GENOTYPIC AND PHENOTYPIC ANALYSES OF TWO MODEL STRAINS OF *CRYPTOCOCCUS NEOFORMANS*

GENOTYPIC AND PHENOTYPIC ANALYSES OF TWO MODEL STRAINS OF *CRYPTOCOCCUS NEOFORMANS*

By WENJING HUA

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Masters of Science

McMaster University

Masters of Science (2017)

Hamilton, Ontario (Department of Biology)

TITLE: Genotypic and phenotypic analyses of two model strains of *Cryptococcus neoformans*

AUTHOR: Wenjing Hua (McMaster University)

SUPERVISOR: Dr. Jianping Xu

NUMBER OF PAGES: xi, 67

Lay Abstract

Cryptococcosis is a globally distributed infection that is prevalent among immune-compromised individuals, such as HIV/AIDS patients. This disease can be attributed to a group of opportunistic fungal pathogens – *Cryptococcus neoformans* species complex. During the past century, significant resources have been put in an effort to understand its ecology, evolution, life cycle, pathogenesis and virulence factors, and molecular and cellular processes. Most of the laboratory-based studies have relied on two model strains assumed to differ only at the mating type locus. My thesis investigated this assumption and found there are several additional significant genetic differences between these two strains and that such differences contribute to the observed phenotypic differences between them. My results highlight the complexity of genotype-phenotype relationships and the continued evolution of strains even in lab environments in *C. neoformans.*

Abstract

The human pathogenic *Cryptococcus neoformans* species complex are agents of a common AIDS-defining disease, which causes about 181,000 deaths each year. There are several specific features distinguishing this species from other fungi, including the presence of a polysaccharide capsule and melanin pigment production, both of which contribute to its virulence. A large number of studies about this pathogen used two model strains JEC20 and JEC21. In these studies, these two strains are assumed to be "isogenic", differ only at the mating type region. Consequently, their phenotypic differences, including virulence, have been attributed to this region. Here, we applied second-generation sequencing and bioinformatics tools to identify sequence polymorphisms between the two genomes. Beside the Mating Type locus, two other regions were found to contain high frequencies of SNPs. To further understand the effects of these loci on the phenotypic differences, four phenotyping assays (mating ability, melanin pigment production, capsule formation, and high temperature growth ability) were conducted on the recombinant progeny obtained from the cross between JEC20 and JEC21. In addition, genomic sequences of these progeny were obtained to identify the complete distributions of other SNPs among the strains. Finally, we identified several novel SNPs contributing to virulence-related traits in this species, which suggest that caution should be placed in attributing phenotypic differences to specific genomic regions in "isogenic" strains derived from classical breeding experiments.

Acknowledgements

I would like to thank my supervisor, JP Xu, for the patient guidance, encouragement and support he has provided throughout the past two years as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and provided great advice and patience about my questions and problems. I would also like to thank all the members of Xu Lab: Himeshi, Eta, Adrian, Sally, Renad, Aly, etc and staff in our departments, who gave me a lot of help and good memories during my time in McMaster. I want to also thank Aaron Vogan, a past member of our lab, for leading me into this project and offering helps when I got questions. Also, thanks to Prof. Singh and Prof. Quinn for being my committee members and providing guidance.

Finally, I would like to thank my parents, my supervisor's wife - Heather, and my friends in Canada and China, who have given me warm-hearted comforts and companies, helping me overcome difficulties through my studies.

Table of Contents

List of Figures

List of Tables

Acronyms

YEPD	Yeast Extract Peptone Dextrose
CNS	Central nervous system
GXM	Glucuronoxylomanan
GalXM	Galactoxylomannan
TNF	Tumour necrosis factor
MAT	Mating Type Locus

Chapter 1

General Introduction

1.1 Background

1.1.1 History and distribution of *Cryptococcus neoformans*

Cryptococcus neoformans is a worldwide human fungal pathogen and causes an estimated 220,000 cases of cryptococcal mengitis annually among people with HIV, resulting in nearly 181,000 deaths globally (Rajasingham, 2017). The majority (almost 90%) of these cases are reported in sub-Saharan Africa, which makes Cryptococcal meningitis one of the most fatal diseases among AIDS patients there, even more than tuberculosis.

Cryptococcus neoformans was first identified by two German pathologists, Otto Busse and Abraham Buschke from a specimen of a patient with chronically inflammed tibia in 1894 (described in Knoke & Schwesinger, 1994). Within the ensuring decades, more clinical and environmental isolates were reported and the basic knowledge about this organism was established, including nomination of species (Benham, 1950), antigenic heterogeneity (Evans, 1950), laboratory diagnosis techniques (by testing capsule presence, melanin production and urease activity), and heterothallic sexuality (Kwon-

Chung, 1975). With the onset of the AIDS pandemic in the early 1980s, *Cryptococcus*-associated disease – cryptococcosis became more and more common in HIV-infected patients. Emergent concerns about this health threat were raised since then. As a result, our understanding about this human pathogen has increased significantly.

Genetic analyses of the human pathogenic *Cryptococcus* complex have revealed significant evolutionary diversity and heterogeneity. As a result, the taxonomy of this complex has undergone significant changes and is still hotly debated. The different taxonomic approaches were mainly due to their differences in how species should be defined and what implications they have on clinical practices. In one school of thought (Kwon-Chung, et al, 2017), the agents of cryptococcosis include two sibling species complexes, the *Cryptococcus neoformans* species complex and the *Cryptococcus gattii* species complex. The *C. neoformans* species complex includes three serotypes: serotype A (also called var. *grubii*), serotype D (also called var. *neoformans*), and serotype AD (the hybrids of serotypes A and D). The *C. gattii* species complex contains two serotypes B and C (Kwon-Chung, 2014). In a different approach, serotypes A and D, the two divergent lineages within the *C. neoformans* species complex are called different species, *C. neoformans* and *C. deneoformans*, respectively. Similarly, the five divergent lineages within the *C. gattii* species complex are called five different species not corresponding to serotype designations but according to their sequence divergence based on seven loci: *C. gattii* (VGI), *C. deuterogattii* (VGII), *C. bacillisporus* (VGIII), *C. tetragattii* (VGIV) and *C. decagattii* (VGIV and VGIIIc) (Hagen, et al, 2015).

The environmental sources of the *C. neoformans* species complex are globally distributed in various ecological niches, including avian excreta (the most common source), soil, water, and plants etc. In contrast, the *C. gattii* species complex was previously thought to be distributed only in tropical and subtropical regions, although the recent outbreaks on Vancouver Island and in the North West Pacific Coast of America expanded its geographical range to temperate regions (Bartlett et al, 2012; Bartlett et al, 2008). Similar to the *C. neoformans* species complex, the *C. gattii* species complex samples are also found in diverse ecological niches but with a predominant association with trees, such as *Eucalyptus* in Australia (Springer, 2010). With the help of molecular makers, numerous *C. neoformans* and *C. gattii* samples can be genotyped into more specific subgroups (molecular types VNI, VNII, VNB belong to serotype A of *C. neoformans*; molecular type VNIV is serotype D of *C. neoformans*; molecular types VNIII is AD hybrids of *C. neoformans*; and molecular types VGI-VGIV belong to *C. gattii*. Cogliati, 2013) for further epidemiological studies of cryptococcosis.

1.1.2 Characteristics of *Cryptococcus neoformans*

From the prespective of morphology, there are several cell forms for the human pathogenic *Cryptococcus*, including yeast-like cells, basidiospores, titan cells, chlamydospores, filamentous structure, etc. Under laboratory conditions, yeast form is the most common for the *C. neoformans* species complex: spherical, encapsulated, and around 5-7 μm in diameter, whereas basidiospores are only produced during sexual reproduction, around 1.8 to 3 μm in diameter. Both types of cells are haploid most of the time. A commonly observed morphology of *Cryptococcus* cells *in vivo* is titan cell

(Zarogoza & Nielsen, 2013). Titan cells are polyploid, with highly cross-linked capsule and thickened cell walls, reaching a diameter of more than 15 μm (including capsule). The specific capsule and cell wall complexed structure mainly consist of β-glucan, chitin, and glucuronoxylomanan (GXM), which, together with polyploidy, enhance its ability for survival and dissemination in mammalian and invertebrate hosts (Zarogoza & Nielsen, 2013). In contrast with titan cells, there is another type of smaller cells in vivo with diameters of 2-4 μm with thickened cell walls, found in macrophages. These small cells are called drop cells and they likely play a role in latency and survival within the host (Alanio, 2015). Besides single-celled forms, cryptococci can also develop into filamentous morphologies, including hyphae (sexual reproduction) and pseudohyphae (intermediate shape between yeast and true hyphae) under certain natural environmental or laboratory conditions.

Cryptococcus neoformans can reproduce through two approaches, either asexual life cycle or sexual life cycle through the bipolar mating system with two mating types, MATa and MATα. During the sexual cycle, filamentous structures will form to develop clamp connections and for production of basidiospores. After germination, the basidiospores grow into recombinant offspring, containing genetic materials of both parental strains (Ellis, 1990). Beside bisexual mating and asexual reproduction, haploid fruiting (and same-sex mating), a process involving filamentous structure as well but only one mating type, is also found in *C. neoformans* involving strains of MATα (Wickes, 1996).

For most cells of *C. neoformans*, there are also two structural characteristics distinguishing them from other human pathogenic fungi. Polysaccharide capsule is a structure outside the cell wall of *C. neoformans* that can be easily observed with negative dye staining. The capsule also contains enzymes and proteins that cross into cell wall structure (Nimrichter, et al, 2005), some of which are antigenic components (for identification of serotypes A and D). However, the detailed process involved in the synthesis of this distinctive structure has not been revealed yet. Another feature of *C. neoformans* is its ability to synthesize a dark pigment (melanin) in cell walls. The synthesis of melanin involves oxidation of exogenous dihydroxyphenols (Williamson, Wakamatsu & Ito, 1998). In addition to those structural phenotypes, several common features of human fungal pathogens are also present in *C. neoformans*, such as growth ability under mammalian host temperature, and the secretion of urease and phospholipase. All of these features have been found essential for the virulence of *C. neoformans* in mammalian hosts.

1.1.3 Pathogenesis and virulence factors of *Cryptococcus neoformans*

Fungal cells causing cryptococcosis are usually acquired by inhalation into host bodies. Depending on the immune status of hosts, the cryptococcal cells can cause pneumonia and invasive systemic infections in immunocompromised patients. For immunocompetent individuals, they will either be cleared out by the host immunity response or stay latent with the host showing no obvious symptoms. Among HIV-infected patients, if infected with cryptococcal cells and not treated in time, the infection

often develops into cryptococcocal meningitis, a central nervous system (CNS) infection, and lead to death (May, 2016).

As a common fungal pathogen, a great number of discoveries have been made about the infection process, including how Cryptococci react against the immune system to dampen inflammation, its survival from host phagocytosis, and its further dissemination into the CNS. Firstly, the presence of the polysaccharide capsule components (GXM and GalXM) can depress the secretion of pro-inflammatory cytokines, such as tumour necrosis factor (TNF, Piccioni, 2013), interleukin-12&13 (IL-12, IL-13, Angkasekwinai, et al, 2014), and inhibit the inflammatory responses of the host immune system. Secondly, *Cryptococcus* can resist against phagocytosis, which involves multiple mechanisms, such as avoiding killing by or escaping from immune cells. Several factors help cryptococcal cells survive in this process, including the formation of capsule, melanisation and the secretion of urease. These factors collectively contribute to its survival in harsh in vivo and in vitro environments (Coelho & Bocca, 2014). Moreover, *Cryptococcus* cells can penetrate the blood-brain barrier and enter into CNS, causing intensive brain tissue damaging and ultimately fatal meningitis.

For a better understanding of pathogenesis of *Cryptococcus neoformans*, many genes have been knocked out and identified as closely related to one or more virulence-related features mentioned above. Table 1 summarizes some of the key genes involved in virulence-associated phenotypes of *C. neoformans* (seen Table 1.1).

