UNIVERSITY OF ZULULAND

Comparative structural analysis of cytochrome P450 family CYP51 in the genus *Cryptococcus*

Olufunmilayo Olukemi Akapo

Student number: 201426758

A thesis submitted in fulfilment of the requirements for the award of Doctor of Philosophy (PhD) in the Department of Biochemistry and Microbiology, Faculty of Sciences, University of Zululand, KwaDlangezwa

Supervisor: Prof Khajamohiddin Syed

KwaDlangezwa

October 2020

DECLARATION

I, Olufunmilayo Olukemi Akapo, declare that this thesis is entirely my own work and has not been taken from the work of others, except where I have appropriately acknowledged and referenced the original source. This thesis has never been submitted for any degree for examination at any university. Considering the quality is more important than the quantity (Gould, 2016), care has been taken to present thesis in a publication format to enable presentation of data in a concise manner and easy understanding of the work. Also, I declare that the work presented in this thesis has been published and the remaining work has been communicated for publishing as indicated in the research outputs section.

_____

Signed on the _10____ day of _October__, 2020

Reference: Gould, J., 2016. Future of the thesis. Nature, 535(7610), pp.26-29.

ABSTRACT

Cytochrome P450 monooxygenase CYP51 (sterol 14α-demethylase) is a well-known target of the azole drug fluconazole for treating cryptococcosis, a life-threatening fungal infection in immune-compromised patients in poor countries. Studies indicate that mutations in *CYP51* confer fluconazole resistance on cryptococcal species. Despite the importance of CYP51 in these species, genome data mining, annotation and phylogenetic analysis of CYPs has not been reported. Furthermore, few studies on the structural analysis of CYP51 and its interactions with different azole drugs have been reported. This study is aimed to address these two research gaps.

Comprehensive genome-wide CYP analysis revealed the presence of 203 CYPs (excluding 16 pseudo-CYPs) in 23 species of *Tremellomycetes* that can be grouped into 38 CYP families and 72 CYP subfamilies. Twenty-three CYP families are new and three CYP families (CYP5139, CYP51 and CYP61) were conserved across 23 species of *Tremellomycetes*. Pathogenic cryptococcal species have 50% fewer CYP genes than non-pathogenic species. After successful identification of CYPs, *in silico* structural analysis of 12 CYP51s from cryptococcal species and other *Tremellomycetes* were carried out. Interactions of 12 CYP51s with nine ligands (three substrates and six azoles) performed by Rosetta docking using 10 000 combinations for each of the CYP51-ligand complex (12 CYP51s x 9 ligands =108 complexes) and hierarchical agglomerative clustering were used for selecting the complexes. A web application for visualization of CYP51s' interactions with ligands was developed (http://bioshell.pl/azoledocking/). The study results indicated that *Tremellomycetes* CYP51s have a high preference for itraconazole, corroborating the *in vitro* effectiveness of itraconazole compared to fluconazole. Amino acids interacting with different ligands were found to be conserved across CYP51s, indicating that the procedure employed in this study is accurate and can be automated for studying P450-ligand interactions to cater for the growing number of P450s. The results of this study will serve as reference for future annotation and characterization of CYPs in species of *Tremellomycetes*.

DEDICATION

This research is dedicated to the Almighty Father for His faithfulness and mercies. This is also dedicated to my dear husband and wonderful children for their unfailing support.

ACKNOWLEDGEMENTS

All glory to the All-sufficient, Powerful, Almighty and Omnipotent God for the special enablement, divine health, divine strength and knowledge He gave to me to undertake this research project. My profound thanks go to Professor Syed Khajamohiddin, who gave me a warm welcome to his research group, which led to my journey of bioinformatics, and for guiding me throughout my research. Prof, as you are fondly called, you have been an amazing person to work with, straightforward and diligent. Your supervisory role allowed me to accomplish one of the most important goals in my life.

To my colleagues. Zinhle and Fanele, who made me feel comfortable in the P450 group, Tiara Padanyachee, Martins, Nokwanda, Nomfundo, Minnie, I sincerely appreciate you for sharing your knowledge with me in this research work. Similarly, I am grateful to Dr Dominic Gront in the Department of Chemistry, University of Warsaw, Poland who gave me the opportunity to visit his laboratory and guided me to complete the homology modelling part of this research. Dr Dominic, words would not be able to express how grateful I am for your willingness to allow me in your laboratory at short notice. You also gave me the platform to gain new knowledge in bioinformatics. I also appreciate Joanna (Asa) Macnar, who make sure my visit to Poland and Dr Dominic's laboratory was memorable. I want to say thank you as well to Natalia for those chocolates and goodies you frequently provided during my stay at the University of Warsaw, just to make sure I was happy. Thank you to Justyna for always lending a helping hand whenever I asked for help during my stay at the laboratory. You are all amazing people who made me learn different things within a short time and made me feel at home in Poland. My appreciation also goes to Professor Abidemi Kappo who gave his support during this research. I also want to thank Ms Barbara Bradley who took time to edit my thesis and make the necessary corrections to bring out the best in my thesis. I want to thank all the lecturers, staff members in the Department of Biochemistry and Microbiology who in one way or another has given me moral support throughout my studies in the department and to Professor Albertus Basson, the head of the Department of Biochemistry and Microbiology, who make sure he guided and supported me throughout my studies in the department.

I thank my senior colleagues, Dr Foluso Osunsanmi and Dr Bolajoko Ogunyinka, for all their efforts, counsel and moral support during this PhD programme. Dr Edwina Uzunuigbe and Dr Adenowo Fatimah encouraged me all the way throughout the steps in this programme. Thank you to my dearest father, Mr Ayinde Ogundipe, for passing on the legacy of education to me right from my youth; I am happy to fulfil your desires. Your labour, your moral support and financial investment in me has eventually paid off in the long run. This effort would never have been possible without my wonderful mother, Mrs Olabisi Ayoola, who has been supporting me morally and spiritually. Sweet Mother, I really appreciate your prayers day and night for me. To my brothers and my sister, I appreciate you all for your moral support and understanding. Words cannot be enough to appreciate my Big Uncle, Pastor Gabriel Kolawole and Pastor (Mrs) Comfort Adejimi for their tremendous efforts to make the journey of this PhD possible.

Special appreciation goes to my pastor, bosom friend, the father of my children, counsellor, my pillar of encouragement, Pastor Charles Akapo, the pastor in charge of the Redeemed Christian Church of God, Richards Bay, who has been there for me when things were bad or good, ready to offer help at all times. Your financial support and the sacrifices you made for me to finish this PhD programme are highly appreciated. May the Lord I serve never forget your labour of love and reward you in Jesus' name. To my three lovely children, Oluwapelumi Timothy, Oluwaremilekun Delight and Ebunoluwa Deborah Akapo, I sincerely appreciate all your sacrifices, understanding and supportive role during the programme; your all gave me the courage to carry on and become what the Lord had ordained me to be. My brethren from the Redeemed Christian Church of God in Richards bay, Empangeni Area and the House of Truth, Warsaw, Poland, I want to say thank you for all your moral and spiritual support.

RESEARCH OUTPUTS

- **Articles**

1. **Akapo, O.O.,** Padayachee, T., Chen, W., Yu, J.H., Nelson, D.R. and Syed, K., 2019. Distribution and diversity of cytochrome P450 monooxygenases in the fungal class *Tremellomycetes*. *International journal of molecular sciences*, *20*(12), p.2889.

2. **Akapo, O.O.**, Macnar, J.M., Kryś J.D., Syed, P.R., Syed, K., Dominik Gront. 2020. *In silico* structural modeling and analysis of interactions of *Tremellomycetes* cytochrome P450 monooxygenases CYP51s with substrates and azoles. Scientific Reports (Submission ID b2d39cce-8a1a-47a9-83c3-c907af353db5).

- **Conference presentation (international)**

30th International congress on Prevention of diabetes and complications & 15th World Congress on Infection Prevention and Control Conference on Infectious diseases and Prevention, Movenpick Conference Center, Zurich, Switzerland. 21- 22, October 2019. **Paper presented: Akapo, O. O.,** Padayachee, T., Chen, W, Yu, J-H & Nelson, D., Syed, K. (2019). Distribution and diversity of cytochrome P450 monooxygenases in the fungal class *Tremellomycetes.*

- **Conference Presentation (Local**)

Conference on Genomics, Proteomics and Metabolomics: All in the Bioinformatics (CGPMB). Umfolozi Hotel, Casino and Convention Center, Empangeni, KZN, South Africa. 27-28 July *2019.* **Paper Presented: Akapo, OO.,** Padayachee, T., Chen, W, Yu, J-H & Nelson, D., Syed, K. (2019). *In silico* analysis of cytochrome P450 monooxygenase proteome in the fungal class *Tremellomycetes.*

ABBREVIATIONS

3D	Three-dimensional
BLAST	Basic Local Alignment Search Tool
NCBI	National Centre for Biotechnology Information
PDB	Protein Data Bank
Phyre	Protein Homology/Analogy Recognition Engine
RCSB	Research Collaborator for Structural Bioinformatics
CYPs	Cytochromes P450 monooxygenases
RMSD	Root Mean-square Deviation
JGI	Joint Genome Institute

TABLE OF CONTENTS

DECLARATION i

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGEMENTS iv

RESEARCH OUTPUTS vi

ABBREVIATIONS vii

TABLE OF CONTENTS viii

LIST OF FIGURES xi

LIST OF TABLES xii

CHAPTER 1 1

Introduction 1

1.2 Problem statement 1

1.3 Novelty 2

1.4 Aim and objectives of the study 2

1.4.1 Aim of the study 2

1.4.2. Objectives of the study 2

1.5. Thesis overview 3

CHAPTER 2 4

Literature review 4

2.1. Cryptococcosis - history 4

2.2. Cryptococcal species 4

2.3. Taxonomic hierarchy and classification of Cryptococcus 6

2.4. Cryptococcus habitat 7

2.5. Morphology of Cryptococcus cells 8

2.6 Mode of transfer and infection onset 10

2.7. Epidemiology 11

2.8. Cryptococcosis and future therapies for cryptococcosis 13
2.9. Cytochrome P450 monooxygenase CYP51 15
CHAPTER 3 17
3.1 Introduction 17
3.2. Methodology 21
3.2.1. Species and databases 21
3.2.2. CYP mining and annotation 22
3.2.3. Phylogenetic analysis of CYPs 22
3.2.4. Generation of CYP profile heat-maps 23
3.2.5. CYP diversity percentage analysis 23
3.2.6. Functional prediction of CYPs 23
3.3. Results and discussion 23
3.3.1. Pathogenic cryptococcal species have few CYPs in their genomes 23
3.3.2 New CYP families were found in *Tremellomycetes* 25
3.3.3. Four CYP families are conserved in pathogenic cryptococcal species 27
3.3.4. Pathogenic cryptococcal species have the highest CYP diversity 30
3.3.5. Most of the CYPs from the species of *Tremellomycetes* are orphans with no known function 31
3.4 Conclusions 31
CHAPTER 4 32
4.1. Introduction 32
4.2. Methods 34
4.2.1. CYP51s used in the study 34
4.2.2. Phylogenetic analysis 35
4.2.3. Homology modelling of CYP51s 35
4.3. Preparation of CYP51s models for docking 36
4.3.1 Ligands and their docking procedure 36

4.4 Hierarchical agglomerative clustering 37

4.5. Analysis of docking poses 37

4.5.1 Protocols 38

4.5.2. Data visualisation 38

4.5.3. Active site cavity and ligand interacting amino acids detection 38

4.6 Results and discussion 38

4.6.1 CYP51s grouped as per *Tremellomycetes* lifestyle 38

4.6.2. CYP51s of *Tremellomycetes* have all the CYP characteristic motifs 40

4.6. 3. *Tremellomycetes* CYP51s active site cavities are highly hydrophobic . 42

4.6.4. A web application for visualization of *Tremellomycetes* CYP51s interactions with ligands 44

4.6.5 *Tremellomycetes* CYP51s have the highest preference for itraconazole 45

4.6.7. High conservation observed in *Tremellomycetes* CYP51s amino acids interacting with ligands..... 50

4.7. Conclusions..... 52

CHAPTER 5..... 53

Overall conclusions and future perspectives 53

REFERENCES 55

APPENDICES 72

APPENDIX A..... 72

APPENDIX B..... 74

LIST OF FIGURES

Figure 2.1. Schematic diagram representing *Cryptococcus gatti* molecular types dominance in different parts of the world 7
Figure 2.2. Habitat and transmission of C*ryptococci* 8
Figure 2.3. Schematic diagram representing infection establishment of cryptococcal cells in an immunocompromised patient 11
Figure 2.4. Prevalence of cryptococcosis in the world 12
Figure 2.5. Current and future therapies for cryptococcosis 15
Figure 2.6. CYP51 role in synthesis of membrane sterols across the phyla 16
Figure 3.1. Comparative analysis of CYPs in the species of *Tremellomycetes.* 24
Figure 3.2. Phylogenetic analysis of CYPs from the species of *Tremellomycetes.* 26
Figure 3.3. The CYP family-level comparative analysis in the species of *Tremellomycetes.* 28
Figure 3.4. Heat map representing the presence or absence of cytochrome P450 families in 23 species of Tremellomycetes. 29
Figure 3.5. CYP diversity percentage analysis in *Tremellomycetes.* 30
Figure 4.1. Phylogenetic analysis of CYP51s from *Tremellomycetes.* 39
Figure 4.2. Superimposition of 3D structural models of 12 CYP51s from *Tremellomycetes.* 42
Figure 4.3. Active site cavities of 12 CYP51s from *Tremellomycetes.* 43
Figure 4.4. Graphical overview of docking results at online web application. 45
Figure 4.5. Graphical representation of results of hierarchical clustering. 50
Appendix B: Analysis of active site cavity amino acids in Tremellomycetes CYP51s 74

LIST OF TABLES

Table 2.1. Taxonomic hierarchy of cryptococcal species. 6
Table 2.2. WHO 2018 guidelines for management of cryptococcal disease in HIV-infected adults 13
Table 3.1. Some species of *Tremellomycetes* and their well-known characteristics.... 18
Table 3.2. Species of Tremellomycetes used in the study 21
Table 4.1 Information on CYP51s from *Tremellomycetes* used for structural analysis in this study. 34
Table 4.2. Summary statistics of CYP51s from Tremellomycetes alignments with CYP51 from Saccharomyces cerevisiae (PDB code: 4LXJ) 40
Table 4.3. Analysis of *Tremellomycetes* CYP51s and ligand complexes. 47
Table 4.4. Analysis of Tremellomycetes CYP51s amino acids interacting with different ligands. 51
Structural alignment of *Tremellomycetes* CYP51s with template 4LXJ 72

CHAPTER 1

Introduction

1.1 Introduction

Cryptococcus is the most common cause of meningitis in adults living with human immunodeficiency virus (HIV) in sub-Saharan Africa (Rajasingham *et al.,* 2017, Williamson *et al.,* 2017). Central nervous system (CNS) cryptococcosis occurs in an estimated 1 million people in the world each year, most of whom are HIV positive and reside in sub-Saharan Africa (Rajasingham *et al.,* 2017, Williamson *et al.,* 2017). An estimated 600 000 deaths due to cryptococcosis occur among these patients annually, an estimate that ranks fourth among causes of death attributable to infection in sub-Saharan Africa (Rajasingham *et al.,* 2017, Williamson *et al.,* 2017). There are over 30 different *Cryptococcus* species (spp.), with *C. neoformans* and *C. gattii* being the two primary human pathogens responsible for the fungal disease (Kwon-Chung *et al.,* 2014). Cryptococcal species have been found to develop resistance to the currently available azole drug fluconazole used in the treatment of these fungal infections (Hsueh *et al.,* 2005, Sar *et al.,* 2004, Pan *et al.,* 2012, Wang *et al.,* 2012, Chen *et al.,* 2015, Smith *et al.,* 2015). Azole compounds exert their anti-fungal activity by inhibiting one of the key enzyme cytochrome P450 monooxygenases (CYP/P450), CYP51, involved in the synthesis of fungal membrane ergosterol (Daum *et al.,* 1998, Kelly *et al.,* 1999). CYP51, also known as sterol 14α-demethylase, is a highly conserved CYP across the phyla and performs a key enzymatic reaction that involves stereo-selective three-step oxidative removal of the 14α-methyl group from the sterol (Lepesheva *et al.,* 2018). Because of this important enzymatic reaction this enzyme has become a drug target against fungal pathogens and protozoan parasites where azole compounds are in use as a drug in clinical practice targeting this enzyme (Zhang *et al.,* 2019, Lepesheva *et al.,* 2018, Choi *et al.,* 2014).

1.2 Problem statement

Non-susceptibility of cryptococcal species to fluconazole has been observed around the world and this has become a growing problem (Chen *et al.,* 2015, Wang *et al.,* 2012, Hsueh *et al.,* 2005, Sar *et al.,* 2004, Pan *et al.,* 2012, Smith *et al.,* 2015). The resistance is attributed to CYP51 (Lamb *et al.,* 1995, Rodero *et al.,* 2003, Sionov *et al.,* 2012, Kano *et al.,* 2017, Zhang *et al.,* 2019). Despite the importance of CYP51, to date, the CYP repertoire in cryptococcal species or in other species of *Tremellomycetes* has not been elucidated and CYP51 structure-activity studies with respect to different ligands (substrates or azole drugs) are scarcely

reported. Interestingly, quite a number of laboratory-based studies reported better performance of other azoles, especially itraconazole, against cryptococcal species compared to fluconazole (Armengou, 1996, Barchiesi *et al.,* 2000, Trilles *et al.,* 2004, Illnait-Zaragozi *et al.,* 2008, Thompson *et al.,* 2009, Hagen *et al.,* 2010, Pan *et al.,* 2012, Hoekstra *et al.,* 2014, Warrilow *et al.,* 2014, Lockhart *et al.,* 2016a, Nielsen *et al.,* 2017a).

This study is aimed at addressing this research gap by performing genome-wide data mining, annotation and phylogenetic analysis of P450s in cryptococcal species and other *Tremellomycetes* and carrying out extensive *in silico* structural analysis of CYP51, including construction of 3D models and identifying and mapping of amino acids interacting with substrates and azoles. The study is also aimed at proposing a hypothesis on the higher effectiveness of itraconazole over fluconazole against cryptococcal species.

1.3 Novelty

- This study is a first of its kind on genome data mining, annotation and phylogenetic analysis of CYPs in cryptococcal species.
- In addition, this study is a first of its kind on comparative modelling and analysis of 12 CYP51s' interactions with substrates and azole drugs in cryptococcal species.
- An advanced aspect of the study is that a web application for the visualisation of CYP51s' interactions with ligands was developed (http://bioshell.pl/azoledocking/).

1.4 Aim and objectives of the study

1.4.1 Aim of the study

Comparative analyses and gaining structural insights into the cytochrome P450 family CYP51 in the genus *Cryptococcus.*

1.4.2. Objectives of the study

- To carry out genome data mining, annotation and phylogenetic analysis of CYPs in cryptococcal species.
- To perform comparative analysis of CYPs between cryptococcal species and other *Tremellomycetes.*
- To make CYPs functional predictions.
- To perform *in silico* structural analysis of CYP51s from cryptococcal species.
- To assess the interaction of CYP51s with substrates and different azole drugs.

- To identify amino acids interacting with binding affinity with CYP51 P450s of cryptococcal species.
- To propose a CYP51-based hypothesis on the effectiveness of itraconazole over fluconazole in cryptococcal species.

1.5. Thesis overview

This thesis is divided into five chapters. Chapter 1 provides a brief overview of the work, problem statement, aims and objective and the novelty of the study. Chapter 2 consist of a literature review. A literature review is also provided as part of Chapters 3 and 4. Care has been taken not to repeat information. Chapter 3 details genome data mining, annotation and phylogenetic analysis of P450s in cryptococcal species and their comparison with other *Tremellomycetes*. Chapter 4 details the comparative structural analysis of CYP51s of cryptococcal species and their interactions with different azole drugs. A concise conclusion was presented as part of Chapters 3 and 4, but overall conclusions and future perspectives were presented as part of Chapter 5. The highlight of the study is that two research articles were generated and a web-based application to visualise CYP51s interactions were developed.

CHAPTER 2

Literature review

2.1. Cryptococcosis - history

Cryptococcosis is a globally distributed invasive fungal infection that is caused by cryptococcal species that pose a challenge therapeutically (May *et al.*, 2016, Rajasingham *et al.*, 2017, Williamson *et al.*, 2017). Recent studies have revealed that cryptococcosis is not transferred through human-human transmission; however, various mechanisms that facilitate cryptococci infection disseminate the pathogen, which eventually leads to the death of the human host in which it has been established (May *et al.*, 2016). This fungus was isolated in 1894 from a bone infection in a young woman by pathologist Otto Busse and physician Abraham Buschke as a *Saccharomyces-like* organism (Busse, 1894). In the same year, a similar organism was isolated from fermenting peach juice and named *Saccharomyces neoformans* by Francesco Sanfelice (Busse, 1894), Subsequently, in 1901, Jean-Paul Vuillemin renamed the organism *Cryptococcus neoformans* because it did not produce ascospores (Barnett, 2010). Cryptococcosis was first reported in man as a result of direct contact with contaminated droppings of pigeon (Littman, 1959). Detailed historical events on this fungus have been documented in the form of a review article (Srikanta *et al.*, 2014).

Cryptococcosis became recognised as a major health threat with the onset of the acquired immunodeficiency syndrome (AIDS) pandemic in the 1980s, when this fungal infection became a common AIDS-defining illness in patients with greatly reduced T cell function (May *et al.*, 2016). Although cryptococcosis is predominantly a disease of immunocompromised patients, a recent outbreak of cryptococcosis in otherwise healthy individuals in North America and Canada (now known as the Pacific Northwest outbreak) has focused attention on the capacity of some lineages of the fungus to act as primary pathogens (Rajasingham *et al.*, 2017).

2.2. Cryptococcal species

Cryptococcal species are normally found in places where there are soil and trees and consequently the urine and faeces of avian species. Since its identification, cryptococcosis has been attributed to a single fungal species, *C. neoformans*. However, improved molecular methods led to another variety of the pathogen, *C. gattii*, being classified as a distinct species. Other pathogens include *Cryptococcus albidus, Cryptococcus laurentii, Cryptococcus luteolus,*

Cryptococcus uniguttulatus, *Cryptococcus curvatus (*former *Candida curvata)*, and *Cryptococcus humicola* (former *Candida humicola* and *Cryptococcus humicolus)*, while others are non-pathogens (Khawcharoenporn *et al.*, 2007).

Among the cryptococcal species, *C. neoformans* is the most prominent medically important cryptococcal species (Kwon-Chung *et al.,* 2014). It is a yeast-like fungus, which causes life-threatening infections in patients living with AIDS (Kwon-Chung *et al.,* 2014). Among other infections, it causes fungal meningoencephalitis, which claims 600 000 lives every year globally, with most cases observed in Africa (Rajasingham *et al.,* 2017; Williamson *et al.,* 2017).

C. gattii, the most common fungal pathogen, causes infection of the CNS (D'Souza *et al.,* 2011; Kwon-Chung *et al.,* 2014) and is similar to *C. neoformans* in that it has the same virulence determinants, such as the polysaccharide capsule, the ability to grow at 37°C, laccase activity, phospholipase B, urease, superoxide dismutase and trehalose. *C. gattii* also has some gene products, including the capsule and phospholipase B (Chen *et al.*, 2014). It was furthermore described that *C. gattii* has an exception in the laccase gene (LAC1), created in the virulent VGIIa strain (R265) from the Vancouver Island outbreak in the mouse inhalation model. It has been reported that this pathogen affects immunocompetent patients, leading to major neurological morbidity, and has the potential to cause fatal infections such as meningoencephalitis and pneumonia (Trilles *et al.*, 2012). *C. neoformans* and *C. gattii* have been identified as the major pathogens in the genus *Cryptococcus* (Kwon-Chung *et al.*, 2014).

The originally identified tropical pathogen, *C. gattii*, is mostly common in Asia, Australia, Africa and South America. The rise in global movement of *C. gattii* has caused a shift in the habitat of this species from one place to another, thus causing it to become endemic in other parts of continents (Datta *et al.*, 2009). Litvintseva and co-workers (Litvintseva *et al.*, 2005) reported *C. gattii* as the cause of more than 13% of cryptococcal disease in HIV-positive humans in the sub-Saharan region; more than half of these patients die as a result of this disease.

C. grubii is an opportunistic pathogen that attacks patients with a compromised immune system, caused by human immunodeficiency virus (HIV) infection, corticosteroid therapy, haematological malignancies and solid-organ transplantation (Vechi *et al.*, 2019). Other cryptococcal species are identified as non-pathogenic. Nonetheless, *C. albidus*, *C. laurentii*, *C. luteolus*, *C. uniguttulatus*, *C. curvatus*, and *C. humicola* are characterised as opportunistic pathogens (Bernal-Martinez *et al.*, 2010).

C. albidus is the second most common pathogen in the genus *Cryptococcus.* Its main route of entry is *via* inhalation. It has been isolated from the eye and cutaneous *C. laurentii* infections because it disseminated disease with invasion of the CNS, fungemias in neonates and cancer patients, oropharyngeal infections in leukaemia patients, pneumonia and pleural effusion in patients living with AID, as well as lesions from disseminated infections in patients with lymphoma or leukaemia (Bernal-Martinez *et al.,* 2010).

C. curvatus has the ability to cause myeloradiculitis in AIDS patients (Bernal-Martinez *et al.,* 2010). *C. humicolus* is closely related to *C. curvatus* and *Trichosporon* sp. It is capable of causing infections of the CNS in HIV patients, and has been involved in systemic infections in patients suffering from cancer and other predisposing diseases (Bernal-Martinez *et al.,* 2010).

2.3. Taxonomic hierarchy and classification of Cryptococcus

Cryptococcal species' taxonomic hierarchy based on The Integrated Taxonomic Information System (ITIS) (https://www.itis.gov/) is shown in Table 2.1.

Table 2.1. Taxonomic hierarchy of cryptococcal species.

Kingdom	Fungi
Sub-kingdom	Dikarya
Division	Basidiomycota
Subdivision	Agaricomycotina
Class	Tremellomycetes
Order	Tremellales
Family	Tremellaceae
Genus	Cryptococcus

C. neoformans exists in two varieties consisting of five serotypes, *C. neoformans* var. *neoformans* (serotypes A, D and AD) and *C. neoformans* var. *gattii* (serotypes B and C) (Bennett *et al.,* 1977, Dromer *et al.,* 1993). Nine molecular types of Cryptococcus species were reported, with five under *C. neoformans* (VNI, VNII, VNB, VNIII, and VNIV) and four under *C. gatti* (VGI, VGII, VGIII, and VGIV) (Cogliati, 2013, Engelthaler and Meyer, 2017). A recent study suggested that *C. gatti* molecular types dominate in different parts of the world (Figure 2.1).

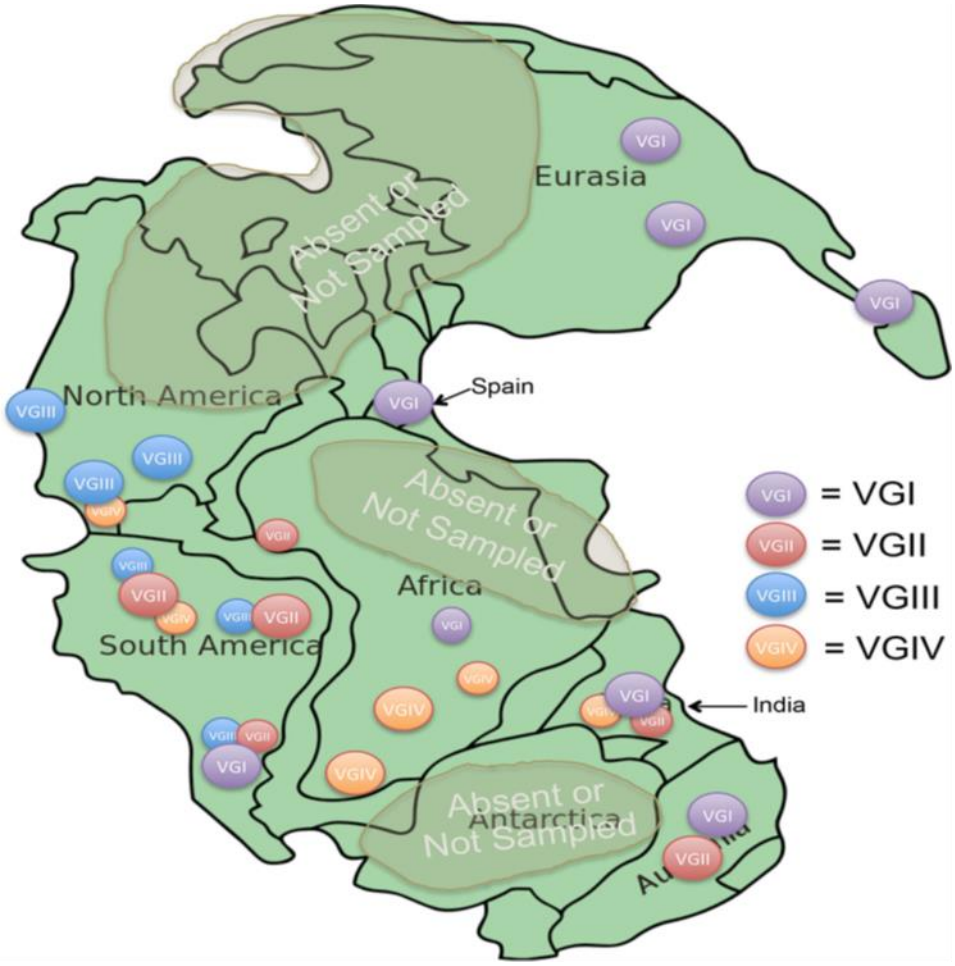


Figure 2.1. Schematic diagram representing *Cryptococcus gatti* molecular types' dominance in different parts of the world (Engelthaler and Meyer, 2017).

