

MOLECULAR EPIDEMIOLOGY OF *CRYPTOCOCCUS NEOFORMANS*

**MOLECULAR EPIDEMIOLOGY OF *CRYPTOCOCCUS NEOFORMANS*  
*VARIETY GRUBII***

**By**

**Dr. SANJAY S. HIREMATH**

**A Thesis**

**Submitted to the School of Graduate Studies**

**in Partial Fulfillment of the Requirements**

**for the Degree**

**Master of Science**

**McMaster University**

**© Copyright by Sanjay S. Hiremath, May 2008**

MASTER OF SCIENCE (2008)

McMaster University  
Hamilton, Ontario

TITLE: Molecular epidemiology of *Cryptococcus neoformans* variety *grubii*

AUTHOR: Sanjay S. Hiremath, M.B.B.S. (Jawaharlal Nehru Medical College, India)

SUPERVISOR: Dr. Jianping Xu (J-P)

NUMBER OF PAGES: x, 65

## Abstract

The basidiomycete yeast *Cryptococcus neoformans* is a cause of significant morbidity and mortality in immunocompromised hosts throughout the world. The sporadic nature of the infection and the limited empirical evidence for direct human-to-human transmission have led to the belief that infections in humans are predominantly caused by the inhalation of basidiospores from environmental sources. Therefore, analyzing the structure of environmental populations of *C. neoformans* can significantly increase our understanding of its ecology, evolution, and epidemiology. Decaying wood is a rich source of organic and inorganic compounds and is known to be a suitable ecological niche for many microorganisms, including *C. neoformans*. However, relatively little is known about the population structure of *C. neoformans* sampled from decaying wood. In this study, we analyzed samples of *C. neoformans* var. *grubii* colonizing decaying wood in tree hollows of nine tree species in five geographical locations (Delhi, Bulandshahar, Hathras, Amritsar and Amrouli) in northwestern India. Multilocus sequence typing was conducted using five gene fragments for each of seventy-eight isolates. All isolates belonged to mating type  $\alpha$ . Population genetic analyses identified no evidence for significant differentiation among populations belonging to either different geographic areas or different host tree species. Interestingly, despite the lack of mating type **a** strains in our survey, we found unambiguous evidence for recombination in our population analyses. Our results are consistent with the hypotheses of long distance dispersal and recombination in environmental populations of this species in India.

## **Acknowledgements**

“Guru” is one who dispels the darkness of ignorance. First I would like to extend my most sincere thanks to my “Guru” Dr. Jianping Xu. I now only endeavor to justify the confidence and kindness reposed on me. This, I believe would justify his generosity. I also thank him for giving me a chance to work in his lab, for his encouragement, guidance with my project and the periodic wake-up calls. To Dr. Elizabeth Weretilnyk for offering a helping hand when I needed it the most. To Dr. Suleiman Igdoura and Dr. Jonathan Stone for their kindness and guidance. I would also like to thank Dr. R. A. Morton for his own unique brand of help and guidance.

Special thanks to Mrs. Pat Hayward for all her assistance and help with all manners of paperwork. To all the members of Dr. J. P. Xu’s lab, especially Sun Sheng, and Mochan Li. Special thanks to our postdoctoral fellow Dr. Timothy James for all the useful suggestions.

Finally, I would like to thank the most important people in my life, my family. Thank you Dad, Mom, Deepak, Namrata, Vinay and my sweet niece Ananya.

## **Thesis organization and format**

In consultation with my supervisor, it was decided that this thesis would be organized in the “sandwich thesis” format approved by McMaster University. Accordingly, this thesis comprises has two chapters. Chapter 1 provides a general introduction to the study organism and to the thesis research. Chapter 2 has been accepted for publication in a peer reviewed scientific journal. It also includes the general conclusions of this research.

### **Chapter 1: General introduction**

### **Chapter 2: Long distance dispersal and recombination in environmental populations of *Cryptococcus neoformans var. grubii* from India**

**Comments:** This chapter has been accepted for publication.

Hiremath SS, Choudhary A, Koushik T, Randhawa HS, Sun S & Xu J (2008). Long distance dispersal and recombination in environmental populations of *C. neoformans var. grubii* from India. *Microbiology* (in press).

This study was done under the supervision of Dr. J.P. Xu. The sample collection was done by Dr. Anuradha Choudhary, Dr. Tusharantak Kowshik and Dr. Harbans Randhawa. Sun Sheng extracted the DNA and sequenced twenty of the samples. I sequenced all the remaining samples, conducted all the remaining experiments, did the analyses and wrote the manuscript. Dr. J.P. Xu guided me with some of the analysis and helped in writing and modifying the manuscript.

## TABLE OF CONTENTS

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Thesis organization and format</b>	<b>v</b>
<b>Table of Contents</b>	<b>vi</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Abbreviations</b>	
<b>Chapter 1. General introduction</b>	
Nomenclature	2
Serotypes and molecular subtypes	3
Reproduction	5
Virulence factors and Disease	7
Multilocus sequence typing	8
Phylogenetic and population genetic analyses	10
Why study <i>C. neoformans</i> in India?	14
References	16
<b>Chapter 2. Long distance dispersal and recombination in environmental populations of <i>Cryptococcus neoformans</i> var. <i>grubii</i> from India</b>	
Abstract	24

Introduction	25
Materials and Methods	28
Results	35
Discussion	39
References	45



## List of figures

### Chapter 2

- Fig. 1:** Maximum parsimony trees showing the relationships among 78 isolates of *Cryptococcus neoformans* var. *grubii* for the combined dataset using mid-point rooting ..... 59
- Fig. 2:** The distribution of  $\theta$  (a measure of genetic differentiation) among geographic populations of *C. neoformans* var. *grubii* from India..... 60
- Fig. 3:** The observed Index of association ( $I_A$ ) within the distribution of  $I_A$  from 10000 artificially recombined datasets..... 61
- Fig. S1:** Maximum parsimony trees showing the relationships among 78 isolates of *Cryptococcus neoformans* var. *grubii* for three of the markers sequenced...62

## List of tables

### Chapter 2

<b>Table 1:</b> <i>Cryptococcus neoformans</i> var. <i>grubii</i> strains used in this study.....	52
<b>Table 2 :</b> Genetic differences among populations of <i>C. neoformans</i> var. <i>grubii</i> in India.....	55
<b>Table 3:</b> Allelic combinations showing evidence of recombination.....	56
<b>Table 4:</b> Multilocus linkage disequilibrium analyses.....	57

## Abbreviations

AFLP	Amplified fragment length polymorphism
<i>C. neoformans</i>	<i>Cryptococcus neoformans</i>
CAP1	Capsule associated protein 1
CPS	capsular polysaccharide
CD	Cryptococcal disease
ITS	Internal transcribed spacer
LAC	Laccase/diphenol oxidase
MLST	Multilocus sequence typing
mtLrRNA	Mitochondrial RNA
RAPD	Random amplification of polymorphic DNA
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism

## **Chapter 1**

### **General introduction**

## General introduction

### Nomenclature

*C. neoformans* is a eukaryotic pathogen belonging to the kingdom fungi, subkingdom Dikarya and phylum Basidiomycota. This filamentous fungus was first isolated more than a hundred years ago from fermented peach juice by the Italian microbiologist Francesco Sanfelice (Sanfelice, 1894). He noticed that the organism formed tumor like lesions and hence named it *Sacharromyces neoformans*. Two contemporary German physicians Otto Busse and Abraham Buschke also recovered the same organism from a tibial sarcoma-like lesion of a thirty-one year old woman (Busse, 1894). The nature of the lesions caused by this fungus resulted in a great deal of confusion, hence several terminologies were used to describe this organism since its discovery. It went from being called *Cryptococcus hominis* by Vuillemin to *Torula histolytica* by Clements and Shear to *Megalococcus myxoides* by Ferdinand Curtis to several others (Benham, 1935a; Drouhet, 1997). Although Vuillemin used the name *Cryptococcus hominis* to describe the species, Kützing was the first to establish the genus *Cryptococcus* in 1833 (Kützing, F. (a) *Algarum aquae dulcis Gernaniae* Decas 111. 1833.(b) *Phycologia germanica*. 1845. (c) *Species algarum*. 1849). It was in 1935 that Rhoda Behnam from Columbia university named this organism *C. neoformans*. This was based on the priority rule of nomenclature and was created by combining the names used by Sanfelice and Vuillemin (Benham, 1935a). Later on, the genus *Cryptococcus* itself was delineated to include only fungi that could utilize inositol, lack fermenting ability, produce urease and be able to produce either hyphae or pseudohyphae (Phaff & Spencer, 1966). Finally, in 1989, the taxonomic nomenclature of designating the genus

*Cryptococcus* and the species originally discovered by Sanfelice as *C. neoformans* was settled upon (Fell et al, 1989).

*C. neoformans* exists both as an anamorph and a teleomorph. This was first noticed when some strains of *C. neoformans* readily produced hyphae accompanied by clamp connections (Shadomy, 1970). Clamp connections maintain the dikaryotic state of this fungus and justified their classification as basidiomycota. This was followed by confirmation of heterothallic mating and characterization of the life cycle of this pathogen (Kwon-Chung, 1975; 1976a; 1976b). Kwon-Chung proposed the binomials *Filobasidiella neoformans* (oval to cylindrical basidiospores) and *Filobasidiella bacillospora* (bacilliform basidiospores) for the teleomorphic forms of this fungus.

### **Serotypes and molecular subtypes**

Benham (Benham, 1935b) first initiated antigenic characterization of the polysaccharide capsule to classify the organism *C. neoformans* into two classes. This was followed in the 1950s, by the identification of several serotypes (A, B & C) within the genus *C. neoformans* based on reciprocal agglutination absorption and microscopic capsular reactions (Evans, 1950). Serotype D was discovered soon after in the 1960s by using adsorbed rabbit immune sera (Wilson et al, 1968). A fifth serotype AD that corresponds to a diploid or aneuploid form of this organism, has also been characterized. Serotype AD is formed by the hybridization of serotypes A & D. The anamorphs *C. neoformans* var. *grubii* (Serotype A) and *C. neoformans* var. *neoformans* (serotype D) correspond to the teleomorph *Filobasidiella neoformans*. While the anamorph *C.*

*neoformans* var. *gattii* (Serotypes B & C) corresponds to the teleomorph *Filobasidiella bacillospora*.

Serotype identification is made possible by eight major antigenic determinants in the capsule that enable immunological subgrouping of the serotypes. Serotyping currently is carried out by using an agglutination test, which is based on polyclonal rabbit sera reactivity to the capsule after differential adsorption of the sera with the nonimmunizing serotypes. This can be done using either commercial (Crypto Check kit; Iatron Labs, Tokyo, Japan) or homemade antisera components (Ikeda et al, 1982). Serotype identification has also been achieved by amplifying fragments of capsule related genes (eg. Enache-Angoulvant et al, 2007).

Extensive variation in CPS and the inability to hybridize efficiently means that the genomes of the different serotypes have undergone extensive divergence. The availability of the genome sequences for Serotypes A, B & D has made it possible to estimate the divergence times between the serotypes. Phylogenetic analysis done in a recent study (Sharpton et al, 2008) has shown that the two species *Cryptococcus neoformans* (comprising serotypes A & D) and *Cryptococcus gattii* (comprising serotypes B & C) diverged approximately 80mya (million years ago). The authors in that study first estimated the dS (synonymous substitution per synonymous site) by a pairwise comparison between Serotypes A/D & B. They then used a neutral mutation rate of  $2 \times 10^{-9}$  to estimate the divergence time.

Based on DNA sequence polymorphism the different serotypes have been further subdivided into nine distinct molecular subtypes. The polymorphic patterns are usually detected by using one of PCR fingerprinting, RAPD, AFLP, RFLP or MLST

(Sorrell et al, 1996; Meyer et al, 2003; Litvintseva et al, 2006). Serotype A can be classified into VNI, VNII and VNB (so far found only in Botswana) patterns. Serotype D so far has only the VNIV pattern, while serotype AD only the VNIII pattern. Serotype B can generate VGI, VGII or VGIII patterns and Serotype C can generate VGIII or VGIV patterns (Meyer et al, 1999).

Epidemiological surveys of this organism have shown that serotype A is distributed worldwide. Majority of the isolates isolated from the environment and clinical infections belong to this serotype. Serotype D has mostly been recovered from India and certain parts of Europe and is relatively rare. Serotypes B & C are mostly found in tropical and subtropical climates. It has also been recovered and was the cause of a major outbreak of CD in Vancouver Island, Canada recently (reviewed in Lin & Heitman, 2006).

## **Reproduction**

*C. neoformans* is a heterothallic, facultative sexual organism. This means that it mostly reproduces asexually and sexual encounters are infrequent and usually brought on by an environmental stimulus. Facultative sex allows the organism to benefit from both asexuality and sex (Wright, 1986; Hastings, 1992). The disadvantage of asexuality is that genetic loci in that case are permanently linked and so more prone to accumulate deleterious mutations. Meiosis and recombinational shuffling on the other hand can break up both favorable and unfavorable combinations of genetic loci. Being heterothallic, the two sexually compatible strains of *C. neoformans* are morphologically



indistinguishable. Understanding the sexual cycle of this pathogen would give useful insights into the production of infectious particles (basidiospores) in this organism. The sexual cycle of this haploid organism is controlled by the *MAT* locus that occurs in two idiomorphic forms corresponding to the opposite mating types  $a$  and  $\alpha$ .

Mating is initiated under conditions of nitrogen limitation, dessication, darkness, etc. when a pheromone (*MF $\alpha$*  or *MFa*) binds to its cognate receptor (*STE3 $\alpha$*  or *STE3a*). This leads to activation of a heterotrimeric G protein and downstream activation of the MAP kinase cascade (Herskowitz, 1995). Studies have shown that in *C. neoformans*, *MATa* produces *MFa* in response to nitrogen deprivation. This induces the formation of a conjugation tube by *MAT $\alpha$*  cells. However, the reverse process wherein *MAT $\alpha$*  signals and *MATa* responds does not occur (reviewed in Mclelland et al, 2004). Cell-cell fusion results in a dikaryotic mycelium (heterokaryosis) linked by fused clamps, which then undergoes a dimorphic transition to a filamentous growth form. Subsequently karyogamy and meiosis occurs in specialized terminal structures called basidia. Four chains of infectious haploid basidiospores are then formed in basipetal series on the apex of the basidium. These spores are readily aerosolized. The spores then germinate to form haploid yeast cells that multiply by budding.

