MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 3559–3571 doi: 10.1111/mec.13260

Comparative analyses of clinical and environmental populations of *Cryptococcus neoformans* in Botswana

YUAN CHEN,*† ANASTASIA P. LITVINTSEVA,†[1] AUBREY E. FRAZZITTA,*[2] MIRIAM R. HAVERKAMP,‡§ LIUYANG WANG,* CHARLES FANG,* CHARLES MUTHOGA,§ THOMAS G. MITCHELL† and JOHN R. PERFECT*

Division of Infectious Diseases, Department of Medicine, Duke University Medical Center, 163 Hanes House, Trent Drive, Durham, NC 27710, USA, †Department of Molecular Genetics and Microbiology, Duke University Medical Center, RM 214 Jones Building, Research Drive, Durham, NC 27710, USA, ‡Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, PA 19104, USA, §Botswana–UPenn Partnership, National Health Laboratory, PO Box AC 157 ACH, Plot 5353 Extension 10 Church Road, Gaborone, Botswana

Abstract

***Cryptococcus neoformans* var. *grubii* (*Cng*) is the most common cause of fungal meningitis, and its prevalence is highest in sub-Saharan Africa. Patients become infected by inhaling airborne spores or desiccated yeast cells from the environment, where the fungus thrives in avian droppings, trees and soil. To investigate the prevalence and population structure of *Cng* in southern Africa, we analysed isolates from 77 environmental samples and 64 patients. We detected significant genetic diversity among isolates and strong evidence of geographic structure at the local level. High proportions of isolates with the rare *MATa* allele were observed in both clinical and environmental isolates; however, the mating-type alleles were unevenly distributed among different subpopulations. Nearly equal proportions of the *MATa* and *MATα* mating types were observed among all clinical isolates and in one environmental subpopulation from the eastern part of Botswana. As previously reported, there was evidence of both clonality and recombination in different geographic areas. These results provide a foundation for subsequent genomewide association studies to identify genes and genotypes linked to pathogenicity in humans.**

Keywords: *Cryptococcus neoformans*, genotyping, population structure, recombination

Received 16 March 2015; revision received 19 May 2015; accepted 2 June 2015

Introduction

Cryptococcosis is a leading cause of mortality in patients with HIV/AIDS and accounts for 13–40% of the annual death rate among these patients in areas of sub-Saharan Africa (Park *et al.* 2009). Most cases of cryptococcosis are caused by *Cryptococcus neoformans* var. *grubii* (*Cng*), an opportunistic basidiomycetous yeast, which has been isolated from avian droppings, trees and soil worldwide. Previous studies have demonstrated that inhaling airborne basidiospores and/or desiccated yeast cells can lead to an initial pulmonary infection with subsequent dissemination to the central nervous system and the manifestation of meningoencephalitis (Giles *et al.* 2009; Velagapudi *et al.* 2009; Botts & Hull 2010). Because transmission of cryptococcosis among humans and other animals is extremely rare, the virulence traits that enable *Cryptococcus* to infect humans undoubtedly evolved in the environment. Therefore, to investigate the genetic profile associated with cryptococcal meningitis, we contemporaneously collected environmental and clinical isolates of *Cng*, determined their genotypes and compared the genetic structure of these populations.

Previously, we described genetic diversity in environmental populations of *C. neoformans* in both the United States and southern Africa (Litvintseva *et al.* 2005; Lit-

Correspondence: Yuan Chen, Fax: +(919) 684 8902;
E-mail: ychenbioinfo@gmail.com
[1] Present address: Mycotic Diseases Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
[2] Present address: School of Medicine, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

vintseva & Mitchell 2012). In the United States, *Cng* isolates are highly clonal, lack geographic structure, are dominated by isolates with a single mating type, *MAT*α, and are primarily associated with pigeon habitats and desiccated pigeon droppings (Litvintseva *et al.* 2005). The same genetic features are also characteristic of the global population of *Cng*, which is dominated by isolates with a relatively small number of molecular genotypes and the *MAT*α mating type (Litvintseva *et al.* 2005). However, isolates in southern Africa also include a unique population of *Cng* that is comprised of genetically diverse isolates and includes isolates with both mating types, *MAT***a** and *MAT*α (Litvintseva *et al.* 2011).

Multiple genetic analyses of globally distributed isolates of *Cng* have identified three distinct molecular types or subpopulations, VNI, VNII and VNB. These differentiated phylogenetic clades may represent cryptic species. VNI isolates are prevalent worldwide in the environment and responsible for most cases of cryptococcosis outside Africa. VNII isolates are globally distributed but rare, and VNB isolates are primarily restricted to sub-Saharan Africa (Litvintseva *et al.* 2006).

Our previous results demonstrated that the African population of *Cng* includes isolates with genotypes of the globally distributed VNI and VNII molecular types. Africa also harbours unique isolates of molecular type VNI and populations of molecular-type VNB, which are unique to southern Africa and associated with trees rather than avian habitats (Litvintseva *et al.* 2006). We also detected an unusually high proportion of isolates with the rare *MAT***a** mating type in the African population, which showed evidence of sexual reproduction and recombination. However, these previous studies were based on analyses of environmental isolates from South Africa that were isolated from pigeon habitats in the metropolitan centres and a sample of uniquely African isolates collected from a small geographic area on the eastern border of Botswana (Litvintseva *et al.* 2011). Here, we report results of the systematic environmental and clinical sampling in Botswana, which is one of the highest HIV burdened countries in the world [Joint United Nations Programme on HIV/AIDS (UNAIDS 2013)], to describe the prevalence and population structure of *C. neoformans* and identify the ecological niche and a potential environmental reservoir of this pathogen.

Materials and methods

Environmental and clinical isolates of Cryptococcus

To determine the environmental prevalence and distribution of *Cryptococcus* in Botswana, we collected a total of 614 samples in April and May, 2012: 571 samples were obtained from trees, 37 were avian droppings, and six were soil samples (Table S1, Supporting information). These samples were collected from rural and urban areas, parks and nature preserves, in regions around the cities of Gaborone, Serowe, Francistown, Nata, Kasane, Maun and Ghanzi. Most of the samples were collected from trees because our previous results demonstrated a significant association between *Cng* and mopane trees. The arboreal samples included decaying woody debris within tree hollows and under the bark as well as soil under the tree canopy. These samples were collected with sterile BBL™ culture swabs containing Amies transport medium (BD Diagnostics, Franklin Lakes, NJ, USA). Samples of soil and excreta were collected in sterile plastic tubes and then stored at room temperature for 1–14 days prior to isolation. Staib agar was supplemented with 0.2 g/L chloramphenicol (Sigma-Aldrich, St. Louis, MO, USA), 0.025 g/L gentamicin (EM Science, Gibbstown, NJ, USA) and 0.1 g/L (0.1 g/10 mL 95% ethanol) biphenyl (Alfa Aesar, Ward Hill, MA, USA). Culture swabs were directly spread on duplicate Staib plates. Samples of soil and droppings were resuspended in 10 mL of sterile water by vortexing. The sediment was allowed to settle for *c.* 10 min, and a 1:10 dilution was prepared in sterile water. A 50 μL portion of the suspension was spread on Staib plates. Inoculated plates were incubated at room temperature for 3–5 days. Brown yeast colonies were selected, grown in pure culture on Staib agar plates without antibiotics, confirmed to be *Cryptococcus neoformans* by standard morphological and physiological criteria, and maintained on yeast extract–peptone–dextrose (YPD) agar (Difco, Baltimore, MD, USA) at 30 °C.