2.4. Cryptococcus habitat

C. neoformans is globally distributed and can be found in droppings of birds, especially pigeon droppings in cosmopolitan settings (Litvintseva and Mitchell, 2012). On the other hand, *C. gattii* is geographically distributed and can be found in tropical and sub-tropical regions where eucalyptus tree species are found (Figure 2.2.). Sorrell and co-worker (Sorrell *et al.,* 1996) reported that *C. neoformans* var. *gattii* was found consistently in eucalyptus trees. This species was also reported to be isolated from eucalyptus trees in India (Chakrabarti *et al.,* 1996). The exportation of eucalyptus seeds from Australia to other countries, especially in the sub-tropics, has also been reported to contribute to the dispersal of these organisms.

The pattern of distribution of eucalyptus and the infection that is caused owing to the association of this tree species with pigeons and other birds that are reservoirs for cryptococcal species (Figure 2.2) has been observed (Kielstein *et al.*, 2000). The recent presence of this species in the Pacific Northwest region of North America, a temperate rainforest region, is giving new insight into the shift in habitat of *C. gattii*, indicating that this species is not only found in the tropics, but also in temperate regions (Figure 2.1). *C. gattii* has also been reported to survive in lesions of succulent plants that can harbour these organisms, which suggests that the invasive insects that cause the lesions may be the mode of transmission for *C. gattii* (Loperena-Alvarez *et al.*, 2010). *C. neoformans* and *C. gattii* can survive and replicate in free-living amoebae and soil nematodes, and it is possible that these alternative hosts may play an important role in determining the distribution and virulence of different cryptococcal lineages around the world (Figure 2.2).

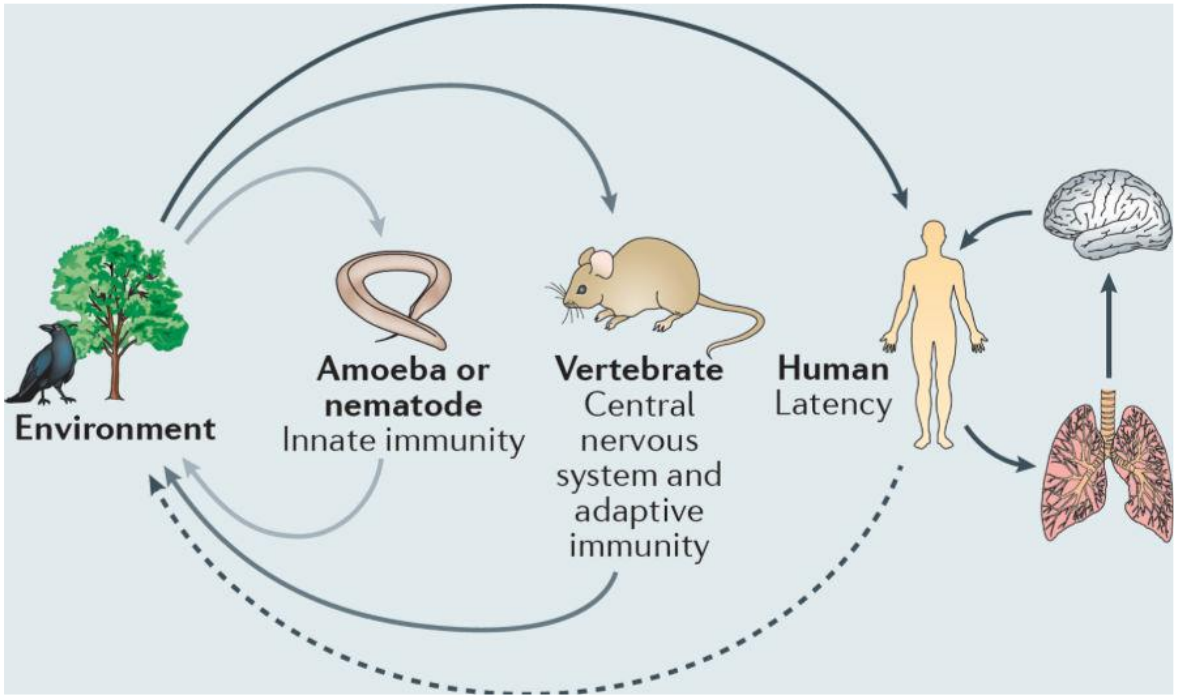


Figure 2.2. Habitat and transmission of C*ryptococci* (May *et al.*, 2016).

2.5. Morphology of Cryptococcus cells

Cryptococci originate from yeast cells and maintain this form. Studies have confirmed that they are round in shape with a diameter of 5–7 μM. Their cell shapes, characteristics and structure tend to differ from one host to another, even within a clonal infection. The survival and

pathogenesis of cryptococcal cells in individual hosts are also influenced by the age of the cryptococcal cells (Bouklas *et al.*, 2013). The founder cell, which is the older cell, is associated with the initial infection in a host; these are resistant to anti-fungal drugs and have the capability to resist phagocytosis. Accumulation of founder cells in the brain and an increase in resistance to phagocyte killing are influenced by changes in the cell wall of cryptococcal cells (Bouklas *et al.*, 2013).

Although a small cryptococcal cell has also been reported, the thickened cell wall present in such small cryptococcal cells enables the cell to adapt to growth in macrophages (Alanio *et al.*, 2015, Feldmesser *et al.*, 2000). This cell can also grow as either pseudo-hyphae or hyphae; however, these morphologies are not present in human infection, unlike other pathogenic fungi. Titan cells have been described as the best typical characterised form of *Cryptococcus* cells. Amazingly, these titan cells are larger than 12 μM in diameter (excluding the capsule), while polyploidy has been described to have a thickened cell wall and highly cross-linked capsules (Okagaki *et al.*, 2010, Zaragoza *et al.*, 2009). A master regulator Znf2 (transcription factor) when over-expressed was reported to trigger the transition of yeast to hyphal growth, which explained the reason for the filamentous morphologies noticeable in mammalian infections.

Vomocytosis is another strategy by which cryptococci escape from phagocyte. They do this by inducing the fusion of the phagosomal membrane with the plasma membrane, causing the elimination of the fungi from phagocyte. After the dissemination of the cryptococci cell into the lung, it encounters a phagocyte-like alveolar macrophage (dendritic cell). However, the cryptococci cells escape being killed by these cells because of their exposure to environmental amoebae and other virulence factors such as melanisation, capsule synthesis and urease secretion, which help in protecting these cells from the harsh environment within phagocytic cells. They do thus do by neutralizing the pH and reactive oxygen species (Ma *et al.*, 2009). Cryptococci have adapted these strategies in facilitating their growth and persistence and regulating the expulsion of the fungi within amoebae, which assist their survival in phagocytes. One interesting way by which a cryptococci cell effects its adaptive immune system is the ability to lie dormant without perturbing the host, thereby preventing total fungus clearance during treatment and causing latent infection.

2.6 Mode of transfer and onset of infection

Infection may be spread to humans through contact with pigeon droppings or unwashed raw fruit (Figure 2.2.). The disease is transmitted by inhalation of a particle that has been infected either from desiccated yeast or spores causing a systemic infection; this could eventually lead to primary pulmonary infection (Figure 2.2) (May *et al.,* 2016). The extent of the fungal infection depends on the immune system of the hostl; for immunocompetent hosts the fungal cells either cause an asymptomatic latent infection or the fungal cell cleared by the immune system, while in immunosuppressed patients they can cause pneumonia; further dissemination to other tissues, especially the CNS, can cause serious infection of the brain tissues and meninges (Gaona-Flores, 2013) (Figure 2.3). Undiagnosed HIV patients are usually identified when they show symptoms of cryptococcal meningoencephalitis (Perfect *et al.,* 2010a). The nature of the virulence ability of *C. neoformans* has made it a suitable model fungal pathogen, which was used in studying iron transport and homeostasis; the deprivation of iron in *C. neoformans* affects not only the synthesis of melanin, but also the formation of the polysaccharide capsule (Choi *et al.*, 2012).

In order for cryptococcal cells to establish an infection in a host, it passes through the lungs; it is then recognised by patrolling phagocytes (Figure 2.3). However, the cryptococcal cells can make use of the anti-phagocytic properties of the fungal capsule in growing to larger titan cells, which can facilitate its survival and persistence in phagocytes (Figure 2.3). The compromised immune system in a host can necessitate the intracellular proliferation of cryptococcal cells (May *et al,* 2017).

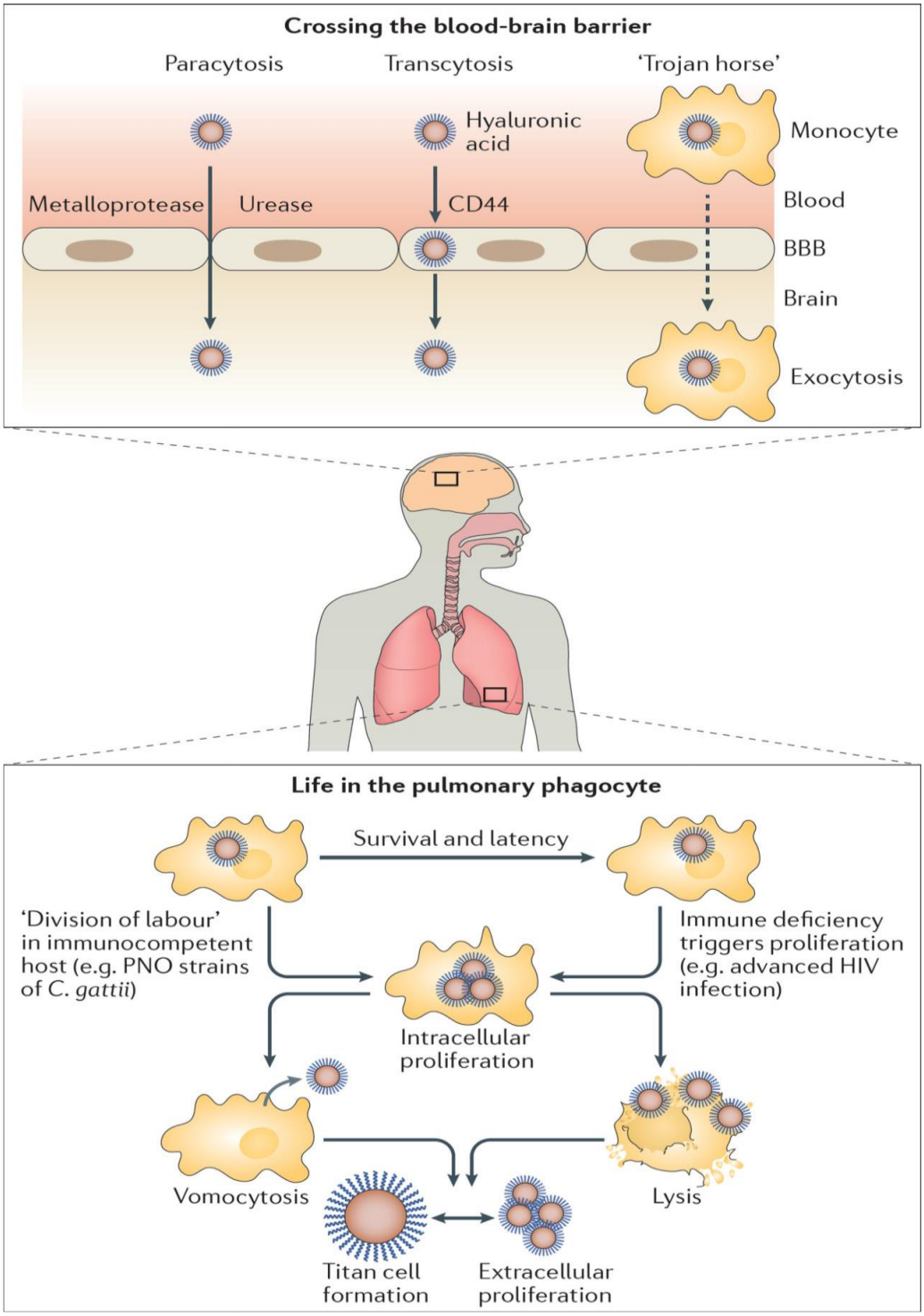


Figure 2.3. Schematic diagram representing infection establishment of cryptococcal cells in an immunocompromised patient (May *et al.*, 2016).

2.7. Epidemiology

Epidemiology statistics have shown that approximately 600 000 people die as a result of this disease every year, 15% of HIV-related deaths globally and an estimated up to 223 100 cases of cryptococcosis results in 181 000 deaths each year among people living with HIV (May *et*

al., 2016, Rajasingham *et al.*, 2017). The number of cryptococcosis cases reported worldwide in the 1950s was only around 300; at the time, the disease was noted to be associated with reticuloendothelial cancers. However, this number has increased drastically in recent years with the rise in HIV/AIDS cases and other immune-compromised states. New cases of cryptococcal meningitis was reported in 2006. These were *C. neoformans* infections, which occurred in sub-Saharan Africa and other developing countries where treatment is limited by infrastructure and cost; poverty and malnutrition remain great challenges (Loyse *et al.*, 2013, WHO, 2018). Various factors have contributed to the rise in the occurrence of cryptococcosis, which include transplant immunology, solid organ transplantations, surgical techniques used during transplantations and biological immunotherapeutics (Snydman *et al.*, 2008). Figure 2.4 shows the prevalence of cryptococcosis around the world.

Figure 2.4. Prevalence of cryptococcosis in the world (May *et al.*, 2016).

2.8. Cryptococcosis and future therapies for cryptococcosis

One of the challenges faced by medical mycology is finding an effective drug that can be used in the treatment of cryptococcal infections. Pharmacology management of cryptococcal infections usually consists of primary therapy with amphotericin B, with or without flucytosine (WHO, 2018). Below, Table 2.2 shows the current recommended treatment for cryptococcosis (WHO, 2018). Amphotericin B when administered intravenously combined with flucytosine clears the cerebrospinal fluid through its fungicidal activity. However, it has been established clinically that amphotericin B is toxic. This procedure requires medical infrastructure, which is not accessible in regions with high HIV prevalence because of limited resources (Loyse *et al.*, 2013, WHO, 2018). Flucytosine is also expensive and not accessible to most patients in regions with limited resources. Fluconazole, a fungistatic, is readily available and used in the treatment of many patients where amphotericin B and flucytosine are not accessible. Fluconazole, which can be taken orally, is confirmed to be cheap, safe, and readily available. However, this drug has lower efficacy than amphotericin B, does not clear the cerebrospinal fluid and has been reported to be associated with high mortality (Bicanic *et al.*, 2009).

Table 2.2. WHO 2018 guidelines for management of cryptococcal disease in HIV-infected adults (WHO, 2018).

	Two-week induction therapy		Consolidation therapy Week 3–10	Maintenance (or secondary prophylaxis) After week 10
	1st week	2nd week		
Preferred regimen	Amphotericin B deoxycholate (1.0 mg/kg/ day) + flucytosine (100 mg/kg/day, QDS)	Fluconazole (1200 mg daily)	Fluconazole (800 mg daily)	Fluconazole (200 mg daily) Until patient is adherent to ART and antifungal maintenance treatment for at least 1 year and has either

				a CD4 count ≥100 cells/ml and a fully suppressed VL or a CD4 count ≥200 cells/ml
Alternative regimens: depending on drug availability	Fluconazole (1200 mg daily) + flucytosine (100 mg/kg/day, QDS)			
	Amphotericin B deoxycholate (1.0 mg/kg/day) + fluconazole (1200 mg daily)			
In all cases	Monitoring and management of raised ICP are recommended Routine use of adjunctive corticosteroid therapy is not recommended			
Initiation of ART	ART initiation should be deferred by 4–6 weeks from the initiation of antifungal treatment			

Abbreviations: **ART**, antiretroviral therapy; **ICP**, intracranial pressure; **QDS**, divided into four doses per day; **VL**, viral load.

Recently, Reis and co-workers also confirmed an alternative plant product, fisetin (3'3/ 4'7 tetrahydroxyflavone), which is commonly found in the resin of *Hymenaea courbaril* (Touil *et al.,* 2011). The plant has been confirmed to possess antimicrobial activities and anti-inflammatory properties; it also has low toxicity compared to other antifungal drugs used in the treatment of cryptococcus. Fisetin has been shown to reduce the metabolic activity of the active cells in *C. gattii* during biosynthesis of ergosterol (Reis *et al.,* 2016). There is a need for more clinical trials of fisetin to evaluate its efficacy compared with other available antifungal drugs. VT-1129 was recently discovered as a new orally available drug; it acts by using tetrazole to bind to the active-site heme iron (Hoekstra *et al.,* 2014). VT- 1129 was also reported to be highly potent against these fungi and has shown low drug-drug interaction with human enzymes, which makes it a suitable future antifungal drug (Lockhart *et al.,* 2016a, Lockhart *et al.,* 2016b, Warrilow *et al.,* 2016, Nielsen *et al.,* 2017a, Nielsen *et al.,* 2017b). Figure 2.5 shows

key current and potential therapeutic targets and examples of anti-fungal drugs acting at each site.

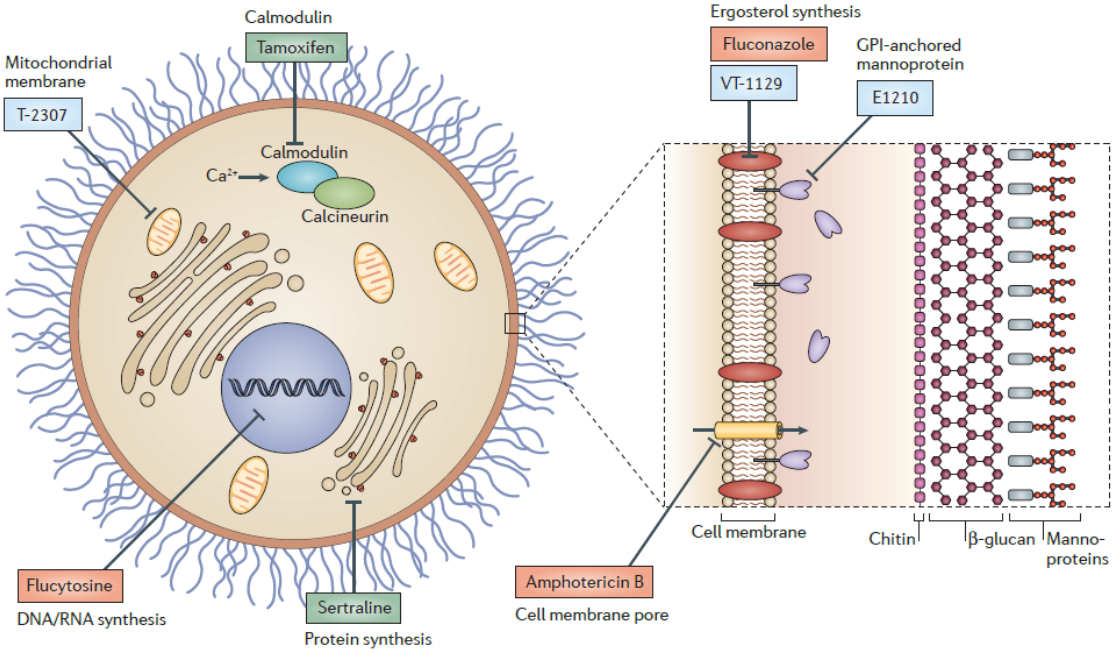


Figure 2.5. Current and future therapies for cryptococcosis (May *et al.*, 2016).

2.9. Cytochrome P450 monooxygenase CYP51

Azole compounds (fluconazole and VT-1129) exert their anti-fungal activity by inhibiting a key enzyme cytochrome P450 monooxygenase (CYP/P450), CYP51, involved in the synthesis of fungal membrane ergosterol (Figure 2.6) (Daum *et al.*, 1998, Kelly *et al.*, 1999). CYP51, also known as sterol 14α-demethylase, is highly conserved across the phyla and performs a key enzymatic reaction that involves stereo-selective three-step oxidative removal of the 14α-methyl group from the sterol (Figure 2.6) (Lepesheva *et al.*, 2018). Because of this important enzymatic reaction this enzyme has become a drug target against fungal pathogens and protozoan parasites where azole compounds are in use as a drug in clinical practice targeting this enzyme (Zhang *et al.*, 2019, Lepesheva *et al.*, 2018, Choi *et al.*, 2014).

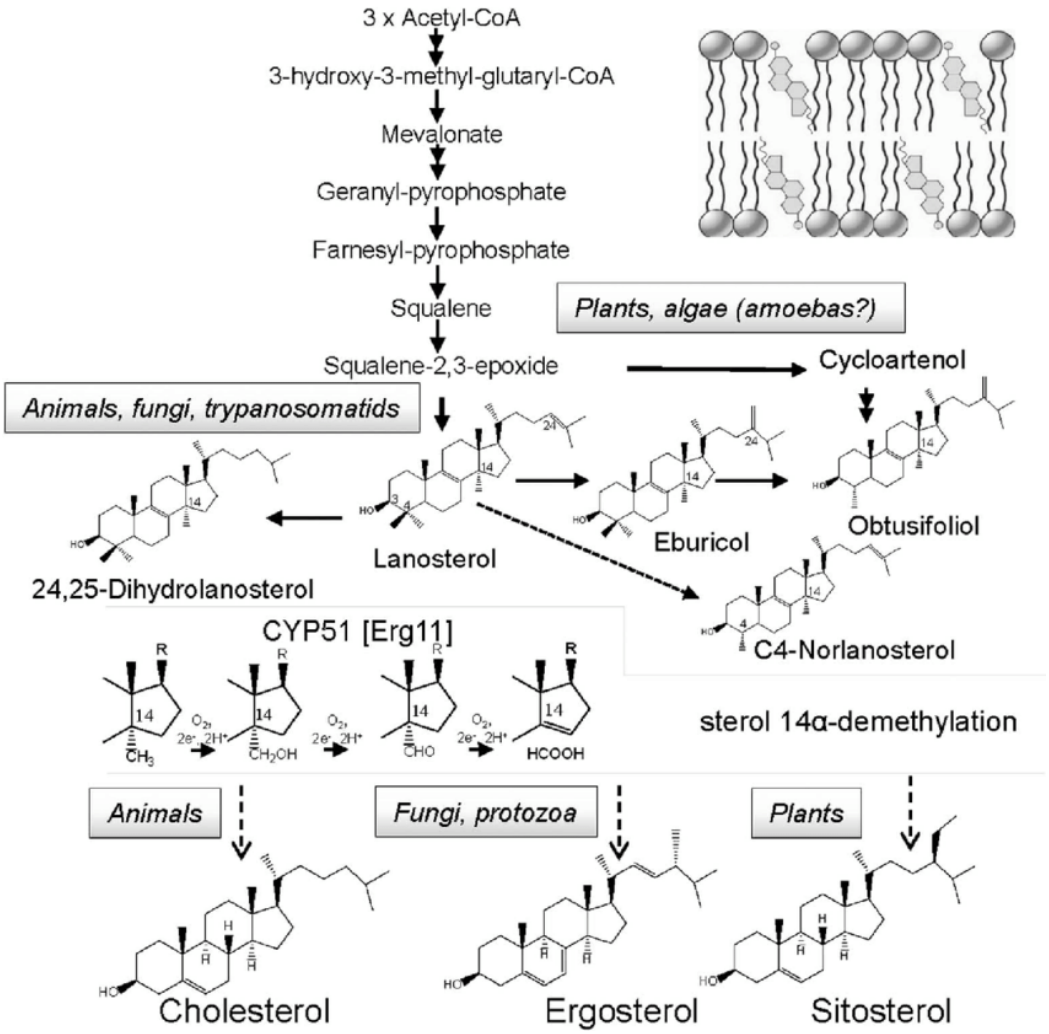


Figure 2.6. CYP51's role in synthesis of membrane sterols across the phyla (Lepesheva *et al.*, 2018).

CHAPTER 3

Distribution and diversity of cytochrome P450 monooxygenases in the genus *Cryptococcus* and their comparative analysis with other *Tremellomycetes*

(This chapter's work has been published in International Journal of Molecular Sciences : *Int. J. Mol. Sci.* **2019**, *20*(12), 2889; https://doi.org/10.3390/ijms20122889)

3.1 Introduction

Cryptococcosis is a fungal infectious disease ubiquitously distributed around the world (May *et al.,* 2016). Two fungal species, *Cryptococcus neoformans* and *C. gattii*, are the main infectious agents causing cryptococcal meningitis in both immunocompetent and immunocompromised humans (Loftus *et al.*, 2005, D'Souza *et al.,* 2011, Farrer *et al.,* 2015, May *et al.,* 2016). This disease is the major cause of morbidity and mortality among people living with advanced HIV and annually accounts for 15% of all HIV-related deaths globally (Rajasingham *et al.*, 2017, WHO, 2019). The burden of HIV-associated cryptococcal disease in Sub-Saharan Africa is alarming, as 73% of deaths in the world are reported in this region (Rajasingham *et al.*, 2017, WHO, 2019). Apart from these opportunistic pathogens, the genus *Cryptococcus* contains species with biotechnological potential (Table 3.1). Among the cryptococcal species, *C. amylolentus* is closely related to the pathogenic *C. neoformans* and is extensively used for comparative studies to identify the pathogenic traits in *C. neoformans* (Passer *et al.,* 2019).

The genus *Cryptococcus* belongs to *Tremellomycetes*, a fungal class in the subphylum *Agaricomycotina*, which contains organisms adapted to different niches and/or having different lifestyles (Table 3.1). Some of the organisms in this class are now regarded as emerging opportunistic human pathogens and some species are adapted to extreme ecological niches, such as cold regions (Table 3.1). Despite being fungi, *Naematella encephala* and *Tremella mesenterica* Fries exhibit fungal parasitism. The diverse lifestyles or characteristics of some species of *Tremellomycetes* are summarized in Table 3.1.

Table 3.1. Some species of *Tremellomycetes* and their well-known characteristics.

Species Name	Information	References
Cryptococcus neoformans	*C. neoformans* causes meningitis in immunocompromised and apparently in immunocompetent humans. This organism is considered a major opportunistic pathogen and a leading cause of mortality in patients infected with HIV.	(Loftus *et al.*, 2005)
Cryptococcus gattii	*C. gattii* causes respiratory (pneumonia) and neurological (meningoencephalitis) diseases in humans and animals and it can infect immunocompetent hosts.	(D'Souza *et al.*, 2011, Farrer *et al.*, 2015)
Cryptococcus terricola JCM 24523	*C. terricola* is oleaginous yeast and has been suggested as a candidate for the consolidated bioprocessing of hydrocarbon chemicals. It has the ability to accumulate unsaturated 18 carbon chain length fatty acids, with additional minor contributions of saturated 18 carbon and 16 carbon fatty acids.	(Kuo *et al.*, 2013, Tanimura *et al.*, 2014, Close *et al.*, 2016)
Cryptococcus curvatus	*C. curvatus* is oleaginous yeast capable of accumulating 18 carbon chain length fatty acids while growing on low or negative cost feed-stock. Thus, it is a potential candidate for the use in industrial fermentation processes. In a rare case *C. curvatus* was found to be involved in peritonitis associated with gastric lymphoma.	(Nowicka *et al.*, 2007, Kuo *et al.*, 2013, Close and Ojumu, 2016)
Naganishia vishniacii (formerly known as *Cryptococcus vishniacii*)	*N. vishniacii* is psychrophilic yeast adapted to live in extreme conditions, such as low-temperature oligotrophic deserts. It also has the ability to grow in a low-nutrient environment, without added vitamins.	(Vishniac and Hempfling, 1979, Kuo *et al.*, 2013, Schmidt *et al.*, 2017)

Cryptococcus wieringae	This is associated with pectin hydrolysis during the dew-wetting process of flax and found at the beginning of grape wine fermentation.	(Kuo *et al.*, 2013, Milanović *et al.*, 2013)
Cryptococcus amylolentus CBS 6273	*C. amylolentus* is the most closely known related species of the pathogenic *Cryptococcus* species complex, and is non-pathogenic.	(Mondo *et al.*, 2017, Passer *et al.*, 2019)
Kockovaella imperatae NRRL Y-17943	*K. imperatae* is a non-pathogenic fungus used in the analysis of widespread adenine N6-methylation of active genes in fungal species.	(Mondo *et al.*, 2017)
Naematella encephela UCDFST	It is a parasite of another fungus, *Stereum sanguinolentum*. This fungus' genome sequencing was carried out for the analysis of widespread adenine N6-methylation of active genes in fungal species.	(Mondo *et al.*, 2017)
Trichosporon asahii	Some species belonging to the genus *Trichosporon* are considered emerging opportunistic human pathogens and are the third most commonly isolated non-*Candida* yeasts from humans. They live in soil and are adapted to colonize the skin, gastrointestinal, respiratory and urinary tracts of humans. *T. asahii* is the most important species causing disseminated disease in immunocompromised patients, while the inhalation of *T. asahii* spores is the most important cause of summer-type hypersensitivity pneumonitis in healthy individuals. Some *Trichosporon* species have also emerged as rare but frequently fatal pathogens causing disseminated infections (trichosporonosis) in immunocompromised individuals and intensive care unit patients.	(Yang *et al.*, 2003, Yang *et al.*, 2012, Davies and Thornton, 2014)
Trichosporon oleaginosus IBC0246	*T. oleaginosus* is oleaginous yeast with the ability to accumulate lipids equivalent to biosynthetic kerosene, and thus is a biotechnologically valuable	(Kuo *et al.*, 2013, Kourist *et al.*, 2015)

	player for the generation of environmentally friendly (carbon-neutral) energy by converting agro-industrial waste to fuel (biodiesel).	
Tremella mesenterica Fries	. It is a parasite of crust fungus of the genus *Peniophora* and has a false appearance, as if it were growing on wood. Whereas in fact, it grows on the crust of fungal mycelium	(Floudas *et al.*, 2012)

In countering cryptococcosis, three classes of antifungal agents are available: polyenes (such as amphotericin B), azoles (such as fluconazole) and the pyrimidine analogue to flucytosine (May *et al.*, 2016). The gold standard induction treatment includes giving amphotericin B along with flucytosine (Perfect *et al.*, 2010b). However, this combination therapy has substantial side effects and the need for intravenous medications poses a problem, as these are not readily available in developing countries, which are most affected by cryptococcosis (Hanson *et al.*, 2017). To overcome this problem, a combination of fluconazole along with flucytosine has been recommended after initial therapy with amphotericin B and flucytosine (Perfect *et al.*, 2010b, May *et al.*, 2016).