The dimorphic transition to a filamentous growth form is also possible by another process called monokaryotic fruiting. This is a form of sexual reproduction between haploid yeast cells belonging to the same mating type (usually  $\alpha$ ). Hyphal cells produced during fruiting are mononucleate, with unfused clamp connections, whereas during mating cells are diploid with fused clamp connections. Karyogamy occurs early during monokaryotic fruiting. The rest of the steps are similar to opposite sex mating.

*C. neoformans* undergoes asexual reproduction by mitosis and budding (Cassone et al, 1974; Mochizuki et al, 1987). The sequence of events is similar to that occurring in other basidiomycetes.

### **Virulence Factors and Disease**

Virulence factors are those products that increase the infectivity and degree of pathogenicity of an organism. The virulence factors that have been proposed for *C. neoformans* are the CPS, cryptococcal products (CPS and the soluble extracellular constituents), production of substances like melanin, mannitol, superoxide dismutase, proteases, phospholipase B lysophospholipase and the ability to grow at 37°C. Of these, the CPS, cryptococcal products and the production of melanin are the most important. Mating type also plays a role as a virulence factor in CD. Studies have shown that majority of the clinical isolates belong to the  $\alpha$  mating type.  $\alpha$  is also more likely to enter the CNS. Also the  $\alpha$  mating type is also more likely to undergo monokaryotic fruiting hence offer it a better survival advantage (Kwon-Chung et al., 1992; Wickes et al., 1996).

The majority of the cases of CD are secondary infections caused by *C. neoformans var. grubii* in patients with a primary compromise in immune functions (T cell mediated immunity) either as a result of disease (eg. AIDS, sarcoidosis, lymphomas, etc.) or as a result of therapy (eg. immune-suppressive therapy following organ transplantation, steroids, chemotherapy etc.). However, it has been known to infect the immunocompetent as well. All the known serotypes can infect either, however Serotype A is the most common amongst the immunocompromised whereas serotype B is more

common amongst the immunocompetent. Regardless of the serotype the vast majority of either clinical or environmental isolates belong to the MAT $\alpha$  mating type. Basidiospores or desiccated yeast cells form the aerosolized inoculums, which are deposited in the distal airways. There they remain dormant until a decline in the host’s immune system sets them free to invade and disseminate via the bloodstream within the host. Infections are almost always acquired from an environmental source. Person to person transmission is extremely rare and has been known to occur only by organ transplantation or sometimes within hospital settings (nosocomial). Most of the lesions in CD are caused as a result of mass lesions or due to inflammation. So, dissemination can cause pulmonary granulomas, meningoencephalitis, cutaneous skin lesions, otitis, sinusitis, osteomyelitis, prostatitis, etc. Lesions in the brain can cause focal neurological deficits due to mass lesions caused by the rapidly multiplying fungus. Of particular concern are prostate infections because the foci may persist even after aggressive therapy for CNS disease and act as a reservoir for relapse in AIDS patients (Casadevall & Perfect, 1998).

### **Multilocus sequence typing**

MLST is a nucleotide sequence-based typing scheme that allows the unambiguous characterization of microbial isolates using the sequences of the internal fragments of genes. It was originally proposed for studying the epidemiology of *Neisseria meningitides* (Maiden et al, 1998). Since then it has been used for other pathogenic and environmental bacterial and eukaryotic species as well (reviewed in Urwin & Maiden, 2003; Cooper & Feil, 2004). Because it is portable, is able to detect

variation at a much higher resolution, and is amenable for ready comparison of datasets between laboratories, it has served as an excellent tool for epidemiological studies. A combination of variable loci like antigenic proteins and slowly evolving loci like housekeeping genes serves as the ideal choice for such studies. This would serve to provide a high resolution localized epidemiology and set that within the context of a broader epidemiological context. The use of multiple markers (fragments of genes) that are scattered around the genome serves two purposes: increases the resolution and discriminatory power of the variation and allows evolutionary reconstruction. Evolutionary reconstruction is possible in this case because it accounts for and is mostly robust to both the effects of homologous recombination within a single gene locus and the possibility that genes that lie close to each other could be physically linked (Feil et al, 2004).

MLST has also been used extensively for the study of *C. neoformans* (eg. Xu et al, 2000; Litvintseva et al, 2006). Variable number and type of nuclear DNA markers have been used. However we found that no common scheme has been agreed upon and hence the purpose of MLST has not been best served in this case. However, we have noticed that one particular marker *LAC* (see below for description) has been used in many of the studies and serves as the best comparative tool at the moment. We used this scheme to study the environmental population structure of *C. neoformans* from India. This is the first time that such a scheme has been applied to study the environmental population structure of *C. neoformans* from India. In this study we used five molecular markers: *CAP1*, *FTR1*, *LAC*, *ITS* and *mtLrRNA* (Xu et al., 2000; Kidd et al., 2005)

*CAP1* encodes a capsule associated protein. The exact function of this protein is not known. It is located within the *MAT* locus on chromosome IV. *FTR1* is located on chromosome III and encodes a high affinity iron permease. It is involved in the transport of iron after it has been reduced to the ferrous form by cell surface reductases (Jung et al., 2006). *LAC* belongs to a class of extracellular proteins called multicopper oxidases. It is located on chromosome VII and encodes laccase involved in the production of melanin. In *C. neoformans* it is involved in the degradation of lignin polymers in rotting wood (Zhu et al., 2001). It is also known to play a role in dissemination to the brain but not in pulmonary persistence. This probably suggests that this protein is secreted in a tissue specific manner (Noverr et al., 2004). *ITS* is the internal transcribed spacer region located on Chromosome II. The *ITS* regions are located in the rDNA gene complex of all eukaryotic cells between the 18S and 5.8S rDNA genes (*ITS1*) and the 5.8S and 28S rDNA genes (*ITS2*). The sequence homology within the rDNA genes of fungi and differences within the spacer regions are the genetic basis of organizing the fungi into taxonomic groups (reviewed in Iwen et al., 2002). Finally we used a mitochondrial marker *mtLrRNA*, which is good to discriminate between species, but conserved within each species.

## **Phylogenetic and population genetic analyses**

### **a) Mode of reproduction**

This essentially involves looking for the distribution of alleles amongst the individuals in a population to uncover any evidence for recombination. Clonality is inferred if there is over representation of certain multilocus genotypes, linkage disequilibrium and the absence of recombinant genotypes. We used three tests to look for recombination in these haploid multilocus genotypes.

1. Partition homogeneity test (PHT) – This test determines whether evidence exists for differing phylogenetic histories underlying two or more gene trees or data sets. In clonal organisms all genes would share a common evolutionary history. In recombining organisms genes maybe exchanged between species and so each gene could have different histories. So basically, given the level of incongruence within each gene tree, this test evaluates if the level of incongruence between two or more gene trees is greater that expected by chance alone (Farris et al., 1995). This test involves three steps: a) Within dataset incongruence is calculated by conducting separate parsimony analysis of the individual genes in a MLST study and summing the tree lengths. b) A simultaneous parsimony analysis of all genes is performed and the resultant tree length obtained to get the total incongruence both within and between trees. This should always equal or be greater than the individual tree lengths. The difference between a) and b) represents the extra incongruence generated by the combined tree. c) The dataset is then randomly partitioned into divisions that mirror the original dataset and the between dataset incongruence is obtained as mentioned before. This procedure is repeated and a distribution of between dataset incongruence is obtained. If the between-data- sets incongruence in five percent of the randomized partitions exceeds that in the real

data sets, it can be concluded that the observed incongruence is statistically significant.

2. Index of association ( $I_A$ ) – This is a statistical test used to measure the extent of linkage disequilibrium based on the distribution of allelic mismatches between organisms over several loci. The mean number of such pair wise mismatches gives us the measure of genetic distance. This index measures if the variance of such associations or pair wise mismatches in the population under study differs from a population that is panmictic.  $I_A$  approaches zero in freely recombining populations but shows a bimodal distribution with an  $I_A$  significantly greater than zero (Maynard-Smith et al., 1993).
3. Phylogenetic incompatibility - It is the inconsistency in the patterns of shared-derived substitutions between loci (Hudson & Kaplan, 1985). In other words two loci would be compatible if we can account for all the observed genotypes by mutation. Incompatibility is inferred if there is either evidence for homoplasy (reversal, parallelism, convergent evolution) or recombination. For example if there are two loci with two alleles each, then it is possible to observe four genotypes. However, in clonal populations if we were to start with a single pair of loci then mutations would only yield three of the four possible combinations. The fourth genotype can only be accounted for if there are repeated substitutions at the same site several times or if there is recombination. Repeated substitutions at the same specific site are unlikely especially if it is a synonymous substitution and if the particular marker is not hypervariable. The most likely scenario then would be recombination.

**b) Population structure**

We used maximum parsimony analysis to construct phylogenetic trees to assess the genetic structure of the population. We then used three tests to assess the population structure of the study samples.

1. Topology dependent permutation tail probability test (T-PTP) – Permutation tail probability is a statistical test in which at first control datasets are generated using within-character randomization to generate trees that simulate recombination. Subsequently, the control dataset is then compared to some aspect of the phylogeny generated by the actual dataset. If the trees derived from the actual dataset are significantly shorter than the ones generated by simulating recombination then clonality is inferred. The test statistic for a T-PTP test is also measured by calculating the difference in tree lengths as mentioned before, but in this case the trees are constrained to include a given subtree (Faith, 1991).
2. Weir and Cockerham’s theta – Differences in allele frequencies between populations can be used to give a measure of the extent of genetic differentiation between populations. This was originally proposed by Wright to measure population subdivision or the degree of gene differentiation between subpopulations by using F-statistics ( $F_{ST}$ ). This statistic can also be calculated for haploid organisms using Weir and Cockerham’s theta ( $\theta$ ). If two populations have different alleles undergoing fixation then  $\theta = 1$ . Genetic differentiation is



determined if the observed values of  $\theta$  is significantly different from one in which individuals are randomly exchanged among populations (Carter et al., 2001).

3. Nei’s genetic distance – This is used to estimate gene differences per locus between populations. Or in other words it is the mean number of codon substitutions per locus corrected for multiple hits (Nei, 1972).

### **Why study *C. neoformans* in India?**

The short answer is that, India is a hotspot for AIDS (acquired immune deficiency syndrome) and CD is an opportunistic infection. In fact, extrapulmonary CD has been classified as one of the AIDS defining illnesses. It is also a place where poverty and patient non-compliance add to the possibility that drug resistant genotypes of this organism may prevail and spread. Since almost all infections are acquired from an environmental source, it is both important and useful from a public health viewpoint that we understand the population dynamics of this organism in the environment in India. CD has been recorded in India since 1941, however there have been no systematic MLST surveys of this organism done from the environment in India.

In addition it is also interesting to know the various ecological niches that this organism thrives on and can complete its life cycle. This is especially true if these habitats are close to human habitation. In my study the samples that we obtained from Dr. J. P. Xu’s colleagues (Dr. Anuradha Choudhary, Dr. Tusharantak Kowshik and Dr. Harbans Randhawa) in India were all obtained from tree hollows that were very close to human habitation.

The objective of my study was to understand the population structure of specifically the most common variety of cryptococcus, *C. neoformans var. grubii* from India. I also wanted to assess the mode of reproduction of this organism in the environment.

## References

1. **Benham, R.W.** (1935a). The terminology of the *Cryptococci* with a note on *Cryptococcus mollis*. *Mycologia* **27**, 496-502.
2. **Benham, R.W.** (1935b). Cryptococci, their identification by morphology and serology. *J Infect Dis* **57**, 255-274.
3. **Busse, O.** (1894). Über parasitäre Zelleinschlüsse und ihre Züchtung. *Zbl Bakt I Abt Orig* **16**, 175–180.
4. **Casadevall, A. & Perfect, J.R.** (1998). *Cryptococcus neoformans*. Washington, DC: ASM press.
5. **Carter, D.A., Taylor, J.W., Koenig, G.I., & White, T.J.** (2001). Amplified single nucleotide polymorphisms and (GA) microsatellite marker reveal genetic differentiation between populations of *Histoplasma capsulatum* from the Americas. *Fungal Genet & Biol* **34**, 37-48.
6. **Cassone, A., Simonetti, N. & Strippoli, V.** (1974). Wall structure and bud formation on *Cryptococcus neoformans*. *Arch Microbiol* **95**, 205-212.
7. **Cooper, J.E. & Feil, E.J.** (2004). Multilocus sequence typing – what is resolved? *Trends Microbiol* **12**, 373-377.
8. **Drouhet, E.** (1997) Milestones in the history of *Cryptococcus* and Cryptococcosis. *J Mycologie Medicale* **7**,10-27.
9. **Enache-Angoulvant, A., Chandener, J., Symoens, F., Lacube, P., Bolognini, J., Douchet, C., Poirot, J.L. & Hennequin, C.** (2007). Molecular identification of *Cryptococcus neoformans* serotypes. *J Clin Microbiol* **45**, 1261-1265.

10. **Evans, E.E.** (1950). The antigenic composition of *Cryptococcus neoformans*. I. A serologic classification by means of the capsular and agglutination reactions. *J Immunol* **64**, 423-430.
11. **Faith, D.L.** (1991). Cladistic permutation tests for monophyly and nonmonophyly. *Syst Zool* **40**, 366-375.
12. **Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C.** (1994). Testing significance of incongruence. *Cladistics* **10**, 315-319.
13. **Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P. & Spratt, B.G.** (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**, 1518-1530.
14. **Fell, J.W., Kurtzman, C.P. & Kwon-Chung, K.J.** (1989). Proposal to conserve *cryptococcus*. *Taxon* **38**, 151-152.
15. **Hastings, I.** (1992). Why is sex so frequent. *Trends Ecol Evol* **7**, 278–279.
16. **Herskowitz, I.** (1995). MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187-197.
17. **Hudson, R.R., & Kaplan, N.L.** (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147-164.
18. **Ikeda, R., Shinoda, T., Fukazawa, Y. & Kaufman, L.** (1982). Antigenic characterization of *Cryptococcus neoformans* serotypes and its application to serotyping of clinical isolates. *J Clin Microbiol* **16**, 22-29.