Under the auspices of Duke IRB-approved protocol and with patient consent, clinical isolates of *Cng* were obtained prospectively from specimens of cerebrospinal fluid (CSF). Beginning in 2012, patients were enrolled at the major public hospitals in the two largest cities, Princess Marina Hospital in the capital city of Gaborone, which is located in southern Botswana, and Nyangabgwe Referral Hospital in Francistown, which is in eastern Botswana. The CSF specimens were collected for routine diagnostic purposes from patients with HIV/AIDS and cryptococcal meningitis. The isolates from patients' CSF were grown on Sabouraud dextrose agar (Oxoid CM0041) with chloramphenicol (0.05 g/L). Colony-purified clinical isolates were then inoculated in 15% glycerol and stored at −80 °C until they were analysed.

A total of 304 isolates of *Cng* from Botswana were analysed in this study (Table S2, Supporting information). They were comprised of 179 environmental isolates obtained from 77 samples, and 125 clinical isolates that were cultured from the CSF specimens of 64 patients. Isolates were maintained on YPD agar medium at 30 °C.

DNA manipulations and mating types

Each isolate of *C. neoformans* was cultured for 2 days on YPD agar, and a portion was collected directly for isolation of genomic DNA. Genomic DNA was extracted and purified from each isolate using a MasterPure yeast DNA purification kit (Epicentre Biotechnologies, Madison, WI, USA) according to the manufacturer's instructions. The mating type of each isolate was determined by PCR analysis using pairs of *STE20* primers with specificity for each of the two *Cng* mating types (Li *et al.* 2012): *STE20*-Aa-f, 5′-CTAACTCTACTACACCT CACGGCA (forward orientation) and *STE20*-Aa-r, 5′-CGCACTGCAAAATAGATAAGTCTG (reverse orientation); *STE20*-Aα-f, 5′-GGCTGCAATCACAGCACCT TAC (forward orientation) and *STE20*-Aα-r, 5′-CTTCA TGACATCACTCCCCTAT (reverse orientation). Each PCR mixture contained 2.5 μL of 10× PCR buffer, 0.75 μL of 50 mM $MgCl_2$, 0.5 μL of 10 mM dNTP Mix, 1.25 μL of each primer, 0.1 μL of Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA), ~2.5 ng of genomic DNA and an appropriate volume of distilled water. The PCR was conducted with the following thermocycling conditions: initial denaturation at 95 °C for 5 min; followed by 35 cycles of 30 s at 95 °C, 30 s at 54 °C and 30 s at 72 °C; and finally 10 min at 72 °C for extension. PCR products were analysed on 1.4% agarose gels.

MLST determination

We randomly picked one isolate from each biological sample for MLST analysis. Eight routinely employed MLST loci were used to analyse genetic diversity of the isolates: *CAP59*, *GPD1*, IGS1, *LAC1*, *PLB1*, *SOD1*, *URA5* and *TEF1*. The primer sequences were used as previously described (Litvintseva *et al.* 2011; Chen *et al.* 2014) except *LAC1*, which was newly designed based on H99 genome (Janbon *et al.* 2014) and included the target region previously used. The forward primer for *LAC1* is 5′-AACATGTTCCCTGGACCTGTG, and reverse primer is 5′-ACGTGGATCTCGGGAGGA. We applied our newly developed NGMLST method in MLST genotyping, and the sequencing libraries were prepared as previously described (Chen *et al.* 2015). In brief, SMRT cell sequencing libraries were prepared using Pacific Biosciences DNA Template Prep Kit 2.0 (Cat# 001-540-835) according to the 3- or 10-kb template preparation and sequencing protocol provided by Pacific Biosciences. Instead of using magnetic beads, the amplicons were loaded by diffusion at a concentration of 300 pM and sequenced using the PacBio RS II platform. In total, three SMRT cells were used to sequence all the isolates in this study. The sequencing run used 1 × 180 min movie with P4-C2 chemistry.

Primary analysis was performed using the PACBIO SMRT ANALYSIS version 2.1 program, and the filtering parameters were as follows: minimum polymerase read quality of 0.75; minimum read length of 50 bp; and minimum subread length of 50 bp. Circular consensus sequencing (CCS) reads with less than four full passes were also filtered in further analysis. We used the MLSTEZ software program to generate all the consensus sequences of each locus and searched for heterozygous loci (Chen *et al.* 2015).

Data analyses

Sequences were aligned using MUSCLE (Edgar 2004); the alignments were edited manually and then trimmed and concatenated using in-house Python scripts. Every unique MLST allele was assigned an allelic type (AT) number at each locus and compared to the existing ATs in MLST database (http://mlst.mycologylab.org) on seven loci (*CAP59*, *GPD1*, IGS1, *LAC1*, *PLB1*, *SOD1* and *URA5*) (Meyer *et al.* 2009).

Sequence diversities were analysed by DNASPV5 (Librado & Rozas 2009). The number of segregating sites (*S*), haplotypes (*h*), haplotypic diversity (Hd), average number of nucleotide differences per site (π), Watterson's estimate of the population scaled mutation rate (θ) (Watterson 1975) and Tajima's *D* (Tajima 1989) were calculated for each locus for both environmental and clinical isolates separately.

Evolutionary descent of the Botswana *Cng* isolates was analysed by EBURST (http://eburst.mlst.net/) (Feil *et al.* 2004) based on the ATs of seven loci (*CAP59*, *GPD1*, IGS1, *LAC1*, *PLB1*, *SOD1*, *URA5*). To assess statistical support for the phylogeny, bootstrapping values were calculated from 1000 resamplings. The genetic relatedness of MLST sequences was evaluated by SMARTPCA (EIGENSOFT version 6.0 beta) (Patterson *et al.* 2006) with default parameters. To better define the relationships among different populations, we used the Bayesian clustering method STRUCTURE (Pritchard *et al.* 2000) based on the single nucleotide polymorphisms (SNP) in the MLST sequences. The analyses compared all clinical and environmental isolates as well as environmental isolates obtained from different areas. Two simulation runs were conducted for each of $K = 1$ to $K = 10$ using a length of burn-in of 10^5 with 10^6 replicates of Markov chain Monte Carlo simulation. In addition, a phylogenetic network was analysed in SPLITSTREE4 (http://splitstree.org/) with the neighbour-net algorithm (Huson & Bryant 2006).

Hierarchical analysis of molecular variance (AMOVA) was performed with ARLEQUIN 3.5 (Excoffier & Lischer 2010). The total variance was partitioned into variance among individuals within populations and variance

among populations. We first calculated the variance components between the clinical and environmental populations, and comparisons were made among three different subsets: (i) the two populations of clinical (n = 64) and environmental (n = 77) isolates; (ii) the clinical (n = 19) and environmental (n = 5) populations of VNI isolates; and (iii) the clinical (n = 33) and environmental (n = 65) populations of VNB isolates. We also calculated the variance components of different environmental isolates based on their geographical origins. Pairwise comparisons were made among the three sampling areas (Fig. 1). Using ARLEQUIN 3.5, pairwise Wright's fixation indexes (F_{ST}) were calculated for each pair of compared populations. The significance of each AMOVA and F_{ST} value was tested by a nonparametric permutation method with 1000 permutations (Excoffier & Lischer 2010).

Linkage disequilibrium among the loci was estimated by determining the index of association (I_A), $\bar{r}_d$ and the most parsimonious tree length (MPTL) (Brown *et al.* 1980; Smith *et al.* 1993; Burt *et al.* 1996). I_A and $\bar{r}_d$ values were calculated using MULTILOCUS 1.3b, and 1000 randomized data sets of shuffled alleles were used to test the null hypothesis. PAUP (Swofford 2003) execution files generated from MULTILOCUS were used to calculate the observed and randomized MPTLs, and 99.9% confidence intervals of the randomized data were calculated in R (Team 2008).