Table 1.1 Virulence-associated phenotypes and their regulation genes

Phenotypes	Regulation gene factors
Mating and filamentation	GPA1 (Alspaugh, 1997); STE12 (Wickes, 1997); GPB1 (Wang, 2000); RAS1 (Alspaugh, 2000); SXI1α and SXI2α (Hull, 2002, 2005); cAMP signaling cascades (D'Souza, 2001)…
Capsule synthesis	CAP gene family: CAP59, CAP64, and CAP60 (Chang, 1996, 1998); CAS gene family: CAS1 and CAS2 (Moyrand, 2002); Nrg1 (Cramer, 2006)…
Melanination	CnLAC1 (Williamson, 1994); WdPKS1 (Feng, 2001); MET3 (Yang, 2002); SIT1 (Tangen, 2007)…
Growth at high temperature	RAS1 and RAS2 (Alspaugh, 2000; Waugh, 2002); SOD2 (Giles, 2005)…

1.2 Objectives of my thesis

Extensive genetic and genomic analyses of the human pathogenic *Cryptococcus* strains have identified significant divergence between the *C. neoformans* species complex and the *C. gatti* species complex (Janbon et al, 2014). These include both chromosomal structural differences as well as nucleotide sequence divergence between homologous genes. For example, homologous genes share about 89% nucleotide sequence identity between model genomes representing *C. neoformans* and *C. gattii*. Between lineages within the *C. neoformans* species complex, fewer chromosomal rearrangements were observed between the model genomes and the nucleotide sequence identity between orthologs of serotypes A and D strains are higher, at 90-95% (Janbon et al, 2014). A recent study from my lab suggested that the phenotypic differences between serotype A and D strains are controlled by multiple genomic regions and some candidate phenotype-related QTL genes were identified based on genome comparison (Vogan et al, 2016). However, because of the reduced recombination between strains of these two serotypes and the large number of open reading frames with unknown functions, we were only able to map the QTLs to regions of hundreds of kilobases and it's extremely difficult to finely map the specific phenotypic traits to individual genes. Using genetically similar strains but with notable phenotypic differences would significantly enhance our ability to identify the relationship between genetic polymorphisms and phenotypic differences. The model "isogenic" strains JEC20 and JEC21 have been reported to show phenotypic differences and such differences have been assumed to be controlled by the mating type locus (Kwon-Chung, 1992b). However, the validity of that assumption has not been

tested. The objectives of my thesis were to examine the genomic differences between JEC20 and JEC21 and if genetic differences were identified, to determine their relationships to the phenotypic differences between these two strains.

These two model strains, JEC20 and JEC21, were constructed from parental strains NIH433 and NIH12 through crossing and repeated backcrossing (Kwon-Chung, 1992a). JEC20 is MATa whereas JEC21 is MATα. Ten years later in 2002, this mating type locus was more precisely defined and sequenced (Figure 1.1, Lengelar, 2002). For the past 2.5 decades, those two strains have been widely used to study various properties and processes concerning *C. neoformans* and played a significant role in our understanding of the genetics and pathogenesis of this species. For example, they were used in studies of mating and filamention (Wickes, 1997; Lengeler, 2001; Lu et al, 2005; Fu, et al, 2011), heterosexual meiotic reproduction (Lin, et al, 2005; Ni, et al, 2013), mitochondrial genetics (Yan & Xu, 2003; Toffaletti, et al, 2004), virulence-related phenotypes (Alspaugh, et al, 2000; Vallim, 2005; Botts, et al, 2009), gene and signal pathway identification (Janbon et al, 2001; Idnurm, et al, 2004; Ekena, et al, 2008), etc. Throughout these studies, the mating type locus has been assumed as the only locus being different between the genetic materials of the two strains and any phenotypic difference between them were attributed to the mating type locus (Kwon-Chung, 1992b). Here, I test this assumption and explore the potential contributions of other genetic loci to the phenotypic variation among the progeny from these two strains.

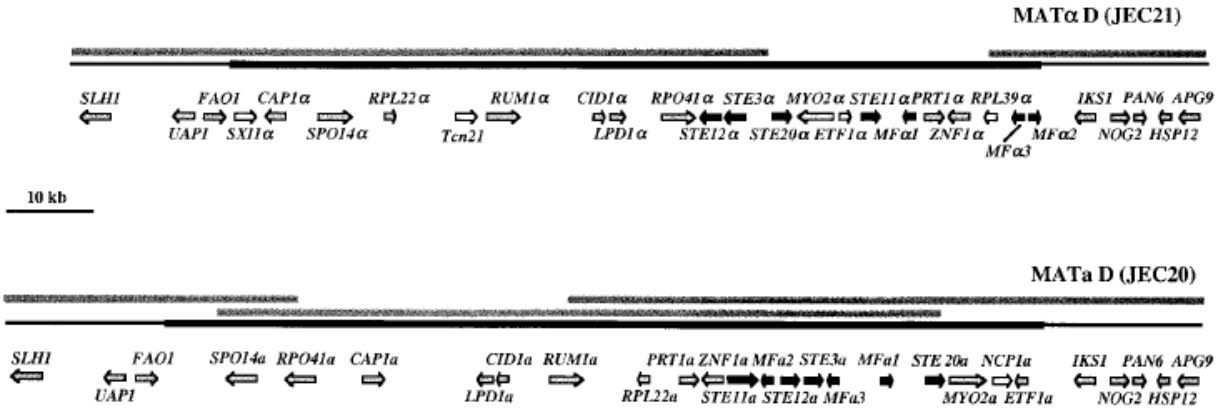


Figure 1.1 Structures of mating-type alleles and adjacent genomic regions of JEC20 and JEC21. The mating-type-specific regions are shown as thick bold lines, and flanking regions are shown as thinner black lines. Sequences were analyzed using BLASTX, and identified genes are shown as arrows in the direction of transcription. Genes encoding pheromone response pathway elements are shown as black arrows, locus-specific genes are shown as white arrows, and all other genes are shown as grey arrows. (Lengeler et al, 2002)

Chapter 2

Genotypic and phenotypic analyses of JEC20 and JEC21

2.1 Preface

The human pathogenic *Cryptococcus neoformans* species complex has a global distribution, causing about 625,000 deaths each year. It's a model species for fungal pathogenesis studies. Several specific features distinguish this species from other fungi, including the presence of a polysaccharide capsule and melanin pigment production, both of which contribute to its virulence. These findings were originally made based on studying two strains JEC20 and JEC21. These two strains have been used as model strains since the early 1990s and are considered isogenic, differ only at the mating type region. Previously, they are thought to be only different at the mating type region, however, their genomes have not been compared yet. In this study, we compared their genomic differences and analyzed their potential impacts on virulence-associated phenotypes. Second-generation sequencing and bioinformatics tools are applied to identify sequence polymorphisms between the two genomes. In total, we revealed three SNP-rich regions, including the previously noted mating-type region, and several SNPs

outside of the three regions in other parts of the genome. To study the effects of these loci on the phenotypic differences between the JEC20 and JEC21, a set of 24 sexual progeny were obtained from the cross between JEC20 and JEC21 and 4 phenotype features were tested among them: mating ability, melanin pigment production, capsule formation, and high temperature growth ability. In addition, we sequenced the 24 progeny using Illumina to identify the complete distributions of other SNPs among the strains. Our research identified additional genetic and genomic differences between the two model "isogenic" strains and revealed their potential contributions to phenotypic differences.

2.2 Introduction

Cryptococcus neoformans is an opportunistic human fungal pathogen that can cause cryptococcal meningitis among immune-compromised individuals, mainly AIDS patients (Del Valle, 2006). A recent estimate suggested that there are about 220,000 new cases of Cryptococcosis each year, resulting in about 181,000 deaths annually (Rajasingham, 2017). *Cryptococcus neoformans* is a basidiomycetous yeast, with a bisexual mating system controlled by a large locus with two alternative mating types termed mating type a (MATa) or mating type α (MATα). Under appropriate conditions such as nitrogen limitation, low humidity and ambient temperature, strains of opposite mating types can form conjugation tubes and fuse with each other to generate a dikaryotic mycelium. At the tip of the dikaryotic mycelium, the two nuclei fuse and undergo meiosis, generating four recombinant haploid nuclei. Chains of basidiospores are generated with each containing one of these nuclei. The spores are released and when encountering appropriate environmental conditions, they germinate and form yeast cells. Because of its

well-defined sexual cycle and its ability to grow on a diversity of media and incubation conditions, *C. neoformans* has become a model organism for fungal genetics and pathogenicity studies. These studies included identifying the mating types and mating type distributions, the relationship between mating type and other phenotypic traits, and signal transduction pathways related to virulence in a diversity of hosts.

Among the strains that have been used as models for our understanding of *C. neoformans* and fungal pathogens in general, two stood out. They are called JEC20 (MATa, also named B-4476) and JEC21 (MATα, also named B-4500), especially the latter as it is often used as a representative strain of serotype D. Of these two strains, JEC20 was a meiotic progeny from a cross between two wild-type isolates (NIH12 and NIH433, MATa and MATα respectively) by Kwon-Chung (1975). JEC21 was the congenic mating partner of JEC20 obtained by selecting MATα progeny in ten rounds of backcross to JEC20 (Kwon-Chung et al, 1992a). Therefore, the genomes of these are assumed to differ only at the mating type locus. Subsequent comparisons demonstrated that JEC21 was more virulence than JEC20 in a mouse model, suggesting that the mating type locus contributes to pathogenicity (Kwon-Chung et al, 1992b). Ten years later, the detailed sequences of the mating-type locus and the adjacent genomic regions of JEC20 and JEC21 were obtained, revealing significant differences in both gene content and gene order between these two strains (Lengeler et al, 2002). For example, genes SX11α, Ten21, and RPL39α were specific to JEC21 while RPL22a was found only in JEC20.

The whole genome sequence of strain JEC21 was published by Loftus et al (2005). This genome contained about 20 Mb distributed among 14 chromosomes. Of the 6574

predicted genes, 80% were associated with a confirmed transcript (Idnurm, 2005). However, while the sequence of the MATa locus in strain JEC20 has been published (Lengeler et al, 2002), the whole genome sequence of JEC20 is not available and aside from the mating type locus, has been assumed to be identical to that of JEC21.

The objective of this study is to obtain the whole genome sequence of strain JEC20 and compare it with that of JEC21. Furthermore, if genomic differences in addition to the mating type locus were found, we were interested in their contributions to the phenotypic differences between these two strains by analyzing the genotypes and phenotypes of recombinant offspring from the cross.

2.3 Materials and methods

2.3.1 Strains

The two model strains JEC20 (Serotype D, MATa) and JEC21 (Serotype D, MATα) were originally obtained from the frozen stocks of Dr. Tracey Moore in 1997 and have been stored in an -80 ℃ freezer since then. Meiotic progeny were generated by mating JEC20 and JEC21 on V8-jiuce agar using the protocol described in Forsythe et al (2016). Basidiospores were isolated from individual basidia using a micromanipulator (MSM System 300, Singer Instruments). Individual spores were transferred to YEPD medium (Forsythe et al. 2016). After incubation at 30 ℃ for 3 days, cells were collected and stored at -80 ℃ for further genotyping and phenotype assays.

2.3.2 Genome sequencing of strain JEC20 and comparison with strain JEC21

The genomic DNA was extracted from JEC20 using a protocol described in Xu et al. (2000). The whole-genome sequencing of JEC20 was performed at Farncombe Institute at McMaster University and 4.15 Gb raw data of the JEC20 genome were generated using the Illumina MiSeq v2 PE250 sequencing platform. Trimming of reads was done with cutadapt (v1.7.1. https://pypi.python.org/pypi/cutadapt/; Martin, 2001) and NGSQCToolkit (v2.3.3. http://www.nipgr.res.in/ngsqctoolkit.html; Patel & Jain, 2012). Assembly of JEC20 genome was done by Velvet (v1.2.10. http://www.ebi.ac.uk/~zerbino/velvet/; Zerbino et al, 2008). The assembled contigs were aligned to the reference genome JEC21 with bwa (v0.7.8. http://bio-bwa.sourceforge.net; Li and Durbin, 2009) and single nucleotide polymorphisms (SNPs) were identified with samtools (v1.3.1. http://samtools.sourceforge.net; Li, 2011). The genome sequence of JEC21 was downloaded from GenBank (http://www.ncbi.nlm.nih.gov/genome/?term=JEC21).

2.3.3 Genotyping progeny

Comparisons between the genome sequence of JEC20 and the published JEC21 sequence identified three regions of genomic differences between the two strains, including the mating type region (See Results below). Based on these differences, three corresponding molecular markers were developed to analyze meiotic progeny to screen for recombinant genotypes involving these regions. Specifically, the mating type region

was assayed using the MATa and MATα - specific primers at the RUM1 locus, following protocols described in Yan et al. (2012). The two remaining polymorphic regions were distinguished based on restriction digest polymorphisms following gene-specific amplification using polymerase chain reaction. The primer sequences for these markers and the restriction enzymes used to differentiate the parental alleles are presented in Table 2.1. The sizes of the restriction-digested fragments for the two parental strains for each of the gene fragments are also shown in Table 2.1. The protocols for PCR, restriction digests, and gel electrophoresis for these two markers followed that in Vogan et al. (2016).