Fluconazole binds to the fungal cytochrome P450 monooxygenase (CYP/P450) enzyme 14α-demethylase, named CYP51, which converts lanosterol into ergosterol, an essential component of the fungal cell membrane (Kelly and Kelly, 2013). *C. neoformans* also has *CYP51* and quite a number of studies have indicated that the development of drug resistance to fluconazole is due to the mutations in the *CYP51* gene and to the elevated levels of CYP51 in cryptococcal species (Lamb *et al.*, 1995, Rodero *et al.*, 2003, Revankar *et al.*, 2004, Sionov *et al.*, 2012, Ngamskulrungroj *et al.*, 2012). In addition to *C. neoformans*, drug resistance in other species of *Tremellomycetes* has also been reported owing to mutations in *CYP51* (Kushima *et al.*, 2012, Kushima *et al.*, 2017). Recent studies have demonstrated that the new anti-cryptococcosis drug named VT-1129 that is in the pipeline strongly binds and inhibits CYP51 of *C. neoformans* and *C. gattii* (Lockhart *et al.*, 2016a, Warrilow *et al.*, 2016, Nielsen *et al.*, 2017a).

Despite the importance of CYPs as drug targets, to date, the CYP repertoire in cryptococcal species or in other species of *Tremellomycetes* has not been elucidated. A few studies reported the CYP contingent of *C. neoformans* and *T. mesenterica* Fries with the purpose of comparing the CYP profiles with wood-degrading fungi (Floudas *et al.*, 2012, Syed *et al.*, 2014, Qhanya

et al., 2015). Thus, in this study we present a comparative analysis of CYPs in species of *Tremellomycetes*.

3.2. Methodology

3.2.1. Species and databases

Twenty-three species of *Tremellomycetes* were used in the study (Table 3.2). Different databases, such as NCBI (https://www.ncbi.nlm.nih.gov/), FungiDB (Basenko *et al.*, 2018) and Joint Genome Institute MycoCosm portal (Kuo *et al.*, 2013), were browsed for cryptococcal species genomes. All the species' genomes used in the study have been published and are available for public use, with the exception of two species, *C. wieringae* and *N. vishniacii*, and permission to use these two species' CYPs was obtained from the principal investigators listed at the respective species' databases at the Joint Genome Institute MycoCosm portal (Kuo *et al.*, 2013).

Table 3.2. Species of *Tremellomycetes* used in the study. The respective genome databases used for analysis of CYPs, along with the reference articles, are listed in the table. Abbreviations: NCBI, National Center for Biotechnology Information and JGI, Joint Genome Institute.

Species name	**Database**	**Reference**
Cryptococcus gattii VGII R265	NCBI	(D'Souza *et al.*, 2011, Farrer *et al.*, 2015)
Cryptococcus gattii NT-10		
Cryptococcus gattii VGII 99/473		
Cryptococcus gattii E566		
Cryptococcus gattii VGII 2001/935-1		
Cryptococcus gattii VGIV IND107		
Cryptococcus gattii VGII CBS 10090		
Cryptococcus gattii VGII 2001/935-1		
Cryptococcus gattii EJB2		
Cryptococcus gattii WM276	NCBI	(D'Souza *et al.*, 2011)
Cryptococcus terricola JCM 24523 v1.0	JGI	(Close *et al.*, 2016)
Cryptococcus curvatus ATCC 20509	JGI	(Close and Ojumu, 2016)
Naganishia vishniacii v1.0 (formerly known as *Cryptococcus vishniacii*)	JGI	(Kuo *et al.*, 2013)

Cryptococcus wieringae	JGI	(Kuo *et al.,* 2013)
Cryptococcus neoformans var. *neoformans* B-3501A	NCBI	(Loftus *et al.,* 2005)
Cryptococcus neoformans var. *neoformans* JEC21	NCBI	(Loftus *et al.,* 2005)
Cryptococcus amylolentus CBS 6273	JGI	(Sun *et al.,* 2017, Passer *et al.,* 2019)
Kockovaella imperatae NRRL Y-17943 v1.0	JGI	(Mondo *et al.,* 2017)
Naematella encephela UCDFST 68-887.2 v1.0	JGI	(Mondo *et al.,* 2017)
Trichosporon asahii var. *asahii* CBS 2479	JGI	(Yang *et al.,* 2003)
Trichosporon asahii var. *asahii* CBS 8904	JGI	(Yang *et al.,* 2012)
Trichosporon oleaginosus IBC0246 v1.0	JGI	(Kourist *et al.,* 2015)
Tremella mesenterica Fries v1.0	JGI	(Floudas *et al.,* 2012)
Cryptococcus neoformans var. *grubii* H99	NCBI	(Loftus *et al.,* 2005)

3.2.2. CYP mining and annotation

Genome mining for CYPs and subsequent annotation was carried out following the protocol described elsewhere (Kgosiemang *et al.,* 2014, Qhanya *et al.,* 2015, Matowane *et al.,* 2018, Ngwenya *et al.,* 2018). Briefly, proteomes cryptococcal species from different genome databases, listed in Section 3.1, were downloaded and subjected to an NCBI conserved domain search (Marchler-Bauer *et al.,* 2017) to classify the proteins into different subfamilies. The proteins grouped under a CYP superfamily were selected and subjected to blast analysis against fungal CYPs (Nelson, 2009) to identify homolog CYPs. Based on percentage identity with a named homolog CYP, the hit proteins were then assigned to different CYP families and CYP subfamilies, following the International P450 Nomenclature Committee rule, that is, >40% identity for a family and >55% identity for a subfamily (Nelson, 1998, Nelson, 2006). CYPs that had less than 40% identity with named fungal CYPs (Nelson, 2009) were assigned to a new family. For each species, CYPs from different databases were compared and duplicate CYPs were removed from the final CYP count.

3.2.3. Phylogenetic analysis of CYPs

Phylogenetic analysis of CYPs was carried out following the procedure described elsewhere (Qhanya *et al.,* 2015, Matowane *et al.,* 2018, Bamal *et al.,* 2018, Senate *et al.,* 2019). Briefly,

first, the protein sequences were aligned by MAFFT v6.864 (Katoh *et al.*, 2005), and embedded on the T-REX web server (Boc *et al.*, 2012). Then, the alignments were automatically subjected to tree inferring and optimization by the T-REX web server. Finally, the best-inferred trees were visualized and colored by iTOL (http://itol.embl.de/about.cgi) (Letunic and Bork, 2016).

3.2.4. Generation of CYP profile heat-maps

The presence or absence of CYPs in species of *Tremellomycetes* was shown with heat-maps generated using CYP family data following the method described elsewhere (Ngwenya *et al.*, 2018, Mthethwa *et al.*, 2018). The data were represented as −3 for gene absence (green) and 3 for gene presence (red). A tab-delimited file was imported into multi-experiment viewer (MeV) (Saeed *et al.*, 2003). Hierarchical clustering using a Euclidean distance metric was used to cluster the data. Twenty-three species of *Tremellomycetes* form the horizontal axis and 38 CYP families form the vertical axis.

3.2.5. CYP diversity percentage analysis

CYP diversity percentage analysis was carried out as described elsewhere (Sello *et al.*, 2015, Parvez *et al.*, 2016, Matowane *et al.*, 2018, Senate *et al.*, 2019). Briefly, the CYP diversity percentage in species of *Tremellomycetes* was measured as a percentage contribution of the number of CYP families in the total number of CYPs.

3.2.6. Functional prediction of CYPs

Literature was searched for characterized CYPs from species of *Tremellomycetes*, if any. Furthermore, functional prediction of CYPs was carried out based on the characterized homolog CYPs from different fungal organisms. CYP family level functional prediction was presented in the article.

3.3. Results and Discussion

3.3.1. Pathogenic cryptococcal species have few CYPs in their genomes

Genome-wide data mining of CYPs in 16 cryptococcal species revealed the presence of 112 CYPs in their genomes (Figure 3.1). *C. curvatus* and *C. terricola* have the highest number of CYPs (16 CYPs each), and *C. gattii* VGIV IND107 has the lowest number of CYPs (Figure 1). An interesting pattern was observed when comparing the CYP count among cryptococcal species. Almost 50% fewer CYPs were found in pathogenic cryptococcal species compared to non-pathogenic cryptococcal species (Figure 3.1). This suggests that adaptation to survive in a

host (mainly animals) that has a rich source of simple nutrients might have led to the loss of CYPs. The same phenomenon was observed in fungal species belonging to the subphylum *Saccharomycotina*, where species lost a considerable number of CYPs owing to their adaptation to simpler carbon sources (Kgosiemang *et al.*, 2014).

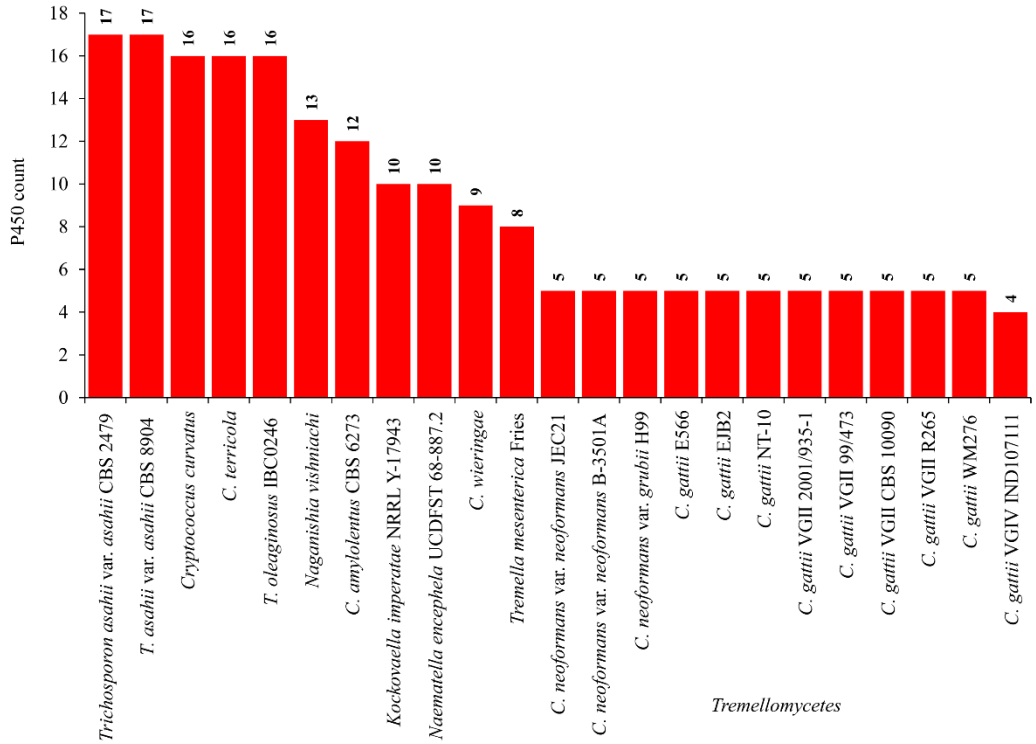


Figure 3.1. Comparative analysis of CYPs in the species of *Tremellomycetes.*

The comparison of cryptococcal species' CYP count with other species belonging to the same subphylum *Agaricomycotina*, especially the well-studied wood-degrading fungi, is not logical, since the wood-degrading species have quite a large number of CYPs in their genomes (Floudas *et al.*, 2012). As cryptococcal species fall under *Tremellomycetes*, in this study, a comprehensive comparative analysis of CYPs in *Tremellomycetes* was carried out (Figure 3.1). As shown in Figure 3.1, the comparison of CYPs among species of *Tremellomycetes* indicated that pathogenic cryptococcal species have a lower number of CYPs compared to other species of *Tremellomycetes*. Fungal parasites such as *T. mesenterica* Fries and *N. encephela* have eight and 10 CYPs in their genomes, somewhat lower than non-pathogens. It is interesting to note that the species belonging to the genus *Trichosporon* have the highest number of CYPs in their genomes, both pathogenic and non-pathogenic (Figure 3.1). It is well-known that most of the species belonging to this genus are considered commensals of the human skin and

gastrointestinal tract, and these species are now increasingly causing superficial and invasive diseases in immunocompromised individuals and intensive care unit patients (Davies and Thornton, 2014, Duarte-Oliveira *et al.*, 2017). This indicates that these organisms have a long way to go to adapt better, similar to the cryptococcal species, and thus, in this process they may lose CYPs as well.

3.3.2 New CYP families were found in *Tremellomycetes*

A total of 203 CYPs were found in 23 species of *Tremellomycetes* (Figure 3.2). Sixteen CYPs were found to be pseudo/false positives, as they lack one of the CYP characteristic motifs and/or short fragments. Thus, these CYPs were not included in the study. The annotation of CYPs as per International P450 Nomenclature Committee rules (Nelson, 1998, Nelson, 2006, Nelson, 2009) in combination with phylogenetic analysis (Figure 3.2) revealed that 203 *Tremellomycetes* CYPs could be grouped into 38 CYP families and 72 CYP subfamilies (Figure 3.2). Phylogenetic analysis of CYPs is critical in assigning the CYP family and subfamily for the CYPs that have a borderline percentage identity of around 40–41% (for a family) and 55–56% (for a subfamily) with the named fungal CYPs.

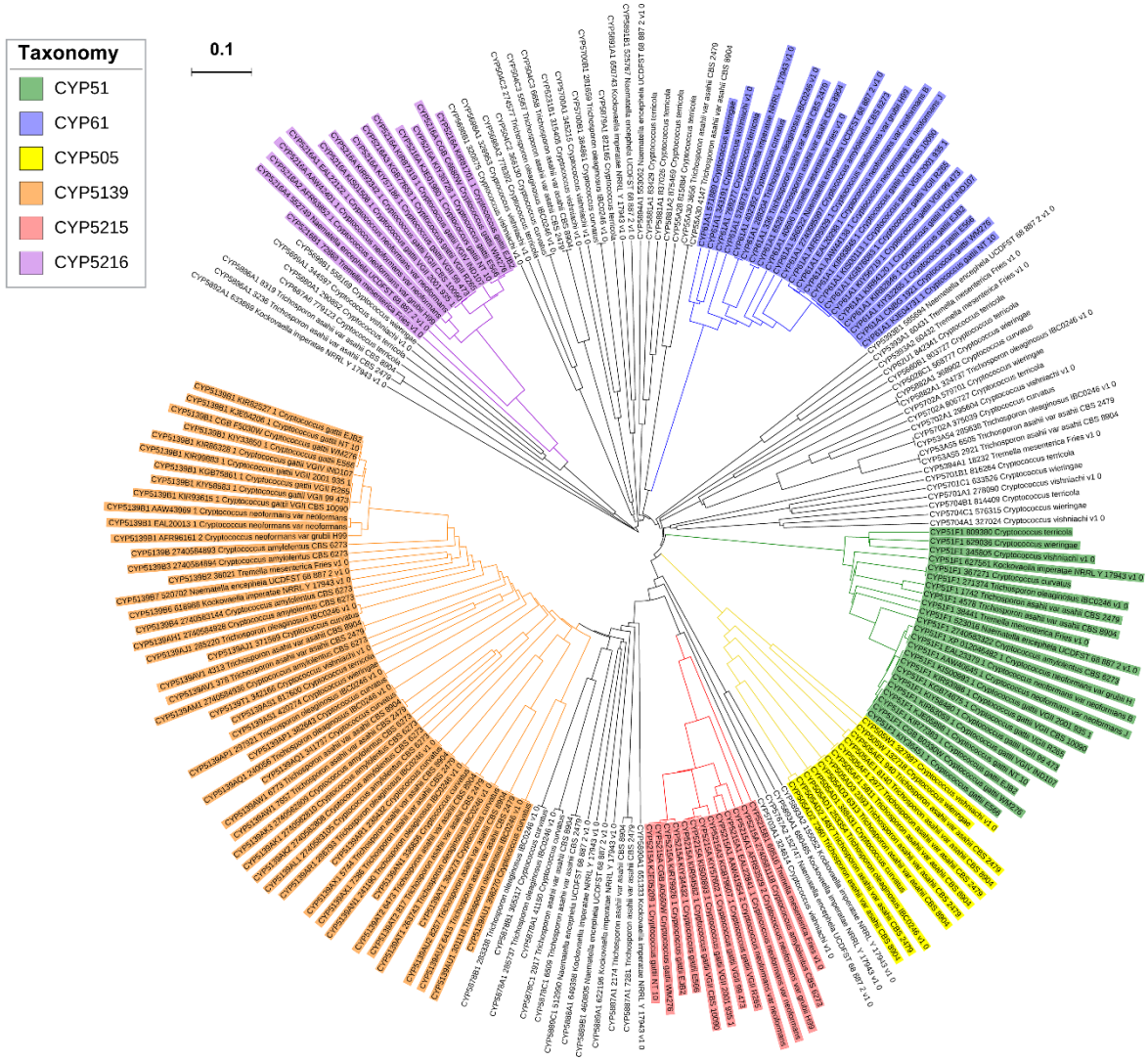


Figure 3.2. Phylogenetic analysis of CYPs from the species of *Tremellomycetes*. CYP families that are highly populated in species of *Tremellomycetes* are indicated in different colors.

A total of 23 new CYP families, named CYP5215, CYP5216, CYP5393, CYP5394, CYP5698-CYP5702, CYP5878-CYP5882, and CYP5886-CYP5894A1, were identified in species of *Tremellomycetes. Kockovaella imperatae* NRRL Y-17943 have the highest number of new CYP families (six CYP families: CYP5888-CYP5893) followed by *Naganishia vishniacii* (five new CYP families: CYP5698-CYP5702), *C. terricola* (three new CYP families: CYP5879-CYP5881), *T. mesenterica* Fries (two new CYP families: CYP5393 and CYP5394), *T. oleaginosus* IBC0246 (CYP5878 and CYP5882) and two new CYP families (CYP5886 and CYP5887) were found in *T. asahii* var. *asahii* CBS 2479 and *T. asahii* var. *asahii* CBS 8904. Three species of *Tremellomycetes* have only one new CYP family: *C. neoformans* var. *grubii*

H99 (CYP5215), *C. neoformans* var. *neoformans* B-3501A (CYP5216), and *N. encephela* UCDFST 68-887.2 (CYP5894).

3.3.3. Four CYP families are conserved in pathogenic cryptococcal species

CYP family-level comparative analysis revealed that among 38 CYP families, the CYP5139 family was found to be dominant in species of *Tremellomycetes* with 51 members, following the CYP51 and CYP61 families each with 23 members, the CYP5216 family with 14 members, the CYP5215 family with 13 members, and the CYP505 family with 12 members (Figure 3.3). Analysis of CYP family conservation revealed that three CYP families, namely CYP5139, CYP51, and CYP61, are conserved in all 23 species of *Tremellomycetes* (Figure 3.4). CYP family comparison among pathogenic cryptococcal species revealed conservation of two more CYP families, CYP5215 and CY5216, in all species except *C. gattii* VGIV IND107, which does not have CYP5215 (Figure 3.4). These two CYP families are also present in fungal parasites, *T. mesenterica* Fries (both families) and *N. encephela* UCDFST 68-887.2 (only CYP5216 family), and non-pathogenic *C. amylolentus* CBS 6273 (Figure 3.4). The CYP family CYP5231 found in *N. vishniacii* is also present in the fungal species, *Melampsora laricis-populina* and *Puccinia graminis,* belonging to the class *Pucciniomycotina,* where this family is bloomed in both species (Qhanya *et al.,* 2015). The presence of the CYP5126 family only in pathogenic or parasitic *Tremellomycetes* indicates that this CYP family might be playing a role in the adaptation of these organisms to their host. The analysis of CYP subfamilies revealed that the CYP5139 family has 17 CYP subfamilies, indicating the blooming of members in this family. The same was observed for quite a number of CYP families in other fungi (Qhanya *et al.,* 2015, Syed *et al.,* 2014).

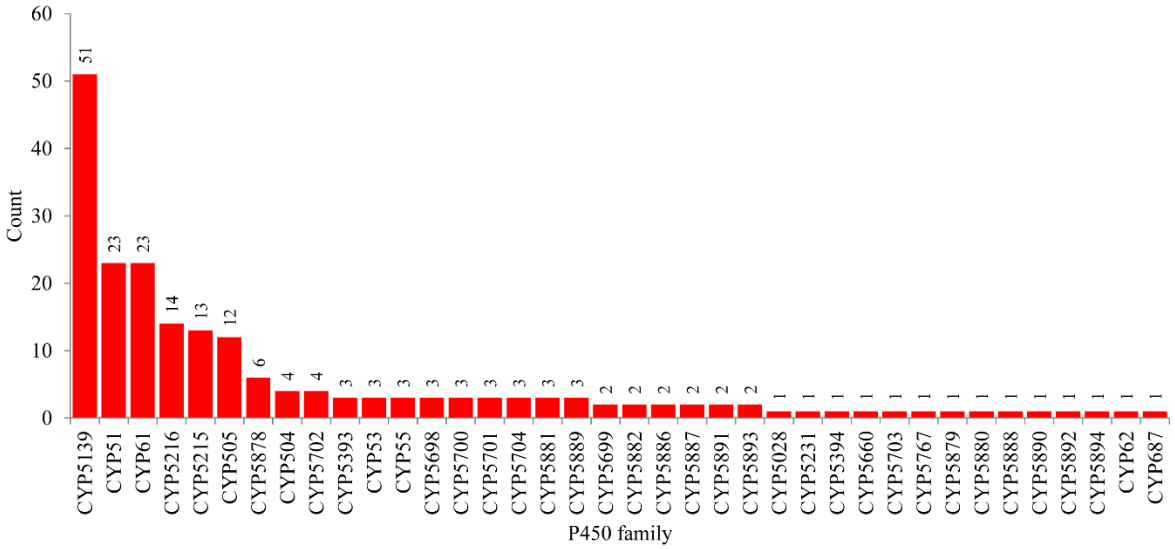


Figure 3.3. The CYP family-level comparative analysis in the species of *Tremellomycetes*. The numbers next to the family bar indicate the total number of CYPs.

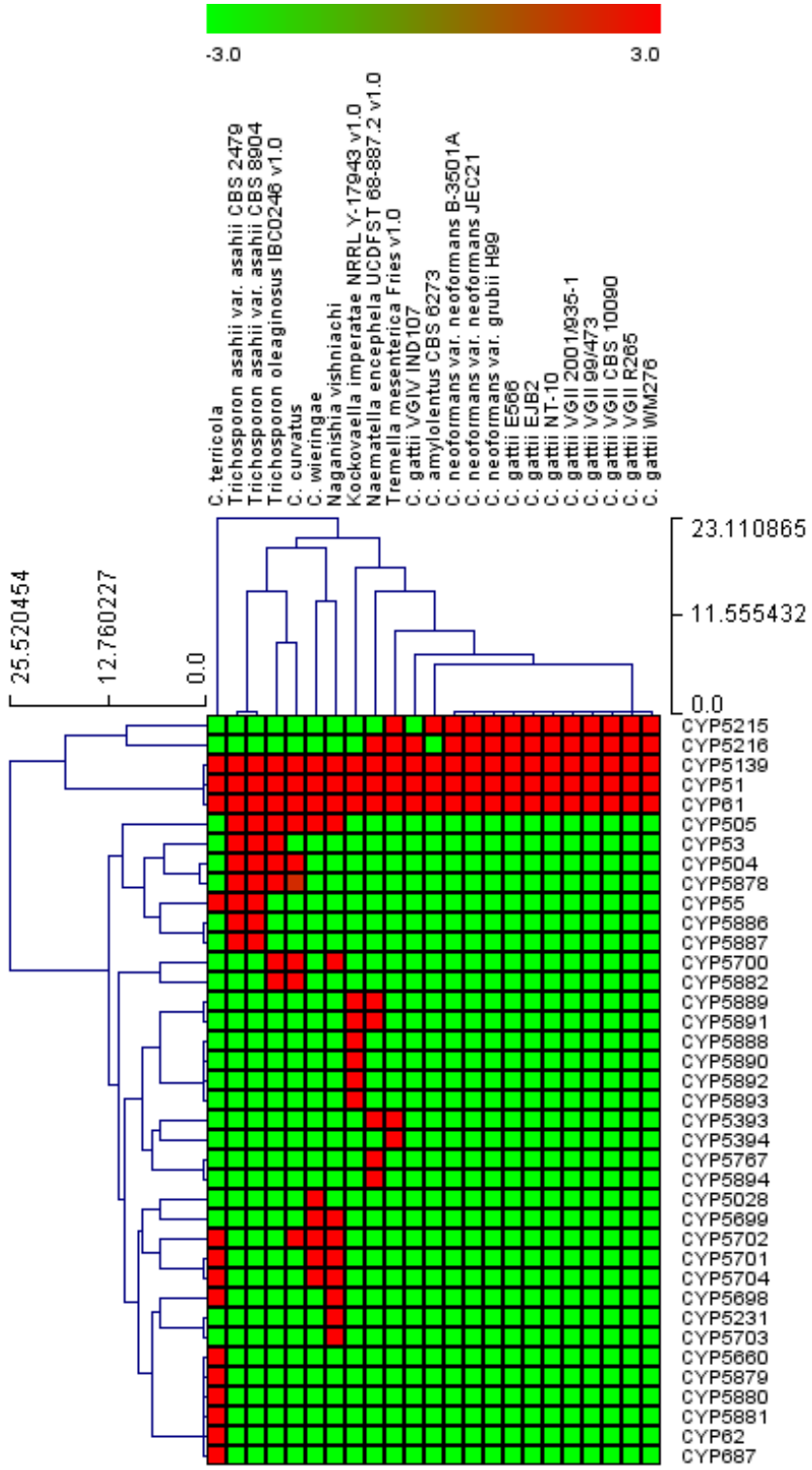


Figure 3.4. Heat map representing the presence or absence of cytochrome P450 families in 23 species of *Tremellomycetes*. The data have been represented as –3 for gene absence (green) and 3 for gene presence (red). Twenty-three species of *Tremellomycetes* form the horizontal axis and CYP families form the vertical axis.

3.3.4. Pathogenic cryptococcal species have the highest CYP diversity

CYP diversity analysis revealed that pathogenic cryptococcal species, along with non-pathogenic *C. wieringae* and the fungal parasite *N. encephela* UCDFST 68-887.2, have 100% CYP diversity in their genomes (Figure 3.5). *Tremellomycetes* such as *C. curvatus*, *C. amylolentus* CBS 6273 and *T. asahii* var. *asahii* strains had the lowest CYP diversity percentage. This is due to the blooming of CYP5139 members in their genome. The highest CYP diversity observed in pathogenic cryptococcal species is perfectly matched with species belonging to the fungal subphylum *Saccharomycotina* (Kgosiemang *et al.*, 2014). One commonality can be found between the species belonging to *Tremellomycetes* and *Saccharomycotina*: It can be assumed that some species of *Tremellomycetes* lost CYPs, compared to their counterparts, which may be due to the adaptation to use simple carbon sources present in the host, as observed for species of *Saccharomycotina*, where the loss of CYPs in response to the adaptation to use simpler carbon sources was observed (Kgosiemang *et al.*, 2014).