The most appropriate model and parameters of DNA sequence evolution for each MLST locus and concatenate data set were determined by JMODELTEST2 (Darriba *et al.* 2012) (Table S7, Supporting information). Neighbour-joining (NJ) phylogenetic trees were constructed based on the Kimura 2-parameter and the inferred model and parameters using PAUP* version 4.0b10 (Swofford 2003). We further used the Shimodaira–Hasegawa (SH) tests (Shimodaira & Hasegawa 1999) to assess phylogenetic incongruence between loci.

Results

Cryptococcus neoformans is widespread in Botswana and primarily associated with mopane trees

Of the 614 environmental sampling sites, 179 isolates of *Cng* were recovered from 77 locations (12.5%) (Fig. 1A and Table S1, Supporting information). Seventy-five of these positive *Cng* sites were trees, and there was one positive site each from soil and avian excreta. In addition, two isolates of *Cryptococcus gattii* were recovered from a single mopane tree (*Colophospermum mopane*) in the Francistown area. The geographical distribution of *Cng* varied in association with the regional prevalence of *C. mopane* trees. In the vicinity of Francistown, in eastern Botswana, 40 of 254 (15.7%) sites were positive, and in around the north central city of Maun, 30 of 175 (17.1%) sites were positive. *C. mopane* trees are endemic around Maun and Francistown, and all but two positive sites were associated with mopane trees. Around Gaborone, in southern Botswana, only seven of 125 sites

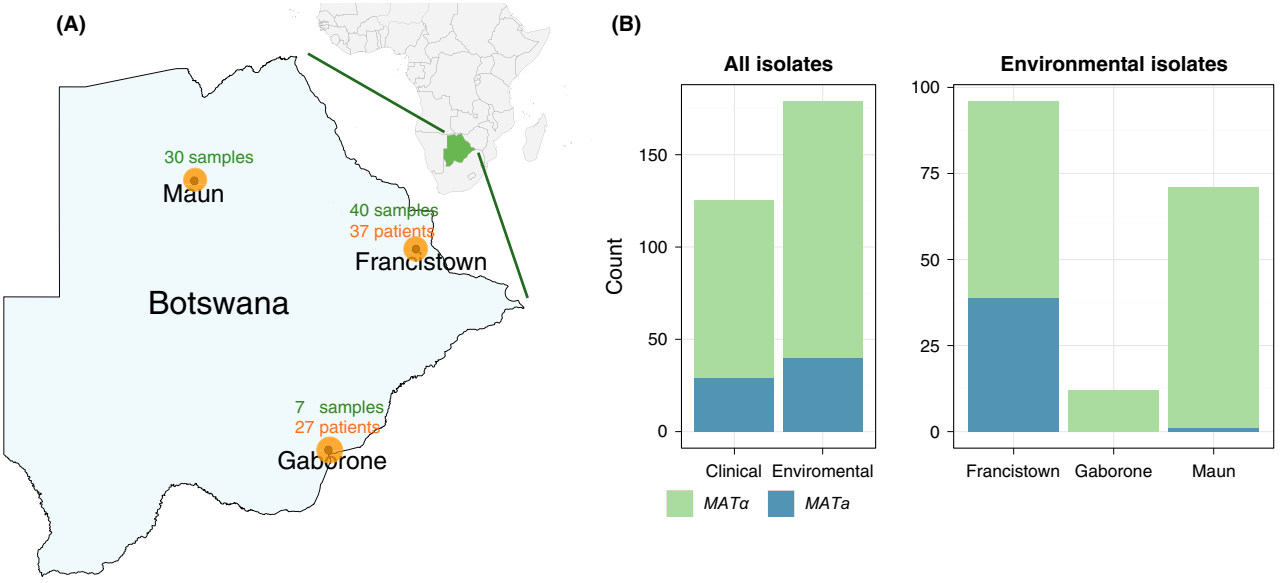

Fig. 1 (A) An outline map of Botswana showing the three geographic locations where environmental samples of *Cryptococcus neoformans* var. *grubii* were obtained and analysed in this study. (B) The ratios of *MAT*α and *MAT***a** isolates in different populations. Similar mating-type ratios were found in all clinical and environmental isolates, but different ratios were found in environmental isolates collected from different areas.

(5.6%) were positive. Mopane trees do not occur in Gaborone, but *Cng* isolates were recovered from an *Acacia* tree and a *Ziziphus mucronata* tree. Fifty-four samples were collected from the western Kalahari Desert region (n = 19), the Okavango delta salt pans in the north central area around Nata (n = 18) and areas near the eastern town of Serowe (n = 17), but no isolates were recovered from these samples.

Clinical specimens of CSF were collected from 83 HIV/AIDS patients who were diagnosed and treated at Princess Maria Hospital in Gaborone or Nyangabgwe Referral Hospital in Francistown, Botswana. From 64 of these patients with cryptococcal meningitis, we recovered 125 isolates of *C. neoformans*. The remaining 19 patients were infected with the sibling species, *C. gattii*, and 30 isolates of *C. gattii* were obtained (data not shown).

Unequal distribution of mating-type alleles in the environmental and clinical Botswana Cng populations

PCR primers specific for *STE20* region of *Cng* were used to determine the mating type of each isolate. The frequency of the rare *MAT**a*** allele in both clinical and environmental populations was higher than previously observed: 40 of the 179 (22.3%) tested environmental isolates and 29 of 125 (21.5%) clinical isolates possessed *MAT**a*** mating type. Among the environmental isolates, the distribution of *MAT**a*** mating type varied among isolates from different geographical areas (Fig. 1B). In Francistown, 39 of 96 (41.6%) isolates possessed *MAT**a***, while only one of 71 (1.5%) isolates from Maun and none of the seven isolates from Gaborone was identified to be *MAT**a***. Fisher's exact tests indicated that the frequency of the *MAT**a*** allele among *Cng* isolates from Francistown was significantly higher than its frequency among isolates from Maun ($P < 0.001$) and Gaborone ($P < 0.001$).

Unequal distribution of VNI, VNB and VNII molecular types among clinical and environmental Botswana populations

One representative *Cng* isolate was selected from each positive environmental site and patient's specimens, and these isolates were genotyped using the NGMLST method (Chen *et al*. 2015) with eight target loci (Litvintseva *et al*. 2011). In total, 141 isolates were analysed (Table S3, Supporting information). The aligned sequences of concatenated loci totalled 5421 base pairs and 202 polymorphic sites (149 parsimony informative and 53 singleton sites). Seven of the eight loci were used in the analysis of AT, and all the sequences were trimmed into the same lengths as the sequences in the MLST database (http://mlst.mycologylab.org). The seven loci (*CAP59*, *GPD1*, *LAC1*, *PLB1*, *SOD1*, *URA5* and IGS1) of 141 isolates yielded 115 ATs, and 49 of them were new to the database (Table S3, Supporting information). To compare the isolates to previous data, sequence type (ST) was generated by the alleles at each of the seven loci for each isolate. In total, 118 STs were generated for all 141 isolates, and only 9 (8.2%) of those matched other existing STs in the database, confirming previously described unique genetic make-up of *Cng* isolates from Botswana.

Table 1 Molecular type statistics in different populations

Molecular type	Clinical	Environmental			Total
		Francistown	Maun	Gaborone	
VNI	19	2	0	3	24
VNB	38	33	28	3	102
VNI/VNB	5	5	2	1	13
VNII	2	0	0	0	2
Total	64	40	30	7	141

The MLST results identified 24 isolates of VNI (17.0%), 102 isolates of VNB (69.5%) and two isolates of VNII (1.4%). In addition, 13 isolates were identified as putative VNI/VNB admixture genotypes that possessed both VNI- and VNB-specific MLST loci. Isolates of VNI and VNII molecular types were significantly overrepresented among clinical compared to environmental isolates, while VNB isolates were more prevalent in the environmental populations. Only 6.5% of the environmental isolates but 30% of clinical isolates were molecular type VNI (Table 1), and the frequency of VNI among clinical isolates was significantly higher than its frequency among environmental isolates (Fisher's exact test, $P < 0.001$).