Table 2.1 Molecular markers distinguishing three SNP-rich regions between JEC20 and JEC21

Region	Chromosome	Location (kb)	Forward primer (5->3)	Reverse primer (5->3)	Fragment size (bp)	Restriction enzyme	Digested fragment sizes respectively
1	2	19.6 ~ 54.6 (35kb)	ACCCCCTCCCACCTGCGA	ACAGGCGGGGACGAGCACT	1023	HaeIII	JEC20: 150, 670, 30, **80, 100** JEC21: 150, 670, 30, **180**
			AATGGCGATCGGAGCTTGC	GCGAGAGGTCAGAGGGTAGA	983	HaeIII	JEC20: 67, 158, **300**, 388, 70 JEC21: 67, 158, **219, 79**, 388, 70
2	2	1390.5 ~ 1432.7 (42.2kb)	CCGGAGGCTCGTTCTTGAGG	AATCTCGGGGGTGCTGAAGC	1232	HaeIII	JEC20: 17,**154, 20,** 52, **990** JEC21: 17, **176**, 51, **571, 417**
			CCAATCCGCTCAGAAGGCCATT	GGAGTTCCCGTAGAGCCG	791	AluI	JEC20: 370, **110, 73**, 70, 132, 11, 25 JEC21: 370, **189**, 70, 132, 11, 25
3	4	1527.7 ~ 1668.9 (141.2kb)	TGAAGATTTTGGATTCGAAGAAGGTGACG	CAAGTGCAGAGCTGATCGGCATGGG	1340	EcoRV	JEC20: **880, 501** JEC21: **370,1020**

2.3.4 Progeny phenotype

To examine the potential contributions of the three genomic regions to phenotypic differences among the progeny, parental strains JEC20 and JEC21 as well as representative recombinant progeny of each genotype were phenotyped. Several phenotypic tests were conducted, including mating test by backcrossing to parental strains, melanin production, capsule production, and high-temperature growth. For these tests, the two parental strains and 24 selected progeny (three representative progeny for each of the eight genotypes) were first grown on YEPD agar to obtain actively growing cells. The cells were suspended and dissolved into ddH_20 and diluted to $1x10^6$ cells/mL estimated using a hemocytometer. Mating ability tests were performed by mixing cells from each progeny separately with the two parental strain cells on V8 mating agar medium. After 2-4 weeks incubation at 23 ℃, mating ability was determined based on whether filamentous morphological change was observed (Kwon-Chung, 1976). Melanin production for each strain was determined by culturing cells on Caffeic acid-ferric citrate medium (Hopfer & Blank, 1975) and incubating at 30 ℃ for 7 days. Variations in the amount of melanin production are determined based on light reflections using the Spot Densitometry function of FluoroChem 8900 (Vogan et al. 2016). Capsule formation was induced by growth at 30 ℃ for 24hr in 10% Sabourand dextrose (SD) broth medium and then 2 days in 100-fold dilution SD broth with 50 mM MOPS. After the induction, cells were stained with nigrosine and observed under microscope to count the proportion of cells with a capsule following the protocol described by Vogan et al. (2016). For each strain, at least 160 cells were screened for capsule production. The high temperature

growth ability of each strain was determined by incubating each sample at a starting concentration of 10^5 ($\pm$10%) cells/mL in RPMI medium at both 30 ℃ (optimal temperature) and 40 ℃ (high temperature) for 5 days. After that, their cell concentrations were measured with a spectrophotometer and a hemocytometer. At least three repeats were performed for each strain for each of the phenotypes.

2.3.5 Genome sequencing of selected progeny

Differences in the above four phenotypes were detected among the progeny (see Results below). To further help explain the genetic bases for their phenotypic differences, we obtained whole-genome sequences for each of the 24 selected meiotic progeny at Farncombe Institute at McMaster University. A total of 8.5 Gb raw data were obtained for the 24 progeny samples using the Illumina MiSeq v3 PE300 sequencing platform, resulting in about 355 Mb per progeny sample. For each sample, trimming of reads and SNP calling followed those described in section 2.2. The output summary files of SNPs were visualized by IGV (v2.3. http://www.broadinstitute.org/software/igv) and Tablet (v1.16.09.16. https://ics.hutton.ac.uk/tablet; Milne et al, 2012).

2.3.6 Data analyses

Aside from genome sequence comparisons to derive genetic and genomic differences between strains, additional analyses were performed to examine the statistical significance of phenotypic differences among progeny and the potential correlation between genotype and phenotypes using ANOVA and the generalized Linear Model

respectively. The statistical analysis was done with R-studio (based on R version 3.3.3. https://www.rstudio.com/products/RStudio/).

2.4 Results and Discussion

2.4.1 Genomic differences between JEC20 and JEC21

On the foundation of genetic relationship, JEC20 and JEC21 should be very similar at genomic level except for their mating type region noted previously. However, by our results, there were two more regions with a length around 40kb each and 1506 other single nucleotide polymorphisms identified between these two "isogenic" strains.

2.4.1.1 Genome of JEC20

After whole genome sequencing, 4.15 Gb raw data of JEC20 genome were used for assembly by Velvet. Assessment with a range of Kmers were done and the Kmer=33 was used as final running parameter. The final assembly (Table 2.2) is estimated to be 1.89 fold of genome length of JEC21 (19 Mbp) and largest contig reachs 25.5kb. The GC content is 48.44%, similar to JEC21 (48.54 %). Statistics were also done with strain JEC21 as reference. Also a rough alignment comparison between the contigs of the JEC20 genome and the published JEC21 genome revealed that 75.5% of total length of JEC20's assembly (only including fully aligned contigs) constituted the same part shared by 90.67% of JEC21 genome. However, the partially aligned and non-aligned contigs might contain more parts that weren't counted in and more precise comparison can be demonstrated with single nucleotide polymorphisms distribution.

Table 2.2 Summary information of genome assembly of strain JEC20.

Statistics without reference (JEC20)	Contigs	6,384
	Contigs (>=200bp)	2,403
	Largest contig	255,194
	Total length	36,046,696
	GC (%)	48.44
	N50*	30,140
	L50*	376
Statistics with reference (JEC21)	Genome fraction (%)	90.673
	Total aligned length	27,214,892
	Largest alignment	100,730
	NG50*	45,306
	LG50*	146
	Fully unaligned contigs	84
	Fully unaligned length	229,452
	Partially unaligned contigs	599
	Partially unaligned length	8,555,924

* N50 is contig length that when using contigs of length not less than it, these contigs account for half of the bases of the assembly. NG50 is the same formula of the reference genome. L50 is the minimum number of contigs that produce half of the bases of the assembly. LG50 is the same formula of bases of reference genome.

2.4.1.2 SNPs information between JEC20 and JEC21

Due to the unexpected large differences between the de novo assembly of the JEC20 genome and the published JEC21 genome, we aligned the raw reads of the JEC20 genome to the reference genome to identify single nucleotide polymorphisms (SNPs). A total of 5322 SNPs were found among the whole genome and with the majority (N=3816, 71.7%) of them located in three regions (Figure 2.1). Two of the SNP-rich regions (regions #1 and #2, Table 2.3) were located on chromosome 2 and one was located on chromosome 4 (region #3, Table 2.3). According to genome location and spanning information, region 3 on chromosome 4 is exactly the mating type region of *C. neoformans* (Lengeler et al., 2002) and some flanking regions (mating type region was estimated to be 105kb). Our sequence comparisons between the mating type region we obtained and that obtained earlier by Lengeler et al. (2002) revealed no new SNPs.

In addition to the three SNP-rich regions, we also identified a total of 1506 SNPs located in other genomic regions. Among these 1506 SNPs, 415 (27.6%) are located in transcribed regions (Figure 2.2), including 271 in exons, 117 in introns, and 27 in UTR regions. These 415 SNPs are distributed in 83 genes with the majority located at the ends of each of the 14 chromosomes. The summary information of these transcribed SNPs is presented in Table 2.4. Twenty of the 83 genes contained multiple SNPs, with eight of the genes annotated to putative known functions and the remaining 12 not annotated to a known functional category (Table 2.5). Most of these genes contained miss-sense mutations, causing amino acid sequence differences between JEC20 and JEC21.

While the three SNP-rich regions most likely reflect the differences between the two parental strains of JEC20 and JEC21, NIH12 and NIH433, the exact sources of the remaining 1506 SNPs are not known. However, there are two possibilities. The first is that these SNPs were present between NIH12 and NIH433, and they were similarly inherited as the other three SNP-rich regions. The second possibility is that there were recently accumulated mutations in laboratory conditions since the initial generation of JEC20 and JEC21. The JEC21 genome was sequenced in 2005, over 10 years after its construction, during which the number of transfers and mitotic divisions that the sequenced strain went through in the laboratory are unknown. We would like to note that these two possibilities are not mutually exclusive and both could have contributed to the observed differences. Regardless of their sources, these SNPs suggest potential candidate genes from which to investigate the genetic bases of phenotypic differences between JEC20 and JEC21.

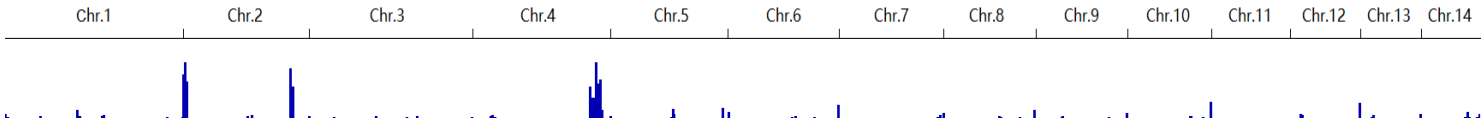


Figure 2.1 Genome sequence alignment between JEC20 and JEC21 to identify the distribution of single nucleotide polymorphisms between these two strains. Height of blue bars represent frequency of SNPs in the location.

Table 2.3 Summary information of the three SNP-rich regions between JEC20 and JEC21 (contrasted with other regions on same chromosome)

Chromosome	2					4		
Regions	**SNP-rich region**		Other region			**SNP-rich region**	Other region	
Region name	**Region1**	**Region 2**	i	ii	iii	**Region 3**	iv	v
Region range (kb)	**19.6 ~ 54.6**	**1390.5 ~ 1432.7**	0~ 19.6	54.6~ 1390.5	1432.7~ 1643.8	**1527.7 ~ 1668.9**	0~ 1527.7	1668.9~ 1803.6
Region length (kb)	**35**	**42.2**	19.6	1335.9	211.1	**141.2**	1527.7	134.7
SNP number	**302**	**252**	8	93	3	**3260**	49	5
SNP frequency (per kb)	**8.63**	**5.97**	0.408	0.070	0.014	**23.08**	0.032	0.037
				0.066			0.032	
ORF region	**CNB00080 ~ CNB00175**	**CNB04890 ~ CNB05020**				**CND05660 ~ CND06050**		
ORF number	**11**	**14**				**40**		

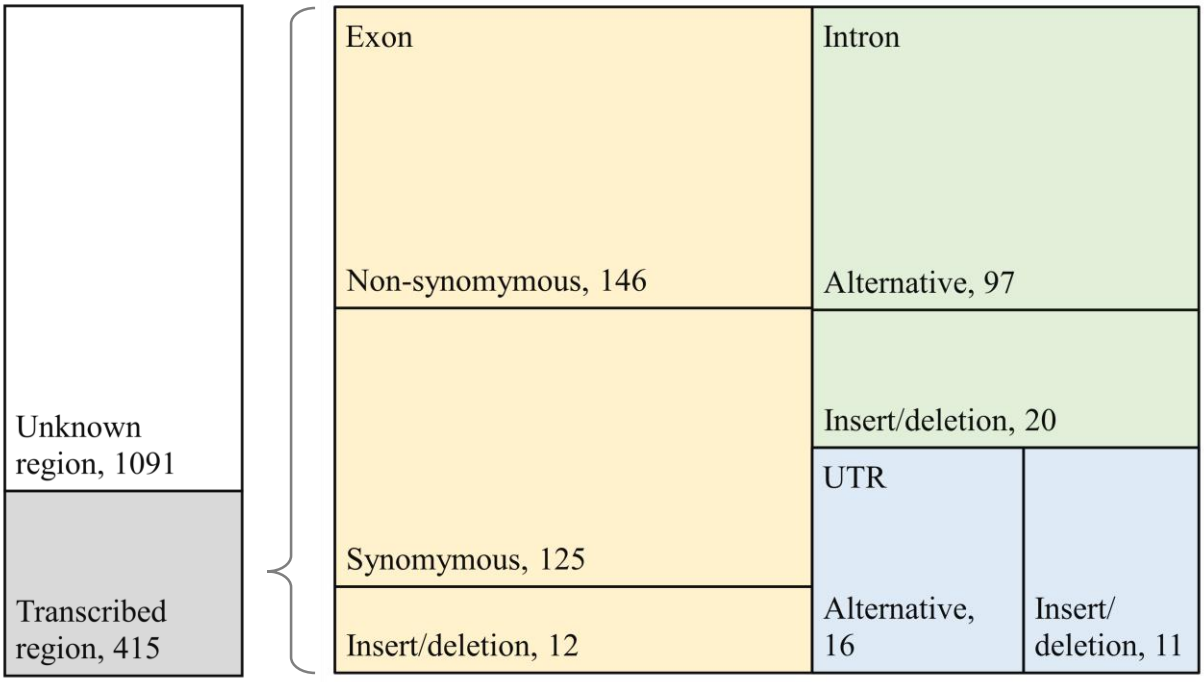


Figure 2.2 Treemap of SNPs types outside the three SNP-rich regions between the JEC20 and JEC21 genomes

Table 2.4 Summary of the types of SNPs in the transcribed regions outside the three SNP-rich regions.