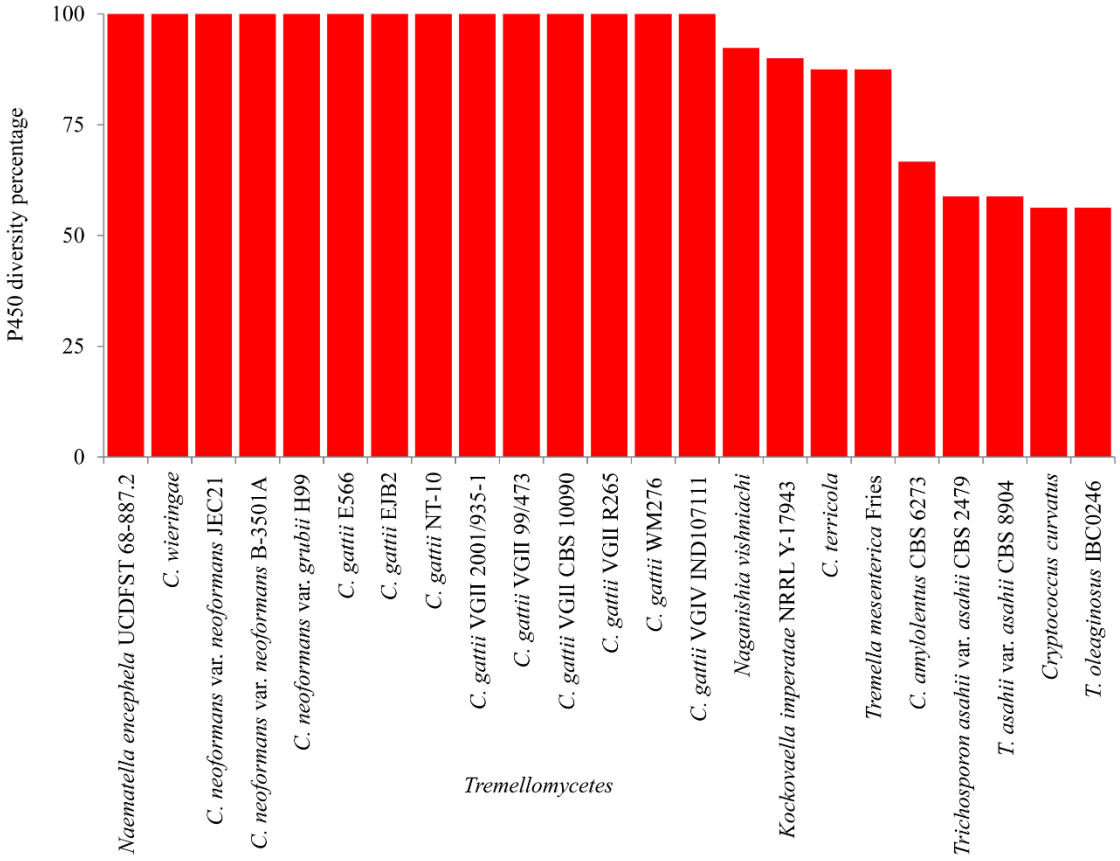


Figure 3.5. CYP diversity percentage analysis in *Tremellomycetes*.

3.3.5. Most of the CYPs from the species of *Tremellomycetes* are orphans with no known function

Among CYPs from the species of *Tremellomycetes,* CYP51F1 of *C. neoformans* has been shown to be involved in 14α-demethylation of lanosterol (Revankar *et al.*, 2004) and the *CYP51F1* gene was cloned from *T. asahii* ATCC MYA-1296 = OMU239 = TIMM4014 (Sionov *et al.*, 2012). Apart from CYP51F1, some of the CYPs' functions can be predicted based on characterized homolog CYPs. CYP61 family members are involved in membrane ergosterol biosynthesis where they catalyze C-22 sterol desaturase activity (Kelly *et al.*, 1997). CYP505 family members are involved in oxidation of fatty acids (Nakayama *et al.*, 1996). CYP504 family members are involved in conversion of phenyl acetate to 2-hydroxyphenylacetate (Mingot *et al.*, 1999). CYP53 family members are involved in detoxification of toxic molecules, including benzoate and its derived compounds (Faber *et al.*, 2001, Matsuzaki and Wariishi, 2005, Durairaj *et al.*, 2015). The primary function of CYP53 is the conversion of benzoate to *para*-hydroxy-benzoate. A study reported that CYP53 could be an alternative antifungal drug target in view of its critical role in fungal organisms (Jawallapersand *et al.*, 2014). It is interesting to note that this CYP family is only present in three species belonging to the genus *Trichosporon* (Figure 3.4). CYP55 family members are involved in the reduction of nitric oxide (NO) to nitrous oxide (N_2O) (Nakahara *et al.*, 1993, Shoun *et al.*, 2012). It is interesting to note that in addition to CYP53, CYP55 family members are also found in two species belonging to the genus *Trichosporon* (Figure 3.4). Apart from the CYP families listed above, the rest of the CYPs found in species of *Tremellomycetes* are orphans.

3.4 Conclusions

In this study, genome data mining, annotation and phylogenetic analysis of CYPs were carried out in cryptococcal species and compared with other *Tremellomycetes*. The study revealed the conservation of the CYP51 family in all species, indicating its important role in the synthesis of membrane ergosterol. For this reason, this CYP is a drug targeted especially by azole compounds in these species. CYP51s annotated in this study need to be analyzed for their structure and interaction analysis with different substrates and azole drugs.

CHAPTER 4

In silico structural modeling and analysis of interactions of cryptococcal species cytochrome P450 monooxygenases CYP51s with substrates and azoles

(This chapter has been submitted to Nature Publication Group Journal Scientific Reports: submission ID b2d39cce-8a1a-47a9-83c3-c907af353db5)

4.1. Introduction

Among fungal diseases cryptococcosis remains a significant cause of morbidity and mortality in immunocompromised people especially in sub-Saharan Africa (Rajasingham *et al.,* 2017). Cryptococcal species such as *Cryptococcus neoformans* and *C. gatti* are primarily responsible for this disease(Kwon-Chung *et al.,* 2014). Across the world, an average of 6% of HIV-infected people have contract this disease, of which 73% of cases were reported in sub-Saharan Africa, with a death rate of 15% (Rajasingham *et al.,* 2017). Treatment options for cryptococcosis are limited to only three drugs (amphotericin B, flucytosine and fluconazole), with few compounds reaching clinical trials (Coelho and Casadevall, 2016). The treatment involves use of an azole drug, fluconazole, as primary or consolidation and maintenance therapy, followed by induction therapy with intravenous amphotericin B with or without flucytosine (WHO, 2018). In developed countries patients are initially subjected to induction therapy and then treatment with fluconazole. However, in resource-limited countries such as low- and middle-income countries, treatment is solely based on fluconazole (WHO, 2018, Loyse *et al.,* 2013).

Non-susceptibility of cryptococcal species to fluconazole has been observed around the world and this has become a growing problem(Chen *et al.,* 2015, Wang *et al.,* 2012, Hsueh *et al.,* 2005, Sar *et al.,* 2004, Pan *et al.,* 2012, Smith *et al.,* 2015). Interestingly, quite a number of laboratory-based studies reported better performance of other azoles (itraconazole, voriconazole, posaconazole, and isavuconazole) including the new class of tetrazole compound, VT-1129, against cryptococcal species(Trilles *et al.,* 2004, Thompson *et al.,* 2009, Pan *et al.,* 2012, Barchiesi *et al.,* 2000, Illnait-Zaragozi *et al.,* 2008, Hagen *et al.,* 2010, Armengou, 1996, Hoekstra *et al.,* 2014, Warrilow *et al.,* 2014, Lockhart *et al.,* 2016a, Nielsen *et al.,* 2017a). Azole compounds exert their anti-fungal activity by inhibiting the enzyme cytochrome P450 monooxygenase (CYP/P450) CYP51 involved in the synthesis of fungal membrane ergosterol(Daum *et al.,* 1998, Kelly *et al.,* 1999). CYP51, also known as sterol 14*α*-

demethylase, is highly conserved across the phyla and stimulates a key enzymatic reaction that involves stereo-selective three-step oxidative removal of the 14α-methyl group from the sterol(Lepesheva *et al.,* 2018). Because of this important enzymatic reaction this enzyme has become a drug target against fungal pathogens and protozoan parasites where azole compounds are in use as drugs in clinical practice targeting this enzyme(Zhang *et al.,* 2019, Lepesheva *et al.,* 2018, Choi *et al.,* 2014). A search using the word "CYP51" at Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (Burley *et al.,* 2019) indicated the presence of 106 crystal structures (as of 30 August 2020), clearly indicating the enormous importance of this enzyme, especially its exploitation as a drug target (Lepesheva *et al.,* 2018, Kelly and Kelly, 2013).

Some progress has been made with unraveling the molecular mechanisms responsible for fluconazole non-susceptibility based on *CYP51* in cryptococcal species(Rodero *et al.,* 2003, Sionov *et al.,* 2012, Zhang *et al.,* 2019, Kano *et al.,* 2017, Lamb *et al.,* 1995). Over-expression of this gene either by sterol regulatory element-binding protein or by virtue of the presence of more than one copy of this gene due to genomic aneuploidy or the mutations in this gene has been to be responsible for fluconazole resistance(Rodero *et al.,* 2003, Sionov *et al.,* 2012, Zhang *et al.,* 2019, Kano *et al.,* 2017, Lamb *et al.,* 1995). A point mutation in CYP51 at G484S resulted in conferring resistance on fluconazole(Rodero *et al.,* 2003). Azole drug resistance patterns were changed by a single mutation Y145F that caused high fluconazole/voriconazole resistance but increased susceptibility to itraconazole and posaconazole(Sionov *et al.,* 2012). G344S mutation caused multi-drug resistance to fluconazole, itraconazole and voriconazole(Kano *et al.,* 2017). Detailed information on different mechanisms employed to deal with fluconazole resistance of cryptococcal species was recently reviewed (Zhang *et al.,* 2019).

The *CYP51* gene mutational studies mentioned above are based on sequencing the *CYP51* gene from the fluconazole non-susceptible strains and comparing its sequence with fluconazole-susceptible strains. To date, structure-activity relationship studies on CYP51 from cryptococcal species have not been reported, with the exception of one study published in 2009 reporting the first three-dimensional (3D) model of CYP51 from *C. neoformans*. Based on the *in silico* results, the authors suggested that G484S substitution changed the orientation of the heme-binding domain, leading to decreased catalytic activity and thus azole binding conferring the drug resistance (Sheng *et al.,* 2009). Recently, our laboratory performed genome-wide analysis and annotation of P450s in the fungal class *Tremellomycetes*, which included quite a

number of cryptococcal species as well (Akapo *et al.*, 2019). The study revealed the presence of the *CYP51* gene in all species of *Tremellomycetes*. Considering that structure-activity studies are scarcely reported on CYP51 of cryptococcal species and the availability of a large number of CYP51 sequences (Akapo *et al.*, 2019), this study is aimed at addressing this research gap by performing comparative modeling of CYP51s, which includes assessing their binding patterns to substrates and azoles. CYP51s from other *Tremellomycetes* were include for comparative analysis with CYP51s of cryptococcal species.

4.2. Methods

4.2.1. CYP51s used in the study

Twenty-one CYP51s from *Tremellomycetes* were used in this study. Among 21 CYP51s, 12 CYP51s from different *Tremellomycetes* representing diverse lifestyles/adaptation to different ecological niches were selected for structural analysis (Table 4.1). A detailed description of the well-known characteristics of *Tremellomycetes* can be found in a recently published article (Akapo *et al.*, 2019). All CYP51 sequences from *Tremellomycetes* were retrieved from a recently published article from our laboratory (Akapo *et al.*, 2019) and used in the study.

Table 4. 1 Information on CYP51s from *Tremellomycetes* used for structural analysis in this study. Amino acid sequences of CYP51s have been retrieved from the published data (Akapo *et al.*, 2019) and used in the study. The databases used for collecting CYP51 sequences were presented in parenthesis under the P450 ID column.

Species name	**Species abbreviation**	**CYP ID**	**CYP abbreviation used in the study**
Cryptococcus neoformans var. *grubii* H99	CngH99	00040 (NCBI)	CYP51F CngH99
Trichosporon asahii var. *asahii* CBS 2479	TaaCBS2479	1742 (JGI)	CYP51F1 TaaCBS2479
Tremella mesenterica Fries v1.0	Tmfri	38441 (JGI)	CYP51F1 Tmfri
Trichosporon oleaginosus	Tole	271374 (JGI)	CYP51F Tole
Cryptococcus curvatus	Ccur	367271 (JGI)	CYP51F Ccur

Naganishia vishniacii v1.0	Nvis	345805 (JGI)	CYP51F1 Nvis
Naematella encephela UCDFST 68-887	Nen UCDFST68-887	523016 (JGI)	CYP51F1 Nen UCDFST68-887
Kockovaella imperatae NRRL Y-17943	Kim NRRL Y-17943	627561 (JGI)	CYP51F1 Kim NRRL Y-17943
Cryptococcus werringae	Cwer	629036 (JGI)	CYP51F Cwer
Cryptococcus terricola	Cter	809380 (JGI)	CYP51F Cter
Cryptococcus neoformans var. *neoformans* B-3501A	Cneo B-3501A	EAL23379 (NCBI)	CYP51F1 Cneo B-3501A
Cryptococcus gattii EJB2	Cgat EJB2	KIR77383 (NCBI)	CYP51F Cgat EJB2

Abbreviations: **NCBI**, National Center for Biotechnology Information; **JGI**, Joint-Genome Institute, USA.

4.2.2. Phylogenetic analysis

Phylogenetic analysis of CYP51s from *Tremellomycetes* was carried out using the maximum likelihood method based on the JTT matrix-based model (Jones *et al.,* 1992) and 500 bootstrap replications. Analysis and visualization of the phylogenetic tree were carried out using molecular evolutionary genetic analysis (Kumar *et al.,* 2016).

4.2.3. Homology modelling of CYP51s

A 3D structural model for each CYP51s (Table 4.1) was constructed by employing a template-based modeling procedure. The template was selected from a non-redundant subset of the Protein Data Bank (Berman, 2008) by searching the database with the HHpred algorithm (Zimmermann *et al.,* 2018) available on the website. CYP51 from *Saccharomyces cerevisiae* (PDB code: 4lxj) (Monk *et al.,* 2014) which has the highest E-value, highest percentage identity and the highest alignment coverage with all 12 CYP51s of *Tremellomycetes* was selected as the best template for modeling of CYP51s. Query to template alignments were also obtained from the HHPred results. Manual editing was introduced when necessary. CYP51s structural models were built based on these alignments with Modeller software (version 9.21) (Webb and

Sali, 2016). Three-dimensional modeling of CYP51s was performed based on structural information from the template 4LXJ where heme co-factor and ligand molecules (lanosterol and heme molecule) were explicitly included during the modelling to ensure correct geometry of an active site.

4.3. Preparation of CYP51s models for docking

Most of calculations related to docking were performed with the Rosetta (Leman *et al.*, 2020) software package, dedicated to biomolecular structure prediction, design and docking. The software has been collaboratively developed by over 60 laboratories from all over the world (Koehler Leman *et al.*, 2020). The clean-pdb.py script (distributed with the software package) (Meiler and Baker, 2006) was used to remove all non-protein molecules from the model. Optimization of bonds and dihedral angles was also performed (Meiler and Baker, 2006). Rosetta's molfile_to_param.py script was used to prepare input files: PDB and params (internal Rosetta format) for each ligand, based on the particular SDF files, downloaded from PubChem following the protocol described elsewhere (Lemmon and Meiler, 2012) obtained from Rosetta Ligand Docking Tutorial online (Meiler and Baker, 2006). Ligand conformers generated with Openbabel (O'Boyle *et al.*, 2011), and BCL (Kothiwale *et al.*, 2015) were used, following the protocol used in Rosetta Community developed at the Meiler's laboratory. A Rosetta program named ligand_rpkmin was used to minimize protein side chains present near the heme molecule following the protocol used in Rosetta Community developed by our laboratory. The protein conformation with the lowest energy was chosen for docking procedures.

4.3.1 Ligands and their docking procedure

CYP51 substrates (lanosterol, obtusifoliol and eburicol), azole drugs such as itraconazole, clotrimazole, voriconazole, fluconazole and ketoconazole including the novel fungal CYP51 inhibitor VT-1129, were used in the study for assessing their binding pattern with 12 CYP51s of *Tremellomycetes* (Table 4.1). *In silico* docking of these nine ligands to each of the CYP51 3D models was performed. Docking simulations were done with a Pyrosetta script (Chaudhury *et al.*, 2010). For this purpose, we adapted a standard script distributed with Rosetta Ligand Docking Tutorial (Meiler and Baker, 2006) and Pyrosetta notebook 08.01-Ligand-Docking-XMLObjects (Le *et al.*, 2020).

The ligand docking procedure was divided into two stages: low and high resolution. During the first, semi-global stage, a ligand was free to sample the protein surface within a sphere of 20 Å, which covers nearly the whole protein. A soft variant of Rosetta score function

called “ligand_soft_rep” was used at this stage to ease energy barrier crossing. Repulsive LJ term and Coulomb interactions were ramped down while hydrogen bonds were promoted. Any side chain within 6 Å from a ligand was also allowed to change its conformation. During the second, high-resolution stage, HighResDocker mover was used for local refinement of a conformation obtained from the previous stage. A final model was subjected to all-atom energy minimization and scoring. The procedure was repeated independently 10 000 for each ortholog-ligand combination and 10 000 resulting poses were subjected to further analysis.

4.4 Hierarchical agglomerative clustering

All clustering analyses in this work were performed with the ap_ligand_clustering program of the BioShell 3.0 package (Macnar *et al.*, 2020). Here, we briefly summarize the procedure as the details of this approach have been reviewed and the software has already been published (Gront and Kolinski, 2005). The procedure begins with placing every structure in a separate cluster of size 1. Then, at every step the two closest clusters are found and merged into one. The newly merged cluster contains all structures from its ancestors and the total number of clusters decreases by one. The procedure is repeated until a convergence condition is met. Here, we stop clustering when the distance between the two closest clusters exceeds a certain distance cutoff, e.g. 5.0 Å. There are a few widely used approaches to compute a distance between two clusters. In this work, we assumed the complete link strategy: the distance between two clusters i.e. groups of structures, is defined as the smallest distance possible between any structure from the first group and a structure from the second group. Clustering with complete link strategy results in relatively many small but tight clusters. A single clustering calculation of 10 000 models takes around 5 minutes on a single CPU core.

It is also very convenient to be able to select a “representative” element of a cluster. Many methods have been devised for this purpose in the literature, e.g. taking an average of all the cluster elements. Averaged conformation however may not represent a chemically correct structure. Therefore, we defined a “middle” structure (often named ‘medoid’ in the literature) as the one for which the sum of distances to all other members of the cluster is the smallest.

4.5. Analysis of docking poses

Each CYP51-ligand combination was analyzed independently. We performed hierarchical agglomerative clustering as described above. In each case, we recorded the five largest clusters having at least 10 structures. Clustering cutoff was chosen to be 5.0 Å after testing values of 1, 2, 3, 5 and 7 Å, followed by manual inspection of the results. We also used the BioShell toolkit

(Gront and Kolinski, 2006) to compute basic statistics on resulting structures, such as the closest distance between a ligand and the iron atom, interatomic contacts between a ligand and a protein, etc.

4.5.1 Protocols

The detailed description of our modeling protocols, comparative modeling, docking and analysis are given on the laboratory Protocols' website: https://labnotes.readthedocs.io. The website provides all the necessary files, scripts and commands that would allow other researchers to our calculations.

4.5.2. Data visualization

In order to analyze the results of the calculations extensively and thoroughly, we built a custom visualization tool (http://bioshell.pl/azoledocking/). The application displays plots, statistics and 3D structural models in a web browser. The server side is run by a Flask application server, also implemented in Python. The client side (i.e. the web browser part) utilizes VisuaLife Python library (visualife.readthedocs.io/) to render plots, charts and tables. NGL biomolecular viewer (Rose and Hildebrand, 2015) is employed to display structural models.

4.5.3. Active site cavity and ligand interacting amino acids detection

Active site cavities in CYP51 models were detected using CavityPlus (Xu *et al.*, 2018). Interactions between a ligand and the protein were detected automatically with a BioShell package (Macnar *et al.*, 2020). Van der Waals interactions were assigned when any atoms of two residues (including hydrogens) were closer to each other than the sum of their VdW radii + 0.5 Å. Hydrogen bonds were assigned when the distance between the donated hydrogen atom and its acceptor was shorter than 3 Å and the planar angle formed by the three atoms, pre-acceptor, acceptor and the hydrogen was wider than 90 deg. Stacking interactions were recorded when the distance between the two aromatic ring centers was shorter than 7 Å.

4.6 Results and discussion

4.6.1 CYP51s grouped as per *Tremellomycetes* lifestyle

Phylogenetic analysis of CYP51s from different *Tremellomycetes* revealed an interesting pattern where CYP51s from species with a similar lifestyle or adaption to similar ecological niches grouped together (Figure 4.1). CYP51s from human pathogenic species such as *C. neoformans* and *C. gatti* grouped together whereas CYP51s from oleaginous yeasts *C. curvatus* and *Trichosporon oleaginosus* IBC0246 were grouped together with the exception of CYP51

from *C. terricola* (Figure 1). The CYP51s from the *Tremellomycetes* that having diverse lifestyles standalone on the tree indicating considerable changes in CYP51 amino acids after speciation. This phenomenon of grouping P450s from species with similar life style/adaption to similar ecological niches was observed in some bacterial (Parvez *et al.*, 2016, Ngcobo *et al.*, 2020, Syed *et al.*, 2019) and fungal species (Jawallapersand *et al.*, 2014, Syed *et al.*, 2014, Qhanya *et al.*, 2015). This indicates that after speciation lifestyle/ecological niches played a key role in changing/preserving the amino acid content, as observed for different P450s families in bacteria (Parvez *et al.*, 2016, Ngcobo *et al.*, 2020, Syed *et al.*, 2019) and fungi (Jawallapersand *et al.*, 2014, Syed *et al.*, 2014, Qhanya *et al.*, 2015).

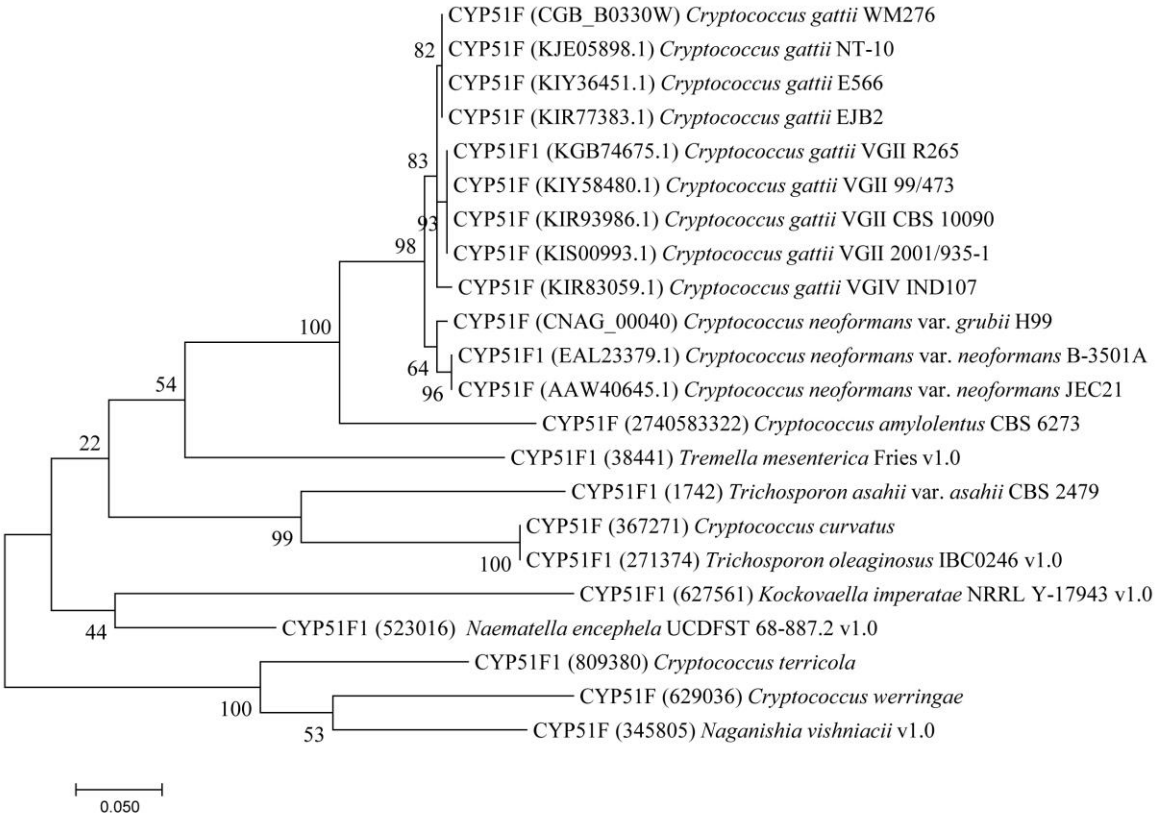


Figure 4.1. Phylogenetic analysis of CYP51s from *Tremellomycetes*. The analysis involved 22 CYP51 protein sequences. Phylogenetic analysis was carried out as described in the methods section and the percentage of trees in which the associated taxa clustered together is shown next to the branches.

4.6.2. CYP51s of *Tremellomycetes* have all the CYP characteristic motifs

Twelve CYP51s from *Tremellomycetes* as representative of diverse lifestyle/ecological niches were selected for structural analysis (Table 1). Three-dimensional structural models were built with template-based modeling using Modeller software version 9.21 (Webb and Sali, 2016). The template 4LXJ (CYP51F from *S. cerevisiae*) (Monk *et al.*, 2014) was selected as the best template for modeling as the sequence identity with CYP51s from *Tremellomycetes* ranged from 41% to 50% at very high alignment coverage above 90% and the sequence similarity was always well above 70% (Table 4.2). Multiple sequence alignment of *Tremellomycetes* CYP51s with the template showed very good alignment, with no major deletions when using the template (Appendix A). Three-dimensional modeling of *Tremellomycetes* CYP51s on such good alignment should produce high-quality models as described elsewhere (Baker and Sali, 2001). Furthermore, a point to be noted is that all these CYPs belong to the same subfamily where they share >55% identity criteria to be part of the same subfamily (Nelson, 2009). Considering these facts, the 12 CYP51 models generated in the study, as expected, were found to have highly similar structures with characteristic CYP helices and beta sheets (Podust *et al.*, 2001, Gotoh, 1992), with only a few minor differences (Figure 4.2 and Appendix A). An amphipathic helix and two transmembrane helices are missing from the N-terminal region of CYP51F Cter whereas an amphipathic helix, two transmembrane helices, αA helix and αL helix and subsequent C-terminal end, are missing from CYP51F1 Nen UCDFST68-887 (protein ID: 523016) (Figure 4.2 and Appendix A). Lack of these helices may be due to genome editing or gene prediction error where prediction of a complete gene sequence impossible at present. Compared to the rest of the 11 CYP51s, the αC helix of CYP51F1 TaaCBS2479 is found to be longer with 34 amino acids (Figure 4.2 and Appendix A). *Tremellomycetes* CYP51s have substrate recognition sites 1-6 (SRS 1-6) with the exception of CYP51F1 Nen UCDFST68-887, where the SRS6 region is missing because of the C-terminal region is missing, as described earlier (Figure 4.2 and Appendix A). Amino acids, part of the SRS, are highly conserved across the *Tremellomycetes* CYP51s (Appendix A).

Table 4.2. Summary statistics of CYP51s from *Tremellomycetes* alignments with CYP51 from Saccharomyces cerevisiae (PDB code: 4LXJ) (Monk *et al.*, 2014).