Genetic relationships among the MLST genotypes were analysed by neighbour-net algorithm (Fig. 2). This analysis confirms the presence of two genetically isolated subpopulations, VNB and VNI, and also provides evidence of recombination and admixture within and between these two groups. Specifically, the neighbour-net algorithm confirmed recombinational events in the phylogenetic history of 13 putative VNI/VNB admixture genotypes, which were identified by comparing topologies of the dendrograms of individual MLST loci (Fig. 2). In addition, several distinct clonal lineages are discernible within each VNB and VNI group, including a previously identified subpopulation of VNB-II (Litvintseva *et al*. 2006) (Fig. 2).

Genetic structure of Botswana Cng populations

Principle component analysis (PCA) was used to assess the genetic relationships among the Botswana isolates

of two global molecular types (Litvintseva *et al.* 2006). Two patient isolates with the VNII molecular type (two STs) and five patient isolates with global VNI genotypes (three STs) were obviously genetically distinct from other isolates (Fig. S1, Supporting information). Therefore, isolates with these global genotypes were removed from subsequent analyses to focus on the uniquely diverse Botswana isolates of molecular types VNB and VNI. PCA of the remaining environmental (n = 77) and clinical (n = 57) isolates was separated into three genetically distinct groups: one group included the majority of isolates with the VNB molecular type VNB-I; the second group contained a mixture of VNB/VNI admixture genotypes and a VNB-II subgroup of the VNB isolates, which was described previously; a third group included most of the endemic isolates with VNI molecular type (Fig. 3A).

The population structure was further evaluated by STRUCTURE (Lawson *et al.* 2012). Based on the PCA results and our previous observations, three subpopulations were assumed (K = 3, VNB-I, VNB-II and VNI) and significant differences were detected between clinical and environmental populations (Fig. 4A) as well as between Gaborone and other geographic regions (Fig. 4B). STRUCTURE also confirmed the existence of isolates with VNB/VNI admixture genotypes, which were identified by phylogenetic analysis (data not shown).

AMOVAS analyses were performed and Wright's fixation indexes (F_{ST}) were calculated to compare the following groups: (i) clinical and environment isolates; (ii) VNI and VNB isolates; (iii) isolates from Francistown, Maun and Gaborone; (iv) VNB isolates with *MAT*α and *MATa* alleles. Consistent with the results of PCA and STRUCTURE analyses, AMOVA and F_{ST} indicated significant genetic differences between clinical and environmental isolates with VNB molecular type, and among isolates from different geographic areas. No significant differences were detected when VNI isolates were compared between clinical and environmental isolates (Table 2). Moreover, there was no genetic difference among the VNB isolates with different mating alleles ($F_{ST} \approx 0$).

Analyses of genetic variation within different Botswana Cng populations

The average nucleotide diversity was compared between clinical and environmental isolates and among environmental isolates from different geographic areas. Calculations of the number of segregating sides (S), haplotypes (h), haplotypic diversity (Hd), the number of nucleotide differences per site (π) and Watterson's estimate of the population scaled mutation rate (θ) revealed significantly higher genetic diversity among clinical compared to the environmental isolates (Table 3). The Tajima's D values for individual loci were not significantly different from 0, which indicated no deviation from the neutral model of evolution (Table 3). As unequal distributions of mating types

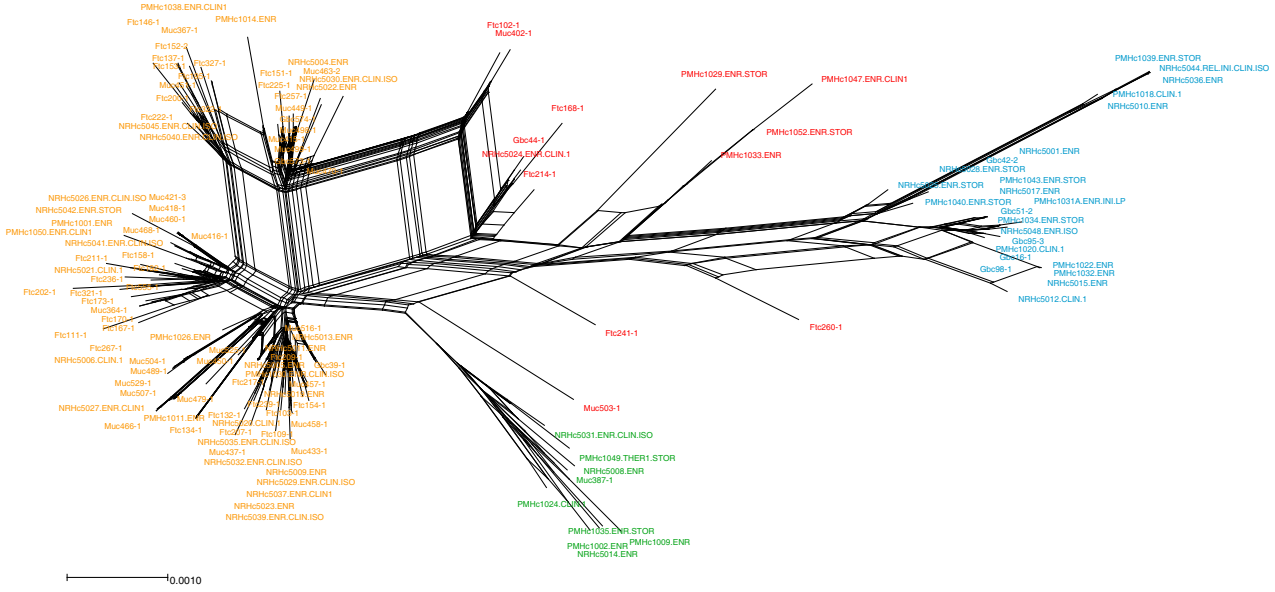

Fig. 2 SPLITSTREE obtained using concatenated sequences of the eight loci for 139 Botswana *Cryptococcus neoformans* var. *grubii* isolates (without VNII isolates). The observation that isolates are linked to each other by multiple pathways, thereby forming an interconnected network rather than a single bifurcating tree, is suggestive of recombination. VNI isolates are labelled blue; VNB-I isolates are labelled orange; VNB-II isolates are labelled green; and VNI/VNB admixture genotypes are labelled red.