19. **Iwen, P.C., Hinrichs, S. & Rupp, M.E.** (2002). Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. *Med Mycol* **40**, 87-109.
20. **Jung, W.H., Sham, A., White, R. & Kronstad, J.W.** (2006). Iron regulation of the major virulence factors in the AIDS-associated pathogen *Cryptococcus neoformans*. *PLoS Biol* **4**, e410.
21. **Kwon-Chung, K.J.** (1975). A new genus, *Filobasidiella*, the perfect state of *Cryptococcus neoformans*. *Mycologia* **67**, 1197-1200.
22. **Kwon-Chung, K.J.** (1976a). Morphogenesis of *Filobasidiella neoformans*, the sexual state of *Cryptococcus neoformans*. *Mycologia* **68**, 821-833.
23. **Kwon-Chung, K.J.** (1976b). A new species of *Filobasidiella*, the sexual state of *Cryptococcus neoformans* B and C serotypes. *Mycologia* **68**, 942-946.
24. **Kwon-Chung, K.J., Edman, J.C. & Wickes, B.L.** (1992). Genetic association of mating types and virulence in *Cryptococcus neoformans*. *Infect Immun* **60**, 602-605.
25. **Litvintseva, A.P., Thakur, R., Vilgalys, R. & Mitchell, T.G.** (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.
26. **Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt, B.G.** (1998). Multilocus sequence typing: a portable approach to the

- identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140-3145.
27. **Maynard-Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993).** How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384-4388.
28. **McClelland, C.M., Chang, Y.C., Varma, A. & Kwon-Chung, K.J. (2004).** Uniqueness of the mating system in *Cryptococcus neoformans*. *Trends Microbiol* **12**, 208-212.
29. **Meyer, W., Marszewska, K., Amirmostofian, M., Igreja, R.P., Hardtke, C., Methling, K., Viviani, M.A., Chindamporn, A., Sukroongreung, S., John, M.A., Ellis, D.H. & Sorrell, T.C. (1999).** Molecular typing of global isolates of *Cryptococcus neoformans var. neoformans* by polymerase chain reaction fingerprinting and randomly amplified polymorphic DNA—a pilot study to standardize techniques on which to base a detailed epidemiological survey. *Electrophoresis* **20**, 1790–1799
30. **Meyer, W., Castaneda, A., Jackson, S., Huynh, M. & Castaneda, E. (2003).** Molecular typing of IberoAmerican *Cryptococcus neoformans* isolates. *Emerg Infect Dis* **9**, 189–195.
31. **Mochizuki, T., Tanaka, S. & Watanabe, S. (1987).** Ultrastructure of the mitotic apparatus in *Cryptococcus neoformans*. *J Med Vet Mycol* **25**, 223-233.
32. **Nei, M. (1972).** Genetic distance between populations. *Am Nat* **106**, 283-292.
33. **Noverr, M. C., P. R. Williamson, R. S. Fajardo, and G. B. Huffnagle. (2004).** *CNLAC1* is required for extrapulmonary dissemination of *Cryptococcus neoformans* but not pulmonary persistence. *Infect. Immun* **72**, 1693-1699.

34. **Phaff, J.H. & Spencer, J.F.T.** (1966). Improved parameters in the separation of species in the genera *Rhodotorula* and *Cryptococcus*. pp. 59-67. Proceedings of the second international symposium on yeasts, Bratislava.
35. **Sanfelice, F.** (1894). Contributo alle morfologia e biologia dei blastomiceti che si sviluppano nei succhi di alcuni frutti. *Ann. Istit. Igiene R. Univ. Roma* **4**, 463–495.
36. **Shadomy, H.J.** (1970) Clamp connections in two strains of *Cryptococcus neoformans*. Spectrum monograph series. Arts and Sciences, Georgia State University **1**, 67-72.
37. **Sharpton, T.J., Neafsey, D.E., Galagan, J.E. & Taylor, J.W.** (2008). Mechanisms of intron gain and loss in *Cryptococcus*. *Genome Biol* **9**, R24.
38. **Sorrell, T.C., Chen, S.C., Ruma, P., Meyer, W., Pfeiffer, T.J., Ellis, D.H. & Brownlee, A.G.** (1996). Concordance of clinical and environmental isolates of *Cryptococcus neoformans* var. *gattii* by random amplification of polymorphic DNA analysis and PCR fingerprinting. *J Clin Microbiol* **34**, 1253-1260.
39. **Urwin, R. & Maiden, M.C.** (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* **11**, 479-487.
40. **Wickes, B.L., Mayorga, M.E., Edman, U. & Edman, J.C.** (1996) Dimorphism and haploid fruiting in *Cryptococcus neoformans*: association with the  $\alpha$  mating type. *Proc Natl Acad Sci U S A*, **93**, 7327-7331.
41. **Wilson, D.E., Bennett, J.E. & Bailey, J.W.** (1968). Serologic grouping of *Cryptococcus neoformans*. *Proc Soc Exp Biol Med* **127**, 820-823.
42. **Wright, S.** (1986). Statistical genetics in relation to evolution. pp. 328 In: Provine WB (ed) Selected papers of Sewall Wright. University of Chicago Press, Chicago.

**43. Zhu, X., Gibbons, J., Garcia-Rivera, J., Casadevall, A. & Williamson, P. R.**

(2001). Laccase of *Cryptococcus neoformans* is a cell wall-associated virulence factor. *Infect Immun* **69**, 5589-5596.



## **Chapter 2**

**Long distance dispersal and recombination in environmental populations of *Cryptococcus neoformans* var. *grubii* from India**

**Long distance dispersal and recombination in environmental populations of *Cryptococcus neoformans* var. *grubii* from India**

**SANJAY S. HIREMATH<sup>\*</sup>, ANURADHA CHOWDHARY<sup>†</sup>, TUSHARANTAK**

**KOWSHIK<sup>†</sup>, HARBANS S. RANDHAWA<sup>†</sup>, SHENG SUN<sup>\*</sup>, JIANPING XU<sup>\*</sup>**

**<sup>\*</sup>Center for Environmental Genomics, Department of Biology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada.**

**<sup>†</sup>Department of Medical Mycology, Vallabhbhai Patel Chest Institute, University of Delhi, 110 007 Delhi, India.**

**Keywords: Recombination, Clonality, Population structure, Cryptococcus, MLST**

**Running title: Population structure in *Cryptococcus neoformans***

**Contents category: Biodiversity and Evolution**

**\*Corresponding author: Dr. Jianping Xu, Department of Biology, McMaster University, 1280 Main St. W., Hamilton, ON, L8S 4K1, Canada; Tel: 1-905-525-9140 x27934; Fax: 1-905-522-6066; Email: [jpxu@mcmaster.ca](mailto:jpxu@mcmaster.ca)**

**GenBank Accession Numbers for DNA sequences obtained here are: CAP1: EU312204-EU312281; FTR1: EU312282-EU312359; ITS: EU312360-EU312437; LAC: EU312438-EU312515; MtLrRNA: EU312516-EU312593.**

## **Abstract**

The basidiomycete yeast *Cryptococcus neoformans* is a cause of significant morbidity and mortality in immunocompromised hosts throughout the world. The sporadic nature of the infection and the limited empirical evidence for direct human-to-human transmission has led to the belief that infections in humans are predominantly caused by the inhalation of basidiospores from environmental sources. Therefore, analyzing the structure of environmental populations of *C. neoformans* can significantly increase our understanding of its ecology, evolution, and epidemiology. Decaying wood is a rich source of organic and inorganic compounds and is known to be a suitable ecological niche for many microorganisms, including *C. neoformans*. However, relatively little is known about the population structure of *C. neoformans* sampled from decaying wood. In this study, we analyzed samples of *C. neoformans* var. *grubii* colonizing decaying wood in tree hollows of nine tree species in five geographical locations (Delhi, Bulandshahar, Hathras, Amritsar and Amrouli) in northwestern India. Multilocus sequence typing was conducted using five gene fragments for each of seventy-eight isolates. All isolates belonged to mating type  $\alpha$ . Population genetic analyses identified no evidence for significant differentiation among populations belonging to either different geographic areas or different host tree species. Interestingly, despite the lack of mating type **a** strains in our survey, we found unambiguous evidence for recombination in our population analyses. Our results are consistent with the hypotheses of long distance dispersal and recombination in environmental populations of this species in India.

## Introduction

*Cryptococcus neoformans* var. *grubii* is the most common causative agent of cryptococcosis and can lead to significant morbidity and mortality in hosts with a defective T-cell mediated immunity (Casadevall & Perfect, 1998). It has a global distribution and has been isolated from both human and environmental sources such as avian guano, soil, fruits, vegetables, and decaying hollows of more than ten species of trees in different parts of the world (Randhawa et al., 2003; Reimão et al., 2007). Current surveys suggest that the prevalence of cryptococcosis caused by strains of *C. neoformans* var. *grubii* varies among geographic regions, likely a reflection of their variable frequencies in different environments (Bennett et al., 1977; Casadevall & Perfect, 1998; Tintelnot et al., 2004). *C. neoformans* var. *grubii* represents strains of serotype A, one of the five serotypes (A, B, C, D, and AD) defined based on their cell surface antigenic properties within the pathogenic cryptococcal species complex.

Like those of other microorganisms in nature, environmental conditions for *C. neoformans* may change and fluctuate significantly, both spatially and temporally, in nutrient levels, temperature, and water availability etc. Under conditions of low nutrient availability, low moisture, and a temperature in the range of 15-33°C, mating and sexual reproduction between strains of opposite mating types may occur, resulting in the production of basidiospores. These sexual spores are smaller than vegetative cells and can disperse easily by wind or other means to other environmental niches, including human hosts.

The mating system of this heterothallic basidiomycete is controlled by one locus with two alternative alleles: *MATa* and *MAT $\alpha$*  (Kwon-Chung, 1975; Kwon-Chung, 1976).

Previous studies have shown that the majority of clinical and environmental isolates belong to the *MAT $\alpha$*  mating type (Kwon-Chung & Bennett, 1978; Yan et al., 2002). Interestingly, a laboratory *MAT $\alpha$*  strain that belonged to serotype D (one of three serotypes, A, D, and AD, in *C. neoformans*) has been shown to be more virulent than its congenic *MAT $\alpha$*  strain in a murine model of cryptococcosis (Kwon-Chung et al., 1992). Although the details and the relative frequencies of mating in natural environments are little known, in the laboratory, both opposite sex mating (*MAT $\alpha$ -MAT $\alpha$* ) and same sex mating (*MAT $\alpha$ -MAT $\alpha$* ) have been demonstrated (Kwon-Chung, 1976; Keller et al., 2003; Nielsen et al., 2003; Lin et al., 2005). A recent study indicated that the *MAT $\alpha$*  locus was among several genomic regions that contribute quantitatively to same sex mating and haploid fruiting in *C. neoformans* (Lin et al., 2006). Evidence for same-sex mating (*MAT $\alpha$ -MAT $\alpha$* ) in nature was recently reported between strains of serotypes A and D (Lin et al., 2007). Same-sex mating has also been hypothesized to be responsible for generating the major genotype of a recent outbreak of cryptococcosis on Vancouver Island, British Columbia, Canada (Fraser et al., 2005).

Sexual reproduction, either between strains of opposite sexes or the same sex, can potentially increase genetic variation in natural populations of organisms. The diversity of genotypes resulting from sexual reproduction can increase the efficiency of natural selection and may enhance the chances of successfully colonizing novel environments (Weismann, 1904; Goddard et al., 2005). However, unlike those in the majority of plants and animals where sexual reproduction can be easily detected, the detection of mating and sexual reproduction in microorganisms requires the analysis of genotypic data from multiple strains in populations. Associations among alleles from within the same or

different loci are used as indicators for sexual or asexual reproduction, typically through the analyses of multilocus linkage disequilibrium (Xu et al., 1999; Xu 2005b). In these tests, random associations (i. e. linkage equilibrium) among alleles at the population level are consistent with sexual reproduction while significant associations (linkage disequilibrium) are indicative of clonal reproduction. Other signatures of clonal reproduction include over- or under- representations of certain genotypes and genealogical congruence among genes from unlinked loci (e.g. Xu, 2005b).

Previous studies have employed a variety of molecular markers to demonstrate the modes of microbial reproduction in nature and to determine how evolutionary as well as ecological factors can influence the modes of reproduction and the structures of microbial populations (Maynard-Smith et al., 1993, Avise, 1994; Xu et al., 2005b), including those of *C. neoformans*. For example, a MLST study of the population structure of *C. neoformans* var. *grubii* involving 102 representative strains from different parts of the world showed that the global population contained three genetically distinct subgroups (Litvintseva et al., 2006). Ten of the 102 strains were *MAT $\alpha$*  and they were all from Botswana. The remaining 92 were *MAT $\alpha$*  and were from different parts of the world, including Botswana. As expected, the Botswana population that contained strains of both mating types showed evidence of recombination and sexual reproduction (Litvintseva et al., 2003; 2006). In contrast, the remaining geographic populations that contained only the *MAT $\alpha$*  strains were largely clonal, with no clear evidence of recombination. Similar analyses of natural populations of a closely related species *Cryptococcus gattii* that contained exclusively or predominantly *MAT $\alpha$*  strains identified at least three distinct lineages in this species but no conclusive evidence for

recombination (Kidd et al., 2005). However, in another study of *C. gattii* populations, evidence for both clonality and recombination were identified, with clonality found for samples belonging to one molecular type VGI, and evidence of recombination found for samples of a different molecular type VGII (Campbell et al. 2005).