CYP51 name	Protein ID	Number of Amino acids in CYP51	Number of amino acids aligned	E-value	Identities	Similarity

			with 4LXJ			
CYP51F CngH99	00040	549	523	7.4E-67	43.0%	78.6%
CYP51F1 TaaCBS2479	1742	562	507	3.5E-63	44.0%	76%
CYP51F Tole	271374	535	512	6.5E-65	44.0%	78.5%
CYP51F1 Nvis	345805	536	514	4.7E-67	44.0%	78.2%
CYP51F Ccur	367271	535	512	6.5E-65	44.0%	78.5%
CYP51F1 Tmfri	38441	540	514	1.9E-65	41.0%	74.3%
CYP51F1 Nen UCDFST68-887	523016	390	364	3.8E-50	50.0%	89.3%
CYP51F1 Kim NRRL Y-17943	627561	548	525	7.5E-63	41.0%	75.7%
CYP51F Cwer	629036	533	516	1.6E-55	45.0%	80.3%
CYP51F Cter	809380	468	445	7.7E-62	46.0%	84.3%
CYP51F1 Cneo B-3501A	EAL23379	546	524	2.00E-66	41.0%	76.8%
CYP51F Cgat EJB2	KIR77383	549	523	1.5E-66	42.0%	78.3%

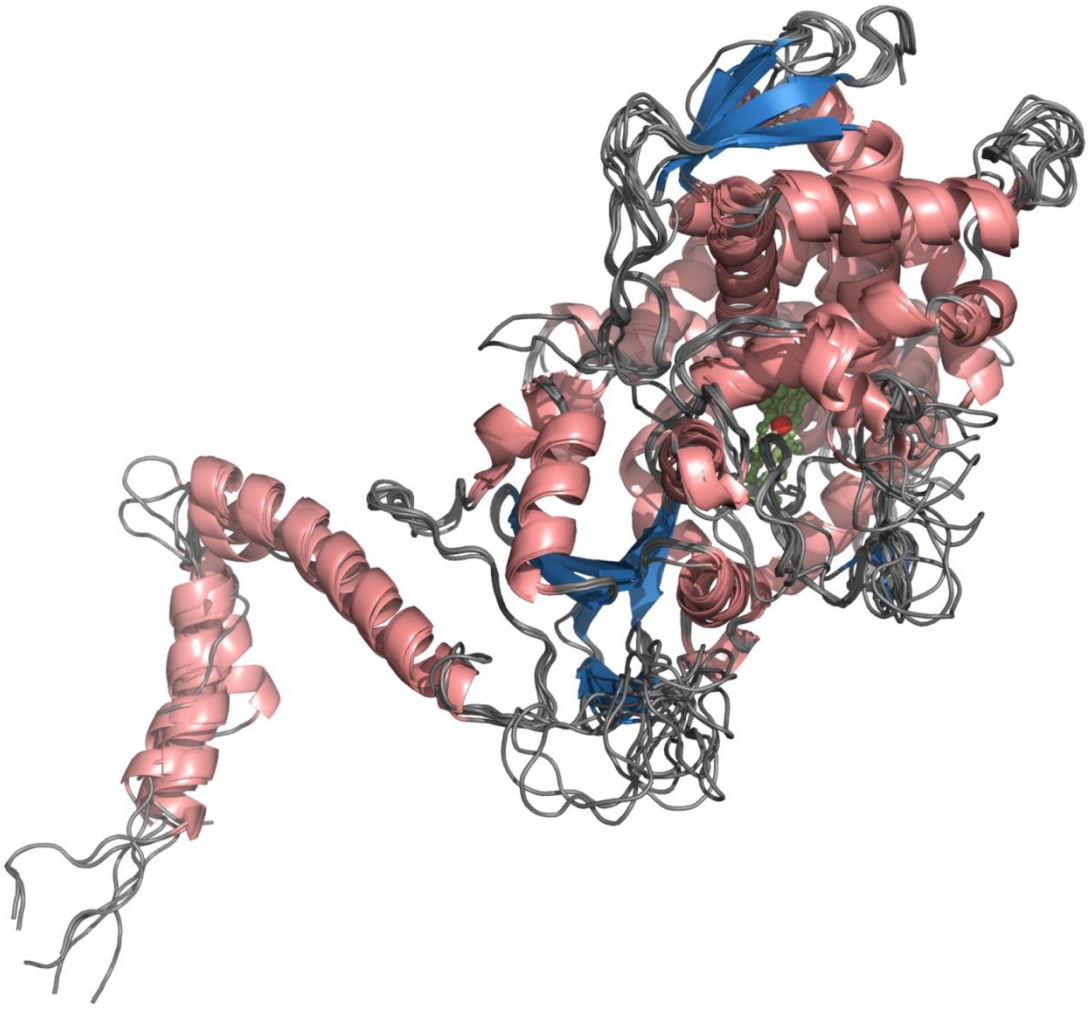


Figure 4.2. Superimposition of 3D structural models of 12 CYP51s from *Tremellomycetes*. Helices and beta sheets are shown in salmon and sky blue. Heme is shown in green along with the Fe atom in red.

4.6. 3. *Tremellomycetes* CYP51s active site cavities are highly hydrophobic

Analysis of active site cavities revealed that *Tremellomycetes* CYP51s active site cavities mostly comprise hydrophobic amino acids (Figure 4.3 and Appendix B). The active site cavity volumes range from 685 $Å^3$ - 4173 $Å^3$, where CYP51F1 Nen UCDFST68-887 has the largest active site cavity and CYP51F1 TaaCBS2479 has the smallest active site cavity (Figure 4.3 and Appendix B). One of the reasons why CYP51F1 Nen UCDFST68-887 has the largest active site cavity is the opening of a channel right next to heme because of the missing C-terminal region that encompasses SRS6, as described earlier. Thus, the active site cavity volume of this P450 is not an actual volume and a full-length CYP sequence is needed to deduce the correct volume of the active site cavity. CYP51s from prominent human pathogens

such as *C. neoformans* var. *grubii* H99, *C. neoformans* var. *neoformans* B-3501A and *C. gattii* EJB2 have similar sizes to active site cavities (Figure 4.3 and Appendix B). CYP51s from other cryptococcal species have larger active site cavities compared to CYP51s from these three prominent human pathogens (Figure 4.3 and Appendix B). Analysis of the composition of active site cavity amino acids revealed that 50-60% of cavity amino acids are hydrophobic (Appendix B), suggesting that *Tremellomycetes* CYP51s' active site cavities are highly hydrophobic. Among *Tremellomycetes* CYP51s, CYP51F from *C. werringae* has the highest percentage of hydrophobic amino acids (60%). CYP51s active site cavities' acidic residues range from 1.64% - 5.73% and the basic residues range from 6.25% - 12.17% where CYP51F1 Nen UCDFST68-887 has most acidic amino acids in its cavity and CYP51F from *C. curvatus* has most basic amino acids in its active site cavity (Appendix B). The hydrophobic nature of these CYP51s' active sites is expected considering that they accept highly hydrophobic substrates such as sterols.

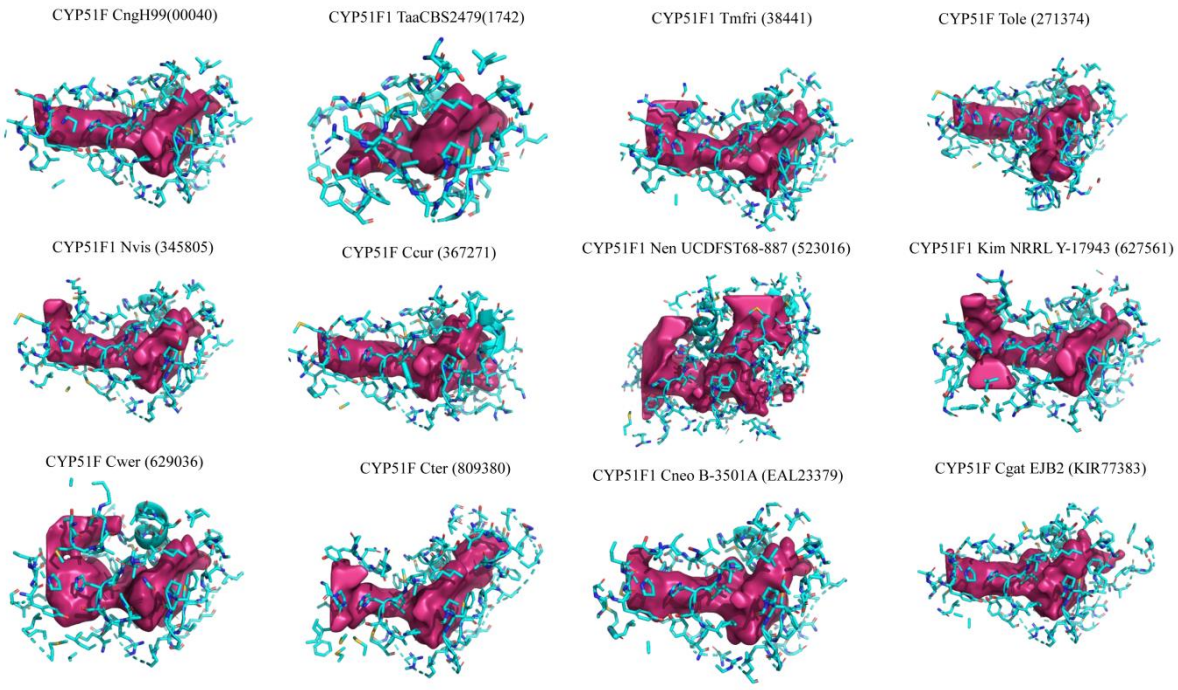


Figure 4.3. Active site cavities of 12 CYP51s from *Tremellomycetes*. The active site cavity surface is shown in pink and amino acids, part of the active site cavity, are represented using a ball and stick model. A detailed list of amino acids, part of the active site cavity, is presented in Appendix B.

4.6.4. A web application for visualization of *Tremellomycetes* CYP51s interactions with ligands

After successful construction of 12 CYP51s 3D models we then proceeded to understand their interactions with ligands such as substrates and azoles. Among substrates, apart from the fungal CYP51 substrates lanosterol and eburicol (Lepesheva *et al.*, 2018) we also included the plant CYP51 substrate obtusifoliol (Lepesheva *et al.*, 2018) as a positive control to assess the accuracy of the study. Among the azoles, five are triazoles (clotrimazole, fluconazole, itraconazole, ketoconazole and voriconazole) and one is tetrazole (VT_1129). The rationale for including these azoles especially itraconazole is to assess if any corroboration occurs between the *in silico* study and the laboratory studies that showed itraconazole performing better than the fluconazole (Trilles *et al.*, 2004, Thompson *et al.*, 2009, Pan *et al.*, 2012, Barchiesi *et al.*, 2000, Illnait-Zaragozi *et al.*, 2008, Hagen *et al.*, 2010, Armengou, 1996).

Docking of each CYP51 vs each ligand in 12 x 9 resulted in generating 108 combinations (Figure 4.4). For each of these combinations, we generated 10 000 models as described in the methods section. Each of these combinations was analyzed independently. The huge amount of data generated during these docking calculations posed a formidable challenge to the analysis. Therefore, we developed a custom web application for this purpose, and made it publicly available at the address: http://bioshell.pl/azoledocking/. The application can display every model that was generated with its basic numerical parameters, such as Rosetta energy, number of hydrogen bonds, etc. The results can be browsed by cluster, sorted, etc.

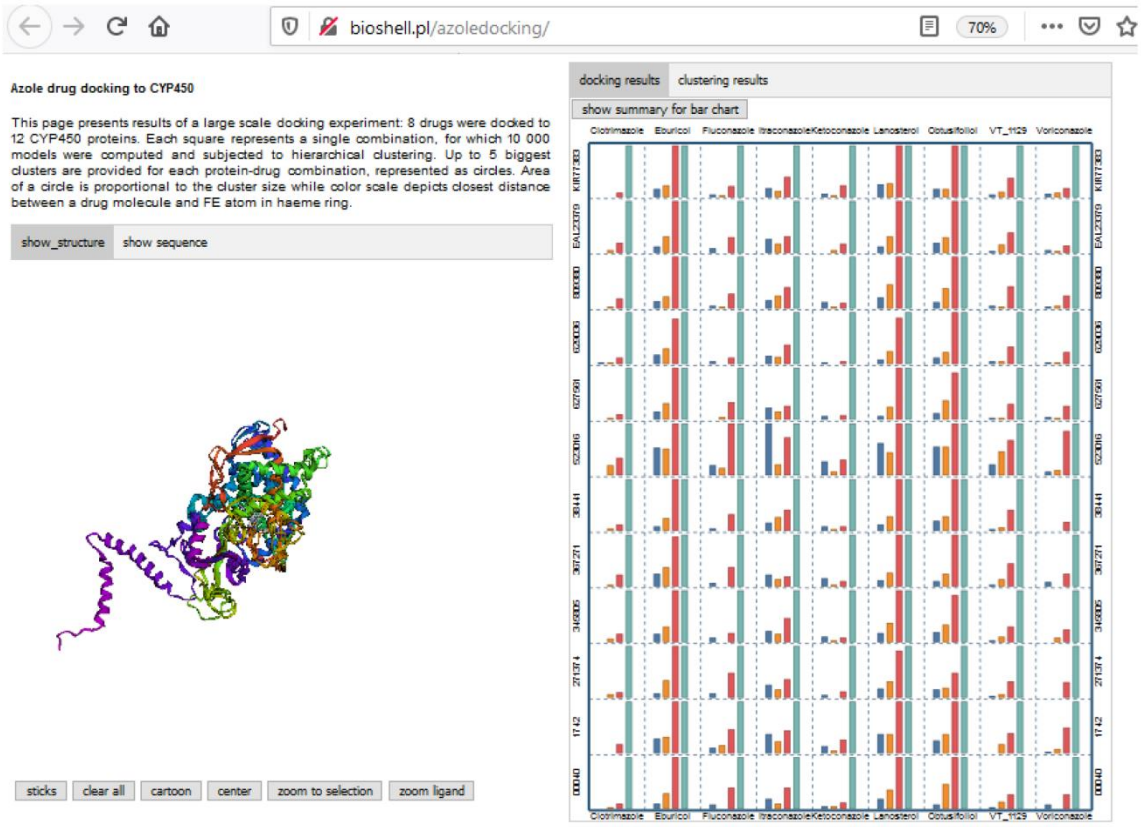


Figure 4.4. Graphical overview of docking results at online web application. There are 108 squares representing the 12 x 9 = 108 CYP51-ligand combinations considered in this work. CYP51s were indicated with their protein IDs. Each square displays up to four bars, corresponding to models classified as ‘clashes’, ‘correct’, ‘near’ and ‘pocket’ (blue, orange, red and cyan, respectively). Bars taller than 60 observations were reduced to 60 to fit into a plotting area; actual values are provided in Table 4.2. Docking results and corresponding information can be accessed at our online web application (http://bioshell.pl/azoledocking/).

4.6.5 *Tremellomycetes* CYP51s have the highest preference for itraconazole

Structures for the complexes (CYP51-ligand) considered in this study have not yet been established experimentally. Therefore, we cannot compare these study results to any reference structure. Nevertheless, the geometry of interaction between a heme moiety of P450s and azole drugs has been studied extensively. Aromatic nitrogen forms a semi-covalent bond with the Fe atom of length d_{NFe} = 2.1 Å (Balding *et al.*, 2008). In our analysis, we report the closest distance between heme's iron and an aromatic nitrogen atom of an azole drug as a measure of success.

For the cases of eburicol, lanosterol and obtusifoliol d_{NFe} denotes the closest distance between Fe and any atom of a ligand. Since we allowed Rosetta to sample a wide range of binding spots, we observe ligands located in virtually any region of a receptor. In order to assess the performance of Rosetta *Ligand Docking* protocol critically, we included all models in our analysis.

The values of d_{NFe} range from below 2 Å to 80 Å. For the analysis, we defined the following four groups, based on the d_{NFe} distance:

- 'clashes' where a ligand was closer to the Fe than 1.5 Å
- 'correct' where d_{NFe} was in the range 1.5 Å - 2.5 Å
- 'near' with d_{NFe} in the range 2.5 Å - 3.5 Å
- 'pocket' with d_{NFe} in the range 3.5 Å - 8.0 Å

The number of models in each of the four categories is provided in Table 4.2, as well as shown in Figure 4.4 by blue, orange, red and cyan bars, respectively. The definition of these categories was based on a visual analysis of the resulting structures using the online application. The distance ranges defining them were manually adjusted to keep the classification meaningful for as many models as possible. Information on each of these four were elaborated below:

- **Models with clashes**

There is a small but noticeable number of clashing conformations observed for many CYP51F-ligand conformations, most likely artifacts of the modelling process. Its fraction ranges from nearly 0% (1 model in 120 000) for clotrimazole to 0.15% for itraconazole. Interestingly, eburicol, lanosterol and obtusifoliol metabolites recorded a significantly higher fraction of incorrect conformations, above the average. These models could, in general, be easily excluded on an early stage of analysis, either by a filtering protocol in Rosetta itself, or by a BioShell script. Here we report them only to draw the complete analysis.

- **Correct conformations**

In this semi-global docking scenario as we applied in our study, ligands explored nearly the whole surface of a protein. The chance of finding a correct solution is therefore relatively low. Here we observe from 14 correct solutions for ketoconazole to 212 correct solutions for lanosterol - from 0.012% to 0.18% of all models generated for a given ligand. The dependence

of these statistics on a protein is considerably weaker than on a ligand. The number of correct solutions ranges from 51 for KIR77383 to 143 cases for 523016 ortholog.

- **Nearly correct conformations**

The 1Å range of d_{NFe} distance we assumed to define correct solutions is pretty tight. For more generous 2Å criterion a large fraction of ligand molecules are located correctly in the active site even though their interaction with an iron atom is not that close. Given such an extended criterion of success, for all 108 combinations "nearly-correct" solutions were found.

- **In the pocket conformations**

The fourth group consists of models where a small molecule agent still resides in the active pocket, but is quite far from the heme co-factor. Orientation of a ligand is very diverse in these models and often different from what might be expected from experimental data. Total energy for models in this group is comparable to the values observed for "correct", and "nearly-correct'' models. The main reason for this is because Rosetta is unaware of the specificity of the system we study. None of the two energy functions that was used by Rosetta modelling software: "soft" for docking and "hard-rep" for final scoring, contain a term that would explicitly assess interactions between a ligand and a heme co-factor, besides standard Van der Waals and Coulomb interactions. Most notably, there is no potential to mimic a semi covalent bond that is formed between a heme iron atom and an azole nitrogen. This interaction has significant contribution to the total energy of a complex. Including this effect in the modelling protocol would certainly favor the correct geometry. In the majority of models therefore an inhibitor is further away from heme that 8 Å and doesn't count to any of the four considered categories.

Table 4.3. Analysis of *Tremellomycetes* CYP51s and ligand complexes. Number of models falling into 'clashes', 'correct', 'near' and 'pocket' group summarized by ligands. These numbers are summed up over all 12 proteins, so the total number of models for each inhibitor is 120000.

Ligand	**no. clashes**	**no. correct**	**no. near**	**no. pocket**
Lanosterol	131	212	809	12032
Obtusifoliol	116	209	720	11670

Eburicol	116	202	828	12511
Itraconazole	185	108	256	9041
average	*76*	*98*	*387*	*7555*
VT_1129	24	76	250	4552
Clotrimazole	1	25	98	2825
Voriconazole	21	22	195	5020
Fluconazole	42	19	240	5040
Ketoconazole	55	14	87	5307

Ligand binding affinity under physiological conditions results from free energy gain upon binding. This free energy change includes entropic contributions as well as other effects such as solvent, ions, etc. Although some of these effects (most notably the interactions with solvent) are included in the Rosetta force field as mean field potentials, the energy value reported by Rosetta primarily accounts for conformational energy (i.e. enthalpy). In order to assess relative binding affinity, one has to take into account the number of observations for a particular conformation. The docking procedure that was applied allowed Rosetta to place a ligand virtually anywhere at the protein surface to avoid any conformational bias. The numbers of correctly docked models given in Table 4.2 therefore provide an estimation of binding affinity for different ligands. Based on these results, it is clear that *Tremellomycetes* CYP51s' natural substrate, lanosterol, achieved the top score, followed by obtusifoliol and eburicol indicating that our method of analysis is correct, as one can expect the highest binding affinity to natural substrates. It is also interesting to see the pattern of preference for azoles by *Tremellomycetes* CYP51s where itraconazole is preferred azole followed by VT_1129 and fluconazole, which is the least favored before ketoconazole (Table 2). This preference pattern perfectly matches wet laboratory data where itraconazole and VT_1129 were shown to perform better than fluconazole (Trilles *et al.,* 2004, Thompson *et al.,* 2009, Pan *et al.,* 2012, Barchiesi *et al.,* 2000, Illnait-Zaragozi *et al.,* 2008, Hagen *et al.,* 2010, Armengou, 1996, Hoekstra *et al.,* 2014, Warrilow *et al.,* 2014, Lockhart *et al.,* 2016a, Nielsen *et al.,* 2017a). Overall, based on these results, we can safely say that our method of analysis is correct.

Given the drawbacks of the Rosetta scoring functions as elaborated on the supplementary section "CYP51-ligand", the chance of selecting a correct complex without prior knowledge of the actual *in vivo* conformation is low. A commonly used remedy to solve this problem is structural clustering. Thus, in this study, we used this method to select the correct CYP-ligand complex.

Clustering analysis is a method commonly used in the field of biomacromolecular modelling. It helps to analyze large amounts of data generated by computer simulations and to overcome inaccuracies of molecular force fields. A group of structurally similar conformations corresponds to a single local energy minimum (Gront *et al.,* 2005). In general, it is expected that a conformation corresponding to a larger cluster will be more favored entropically. Entropic contributions, necessary to estimate the free energy of a given conformation, are typically not described properly by molecular modeling force fields. Giving preference to larger clusters during the process of selecting the final docking result, implicitly allows inclusion of these entropic contributions. Here, we employed a hierarchical agglomerative clustering algorithm, as described in the methods section. Following the standard protocols in the field, we report the five largest clusters for every protein-ligand combination. The clusters are indicated graphically in Figure 4.5 as circles. A circular area is proportional to a cluster size while the color scale represents the d_{NFe} values, in a range from 1 to 80 Å (see Figure 4.5). In 3 cases no suitable clusters were found (Figure 4.5). All these clustering results can be browsed online at: http://bioshell.pl/azoledocking/.

All the following analyses require selection of a single representative model for each of the 108 modelling cases. The clustering procedure described above is certainly helpful, but on its own is unable to select the final solution in every case. Therefore, we used a web application (http://bioshell.pl/azoledocking/) to browse the results and manually selected final conformations. In this process we took into account the classification resulting from the clustering procedure, the chemical correctness of a conformation and the number of interactions (stacking, H-bonding and Van der Waals) between a ligand and a protein. The number of such interactions listed in a table of the web application that facilitates the selection process.

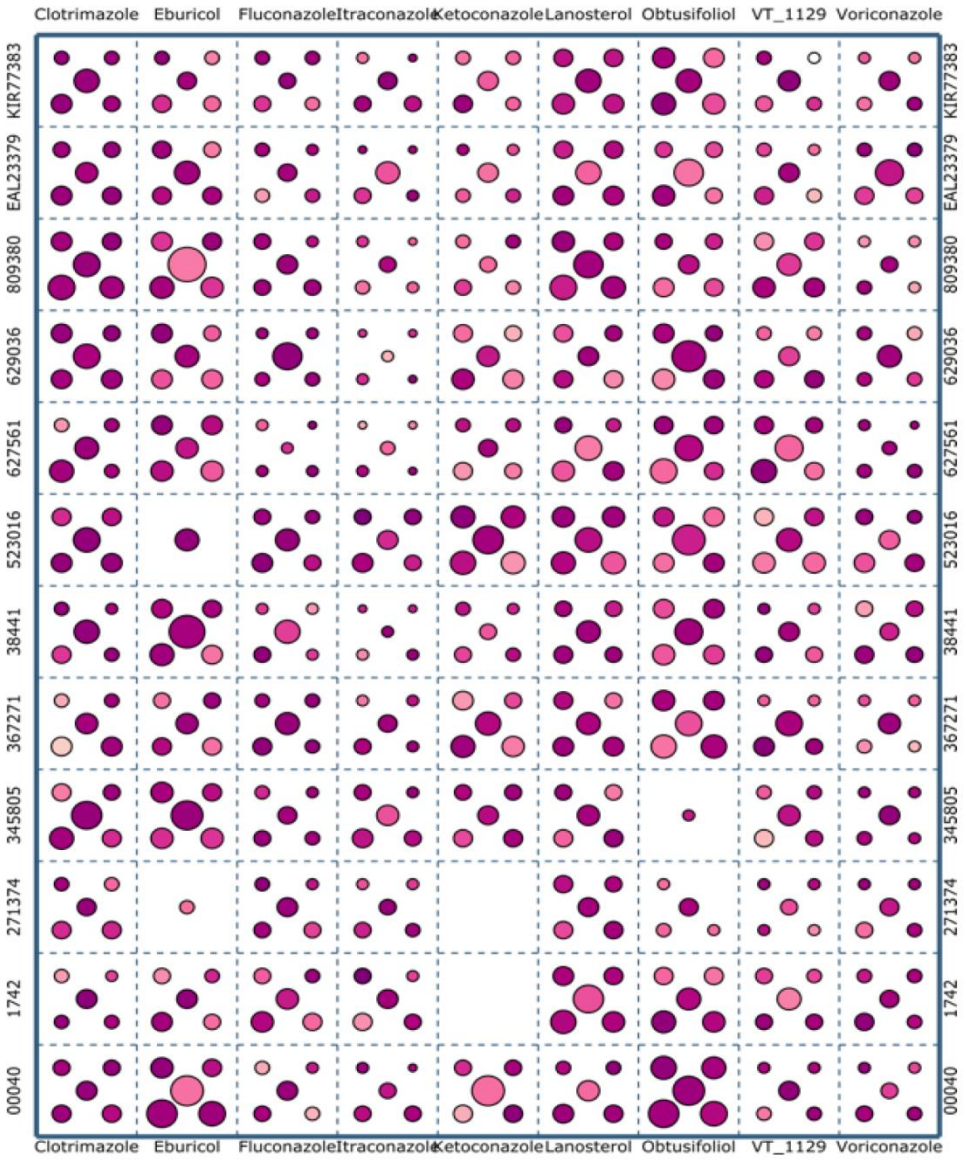


Figure 4.5. Graphical representation of results of hierarchical clustering. Each of the 108 squares displays up to five clusters, represented by circles. The circular area is proportional to the size of the respective cluster, while the color scale denotes the d_{NFe} distance, with darker colors representing shorter values. All clusters of at least 10 structures found by hierarchical clustering are marked on this figure.

4.6.7. High conservation observed in *Tremellomycetes* CYP51s amino acids interacting with ligands

After selection of each of the CYP51-ligand complex as described above we further analyzed the amino acids interacting with different ligands and their location. This is a very important

aspect to check the accuracy of our method employed to dock and select CYP51-ligand complexes. As all of these CYPs belong to the same subfamily and high conservation of amino acids is observed in SRS regions (Appendix A), if our analysis is correct we should observe high conservation of amino acids interacting with analyzed ligands across the CYPs. As predicted, our analysis revealed conservation of a large number of amino acids interacting with docked ligands across the CYPs (Tables 4.3 and Appendix B). Mapping of the amino acids interacting with ≥5 ligands revealed that these amino acids are part of SRSs (Appendix A). We also identified amino acids across the CYPs that are able to form hydrogen bonds with ligands and interact with ligands *via* stacking and Van der Waals interactions (Appendix B). This suggests that our method of docking and selecting the CYP-ligand complexes is correct and in future these aspects can be automated to understand the CYP-ligand interactions to cater for the growing number of CYPs.

Table 4.4. Analysis of *Tremellomycetes* CYP51s amino acids interacting with different ligands. Amino acids interacting with ≥5 ligands are presented in the table. Amino acids are colored based on their interactions with nine ligands (red), eight ligands (blue), seven ligands (brown), six ligands (green) and five ligands (pink). A complete list of amino acids interacting with different ligands is listed in Appendix B.

CYP51F	Amino acids interacting with ligands (≥5 ligands)
CYP51F CngH99	F240, Y145, A317, F139, I386, M528, T321, V144, Y131
CYP51F1 TaaCBS2479	S159, A158, A334, F127, F257, H337, I403, M333, M545, P402, S157, T123, T338, T544
CYP51F1 Tmfri	A307, F126, F228, I376, M521, T311, V131, Y118, Y132
CYP51F Tole	I376, A307, F230, M518, T124, T311, Y120, Y134
CYP51F Ccur	A307, F128, F230, I376, M518, T124, T311, Y120, Y134
CYP51F1 Nvis	F228, Y120, A305, F128, I374, T309, Y134
CYP51F1 Nen UCDFST68-887	A210, A214, F137, F35, H217, I283, T218, T31, V40, Y27, Y41

CYP51F1 Kim NRRL Y-17943	A313, F235, H316, I382, L129, M526, T130, T317, T525, Y126, Y140
CYP51F Cwer	A304, F128, F227, I373, S375, T124, T308, Y120, Y134
CYP51F Cter	A237, F160, F60, H240, L306, L55, M236, M451, T241, T56, Y52, Y66
CYP51F1 Cneo B-3501A	F234, I380, A311, F133, H314, I523, M310, T129, T315, V524, Y125, Y139
CYP51F Cgat EJB2	F240, A317, F139, I386, M316, M528, T135, T321, Y131, Y145

Analysis of amino acid substitutions in CYP51s that are found to be critical in response to different azole drugs (Rodero *et al.*, 2003, Sionov *et al.*, 2012, Kano *et al.*, 2017) revealed that Y145 and G484 amino acids are absolutely conserved in the *Tremellomycetes* CYP51s (Appendix A indicated with "X"). However, substitution of glycine at 410 position (formerly reported as G344)(Kano *et al.*, 2017) with alanine (in three CYP51s), serine (in two CYP51s), lysine (in two CYP51s) and arginine (in a single CYP51) is found (Appendix A indicated with "X"). This is quite an interesting observation as it has been reported that G344S substitution in six strains of *C. neoformans* var. *grubii* strain, NUBS14020 resulted in the development of multi-drug resistance to azoles such as fluconazole, itraconazole and voriconazole (Kano *et al.*, 2017). Considering these findings, it would be interesting to see the response of the strains containing G344 substitutions in their CYP51s.

4.7. Conclusions

In this study, 12 CYP51s structural models were generated. Analysis of these models revealed the presence of all CYP characteristic motifs with minor differences. Active site cavity analysis revealed that all 12 CYPs cavities were highly hydrophobic. Interaction of 12 CYPs with three different substrates and six azole compounds was carried out. To analyze the huge data, an online visualization tool was created. Analysis of CYP51 interactions revealed their highest affinity to the natural substrate, lanosterol. An interesting pattern was observed in azole drug interactions, where 12 CYPs showed the highest preference for itraconazole compared to fluconazole. These study results provided a reason for the better performance of itraconazole compared to fluconazole. Key amino acids interacting with different ligands were identified.

CHAPTER 5

Overall conclusions and future perspectives

Cytochrome P450 monooxygenases (CYPs/P450s) are some of the well-studied enzymes in the field of biology. Among CYPs, CYP51 serves as drug target against fungal pathogens. In the case of cryptococcosis, a life-threatening fungal disease in immunocompromised patients, studies indicated that cryptococcal species are developing resistance to the currently used azole drug fluconazole. Despite knowing that CYP51 is the prime target of this drug, mutations in this gene conferring resistance on fluconazole and a large number of *in vitro* studies indicating that other azoles perform better than fluconazole, to date very few studies on CYP51s structure from cryptococcal or other *Tremellomycetes* and their interactions with azoles have been reported. To address this research gap, two tasks were undertaken in this study: (i) Genome data mining, annotation and phylogenetic analysis of CYPs in cryptococcal species. (ii) Phylogenetic, comparative modelling of 12 CYP51s and their interactions with different ligands and development of an online visualization tool.