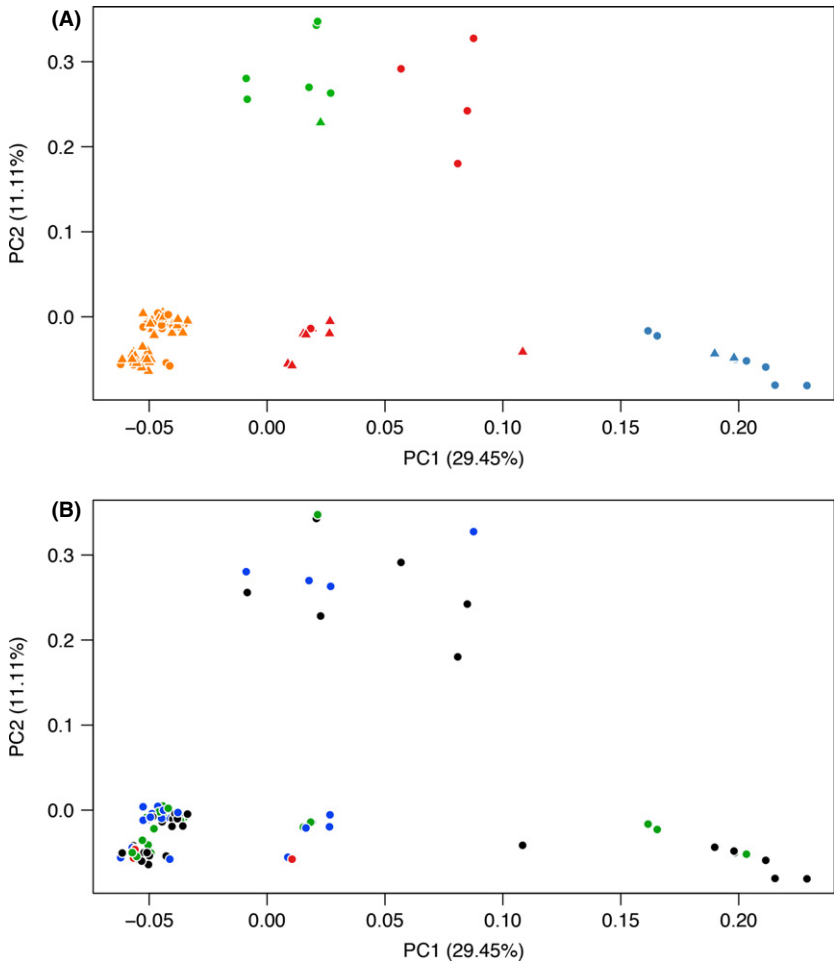

Fig. 3 The genetic relationships among 134 Botswana *Cryptococcus neoformans* var. *grubii* (*Cng*) isolates (two VNII isolates and five VNI isolates with global genotypes, which are distantly related to the Botswana isolates were removed) are visualized by principle components analysis. (A) Circles represent clinical isolates, and triangles represent environmental isolates. Different colours represent different *Cng* molecular types: blue, VNI; orange, VNB-I; green, VNB-II; red, VNB/VNI admixture genotypes. (B) Different colours represent the sources of the isolates. Black dots represent clinical isolates, and the other coloured symbols represent environmental isolates from different areas: blue, Francistown; green, Maun; red, Gaborone.

were detected in different geographic areas, we tested genetic diversities made among the isolates from different areas. Except for the isolates collected from Gaborone that showed higher diversity at the *GPD1* locus, there was no significant difference in nucleotide diversity among isolates recovered from areas with high or low prevalence of the *MAT***a** allele (Table S4, Supporting information). In addition, we compared the genetic diversity between VNI and VNB subpopulations. The result exhibits that the VNB isolates have more haplotypes than the VNI isolates in most tested loci (Table S5, Supporting information).

Evidence of recombination and clonality in Botswana Cng populations

EBURST analysis detected an extended clonal network comprised of several MLST genotypes (Fig. S2, Supporting information); however, STRUCTURE and SPLITSTREE analysis identified multiple admixture/recombination events within and between the VNI and VNB groups, as well as some evidence of clonality (Figs 2 and 4). To further evaluate the extent of clonality and recombination, we calculated index of association (I_A) (Brown *et al.* 1980) and $\bar{r}_d$ (Agapow & Burt 2001) statistics for individual subgroups (Litvintseva *et al.* 2006; Simwami *et al.* 2011). Significant linkage disequilibrium was identified in the VNB subpopulation (Table 4) and also within VNB isolates with different mating types (Table S6, Supporting information). In addition, the possibility of recombination was suggested in the clone-corrected VNI subpopulation (Table 4). The branch lengths of the most parsimonious tree tests showed the observed tree length was significantly shorter than the tree lengths of recombined data in all tested groups ($P < 0.01$), which also rejected the null hypothesis of free recombination in the population. Conversely, approximate likelihood ratio tests based on SH computations provided statistically significant support for phylogenetic conflicts in all pairwise comparisons between loci, which was consistent with recombination (Table S8, Supporting information). Overall, the results demonstrated the evidence of both recombination and clonality.

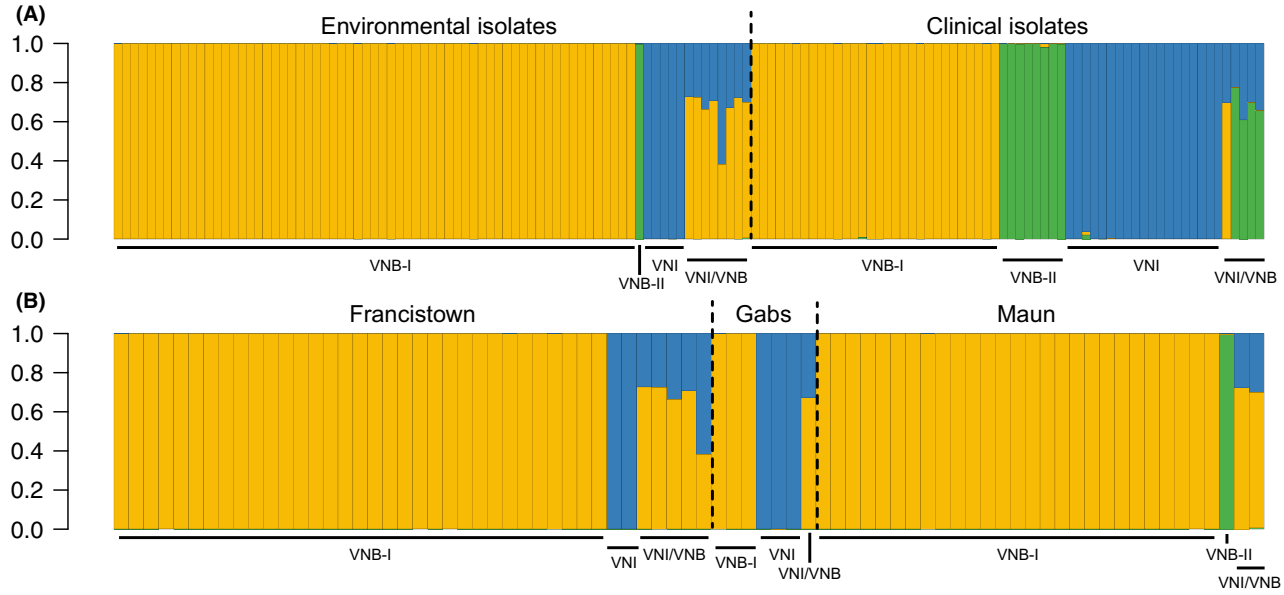

Fig. 4 Results of Botswana *Cryptococcus neoformans* var. *grubii* isolates ancestry analysis ($K = 3$) for all clinical and environmental isolates (A) and ancestry analysis for environmental isolates with different geographical origins: Francistown, Gaborone (Gabs) and Maun (B). Each individual is represented by a thin vertical line, which is partitioned into three coloured segments that represent the individual's estimated membership fractions in three clusters. Black dashed lines separate individuals of different populations. Molecular type of each isolate was labelled at the bottom.

Discussion

Cryptococcosis is a leading cause of morbidity and mortality of patients with HIV/AIDS in sub-Saharan Africa (Park *et al.* 2009). Our previous studies demonstrated that the southern African population of *Cryptococcus neoformans* is largely comprised of genetically unique isolates, which are not found elsewhere in the world, and unlike most global isolates, they are associated with endemic trees rather than with pigeon droppings (Litvintseva *et al.* 2011). However, these results were based on the analyses of a limited number of environmental isolates collected from the Tuli block, a small geographic area in the eastern part of Botswana. Here, we describe results of a broad environmental sampling conducted throughout Botswana and aimed at identifying an ecological niche of the endemic isolates. In addition, a comprehensive analysis of clinical and environmental isolates was undertaken to determine the population structure and frequency of recombination/sexual reproduction in the *Cng* population in Botswana. These results will provide a foundation for the subsequent genomewide association studies (GWAS) intended to identify genes and genotypes linked to pathogenicity and human disease.

