Imperial College London
School of Public Health
Department of Infectious Diseases and Epidemiology

Genomic Epidemiology and Ecology of *Cryptococcus neoformans* var. *grubii* in southern Africa

Mathieu Vanhove

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Biology of the Diploma of Imperial College, April 2016

Abstract

Despite the increased availability of antiretroviral therapy, cryptococcal meningitis remains a neglected disease that is responsible for up to 30% of AIDS related deaths. In the light of next generation sequencing technologies, this thesis investigates the molecular epidemiology of *Cryptococcus neoformans* var. *grubii* (*Cng*) in southern Africa. Sampling in the field and laboratory isolation yielded 37 environmental *Cng* and 38 *C. gattii* (*Cg*) from various ecotypes in Zambia and South Africa. In Zambia, the two sister-species were associated with different ecoregions (p <0.001). *Cng* was predominantly found inhabiting the Zambezi Mopane Woodlands, a previously described ecological niche for the pathogen, whereas *Cg* was more frequently isolated from wet Miombo Woodlands. The fungal community structure associated with these two saprophytes was explored using a high-throughput metabarcoding approach. Changes in the microbial assemblages were observed between the different ecoregions (R_{ANOSIM} >0.350, p <0.001) and these differences were more pronounced when climatic (temperature and precipitation) and spatial factors were considered. This work confirmed the local adaptation of these fungal species to their specific ecoregions, both at the level of the fungal community and their associated environmental variables. The characterisation of *Cryptococcus* biotic environments will help to understand where the pathogen might emerge and to predict a potential spillover of the pathogen into susceptible host populations as well as uninfected ecoregions. Whole-genome sequencing of 50 environmental *Cng* genomes revealed the presence of 26 isolates that grouped into lineage VNI, and 24 isolates that grouped into lineage VNB. Genome-scans were performed to uncover genetic 'outliers'that were putatively under selection within each molecular type. Genomes of environmental VNB isolates were found to group into two statistically-supported clades and 32% of VNB isolates possessed the MAT a mating-type locus. Recombination was shown to occur frequently within the VNB molecular type, whereas VNI isolates displayed a more clonal profile with the MAT a mating-type being much more rare (4%). VNI infections were shown to be more often associated with urban centres while VNB *Cng* are mainly present across the *Colophospermum mopane* region, therefore VNB infection most likely occurs in rural settings. Subsequently, a comparison between environmental and clinical VNB isolates was performed to understand whether the genetic makeup of *Cryptococcus neoformans* isolated in patients is linked to specific environmental genotypes. Genetic variants associated with clinical isolates were uncovered. Nineteen genes with a known-virulence function were found to be under selection in the environment and were also found to be also associated with clinical isolates.

In vitro phenotypic assays were performed to identify variations between clinical and environmental isolates. Clinical isolates were found to be thermotolerant, produced larger capsule and were more easily phagocytosed than environmental isolates. Clinical VNI isolates replicated faster at 37°C than clinical VNB isolates (p <0.002). This finding might be a reflection of the evolutionary relationship between the VNI molecular type and its association with domestic and feral pigeons. Finally, the epidemiology of the VNI molecular type with isolates from the African and Asian continents (n=155) was investigated in the light of whole-genome sequence data. Phylogenetic analyses revealed eight genetic clusters. A ten-fold increase in population size was observed in each VNI cluster in the past 200 to 8,000 years, suggesting the role of the domestication of the pigeon in the global spread of the VNI molecular type. These findings indicate that each cluster is likely to contribute differently to the neglected epidemic of *Cryptococcus neoformans* var. *grubii* worldwide.

Acknowledgements

First, I would like to thank Prof. Matthew Fisher, who allowed this project to take place. I would like to thank everyone in the lab who helped me during this project: Jo for the key help and advices in bioinformatics, Susan for her crucial help in the lab. I extend my thanks to our collaborators at the St Georges University of London (SGUL), Tihana Bicanic (Second supervisor) who reviewed the thesis chapters, and Tom Harrison. Mat Beale for his general help and remarks throughout the project. Also, I would like to acknowledge all the member of the Fisher lab for their general support, Julia, Simon, Rhys, Jen, Pria and even Kieran. I also need to thank Shichina who guided me throughout the use of ImageStream.