This study revealed that cryptococcal species have almost 50% fewer CYP genes than their non-pathogenic counterparts and furthermore have the highest CYP diversity. Four CYP families were found to be conserved in pathogenic *Cryptococcus* species, indicating their important role in these pathogens. Interestingly, the CYP5139 family was bloomed with 17 CYP subfamilies in species of *Tremellomycetes,* indicating its possible key role in the physiology of these organisms. This study's results provide insight into the CYP enzymes in the species of *Tremellomycetes*. This study serves as a reference for future annotation of CYPs and has opened new vistas for the characterization of CYPs in the species of *Tremellomycetes*.

Phylogenetic analysis of CYP51s indicated some kind of amino acid conservation after speciation, as CYP51s grouped as per their species lifestyle/ecological niches. Structural models built for 12 CYP51s revealed the presence of P450 characteristic motifs and high conservation of amino acids in substrate recognition sites (SRSs) 1-6. The active site cavities of CYP51s were found to be highly hydrophobic to accommodation of hydrophobic sterol substrates. Since the structures for the complexes (CYP51-ligand) we considered have not yet been established experimentally, we performed global docking of ligands (three substrates and six azoles) with 10 000 combinations for each of the complex and the correct complexes were selected by hierarchical agglomerative clustering using the cluster program of the BioShell package. To analyze the huge data generated in this study, we developed an online visualization

tool (http://bioshell.pl/azoledocking/). These study results indicated that *Tremellomycetes* CYP51s have the highest preference for its natural substrate lanosterol, followed by obtusifoliol and eburicol. Among azoles, *Tremellomycetes* CYP51s have the highest preference for itraconazole, followed by VT_1129, clotrimazole, voriconazole, fluconazole and ketoconazole. *Tremellomycetes* CYP51s' high preference for itraconazole is possibly the reason why this azole perform better than fluconazole, as observed in a large number of *in vitro* studies. The methodology followed in this study is correct, as the amino acids we identified as interacting with different ligands are highly conserved across the CYP51s, indicating no errors in the procedure. Furthermore, most of these amino acids were found to be part of SRSs. In this study, we focused on a single subfamily of CYPs simply to assess the methods employed in the study, but this method needs to be validated for the analysis of different CYP families and their interactions with different ligands. Studies are in progress to develop a website on which this program will be automated to study CYPs' interactions with different compounds. This type of dedicated website for this family of enzymes is desperately needed, as the number of CYPs is growing exponentially every day. The results of this study further strengthen the previous laboratory results that itraconazole and VT_1129 will be potential compounds in the fight against cryptococcosis and their efficacy against cryptococcal species needs to be determined. One of the interesting results from this study is the identification of non-conservation of G344 amino acid in *Tremellomycetes* CYP51s, including in some cryptococcal species. It would be interesting to see if these *Tremellomycetes* are resistant to fluconazole, itraconazole and voriconazole, as was reported for G344S substituted species.

REFERENCES

AKAPO, O. O., PADAYACHEE, T., CHEN, W., KAPPO, A. P., YU, J. H., NELSON, D. R. & SYED, K. 2019. Distribution and diversity of Cytochrome P450 monooxygenases in the fungal class *Tremellomycetes*. *Int J Mol Sci,* 20.

ALANIO, A., VERNEL-PAUILLAC, F., STURNY-LECLÈRE, A. & DROMER, F. 2015. Cryptococcus neoformans host adaptation: Toward biological evidence of dormancy. *MBio,* 6.

ARMENGOU, A. 1996. Possible development of resistance to fluconazole during suppressive therapy for AIDS-associated cryptococcal meningitis. *Clin Infect Dis,* 23**,** 1337-1338.

BAKER, D. & SALI, A. 2001. Protein structure prediction and structural genomics. *Science,* 294**,** 93-96.

BALDING, P. R., PORRO, C. S., MCLEAN, K. J., SUTCLIFFE, M. J., MARÉCHAL, J.-D., MUNRO, A. W. & VISSER, S. P. D. 2008. How do azoles inhibit cytochrome P450 enzymes? A density functional study. *The Journal of Physical Chemistry A,* 112**,** 12911-12918.

BAMAL, H. D., CHEN, W., MASHELE, S. S., NELSON, D. R., KAPPO, A. P., MOSA, R. A., YU, J. H., TUSZYNSKI, J. A. & SYED, K. 2018. Comparative analyses and structural insights of the novel cytochrome P450 fusion protein family CYP5619 in Oomycetes. *Sci Rep,* 8**,** 6597.

BARCHIESI, F., SCHIMIZZI, A. M., CASELLI, F., NOVELLI, A., FALLANI, S., GIANNINI, D., ARZENI, D., DI CESARE, S., DI FRANCESCO, L. F. & FORTUNA, M. 2000. Interactions between triazoles and amphotericin B against Cryptococcus neoformans. *Antimicrobial Agents and Chemotherapy,* 44**,** 2435-2441.

BARNETT, J. A. 2010. A history of research on yeasts 14: Medical yeasts, part 2, Cryptococcus neoformans. *Yeast (Chichester),* 27**,** 875-904.

BASENKO, E. Y., PULMAN, J. A., SHANMUGASUNDRAM, A., HARB, O. S., CROUCH, K., STARNS, D., WARRENFELTZ, S., AURRECOECHEA, C., STOECKERT, C. J., JR., KISSINGER, J. C., ROOS, D. S. & HERTZ-FOWLER, C. 2018. FungiDB: An integrated bioinformatic resource for fungi and Oomycetes. *J Fungi (Basel),* 4.

BENNETT, J. E., KWON-CHUNG, K. & HOWARD, D. H. 1977. Epidemiologic differences among serotypes of Cryptococcus neoformans. *American Journal of Epidemiology,* 105**,** 582-586.

BERMAN, H. M. 2008. The protein data bank: A historical perspective. *Acta Crystallographica Section A: Foundations of Crystallography,* 64**,** 88-95.

BERNAL-MARTINEZ, L., GOMEZ-LOPEZ, A., CASTELLI, M. V., MESA-ARANGO, A. C., ZARAGOZA, O., RODRIGUEZ-TUDELA, J. L. & CUENCA-ESTRELLA, M. 2010. Susceptibility profile of clinical isolates of non-Cryptococcus neoformans/non-Cryptococcus gattii Cryptococcus species and literature review. *Medical Mycology,* 48**,** 90-96.

BOC, A., DIALLO, A. B. & MAKARENKOV, V. 2012. T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res,* 40**,** W573-9.

BOUKLAS, T., PECHUAN, X., GOLDMAN, D. L., EDELMAN, B., BERGMAN, A. & FRIES, B. C. 2013. Old Cryptococcus neoformans cells contribute to virulence in chronic cryptococcosis. *MBio,* 4.

BURLEY, S. K., BERMAN, H. M., BHIKADIYA, C., BI, C., CHEN, L., DI COSTANZO, L., CHRISTIE, C., DALENBERG, K., DUARTE, J. M. & DUTTA, S. 2019. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids esearch,* 47**,** D464-D474.

BUSSE, O. 1894. Uber parasitare Zelleinschlusse und ihre Zuchtung. *Zentralbl Bakteriol,* 16**,** 175-180.

CHAKRABARTI, A., GHOSH, A., BATRA, R., KAUSHAL, A., ROY, P. & SINGH, H. 1996. Antifungal susceptibility pattern of non-albicans Candida species and distribution of species isolated from Candidaemia cases over a 5 year period. *The Indian Journal of Medical Research,* 104**,** 171-176.

CHAUDHURY, S., LYSKOV, S. & GRAY, J. J. 2010. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics,* 26**,** 689-691.

CHEN, S. C.-A., MEYER, W. & SORRELL, T. C. 2014. Cryptococcus gattii infections. *Clinical microbiology reviews,* 27**,** 980-1024.

CHEN, Y.-C., CHANG, T.-Y., LIU, J.-W., CHEN, F.-J., CHIEN, C.-C., LEE, C.-H. & LU, C.-H. 2015. Increasing trend of fluconazole-non-susceptible Cryptococcus neoformans in patients with invasive cryptococcosis: A 12-year longitudinal study. *BMC Infectious Diseases,* 15**,** 277.

CHOI, J. N., KIM, J., KIM, J., JUNG, W. H. & LEE, C. H. 2012. Influence of iron regulation on the metabolome of Cryptococcus neoformans. *PLoS One,* 7**,** e41654.

CHOI, J. Y., PODUST, L. M. & ROUSH, W. R. 2014. Drug strategies targeting CYP51 in neglected tropical diseases. *Chemical Reviews,* 114**,** 11242-11271.

CLOSE, D. & OJUMU, J. 2016. Draft genome sequence of the oleaginous yeast *Cryptococcus curvatus* ATCC 20509. *Genome Announc,* 4.

CLOSE, D., OJUMU, J. & ZHANG, G. 2016. Draft genome sequence of *Cryptococcus terricola* JCM 24523, an oleaginous yeast capable of expressing exogenous DNA. *Genome Announc,* 4.

COELHO, C. & CASADEVALL, A. 2016. Cryptococcal therapies and drug targets: The old, the new and the promising. *Cellular Microbiology,* 18**,** 792-799.

COGLIATI, M. 2013. Global molecular epidemiology of Cryptococcus neoformans and Cryptococcus gattii: An atlas of the molecular types. *Scientifica,* 2013.

D'SOUZA, C. A., KRONSTAD, J. W., TAYLOR, G., WARREN, R., YUEN, M., HU, G., JUNG, W. H., SHAM, A., KIDD, S. E., TANGEN, K., LEE, N., ZEILMAKER, T., SAWKINS, J., MCVICKER, G., SHAH, S., GNERRE, S., GRIGGS, A., ZENG, Q., BARTLETT, K., LI, W., WANG, X., HEITMAN, J., STAJICH, J. E., FRASER, J. A., MEYER, W., CARTER, D., SCHEIN, J., KRZYWINSKI, M., KWON-CHUNG, K. J., VARMA, A., WANG, J., BRUNHAM, R., FYFE, M., OUELLETTE, B. F., SIDDIQUI, A., MARRA, M., JONES, S., HOLT, R., BIRREN, B. W., GALAGAN, J. E. & CUOMO, C. A. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *MBio,* 2**,** e00342-10.

DATTA, K., BARTLETT, K. H. & MARR, K. A. 2009. Cryptococcus gattii: Emergence in Western North America: Exploitation of a novel ecological niche. *Interdisciplinary Perspectives on Infectious Diseases,* 2009.

DAUM, G., LEES, N. D., BARD, M. & DICKSON, R. 1998. Biochemistry, cell biology and molecular biology of lipids of Saccharomyces cerevisiae. *Yeast,* 14**,** 1471-1510.

DAVIES, G. E. & THORNTON, C. R. 2014. Differentiation of the emerging human pathogens *Trichosporon asahii* and *Trichosporon asteroides* from other pathogenic yeasts and moulds by using species-specific monoclonal antibodies. *PLoS One,* 9**,** e84789.

DROMER, F., GUEHO, E., RONIN, O. & DUPONT, B. 1993. Serotyping of Cryptococcus neoformans by using a monoclonal antibody specific for capsular polysaccharide. *Journal of Clinical Microbiology,* 31**,** 359-363.

DUARTE-OLIVEIRA, C., RODRIGUES, F., GONCALVES, S. M., GOLDMAN, G. H., CARVALHO, A. & CUNHA, C. 2017. The cell biology of the *Trichosporon*-host interaction. *Front Cell Infect Microbiol,* 7**,** 118.

DURAIRAJ, P., JUNG, E., PARK, H. H., KIM, B. G. & YUN, H. 2015. Comparative functional characterization of a novel benzoate hydroxylase cytochrome P450 of *Fusarium oxysporum*. *Enzyme Microb Technol,* 70**,** 58-65.

ENGELTHALER, D. M. & MEYER, W. 2017. Furthering the continental drift speciation hypothesis in the pathogenic Cryptococcus species complexes. *Msphere,* 2.

FABER, B. W., VAN GORCOM, R. F. & DUINE, J. A. 2001. Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. *Arch Biochem Biophys,* 394**,** 245-54.

FARRER, R. A., DESJARDINS, C. A., SAKTHIKUMAR, S., GUJJA, S., SAIF, S., ZENG, Q., CHEN, Y., VOELZ, K., HEITMAN, J., MAY, R. C., FISHER, M. C. & CUOMO, C. A. 2015. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. *MBio,* 6**,** e00868-15.

FELDMESSER, M., RIVERA, J., KRESS, Y., KOZEL, T. R. & CASADEVALL, A. 2000. Antibody interactions with the capsule of Cryptococcus neoformans. *Infection and Immunity,* 68**,** 3642-3650.

FLOUDAS, D., BINDER, M., RILEY, R., BARRY, K., BLANCHETTE, R. A., HENRISSAT, B., MARTÍNEZ, A. T., OTILLAR, R., SPATAFORA, J. W. & YADAV, J. S. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science,* 336**,** 1715-1719.

GAONA-FLORES, V. A. 2013. Central nervous system and Cryptococcus neoformans. *North American Journal of Medical Sciences,* 5**,** 492.

GOTOH, O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry,* 267**,** 83-90.

GRONT, D., HANSMANN, U. H. & KOLINSKI, A. 2005. Exploring protein energy landscapes with hierarchical clustering. *International Journal of Quantum Chemistry,* 105**,** 826-830.

GRONT, D. & KOLINSKI, A. 2005. HCPM — Program for hierarchical clustering of protein models. *Bioinformatics,* 21**,** 3179-3180.

GRONT, D. & KOLINSKI, A. 2006. BioShell — A package of tools for structural biology computations. *Bioinformatics,* 22**,** 621-622.

HAGEN, F., ILLNAIT-ZARAGOZI, M.-T., BARTLETT, K. H., SWINNE, D., GEERTSEN, E., KLAASSEN, C. H., BOEKHOUT, T. & MEIS, J. F. 2010. In vitro antifungal susceptibilities and amplified fragment length polymorphism genotyping of a worldwide collection of 350 clinical, veterinary, and environmental Cryptococcus gattii isolates. *Antimicrobial Agents and Chemotherapy,* 54**,** 5139-5145.

HANSON, K. E., CATANIA, J., ALEXANDER, B. D. & PERFECT, J. R. 2017. Drug resistance in Cryptococcosis. *In:* MAYERS, D. L., SOBEL, J. D., OUELLETTE, M., KAYE, K. S. & MARCHAIM, D. (eds.) *Antimicrobial Drug Resistance: Clinical and Epidemiological Aspects, Volume 2.* Cham: Springer International Publishing.

HOEKSTRA, W. J., GARVEY, E. P., MOORE, W. R., RAFFERTY, S. W., YATES, C. M. & SCHOTZINGER, R. J. 2014. Design and optimization of highly-selective fungal CYP51 inhibitors. *Bioorganic & Medicinal Chemistry Letters,* 24**,** 3455-3458.

HSUEH, P.-R., LAU, Y.-J., CHUANG, Y.-C., WAN, J.-H., HUANG, W.-K., SHYR, J.-M., YAN, J.-J., YU, K.-W., WU, J.-J. & KO, W.-C. 2005. Antifungal susceptibilities of clinical isolates of Candida species, Cryptococcus neoformans, and Aspergillus species from Taiwan: Surveillance of multicenter antimicrobial resistance in Taiwan program data from 2003. *Antimicrobial Agents and Chemotherapy,* 49**,** 512-517.

ILLNAIT-ZARAGOZI, M.-T., MARTÍNEZ, G. F., CURFS-BREUKER, I., FERNÁNDEZ, C. M., BOEKHOUT, T. & MEIS, J. F. 2008. In vitro activity of the new azole isavuconazole (BAL4815) compared with six other antifungal agents against 162 Cryptococcus neoformans isolates from Cuba. *Antimicrobial Agents and Chemotherapy,* 52**,** 1580-1582.

JAWALLAPERSAND, P., MASHELE, S. S., KOVACIC, L., STOJAN, J., KOMEL, R., PAKALA, S. B., KRASEVEC, N. & SYED, K. 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: Comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. *PLoS One,* 9**,** e107209.

JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci,* 8**,** 275-82.

KANO, R., OKUBO, M., HASEGAWA, A. & KAMATA, H. 2017. Multi-azole-resistant strains of Cryptococcus neoformans var. grubii isolated from a FLZ-resistant strain by culturing in medium containing voriconazole. *Medical Mycology,* 55**,** 877-882.

KATOH, K., KUMA, K., TOH, H. & MIYATA, T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res,* 33**,** 511-8.

KELLY, S. L. & KELLY, D. E. 2013. Microbial cytochromes P450: Biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20120476.

KELLY, S. L., LAMB, D. C., BALDWIN, B. C., CORRAN, A. J. & KELLY, D. E. 1997. Characterization of *Saccharomyces cerevisiae* CYP61, sterol delta22-desaturase, and inhibition by azole antifungal agents. *J Biol Chem,* 272**,** 9986-8.

KELLY, S. L., LAMB, D. C. & KELLY, D. E. 1999. Inhibitors of CYP51 as antifungal agents and resistance to azole antifungals. *Molecular and Applied Aspects of Oxidative Drug Metabolizing Enzymes.* Springer.

KGOSIEMANG, I. K. R., SYED, K. & MASHELE, S. S. 2014. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum *Saccharomycotina*. *J Pure Appl Microbiol,* 8**,** 12.

KHAWCHAROENPORN, T., APISARNTHANARAK, A. & MUNDY, L. M. 2007. Non-neoformans cryptococcal infections: A systematic review. *Infection,* 35**,** 51.

KIELSTEIN, P., HOTZEL, H., SCHMALRECK, A., KHASCHABI, D. & GLAWISCHNIG, W. 2000. Occurrence of Cryptococcus spp. in excreta of pigeons and pet birds. *Mycoses,* 43**,** 7-15.

KOEHLER LEMAN, J., WEITZNER, B. D., RENFREW, P. D., LEWIS, S. M., MORETTI, R., WATKINS, A. M., MULLIGAN, V. K., LYSKOV, S., ADOLF-BRYFOGLE, J. & LABONTE, J. W. 2020. Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Computational Biology,* 16**,** e1007507.

KOTHIWALE, S., MENDENHALL, J. L. & MEILER, J. 2015. BCL:: C onf: Small molecule conformational sampling using a knowledge based rotamer library. *Journal of Cheminformatics,* 7**,** 1-15.

KOURIST, R., BRACHARZ, F., LORENZEN, J., KRACHT, O. N., CHOVATIA, M., DAUM, C., DESHPANDE, S., LIPZEN, A., NOLAN, M., OHM, R. A., GRIGORIEV, I. V., SUN, S., HEITMAN, J., BRUCK, T. & NOWROUSIAN, M. 2015. Genomics and transcriptomics analyses of the oil-accumulating basidiomycete yeast *Trichosporon oleaginosus*: Insights into substrate utilization and alternative evolutionary trajectories of fungal mating systems. *MBio,* 6**,** e00918.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol,* 33**,** 1870-4.

KUO, A., SALAMOV, A., KORZENIEWSKI, F., NORDBERG, H., SHABALOV, I., DUBCHAK, I., OTILLAR, R., RILEY, R., OHM, R., NIKITIN, R., HARIDAS, S., SMIRNOVA, T., ZHAO, X. & GRIGORIEV, I. V. 2013. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research,* 42**,** D699-D704.

KUSHIMA, H., TOKIMATSU, I., ISHII, H., KAWANO, R., SHIRAI, R., KISHI, K., HIRAMATSU, K. & KADOTA, J. 2012. Cloning of the lanosterol 14-alpha-demethylase (ERG11) gene in *Trichosporon asahii*: A possible association between G453R amino acid substitution and azole resistance in *T. asahii*. *FEMS Yeast Res,* 12**,** 662-7.

KUSHIMA, H., TOKIMATSU, I., ISHII, H., KAWANO, R., WATANABE, K. & KADOTA, J. I. 2017. A new amino acid substitution at G150S in lanosterol 14-alpha eemethylase (Erg11 protein) in multi-azole-resistant *Trichosporon asahii*. *Med Mycol J,* 58**,** E23-e28.

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**,** a019760.

LAMB, D., CORRAN, A., BALDWIN, B. C., KWON-CHUNG, J. & KELLY, S. 1995. Resistant P45051A1 activity in azole antifungal tolerant Cryptococcus neoformans from AIDS patients. *FEBS Letters,* 368**,** 326-330.

LE, K., ADOLF-BRYFOGLE, J., KLIMA, J., LYSKOV, S., LABONTE, J., BERTOLANI, S., BURMAN, S. R., LEAVER-FAY, A., WEITZNER, B. & MAGUIRE, J. 2020. PyRosetta Jupyter Notebooks Teach Biomolecular Structure Prediction and Design. *Preprints*, 2020020097 (doi: 10.20944/preprints202002.0097.v1).

LEMAN, J. K., WEITZNER, B. D., LEWIS, S. M., ADOLF-BRYFOGLE, J., ALAM, N., ALFORD, R. F., APRAHAMIAN, M., BAKER, D., BARLOW, K. A. & BARTH, P. 2020. Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nature Methods***,** 1-14.

LEMMON, G. & MEILER, J. 2012. Rosetta ligand docking with flexible XML protocols. *Computational Drug Discovery and Design.* Methods in Molecular Biology, P143-155, Springer. DOI: 10.1007/978-1-61779-465-0_10.

LEPESHEVA, G. I., FRIGGERI, L. & WATERMAN, M. R. 2018. CYP51 as drug targets for fungi and protozoan parasites: Past, present and future. *Parasitology,* 145**,** 1820-1836.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

LITTMAN, M. L. 1959. Cryptococcosis (Torulosis): Current concepts and therapy. *The American Journal of Medicine,* 27**,** 976-998.

LITVINTSEVA, A. P. & MITCHELL, T. G. 2012. Population genetic analyses reveal the African origin and strain variation of Cryptococcus neoformans var. grubii. *PLoS Pathog,* 8**,** e1002495.

LITVINTSEVA, A. P., THAKUR, R., RELLER, L. B. & MITCHELL, T. G. 2005. Prevalence of clinical isolates of Cryptococcus gattii serotype C among patients with AIDS in Sub-Saharan Africa. *The Journal of Infectious Diseases,* 192**,** 888-892.

LOCKHART, S. R., FOTHERGILL, A. W., IQBAL, N., BOLDEN, C. B., GROSSMAN, N. T., GARVEY, E. P., BRAND, S. R., HOEKSTRA, W. J., SCHOTZINGER, R. J. & OTTINGER, E. 2016a. The investigational fungal Cyp51 inhibitor VT-1129 demonstrates potent in vitro activity against Cryptococcus neoformans and Cryptococcus gattii. *Antimicrobial Agents and Chemotherapy,* 60**,** 2528-2531.

LOCKHART, S. R., FOTHERGILL, A. W., IQBAL, N., BOLDEN, C. B., GROSSMAN, N. T., GARVEY, E. P., BRAND, S. R., HOEKSTRA, W. J., SCHOTZINGER, R. J., OTTINGER, E., PATTERSON, T. F. & WIEDERHOLD, N. P. 2016b. The investigational fungal Cyp51 inhibitor VT-1129 demonstrates potent *in vitro* activity against *Cryptococcus neoformans* and *Cryptococcus gattii*. *Antimicrob Agents Chemother,* 60**,** 2528-31.

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., PERTEA, 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**,** 1321-4.

LOPERENA-ALVAREZ, Y., REN, P., LI, X., BOPP, D., RUIZ, A., CHATURVEDI, V. & RIOS-VELAZQUEZ, C. 2010. Genotypic characterization of environmental isolates of Cryptococcus gattii from Puerto Rico. *Mycopathologia,* 170**,** 279-285.

LOYSE, A., THANGARAJ, H., EASTERBROOK, P., FORD, N., ROY, M., CHILLER, T., GOVENDER, N., HARRISON, T. S. & BICANIC, T. 2013. Cryptococcal meningitis: improving access to essential antifungal medicines in resource-poor countries. *The Lancet Infectious Diseases,* 13**,** 629-637.

MA, H., HAGEN, F., STEKEL, D. J., JOHNSTON, S. A., SIONOV, E., FALK, R., POLACHECK, I., BOEKHOUT, T. & MAY, R. C. 2009. The fatal fungal outbreak on Vancouver Island is characterized by enhanced intracellular parasitism driven by mitochondrial regulation. *Proceedings of the National Academy of Sciences,* 106**,** 12980-12985.

MACNAR, J. M., SZULC, N. A., KRYŚ, J. D., BADACZEWSKA-DAWID, A. E. & GRONT, D. 2020. BioShell 3.0: Library for Processing Structural Biology Data. *Biomolecules,* 10**,** 461.

MARCHLER-BAUER, A., BO, Y., HAN, L., HE, J., LANCZYCKI, C. J., LU, S., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., LU, F., MARCHLER, G. H., SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C., GEER, L. Y. & BRYANT, S. H. 2017. CDD/SPARCLE: FFunctional classification of proteins *via* subfamily domain architectures. *Nucleic Acids Res,* 45**,** D200-d203.

MATOWANE, R. G., WIETESKA, L., BAMAL, H. D., KGOSIEMANG, I. K. R., VAN WYK, M., MANUME, N. A., ABDALLA, S. M. H., MASHELE, S. S., GRONT, D. & SYED, K. 2018. *In silico* analysis of cytochrome P450 monooxygenases in chronic granulomatous infectious fungus *Sporothrix schenckii*: Special focus on CYP51. *Biochim Biophys Acta Proteins Proteom,* 1866**,** 166-177.

MATSUZAKI, F. & WARIISHI, H. 2005. Molecular characterization of cytochrome P450 catalyzing hydroxylation of benzoates from the white-rot fungus *Phanerochaete chrysosporium*. *Biochem Biophys Res Commun,* 334**,** 1184-90.

MAY, R. C., STONE, N. R., WIESNER, D. L., BICANIC, T. & NIELSEN, K. 2016. *Cryptococcus*: From environmental saprophyte to global pathogen. *Nat Rev Microbiol,* 14**,** 106-17.

MEILER, J. & BAKER, D. 2006. ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins: Structure, Function, and Bioinformatics,* 65**,** 538-548.

MILANOVIĆ, V., COMITINI, F. & CIANI, M. 2013. Grape berry yeast communities: Influence of fungicide treatments. *International Journal of Food Microbiology,* 161**,** 240-246.

MINGOT, J. M., PENALVA, M. A. & FERNANDEZ-CANON, J. M. 1999. Disruption of *phacA,* an *Aspergillus nidulans* gene encoding a novel cytochrome P450 monooxygenase catalyzing phenylacetate 2-hydroxylation, results in penicillin overproduction. *J Biol Chem,* 274**,** 14545-50.

MONDO, S. J., DANNEBAUM, R. O., KUO, R. C., LOUIE, K. B., BEWICK, A. J., LABUTTI, K., HARIDAS, S., KUO, A., SALAMOV, A., AHRENDT, S. R., LAU, R., BOWEN, B. P., LIPZEN, A., SULLIVAN, W., ANDREOPOULOS, B. B., CLUM, A., LINDQUIST, E., DAUM, C., NORTHEN, T. R., KUNDE-RAMAMOORTHY, G., SCHMITZ, R. J., GRYGANSKYI, A., CULLEY, D., MAGNUSON, J., JAMES, T. Y., O'MALLEY, M. A., STAJICH, J. E., SPATAFORA, J. W., VISEL, A. & GRIGORIEV, I. V. 2017. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet,* 49**,** 964-968.

MONK, B. C., TOMASIAK, T. M., KENIYA, M. V., HUSCHMANN, F. U., TYNDALL, J. D., O'CONNELL, J. D., CANNON, R. D., MCDONALD, J. G., RODRIGUEZ, A. & FINER-MOORE, J. S. 2014. Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proceedings of the National Academy of Sciences,* 111**,** 3865-3870.

MTHETHWA, B. C., CHEN, W., NGWENYA, M. L., KAPPO, A. P., SYED, P. R., KARPOORMATH, R., YU, J. H., NELSON, D. R. & SYED, K. 2018. Comparative analyses of cytochrome P450s and those associated with secondary metabolism in *Bacillus* species. *Int J Mol Sci,* 19.

NAKAHARA, K., TANIMOTO, T., HATANO, K., USUDA, K. & SHOUN, H. 1993. Cytochrome P-450 55A1 (P-450dNIR) acts as nitric oxide reductase employing NADH as the direct electron donor. *J Biol Chem,* 268**,** 8350-5.

NAKAYAMA, N., TAKEMAE, A. & SHOUN, H. 1996. Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum*. *J Biochem,* 119**,** 435-40.

NELSON, D. R. 1998. Cytochrome P450 nomenclature. *Methods Mol Biol,* 107**,** 15-24.

NELSON, D. R. 2006. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol,* 320**,** 1-10.

NELSON, D. R. 2009. The cytochrome p450 homepage. *Hum Genomics,* 4**,** 59-65.

NGAMSKULRUNGROJ, P., CHANG, Y., HANSEN, B., BUGGE, C., FISCHER, E. & KWON-CHUNG, K. J. 2012. Characterization of the chromosome 4 genes that affect fluconazole-induced disomy formation in *Cryptococcus neoformans*. *PLoS One,* 7**,** e33022.