The results of our environmental sampling indicated that *Cng* is highly prevalent in eastern and northern Botswana and significantly associated with the mopane woodland ecoregion dominated by *Colophospermum mopane* trees. In Gaborone, where mopane trees do not occur, *C. neoformans* was isolated from other tree species; however, the prevalence of *C. neoformans* in Gaborone was significantly lower, and no isolates were recovered from any of the samples obtained from areas of western and central Botswana, which are not part of the mopane woodland. The vast majority of isolates isolated from mopane trees possessed the VNB molecular type, while half of the isolates from Gaborone, where mopane do not occur, possessed the VNI molecular type, suggesting that VNB isolates are associated with the mopane woodland.

Population genetic analysis detected significant diversity among *Cng* isolates in Botswana. Of 118 STs identified in this study, only nine have been previously described in MLST Database (http://mlst.mycology-lab.org), and only five (three VNI STs and two VNII STs) were found to be the same or very similar to the strains isolated outside Africa. Furthermore, our analysis indicated a significant difference in the genetic compositions of clinical and environmental populations in Botswana. Clinical isolates were significantly more diverse and included genotypes that were either rare or absent in the environmental sample.

Of 125 ATs identified from the seven loci in this study, 48 were shared by both clinical and environmental isolates, representing 69.6% and 46.2% of all ATs among environmental and clinical isolates, respectively (Table S3, Supporting information). Furthermore, only

Table 2 AMOVAs of MLST genotypes of different *Cryptococcus neoformans* var. *grubii* populations

	d.f.	Sum of squares	Variance components (%)	F_{ST}
(i) All isolates: clinical (64), environmental (77)				
Between population	1	122.3	1.56 (10.6)	0.106***
Within population	139	1832.2	13.18 (89.4)	
Total	140	1954.5	14.74 (100)	
(ii) VNI isolates: clinical (19), environmental (5)				
Between population	1	12.5	0.82 (12.0)	0.120
Within population	22	132.6	6.03 (88.0)	
Total	23	145.2	6.85 (100)	
(iii) VNB isolates: clinical (38), environmental (64)				
Between population	1	26.3	0.38 (4.6)	0.046***
Within population	100	797.9	7.98 (95.4)	
Total	101	824.1	8.36 (100)	
(iv) Environmental isolates: Francistown (40), Maun (30)				
Between population	1	22.4	0.39 (4.2)	0.042**
Within population	68	609.4	8.96 (95.8)	
Total	69	631.8	9.35 (100)	
(v) Environmental isolates: Francistown (40), Gaborone (7)				
Between population	1	42.5	2.68 (20.3)	0.203**
Within population	45	472.8	10.51 (79.7)	
Total	46	515.3	13.19 (100)	
(vi) Environmental isolates: Maun (30), Gaborone (7)				
Between population	1	48.8	3.52 (28.6)	0.286***
Within population	35	308.0	8.80 (71.4)	
Total	36	356.8	12.32 (100)	

*$0.01 < P < 0.05$; **$0.001 < P \leq 0.01$; ***$P \leq 0.001$.

two STs were found in both environmental and clinical samples in this study (Table S1, Supporting information). Similar differences between clinical and environmental populations of *C. neoformans* in the USA have been described before (Litvintseva *et al.* 2005). The most likely explanation for the unequal distribution of genotypes in clinical and environmental populations is the limitation of environmental sampling. For example, it is possible that some isolates, which were significantly overrepresented in the clinical sample but rare or absent from the environmental sample, are restricted to specific geographic areas and/or ecological niches that have not been adequately sampled. Therefore, additional environmental sampling from broader areas of Botswana might reveal a potential link between clinical and environmental genotypes. Extensive environmental sampling in and around patients' households is needed to investigate the hypothesis of local acquisition of *Cng*. However, considering the evidence of a lengthy incubation period between infection with *C. neoformans* and the manifestation of disease (Dromer *et al.* 1992; Saha *et al.* 2007), we cannot exclude the possibility that people may become infected early in life and beyond their residence in Botswana. For example, seven isolates with cosmopolitan genotypes of VNII and VNI molecular types may have been acquired from South Africa, which has a large population of *Cng* isolates associated with urban environment and pigeon habitats (Litvintseva *et al.* 2011; Miglia *et al.* 2011; Van Wyk *et al.* 2014). Moreover, it is also possible that some isolates are more likely to cause infections.

Population genetic analyses detected evidence of a complex population structure of *Cng* in Botswana. Depending upon the size used to define a subpopulation, no fewer than three and up to five subpopulations were identified by PCA (Fig. 3). Three subpopulations were detected by STRUCTURE (Fig. 4), and three to five subpopulations or clonal lineages were detected by the neighbour-net algorithm (Fig. 2). These subgroups were supported by F_{ST} analysis and AMOVA, which confirmed that limited genetic exchange occurred among genetic subgroups and geographic regions (Table 2). However, we also observed evidence of genetic exchange between groups. Specifically, STRUCTURE and SPLITSTREE4 detected evidence of mixing between subpopulations and VNI/VNB admixture genotypes among both clinical and environmental isolates (Figs 2 and 4). These isolates are different from the previously described AD or VNB/VNI hybrids (Xu *et al.* 2000; Lin *et al.* 2009) as they possess a mixture of VNB and VNI loci with single alleles at each locus, which indicates that they are products of admixture/recombination rather than hybridization between subpopulations. This complexity of population structure must be considered in the future GWAS.

Our previous investigations, which involved relatively small populations, also detected evidence of recombination among isolates of *Cng* in Botswana (Litvintseva *et al.* 2011; 2003). Here, we have analysed a much larger number of environmental and clinical isolates, and the results highlight a complex mating history of *Cng* in Botswana that includes both sexual and asexual propagation. EBURST (Fig. S2, Supporting information) and the neighbour-net algorithm (Fig. 2) detected clonal lineages within the populations. Moreover, tests for linkage equilibrium on clone-corrected genotypes revealed significant disequilibrium among loci in the VNB isolates with different mating types, which is consistent with clonality; however, linkage equilibrium was detected in the VNI subpopulation, which may suggest recombination (Table 4). In addition, SH tests for conflict among phylogenies detected significant incongruence among all eight loci, which is indicative of recombination (Table S6, Supporting information). The result is consistent with the previous study by Hiremath *et al.* (2008) that both clonality and recombination were observed in the VNI isolates with *MAT*α allele. The STRUCTURE and SPLITSTREE analyses detected the presence of admixture genotypes with a

Table 3 Summary of polymorphism among clinical and environmental isolates at each locus

Locus	pb*	Source	S†	*h*‡	Hd§	π¶	θ**	Tajima's *D*
CAP59	561	Clinical	16	13	0.814	0.00361	0.00641	−1.31
		Environmental	13	14	0.850	0.00343	0.00472	−0.76
GPD1	558	Clinical	15	11	0.736	0.00391	0.00570	−0.93
		Environmental	7	6	0.197	0.00076	0.00256	−1.72
IGS1	853	Clinical	29	15	0.846	0.00508	0.00796	−1.17
		Environmental	8	9	0.752	0.00167	0.00198	−0.40
LAC1	672	Clinical	24	17	0.821	0.00580	0.00790	−0.84
		Environmental	16	13	0.731	0.00312	0.00487	−1.04
PLB1	632	Clinical	20	16	0.891	0.00526	0.00699	−0.66
		Environmental	7	6	0.536	0.0017	0.00225	−0.60
SOD1	609	Clinical	29	17	0.844	0.01114	0.01075	0.11
		Environmental	20	10	0.633	0.00687	0.00712	−0.11
TEF1	804	Clinical	23	16	0.888	0.00389	0.00610	−1.13
		Environmental	16	22	0.878	0.00411	0.00413	−0.01
URA5	732	Clinical	28	13	0.881	0.00923	0.00842	0.31
		Environmental	19	9	0.763	0.00725	0.00558	0.90

*Total number of sites in alignments.
†Number of segregating sites.
‡Number of haplotypes.
§Haplotypic diversity.
¶Average number of nucleotide differences per site.
**Watterson's estimate of the population scaled mutation rate, expressed per site.