		Intron	UTR regions			Exon	
			3' UTR	5' UTR			
SNPs number		117	14	13		271	
Mutation	Alternative	97	7	9	259	Synonymous	125
						Non-synonymous	146
	Insertion *	9	2	3		5	
	Deletion *	11	5	1		7	
Genes		28	10	8		37	

* Insertions and deletions were derived based on comparisons with JEC21.

Table 2.5 Gene regions outside SNP-rich regions with accumulation of multiple SNPs and predicted mutations

Gene region	Chromosome	Location (bp)	notation	SNPs	Exon	Intron	UTR	Predicted mutations
CNA00180	1	54488 ~ 57199	calcium ion transporter	4	1	3		mutate a 6-AA sequence into another 20-AA sequence
CNA04790	1	1272537 ~ 1272973	hypothetical protein	6	5	1		Alternation of 5 single AAs
CNE03210	5	912245 ~ 913393	hypothetical protein	6	5	2		Alternation of 5 single AAs
CNE03750	5	1071424 ~ 1073239	hypothetical protein	2	2			Alternation of 2 single AAs
CNE05190	5	1457652 ~ 1460962	hypothetical protein	17	17			Alternation of 17 single AAs
CNF00650	6	216065 ~ 220767	pyruvate carboxylase	2	2			Alternation of 1 AA and a STOP
CNG00010	7	197 ~ 1723	hypothetical protein	26	3	23		Alternation of 3 single AAs
CNG02140	7	597916 ~ 605745	mitotic spindle assembly -related protein	2	2			Insertion cause dislocation of AA
CNI00010	9	754 ~ 2253	fungal specific transcription factor	26	11	15		Alternation of 11 single AAs
CNJ00010	10	172 ~ 4804	hypothetical protein	3	3			Alternation of 3 single AAs

CNJ03440	10	1081812 ~ 1085004	hypothetical protein	4	3		1	Alternation of 3 single AAs and a site in UTR
CNK00010	11	905 ~ 4960	hypothetical protein	9	9			Alternation of 9 single AAs
CNK00650	11	214880 ~ 216565	hypothetical protein	30	30			Alternation of 30 single AAs
CNK03480	11	1018826 ~ 1019636	hypothetical protein	7	2		5	Alternation of 2 single AAs and 5 sites in UTR
CNL04160	12	137655 ~ 141899	retrotransposon nucleocapsid protein	2	2			Alternation of 2 single AAs
CNM00010	13	213 ~ 2368	transporter	54	8	46		Alternation of 8 single AAs
CNM00580	13	182012 ~ 183433	hypothetical protein	4	4			Alternation of 5 single AAs
CNM02480	13	746928 ~ 749661	hypothetical protein	2	2			Alternation of 1 AA and 1 insertion cause dislocation of downstream AA
CNN02320	14	714977 ~ 721654	1,3-beta-glucan synthase	4	3	1		Alternation of 3 single AAs
CNN02460	14	760649 ~ 762648	retrotransposable element slacs 132 kda protein	5	2		3	Alternation of 2 single AAs and 3 sites in UTR

AA refers to amino acid.

2.4.2 Phenotype assays

To analyze the potential contributions of the three SNP-rich genomic regions to phenotypic differences between JEC20 and 21, we crossed these two strains and obtained their meiotic progeny. A total of 46 progeny were isolated by micromanipulation and these offspring were genotyped using PCR-based molecular markers described in Table 2.1. All eight genotypes representing the different allelic combinations at these three loci were found among the 46 progeny. For each of the eight genotypes, we randomly selected 3 representative progeny for following phenotypic analyses. The genotypic classes for the 24 progeny at the three SNP-rich regions are shown seen in Table 2.6.

Table 2.6 Genotype identification of progeny and parental strains covering 3 SNP-rich regions

Progeny sample	Genotype	Region3 (mating type)	Region 1	Region 2
S1, S2, S3, P1	GENO#1	**a**	**B**	**C**
S4, S5, S6	GENO#2	**a**	**B**	γ
S7, S8, S9	GENO#3	**a**	β	**C**
S10, S11, S12	GENO#4	**a**	β	γ
S13, S14, S15	GENO#5	α	**B**	**C**
S16, S17, S18	GENO#6	α	**B**	γ
S19, S20, S21	GENO#7	α	β	**C**
S22, S23, S24, P2	GENO#8	α	β	γ

2.4.2.1 Mating success

Consistent with the molecular typing results at the mating type locus, the mating test results showed that these 24 progeny belonged to two distinct mating types (Table 2.7). Neither of the two other polymorphic regions was found to influence mating compatibility among these progeny strains, consistent with the bipolar mating system in *C. neoformans.* In addition, all the MATa progeny successfully mated with CDC15, a serotype A, MATα strain. Not surprisingly, the results are consistent with previous knowledge that sexual reproduction normally occurs between opposite mating types (**a** and **α**) with sexual compatibility between different serotypes (Schmeding et al, 1981).

However, none of the progeny showed evidence of haploid fruiting on V8-jiuce agar medium, regardless of their mating types. Similarly, neither parental strains JEC20 or JEC21 formed filaments on V8-jiuce agar medium separately by themselves after almost four weeks of incubation. In a haploid fruiting study in 1996 (Wickes et al, 1996), the optimal conditions for hyphal induction and several critical factors for the process were described, including the using of YNB-AS media (yeast nitrogen base without amino acids and without ammonium sulfate), inoculation method (by streaking), optimal temperature (25-30 ℃), the type and concentration of agar. Because we didn't test all the conditions as suggested, it's possible that the capacity of these progeny strains to undergo haploid fruiting still exist.

Table 2.7 Mating ability assay on V8-juice agar

Sample	Genotype	Region3 (mating type)	Mating with strains:			
			Self-crossing	JEC20 (**a**D)	JEC21 (αD)	CDC15 (αA)
S1, S2, S3	GENO#1	**a**	-	-	+	+
S4, S5, S6	GENO#2	**a**	-	-	+	+
S7, S8, S9	GENO#3	**a**	-	-	+	+
S10, S11, S12	GENO#4	**a**	-	-	+	+
JEC20	GENO#1	**a**	-	-	+	+
S13, S14, S15	GENO#5	α	-	+	-	-
S16, S17, S18	GENO#6	α	-	+	-	-
S19, S20, S21	GENO#7	α	-	+	-	-
S22, S23, S24	GENO#8	α	-	+	-	-
JEC21	GENO#8	α	-	+	-	-

2.4.2.2. Melanin production

The growth medium containing caffeic acid and ferric citrate was developed initially for identification of *C. neoformans* because it was efficient and specific for inducing melanin formation (Hopfer & Blank, 1975). Through incubation on this medium, the melanin pigment accumulates in cell walls and turns colonies grey to black colour (Wang et al, 1995). Compared to serotype A strains such as H99 and CDC15 (Vogan et al. 2016), both parental serotype D strains JEC20 and JEC21 as well as their 24 meiotic progeny all produce relatively low amounts of melanin. For example, on the caffeic acid medium and at 30 ℃, both H99 and CDC15 produce abundant melanin and turn the colonies black within 3 days. In contrast, for strains JEC20 and JEC21, there was little melanin production within 3 days on the same medium and under the same incubation condition. However, when incubation time was prolonged to 7 days, colonies of both parental strains and their progeny turned grey/black and differences between the two parental strains can be quantified using spot densitometry. Here, the 7-day incubation time point was chosen to quantify melanin production of each strain.

Our results showed that even though the two parental strains differed relatively little in melanin production, their difference was statistically significant (Figure 2.3). Between this pair, JEC20 produced more melanin that JEC21 after 7-days incubation. Interestingly, two progeny (S15 and S24) produced significantly more melanin than both parental strains (Figure 2.3). In contrast, four progeny (S7, S8, S9, S22) produced significantly less melanin than both parental strains. For the remaining 19 progeny, they didn’t show distinction with either one or both parental strains. Among the six strains that

showed significant differences from both parents, progeny S7, S8, and S9 have the same genotype (GENO#3) at the three SNP-rich regions, suggesting that the specific allelic combination (aβC) at the three SNP-rich regions might be associated with low melanin production. However, progeny S22, S23, and S24 shared a different genotype (GENO#8) as one of the parental strains - JEC21 but they had significantly different melanin productions, with S22 being the lowest melanin producer, S23 an intermediate producer, and S24 the highest producer among the progeny. This result suggests that SNPs in other gene regions are also likely involved in influencing melanin production in this species.

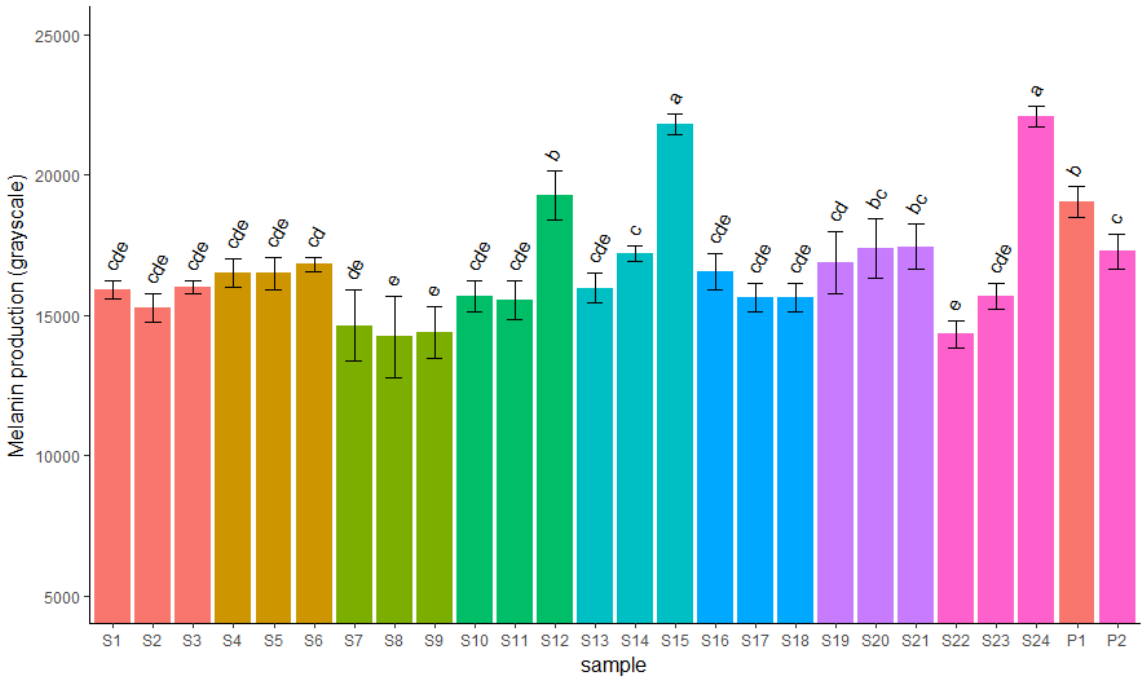


Figure 2.3. Melanin production among segregating progeny from the isogenic cross of *C. neoformans* at seven days after incubation on caffeic acid medium. The X-axis represents 24 progeny (S1-S24) of eight genotypes (colour differentiated) and two parental strains (P1: JEC20 and P2: JEC21). The Y-axis shows the relative amount of melanin production assayed using spot densitometry where a value of "0" means no melanin production with all light is reflected and a value of 65535 means completely black with no light reflection. Each strain was assayed at least five times, with each value represents the mean and the scale bar represents standard deviation. Letters represent statistically significant groups with progeny sharing the same letter having non-significantly different melanin production abilities. P values are corrected for multiple tests ($p<0.0001$).