In this study, we analyzed 78 strains of *C. neoformans* var. *grubii* isolated from nine different tree species distributed in five geographical areas in northwestern India. The mating types of these strains were determined based on direct PCR using mating type specific primers (Yan et al., 2002) and all strains were found to have the *MAT $\alpha$*  mating type. A multiple gene genealogical approach was used to analyze the population samples. DNA sequences were obtained for each strain from each of the five genetic loci distributed in different parts of the genome. Our population genetic analyses identified no evidence of population subdivision based on either the host tree species or their geographic origins. Interestingly, while clonality and clonal dispersal among geographic regions were clearly evident in our analyzed populations, there was also unambiguous evidence for recombination in this *MAT $\alpha$*  population of *C. neoformans* var. *grubii*.

## Materials and Methods

**Samples of *C. neoformans*.** The decaying wood inside the trunk hollows of nine species of trees was sampled for *Cryptococcus neoformans* (serotypes A and D) between 2003 and 2006. Strains were collected by taking swabs from the decayed wood and inoculating the swabs on modified Niger Seed Agar (Randhawa et al., 2005). The colonies were screened for morphological (microscopy) and physiological characteristics consistent with *C. neoformans* (VITEK 2 system, bioMérieux, l’Etiolle, France). Isolates

were then serotyped using the Crypto-check slide agglutination test (Iatron, Tokyo, Japan). Details of the seventy-eight strains obtained for this study are presented in Table 1. The strains analyzed here are available upon request from Drs. Harbans Randhawa and JP Xu.

**DNA Extraction and PCR Amplification.** A previously described technique (Xu et al., 2000) was used to extract high molecular weight genomic DNA from all 78 strains. Fragments of four nuclear DNA genes and one mitochondrial gene were chosen for PCR amplification and analyses. To avoid the potential bias in linkage disequilibrium due to physical linkage between markers, the five marker loci used in this study were located on separate chromosomes, as determined based on the completed JEC21 genome (Loftus et al., 2005). The nuclear DNA markers used for amplification were: (1). *CAP1* (on chromosome 4) located within the mating locus and encodes a capsule-synthesis associated protein; (2). *FTR1* (on chromosome 3) which encodes a high affinity iron permease; (3). *LAC* (on chromosome 7) which encodes the laccase/diphenol oxidase; (4). the internal transcribed spacer (*ITS*) region which comprises ITS1, 5.8S rRNA and ITS2 (on chromosome 2); and (5). the mitochondrial large subunit RNA (*mtLrRNA*) (Xu et al. 2000). The JEC21 genome assembly from the Stanford Genome Technology Center (<http://www.tigr.org/tdb/e2k1/cna1/>) was used to identify the loci and design primers when required. Fragments of 731bp (*CAP1*), 788bp (*FTR1*), 486bp (*LAC*), 468bp (*ITS*) and 439bp (*mtLrRNA*) were obtained by using the following primer pairs (5’->3’): *CAP1*, CGTTCGCGATAGAGAGAGGA (forward) and CCTTACCTTCACAGTCGCCC (reverse); *FTR1*, GTTCTCGGTCACCATCTTC (forward) and TCTCAGGCTCGCCATCTTC (reverse); *LAC*, GGCGATACTATTATCGTA (forward)



and TTCTGGAGTGGCTAGAGC (reverse); *ITS*, TCCGTAGGTGAACCTGCGG (forward) and TCCTCCGCTTATTGATATGC (reverse); *mtLrRNA*, CGCTAAATACGATCCTGCATG (forward) and TTATCCCTAGCGTAACTTTTATC (reverse). Each PCR reaction comprised of 2.1µL (~1ng) template DNA, 8µL *GoTaq* (*Taq* DNA polymerase + MgCl<sub>2</sub> at a final concentration of 1.5mM, supplied by Promega, USA), and 0.2µM of each primer to make up a total volume of 16µL. The thermo cycling profile included an initial denaturing temperature of 95°C for 4min followed by 40 cycles of 95°C for 1min, annealing temperature for 1 min, and 72°C for 1min. The final extension step was 72°C for 7min. Annealing temperatures for amplifying the five genes were: 59.2°C for *CAP1*, 58.6°C for *FTR1*, 50°C for *LAC*, 50°C for *ITS*, and 46°C for *mtLrRNA*. The resulting PCR products were purified using the MicroCLEAN kit (Microzone, Ottawa, Canada) and sequenced at the MoBix Laboratory, McMaster University, using the ABI3100 automated DNA sequencer.

**Mating type determination using molecular methods.** The mating types of isolates were determined by PCR amplification using primer pairs designed from the sequences of the mating type-specific *STE12* and *STE20* genes. These primers were serotype and/or mating type-specific. The specific primer sequences, amplification conditions, gel electrophoresis and data screening followed those described in Yan et al. (2002).

**Phylogenetic analysis.** Sequences from the 78 strains for each gene locus were aligned using Clustal X version 2 (Thompson et al., 1997). The alignments were then imported into Seaview (Galtier et al., 1996) and edited manually. Orthologs for the five genes in two model strains were retrieved for comparison from the completed genomes of

Serotype A (H99 strain, [http://www.broad.mit.edu/annotation/genome/Cryptococcus\\_neoformans](http://www.broad.mit.edu/annotation/genome/Cryptococcus_neoformans)) and Serotype D (JEC21 strain, <http://www.tigr.org/>). Phylogenetic analysis was performed using PAUP\*4.0b10 (Swofford, 2002). Maximum-parsimony trees were constructed for all the individual fragments by conducting heuristic searches based on 500 random sequence additions. Bootstrap analyses to assess the support for the clades were also conducted using 1000 replicate samples of phylogenetically informative characters. The serotype D strain JEC21 was used as an out-group for all the analysis. For the LAC gene fragment where extensive sequence information is available and is used to delineate the molecular subtypes (VNI, VNII, and VNB) within *C. neoformans* var. *grubii*, strains representative of each of the three types from Litvintseva et al. (2003) are analyzed together with our data.

**Clonality and recombination.** Three complementary tests were conducted to examine evidence for clonality and recombination in the Indian population of *C. neoformans* var. *grubii*. In the first test, we examined evidence for genealogical incongruence among the five genes. This test was conducted using the partition homogeneity test (PHT) or incongruence length difference (ILD) test (Farris et al., 1994) implemented in PAUP\*4.0b10 (Swofford, 2002). The assumption made in this test is that the whole genome evolves as one single unit in strictly clonally reproducing organisms and as a result, genealogies of different gene fragment should be congruent in such organisms. In contrast, in case of a recombining population structure, there would be incongruence among the different gene genealogies because different genes could have different histories. For this test, we used 100 replicates, each with 10 random sequence

addition replicates using the NNI branch swapping algorithm and saving a maximum of 1000 trees per replicate. This test was conducted for both the entire dataset and also the two largest individual populations, from Delhi and Amritsar, respectively. It has been shown that relying on this test alone for inferences of recombination could produce false-positive results when the consistency indices for individual gene trees are not high (Barker and Lutzoni 2002). To avoid the potential problem, we also conducted the following two population genetic tests for evidence of recombination.

In the second test, we examined allelic associations among alleles from different loci using two common statistics in population genetics: the index of association ( $I_A$ ) (Maynard-Smith et al., 1993) and RbarD (Agapow & Burt, 2001) tests.  $I_A$  and RbarD were estimated using Multilocus, version 1.0b (Agapow & Burt, 2001). The  $I_A$  analyzes the variance of the distances between all possible pairs of multilocus genotypes. The null hypothesis is that there is random association of alleles. Statistical significance was derived by comparing the observed dataset to 10000 artificially recombined datasets. During the process of generating the artificially recombined datasets, we permitted random shuffling of alleles at each locus within a population but kept the proportions of alleles at each locus constant. If there were linkage disequilibrium due to a lack of recombination, the observed  $I_A$  should be significantly higher than the randomized recombining data set. RbarD is a modified measure of  $I_A$ , but adjusted for the number of loci. The adjustment helps facilitate better comparison between populations and studies.

In the third test, we calculated the phylogenetic incompatibility by looking for the proportion of loci with all possible recombinant genotypes. This test looks for the evidence of recombination in a population by essentially looking for incongruence in

different regions of the genome. The statistical significance of this test was inferred by comparing the number of incompatible pairs of loci in the observed dataset to those from a randomly shuffled dataset, using the program Multilocus, version 1.0b (Agapow & Burt, 2001).

**Population structure.** We assessed the genetic structure of the Indian samples of *C. neoformans* using three different analyses. In the first, we conducted the T-PTP (topology-dependent permutation tail probability) test (Faith, 1991) to determine whether host tree species or the geographical locations of strains had a significant influence on the topologies of phylogeny. For this test, reshuffled datasets were generated using Multilocus version 1.0b (Agapow & Burt, 2001) and the tree lengths were estimated using PAUP\* version 4.0b10 (Swofford, 2002). If there was no significant structure identified with respect to either host tree species or geographic origins, the optimal tree score for the observed data without any constraint would be significantly shorter than trees with constrained topologies. In the second test, population differentiation analysis was done by calculating  $\theta$  (Weir, 1996), using Multilocus version 1.0b (Agapow & Burt, 2001). The null hypothesis of this test was no genetic differentiation between populations as defined by either host tree species or geographic origins. Statistical significance for this test was derived by comparing the observed dataset to 10000 randomized datasets. During randomization, any linkage disequilibria present in the observed data were maintained. The statistic  $\theta$  ranges from 0 (for no population differentiation due to frequent gene flow) to 1 (total isolation of the populations from each other). In the third analyses, we examined the relationship between genetic distance and geographic distance between populations, using the Isolation by Distance Web Service

(<http://ibdws.sdsu.edu/>) (Jensen et al., 2005). A spatially limited gene flow would result in decreased genetic similarity between populations as their geographic distances increase and this is commonly referred to as isolation by distance (IBD) (Wright, 1938). This test plots and correlates genetic distances and geographic distances between pairs of populations. The genetic distance measure used here is the standard Nei’s genetic distance (D) between populations, established based on either host tree species or geographic origins.

**Mating assays.** The abilities to mate and reproduce sexually were examined for representative strains in this study. Briefly, twenty strains representing all 15 multilocus genotypes (MG) identified in this study were chosen (see below; these strains are marked in Table 1). Each of the 20 strains was mated with three standard mating testers for *C. neoformans* independently. The three testers were JEC20 (*MATa*), JEC21 (*MAT $\alpha$* ), and H99 (*MAT $\alpha$* ). JEC20 and JEC21 belong to *C. neoformans* var. *neoformans* (serotype D) while H99 belongs to *C. neoformans* var. *grubii* (serotype A). To prepare mating, strains were first grown on YEPD (yeast extract-peptone-dextrose) medium at room temperature for 3 days. Approximately equal number of cells from mating partners were then mixed on V8-juice agar medium (5% [v/v] V8 juice, 3 mM  $\text{KH}_2\text{PO}_4$  and 4% [w/v] agar at pH 7). The plates were then incubated in the absence of light at room temperature ( $\sim 22^\circ\text{C}$ ) for 21 days. For negative controls, we separately inoculated the three tester strains as well as the 20 representative strains from India onto V8-Juice agar media and incubated under the same condition as described above for mating mixtures. For positive controls, we mated JEC20 and H99. In addition, we constructed several pairs of same-sex mating

between strains from India (all are MAT $\alpha$ ). Filamentation and basidiospore formation were recorded under light microscopy for each of the mating and controls.

## Results

### Molecular variation

We were able to amplify all five DNA fragments for all the seventy-eight isolates collected for this study. The results of PCR using the *STE12* and *STE20* gene primers showed that all isolates belonged to the  $\alpha$  mating type. Our direct sequencing of PCR products identified no ambiguous nucleotide sites at any of the five genes for any of the 78 strains, consistent with haploidy. In total, we obtained 2912 nucleotides for each strain. The entire alignment had 11 parsimony informative sites and 5 parsimony uninformative sites. No polymorphic site was observed within the sequenced regions for the *ITS* and *mtLrRNA* gene fragments. The GenBank accession numbers for the 390 DNA sequences are: *CAP1*: EU312204-EU312281; *FTR1*: EU312282-EU312359; *ITS*: EU312360-EU312437; *LAC*: EU312438-EU312515; *MtLrRNA*: EU312516-EU312593.

We used both maximum parsimony and maximum likelihood analyses to construct the evolutionary relationships among strains for each of the three gene loci that contained nucleotide polymorphisms, both individually and in combination (Fig. 1). Also see attached supplementary material (Fig. S1a, S1b & S1c). Both analyses generated identical phylogenetic patterns. All the isolates clustered more closely with the *C. neoformans* var. *grubii* model strain H99 (serotype A) than to the *C. neoformans* var. *neoformans* model strain JEC21 (serotype D), consistent with our serological identification that all the strains analyzed here belonged to serotype A. Aside from the

LAC phylogeny, midpoint rooting was used to display the topologies of all the phylogenies to increase resolution among genotypes (Fig. S1). However, we would like to mention that the topologies did not change when we rooted the trees using the orthologs from the serotype D strain JEC21 for either CAP1 or FTR1 gene tree.] In each of the three polymorphic genetic loci, we saw no obvious evidence of strain clustering based either on geographic origin or host tree species. Indeed, all major multilocus genotypes are shared among strains from multiple geographic regions or host tree species.

Among the 78 strains, the number of haplotypes for the individual gene fragments were as follows: four for *LAC*, three for *CAP1*, eight for *FTR1* (Fig. S1a, S1b, and S1c), and one each for *ITS* and *mtLrRNA* (Table 1). In total, the entire dataset yielded 16 multilocus genotypes among the seventy-eight strains (Fig. 1). Almost all the genotypes identified here were represented in the largest subpopulation from Delhi (Fig. 1). The second largest subpopulation (Amritsar) had two haplotypes for each of *CAP1*, *FTR1*, and *LAC*.