Table 4 Statistical test of association among the loci in different Botswana *Cryptococcus neoformans* var. *grubii* populations

	Total samples				Clone-corrected samples††			
Population	*n*†	PcP‡	I_A§	$\bar{r}_d$¶	*n*	PcP	I_A	$\bar{r}_d$
All	141	0.869***	12.191***	0.056***	110	0.870***	11.344***	0.052***
VNI All	24	0.997***	8.661***	0.155***	15	0.997***	6.801	0.122
VNI Clinical	19	0.998***	9.039***	0.184***	11	0.999***	6.697	0.141
VNI Environmental	5	1	0.917	0.061	4	1.000	0.917	0.061
VNB All	102	0.939***	6.612***	0.050***	80	0.940***	6.601***	0.049***
VNB Clinical	38	0.966***	7.572***	0.063***	36	0.966***	7.572***	0.063***
VNB Environmental	64	0.959***	6.020***	0.063***	44	0.960*	5.738**	0.058**
Francistown	40	0.946***	9.993***	0.089***	35	0.947***	9.790***	0.087***
Maun	30	0.973***	6.945**	0.070**	15	0.974*	5.137	0.049
Gaborone	7	0.994***	15.647***	0.204***	6	0.994***	15.647***	0.204***

†Number of isolates.
‡Percentage of phylogenetically compatible pairs (PcP) of loci.
§Index of association.
¶Scaled index of association (I_A) by the number of loci ($m - 1$).
††Excluding replicate haplotypes.
*$0.01 < P < 0.05$; **$0.001 < P \leq 0.01$; ***$P \leq 0.001$.

mixture of loci from VNI, VNB-I and VNB-II genetic groups, which provides a strong indication of at least occasional mating and recombination (Figs 2 and 4). Other factors, such as inbreeding, nonrandom mating, same-sex mating or strong underlying population structure, may have contributed to the observed linkage disequilibrium in the VNB population (Taylor *et al.* 1999; Pritchard & Przeworski 2001; Lin *et al.* 2005). Overall, our results indicate that both clonal reproduction and sexual reproduction occur in Botswana *Cng* populations; however, the frequency of sexual reproduction in these natural populations needs to be further evaluated.

We observed unequal distribution of *MAT**a*** mating types among different geographic areas. Although isolates with the same genotypes were isolated from Maun and Francistown, only one of 71 (1.5%) environmental

isolates from Maun possessed the rare *MAT**a*** mating type, while 39 of 96 (41.6%) environmental isolates from Francistown possessed the rare *MAT**a*** allele (Fig. 1B). In addition, no environmental isolates with the *MAT**a*** mating type were isolated from Gaborone; however, the Gaborone population of *Cng* lacked VNB isolates, which are more likely to possess *MAT**a***. The proportions of the two mating types, *MAT**a*** and *MAT*α, were almost equal in the *Cng* isolates from Francistown. This prevalence of *MAT**a*** isolates is the highest yet detected in populations of *Cng*, and the presence of both mating types is certainly compatible with sexual reproduction. A similar distribution of the mating types was observed in the environmental population of *Cryptococcus gattii* in Australia, which nonetheless had characteristics of clonality (Vilcins *et al.* 2002; Halliday & Carter 2003). The observed difference in numbers of *MAT**a*** isolates between Francistown and Maun environmental populations is unusual because the two regions share a similar ecological niche in the mopane woodland ecoregion. However, the two regions are physically separated by the Kalahari Desert and the Okavango Delta salt pans that do not support growth of *C. mopane*, and the harsh climate conditions may prevent the exchange of isolates between the regions. Although the isolates with *MAT**a*** alleles are unevenly distributed, we detected no evidence of genetic differentiation between isolates with the opposite mating types in the VNB isolates. This observation further indicates the uneven distribution of *MAT**a*** isolates might be mainly caused by geographic isolation.

We also observed significant overrepresentation of the rare *MAT**a*** allele among clinical isolates, even though less than half of the patients resided in Francistown, and many came from Gaborone and Maun, which had a low prevalence of the *MAT**a*** isolates in the environment. There are at least three possible explanations for the high prevalence of *MAT**a*** mating-type allele in clinical population. First, the majority of patients may have acquired infection while away from home and visiting the Francistown area, which has high prevalence of the *MAT**a*** in the environment. However, results of the population genetic analyses do not support this hypothesis as the genetic compositions of the Francistown environmental population and clinical population were different ($F_{ST} = 0.10$, P-value < 0.001). Second, patients may have been infected with a currently unrecognized and unsampled environmental population possessing a greater frequency of *MAT**a***. This explanation is plausible because PCA and other structure analyses detected considerable underlying structure in the clinical population, which was not revealed by our environmental sampling. Third, the high prevalence of *MAT**a*** isolates among clinical isolates may indicate that people are more likely to be infected by basidiospores than desiccated yeasts or other asexual structures. (Sexual reproduction yields an equal number of basidiospores with each mating type.) Several recent studies indicated that *Cng* basidiospores are capable of initiating infections (Giles *et al.* 2009; Velagapudi *et al.* 2009). Although linkage equilibrium among loci, which is usually expected in sexual population, was not detected in the clinical population, other factors such as underlying population structure, inbreeding and/or nonrandom mating might have contributed to the apparent disequilibrium (Pritchard & Przeworski 2001). Furthermore, clearly the *MAT**a*** isolates are fully virulent and the suggestion that *MAT*α isolates and their locus are linked to virulence in congenic murine studies (Kwon-Chung *et al.* 1992) has less relevance to disease production when *MAT**a*** isolates are found in the local environment. More research is necessary to address this question.

Our results here describe the detailed ecological and population genetic analyses of *C. neoformans* in Botswana at the epicentre of HIV and cryptococcosis. These results support our previous observations that *Cng* is widespread in the environment and associated with the mopane tree ecoregion. Natural populations of *Cng* in Botswana have high genetic diversity and clear population structure with the evidence of admixture between different genetic groups. The results also demonstrate that some geographic regions contain almost equal proportions of the two mating types; however, the extent of sexual reproduction in the environment needs to be further evaluated. These results confirm that the genetic profile of the Botswana population is very different from that of global populations of *Cng*. Different population structures between clinical and environmental isolates may indicate some clinical isolates are more likely to cause infections. Genomewide association study of selected isolates could reveal important genetic differences between the two populations.

Acknowledgements

The authors thank Prof. Rytas Vilgalys for the comments and suggestions on the manuscript, Dr. Olivier Fedrigo and Graham Alexander (Duke Center for Genomic and Computational Biology) for discussion and sample sequencing. This work was supported in part by US Public Health Service NIH grants AI93257, AI73896, AI25783, AI0055412-04, AI0055412-04 and T32DA13911-10.

References

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.

Botts MR, Hull CM (2010) Dueling in the lung: how *Cryptococcus* spores race the host for survival. *Current Opinion in Microbiology*, **13**, 437–442.

Brown AH, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW (1996) Molecular markers reveal cryptic sex in the human pathogen Coccidioides immitis. *Proceedings of the National Academy of Sciences, USA*, **93**, 770–773.