NGCOBO, N. S., CHILIZA, Z. E., CHEN, W., YU, J.-H., NELSON, D. R., TUSZYNSKI, J. A., PRETO, J. & SYED, K. 2020. Comparative analysis, structural insights, and substrate/drug interaction of CYP128A1 in *Mycobacterium tuberculosis*. *International Journal of Molecular Sciences,* 21.

NGWENYA, M. L., CHEN, W., BASSON, A. K., SHANDU, J. S., YU, J. H., NELSON, D. R. & SYED, K. 2018. Blooming of unusual cytochrome P450s by tandem duplication in the pathogenic fungus *Conidiobolus coronatus*. *Int J Mol Sci,* 19.

NIELSEN, K., VEDULA, P., SMITH, K. D., MEYA, D. B., GARVEY, E. P., HOEKSTRA, W. J., SCHOTZINGER, R. J. & BOULWARE, D. R. 2017a. Activity of VT-1129 against *Cryptococcus neoformans* clinical isolates with high fluconazole MICs. *Med Mycol,* 55**,** 453-456.

NIELSEN, K., VEDULA, P., SMITH, K. D., MEYA, D. B., GARVEY, E. P., HOEKSTRA, W. J., SCHOTZINGER, R. J. & BOULWARE, D. R. 2017b. Activity of VT-1129 against Cryptococcus neoformans clinical isolates with high fluconazole MICs. *Medical mycology,* 55**,** 453-456.

NOWICKA, J., NAWROT, U., HAUS, O., KULICZKOWSKI, K., FONTEYNE, P.-A. & NOLARD, N. 2007. *Cryptococcus curvatus* in peritoneal fluid of gastric lymphoma patient with complex chromosome aberrations—case report. *Medical Mycology/Mikologia,* 14.

O'BOYLE, N. M., BANCK, M., JAMES, C. A., MORLEY, C., VANDERMEERSCH, T. & HUTCHISON, G. R. 2011. Open Babel: An open chemical toolbox. *Journal of Cheminformatics,* 3**,** 33.

OKAGAKI, L. H., STRAIN, A. K., NIELSEN, J. N., CHARLIER, C., BALTES, N. J., CHRÉTIEN, F., HEITMAN, J., DROMER, F. & NIELSEN, K. 2010. Cryptococcal cell morphology affects host cell interactions and pathogenicity. *PLoS Pathog,* 6**,** e1000953.

PAN, W., KHAYHAN, K., HAGEN, F., WAHYUNINGSIH, R., CHAKRABARTI, A., CHOWDHARY, A., IKEDA, R., TAJ-ALDEEN, S. J., KHAN, Z. & IMRAN, D. 2012.

Resistance of Asian Cryptococcus neoformans serotype A is confined to few microsatellite genotypes. *PloS One,* 7**,** e32868.

PARVEZ, M., QHANYA, L. B., MTHAKATHI, N. T., KGOSIEMANG, I. K., BAMAL, H. D., PAGADALA, N. S., XIE, T., YANG, H., CHEN, H., THERON, C. W., MONYAKI, R., RASELEMANE, S. C., SALEWE, V., MONGALE, B. L., MATOWANE, R. G., ABDALLA, S. M., BOOI, W. I., VAN WYK, M., OLIVIER, D., BOUCHER, C. E., NELSON, D. R., TUSZYNSKI, J. A., BLACKBURN, J. M., YU, J. H., MASHELE, S. S., CHEN, W. & SYED, K. 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci Rep,* 6**,** 33099.

PASSER, A. R., COELHO, M. A., BILLMYRE, R. B., NOWROUSIAN, M., MITTELBACH, M., YURKOV, A. M., AVERETTE, A. F., CUOMO, C. A., SUN, S. & HEITMAN, J. 2019. Genetic and genomic analyses reveal boundaries between species closely related to *Cryptococcus* pathogens. *bioRxiv***,** 557140.

PERFECT, J. R., DISMUKES, W. E., DROMER, F., GOLDMAN, D. L., GRAYBILL, J. R., HAMILL, R. J., HARRISON, T. S., LARSEN, R. A., LORTHOLARY, O. & NGUYEN, M.-H. 2010a. Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. *Clinical Infectious Diseases,* 50**,** 291-322.

PERFECT, J. R., DISMUKES, W. E., DROMER, F., GOLDMAN, D. L., GRAYBILL, J. R., HAMILL, R. J., HARRISON, T. S., LARSEN, R. A., LORTHOLARY, O., NGUYEN, M. H., PAPPAS, P. G., POWDERLY, W. G., SINGH, N., SOBEL, J. D. & SORRELL, T. C. 2010b. Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the infectious diseases society of america. *Clin Infect Dis,* 50**,** 291-322.

PODUST, L. M., STOJAN, J., POULOS, T. L. & WATERMAN, M. R. 2001. Substrate recognition sites in 14α-sterol demethylase from comparative analysis of amino acid sequences and X-ray structure of Mycobacterium tuberculosis CYP51. *Journal of Inorganic Biochemistry,* 87**,** 227-235.

QHANYA, L. B., MATOWANE, G., CHEN, W., SUN, Y., LETSIMO, E. M., PARVEZ, M., YU, J. H., MASHELE, S. S. & SYED, K. 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete biotrophic plant pathogens. *PLoS One,* 10**,** e0142100.

RAJASINGHAM, R., SMITH, R. M., PARK, B. J., JARVIS, J. N., GOVENDER, N. P., CHILLER, T. M., DENNING, D. W., LOYSE, A. & BOULWARE, D. R. 2017. Global

burden of disease of HIV-associated cryptococcal meningitis: An updated analysis. *Lancet Infect Dis,* 17**,** 873-881.

REIS, M., CARVALHO, C., ANDRADE, F., FERNANDES, O. D. F. L., ARRUDA, W. & SILVA, M. 2016. Fisetin as a promising antifungal agent against Cryptococococcus neoformans species complex. *Journal of Applied Microbiology,* 121**,** 373-379.

REVANKAR, S. G., FU, J., RINALDI, M. G., KELLY, S. L., KELLY, D. E., LAMB, D. C., KELLER, S. M. & WICKES, B. L. 2004. Cloning and characterization of the lanosterol 14alpha-demethylase (ERG11) gene in *Cryptococcus neoformans*. *Biochem Biophys Res Commun,* 324**,** 719-28.

RODERO, L., MELLADO, E., RODRIGUEZ, A. C., SALVE, A., GUELFAND, L., CAHN, P., CUENCA-ESTRELLA, M., DAVEL, G. & RODRIGUEZ-TUDELA, J. L. 2003. G484S amino acid substitution in lanosterol 14-alpha demethylase (ERG11) is related to fluconazole resistance in a recurrent *Cryptococcus neoformans* clinical isolate. *Antimicrob Agents Chemother,* 47**,** 3653-6.

ROSE, A. S. & HILDEBRAND, P. W. 2015. NGL Viewer: A web application for molecular visualization. *Nucleic Acids Research,* 43**,** W576-W579.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374-8.

SAR, B., MONCHY, D., VANN, M., KEO, C., SARTHOU, J. L. & BUISSON, Y. 2004. Increasing in vitro resistance to fluconazole in Cryptococcus neoformans Cambodian isolates: April 2000 to March 2002. *Journal of Antimicrobial Chemotherapy,* 54**,** 563-565.

SCHMIDT, S. K., VIMERCATI, L., DARCY, J. L., ARÁN, P., GENDRON, E. M., SOLON, A. J., PORAZINSKA, D. & DORADOR, C. 2017. A Naganishia in high places: Functioning populations or dormant cells from the atmosphere? *Mycology,* 8**,** 153-163.

SELLO, M. M., JAFTA, N., NELSON, D. R., CHEN, W., YU, J. H., PARVEZ, M., KGOSIEMANG, I. K., MONYAKI, R., RASELEMANE, S. C., QHANYA, L. B., MTHAKATHI, N. T., SITHENI MASHELE, S. & SYED, K. 2015. Diversity and evolution of cytochrome P450 monooxygenases in *Oomycetes*. *Sci Rep,* 5**,** 11572.

SENATE, L. M., TJATJI, M. P., PILLAY, K., CHEN, W., ZONDO, N. M., SYED, P. R., MNGUNI, F. C., CHILIZA, Z. E., BAMAL, H. D., KARPOORMATH, R., KHOZA, T., MASHELE, S. S., BLACKBURN, J. M., YU, J. H., NELSON, D. R. & SYED, K. 2019. Similarities, variations, and evolution of cytochrome P450s in *Streptomyces* versus *Mycobacterium*. *Sci Rep,* 9**,** 3962.

SHENG, C., MIAO, Z., JI, H., YAO, J., WANG, W., CHE, X., DONG, G., LÜ, J., GUO, W. & ZHANG, W. 2009. Three-dimensional model of lanosterol 14α-demethylase from Cryptococcus neoformans: Active-site characterization and insights into azole binding. *Antimicrobial agents and chemotherapy,* 53**,** 3487-3495.

SHOUN, H., FUSHINOBU, S., JIANG, L., KIM, S. W. & WAKAGI, T. 2012. Fungal denitrification and nitric oxide reductase cytochrome P450nor. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1186-94.

SIONOV, E., CHANG, Y. C., GARRAFFO, H. M., DOLAN, M. A., GHANNOUM, M. A. & KWON-CHUNG, K. J. 2012. Identification of a *Cryptococcus neoformans* cytochrome P450 lanosterol 14alpha-demethylase (Erg11) residue critical for differential susceptibility between fluconazole/voriconazole and itraconazole/posaconazole. *Antimicrob Agents Chemother,* 56**,** 1162-9.

SMITH, K. D., ACHAN, B., HULLSIEK, K. H., MCDONALD, T. R., OKAGAKI, L. H., ALHADAB, A. A., AKAMPURIRA, A., RHEIN, J. R., MEYA, D. B. & BOULWARE, D. R. 2015. Increased antifungal drug resistance in clinical isolates of Cryptococcus neoformans in Uganda. *Antimicrobial Agents and Chemotherapy,* 59**,** 7197-7204.

SNYDMAN, D. R., SINGH, N., DROMER, F., PERFECT, J. R. & LORTHOLARY, O. 2008. Cryptococcosis in solid organ transplant recipients: Current state of the science. *Clinical Infectious Diseases,* 47**,** 1321-1327.

SORRELL, T. C., BROWNLEE, A. G., RUMA, P., MALIK, R., PFEIFFER, T. J. & ELLIS, D. H. 1996. Natural environmental sources of Cryptococcus neoformans var. gattii. *Journal of Clinical Microbiology,* 34**,** 1261-1263.

SRIKANTA, D., SANTIAGO-TIRADO, F. H. & DOERING, T. L. 2014. Cryptococcus neoformans: Historical curiosity to modern pathogen. *Yeast,* 31**,** 47-60.

SUN, S., YADAV, V., BILLMYRE, R. B., CUOMO, C. A., NOWROUSIAN, M., WANG, L., SOUCIET, J. L., BOEKHOUT, T., PORCEL, B., WINCKER, P., GRANEK, J. A., SANYAL, K. & HEITMAN, J. 2017. Fungal genome and mating system transitions

facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol,* 15**,** e2002527.

SYED, K., SHALE, K., PAGADALA, N. S. & TUSZYNSKI, J. 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One,* 9**,** e86683.

SYED, P. R., CHEN, W., NELSON, D. R., KAPPO, A. P., YU, J. H., KARPOORMATH, R. & SYED, K. 2019. Cytochrome P450 Monooxygenase CYP139 Family Involved in the Synthesis of Secondary Metabolites in 824 Mycobacterial Species. *Int J Mol Sci,* 20.

TANIMURA, A., TAKASHIMA, M., SUGITA, T., ENDOH, R., KIKUKAWA, M., YAMAGUCHI, S., SAKURADANI, E., OGAWA, J., OHKUMA, M. & SHIMA, J. 2014. *Cryptococcus terricola* is a promising oleaginous yeast for biodiesel production from starch through consolidated bioprocessing. *Sci Rep,* 4**,** 4776.

THOMPSON, G. R., WIEDERHOLD, N. P., FOTHERGILL, A. W., VALLOR, A. C., WICKES, B. L. & PATTERSON, T. F. 2009. Antifungal susceptibilities among different serotypes of Cryptococcus gattii and Cryptococcus neoformans. *Antimicrobial Agents and Chemotherapy,* 53**,** 309-311.

TOUIL, Y. S., AUZEIL, N., BOULINGUEZ, F., SAIGHI, H., REGAZZETTI, A., SCHERMAN, D. & CHABOT, G. G. 2011. Fisetin disposition and metabolism in mice: Identification of geraldol as an active metabolite. *Biochemical pharmacology,* 82**,** 1731-1739.

TRILLES, L., FERNÁNDEZ-TORRES, B., DOS SANTOS LAZÉRA, M., WANKE, B. & GUARRO, J. 2004. In vitro antifungal susceptibility of Cryptococcus gattii. *Journal of Clinical Microbiology,* 42**,** 4815-4817.

TRILLES, L., MEYER, W., WANKE, B., GUARRO, J. & LAZÉRA, M. 2012. Correlation of antifungal susceptibility and molecular type within the Cryptococcus neoformans/C. gattii species complex. *Medical Mycology,* 50**,** 328-332.

VECHI, H. T., THEODORO, R. C., DE OLIVEIRA, A. L., DA SILVA GOMES, R. M. O., DE ALMEIDA SOARES, R. D., FREIRE, M. G. & BAY, M. B. 2019. Invasive fungal infection by Cryptococcus neoformans var. grubii with bone marrow and meningeal involvement in a HIV-infected patient: a case report. *BMC Infectious Diseases,* 19**,** 1-8.

VISHNIAC, H. S. & HEMPFLING, W. P. 1979. *Cryptococcus vishniacii* sp. nov., an Antarctic yeast. *International Journal of Systematic and Evolutionary Microbiology,* 29**,** 153-158.

WANG, H., XIAO, M., CHEN, S. C.-A., KONG, F., SUN, Z.-Y., LIAO, K., LU, J., SHAO, H.-F., YAN, Y. & FAN, H. 2012. In vitro susceptibilities of yeast species to fluconazole and voriconazole as determined by the 2010 National China Hospital Invasive Fungal Surveillance Net (CHIF-NET) study. *Journal of Clinical Microbiology,* 50**,** 3952-3959.

WARRILOW, A., HULL, C., PARKER, J., GARVEY, E., HOEKSTRA, W., MOORE, W., SCHOTZINGER, R., KELLY, D. & KELLY, S. 2014. The clinical candidate VT-1161 is a highly potent inhibitor of Candida albicans CYP51 but fails to bind the human enzyme. *Antimicrobial Agents and Chemotherapy,* 58**,** 7121-7127.

WARRILOW, A. G. S., PARKER, J. E., PRICE, C. L., NES, W. D., GARVEY, E. P., HOEKSTRA, W. J., SCHOTZINGER, R. J., KELLY, D. E. & KELLY, S. L. 2016. The investigational drug VT-1129 is a highly potent inhibitor of *Cryptococcus* species CYP51 but only weakly inhibits the human enzyme. *Antimicrobial Agents and Chemotherapy,* 60**,** 4530-4538.

WEBB, B. & SALI, A. 2016. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics,* 54**,** 5.6.1-5.6.37.

WHO 2018. Guidelines for the diagnosis, prevention, and management of cryptococcal disease in HIV-infected adults, adolescents and children, March 2018: Supplement to the 2016 consolidated guidelines of the use of antiretroviral drugs for treating and preventing HIV infection. Switzerland, Geneva.

WHO 2019. Cryptococcal disease: What's new and important. World Health Organization (WHO). Available online: https://www.who.int/hiv/mediacentre/news/cryptococcal-disease-key-messages/en/ (accessed on 8 April, 2019).

WILLIAMSON, P. R., JARVIS, J. N., PANACKAL, A. A., FISHER, M. C., MOLLOY, S. F., LOYSE, A. & HARRISON, T. S. 2017. Cryptococcal meningitis: Epidemiology, immunology, diagnosis and therapy. *Nature Reviews Neurology,* 13**,** 13.

XU, Y., WANG, S., HU, Q., GAO, S., MA, X., ZHANG, W., SHEN, Y., CHEN, F., LAI, L. & PEI, J. 2018. CavityPlus: A web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Research,* 46**,** W374-W379.

YANG, R., AO, J., WANG, W., SONG, K., LI, R. & WANG, D. 2003. Disseminated trichosporonosis in China. *Mycoses,* 46**,** 519-23.

YANG, R. Y., LI, H. T., ZHU, H., ZHOU, G. P., WANG, M. & WANG, L. 2012. Genome sequence of the Trichosporon asahii environmental strain CBS 8904. *Eukaryot Cell,* 11**,** 1586-7.

ZARAGOZA, O., RODRIGUES, M. L., DE JESUS, M., FRASES, S., DADACHOVA, E. & CASADEVALL, A. 2009. The capsule of the fungal pathogen Cryptococcus neoformans. *Advances in Applied Microbiology,* 68**,** 133-216.

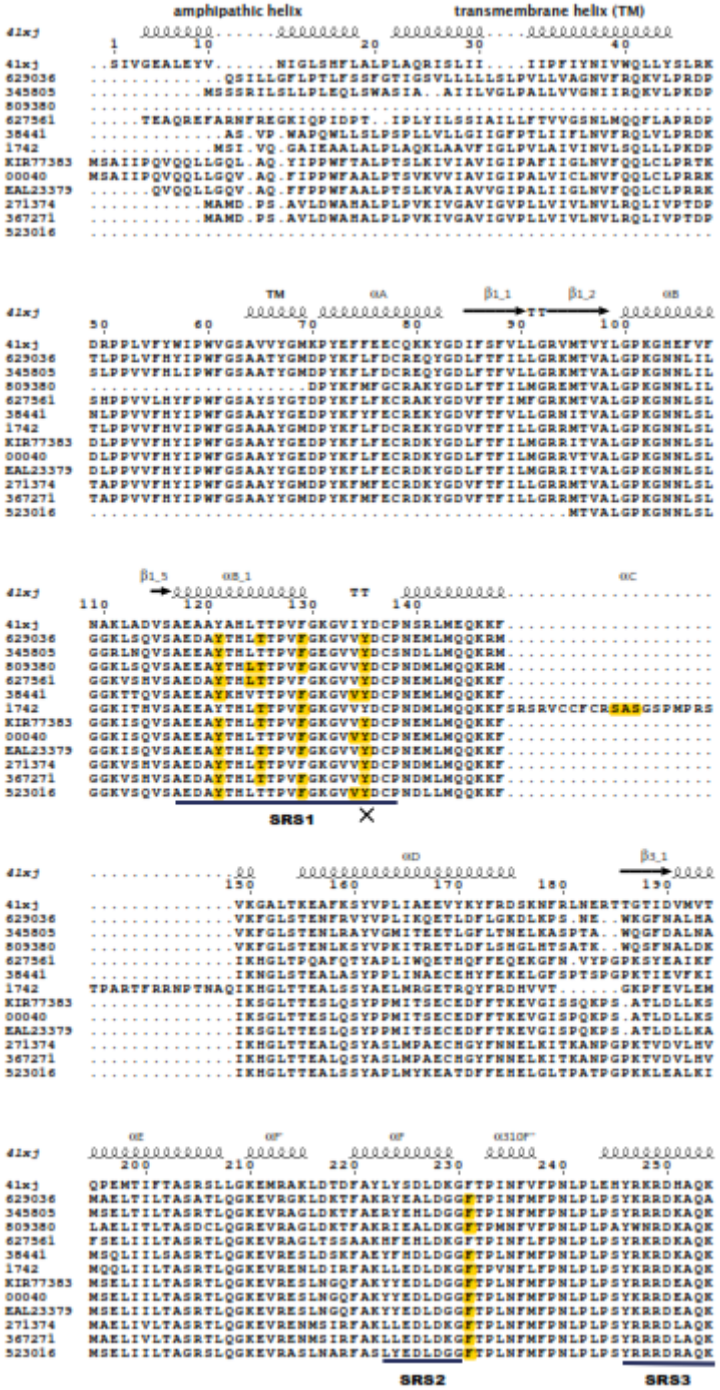
ZHANG, J., LI, L., LV, Q., YAN, L., WANG, Y. & JIANG, Y. 2019. The fungal CYP51s: Their functions, structures, related drug resistance, and inhibitors. *Frontiers in Microbiology,* 10**,** 691.

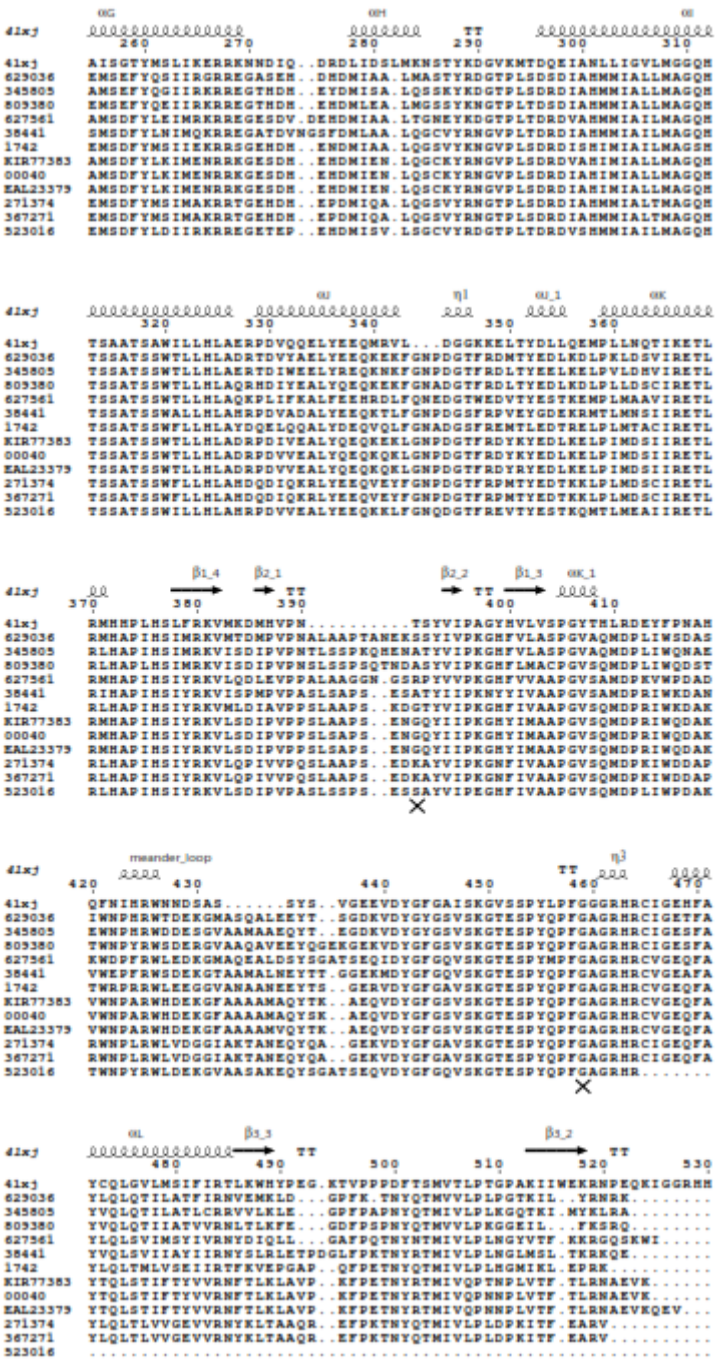
ZIMMERMANN, L., STEPHENS, A., NAM, S.-Z., RAU, D., KÜBLER, J., LOZAJIC, M., GABLER, F., SÖDING, J., LUPAS, A. N. & ALVA, V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of Molecular Biology,* 430**,** 2237-2243.

APPENDICES

APPENDIX A

Structural alignment of *Tremellomycetes* CYP51s with template 4LXJ. P450 characteristic notation for α-helices and β strands and substrate recognition sequences (SRS) were mapped based on the template. Residues interacting with ≥5 ligands highlighted in yellow. Residues playing a role in azole drug resistance are indicated with "X".




αG
αH
αI
4lxj
260
270
280
290
300
310
4lxj AISGTYMSLIKERRKNNDIQ..DRDLIDSLMKNSTYKDGVKMTDQEIANLLIGVLMGGQH
629036 EMSEFYQSIIRGRREGASEM..DHDMIAA.LMASTYRDGTPLSDSDIAHMMIALLMAGQH
345805 AMSEFYQGIIRKRREGTHDM..EYDMISA.LQSSKYKDGTPLSDRDIAHMMIALLMAGQH
809380 EMSEFYQEIIRKRREGTHDM..EHDMLEA.LMGSSYKNGTPLTDSDIAHMMIALLMAGQH
627561 AMSDFYLEIMRKRREGESDV.DEHDMIAA.LTGNEYKDGTPLTDRDVAHMMIAILMAGQH
38441 SMSDFYLNIMQKRREGATDVNGSFDMLAA.LQGCVYRNGVPLTDRDIAHMMIAILMAGQH
i742 EMSDFYMSIIEKRRSGEHDM..ENDMIAA.LQGSVYKNGVPLSDRDISHIMIAILMAGSH
KIR77383 AMSDFYLKIMENRRKGESDM..EHDMIEN.LQGCKYRNGVPLSDRDVAHIMIALLMAGQH
00040 AMSDFYLKIMENRRKGESDM..EHDMIEN.LQSCKYRNGVPLSDRDIAHIMIALLMAGQH
EAL23379 AMSDFYLKIMENRRKGESDM..EHDMIEN.LQSCKYRNGVPLSDRDIAHIMIALLMAGQH
271374 EMSDFYMSIMAKRRTGEHDM..EPDMIQA.LQGSVYRNGTPLSDRDIAHMMIALTMAGQH
367271 EMSDFYMSIMAKRRTGEHDM..EPDMIQA.LQGSVYRNGTPLSDRDIAHMMIALTMAGQH
523016 EMSDFYLDIIRKRREGETEP..EHDMISV.LSGCVYRDGTPLTDRDVSHMMIAILMAGQH
αJ
η1
αJ_1
αK
4lxj
320
330
340
350
360
4lxj TSAATSAWILLHLAERPDVQQELYEEQMRVL...DGGKKELTYDLLQEMPLLNQTIKETL
629036 TSSATSSWTLLHLADRTDVYAELYEEQKEKFGNPDGTFRDMTYEDLKDLPKLDSVIRETL
345805 TSSATSSWTLLHLAERTDIWEELYREQKNKFGNPDGTFRDLTYEELKELPVLDHVIRETL
809380 TSSATSSWTLLHLAQRHDIYEALYQEQKEKFGNADGTFRDLTYEDLKDLPLLDSCIRETL
627561 TSSATSSWTLLHLAQKPLIFKALFEEHRDLFQNEDGTWEDVTYESTKEMPLMAAVIRETL
38441 TSSATSSWALLHLAHRPDVADALYEEQKTLFGNPDGSFRPVEYGDEKRMTLMNSIIRETL
i742 TSSATSSWFLLHLAYDQELQQALYDEQVQLFGNADGSFREMTLEDTRELPLMTACIRETL
KIR77383 TSSATSSWTLLHLADRPDIVEALYQEQKEKLGNPDGTFRDYKYEDLKELPIMDSIIRETL
00040 TSSATSSWTLLHLADRPDVVEALYQEQKQKLGNPDGTFRDYKYEDLKELPIMDSIIRETL
EAL23379 TSSATSSWTLLHLADRPDVVEALYQEQKQKLGNPDGTFRDYRYEDLKELPIMDSIIRETL
271374 TSSATSSWFLLHLAHDQDIQKRLYEEQVEYFGNPDGTFRPMTYEDTKKLPLMDSCIRETL
367271 TSSATSSWFLLHLAHDQDIQKRLYEEQVEYFGNPDGTFRPMTYEDTKKLPLMDSCIRETL
523016 TSSATSSWILLHLAHRPDVVEALYEEQKKLFGNQDGTFREVTYESTKQMTLMEAIIRETL
β1_4
β2_1
β2_2
β1_3
αK_1
4lxj
370
380
390
400
410
4lxj RMHHPLHSLFRKVMKDMHVPN..........TSYVIPAGYHVLVSPGYTHLRDEYFPNAH
629036 RMHAPIHSIMRKVMTDMPVPNALAAPTANEKSSYIVPKGHFVLASPGVAQMDPLIWSDAS
345805 RLHAPIHSIMRKVISDIPVPNTLSSPKQHENATYVIPKGHFVLASPGVAQMDPLIWQNAE
809380 RLHAPLHSIMRKVISDIPVPNSLSSPSQTNDASYVIPKGHFLMACPGVSQMDPLIWQDST
627561 RMHAPIHSIYRKVLQDLEVPPALAAGGN.GSRPYVVPKGHFVVAAPGVSAMDPKVWPDAD
38441 RIHAPIHSIYRKVISPMPVPASLSAPS..ESATYIIPKNYYIVAAPGVSAMDPRIWKDAN
i742 RLHAPIHSIYRKVMLDIAVPPSLAAPS..KDGTYVIPKGHFIVAAPGVSQMDPRIWKDAK
KIR77383 RMHAPIHSIYRKVLSDIPVPPSLAAPS..ENGQYIIPKGHYIMAAPGVSQMDPRIWQDAK
00040 RMHAPIHSIYRKVLSDIPVPPSLSAPS..ENGQYIIPKGHYIMAAPGVSQMDPRIWQDAK
EAL23379 RMHAPIHSIYRKVLSDIPVPPSLSAPS..ENGQYIIPKGHYIMAAPGVSQMDPRIWQDAK
271374 RLHAPIHSIYRKVLQPIVVPQSLAAPS..EDKAYVIPKGNFIVAAPGVSQMDPKIWDDAP
367271 RLHAPIHSIYRKVLQPIVVPQSLAAPS..EDKAYVIPKGNFIVAAPGVSQMDPKIWDDAP
523016 RLHAPIHSIYRKVLSDIPVPASLSSPS..ESSAYVIPEGHFIVAAPGVSQMDPLIWPDAK
meander_loop
η3
4lxj
420
430
440
450
460
470
4lxj QFNIHRWNNDSAS.....SYS..VGEEVDYGFGAISKGVSSPYLPFGGGRHRCIGEHFA
629036 IWNPHRWTDEKGMASQALEEYT..SGDKVDYGYGSVSKGTESPYQPFGAGRHRCIGETFA
345805 EWNPHRWDDESGVAAMAAEQYT..EGDKVDYGYGSVSKGTESPYQPFGAGRHRCIGESFA
809380 TWNPYRWSDERGVAAQAVEEYQGEKGEKVDYGFGSVSKGTESPYQPFGAGRHRCIGESFA
627561 KWDPFRWLEDKGMAQEALDSYSGATSEQIDYGFGQVSKGTESPYMPFGAGRHRCVGEQFA
38441 VWEPFRWSDEKGTAAMALNEYTT.GGEKMDYGFGQVSKGTESPYQPFGAGRHRCVGEAFA
i742 TWRPRRWLEEGGVANAANEEYTS..GERVDYGFGAVSKGTESPYQPFGAGRHRCVGEQFA
KIR77383 VWNPARWHDEKGFAAAAMAQYTK..AEQVDYGFGSVSKGTESPYQPFGAGRHRCVGEQFA
00040 VWNPARWHDEKGFAAAAMAQYSK..AEQVDYGFGSVSKGTESPYQPFGAGRHRCVGEQFA
EAL23379 VWNPARWHDEKGFAAAAMVQYTK..AEQVDYGFGSVSKGTESPYQPFGAGRHRCVGEQFA
271374 RWNPLRWLVDGGIAKTANEQYQA..GEKVDYGFGAVSKGTESPYQPFGAGRHRCIGEQFA
367271 RWNPLRWLVDGGIAKTANEQYQA..GEKVDYGFGAVSKGTESPYQPFGAGRHRCIGEQFA
523016 TWNPYRWLDEKGVAASAKEQYSGATSEQVDYGFGQVSKGTESPYQPFGAGRHR.......
αL
β3_3
β3_2
4lxj
480
490
500
510
520
530
4lxj YCQLGVLMSIFIRTLKWHYPEG.KTVPPPDFTSMVTLPTGPAKIIWEKRNPEQKIGGRHH
629036 YLQLQTILATFIRNVEMKLD...GPFK.TNYQTMVVLPLPGTKIL..YRNRK.........
345805 YVQLQTILATLCRRVVLKLE...GPFPAPNYQTMIVLPLKGQTKI.MYKLRA........
809380 YVQLQTIIATVVRNLTLKFE...GDFPSPNYQTMVVLPKGGEIL...FKSRQ........
627561 YLQLSVIMSYIVRNYDIQLL...GAFPQTNYNTMIVLPLNGYVTF.KKRGQSKWI.....
38441 YVQLSVIIAYIIRNYSLRLETPDGLFPKTNYRTMIVLPLNGLMSL.TKRKQE........
i742 YLQLTMLVSEIIRTFKVEPGAP..QFPETNYQTMIVLPLHGMIKL.EPRK..........
KIR77383 YTQLSTIFTYVVRNFTLKLAVP..KFPETNYRTMIVQPTNPLVTF.TLRNAEVK......
00040 YTQLSTIFTYVVRNFTLKLAVP..KFPETNYRTMIVQPNNPLVTF.TLRNAEVK......
EAL23379 YTQLSTIFTYVVRNFTLKLAVP..KFPETNYRTMIVQPNNPLVTF.TLRNAEVKQEV...
271374 YLQLTLVVGEVVRNYKLTAAQR..EFPKTNYQTMIVLPLDPKITF.EARV..........
367271 YLQLTLVVGEVVRNYKLTAAQR..EFPKTNYQTMIVLPLDPKITF.EARV..........
523016 ....