The project would never have been possible without the help of Duncan Chanda, Shabbir Lakhi and Natasha Karunaharan in Zambia, who were very helpful and receive us very well. I wish to acknowledged Robert Wilkinson and Graeme Meintjes and the every member of the Wilkinson group including Brian, Rene and Kathryn who made my stay in Cape Town very pleasant.

I would like to thank the National Environment Research Council (NERC, UK) which funded this work.

On a personal note, I would like to thank my family and friends in London, Scotland and France who have been very supportive during the thesis. Just to name a few, I would like to thank Simon, Raf, Doug, Laurne, Alice Todd.

Declaration

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

Mathieu Vanhove

Copyright Declaration

'The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work'

Contents

Abstract **i**

Acknowledgements **iii**

Declaration **iii**

1 Introduction **2**

1.1 General Background 2

1.2 History and Taxonomy of *Cryptococcus* 4

1.2.1 History of the *Cryptococcus* Nomenclature 4

1.2.2 Molecular-based Typing 5

1.3 Reproduction and Infectious Cycle 6

1.3.1 The reproduction of the *Cryptococcus*-species complex 7

1.4 Infection cycle 8

1.4.1 Human infection and dissemination 8

1.4.2 Non-Human Infection 10

1.5 Management of Cryptococcal Meningitis 11

1.5.1 Clinical Manifestation 11

1.5.2 Diagnostic Tests . . . 12
1.5.3 Treatment . . . 13
1.6 Epidemiology and Ecology . . . 14
1.6.1 Distribution of *Cryptococcus*-species complex . . . 15
1.6.2 Ecology of the *Cryptococcus neoformans/gattii*-species complex . . . 16
1.7 *Cryptococcus* in the Genomic Era . . . 19
1.8 How do Virulent Fungal Pathogens Emerge? . . . 20
1.8.1 *Cryptococcus* fitness . . . 20
1.8.2 How does an environmental yeast become a human pathogen? . . . 21
1.8.3 Virulence factors in *C. neoformans* . . . 22
1.9 Thesis Aims and Objectives . . . 25

2 Environmental Sampling **29**

2.1 Introduction . . . 29
2.2 Methods . . . 31
2.2.1 Environmental Sampling . . . 31
2.2.2 Isolation and identification of *C. neoformans* . . . 32
2.2.3 Differentiation between *C. neoformans* and *C. gattii* . . . 34
2.3 Results . . . 37
2.3.1 Recovery of *Cryptococcus* species in Zambia . . . 37
2.3.2 South Africa . . . 40
2.4 Discussion . . . 42
2.4.1 Trees Species and Decaying Wood . . . 42

2.4.2 Soil . . . 43
2.4.3 Pigeon Guano and Other Sources . . . 43
2.4.4 Variation in Species Geography . . . 44
2.4.5 Seasonal Variation . . . 45
2.5 Conclusion . . . 46

3 Influences of Fungal Diversity on *Cryptococcus* Distribution 48

3.1 Introduction . . . 48
3.1.1 Background . . . 48
3.1.2 Importance of understanding fungal biodiversity . . . 50
3.1.3 Benefits of Understanding the Ecology of Yeasts in Soil . . . 51
3.1.4 Use of Internal transcribed spacer (*ITS*) to Investigate Fungal Diversity . 51
3.2 Methods . . . 52
3.2.1 Sequencing of the *Internal transcribed spacer 1* (*ITS1*) in Isolated Yeast 52
3.2.2 Metabarcoding of *ITS2* . . . 56
3.2.3 Statistical Analysis . . . 60
3.2.4 Ecological Niche Modelling Maxent . . . 62
3.3 Results . . . 63
3.3.1 Fungal Diversity using Metabarcoding *ITS2* Approach . . . 63
3.3.2 Microbial patterns within Ecoregions . . . 65
3.3.3 Differences in Fungal Community Structure between the Mopane and Miombo Regions . . . 66
3.3.4 Fungal Diversity using Sanger Sequencing *ITS1* Approach . . . 70