2.4.2.3 Capsule formation

The polysaccharide capsule is a structure surrounding the cell wall of *C. neoformans* and has multiple effects against the immune system of infected hosts, such as interfering with phagocytosis and inhibiting the generation of pro-inflammatory cytokines (Buchana, 1998). Using light microscopy, the capsule can be observed using negative straining. However, for still unknown reasons, not all cells in each population produce capsule under the testing condition. After 2 days of starvation in diluted Sabourand broth medium, cells of each progeny were stained with nigrosine and the numbers of cells with and without the capsule were counted. A previous study identified multiple QTLs influencing the proportion of hybrid progeny cells with and without capsule under the specific growth and inducing condition (Vogan et al. 2016). Figure 2.4 shows the percentage of capsulated cells for each progeny and for the two parental strains.

Our results showed that the two parental strains had similar percentages of cells with capsules, at about 90%. However, there were significant differences among the 24 progeny. While 15 of the 24 progeny had similar percentages as the two parental strains, the remaining nine (S3, S10, S12, S14, S15, S16, S17, S20, and S24) had significantly less than the parental strains. These nine progeny belonged to six different genotypes based on alleles at the three SNP-rich regions (Figure 2.4). Interestingly, these nine progeny could be further grouped into several categories based on their percentages of cells with capsules (Figure 2.4). Very few cells of progeny S12 and S15 produced capsule while about 40% of the cells in progeny S10 and S17 produced capsule. Different from that for melanin production, none of the eight genotypes showed a consistently reduced

percentage of capsule production over that of the parental strains. Taken together, our results suggest that SNPs in other gene regions are likely involved in influencing capsule production in this species as well.

Another observation during our experiment was that cells with capsules usually clumped together, which could be resulted from the budding process that directs capsule synthesis into daughter cells (Pierini & Doering, 2001).

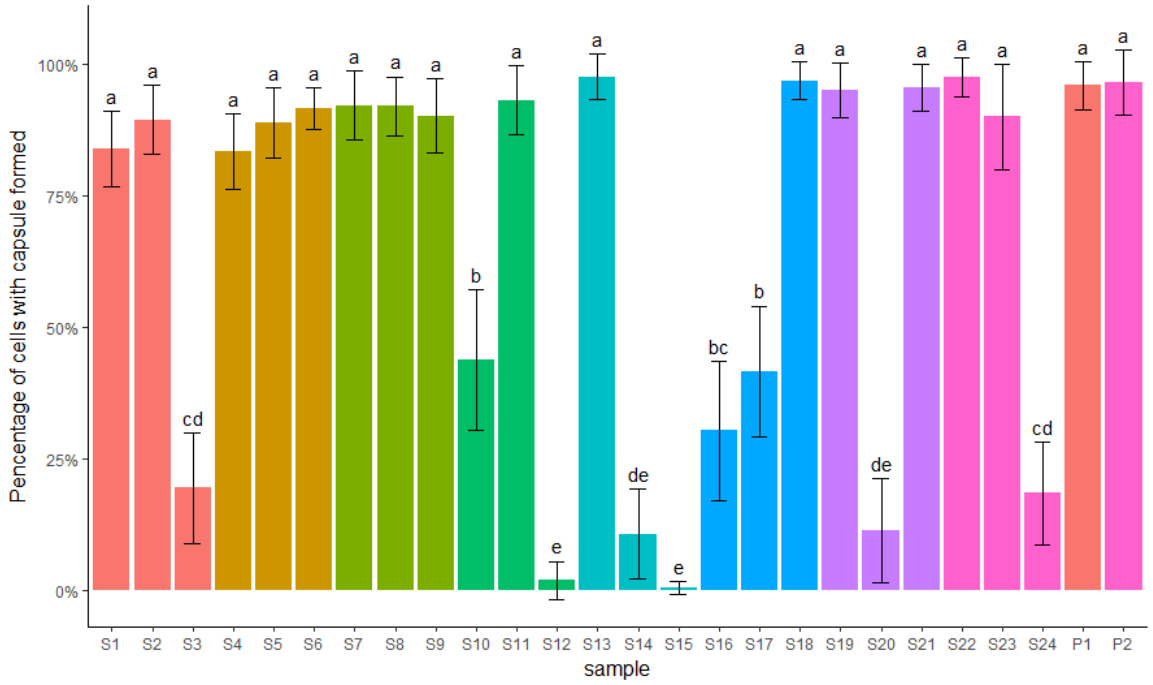


Figure 2.4 Percentage of cells producing capsules among segregating progeny from the "isogenic" cross of *C. neoformans*. The X-axis represents 24 progeny (S1-S24) of eight genotypes (colour differentiated) and two parental strains (P1: JEC20 and P2: JEC21). The Y-axis shows the percentage of cells producing the capsule. For each progeny, five repeats were done and each repeat counted a total of 160 cells under microscope. Letters represent statistically significant groups with progeny sharing the same letter having non-significantly different capsule production abilities. P values are corrected for multiple tests ($p<0.0001$).

2.4.2.4. Effects of temperature

The ability to survive and grow at a relatively high temperature (> 37 ℃) is a prerequisite for *C. neoformans* as a human pathogen. Here, we assessed the progeny growth at 40 ℃, a common temperature used for examining the high temperature growth ability in the pathogenic *Cryptococcus* species complex. For comparison, we also examined their growth abilities at 30 ℃ (optimal temperature for yeast cell growth). For each progeny, eight repeats were performed and each repeat was assayed for cell density at four time points, at 0hr, 24hr, 72hr, and 120hr after inoculation. The summary results are presented in Figure 2.5A. At 30 ℃, all progeny showed significant growth, reaching at least 100-fold of the initial inoculated cell number. At this temperature, the two parental strains showed relatively little difference in their growths. However, significant differences were found among the progeny and between some of the progeny and the parental cells (Figure 2.5B). For example, after five days of incubation, progeny S15 grew significantly better than all other progeny and both parents.

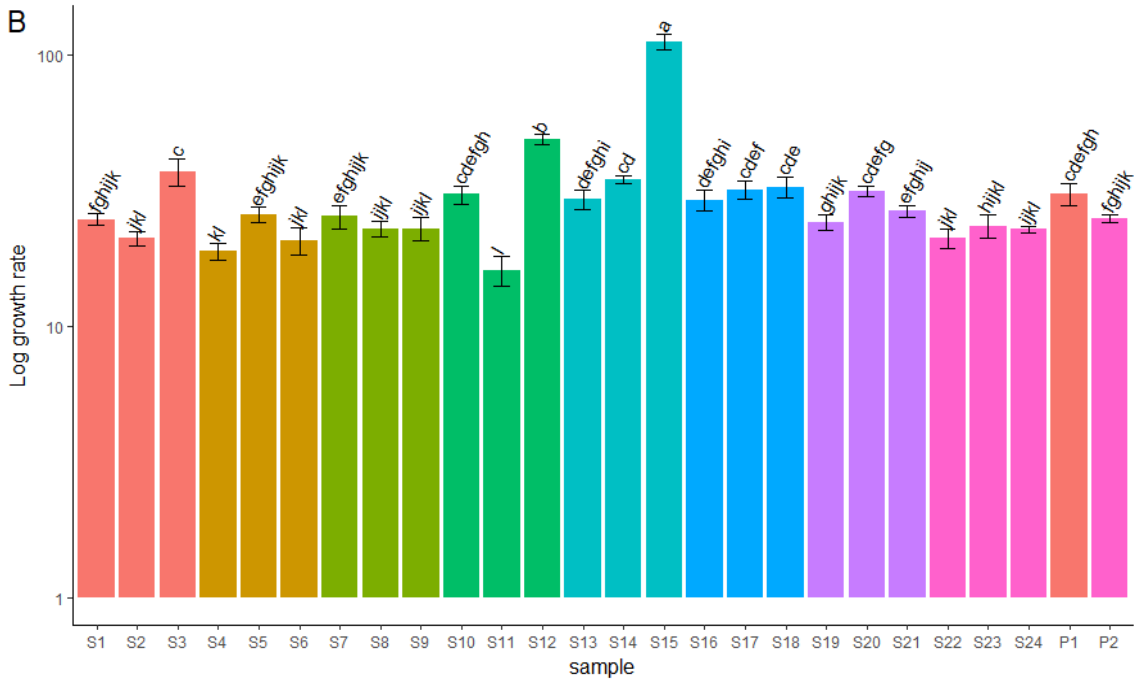
Different from the rigorous growths exhibited by all 24 progeny at the 30℃ environment, 16 of the 24 examined progeny showed almost no growth at 40 ℃ (Figure 2.5C). Of the remaining 8 progeny, seven showed similar growths as either of the two parental strains. Interestingly, one progeny S18 showed significantly better growth than both parental strains and all other 23 tested progeny. Among the progeny, three genotypes - GENO#1 (progeny S1, S2, S3), GENO#3 (progeny S7, S8, S9), and GENO#4 (progeny S10, S11, S12) showed consistently lower growth than both parental strains while the remaining five genotypes showed variable patterns among the progeny

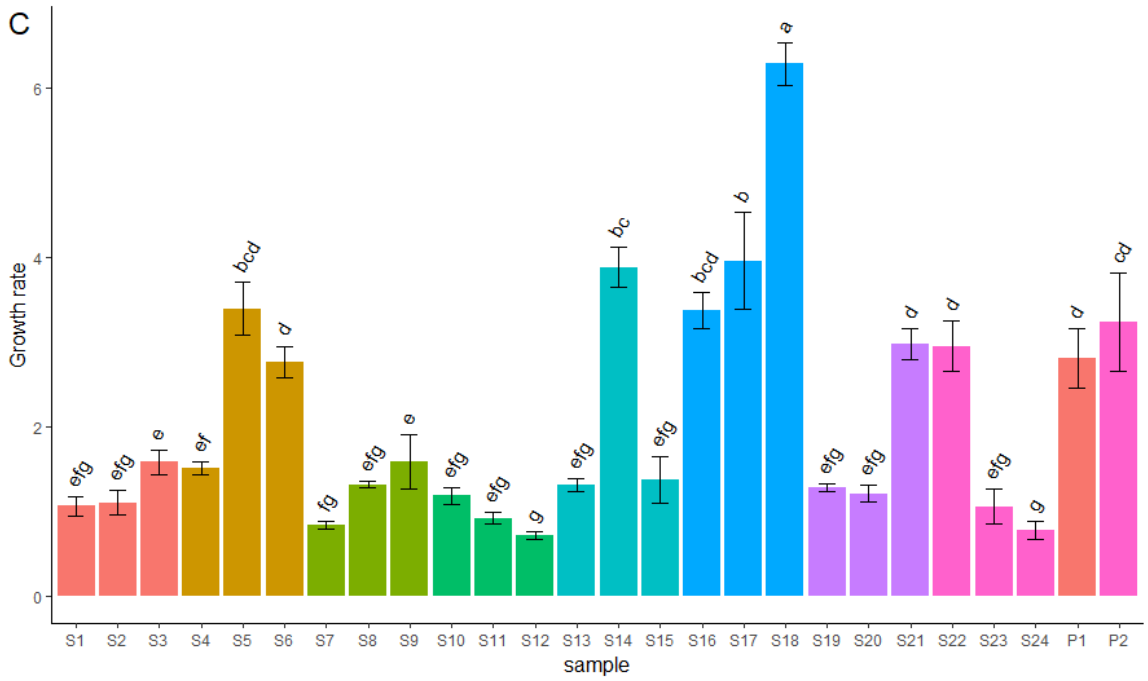
within each of the genotypes. The genotype with the highest average growths at 40 ℃ was genotype GENO#6 (allelic combination: αBγ), which also had great distinction with its opposite genotype GENO#3 (allelic combination: aβC). These results suggest that allelic combinations at the three SNP-rich regions had some influence on growth at the 40 ℃ environment. However, potential mutations in other genomic regions were also likely involved. Microscopic observations of cells identified that progeny with vigorous growths at the 40 ℃ environment had regular spherical cells while those without obvious growth had irregularly shaped cells (Figure 2.5D).