Chen Y, Toffaletti DL, Tenor JL *et al.* (2014) The *Cryptococcus neoformans* transcriptome at the site of human meningitis. *mBio*, **5**, e01087–13.

Chen Y, Frazzitta AE, Litvintseva AP *et al.* (2015) Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequencing typing. *Fungal Genetics and Biology*, **75C**, 64–71.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Dromer F, Ronin O, Dupont B (1992) Isolation of *Cryptococcus neoformans* var. *gattii* from an Asian patient in France: evidence for dormant infection in healthy subjects. *Medical Mycology*, **30**, 395–397.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Feil EJ, Li BC, Aanensen DM, Hanage WP (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology*, **186**, 1518–1530.

Giles SS, Dagenais TRT, Botts MR, Keller NP, Hull CM (2009) Elucidating the pathogenesis of spores from the human fungal pathogen *Cryptococcus neoformans*. *Infection and Immunity*, **77**, 3491–3500.

Halliday CL, Carter DA (2003) Clonal Reproduction and Limited Dispersal in an Environmental Population of *Cryptococcus neoformans* var. *gattii* Isolates from Australia. *Journal of Clinical Microbiology*, **41**, 703–711.

Hiremath SS, Chowdhary A, Kowshik T *et al.* (2008) Long-distance dispersal and recombination in environmental populations of *Cryptococcus neoformans* var. *grubii* from India. *Microbiology (Reading, England)*, **154**, 1513–1524.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Janbon G, Ormerod KL, Paulet D *et al.* (2014) Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. *PLoS Genetics*, **10**, e1004261.

Joint United Nations Programme on HIV/AIDS (UNAIDS) (2013) *Global Report: UNAIDS Report on the Global AIDS Epidemic 2013*. UNAIDS, Geneva.

Kwon-Chung KJ, Edman JC, Wickes BL (1992) Genetic association of mating types and virulence in *Cryptococcus neoformans*. *Infection and Immunity*, **60**, 602–605.

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genetics*, **8**, e1002453.

Li W, Averette AF, Desnos-Ollivier M *et al.* (2012) Genetic diversity and genomic plasticity of *Cryptococcus neoformans* AD hybrid strains. *G3 (Bethesda, Md.)*, **2**, 83–97.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*, **25**, 1451–1452.

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

Lin X, Patel S, Litvintseva AP *et al.* (2009) Diploids in the *Cryptococcus neoformans* serotype A population homozygous for the alpha mating type originate via unisexual mating. (A Andrianopoulos, Ed,). *PLoS Pathogens*, **5**, e1000283.

Litvintseva AP, Mitchell TG (2012) Population genetic analyses reveal the African origin and strain variation of *Cryptococcus neoformans* var. *grubii*. *PLoS Pathogens*, **8**, e1002495.

Litvintseva AP, Marra RE, Nielsen K *et al.* (2003) Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. *Eukaryotic Cell*, **2**, 1162–1168.

Litvintseva AP, Kestenbaum L, Vilgalys R, Mitchell TG (2005) Comparative analysis of environmental and clinical populations of *Cryptococcus neoformans*. *Journal of Clinical Microbiology*, **43**, 556–564.

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

Litvintseva AP, Carbone I, Rossouw J *et al.* (2011) Evidence that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have evolved in Africa. *PLoS ONE*, **6**, e19688.

Meyer W, Aanensen DM, Boekhout T *et al.* (2009) Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Medical Mycology: Official Publication of the International Society for Human and Animal Mycology*, **47**, 561–570.

Miglia KJ, Govender NP, Rossouw J, Meiring S, Mitchell TG (2011) Analyses of pediatric isolates of *Cryptococcus neoformans* from South Africa. *Journal of Clinical Microbiology*, **49**, 307–314.

Park BJ, Wannemuehler KA, Marston BJ *et al.* (2009) Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS (London, England)*, **23**, 525–530.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics*, **69**, 1–14.

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

Saha DC, Goldman DL, Shao X *et al.* (2007) Serologic evidence for reactivation of cryptococcosis in solid-organ transplant recipients. *Clinical and Vaccine Immunology: CVI*, **14**, 1550–1554.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Simwami SP, Khayhan K, Henk DA *et al.* (2011) Low diversity *Cryptococcus neoformans* variety *grubii* multilocus sequence types from Thailand are consistent with an ancestral African origin. *PLoS Pathogens*, **7**, e1001343.

Smith JM, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA*, **90**, 4384–4388.

Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Taylor JW, Jacobson DJ, Fisher MC (1999) The evolution of asexual fungi: reproduction, speciation and classification. *Annual Review of Phytopathology*, **37**, 197–246.

Team RDC (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Van Wyk M, Govender NP, Mitchell TG, Litvintseva AP, GERMS-SA (2014) Multilocus sequence typing of serially collected isolates of *Cryptococcus* from HIV-infected patients in South Africa. *Journal of Clinical Microbiology*, **52**, 1921–1931.

Velagapudi R, Hsueh Y-P, Geunes-Boyer S, Wright JR, Heitman J (2009) Spores as infectious propagules of *Cryptococcus neoformans*. *Infection and Immunity*, **77**, 4345–4355.

Vilcins I, Krockenberger M, Agus H, Carter D (2002) Environmental sampling for *Cryptococcus neoformans* var. *gattii* from the Blue Mountains National Park, Sydney, Australia. *Medical Mycology: Official Publication of the International Society for Human and Animal Mycology*, **40**, 53–60.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Xu J, Vilgalys R, Mitchell TG (2000) Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Molecular Ecology*, **9**, 1471–1481.

Y.C. and A.P.L. designed the study and wrote the manuscript. A.E.F. and C.F. were involved in the genotyping of the isolates. Y.C. and L.W. analysed the data. A.P.L, T.G.M., M.R.H. and C.M. collected samples. J.R.P. and T.G.M. contributed to the discussion. J.R.P. was the lead P.I. on the grant. None of the authors have a conflict of interest.

Data accessibility

DNA sequences from this investigation were deposited in GenBank with the following Accession nos: *CAP59*, KR810602–KR810742; *GPD1*, KR810461–KR810601; IGS1, KR809897–KR810037; *PLB1*, KR810179–KR810319; *LAC1*, KR810743–KR810883; *SOD1*, KR810884–KR811024; *TEF1*, KR810038–KR810178; and *URA5*, KR810320–KR810460. Aligned DNA sequences of each loci, concatenated DNA sequences, STRUCTURE, SPLITTREE, EBURST input files, raw PACBIO CCS reads and barcode information: Dryad doi:10.5061/dryad.tn28f. All of the allele type and sequence type information have been submitted to Fungal MLST Database (http://mlst.mycologylab.org).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 The genetic relationships among 141 Botswana *Cng* isolates are visualized by Principle Components Analysis.

Fig. S2 EBURST analysis of the *Cng* isolates from Botswana.

Table S1 Summary information of 614 environmental sampling sites in Botswana.

Table S2 Mating types of 304 *Cng* isolates. City information for clinical strains is based on the patients' residence areas.

Table S3 The MLST allelic profiles of 141 *Cng* isolates from Botswana.

Table S4 Summary of polymorphism among environmental isolates from different areas at each locus.

Table S5 Summary of polymorphism among VNI and VNB isolates at each locus.

Table S6 Indices of Association for two mating types of VNB strains.

Table S7 Best models for individual locus and concatenate sequence estimated by JMODELTEST2.

Table S8 Shimodaira–Hasegawa (SH) tests for phylogenetic incongruence for all pairwise loci and the concatenated sequences of all loci (Merge). This test assumes alternative topologies are equal supported by data set of a locus. Here the * represents *P*-value is <0.05, indicating significant difference of tree topologies for that pair loci.