3.3.5 Impacts of Environmental Conditions on Soil Community Structures . . 85
3.4 Discussion . . . 90
3.4.1 Illumina Sequencing . . . 91
3.4.2 The Use of *ITS* to Assess Fungal Diversity . . . 92
3.4.3 Fungal Diversity Present in Zambia . . . 93
3.4.4 Fungal Diversity within in Zambian Ecoregions . . . 94
3.4.5 The Role of Seasonality on Fungal Communities . . . 95
3.4.6 Environmental Conditions and Soil Community Structure . . . 95
3.4.7 Ecological niche modelling . . . 98
3.5 Conclusion . . . 98

4 African Environmental *Cryptococcus* populations 100

4.1 Introduction . . . 100
4.1.1 Whole Genome Sequencing . . . 101
4.1.2 Investigating Selection . . . 101
4.1.3 Biogeography . . . 102
4.1.4 Landscape Genomics . . . 103
4.2 Methods . . . 104
4.2.1 Whole-Genome Sequencing of Environmental *Cryptococcus neoformans* . 104
4.2.2 Phylogeny . . . 107
4.2.3 Population Assignment and Admixture . . . 108
4.2.4 Mating types . . . 109
4.2.5 Biogeography: Effect of Distance on Genetic Divergence . . . 110

4.2.6 Population Genomic Analyses . . . 111
4.2.7 Linkage Disequilibrium (LD) decay . . . 112
4.2.8 BayeScan . . . 112
4.2.9 Gene Content Comparison . . . 113
4.2.10 Identifying Signals of Adaptive Selection . . . 113
4.3 Results . . . 115
4.3.1 WGS and SNP analysis . . . 115
4.3.2 Phylogeny and Population Assignment . . . 116
4.3.3 Population Genomic Analyses . . . 123
4.3.4 Intra-specific variations in the VNB clade . . . 132
4.3.5 Environmental VNI Population . . . 152
4.4 Discussion . . . 165
4.4.1 Population Structure of Environmental Populations . . . 165
4.4.2 Population Genomic Analyses . . . 166
4.4.3 Biogegraphy . . . 169
4.5 Conclusion . . . 172

5 Clinical and Environmental Cryptococcal populations 174

5.1 Introduction . . . 174
5.1.1 Epidemiology and Population Structure . . . 175
5.1.2 Study of virulence in *Cryptococcus neoformans* var. *grubii* . . . 175
5.1.3 Genetic factors underpinning virulence . . . 177
5.2 Methods . . . 178

5.2.1 *Cryptococcus neoformans* var. *grubii* Isolates . . . 178
5.2.2 Whole-Genome Sequencing . . . 178
5.2.3 Phylogeny . . . 178
5.2.4 Population Assignment and Admixture . . . 178
5.2.5 Mating Type . . . 178
5.2.6 Population Genomic Analyses . . . 179
5.2.7 Chromosome Copy Number Variation (CCNV) . . . 179
5.2.8 Genome-wide association study (GWAS) . . . 180
5.2.9 Phenotypic Characterisation . . . 180
5.3 Results . . . 182
5.3.1 VNB *Cng*: Clinical vs. Environmental isolates . . . 182
5.3.2 The *Cryptococcus neoformans* VNI Molecular Type . . . 205
5.3.3 Phenotype . . . 214
5.4 Discussion . . . 221
5.4.1 VNB Molecular Type . . . 221
5.4.2 The VNI molecular Type . . . 223

6 Discussion and Conclusions 226

6.1 Introduction . . . 226
6.2 *Cryptococcus neoformans* and *C. gattii* have distinct ecological niches . . . 227
6.2.1 *C. neoformans* is adapted to the drier Mopane Woodlands . . . 227
6.2.2 Our knowledge regarding the fundamental niche of the *Cryptococcus*-species complex remains limited . . . 228