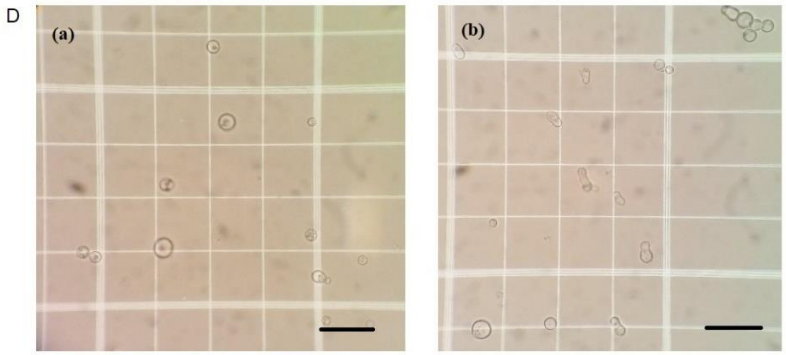
Interestingly, the progeny (S15) with the most growth at the 30 ℃ environment had among the least growths at the 40 ℃ environment. Similarly, five progeny (S5, S14, S16, S17, S18) that showed vigorous growth at the 40 ℃ environment had average growths at the 30 ℃ environment. However, a correlation test considering all the samples (Figure 2.5E) doesn't support a strong positive or negative correlation between these two temperatures.

Compared with capsule formation or melanin production, the effects of temperature are likely much more complicated with many more processes and mechanisms involved. These include the synthesis of proteins, enzyme activities, membrane properties, and cell divisions and mitosis. However, strains identified here that can grow vigorously at high temperature could help reveal the underlying genetic mechanisms, and bring valuable information for understanding the pathogenesis in hosts.


A
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
P1
P2
Log growth rate
Day
Temp
30
40


B
Log growth rate
sample
S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 P1 P2


C
Growth rate
0
2
4
6
efg
efg
e
ef
bcd
d
fg
efg
e
efg
efg
g
efg
bc
efg
bcd
b
a
efg
efg
d
d
efg
g
d
cd
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20
S21
S22
S23
S24
P1
P2
sample


D
(a)
(b)

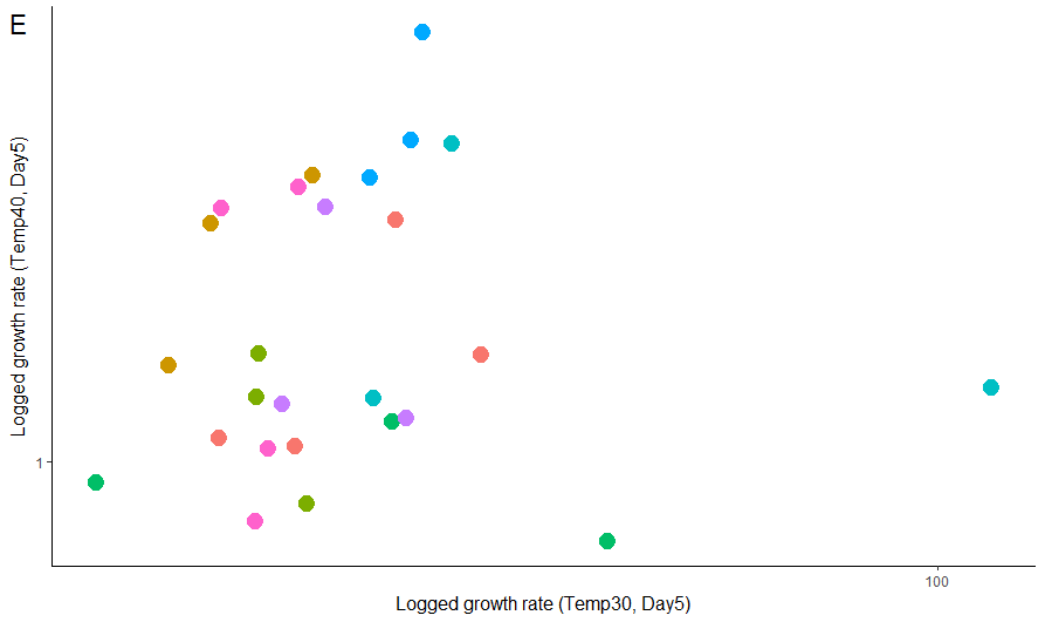


Figure 2.5 Growth patterns of segregating progeny from the "isogenic" cross of *C. neoformans* under different temperatures in RPMI. Growth rate is calculated by dividing cell density at observation point by start amount [10^5 (± 10%) cells/mL]. 24 progeny (S1-S24) of eight genotypes (colour differentiated) and two parental strains (P1: JEC20 and P2: JEC21) are presented. For each observation of growth condition for each progeny, eight repeats were done. A. Observation on day 0, 1, 3, 5 respectively with both temperature conditions. The Y-axis represents growth rate transformed by log. B. Growth condition on day 5 at 30 °C. The X-axis represents 26 samples and the Y axis respresents growth rate (logged). Letters represent statistically significant groups with progeny sharing the same letter having non-significantly different capsule production abilities. P values are corrected for multiple tests ($p<0.0001$). C. Growth condition on day 5 at 40 °C. The Y axis respresents growth rate (unlogged) ($p<0.0001$). D. Cell morphology under 100x microscope after 5-day incubation at 40 °C. (a) progeny S17 and (b) progeny S23. Length of bar is 50 µm. E. Relationship between growth rates at 40 °C and 30 °C after 5 days incubation. Correlation coefficient = -0.037.

2.4.3 Genotypes of progeny samples and correlation tests

The above results suggest that genetic differences between JEC20 and JEC21 outside of the three SNP-rich regions contribute to phenotypic variations in the progeny population. To identify the potential roles of those additional SNPs, we obtained the genome sequences of the 24 progeny and mapped their reads back to the SNPs of both parental genomes. The genotypes of the 24 progeny at the three SNP-rich regions were all confirmed by our whole genome sequencing results. In addition, 6 SNPs were identified among those progeny (Table 2.8) and given 18 genotypes together with 3 SNP-rich regions (Table 2.9). All 6 SNPs have already been found within the 1509 SNPs revealed earlier by genomic comparison of JEC20 and JEC21. Three of the six SNPs are located in genes on chromosomes 5, 10, 12 and only SNP3 caused an alternation of amino acid (from Asparagine in JEC21 to Lysine in JEC20) in a hypothetical protein. The two other SNPs are synonymous changes in exon regions of two other hypothetical proteins. The remaining three SNPs are located in unknown regions according to transcriptome information (among those 1091 SNPs outside transcribed regions identified before). Within these 6 SNPs, only two SNPs (SNP#1 and SNP#3) had relatively equal allele distributions among the 24 progeny sample while the remaining four had significantly biased allele distributions.

The 18 genotypes identified based on the three SNP-rich regions and the six SNPs were further analyzed to determine if they contributed significantly to the phenotypic differences among the progeny. The summary results are presented in Table 2.10. Our analyses showed these genotypic differences did have effects on all three virulence-

related phenotypic traits examined here. We further conducted a simple correlation test to determine if there was any association between each of the three phenotypic traits and the genetic loci at each of the six SNPs and each of the three SNP-rich regions individually. Only one SNP-rich region (#1) was found to be significantly associated with high temperature growth ability (<0.05) (Table 2.11). Three other potential correlations (<0.1 and >0.05) were also found between SNP#1 and high temperature growth, between the mating type region and high temperature growth, and between SNP#5 and capsule formation. A greater sample size may help to reveal whether these three correlations would be statistically significant.

Linkage disequilibrium among the variant sites and regions in this progeny population could affect our tests of the relationships between genotype and phenotype. Thus, we further determined whether there was significant association among the nine variant sites/regions (six SNPs and three SNP-rich regions). The results of the analyses are shown in Table 2.12. Indeed, we found significant linkage disequilibrium between SNP1 and SNP3, SNP1 and SNP6, SNP5 and SNP6, and SNP3 and SNP6 (especially the former two that further confirmed with Chi-squared test as well). The significant linkage disequilibrium observed here suggest that multiple loci might together contribute to the phenotypic variations among the progeny here. To test this possibility, we first employed a generalized linear model to evaluate the potential effects of all the genotype regions on each of the three phenotypic traits (original models in Table 2.13). Our results showed that these variant SNPs or SNP-rich regions contributed variably to each of the traits. Overall, SNP#3 and SNP#5 contributed the most to capsule production while SNP-rich

region #1 contributed the most to high temperature growth. Further improvements of the linear model were made by filtering higher p-value factors and also comparing R^2 value. The top several models were selected to describe the possible relationships and reveal their contributions (other numbered models in Table 2.13). Through these analyses, SNP#5 and the mating type region (SNP-rich region #3) was found to have likely contributed to melanin production, SNP#2 and SNP-rich region #2 might play a role in the growth ability at high temperature. While statistically not as robust, these new models suggest candidate relationships worthy of further investigation in future studies.

Because of the sample size restrictions and the nature of the cross, the correlation results obtained above are preliminary for revealing the relationship between genetic and phenotypic variations. Specifically, the progeny selected for the analyses were chosen based on their genotypes at the three SNP-rich regions. As a result, other SNPs were biased in their frequencies, leaving some of our analyses with little power to detect statistical differences. And the limited number of progeny didn't cover all kinds of combinations of 9 loci/regions, which made it hard to separately analyse the effect of each single factor. Secondly, around 1509 SNPs were identified outside the three SNP-rich regions between JEC20 and JEC21 whereas only 6 of them were detected among limited numbers of progeny. Other potential polymorphic markers contributing to the differences between JEC20 and JEC21 might exist but were monomorphic among our 24 progeny. A ramdon progeny sample with significantly larger population size would allow us to detect more polymorphic regions and potentially more correlations between genotypes and phenotypes within the progeny of JEC20 and JEC21.

Table 2.8 SNPs identified by genotyping of progeny using whole-genome sequencing

NO.	Chr.	Position	Gene	Expected change
SNP#1	5	16608	CNE00100	synonymous
SNP#2	5	1460209	CNE05190	synonymous
SNP#3	10	620	CNJ00010	N (JEC21) to K (JEC20)
SNP#4	10	845413	CNJ02800	Unknown function
SNP#5	12	156060	CNL04180	Unknown function
SNP#6	12	156156	CNL04180	Unknown function

Table 2.9 Genotyping of 24 progeny for SNPs including SNP-rich regions

Samples	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Region3	Region1	Region2	Genotype
S1	C	G	G	T	A	A	**a**	**B**	**C**	1
S2	T	G	C	T	A	G	**a**	**B**	**C**	2
S3	T	G	C	T	A	G	**a**	**B**	**C**	2
S4	C	G	G	T	A	G	**a**	**B**	γ	3
S5	T	G	G	T	A	G	**a**	**B**	γ	4
S6	T	G	C	T	A	G	**a**	**B**	γ	5
S7	T	G	C	T	A	G	**a**	β	**C**	6
S8	C	G	G	T	A	G	**a**	β	**C**	7
S9	T	G	C	T	A	G	**a**	β	**C**	6
S10	T	G	C	T	A	G	**a**	β	γ	8
S11	C	G	G	T	A	G	**a**	β	γ	9
S12	C	G	G	T	G	A	**a**	β	γ	10
S13	T	G	G	T	A	G	α	**B**	**C**	11
S14	T	G	C	T	A	G	α	**B**	**C**	12
S15	T	G	C	T	A	G	α	**B**	**C**	12
S16	T	G	C	T	A	G	α	**B**	γ	13
S17	T	G	C	C	A	G	α	**B**	γ	14
S18	T	G	C	T	A	G	α	**B**	γ	13
S19	T	G	C	T	A	G	α	β	**C**	15
S20	C	G	C	T	A	G	α	β	**C**	16
S21	T	A	G	T	A	G	α	β	**C**	17
S22	T	G	C	T	A	G	α	β	γ	18
S23	T	G	C	T	A	G	α	β	γ	18
S24	T	G	C	T	A	G	α	β	γ	18
Genotype same with JEC20/JEC21	6/18	1/23	8/16	1/23	1/23	2/22	12/12	12/12	12/12	

Table 2.10 One-way ANOVA test result based on 26 strains individually and 18 genotypes using existing markers

