

851 Shea JM, Henry JL, Poeta MD. 2006. Lipid metabolism in *Cryptococcus neoformans*. *FEMS*
852 *Yeast Res* **6**: 469–479.

853 Shin J-H, Blay S, McNeney B, Graham J. 2006. LDheatmap: an R function for graphical display
854 of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw*
855 **16**. https://www.jstatsoft.org/article/view/v016c03 (Accessed August 12, 2016).

856 Specht CA, Lee CK, Huang H, Tipper DJ, Shen ZT, Lodge JK, Leszyk J, Ostroff GR, Levitz SM.
857 2015. Protection against experimental cryptococcosis following vaccination with glucan
858 particles containing *Cryptococcus* alkaline extracts. *mBio* **6**: e01905-15.

859 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
860 large phylogenies. *Bioinformatics* **30**: 1312–1313.

861 Steenbergen JN, Shuman HA, Casadevall A. 2001. *Cryptococcus neoformans* interactions with
862 amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in
863 macrophages. *Proc Natl Acad Sci* **98**: 15245–15250.

864 Stewart JE, Timmer LW, Lawrence CB, Pryor BM, Peever TL. 2014. Discord between
865 morphological and phylogenetic species boundaries: incomplete lineage sorting and
866 recombination results in fuzzy species boundaries in an asexual fungal pathogen. *BMC*
867 *Evol Biol* **14**: 38.

868 Sun S, Billmyre RB, Mieczkowski PA, Heitman J. 2014. Unisexual reproduction drives meiotic
869 recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*.
870 *PLoS Genet* **10**: e1004849.

871 Sun S, Hsueh Y-P, Heitman J. 2012. Gene conversion occurs within the mating-type locus of
872 *Cryptococcus neoformans* during sexual reproduction. *PLoS Genet* **8**: e1002810.

873 Tajima F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* **123**:
874 597–601.

875 Upadhya R, Campbell LT, Donlin MJ, Aurora R, Lodge JK. 2013. Global transcriptome profile of
876 *Cryptococcus neoformans* during exposure to hydrogen peroxide induced oxidative
877 stress. *PLOS ONE* **8**: e55110.

878 Vanhove M, Beale MA, Rhodes J, Chanda D, Lakhi S, Kwenda G, Molloy S, Karunaharan N,
879 Stone N, Harrison TS, et al. 2016. Genomic epidemiology of *Cryptococcus* yeasts
880 identifies adaptation to environmental niches underpinning infection across an African
881 HIV/AIDS cohort. *Mol Ecol*. doi:10.1111/mec.13891.

882 Wagih O, Parts L. 2014. gitter: a robust and accurate method for quantification of colony sizes
883 from plate images. *G3 GenesGenomesGenetics* **4**: 547–552.

884 Wang L, Zhai B, Lin X. 2012. The link between morphotype transition and virulence in
885 *Cryptococcus neoformans*. *PLOS Pathog* **8**: e1002765.

886 Wang Y, Liu T-B, Patel S, Jiang L, Xue C. 2011. The casein kinase I protein Cck1 regulates
887 multiple signaling pathways and is essential for cell integrity and fungal virulence in
888 *Cryptococcus neoformans*. *Eukaryot Cell* **10**: 1455–1464.

889 Xue C. 2012. *Cryptococcus* and beyond—inositol utilization and its implications for the
890 emergence of fungal virulence. *PLoS Pathog* **8**: e1002869.

891 Xue C, Liu T, Chen L, Li W, Liu I, Kronstad JW, Seyfang A, Heitman J. 2010. Role of an
892 expanded inositol transporter repertoire in *Cryptococcus neoformans* sexual
893 reproduction and virulence. *mBio* **1**: e00084-10.

894 Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association
895 studies. *Nat Genet* **44**: 821–824.

896