6.3 Associations between *Cryptococcus* genetic variations and the environment . . . 229

6.3.1 The fungal community structure changed between the Zambian ecoregions 229

6.3.2 Environmental niche model conforms with biodiversity studies 230

6.3.3 Biogeography: variations in the distribution of *Cng* molecular types in Southern Africa 231

6.4 Population genomics of *Cryptococcus neoformans* var. *grubii* 232

6.4.1 Signature of selection 232

6.4.2 The rise of Genome-scans 233

6.4.3 The presence of MAT a locus supports the evidence of recombination in the VNB clade 234

6.4.4 The key role of CCNV in the infection process 235

6.5 Aiming to link the genotype to the clinically-relevant phenotype 236

6.6 Epidemiology of *Cryptococcus neoformans* var. *grubii* 237

6.6.1 Local infection VNI is an urban acquired-infection and VNB is a rural . 237

6.6.2 Clinical isolates do not belong to a particular genetic cluster 238

6.6.3 Inference on the population dynamics and the 'Out of Africa' hypothesis 238

6.7 Conclusion 239

6.8 Principal findings 240

Bibliography **241**

List of Tables

1.1 Virulence factors for *C. neoformans* - adapted from (Casadevall et al., 2003). . . 24

2.1 Primer sequences for the PCR reactions 36

2.2 *Cryptococcus*species complex recovery in Zambia 37

2.3 *Cryptococcus* species complex Recovery in South Africa 41

2.4 *Cryptococcus*species complex Recovery from the Environment: recovery types were not found to be significantly different using a chi-squared test. 42