Phenotypes	Grouping	**Source of Variation**	**df**	**SS**	**MS**	**F**	**P-value**
Melanin production	Among individuals	Between Groups	25	651981677	26079267	60.88	<2e-16
		Within Groups	174	74540850	428396		
	Among 18 genotypes	Between Groups	17	379294880	22311464	11.70	<2e-16
		Within Groups	182	347227647	1907844		
Capsule formation	Among individuals	Between Groups	25	53.37	2.13	363.7	<2e-16
		Within Groups	390	2.29	0.01		
	Among 18 genotypes	Between Groups	17	38.66	2.27	53.25	<2e-16
		Within Groups	398	17.00	0.04		
Growth ability at high temperature	Among individuals	Between Groups	25	368.30	14.73	248.5	<2e-16
		Within Groups	182	10.80	0.06		
	Among 18 genotypes	Between Groups	17	255.00	15.00	22.98	<2e-16
		Within Groups	190	124.00	0.65		

Table 2.11 P-values of correlation test between three phenotypes and each of the polymorphic regions

	SNP1	**SNP2**	**SNP3**	**SNP4**	**SNP5**	**SNP6**	**Region3**	**Region1**	**Region2**
Melanin production	0.9095	0.6620	0.8240	0.6508	0.1747	0.4620	0.1104	0.8269	0.7594
Capsule formation	0.7765	0.4068	0.1724	0.5338	0.0822	0.3945	0.3193	0.6409	0.9986
High temperature growth ability	0.0655	0.4916	0.3695	0.1560	0.3469	0.2365	0.0644	0.0253	0.1719

Table 2.12 P-values of correlation test and Chi-squared test between each pairs of polymorphic gene regions

		SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Region3	Region1	Region2
Correlation test	**SNP1**	X	0.5753	0.0015	0.5753	0.0829	0.0089	0.0633	0.3676	1.0000
	SNP2	X	X	0.1619	0.8401	0.8401	0.7704	0.3282	0.3282	0.3282
	SNP3	X	X	X	0.4918	0.1619	0.0377	0.0901	1.0000	1.0000
	SNP4	X	X	X	X	0.8401	0.7704	0.3282	0.3282	0.3282
	SNP5	X	X	X	X	X	0.0002	0.3282	0.3282	0.3282
	SNP6	X	X	X	X	X	X	0.1522	1.0000	1.0000
	Region3	X	X	X	X	X	X	X	1.0000	1.0000
	Region1	X	X	X	X	X	X	X	X	1.0000
	Region2	X	X	X	X	X	X	X	X	X
		SNP1	**SNP2**	**SNP3**	**SNP4**	**SNP5**	**SNP6**	**Region3**	**Region1**	**Region2**
Chi-squared test	**SNP1**	X	1.0000	0.0124	1.0000	0.5553	0.0881	0.1573	0.6374	1.0000
	SNP2	X	X	0.7180	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	SNP3	X	X	X	1.0000	0.7180	0.1917	0.1939	1.0000	1.0000
	SNP4	X	X	X	X	1.0000	1.0000	1.0000	1.0000	1.0000
	SNP5	X	X	X	X	X	0.1236	1.0000	1.0000	1.0000
	SNP6	X	X	X	X	X	X	0.4602	1.0000	1.0000
	Region3	X	X	X	X	X	X	X	1.0000	1.0000
	Region1	X	X	X	X	X	X	X	X	1.0000
	Region2	X	X	X	X	X	X	X	X	X

Table 2.13 Original and improved linear model between 3 phenotypes and all genotype regions of each progeny strain.

Melanin production

Linear model	Factors	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Region 3	Region 1	Region 2	R^2 value
Original	Contribution	5.08%	11.72%	6.18%	19.26%	33.83%	0.13%	14.08%	7.80%	1.93%	0.3181
	P-value	0.7078	0.6338	0.6269	0.3428	0.2412	0.9953	0.1194	0.3852	0.8137	
1	Contribution	5.10%	11.72%	6.17%	19.26%	33.94%		14.08%	7.81%	1.92%	0.3181
	P-value	0.6837	0.6215	0.6142	0.3255	0.1067		0.106	0.3521	0.8044	
2	Contribution		7.95%	3.42%	21.36%	42.75%		16.67%	7.85%		0.2758
	P-value		0.73489	0.74534	0.32051	0.06619		0.08828	0.37502		
3	Contribution				24.22%	46.66%		21.08%	8.05%		0.2743
	P-value				0.301	0.055		0.032	0.389		

Capsule formation

Linear model	Factors	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Region 3	Region 1	Region 2	R^2 value
Original	Contribution	13.38%	5.01%	16.69%	1.55%	39.43%	6.89%	5.77%	8.30%	2.99%	0.4840
	P-value	0.148	0.757	0.061	0.906	0.048	0.63	0.32	0.169	0.581	
1	Contribution	13.60%	5.21%	17.00%		39.90%	6.94%	5.93%	8.54%	2.88%	0.4834
	P-value	0.131	0.741	0.049		0.04	0.619	0.289	0.138	0.577	
2	Contribution	13.47%		16.79%		43.02%	7.53%	7.04%	8.58%	3.58%	0.4795
	P-value	0.119		0.026		0.036	0.61	0.215	0.13	0.499	
3	Contribution	14.50%		20.34%		43.60%		8.61%	9.32%	3.62%	0.4707
	P-value	0.128		0.02		0.016		0.193	0.141	0.548	

High temperature growth ability

Linear model	Factors	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	Region 3	Region 1	Region 2	R^2 value
Original	Contribution	0.63%	33.08%	6.76%	6.23%	1.73%	7.50%	8.98%	21.09%	14.00%	0.5355
	P-value	0.964	0.206	0.608	0.764	0.953	0.739	0.325	0.035	0.115	
1	Contribution		34.25%	7.26%	6.30%		7.01%	9.19%	21.53%	14.47%	0.5353
	P-value		0.133	0.471	0.75		0.643	0.29	0.012	0.075	
2	Contribution		36.13%	7.93%			7.12%	10.04%	23.07%	15.71%	0.5323

	P-value	0.119	0.437		0.643	0.251	0.007	0.052	
3	Contribution	36.89%	8.97%	5.39%		10.19%	22.47%	16.09%	0.5281
	P-value	0.1067	0.356	0.7937		0.2399	0.0097	0.0499	
4	Contribution	39.04%	10.05%			10.74%	23.88%	16.30%	0.5261
	P-value	0.0893	0.2882			0.2193	0.0055	0.0449	

Original linear model includes all 9 loci/regions as independent variables and each phenotype as dependent variable. The model numbers 1, 2, 3, and 4 refer to improved linear models. All linear models including n loci/regions (n=1~9) as independent variables were made and ordered by R^2 value of each model. The top 10% models were further compared by filtering the ones containing high p-value factors and ensuring their diagnostic plots are relatively evenly distributed. By these approaches, the improved linear models were selected.

2.5 Conclusions

My thesis work identified the genome sequence differences between a pair of "isogenic" model strains of *C. neoformans*, JEC20 and JEC21 and investigated their potential effects on phenotypes. We found three regions significantly enriched SNPs. Based on markers developed from these three regions, I constructed a cross between JEC20 and JEC21 and screened a set of meiotic progeny to test the relationships between these polymorphic genomic regions and four phenotypes. While the mating test confirmed what was expected of the *C. neoformans* mating system, the analyses of three other traits revealed unexpected diversity and genetic relationships among the progeny. My results add to the complexity of genotype-phenotype relationships in this important human fungal pathogen and suggested potential further work that could be done to help dissect such complexity.

For the genome sequence comparison, we obtained the whole genome sequence of strain JEC20 and then compared it to strain JEC21 to detect sequence polymorphisms between the two genomes. By this approach, we revealed three SNP-rich regions together with several SNPs outside these regions in other parts of the genome. The comparison revealed that there were more genomic differences between JEC20 and JEC21 than previously assumed. Of these three SNP-rich regions, one was located on chromosome 4, consistent with what was noted previously (Kwon-Chung, 1992a; Lengelar, 2002). The other two were both located on chromosome 2. In addition, we found a large number of SNPs outside of these two regions. However, only a few of those were confirmed in the progeny. My analyses suggested that these novel mutations were most likely the results

of mutation accumulation between the creation of strains JEC20 and JEC21 and the more than ten years of sub-culturing in laboratory conditions before they were sequenced.

My analyses of the meiotic progeny from the JEC20 and JEC21 cross indicated a diversity of relationships between genotypes and phenotypes. Overall, the assay results showed that (i) mating ability was only related to mating-type region on chromosome 4; (ii) the other three phenotype features (melanin production, capsule formation, and high temperature growth ability) were quantitative traits. While some contributions concerning the three SNP-rich regions were found, and effects other than theses were not exclusive; (iii) instead, other SNPs outside the SNP-rich regions as well as locus-locus interactions could contribute to the observed phenotypic variations. Therefore, additional genome sequencing of these progeny were also done to identify the complete distribution of other SNPs among the strains. We found 6 additional SNPs among those progeny and assessed their effects on the phenotypic variation. Our results suggest SNP#5 was correlated with melanin production and capsule formation, and that SNP-rich region #1 was associated with high temperature growth ability.

However, because of the limited progeny size, we were unable to perform full locus-locus interaction effects and that certain polymorphic regions between JEC20 and JEC21 were not segregating among the 24 progeny. Thus, a greater progeny sample sizes are needed in order to further explore the relationships between genotype and phenotype. Regradless, our study revealed that we should exercise caution when making conclusions based on results from “isogenic” strains derived from classical genetic crosses.

Bibliography

Alanio, A., Vernel-Pauillac, F., Sturny-Leclère, A., & Dromer, F. (2015). *Cryptococcus neoformans* host adaptation: toward biological evidence of dormancy. MBio, 6(2), e02580-14.

Alspaugh, J. A., Perfect, J. R., & Heitman, J. (1997). *Cryptococcus neoformans* mating and virulence are regulated by the G-protein α subunit GPA1 and cAMP. Genes & Development, 11(23), 3206-3217.

Alspaugh, J. A., Cavallo, L. M., Perfect, J. R., & Heitman, J. (2000). RAS1 regulates filamentation, mating and growth at high temperature of *Cryptococcus neoformans*. Molecular microbiology, 36(2), 352-365.

Angkasekwinai, P., Sringkarin, N., Supasorn, O., Fungkrajai, M., Wang, Y. H., Chayakulkeeree, M., ... & Pattanapanyasat, K. (2014). Cryptococcus gattii infection dampens Th1 and Th17 responses by attenuating dendritic cell function and pulmonary chemokine expression in the immunocompetent hosts. Infection and immunity, 82(9), 3880-3890.

Bartlett, K. H., Kidd, S. E., & Kronstad, J. W. (2008). The emergence of Cryptococcus gattii in British Columbia and the Pacific Northwest. Current Infectious Disease Reports, 10(1), 58-65.

Bartlett, K. H., Cheng, P. Y., Duncan, C., Galanis, E., Hoang, L., Kidd, S., ... & Morshed, M. (2012). A decade of experience: Cryptococcus gattii in British Columbia. Mycopathologia, 173(5-6), 311-319.

Benham, R. W. (1950). Cryptococcosis and blastomycosis. Annuals of the New York Academy of Sciences, 50, 1299-1314.

Botts, M. R., Giles, S. S., Gates, M. A., Kozel, T. R., & Hull, C. M. (2009). Isolation and characterization of *Cryptococcus neoformans* spores reveal a critical role for capsule biosynthesis genes in spore biogenesis. Eukaryotic cell, 8(4), 595-605.

Buchanan, K. L., & Murphy, J. W. (1998). What makes *Cryptococcus neoformans* a pathogen?. Emerging infectious diseases, 4(1), 71.

Chang, Y. C., Penoyer, L. A., & Kwon-Chung, K. J. (1996). The second capsule gene of *Cryptococcus neoformans*, CAP64, is essential for virulence. Infection and Immunity, 64(6), 1977-1983.

Chang, Y. C., & Kwon-Chung, K. J. (1998). Isolation of the third capsule-associated gene, CAP60, required for virulence in *Cryptococcus neoformans*. Infection and Immunity, 66(5), 2230-2236.

Coelho, C., Bocca, A. L., & Casadevall, A. (2014). The intracellular life of *Cryptococcus neoformans*. Annual Review of Pathology: Mechanisms of Disease, 9, 219-238.

Cogliati, M. (2013). Global molecular epidemiology of *Cryptococcus neoformans* and Cryptococcus gattii: an atlas of the molecular types. Scientifica, 2013.

Cramer, K. L., Gerrald, Q. D., Nichols, C. B., Price, M. S., & Alspaugh, J. A. (2006). Transcription factor Nrg1 mediates capsule formation, stress response, and pathogenesis in *Cryptococcus neoformans*. Eukaryotic Cell, 5(7), 1147-1156.