APPENDIX B

Appendix B: Analysis of active site cavity amino acids in *Tremellomycetes* CYP51s**.** CYP51s represented with their protein IDs. The nature of active site cavity amino acid composition was calculated at the web-server https://www.peptide2.com/ using the Peptide Hydrophobicity/Hydrophilicity Analysis tool. Hydrogen bonds, stacking interactions and Van der Waals interactions are marked in blue, green and black font, respectively.

CYP51F	Active site cavity amino acids	No. of Amino acids in cavity	Cavity volume [Å^3]	Nature of active site	Amino acids interacting with ligands								
					Clotrimazole	Eburicol	Fluconazole	Itraconazole	Ketoconazole	Lanosterol	Obtusifoliol	VT-1129	Voriconazole
00040	F71 S73 A74 A75 Y76 Y77 G78 E79 D80 P81 L100 M101 R103 L118 A127 A130 Y131 L134 T135 F139 V143 V144 Y145 M151 L152 M153 Q155 K156 K157 I159 K160	99	1518	Hydrophobic: 54.55% Acidic: 3.03% Basic: 12.12% Neutral: 30.3%	Y131 HEM1 F139 F240 Y145 T135 T321 M316 M528 A313 I386 A317 V144	Y131 Y145 I389 A317 F240 V144 I386 A313 HEM F139 T321	W445 HEM Y480 F483 Q481 R377 A384 F245 Y145 A457 L380 R381 M528 A317 Y131 Q428 T321 W445 P482 E448 I386 P385	Y390 F240 F245 H387 Y131 P242 L100 Y390 HEM Y77 V144 I386 T321 F139 M528 G78 A317 T527	Y390 F240 F245 H387 Y131 P242 L100 Y390 HEM Y77 V144 I386 T321 F139 M528 G78 A317 T527	M528 A324 S322 F240 A317 L380 P385 V144 T325 I386 Y145 Y131 HEM H320 T321 T135	L134 A130 H133 F240 A317 F245 T321 M528 Y131 L100 Y390 HEM I386 Y145	Y131 F139 H320 F240 Y145 Q319 Y131 M316 T321 HEM I386 D237 A317 Y233 M528	I386 Y131 Y145 HEM F240 P385 T135 A317 G318 T321 I386 M528

	L163 A212 L216 F240 T241 P242 L243 N244 F245 M246 M287 H309 I310 M311 I312 A313 L314 L315 M316 A317 G318 Q319 H320 T321 S322 S323 T325 L380 A384 P385 I386 H387 S388 I389 Y390 R391 A422 P424 G425 F451 A452 A453 A454 Q481 P482 F483 G484 A485												

	G486 R487 H488 R489 C490 V491 G492 E493 Q494 F495 A496 Y497 Q499 L500 N524 Y525 R526 T527 M528 I529												
1742	W58 F59 S61 A62 A63 A64 Y65 G66 M67 D68 P69 F86 L88 L89 M93 C155 R156 T258 P259 V260 N261 F262 L263 F264 Y272 H404 S405 Y407	32	685	Hydrophobic: 50% Acidic: 3.13% Basic: 6.25% Neutral: 40.63%	HEM1 S159 F257 N137 Y133 I403 P402 T544 HEM1 A334 T338 M333 H337 I406 A158	F257 Y119 F127 I403 F257 T258 A158 S159 M545 T544 S157 F262 A334 T338 G160 L122 HEM1 R156 P259 V126 T123	HEM1 S159 F127 F257 T544 G160 M545 I403 T338 P402 T123 A334 S157 A158	I110 S157 S159 Y484 H505 Y119 A158 T338 I406 R408 P136 G108 Y407 I403 N137 R504 S405 L106 I110 S105 F257 M545 K109 HEM1 G107	F257 H337 F127 T123 F262 I226 A334 M333 T338 D254 M545 L122 T544 S157 F257 S336 HEM1 A158 R156 S159 G160 Y119 L250	V126 G335 S339 A158 HEM1 H337 F127 T338 M545 A334 A513 F257 T342 A330 Y133 I403 S157 P402 M333 S159	G160 T123 I403 T338 L122 P402 Y119 S159 M333 G335 F257 S157 HEM1 H121 V126 A334 P125 T124 A158 P162 T120 F127 M545	A158 H337 F257 Y407 F262 F127 R156 M333 A334 M545 HEM1 S157 L122 T123 A158 V126 T338 T544	H337 I546 T338 HEM1 F257 F127 Y133 P402 T544 A330 A334 HEM1 S157 R156 H337 M545 I403 M333 A158

	N541 Y542 Q543 T544												
271374	F60 S62 A63 A64 Y65 Y66 G67 M68 P70 F87 L89 L90 R92 M94 A97 G99 P100 G102 N103 N104 L105 S106 L107 G108 A116 A119 Y120 L123 T124 F128 V132 V133 Y134 L141 M142 Q144 K145 I148 L152 I198 A202 L206 F230	118	1746	Hydrophobic: 51.69% Acidic: 1.69% Basic: 7.63% Neutral: 38.98%	F230 HEM1 Y120 A307 T124 T311 M518 F128 I379 T517 I376 V133	Y134 Y120 F230 A303 G131 T124 M518 V132 T125 I376 K130 HEM1 F128 T311 T121 A307 G129 Y134 D135 V133	A307 T311 Y120 L123 F128 M518 I376 M306 F230 HEM1	I519 W318 Y71 H373 H377 I376 K512 P511 M518 T124 T517 I519 P375 L123 T311 W318 Y120 A307 H310 HEM1 L372 F235 F510 Y134 A374 F230	F230 H310 HEM1 Q309 A307 D227 A195 T311 Y134 I376 T517 L123 M518 Y120 E224 T124 M306	T124 Y134 Y120 A307 P375 I376 I198 Q309 T311 M518 F485 H310 HEM1 M306 F230 G308 Q489 S312	T311 S312 L304 Y380 Y120 HEM1 M518 F230 A307 I376 Y134 T305 M306 S378 G308 Q309	Y120 H377 Y380 F235 F230 A307 M518 HEM1 I379 I376 Q516 T311 T517	Y134 F230 Y120 HEM1 F128 M518 Y134 T311 A307 V133 I376 L123 T124

	T231 P232 L233 N234 F235 M236 F237 M277 M300 I302 A303 L304 T305 M306 A307 G308 Q309 H310 T311 S312 S313 T315 S316 I366 L370 H373 A374 P375 I376 H377 S378 I379 Y380 R381 V411 A412 A413 P414 G415 S417 Q418 I441 A442 S463 G465 T466 E467 S468												

	P469 Y470 Q471 P472 F473 G474 A475 G476 R477 H478 R479 C480 I481 G482 E483 Q484 F485 A486 Y487 Q489 L490 N514 Y515 Q516 T517 M518 I519												
345805	S62 A63 A64 T65 Y66 G67 M68 D69 P70 F87 L89 L90 M94 L107 A116 A119 Y120 L123 T124 F128 V133	98	1696	Hydrophobic: 56.12% Acidic: 2.04% Basic: 8.16% Neutral: 33.67%	HEM1 Y120 F228 F128 H308 V133 M304 T309 HEM1 I374 V519 A301 Y134 A305 I518	T309 H308 F228 Y120 Y120 T124 P373 S310 L123 HEM1 Y134 F228 A305 I374 F128 M304	A301 F128 T309 F228 V133 I518 Y120 Y134 A305 HEM1	Y134 F128 HEM1 L123 L140 F228 V133 Q143 P289 V132 C136 L141 Q144 T288 M147 I518 K285 T309 L290 M298 Y284 K145 A305 K283 Y134 I374	F128 H308 F228 S193 Q307 D225 L123 A305 HEM1 Y221 T309 Y120 M304 T124 I197	H308 S310 T309 F228 Y120 A305 F485 V519 T196 Q489 Y134 Q307 L303 I374 G306 HEM1	Q307 Y134 A305 G306 L302 Y120 L123 I374 HEM1 F128 F228 T124 M304 L303 L204 T309	Y120 F233 Y134 F128 F228 I518 M517 HEM1 I374 T309 T124 A305 M378 P230	F485 A486 E483 F473 Y487 Y470 HEM1 I481 C480 A442 V441 L368 P472 F485 F473 S484 G482

	Y134 N138 L141 M142 Q144 K145 V148 L152 T196 A200 L204 F228 T229 P230 I231 N232 F233 M234 F235 M275 M298 I300 A301 L302 L303 M304 A305 G306 Q307 H308 T309 S310 S311 T313 I364 L368 H371 A372 P373 I374 H375 S376 I377 M378 R379 L411 P414 G415					T309 I377							

	Q418 V441 Y470 Q471 P472 F473 G474 A475 G476 R477 H478 R479 C480 I481 G482 E483 S484 F485 A486 Y487 Q489 L490 Y514 Q515 T516 M517 I518 V519 L520												
367271	F60 S62 A63 A64 Y65 Y66 G67 M68 P70 F87 L89 L90 R92 M94 L107 A116 A119 Y120 L123	115	1873	Hydrophobic: 52.17% Acidic: 1.74% Basic: 12.17% Neutral: 33.91%	Y134 F128 HEM1 F230 Y120 A303 A307 M518 I376 M306 H310 T311	Y134 Y120 Y120 A119 A116 D118 E117 T124 M518 I376 HEM1 Y134 T311 T121	L123 Y120 A307 HEM1 I376 M518 T311 F230 T517	Y66 Y380 F87 HEM1 H377 P414 M518 P375 S378 F230 M74 F85 T124 V96 I376 A412 F73 I379 Y66 C77 V411 A413 F128 L123 A307	HEM1 H310 M306 G308 T517 S312 T201 T311 L206 F230 M518 Q309 I376 HEM1 A307 Q489 A202 I198 Y120	G308 F128 T311 A307 F230 Y134 V133 A303 H310 I376 M518 T124 HEM1 Y120 P375	T311 P375 Y134 I376 A307 M518 F230 A119 F235 Y120 L123 H122 T311 R92 HEM1	H377 Y380 F235 F128 H310 F230 A307 L123 M518 HEM1 P232 T124 I376 M306	F230 HEM1 Y134 F128 Y120 T311 M518 F230 A307 A116 T124 I376 R381

	T124 F128 V132 V133 Y134 M140 L141 M142 Q143 Q144 K145 K146 F147 I148 K149 H150 G151 L152 I198 A202 L206 F230 T231 P232 L233 N234 F235 M236 F237 M277 H299 M300 M301 I302 A303 L304 T305 M306 A307 G308 Q309 H310 T311 S312 S313 T315 S316 D363					A307 F230		T311 L98 V416				T311 V127 T517	

	I366 R367 L370 H373 A374 P375 I376 H377 S378 I379 Y380 R381 A412 P414 G415 Q418 I441 A442 G458 F459 G460 Y470 Q471 P472 F473 G474 A475 G476 R477 H478 R479 C480 I481 G482 E483 Q484 F485 A486 Y487 L488 Q489 L490 N514 Y515 Q516 T517 M518 I519												

38441	F58 S60 A61 A62 Y63 Y64 G65 E66 D67 P68 L87 L88 R90 I92 N101 L105 A114 A117 Y118 K119 V121 T122 F126 V130 V131 Y132 N136 L139 M140 Q142 K143 I146 K147 G149 L150 I196 S199 A200 S201 T203 L204 D225 G226 F228 T229 P230 L231 N232 F233	107	1762	Hydrophobic: 50.47% Acidic: 3.74% Basic: 11.21% Neutral: 34.58%	F126 Y118 H310 HEM1 F228 V131 A307 Y132 F126 T311 M306 I522 A303 I376	Y118 F228 F126 F233 T122 I522 V121 I376 Y132 H310 P230 HEM1 M306 T311 A307 V125 T229 M521 G227	HEM1 Y118 F126 T311 M521 A307 T122 F228 I376 Y132	H310 H222 F228 Y118 T520 A307 D225 M521 I522 V131 HEM1 L524 I376 Y132 T311	F228 H310 HEM1 I376 V131 T122 F126 M521 Y132 M306 D225 Q309 A307 L524	G308 F126 A307 S312 Y118 Q309 I376 I522 P375 V131 F228 T122 Y132 I196 T311 HEM1	F228 H377 G308 S378 Y380 F233 Y132 P230 A307 M306 T311 A303 F126 I376 Y118 HEM1 M521 I379	E115 A114 Y118 F126 Y132 H479 S113 Y118 A307 L139 HEM1 T109 R381 E115 N136 A303 V131 I376 R478 T311	A307 Y132 Y118 HEM1 F126 G308 I376 R381 A307 A303 F228 T311 Y132

	M234 F235 M277 M300 I302 A303 I304 L305 M306 A307 G308 Q309 H310 T311 S312 S313 T315 L370 H373 A374 P375 I376 H377 S378 I379 Y380 R381 A412 P414 G415 F460 Q462 Q472 P473 F474 G475 A476 G477 R478 H479 R480 C481 V482 G483 E484 A485 F486 A487												

	Y488 Q490 L491 N517 Y518 R519 T520 M521 I522 V523												
523016	M1 L14 A23 D25 A26 Y27 T28 H29 L30 T31 V34 F35 V39 V40 Y41 L47 L48 M49 Q51 K52 K53 F54 I55 K56 H57 G58 L59 T60 T61 A63 L64 S65 Y67 A68 P69 L70 M71 Y72 K73 E74	157	4173	Hydrophobic: 49.68% Acidic: 5.73% Basic: 11.46% Neutral: 33.12%	HEM F35 Y41 Y27 F137 Y27 HEM1 I283 V40 H217 Y41 A214 T218 F35	H217 F137 F35 Y27 T218 V40 HEM1 A214 L48 S219 P282 I283 G215 Y41 T31	T31 Y27 I283 F35 A210 M213 A214 F137 Y41 HEM1 L30 T218	N45 Y27 H387 F35 Y41 HEM1 I283 Y27 N45 F137 V40 A214 L48 T218 V18 E24 A210 P44 C43 A23	E354 Q355 A352 R388 K353 K56 HEM1 S351 F382 L59	G36 V40 Y41 A214 H206 T32 M207 F35 T28 I283 T218 T31 V39 A210 Y27 F137 HEM1 H217	F35 HEM1 T218 T222 I283 V40 A214 L277 Y41 H217 A281 A210 S219 T31 P282	F137 Y41 Y27 H217 D134 T218 HEM1 F137 A214 F35 I283	H217 F137 HEM1 F35 Y27 Y41 A214 T31 T218 I283

	A75 L98 M101 S102 L104 I105 I106 L107 T108 A109 G110 S112 L113 Q114 Y130 E131 D132 L133 D134 G135 G136 F137 T138 P139 N141 F142 M143 F144 P145 D183 H206 M207 M208 I209 A210 I211 L212 M213 A214 G215 Q216 H217 T218 S219 S220 A221 T222 S223												

	S224 W225 I226 L227 L228 L230 L240 Q244 F248 T254 F255 V258 T259 Y260 E261 S262 T263 K264 Q265 M266 M269 E270 I272 I273 R274 L277 H280 A281 P282 I283 H284 S285 I286 Y287 R288 K289 F316 I317 V318 P321 W342 G347 V348 A349 A350 S351 A352 K353												

	E354 Q355 Y356 S357 G358 Y366 F368 G369 Q370 V371 S372 Y379 Q380 P381 F382 G383 A384 G385 R386 H387 R388												
627561	V59 W65 F66 S68 A69 Y70 S71 Y72 G73 T74 D75 P76 Y77 K78 F79 L80 F81 C83 F91 F93 I94 M95 F96 R98 K99 M100 T101 V102	61	1839	Hydrophobic: 55.74% Acidic: 1.64% Basic: 9.84% Neutral: 32.79%	Y126 F134 HEM1 F235 M526 A309 T317 I382 T525 A313 M312	F235 Y126 P237 Y140 M526 F235 G314 T317 L129 M312 A313 I382 HEM1 H316 P381 T130 F240 Y386	T317 Y126 H316 T130 F235 I382 L129 A313 HEM1 T525	F235 H316 H316 T525 T317 Y386 S200 L129 E201 Q315 F240 M312 F235 Y126 D232 A313 E229 HEM1 I382 M526 I204	H379 N522 T521 M378 P381 I527 L129 H379 HEM1 T317 A320 T525 L529 V528 W324 I382 F235 H316 M526 A313	Q315 HEM1 I382 T130 Y140 Y126 I310 T317 P381 F235 A313 V139 G314 F134 S318 M526 L311	F494 A313 G314 Y140 I382 F235 S318 T317 Y126 H316 Q498 HEM1 A495 M526 I203	Y386 F240 F235 Y140 F134 Y126 V139 A309 HEM1 L129 I382 N524 H383 T525 P237 A313 T130	Y140 F235 HEM1 Y126 M526 T130 Y140 L129 I385 I382 T317 A313

	L104 A125 Y126 L129 T130 Y140 F235 T236 P237 I238 N239 F240 L241 F242 H316 P381 I382 H383 S384 I385 Y386 R387 V418 A419 A420 P421 V423 N522 Y523 N524 T525 M526 I527												
629036	A63 Y66 G67 M68 D69 P70 F87 L89 L90 R92 M94 L107 A116 A119	110	2123	Hydrophobic: 60% Acidic: 4.55% Basic: 9.09% Neutral: 26.36%	HEM1 H477 Y458 A474 V461 F472 I480 C479 G481 R478 G473 HEM1 K145	A304 Y120 T124 S375 L141 G305 I373 A300 I376	F227 HEM1 T124 F128 A304 L123 I373	F472 HEM1 Y469 W434 Y448 E437 F472 L367 Y120 R368 R364 V516 I373 P372 Q417 A371 A304 T308	F227 H307 F128 L123 M303 D224 A304 Y134 L302 I373 V127 Y120 F232 Q306 F227 I299 T124	Y120 L203 F227 S375 I376 Y134 F128 A300 Q306 A304 L301 G305 T308	H122 F232 S375 M377 I373 Y120 A304 HEM1 I376 A119 T308	Y120 F227 Y134 F128 F232 HEM1 A304 S375 A300 V133	HEM1 Y134 Y120 F227 S375 HEM1 A304 T308 T124 I373 I376

	Y120 H122 L123 T124 F128 V132 V133 Y134 L141 Q144 K145 V148 L152 F157 L188 A192 T195 A199 L203 E221 A222 L223 D224 G225 G226 F227 T228 P229 I230 N231 F232 M233 F234 M274 M297 M298 I299 A300 L301 L302 M303 A304 G305 Q306 H307 T308 S309 S310					HEM1 Y134 F128 V133 L301 T308 H374		P471 F227 Q470	M377 Y220 L223	I373 HEM1 L302 M303		Y66 T124 T308 I373	

	A311 T312 I363 L367 H370 A371 P372 I373 H374 S375 I376 M377 R378 F408 V409 L410 A411 S412 P413 G414 Q417 E446 Q470 P471 F472 G473 A474 R476 H477 R478 C479 I480 G481 E482 T483 F484 A485 Y486 Q488 L489 T514 M515 V516 V517 L518 P519 L520												

	P521												
809380	D1 P2 Y3 F5 M6 F19 L21 M22 M26 V28 L39 A48 A51 Y52 L55 T56 F60 V64 V65 Y66 N70 L73 M74 Q76 K77 V80 L84 N88 L89 Y92 V93 I96 L127 I128 T129 L130 T131 A132 S133 D134 C135 L136 Q137 F160 T161 P162 M163 N164 F165	110	1769	Hydrophobic: 57.27% Acidic: 2.73% Basic: 7.27% Neutral: 32.73%	F60 Y52 Y66 F160 HEM1 M236 A233 I309 T450 M451 A237 T241 H240 L306	Y52 F60 T241 E49 V65 T56 Y66 L306 F160 Y52 C68 HEM1 A237 T53 A48 M451	M451 F60 A237 F160 L306 L55 T241 Y66 V65 T56 HEM1 A233	F165 F19 H240 Y52 H307 F160 HEM1 P2 L55 L21 A237 P162 F5 Y66 T161 F60 T450 M451 L306 Q449	F160 F60 H240 A125 Q239 M310 D157 Y66 HEM1 A237 T241 L306 Y52 F160 M236 M451 I153	S242 T56 H240 HEM1 A237 Y66 Q239 Q423 G238 F160 I128 T241 Y52 M451 F419	A237 L234 M451 HEM1 M236 H307 F165 M310 S308 F160 Y52 T56 L306 L55 A233 G238 Y66 F60	H240 F160 Y52 T450 HEM1 D157 A237 L306 T56 M236 H240 M451 L55 F60 T241 T161 Q449	F165 HEM1 H240 Y52 Y66 F160 L55 A237 T56 L306 M236 F60 M451 T241

	V166 M207 L208 M230 M231 I232 A233 L234 L235 M236 A237 G238 Q239 H240 T241 S242 S243 T245 I296 L300 H303 A304 P305 L306 H307 S308 I309 M310 R311 M343 C345 P346 G347 V373 Q405 P406 F407 G408 A409 G410 R411 H412 R413 C414 I415 G416 E417 S418												

	F419 A420 Y421 V422 Q423 L424 N447 Y448 Q449 T450 M451 V452 V453												
EAL23379	S67 A68 A69 Y70 Y71 G72 E73 P75 F92 L94 M95 I99 L112 A121 A124 Y125 L128 T129 V132 F133 V137 V138 Y139 N143 M145 L146 M147 Q148 Q149 K150 K151 I153 G156 L157	104	1566	Hydrophobic: 53.85% Acidic: 1.92% Basic: 10.58% Neutral: 33.65%	F133 F234 HEM1 Y125 F234 A307 H314 I380 I523 M310 A311 T315 F133 T129 I383	A378 H377 A311 I380 Y139 T129 T315 V138 A318 T319 P379 HEM1 F133 Y125 F234 H314 V524 L374	T315 T129 HEM1 I380 Y139 M310 F133 I523 A311 F234 H314	F234 H314 I523 A311 V138 V524 Y139 T235 HEM1 T315 M522 D231 Y125	HEM1 Y139 F133 I380 A311 Y384 L308 L146 I153 T129 F239 A307 Y125 F234 T315 Q149 HEM1 I304 V138	H314 S316 T315 L128 A490 V524 I380 G312 M310 Q313 HEM1 Q493 Y139 I523 T129 Y125 A311 P379 F234	T315 M310 A124 G312 F234 A311 Y125 H314 Y384 V524 HEM1 I380 T315 F239 I523 S382	Y125 Y139 Y384 F239 F234 T129 I523 Y125 L128 T315 HEM1 R520 A311 Y71 I380 G72	Y125 H314 HEM1 F234 L128 F133 A311 T315 Y125 V524 M310 I380 I523

	I202 A206 S207 L210 F234 T235 P236 L237 N238 F239 M240 M281 K289 H303 I304 M305 I306 A307 L308 L309 M310 A311 G312 Q313 H314 T315 S316 S317 T319 S320 I370 L374 A378 P379 I380 H381 S382 I383 Y384 R385 P418 A446 V451 Y474 Q475 P476 F477 G478												

	A479 G480 R481 H482 R483 C484 V485 G486 E487 Q488 F489 A490 Y491 Q493 L494 Y519 R520 T521 M522 I523 V524 Q525												
KIR77383	S73 A74 A75 Y76 Y77 G78 E79 D80 P81 F98 L100 M101 I105 L118 A127 A130 Y131 H133 L134 T135 F139 V143 V144 Y145 N149	111	1662	Hydrophobic: 50.45% Acidic: 2.7% Basic: 11.71% Neutral: 35.14%	Y145 HEM1 Y131 F240 F139 V144 T135 Y145 T321 H320 A317 I386 M316	Y131 F240 Y145 T136 F139 L134 M316 Y131 G140 T321 M528 A317 V143 I386 HEM1 D146 T135 A313	HEM1 F139 Y145 F240 A317 L134 I386 M528 M316 T321 Y131	HEM1 Y145 F139 F158 I310 A313 A317 L302 I386 L314 HEM1 I288 K295 M311 Y390 Q155 L152 M273 Y131 T321 I159 M528 F240 Q292 V307 V144 C294 L291 K156	H320 F240 F139 F240 M528 HEM1 M316 H320 I389 T321 Y233 I209 Q319 I386 E234 Y390 D237 Y131 A317	T135 S322 Y131 F495 A317 F240 HEM1 Q499 F139 I208 M528 L315 Q319 I386 T321 Y145 G318 H320	HEM1 I386 Y131 S322 S388 G318 Y390 Y145 M528 A317 L315 A313 L134 F240 F139 Q319 L314 T321 M316	Y145 Y131 F139 F245 Y390 F240 Y145 L134 Y131 HEM1 R103 I386 A317 T135 S388 T527 P242 M528	Y145 HEM1 F240 Y131 A317 T321 F139 Y145 T135 I386 T527 P385 M528

	M151 L152 M153 Q154 Q155 K156 K157 I159 K160 S161 G162 L163 T164 S167 L168 Y171 I208 I209 T211 A212 S213 L216 F240 T241 P242 L243 N244 F245 M246 F247 M287 H309 I310 M311 I312 A313 L314 L315 M316 A317 G318 Q319 H320 T321 S322 S323 T325 I376					T132 V144							

	L380 H383 A384 P385 I386 H387 S388 I389 Y390 R391 P424 F451 Y480 Q481 P482 F483 G484 A485 G486 R487 H488 R489 C490 V491 G492 E493 Q494 F495 A496 Y497 Q499 L500 N524 Y525 R526 T527 M528 I529												