3.1 ANOSIM and permutation MANOVA of microbial diversity patterns across Zambian Ecoregions 64

3.2 Relative OTUs abundance 64

3.3 ANOVA - Important Taxonomic Variations between Ecoregions 64

3.4 Microbial patterns within Ecoregions 66

3.5 ANOSIM and ADONIS of microbial diversity patterns within Zambian Ecoregions 67

3.6 ANOVA - Important Taxa Variations between the Mopane and Miombo Region 67

3.7 ANOVA - Important Taxa Variations between the Mopane and Miombo Region in Dry Season 68

3.8 Taxa Variations across space in the Mopane Region (Dry Season) 69

3.9 Relative *Cryptococcus* Genus Abundance 70

3.10 ANOSIM and ADONIS in *Cryptococcus*-species Complex Positive samples . . . 70

3.11 ANOSIM and ADONIS in *Cryptococcus*-species Complex Positive samples . . . 71

3.12 ANOSIM of microbial diversity using the *ITS1* approach 71

3.13 Bray-Curtis Diversity across Season 72

3.14 Microbial patterns within ZMW 72

3.15 Feature Importance Taxa within the Mopane Region 73

3.16 Microbial patterns within CMW 74

3.17 Feature Importance Taxa within the Miombo Region 74

3.18 Microbial patterns within SMW 85

3.19 Permutational MANOVA of Environmental Effects on Microbial Diversity Patterns Between Regions 88

3.20 BEST Analysis within Ecoregions 89

3.21 BEST Analysis within Mopane Woodlands 89

3.22 BEST Analysis within Miombo Woodlands 90

4.1 Population Genetics Statistics of environmental *Cng* 116

4.2 SNPs differences among the different environmental *Cng* Populations 116

4.3 F_{ST} between populations 121

4.4 Gene annotation for gene identfied using Gene content analysis 127

4.5 Significant Virulence-associated Genes in the 0.1% quantile in VNB isolates (n=24) 129

4.6 Significant Virulence-associated Genes in the 0.1% quantile in VNI isolates (n=26) 130

4.7 Direction of Selection in VNB and VNI isolates 130

4.8 Virulence-associtaed genes under selection in the VNB and VNI clade 131

4.10 Partial Mantel Tests in VNB population 138

4.11 RDA analysis in VNB populations 138

4.12 Virulence-associated genes correlated with Environmental Variables in VNB clade140

4.14 Virulence-associated genes under selection in the VNBa clade 145

4.15 Virulence-associated genes under selection in the VNBb clade 146

4.16 Direction of Selection in VNB isolates 147

4.17 Virulence-associated genes under selection in the VNBa clade (gene-based methods) 148

4.18 Virulence-associated genes under selection in the VNBb clade (gene-based methods) 150

4.19 Partial Mantel Tests in VNI populations 155

4.20 RDA analysis in VNI populations 155

4.21 Virulence-associated genes with significant SNPs after accounting for multiple testing methods 157

4.22 Gene containing SNPs under selection in at least 3 climatic variables 159

4.23 Virulence-associated genes under selection in the VNIb clade (n=13) 163

5.1 F_{ST} statistics among VNB clusters 187

5.2 Population Genetics Statistics of clinical and environmental *Cng* populations . . 187

5.3 Population Genetics Statistics of the VNBa *Cng* cluster 188

5.4 Virulence-associated genes identified using BayeScan in the VNBa cluster 1 (n=14).189

5.5 Genes identified only in the clinical genes of the VNBa cluster 1 191

5.6 Population Genetics Statistics of in the VNBa *Cng* - cluster 1 environmental and cluster 3 191

5.7 Genes identified only in the clinical genes of the VNB cluster 3 193

5.8 Virulence-associated genes under balancing selection in the VNBa clinical isolates (Tajima's D) 196

5.9 Virulence-associated Genes under positive selection in VNBa clinical isolates . . 198

5.11 Virulence-associated genes associated with Clinical and Environmental VNB isolates 200

5.13 Genes under positive selection in environmental population and of interest in clinical isolates 201

5.14 Population Genetics Statistics of VNI *Cng* genetic clusters 209

5.15 SNPs uniquely shared by two clusters 209

5.16 F_{ST} Statistics among VNI clusters 210

5.17 Recombinating regions among VNI genetic clusters 211

5.18 Divergence Estimations of chromosome 12 using BEAST 212

5.19 SNPs uniquely shared by two clusters 214

S1 Environmental Isolates Collected in Southern Africa, 49 *Cryptococcus neoformans* var. *grubii* and 38 *Cryptococcus gattii* 295

S2 List of sample used in ITS2 metabarcoding 297

S3 Microbial patterns within Ecoregions per season 297

S4 Relative contribution of Bioclim layers to the Maxent model (Busby, 1991) . . . 301

S5 Environmental Isolates Collected in Southern Africa (n=49) 304

S6 Sequencing Results for Environmental *C. neoformans* Genomes (n=50) 305

S7 SNPs found using VNI H99 reference and using a VNB SAe9 reference 305

S8 SNPs unique to each *Cng* Populations 306

S9 Strong Gene Candidates under selection in the VNB lineage 306

S10 Gene identify using the BSR analyses 306

S11 VNB genes under selection 307

S12 direction of selection VNB 308

S13 Gene identified in the VNBa clade using the Gene content analyses 310

S14 Gene identified in the VNIa clade using the Gene content analyses 311

S15 Gene identified in the VNBb clade using the Gene content analyses 311

S16 Gene identified in the VNIb clade using the Gene content analyses 311

S17 *Cryptococcus neoformans* var. *grubii* - Virulence genes (n = 470) 312

S18 Signficant SNPs identified using different multiple testing methods (VNI isolates) 313

S19 Genes Detected using a Landscape Genomics Approach 314

S20 Virulence-associated genes under selection in the VNBb clade 314

S21 Genes under selection in environmental population 314

S23 SNPs variations in the VNB molecular clade 315

S24 Genes identified in the VNBa population using a BayeScan Approach 316

S25 Mating Type in Clinical VNB isolates 317

S26 Gene under selection in Clinical VNBa 317

S27 Unique SNPs in each VNI genetic cluster 321

S28 Shared SNPs in each VNI genetic cluster 321

S29 Unique SNPs Ciculating in the Environmental African Population (n=23) 325

S30 Unique SNPs Ciculating in the Environmental Indian Population (n=23) 326

S31 Samples Location 332

List of Figures

1.1 The *C. neoformans* species complex - According to Xu et al. (2001), *Cryptococcus neoformans* and *C. gattii* diverged 37 Mya. The serotypes A and D diverged 18.5 Mya, whereas the division between Serotype B and C is estimated at 9.5 Mya. VNI remains the genotype the most widely spread and responsible for most of the burden of cryptococcal meningitis (Lin et al., 2006). Recently two sub-populations of *Cng* have been described in southern Africa by Litvintseva (2011): an endemic arboreal population which appears highly genetically diverse and an avian population linked to pigeon excreta. 26

1.2 *C. neoformans* reproduction (A) Pseudohyphal differentiation. (B) Mating of the haplotype of opposite mating type. Under nutrient starvation, cell secrets pheromones which lead to cell fusion. The resulting dikaryon starts filamentous growth. Blastosportes can bud from the hyphae and undergo mitosis to form yeast cells. In the basidium, the two nuclei fuse and meiosis occurs leading to the production of four chains of basidiospores (by mitosis), which bud out of the basidium. (C) Monokaryotic fruiting cells of one mating type can undergo either cell fusion and nuclei fusion or endoreplication. Blastospores are also formed and haploid basidiospores can be produce in four chains (Lin, 2009) 27