To compare the multilocus genotypes identified here with those reported previously, we compared our sequences with those in the GenBank using BLASTn. Our analyses identified that all our sequences had the closest matches to sequences of strains of *C. neoformans* var. *grubii*, consistent with our expectation. Of the three gene fragments analyzed here, *LAC* has been the most widely used in previous studies. As a result, a significant number of allelic sequences exist for the *LAC* gene in the GenBank and some of these sequences have been used to define several molecular types within *C. neoformans* (Litvintseva et al. 2003; 2006). We therefore analyzed our sequences with

eight representative sequences from the three different molecular types to determine the likely molecular type(s) of our strains. The eight representative sequences of the LAC gene included five of VNI, two of VNB, and one of VNII. Our analyses identified that all 78 strains showed the highest similarity to strains of the VNI molecular type (Fig. S1a).

### **Lack of genetic differentiation between populations**

The  $\theta$  value, a measure of population subdivision, showed no significant differentiation among geographic or host tree-based populations and the observed value lies well within the distribution of the randomized datasets (only results based on geographic origins are shown in Fig. 2). The amount of genetic differentiation observed between populations based on either geographic location (Table 2a) or host tree species (Table 2b) was also measured by Nei’s D and we observed no significant difference between these populations. The lack of significant genetic differentiation is also supported by a lack of monophyletic pattern in the T-PTP tests based on either geographic origins or host tree species, with T-PTP values of 1.00 and 0.60 respectively. While we observed a positive correlation between geographic distance and pair-wise population genetic distance, the correlation was statistically not significant (Data not shown).

### **Evidence for both clonality and recombination**

As was described above, the evidence for clonality and recombination was examined using three tests. The first was the partition homogeneity test or incongruence length difference test. This test showed no statistically significant heterogeneity among



the five markers used in this study. This test was conducted for the total sample of 78 strains as well as separately for each of the two largest geographic populations (Delhi and Amritsar). None of the samples showed any significant incongruence between genealogies from different markers [Entire dataset (n=78):  $p=0.27$ ; Delhi (n=47):  $p=0.23$ ; Amritsar (n=21):  $p=0.75$ ]. This result suggests that the overall population structure is consistent with significant clonality and clonal expansion. We also found other evidence for clonality based on the distribution of different genotypes between different geographic locations (Fig. 1). For example, two multilocus genotypes (MG), MGIX and MGXI, were over-represented in the total sample with 41 and 10 isolates each, both from different host trees and geographic areas (Fig. 1). Furthermore, with the exception of MGIV and MGXVI (Fig. 1), all the other genotypes in our study were represented in the strains from Delhi. The MGXVI genotype was recovered from three different cities: Amritsar, Bulandshahar and Hathras. The lone genotype MGIV was from Bulandshahar. While genotypes MGI, MGII, MGIII, MGVI, MGVIII, MGX, MGIV and MGXV were exclusive to Delhi, the remaining genotypes were each sampled from more than one geographic location. This pattern of distribution and the sharing of genotypes between different geographic locations are consistent with long-distance dispersal and clonal expansion within and among the geographic populations of *C. neoformans* var. *grubii* from India.

The second test assessed the overall associations among alleles at different loci. Our population genetic analyses revealed that both  $I_A$  and  $R_{barD}$  values were low for both the total population sample as well as for the two largest subpopulations (Delhi and Amritsar) (Table 4). In fact, the null hypothesis of random recombination could not be

rejected and the  $I_A$  values were well within the randomized recombining datasets (Fig. 3). We repeated these tests for samples specific to each host tree species and all failed to reject the null hypothesis of random recombination. More importantly, our third test using phylogenetic incompatibility also found clear evidence for recombination. Specifically, among three polymorphic loci *LAC*, *CAP1* and *FTR1*, we found all possible allelic combinations between pairs of loci for several alleles (raw data in Table 1 and examples shown in Table 3). Our results thus suggest that recombination is occurring within all populations analyzed in this study.

### **Fertility of *C. neoformans var. grubii* strains from India**

Among the 60 crosses that we set up between 20 representative strains and the three testers, only seven successfully produced hyphal filaments and basidiospores. All seven were with strain JEC20 (the *MATa* mating tester) and none with JEC21 or H99. The results are consistent with our molecular determination of mating types and suggest that at least some of the strains are sexually fertile. While the positive control cross JEC20xH99 mated successfully, none of the negative controls worked. In addition, after more than three weeks of incubation, none of the crosses between strains of the same sex, *MAT $\alpha$* , produced any hyphal filaments.

### **Discussion**

*C. neoformans var. grubii* is among the most important fungal pathogens in humans. Understanding its population structure and ecology in natural environments could have significant medical implications. In this study, our population genetic

analyses showed evidence for extensive gene flow among the environmental populations of this organism in India. Furthermore, we identified evidence for both clonality and recombination in this sample that contained exclusively the *MAT $\alpha$*  mating type. Our results are consistent with the hypothesis that *C. neoformans* might be able to complete its sexual life cycle on decaying wood of several tree species in India. A recent study by Xue et al. (2007) demonstrated that mating and sexual reproduction between strains of opposite mating types can occur for both *C. neoformans* and *C. gattii* on the model plant *Arabidopsis thaliana* and a natural tree host of *C. gattii*, the Eucalyptus trees.

All of the host tree species investigated in this study are native to India. These tree species included *Syzygium cumini* (Java plum /Jamun tree), *Polyalthia longifolia* (Indian mast tree), *Mimusops elengi* (Bullet wood), *Azadirachta indica* (Neem tree), *Manilkara hexandra* (Margosa), *Acacia nilotica* (Thorn mimosa), *Cassia fistula* (Golden shower tree), and *Mangifera indica* (Mango). In particular, one tree species, *Polyalthia longifolia*, is found around many homes throughout India. All the trees from which *C. neoformans* var. *grubii* was sampled were very close to human habitation. In addition to these tree species, *C. neoformans* var. *grubii* has also been recovered from the hollows of other tree species such as *Cassia grandis* (Pink shower tree), *Senna multijuga* (November shower tree), and *Ficus microcarpa* (Fig tree) (Lazera et al., 1996). Our results demonstrated that the populations of *C. neoformans* var. *grubii* in different tree species are not genetically different and that there is frequent gene flow among different tree species. The spread of this microorganism between different hosts could have been mediated by wind-aided dispersal of basidiospores or desiccated vegetative cells. In this study, the five cities from where the strains were isolated were among the most densely

populated regions in India. This is especially true of Delhi where human population density and migration are both very high. The high rate of human migration could also contribute to the dispersal of *C. neoformans* var. *grubii* among different locations. In addition, other tree dwelling animals in these regions such as birds and squirrels can potentially aid in the spread and dispersal of this organism. At the global level, there is evidence for limited genetic differentiation and frequent gene flow among geographic populations of *C. neoformans* var. *grubii* (Xu et al., 2000; Litvintseva et al., 2003; 2006). Short- and long distance dispersals have been reported previously for many fungal species, including saprobes, and plant and animal pathogens (Xu, 2005b).

The limited genetic variation found here among strains from the different hosts and geographic regions suggests that the Indian population of *C. neoformans* var. *grubii* might be of recent origin and/or that the dispersals were very frequent and recent. While the widespread over-representation of certain genotypes is consistent with clonality, the identification of phylogenetic incompatibility and low  $I_A$  values also suggested the presence of recombination in the Indian environmental population of *C. neoformans* var. *grubii*. These results are consistent with the structures of many microbial populations, characterized by unambiguous recombination, clonal dispersal and expansion of a few genotypes (e.g. Maynard-Smith et al., 1993; Xu et al. 2005b).

The evidence for recombination in a *MAT $\alpha$*  population of *C. neoformans* var. *grubii* is surprising. Previous analyses of samples of *C. neoformans* var. *grubii* have indicated that clonality was the dominant feature of reproduction in natural populations. This was somewhat expected because in most previous studies, the strains analyzed from either human or environmental sources belonged to mating type  $\alpha$ . There were two

exceptions where evidence for recombination had been detected in *C. neoformans* var. *grubii*. The first was an analysis of 14 hybrid strains of serotypes AD where both mating types *a* and  $\alpha$  strains from both serotypes A and D were reconstructed using gene genealogies. The reconstructed serotype A and D populations both showed evidence of recombination (Xu & Mitchell, 2003). The second example was from humans in Botswana where both mating types were found and not surprisingly, evidence for recombination as well (Litvintseva et al., 2003).

Our current study is unique in that there was clear evidence for recombination in a sample where all the isolates belonged to one mating type, the *MAT $\alpha$*  mating type. At present, the detailed mechanism for our observed recombination is not known. However, there are several possibilities. The first is that, unlike most previous studies (e.g. Litvintseva et al., 2003; Xu et al. 2000), the samples analyzed here were all from a geographically more restricted region in northwestern India. The close geographic proximity could have contributed to frequent mating and genetic mixing and resulting in recombinant genotypes. The second is that all strains analyzed here are genetically very similar to each other. In fact, all 78 strains here were found to belong to the same molecular type VNI. Strains with similar genotypes might have a greater opportunity of producing viable genetic recombinants capable of surviving in nature, as suggested previously by Campbell et al. (2005). The third is that the environmental niche, the tree hollows, might be highly conducive for mating. A recent study indicated that both *C. neoformans* and *C. gattii* could complete their sexual reproductive life cycle in plants (Xue et al. 2007). Most previous studies of environmental samples of *C. neoformans* var. *grubii* have focused on pigeon guano. Whether populations of *C. neoformans* var. *grubii*

from different ecological niches (e.g. pigeon guano and soil) in India show population structures similar to that from tree hollows remain to be examined.

Sexual recombination detected here could be achieved through one of two pathways. The first is that recombination observed here was due to same-sex mating between *MAT $\alpha$*  strains in natural populations of *C. neoformans* var. *grubii*. Same sex mating has been observed in the laboratory (Lin et al., 2005; Yan et al., 2007) and in nature between strains of serotypes A and D (Lin et al. 2007). The second hypothesis is that the recombination observed here was between strains of opposite mating types. Even though we were unable to isolate any *MATa* strains from India, it’s possible that *MATa* strains are there but are very rare in nature in India and difficult to isolate. Our result that some of these isolates can mate with *MATa* but not *MAT $\alpha$*  strains in the lab suggests that these *MATa* strains might be present in the environment in India to help maintain the sexual fertility of *MAT $\alpha$*  strains. However, more extensive sampling would be required to distinguish between these two hypotheses. We would like to emphasize that the evidence for recombination was found not only for the whole sample but also for the subpopulations in Delhi as well as Amristar. This finding suggests that many geographic regions in India (and potentially other places in other parts of the world) could be hotspots for sexual mating and the dispersal of basidiospores of this organism.

A combined sexual and asexual mode of reproduction is common for many microbial species, including many human pathogens (Xu, 2005b). It was recently shown that sexual reproduction could incur a cost in fitness (Xu, 2005a). It is also well known that sexual reproduction could produce a large number of recombinant genotypes that could speed up adaptation to novel environments (Weismann, 1904; Goddard et al.,

2005). In *C. neoformans*, while mating could occur between strains of the same or different sexes, it remains to be seen if the two alternative forms of mating have different adaptive significance. In addition, we know very little of the specific conditions in nature under which this organism undergoes same-sex mating as opposed to mating between opposite mating types. In the laboratory setting, low nitrogen, low moisture, and ambient temperature are conducive for mating. The nutritional and other physical/chemical features of tree hollows from where *C. neoformans* were isolated are unknown at present. Ecological and physiological studies of natural environments conducive for *C. neoformans* mating will provide further insights into the mechanisms of evolution and population structuring of this important fungal pathogen.

### **Acknowledgements**

This work was supported by grants from the Natural Science and Engineering Research Council of Canada (JX), Genome Canada (JX), and the Premier’s Research Excellence Award (JX). We thank the Indian National Science Academy for the award of a Honorary Scientist position to H.S.R. and to the Department of Science & Technology, New Delhi, for the award of a Young Scientist research grant to A.C.

## References

1. **Agapow, P.M. & Burt, A. (2001)** Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* **1**, 101-102.
2. **Avise, J.C. (1994)** *Molecular Markers, Natural History and Evolution* Chapman & Hall: New York.
3. **Barker, F.K. and Lutzoni, F. M. (2002).** The utility of the incongruence length difference test. *Syst Biol.* **51**, 625-637.
4. **Bennett, J.E., Kwon-Chung, K.J. & Howard, D.H. (1977)** Epidemiologic differences among serotypes of *Cryptococcus neoformans*. *Am J Epidemiol* **105**, 582–586.
5. **Burt, A., Carter, D.A., Koenig, G.L., White, T.J. & Taylor, J.W. (1996)** Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *P Natl Acad Sci USA* **93**, 770-773.
6. **Campbell, L.T., Currie, B.J., Krockenberger, M., Malik, R., Meyer, W., Heitman, J. & Carter, D. (2005)** Clonality and recombination in genetically differentiated subgroups of *Cryptococcus gattii*. *Euk Cell* **8**, 1403-1409.
7. **Casadevall, A. & Perfect, J.R. (1998)** *Cryptococcus neoformans*. ASM Press: Washington, DC.
8. **Casadevall, A., Steenbergen, J.N. & Nosanchuk, J.D. (2003)** Ready-made virulence and dual use virulence factors in pathogenic environmental fungi—the *Cryptococcus neoformans* paradigm. *Curr Opin Microbiol* **6**, 332–337.
9. **Faith, D.L. (1991)** Cladistic permutation tests for monophyly and nonmonophyly. *Syst Zool* **40**, 366-375.