D'Souza, C. A., & Heitman, J. (2001). Conserved cAMP signaling cascades regulate fungal development and virulence. FEMS microbiology reviews, 25(3), 349-364.

Ekena, J. L., Stanton, B. C., Schiebe-Owens, J. A., & Hull, C. M. (2008). Sexual development in *Cryptococcus neoformans* requires CLP1, a target of the homeodomain transcription factors Sxi1α and Sxi2a. Eukaryotic cell, 7(1), 49-57.

Ellis, D., & Pfeiffer, T. (1990). Ecology, life cycle, and infectious propagule of *Cryptococcus neoformans*. The Lancet, 336(8720), 923-925.

Evans, E. E. (1950). The antigenic composition of *Cryptococcus neoformans*: 1. A serologic classification by means of the capsular and agglutination reactions. The Journal of Immunology, 64, 423-430.

Feng, B., Wang, X., Hauser, M., Kaufmann, S., Jentsch, S., Haase, G., ... & Szaniszlo, P. J. (2001). Molecular cloning and characterization ofWdPKS1, a gene involved in dihydroxynaphthalene melanin biosynthesis and virulence in Wangiella (Exophiala) dermatitidis. Infection and immunity, 69(3), 1781-1794.

Forsythe, A., Vogan, A., & Xu, J. (2016). Genetic and environmental influences on the germination of basidiospores in the *Cryptococcus neoformans* species complex. Scientific Reports, 6.

Fu, J., Mares, C., Lizcano, A., Liu, Y., & Wickes, B. L. (2011). Insertional mutagenesis combined with an inducible filamentation phenotype reveals a conserved STE50 homologue in *Cryptococcus neoformans* that is required for monokaryotic fruiting and sexual reproduction. Molecular microbiology, 79(4), 990-1007.

Giles, S. S., Batinić-Haberle, I., Perfect, J. R., & Cox, G. M. (2005). *Cryptococcus neoformans* mitochondrial superoxide dismutase: an essential link between antioxidant function and high-temperature growth. Eukaryotic cell, 4(1), 46-54.

Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., ... & Boekhout, T. (2015). Recognition of seven species in the Cryptococcus gattii/*Cryptococcus neoformans* species complex. Fungal Genetics and Biology, 78, 16-48.

Hopfer, R. L., & Blank, F. (1975). Caffeic acid-containing medium for identification of *Cryptococcus neoformans*. Journal of clinical microbiology, 2(2), 115-120.

Hull, C. M., Davidson, R. C., & Heitman, J. (2002). Cell identity and sexual development in *Cryptococcus neoformans* are controlled by the mating-type-specific homeodomain protein Sxi1α. Genes & development, 16(23), 3046-3060.

Hull, C. M., Boily, M. J., & Heitman, J. (2005). Sex-specific homeodomain proteins Sxi1α and Sxi2a coordinately regulate sexual development in *Cryptococcus neoformans*. Eukaryotic cell, 4(3), 526-535.

Idnurm, A., Reedy, J. L., Nussbaum, J. C., & Heitman, J. (2004). *Cryptococcus neoformans* virulence gene discovery through insertional mutagenesis. Eukaryotic cell, 3(2), 420-429.

Idnurm, A., Bahn, Y. S., Nielsen, K., Lin, X., Fraser, J. A., & Heitman, J. (2005). Deciphering the model pathogenic fungus *Cryptococcus neoformans*. Nature Reviews Microbiology, 3(10), 753-764.

Janbon, G., Himmelreich, U., Moyrand, F., Improvisi, L., & Dromer, F. (2001). Cas1p is a membrane protein necessary for the O-acetylation of the *Cryptococcus neoformans* capsular polysaccharide. Molecular microbiology, 42(2), 453-467.

Janbon, G., Ormerod, K. L., Paulet, D., Byrnes III, E. J., Yadav, V., Chatterjee, G., ... & Bahn, Y. S. (2014). Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS genetics, 10(4), e1004261.

Knoke, M., & Schwesinger, G. (1994). One hundred years ago: the history of crytococcosis in Greifswald. Medical mycology in the nineteenth century. Mycoses, 37(7-8), 229-233.

Kwon-Chung, K. J. (1975). Description of a new genus, Filobasidiella, the perfect state of *Cryptococcus neoformans*. Mycologia 67: 1197–1200

Kwon-Chung, K. (1976). Morphogenesis of Filobasidiella neoformans, the Sexual State of *Cryptococcus neoformans*. Mycologia, 68(4), 821-833. doi:10.2307/3758800

Kwon-Chung, K. J., and Bennett, J. E. (1992a). Medical Mycology. Lea & Febiger, Philadelphia.

Kwon-Chung, K. J., Edman, J. C., & Wickes, B. L. (1992b). Genetic association of mating types and virulence in *Cryptococcus neoformans*. Infection and Immunity, 60(2), 602-605.

Kwon-Chung, K. J., Fraser, J. A., Doering, T. L., Wang, Z. A., Janbon, G., Idnurm, A., & Bahn, Y. S. (2014). *Cryptococcus neoformans* and Cryptococcus gattii, the etiologic agents of cryptococcosis. Cold Spring Harbor perspectives in medicine, 4(7), a019760.

Kwon-Chung, K. J., Bennett, J. E., Wickes, B. L., Meyer, W., Cuomo, C. A., Wollenburg, K. R., ... & Cogliati, M. (2017). The case for adopting the "species complex" nomenclature for the etiologic agents of cryptococcosis. mSphere, 2(1), e00357-16.

Lengeler, K. B., Cox, G. M., & Heitman, J. (2001). Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infection and immunity, 69(1), 115-122.

Lengeler, K. B., Fox, D. S., Fraser, J. A., Allen, A., Forrester, K., Dietrich, F. S., & Heitman, J. (2002). Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. Eukaryotic Cell, 1(5), 704-718.

Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-1760.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987-2993.

Lin, X., Hull, C. M., & Heitman, J. (2005). Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature, 434(7036), 1017-1021.

Loftus, B. J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., ... & Allen, J. E. (2005). The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science, 307(5713), 1321-1324.

Lu, Y. K., Sun, K. H., & Shen, W. C. (2005). Blue light negatively regulates the sexual filamentation via the Cwc1 and Cwc2 proteins in *Cryptococcus neoformans*. Molecular microbiology, 56(2), 480-491.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1), pp. 10-12. doi:http://dx.doi.org/10.14806/ej.17.1.200

May, R. C., Stone, N. R., Wiesner, D. L., Bicanic, T., & Nielsen, K. (2016). Cryptococcus: from environmental saprophyte to global pathogen. Nature reviews. Microbiology, 14(2), 106-117.

Meyer, W., Aanensen, D. M., Boekhout, T., Cogliati, M., Diaz, M. R., Esposto, M. C., ... & Litvintseva, A. P. (2009). Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and Cryptococcus gattii. Medical mycology, 47(6), 561-570.

Milne, I., Stephen, G., Bayer, M., Cock, P. J., Pritchard, L., Cardle, L., ... & Marshall, D. (2012). Using Tablet for visual exploration of second-generation sequencing data. Briefings in bioinformatics, bbs012.

Moyrand, F., Klaproth, B., Himmelreich, U., Dromer, F., & Janbon, G. (2002). Isolation and characterization of capsule structure mutant strains of *Cryptococcus neoformans*. Molecular microbiology, 45(3), 837-849.

Ni, M., Feretzaki, M., Li, W., Floyd-Averette, A., Mieczkowski, P., Dietrich, F. S., & Heitman, J. (2013). Unisexual and heterosexual meiotic reproduction generate aneuploidy and phenotypic diversity de novo in the yeast *Cryptococcus neoformans*. PLoS Biology, 11(9), e1001653.

Nimrichter, L., Rodrigues, M. L., Rodrigues, E. G., & Travassos, L. R. (2005). The multitude of targets for the immune system and drug therapy in the fungal cell wall. Microbes and infection, 7(4), 789-798.

Patel RK, Jain M (2012) NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. PLoS ONE 7(2): e30619. https://doi.org/10.1371/journal.pone.0030619

Park, B. J., Wannemuehler, K. A., Marston, B. J., Govender, N., Pappas, P. G., & Chiller, T. M. (2009). Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. Aids, 23(4), 525-530.

Piccioni, M., Monari, C., Kenno, S., Pericolini, E., Gabrielli, E., Pietrella, D., ... & Vecchiarelli, A. (2013). A purified capsular polysaccharide markedly inhibits inflammatory response during endotoxic shock. Infection and immunity, 81(1), 90-98.

Pierini, L. M., & Doering, T. L. (2001). Spatial and temporal sequence of capsule construction in *Cryptococcus neoformans*. Molecular microbiology, 41(1), 105-115.

Rajasingham, R., Smith, R. M., Park, B. J., Jarvis, J. N., Govender, N. P., Chiller, T. M., ... & Boulware, D. R. (2017). Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. The Lancet Infectious Diseases.

Schmeding, K. A., Jong, S. C., & Hugh, R. (1981). Sexual compatibility between serotypes ofFilobasidiella neoformans (*Cryptococcus neoformans*). Current Microbiology, 5(3), 133-138.

Springer, D. J., & Chaturvedi, V. (2010). Projecting global occurrence of Cryptococcus gattii. Emerging infectious diseases, 16(1), 14.

Tangen, K. L., Jung, W. H., Sham, A. P., Lian, T., & Kronstad, J. W. (2007). The iron- and cAMP-regulated gene SIT1 influences ferrioxamine B utilization, melanization and cell wall structure in *Cryptococcus neoformans*. Microbiology, 153(1), 29-41.

Toffaletti, D. L., Nielsen, K., Dietrich, F., Heitman, J., & Perfect, J. R. (2004). *Cryptococcus neoformans* mitochondrial genomes from serotype A and D strains do not influence virulence. Current genetics, 46(4), 193-204.

Vallim, M. A., Nichols, C. B., Fernandes, L., Cramer, K. L., & Alspaugh, J. A. (2005). A Rac homolog functions downstream of Ras1 to control hyphal differentiation and high-

temperature growth in the pathogenic fungus *Cryptococcus neoformans*. Eukaryotic cell, 4(6), 1066-1078.

Vogan, A. A., Khankhet, J., Samarasinghe, H., & Xu, J. (2016). Identification of QTLs Associated with Virulence Related Traits and Drug Resistance in *Cryptococcus neoformans*. G3: Genes, Genomes, Genetics, 6(9), 2745-2759.

Wang, P., Perfect, J. R., & Heitman, J. (2000). The G-protein β subunit GPB1 is required for mating and haploid fruiting in *Cryptococcus neoformans*. Molecular and Cellular Biology, 20(1), 352-362.

Wang, Y., Aisen, P., & Casadevall, A. (1995). *Cryptococcus neoformans* melanin and virulence: mechanism of action. Infection and immunity, 63(8), 3131-3136.

Waugh, M. S., Nichols, C. B., DeCesare, C. M., Cox, G. M., Heitman, J., & Alspaugh, J. A. (2002). Ras1 and Ras2 contribute shared and unique roles in physiology and virulence of *Cryptococcus neoformans*. Microbiology, 148(1), 191-201.

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. Proceedings of the National Academy of Sciences, 93(14), 7327-7331.

Wickes, B. L., Edman, U., & Edman, J. C. (1997). The *Cryptococcus neoformans* STE12α gene: a putative Saccharomyces cerevisiae STE12 homologue that is mating type specific. Molecular microbiology, 26(5), 951-960.

Williamson, P. R., Wakamatsu, K., & Ito, S. (1998). Melanin biosynthesis in *Cryptococcus neoformans*. Journal of bacteriology, 180(6), 1570-1572.

Xu J, Ali RY, Gregory DA, Amick D, Lambert SE, et al. (2000) Uniparental mitochondrial transmission in sexual crosses in *Cryptococcus neoformans*. Current Microbiology 40: 269–273. doi:10.1007/s002849910053.

Yan, Z., & Xu, J. (2003). Mitochondria are inherited from the MATa parent in crosses of the basidiomycete fungus *Cryptococcus neoformans*. Genetics, 163(4), 1315-1325.

Yang, Z., Pascon, R. C., Alspaugh, A., Cox, G. M., & McCusker, J. H. (2002). Molecular and genetic analysis of the *Cryptococcus neoformans* MET3 gene and a met3 mutanta. Microbiology, 148(8), 2617-2625.

Zaragoza, O., & Nielsen, K. (2013). Titan cells in *Cryptococcus neoformans*: cells with a giant impact. Current opinion in microbiology, 16(4), 409-413.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research, 18(5), 821-829.