1.3 Life-cycle of *Cryptococcus neoformans* var. *grubii*. Pigeons (*Columba livia*) act as mechanical transporters in the dissemination of *Cng*. The infection is acquired by inhaling spores from the environment. The basidiomycete fungal yeast is ubiquitous in the environment and is thought to be associated with many environmental niches. Individuals with a suppressed immune system, such as HIV-positive population, can hardly control the dissemination of the fungal pathogen and are at high risk cryptococcal meningitis. (Lin et al., 2006) 28

1.4 Methodogical framework to study genetic variations linking genetic, phenotype and environmental factors (Lepais and Bacles, 2014) 28

2.1 Sampling sites across southern Africa (n=1,777). Each site was marked and GPS coordinates were taken. 32

2.2 Niger seed plate showing characteristic growth of *Cryptococcus neoformans*. , Individual colonies were picked and spread onto a new Niger-seed agar after 48 hours to prevent filamentous growth. After a 48 hours incubation period, the plates were stored at 4C and dark brown colonies (number 5-6-8) were cryopreserved at -80C. 33

2.3 PCR using *CAP59* and *SOD1* primers to distinguish *Cryptococcus neoformans* and *C.gattii*. On the three isolated tested the first two amplify both primers meaning that they are *Cng*, whereas the last isolate failed to amplify *SOD1* and is therefore *C. gattii* 36

2.4 Recovery of *Cryptococcus* species-complex from Zambia. The map displays the different Zambian Ecoregions (White 1983). *Cryptococcus neoformans* isolates appear in red and *C. gattii* recovered are in blue. 40

2.5 Recovery of *Cryptococcus* species-complex in Southern Africa - *Cng* isolates are in red (n=37) and *Cg* (n=38) are in green. 47

3.1 Fungal *internal transcribed spacer* (*ITS*) of ribosomal RNA and its associated primers - *ITS3* and *ITS4* were used for the metabarcode of the *ITS2*. *ITS1F* and *ITS4* were used using the Sanger-Sequencing method. 53

3.2 Soil Sampling sites used for the Metabarcoding Approach (n = 74) 76

3.3 *ITS* Primers with Illumina adapter overhang sequences and linker sequence. . . 76

3.4 TapeStation D1000 electropherogram Final Cleaned Fragment with an average band size of 509 bp, with region plot defined (red bars), and upper and lower reference peaks visible 76

3.5 Analysis of Kapa Real-time PCR 77

3.6 Variation between the different Ecoregions (a) OTUs variation between the different Zambian ecoregions,; (b) Chao1 distribution for the three ecoregions (ZMW - blue, CMW - Red, SMW - orange). 77

3.7 Fungal Community Structure Variations within Mopane and Miombo Region in the Rainy Season - (a) UPGMA tree of samples from the CMW (blue) and ZMW (red) showing a clear clustering; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profile for CMW (left) and ZMW (right); (c) PCoA of samples coloured by ecoregions: CMW (blue), ZMW (red). 78

3.8 Fungal Community Structure Variations within Mopane and Miombo Region in Dry Season - (a) Clustering of samples in each of the ecoregions CMW (blue) ZMW (red); Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profile for CMW (left) and ZMW (right); (c) PCoA of samples coloured by ecoregions: CMW (blue), ZMW (red). 79

3.9 Fungal Community Structure Variations across Space within the Mopane Region (Dry Season)(a) On the UPGMA, samples appear mixed between the north (red) and the south (blue) of the ZMW; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the north (left) and south (right) show a similar profile; (c) PCoA of samples appear mixed: northern samples (red), southern samples (blue). 80

3.10 Distribution of the different Genus identified using culture-based *ITS1* sequencing method 81

3.11 Species distribution of the within the *Cryptococcus* Genus identified using culture-based *ITS1* sequencing method 82

3.12 Seasonal Variation within the Mopane Region - (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue), *Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA reveals have a mixed profile. 83

3.13 Seasonal Variation within the Miombo Region (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA of samples appear mixed. 84

3.14 Seasonal Variation within the SMW - (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA reveals have a mixed profile. 86