10. **Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C. (1994)** Testing significance of incongruence. *Cladistics* **10**, 315-319.
11. **Fraser, J.A., Giles, S.S., Wenink, E.C., Geunes-Boyer, S.G., Wright, J.R., Diezmann, S., Allen, A., Stajich, J.E., Dietrich, F.S., Perfect, J.R. & Heitman, J. (2005)** Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. *Nature*, **437**, 1360-1364.
12. **Galtier, N., Gouy, M. & Gautier, C. (1996)** SeaView and Phylo\_win, two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**, 543-548.
13. **Goddard, M.R., Godfray, H.C.J. & Burt, A. (2005)** Sex increases the efficacy of natural selection in experimental yeast populations. *Nature*, **434**, 636–640.
14. **Jensen, J.L., Bohonak, A.J. & Kelley, S.T. (2005).** Isolation by distance, web service v.3.14. *BMC Genetics* **6**, 13.
15. **Keller, S. M., Viviani, M. A. Esposito, M. C. Cogliati, M. & Wickes, B. L. (2003).** Molecular and genetic characterization of a serotype A MATa *Cryptococcus neoformans* isolate. *Microbiology* **149**, 131-142.
16. **Kidd, S.E., Guo, H, Bartlett, K.H., Xu, J. & Kronstad, J.W. (2005).** Comparative gene genealogies indicate that two clonal lineages of *Cryptococcus gattii* in British Columbia resemble strains from other geographical areas. *Eukaryot Cell* **10**, 1629-1638.
17. **Kwon-Chung, K. J. (1975).** A new genus, *Filobasidiella*, the perfect state of *Cryptococcus neoformans*. *Mycologia* **67**, 1197–1200.

18. **Kwon-Chung, K. J. (1976).** Morphogenesis of *Filobasidiella neoformans*, the sexual state of *Cryptococcus neoformans*. *Mycologia* **68**, 821–833.
19. **Kwon-Chung, K. J. & Bennett, J. E. (1978).** Distribution of  $\alpha$  and a mating types of *Cryptococcus neoformans* among natural and clinical isolates. *Am J Epidemiol* **108**, 337-340.
20. **Kwon-Chung, K.J., Wickes, B.L., Stockman, L., Roberts, G.D., Ellis, D. & Howard, D.H. (1992).** Virulence, serotype, and molecular characteristics of environmental strains of *Cryptococcus neoformans* var. *gattii*. *Infect Immun* **60(5)**, 1869-1874.
21. **Lazera, M.S., Pires, F.D., Camillo-Coura, L., Nishikawa, M.M., Bezerra, C.C., Trilles, L. & Wanke, B. (1996)** Natural habitat of *Cryptococcus neoformans* var. *neoformans* in decaying wood forming hollows in living trees. *J Med Vet Mycol* **34**, 127-131.
22. **Lin, X., Hull, C.M. & Heitman, J. (2005).** Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* **434**, 1017-21.
23. **Lin, X., Huang, J.C., Mitchell, T.G., & Heitman, J. (2006).** Virulence attributes and hyphal growth of *C. neoformans* are quantitative traits and the *MAT* $\alpha$  allele enhances filamentation. *PLoS Genetics* **2**, e187.
24. **Lin, X., Litvintseva, A. P., Nielsen K, Patel, S., Floyd, A., Mitchell, T. G. & Heitman J. 2007.**  $\alpha$ AD $\alpha$  hybrids of *Cryptococcus neoformans*: evidence of same-sex mating in nature and hybrid fitness. *PLoS Genetics* **3**, 1975-1990.

25. **Litvintseva, A.P., Marra, R.E., Nielsen, K., Heitman, J., Vilgalys, R.J. & Mitchell, T.G. (2003).** Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. *Eukaryot Cell* **2**, 1162-1168.
26. **Litvintseva, A.P., Thakur, R., Vilgalys, R. & Mitchell, T.G. (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.
27. **Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A., Allen, J.E., Bosdet, I.E., Brent, M.R., Chiu, R., Doering, T.L., Donlin, M.J., D'Souza, C.A., Fox, D.S., Grinberg, V., Fu, J., Fukushima, M., Haas, B.J., Huang, J.C., Janbon, G., Jones, S.J., Koo, H.L., Krzywinski, M.I., Kwon-Chung, J.K., Lengeler, K.B., Maiti, R., Marra, M.A., Marra, R.E., Mathewson, C.A., Mitchell, T.G., Perlea, M., Riggs, F.R., Salzberg, S.L., Schein, J.E., Shvartsbeyn, A., Shin, H., Shumway, M., Specht, C.A., Suh, B.B., Tenney, A., Utterback, T.R., Wickes, B.L., Wortman, J.R., Wye, N.H., Kronstad, J.W., Lodge, J.K., Heitman, J., Davis, R.W., Fraser, C.M. & Hyman, R.W. (2005).** The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321-1324.
28. **Maynard-Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993).** How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384-4388.
29. **Nei, M. (1972)** Genetic distance between populations. *Am Nat* **106**, 283-292.

30. **Nielsen, K., Cox, G.M., Wang, P., Toffaletti, D.L., Perfect, J.R., & Heitman, J. (2003).** Sexual cycle of *Cryptococcus neoformans* variety *grubii* and virulence of congenic **a** and alpha isolates. *Infect Immun* **71**, 4831-4841.
31. **Randhawa, H.S., Kowshik, T. & Khan, Z.U. (2003)** Decayed wood of *Syzygium cumini* and *Ficus religiosa* living trees in Delhi / New Delhi metropolitan area as natural habitat of *Cryptococcus neoformans*. *Med Mycol* **41**: 199-209.
32. **Randhawa, H.S., Kowshik, T. & Khan, Z.U. (2005).** Efficacy of swabbing versus a conventional technique for isolation of *Cryptococcus neoformans* from decayed wood in tree trunk hollows. *Med Mycol* **43**, 67-71.
33. **Reimão, J.Q., Drummond, E.D., Terceti, Mde. S., Lyon, J.P., Franco, M. C. & de Siqueira, A.M. (2007).** Isolation of *Cryptococcus neoformans* from hollows of living trees in the city of Alfenas, MG, Brazil. *Mycoses* **50**, 261-264.
34. **Swofford, D. L. (1996)** Phylogenetic inference. In *Molecular Systematics*, pp. 407–514. Edited by D. Hill, C. Moritz & B. Mable: Sinauer Associates, Inc., Sunderland, MA.
35. **Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997).** The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**, 4876-4882.
36. **Tintelnot, K., Lemmer, K., Losert, H., Schar, G. & Polak. A. (2004).** Follow-up of epidemiological data of cryptococcosis in Austria, Germany and Switzerland with special focus on the characterization of clinical isolates. *Mycoses* **47**, 455-464.
37. **Weir, B. S. (1996).** *Genetic data analysis II*: Sinauer, Sunderland.

38. **Weismann, A. (1904).** *The evolution theory*, 2 vols. Translated from the 1904 2<sup>nd</sup> German edition by J. A. Thomson & M.R. Thomson. London:Edward Arnold.
39. **Wright, S. (1938).** Size of population and breeding structure in relation to evolution. *Science* **87**, 430-431.
40. **Xu, J., Mitchell, T.G. & Vilgalys, R. (1999).** PCR-restriction fragment length polymorphism (RFLP) analyses reveal both extensive clonality and local genetic differentiation in *Candida albicans*. *Mol Ecol* **8**, 59-73.
41. **Xu, J., Vilgalys, R. & Mitchell, T.D. (2000).** Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol Ecol* **9**, 1471-1481.
42. **Xu, J. & Mitchell, T.G. (2003).** Comparative gene genealogical analyses of strains of serotype AD identify recombination in populations of serotypes A and D in the human pathogenic yeast *Cryptococcus neoformans*. *Microbiology* **149**, 2147-2154.
43. **Xu, J. (2005a)** Cost of interacting with sexual partners in a facultative sexual microbe. *Genetics* **171**, 1597-1604.
44. **Xu, J. (2005b)** *Evolutionary genetics of fungi*. **Horizon Bioscience: U.K.**
45. **Xue, C., Tada, Y., Dong, X. & Heitman J. (2007).** The human fungal pathogen *Cryptococcus* can complete its sexual cycle during a pathogenic association with plants. *Cell Host Microbe* **1**, 263–273.
46. **Yan, Z., Li, X. & Xu, J. (2002).** Geographic distribution of mating type alleles of *Cryptococcus neoformans* in four areas of the United States. *J Clin Microbiol* **40**, 965-972.

47. **Yan, Z., Hull, C. M., Sun, S., Heitman, J., and Xu, J.** (2007). The mating type-specific homeodomain genes *SX11 $\alpha$*  and *SX12a* coordinately control uniparental mitochondrial inheritance in *Cryptococcus neoformans*. *Current Genetics* **51**, 187-195.

**Table 1.** *Cryptococcus neoformans* var. *grubii* strains used in this study

Strain ID	Source tree/Geographical origin	Year Isolated	Multilocus genotype <sup>2</sup>				
INN3_279DDD_SC_DL_India <sup>1</sup>	Syzygium cumini /Delhi	2003	1	1	1	1	2
INN4_280U-DDD_SC_DL_India	Syzygium cumini /Delhi	2003	1	1	2	1	1
INN5_S284B_SC_DL_India	Syzygium cumini /Delhi	2003	1	2	2	1	1
INN6_320-1DDD'_SC_DL_India <sup>1</sup>	Syzygium cumini /Delhi	2004	1	3	1	1	1
INN8_320-2A_ii_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	4	2	1	1
INN9_320-3DDD'_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN11_320-4-DDD'_3_PL_DL_India <sup>1</sup>	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN12_322AA_1__PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN13_322AA_2__PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN14_323A-AA_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN15_323A-DDD_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN16_323B-AA_2__PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	1	1	1
INN17_323B-DDD'_1__PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN18_323B-DDD'_2__PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN19_325A_PL_DL_India <sup>1</sup>	Polyalthia longifolia/Delhi	2004	2	5	2	1	1
INN20_326B-DDD_1__PL_DL_India <sup>1</sup>	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN26_373DDD'_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN27_374A/2-10_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	1	2	1	1
INN28_374A/2-100_PL_DL_India	Polyalthia longifolia/Delhi	2004	1	3	2	1	1
INN29_540S-BB_1__PL_DL_India	Polyalthia longifolia/Delhi	2005	1	1	2	1	1
INN30_540S-BB_2__PL_DL_India	Polyalthia longifolia/Delhi	2005	1	5	2	1	1
INN33_572S_ME_DL_India	Mimusops elengi/Delhi	2005	1	1	2	1	1
INN34_572S-C_ME_DL_India	Mimusops elengi/Delhi	2005	1	1	2	1	1
INN35_580A-d_ME_DL_India	Mimusops elengi/Delhi	2005	2	3	2	1	1
INN36_580A-S_ME_DL_India	Mimusops elengi/Delhi	2005	2	3	2	1	1

“Master’s thesis – Sanjay S. Hiremath” “McMaster – Biology”

INN37_626S-C_AI_DL_India <sup>1</sup>	<i>Azadirachta indica</i> /Delhi	2005	1	6	2	1	1
INN38_627d-BB_AI_DL_India <sup>1</sup>	<i>Azadirachta indica</i> /Delhi	2005	2	1	2	1	1
INN39_628d-C_AI_DL_India <sup>1</sup>	<i>Azadirachta indica</i> /Delhi	2005	1	1	2	1	1
INN41_618A-d-C_MH_DL_India	<i>Manilkara hexandra</i> /Delhi	2005	1	1	2	1	1
INN43_618A-d-BB_MH_DL_India	<i>Manilkara hexandra</i> /Delhi	2005	1	1	2	1	1
INN44_618B-d-BB_MH_DL_India	<i>Manilkara hexandra</i> /Delhi	2005	2	1	2	1	1
INN46_611d-C_1__AN_DL_India	<i>Acacia nilotica</i> /Delhi	2005	1	1	2	1	1
INN48_692d-C_1__CF_DL_India	<i>Cassia fistula</i> /Delhi	2005	1	3	1	1	1
INN49_692d-C_2__CF_DL_India <sup>1</sup>	<i>Cassia fistula</i> /Delhi	2005	1	3	2	1	1
INN50_747S_2__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN51_748S_CF_DL_India <sup>1</sup>	<i>Cassia fistula</i> /Delhi	2006	1	2	2	1	1
INN52_748d_1__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	5	2	1	1
INN53_748d_2__CF_DL_India <sup>1</sup>	<i>Cassia fistula</i> /Delhi	2006	3	1	3	1	1
INN54_749S_1__CF_DL_India <sup>1</sup>	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN55_749S_2__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN56_750S_1__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	7	2	1	1
INN57_750S_2__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN58_751d_CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN59_753d_1__CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN60_753d_2__CF_DL_India <sup>1</sup>	<i>Cassia fistula</i> /Delhi	2006	1	1	2	1	1
INN61_757d_CF_DL_India	<i>Cassia fistula</i> /Delhi	2006	2	3	2	1	1
INN62_272DDD'_FR_DL_India <sup>1</sup>	<i>Ficus religiosa</i> /Delhi	2003	1	1	4	1	1
INN63_790S_1_A_SC_UP_India	<i>Syzygium cumini</i> /Bulandshahar	2006	1	1	2	1	1
INN64_806S_1__AI_UP_India <sup>1</sup>	<i>Azadirachta indica</i> /Bulandshahar	2006	1	1	5	1	1
INN65_809S_1__AI_UP_India <sup>1</sup>	<i>Azadirachta indica</i> /Bulandshahar	2006	2	1	2	1	1
INN66_713S_3__AI_UP_India	<i>Azadirachta indica</i> /Hathras	2006	1	1	1	1	1
INN67_717S_AI_UP_India	<i>Azadirachta indica</i> /Hathras	2006	1	1	2	1	1
INN68_425Sd-BB_SC_PB_India	<i>Syzygium cumini</i> /Amritsar	2004	1	1	1	1	1



“Master’s thesis – Sanjay S. Hiremath” “McMaster – Biology”

INN69_425SS-C_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN70_425SS-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	2	1	2	1	1
INN71_428B-WC_SC_PB_India	Syzygium cumini/Amritsar	2004	2	1	2	1	1
INN72_428B-WC-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN73_428B-WC-BBT_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN75_428B-WS-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN76_430SS-C_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN77_430SS-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN78_432WS-C_SC_PB_India	Syzygium cumini/Amritsar	2004	1	3	2	1	1
INN79_433Sd-C_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN80_436WS-C_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN81_437SS-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	3	2	1	1
INN82_439WS-C_SC_PB_India	Syzygium cumini/Amritsar	2004	2	3	2	1	1
INN83_440SS-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	2	3	2	1	1
INN84_441Sd-BB_1__SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN85_441Sd-BB_2__SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN86_441S-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN87_441S-BB_SC_PB_India	Syzygium cumini/Amritsar	2004	1	1	2	1	1
INN88_449DDD_SC_PB_India	Syzygium cumini/Amritsar	2005	1	1	2	1	1
INN89_455DDD'_SC_PB_India	Syzygium cumini/Amritsar	2005	1	3	2	1	1
INN94_782d_1__MI_HR_India	Mangifera indica/Amrouli	2006	1	1	2	1	1
INN95_782d_2__MI_HR_India	Mangifera indica/Amrouli	2006	1	5	2	1	1
INN96_782d_3__MI_HR_India	Mangifera indica/Amrouli	2006	3	1	3	1	1
INN97_782S_1__MI_HR_India	Mangifera indica/Amrouli	2006	1	5	2	1	1
INN98_782S_2__MI_HR_India	Mangifera indica/Amrouli	2006	1	1	2	1	1

<sup>1</sup>, indicate strains that have been tested for mating ability. <sup>2</sup>, Alleles for markers (from left to right) *CAP1*,

*FTR1*, *LAC*, *ITS*, *MtLrRNA*

**Table 2.** Genetic differences among populations of *C. neoformans* var. *grubii* in India.

(a) Nei’s genetic distance D between geographical populations; (b) Nei’s genetic distance

D between populations from different host tree species.

(a)

Geographical Region	Delhi (n=47)	Bulandshahar (n=3)	Hathras (n=2)	Amritsar (n=21)	Amrouli (n=5)
Delhi 28°36'0"E 77°19'0"N	0				
Bulandshahar 28°24'0"E 77°51'0"N	0.053	0			
Hathras 27°36'0"E 78°3'0"N	0.076	0.058	0		
Amritsar 31°37'59"E 74°51'56"N	0.007	0.047	0.070	0	
Amrouli 28° 25' 59"E 77° 19' 0"N	0.038	0.084	0.095	0.048	0

(b)

Host tree species	<i>Syzygium cumini</i> (n=26)	<i>Polyalthia longifolia</i> (n=14)	<i>Mimusops elengi</i> (n=4)	<i>Azadirachta indica</i> (n=7)	<i>Manilkara hexandra</i> (n=3)	<i>Cassia fistula</i> (n=14)	<i>Mangifera indica</i> (n=5)
<i>Syzygium cumini</i>	0						
<i>Polyalthia longifolia</i>	0.022	0					
<i>Mimusops elengi</i>	0.050	0.054	0				
<i>Azadirachta indica</i>	0.024	0.068	0.079	0			
<i>Manilkara hexandra</i>	0.024	0.082	0.065	0.031	0		
<i>Cassia fistula</i>	0.007	0.013	0.059	0.035	0.044	0	
<i>Mangifera indica</i>	0.046	0.053	0.118	0.064	0.067	0.028	0

**Table 3.** Allelic combinations showing evidence of recombination in the environmental population of *C. neoformans* var. *grubii* from India. (a) All four possible allelic combinations are found between alleles 1 and 3 of locus *FTR1* and alleles 1 and 2 of locus *LAC*. (b) All six possible allelic combinations are found between alleles 1, 3, and 5 of locus *FTR1* and alleles 1 and 2 of locus *CAP1*. The number of strains for each genotype is shown in Table.

(a)

	<u>LAC (Allele 1)</u>	<u>LAC (Allele 2)</u>
FTR1 (Allele 1)	<b>4</b>	<b>42</b>
FTR1 (Allele 3)	<b>3</b>	<b>14</b>

(b)

	<u>CAP1 (Allele 1)</u>	<u>CAP1 (Allele 2)</u>
FTR1 (Allele 1)	<b>43</b>	<b>4</b>
FTR1 (Allele 3)	<b>11</b>	<b>5</b>
FTR1 (Allele 5)	<b>4</b>	<b>1</b>

**Table 4.** Multilocus linkage disequilibrium analyses for samples of *Cryptococcus neoformans* var. *grubii* from India (sub-populations with sample sizes greater than 5 are also shown)

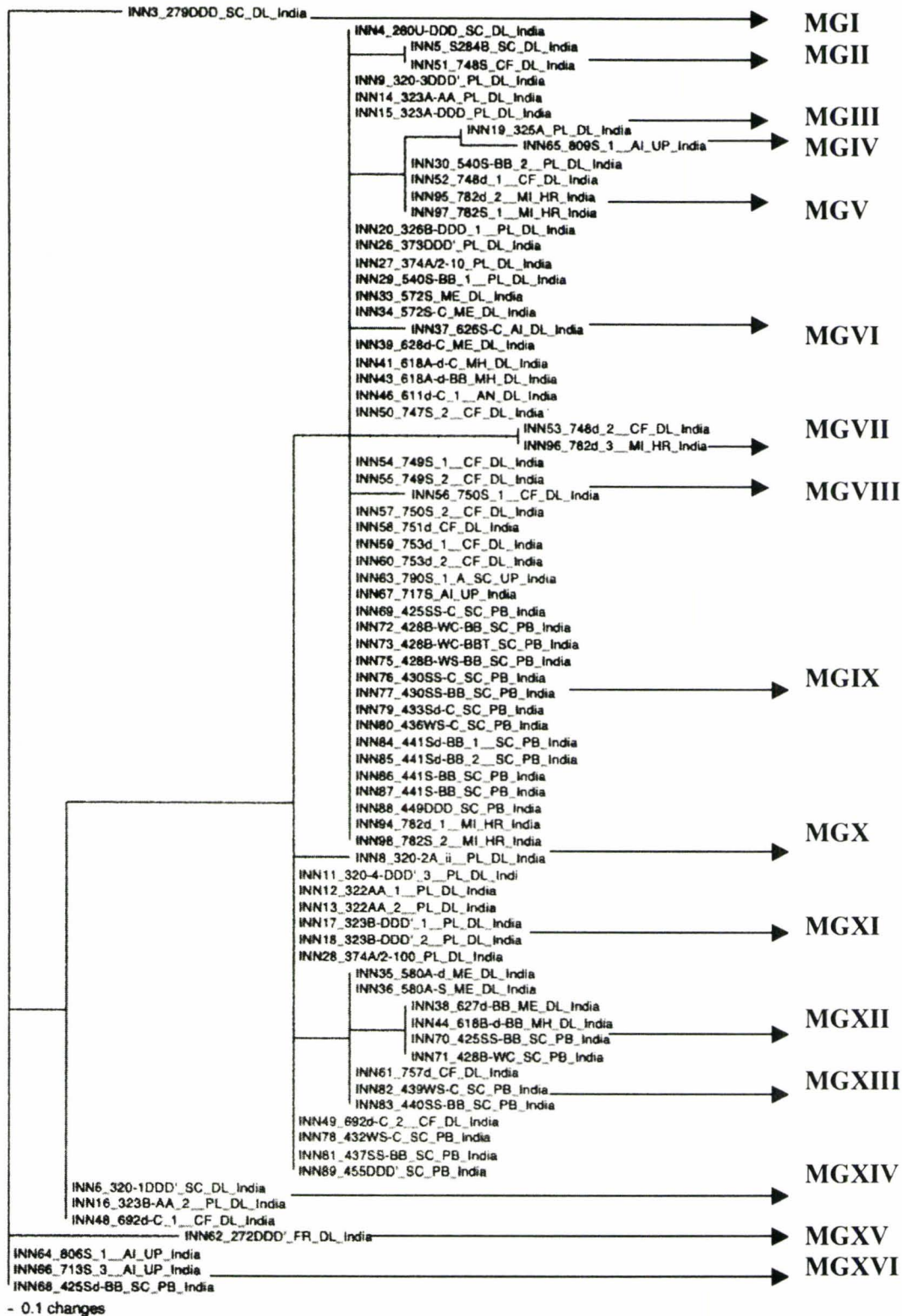
<b>Population</b>	<b><math>I_A^*</math></b> <b>(<i>p</i> value)</b>	<b>rBarD</b> <b>(<i>p</i> value)</b>	<b>Phylogenetic</b> <b>incompatibility</b> <b>(<i>p</i> value)</b>
Total (n=78)	0.1433 (0.138)	0.0360 (0.138)	0.5 (<0.0001)
Delhi (n=47)	0.1519 (0.102)	0.0385 (0.102)	0.7 (<0.0001)
Amritsar (n=21)	0.0454 (0.363)	0.0114 (0.363)	0.8 (0.108)

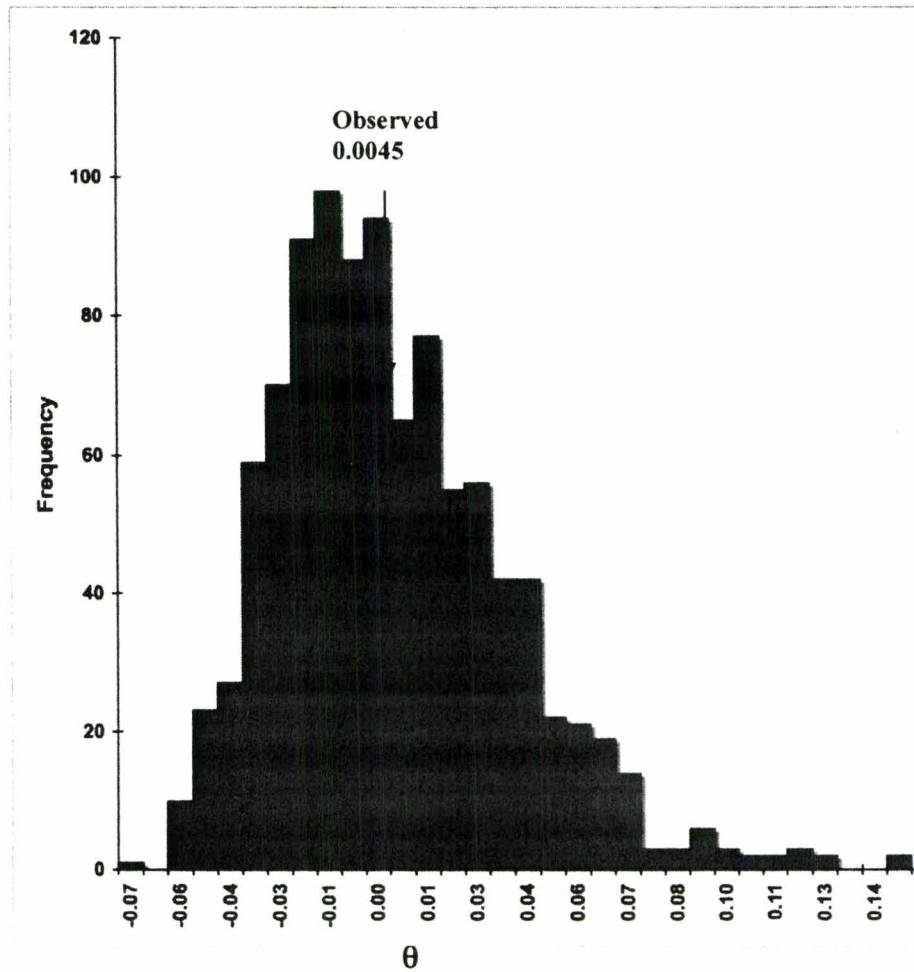
\*  $I_A$ , index of association

## Figure Legends

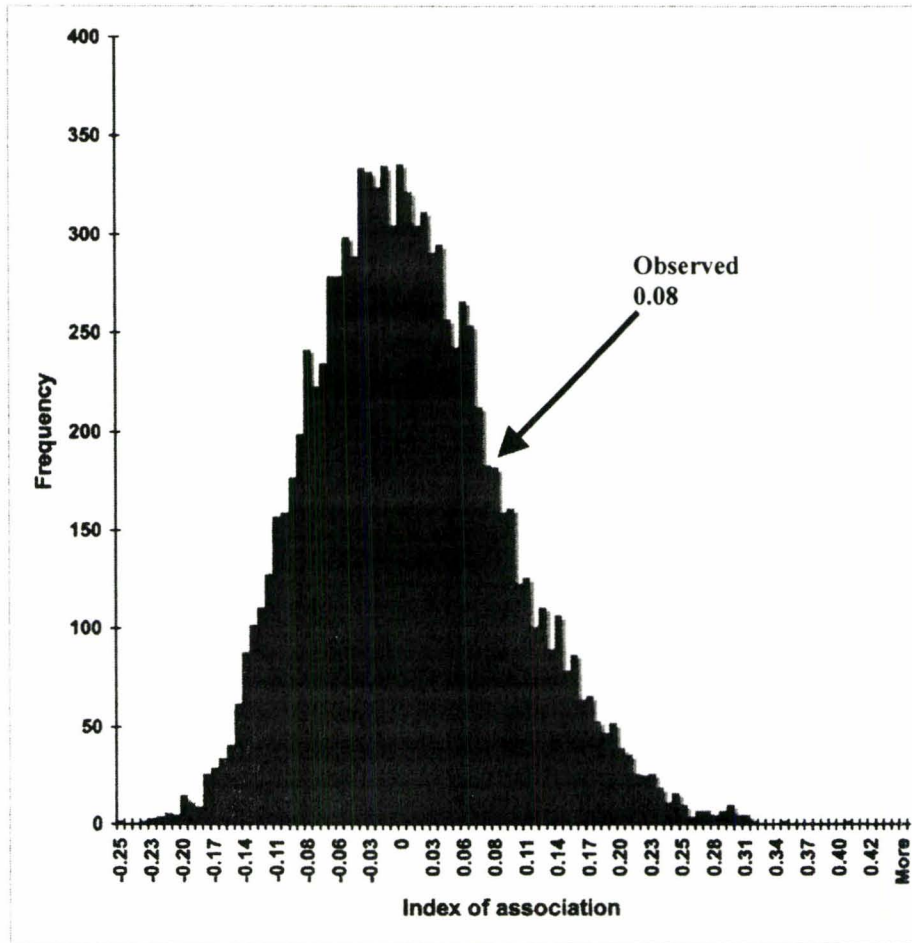
**Figure 1.** Maximum parsimony trees showing the relationships among 78 isolates of *Cryptococcus neoformans* var. *grubii* for the combined dataset using mid-point rooting (constructed using five markers: *CAP1*, *FTR1*, *LAC*, *ITS* & *mtLrRNA*). MPTs, most parsimonious trees; CI, consistency index; RI, retention index; HI, homoplasy index. Combined dataset (2912bp aligned; 84 MPTs; Tree length=20; CI=0.850, RI=0.953, HI=0.150). The combined dataset showed a phylogenetic pattern that did not distinguish populations based on either host trees or geographical locations. Genotype labels MGI to MGXVI represent the different multilocus haploid genotypes found in this study based on the five gene fragments. The midpoint-rooted tree is identical in topology to those based on outgroup-rooted tree using sequences from the serotype D strain JEC21. The midpoint-rooted tree is shown to increase the proportional lengths of branches for better visualization.

**Figure 3.** The observed Index of association ( $I_A$ ) value was within the distribution of  $I_A$  obtained from 10000 artificially recombined datasets ( $p=0.1816$ ).





**Figure 2.** The distribution of  $\theta$  (a measure of genetic differentiation) among geographic populations of *C. neoformans* var. *grubii* from India. The  $\theta$  value was well within the distribution of the permuted datasets that assumed no genetic differentiation ( $p=0.366$ ).



**Figure 3.** The observed Index of association (IA) value was within the distribution of IA obtained from 10000 artificially recombined datasets ( $p=0.1816$ ).



Figure S1

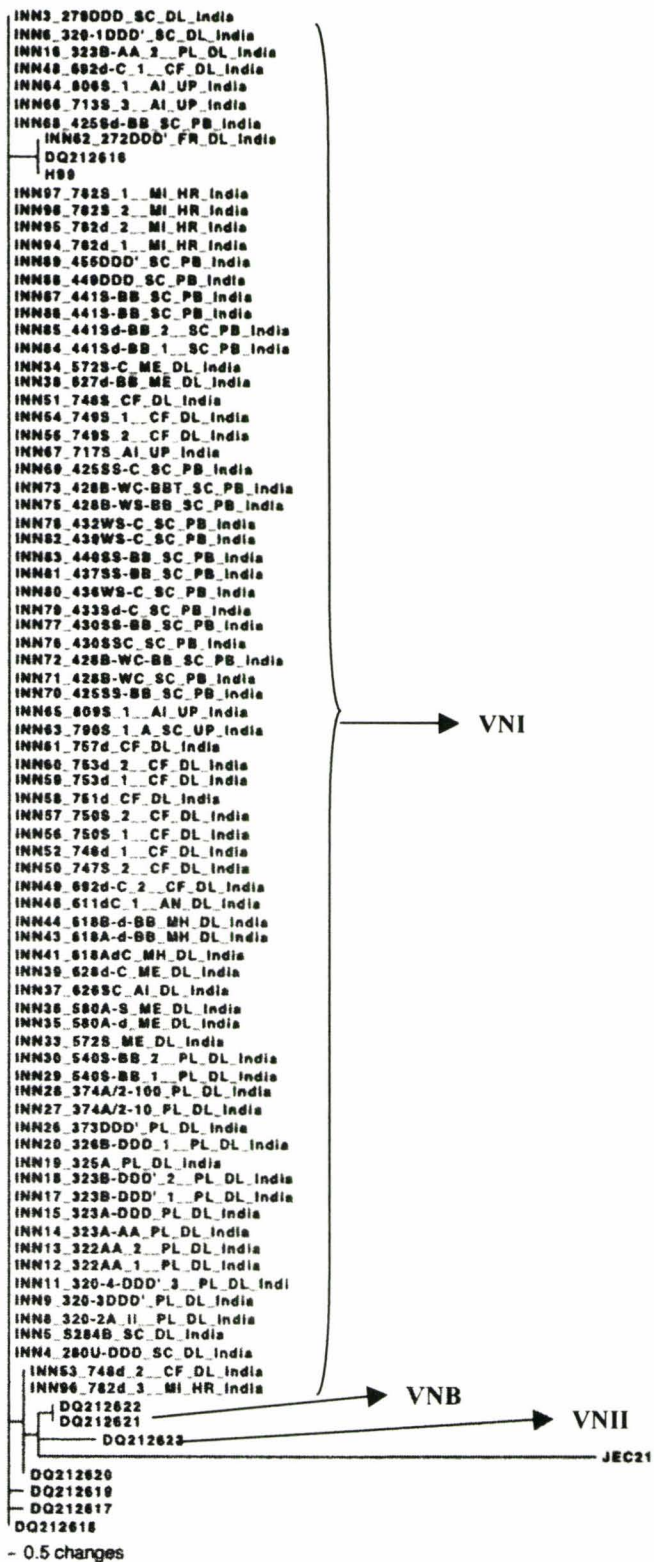
Maximum parsimony trees showing the relationships among 78 isolates of *Cryptococcus neoformans* var. *grubii* for three of the five markers sequenced (1a, 1b, and 1c). MPTs, most parsimonious trees; CI, consistency index; RI, retention index; HI, homoplasy index. (a) *LAC* (486bp aligned; 1 MPT; Tree length=45; CI=1, RI=1, HI=0)

(b) *CAP1* (731 bp aligned; Tree length=54; 1 MPT; CI=1, RI=1, HI=0)

(c) *FTR1* (788bp aligned; 1MPT; Tree length=100; CI=1, RI=1, HI=0).

Except the *LAC* tree, all other trees shown here are midpoint-rooted. The midpoint-rooted trees are identical in topology to those based on outgroup-rooted tree using sequences from the serotype D strain JEC21. The midpoint-rooted trees are shown to increase the proportional lengths of branches for better visualization. Numbers on branches indicate the number of nucleotide substitutions between the genotypes. The eight reference sequences (DQ212616 – DQ212623) for the three molecular types VNI, VNII and VNB are included in Figure 1a. Specifically, they are DQ212616 (strain A135-8, North Carolina, pigeon excreta, MAT $\alpha$ ; VNI); DQ212617 (strain c26, North Carolina, blood/HIV+, MAT $\alpha$ ; VNI); DQ212618 (strain A2 102-5, Texas, pigeon excreta, MAT $\alpha$ , VNI); DQ212619 (strain mal 104, Malawi, blood/AIDS, MAT $\alpha$ ; VNI); DQ212620 (strain A5 35-17, North Carolina, pigeon excreta, MAT $\alpha$ ; VNI); DQ212621 (strain bt1, Botswana, CSF/AIDS, MAT $\alpha$ , VNB); DQ212622 (strain bt60, Botswana, CSF/AIDS, MAT $\alpha$ , VNB); and DQ212623 (strain JH8-1, North Carolina, source unidentified, MAT $\alpha$ , VNII). Reference strains H99 and JEC21 are also included in Figure 1a.

1a)



1b)

INN3_279DDD_SC_DL_India	
INN4_280U-DDD_SC_DL_India	
INN6_320-1DDD'_SC_DL_India	
INN8_320-2A_II_PL_DL_India	
INN9_320-3DDD'_PL_DL_India	
INN11_320-4-DDD'_3_PL_DL_India	
INN12_322AA_1_PL_DL_India	
INN13_322AA_2_PL_DL_India	
INN14_323A-AA_PL_DL_India	
INN15_323A-DDD_PL_DL_India	
INN16_323B-AA_2_PL_DL_India	
INN17_323B-DDD'_1_PL_DL_India	
INN18_323B-DDD'_2_PL_DL_India	
	INN19_325A_PL_DL_India
	INN35_580A-d_ME_DL_India
	INN36_580A-B_ME_DL_India
	INN38_627d-BB_ME_DL_India
	INN44_618B-d-BB_MH_DL_India
	INN61_757d_CF_DL_India
	INN65_809S_1_AI_UP_India
	INN70_425SS-BB_SC_PB_India
	INN71_428B-WC_SC_PB_India
	INN82_439WS-C_SC_PB_India
	INN83_440SS-BB_SC_PB_India
INN20_326B-DDD_1_PL_DL_India	
INN26_373DDD'_PL_DL_India	
INN27_374A/2-10_PL_DL_India	
INN28_374A/2-100_PL_DL_India	
INN29_548S-BB_1_PL_DL_India	
INN30_548S-BB_2_PL_DL_India	
INN33_672S_ME_DL_India	
INN34_572S-C_ME_DL_India	
INN37_626S-C_AI_DL_India	
INN39_628d-C_ME_DL_India	
INN41_618A-d-C_MH_DL_India	
INN43_618A-d-BB_MH_DL_India	
INN45_611d-C_1_AN_DL_India	
INN48_682d-C_1_CF_DL_India	
INN49_682d-C_2_CF_DL_India	
INN50_747S_2_CF_DL_India	
INN51_748S_CF_DL_India	
INN52_748d_1_CF_DL_India	
	INN53_748d_2_CF_DL_India
	INN96_782d_3_MI_HR_India
INN54_749S_1_CF_DL_India	
INN55_749S_2_CF_DL_India	
INN56_760S_1_CF_DL_India	
INN57_760S_2_CF_DL_India	
INN58_761d_CF_DL_India	
INN59_763d_1_CF_DL_India	
INN60_763d_2_CF_DL_India	
INN62_272DDD'_FR_DL_India	
INN63_790S_1_A_SC_UP_India	
INN64_806S_1_AI_UP_India	
INN66_713S_3_AI_UP_India	
INN67_717S_AI_UP_India	
INN68_425Sd-BB_SC_PB_India	
INN69_425SS-C_SC_PB_India	
INN72_428B-WC-BB_SC_PB_India	
INN73_428B-WC-BBT_SC_PB_India	
INN75_428B-W6-BB_SC_PB_India	
INN76_430SS-C_SC_PB_India	
INN77_430SS-BB_SC_PB_India	
INN78_432WS-C_SC_PB_India	
INN79_433Sd-C_SC_PB_India	
INN80_436WS-C_SC_PB_India	
INN81_437SS-BB_SC_PB_India	
INN84_441Sd-BB_1_SC_PB_India	
INN85_441Sd-BB_2_SC_PB_India	
INN86_441S-BB_SC_PB_India	
INN87_441S-BB_SC_PB_India	
INN88_449DDD'_SC_PB_India	
INN89_455DDD'_SC_PB_India	
INN94_782d_1_MI_HR_India	
INN95_782d_2_MI_HR_India	
INN97_782S_1_MI_HR_India	
INN98_782S_2_MI_HR_India	

— 0.05 changes

1c)

INN3_279DDD_SC_DL_India INN4_290U-DDD_SC_DL_India INN5_5284B_SC_DL_India	INN6_320-1DDD'_SC_DL_India INN11_320-4-DDD'_3_PL_DL_India INN12_322AA_1_PL_DL_India INN13_322AA_2_PL_DL_India INN16_323B-AA_2_PL_DL_India INN17_323B-DDD'_1_PL_DL_India INN18_323B-DDD'_2_PL_DL_India INN28_374A/2-100_PL_DL_India INN35_580A-d_ME_DL_India INN36_580A-5_ME_DL_India INN48_692d-C_1_CF_DL_India INN49_692d-C_2_CF_DL_India INN61_767d_CF_DL_India INN78_432WS-C_SC_PB_India INN81_437SS-BB_SC_PB_India INN82_439WS-C_SC_PB_India INN83_440SS-BB_SC_PB_India INN89_465DDD'_SC_PB_India	INN9_320-2A-II_PL_DL_India
INN9_320-3DDD'_PL_DL_India INN14_323A-AA_PL_DL_India INN15_323A-DDD_PL_DL_India	INN19_325A_PL_DL_India INN30_540S-BB_2_PL_DL_India INN52_748d_1_CF_DL_India	
INN20_326B-DDD_1_PL_DL_India INN26_373DDD'_PL_DL_India INN27_374A/2-10_PL_DL_India INN29_540S-BB_1_PL_DL_India INN33_572S_ME_DL_India INN34_572S-C_ME_DL_India	INN95_782d_2_MI_HR_India INN97_782S_1_MI_HR_India	INN65_809S_1_AI_UP_India
INN38_627d-BB_ME_DL_India INN39_628d-C_ME_DL_India INN41_618A-d-C_MH_DL_India INN43_618A-d-BB_MH_DL_India INN44_618B-d-BB_MH_DL_India INN46_611d-C_1_AN_DL_India INN50_747S_2_CF_DL_India INN51_748S_CF_DL_India INN53_748d_2_CF_DL_India INN54_749S_1_CF_DL_India INN55_749S_2_CF_DL_India INN56_750S_1_CF_DL_India INN57_750S_2_CF_DL_India INN58_751d_CF_DL_India INN59_753d_1_CF_DL_India INN60_753d_2_CF_DL_India INN62_272DDD'_FR_DL_India INN63_790S_1_A_SC_UP_India INN64_806S_1_AI_UP_India INN66_713S_3_AI_UP_India INN67_717S_AI_UP_India INN68_425Sd-BB_SC_PB_India INN69_425SS-C_SC_PB_India INN70_425SS-BB_SC_PB_India INN71_428B-WC_SC_PB_India INN72_428B-WC-BB_SC_PB_India INN73_428B-WC-BBT_SC_PB_India INN75_428B-WS-BB_SC_PB_India INN76_430SS-C_SC_PB_India INN77_430SS-BB_SC_PB_India INN79_433Sd-C_SC_PB_India INN80_436WS-C_SC_PB_India INN84_441Sd-BB_1_SC_PB_India INN85_441Sd-BB_2_SC_PB_India INN86_441S-BB_SC_PB_India INN87_441S-BB_SC_PB_India INN88_448DDD'_SC_PB_India INN94_782d_1_MI_HR_India INN96_782d_3_MI_HR_India INN98_782S_2_MI_HR_India	INN37_628S-C_AI_DL_India	

— 0.05 changes