3.15 Maxent Species Distribution Modelling based on Bioclim Layers - Ecoregion map (bottom-right) portrays the ecoregion associated with each sister species, ZMW and SMW for *Cng* and CMW for *Cg*. 90

3.16 Principal physical and bioclimatic variables (a) Altitude (b) Temperature of Warmest Quarter (bio 10)(c) Precipitation of Wettest Quarter (bio 16)(d) Precipitation Seasonality (bio 15) (e) Precipitation of Wettest Month (bio 13)and (f) Isothermality (bio 3). Sample locations are given on figure a. 97

4.1 Population Genomics Framework - (a) Detection of genes which varies between linages using Bayescan and LS-BSR analysis. (b) Detection of gene under selection. Candidate genes were considered as strong candidates undergoing positive selection if they were detected by at least two statistic parameters, either frequency-based parameters (Tajima's D or D') or gene-based parameters (Mk test or ω). (c) Detection of loci associated with environmental variables using LFMM 118

4.2 Phylogeny using RAxML - Isolates are coloured by populations assigned using *fineStructure* and Admixure. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: namibian isolates, H99 is the VNI reference. Bootstrap values were averaging 100% for each node. 119

4.3 Admixture Results - In the VNI (a) and VNB (b) clade, two sub-populations can be observed as shown by the admixture cross-validation plot (c and d). Potential hybrids can be observed in both molecular types, SAe10 in VNI and Ze20, Ze10 and Ze7 in the VNB clade. 120

4.4 FineStructure Analysis (Unlinked Model) - *fineStructure* allow to investigate population structure. On the x-axis, each genome is considered as a recipient, and on the y-axis, the *Cng* isolate is considered a donor of genomic region. The VNB and VNI populations are clearly segregated with very limited sharing of genomic regions between the VNI and VNB populations. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow. 121

4.5 Sampling distribution of the different VNI and VNB sub-populations across Southern Africa. 122

4.6 Recombination analysis, using LDhat interval - The peaks represents the recombination hotspots across the genome. 124

4.7 Population decay of Linkage Disequilibrium - The rate of decay of r^2 in the VNI population, in red, appears slower than in the VNB population, in blue. Recombination is therefore acting in VNB isolates to break up LD regions. . . . 125

4.8 Bayescan output comparing VNB and VNI molecular types - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}). 125

4.9 Population-level gene content analysis of environmental isolates (n=50) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two molecular types. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0). 126

4.10 Variation of Tajima's D across the environmental *Cng* (Vilella et al., 2005) - Tajima's D values were calculated over a 100 bp non-overlapping sliding window. Negative value in the 0.1% quantile are highlighted in green. 128

4.11 Variation of LD across the environmental *Cng* (Vilella et al., 2005) - D'was calculated over a 100 bp non-overlapping window. Loci in the 0.1% quantile are highlighted in green. 129

4.12 Population Genetics analyses of environmental VNB isolates across the genome. In each chromosome, the value of each statistic was averaged over a non-overlapping 10-Kb window - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n = 470) ; (b) SNPs density histogram and heatmap ; (c) Histogram of F_{ST} statistics ; (d) line plot of the LD-hat rho statistic ; (e) heatmap and histogram Tajima's D ; (f) histogram of D'; The position of significant virulence genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. Genes highlighted blue are virulence genes under positive selection identified using frequency-based methods (Tajima's D and D'), genes identified using gene-based methods are in black; Strong gene candidates (identified using two statistical parameters) are in red. The band in yellow on chromosome five (containing the virulence-associated CNAG_01106) is a region which was detected under strong selection forces using frequency-based methods. 133

4.13 Population Genetics analyses of environmental VNI isolates across the genome. In each chromosome, the value of each statistic was averaged over a non-overlapping 10-Kb window. - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n = 470) ;(b) SNPs density histogram and heatmap ; (c) Histogram of F_{ST} statistics ; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; (f) histogram of D'; The position of significant virulent genes (at the 0.1% quantile) is highlighted on chromosome ideograms. Genes highlighted blue are virulent genes under positive selection identified using frequency-based methods (Tajima's D and D'), genes identified using gene-based methods are in black; Strong gene candidates (identified using two parameters) are in red. 134

4.14 Phylogenic analysis of environmental *Cryptococcus neoformans* VNB isolates (n=24) - VNBa isolates are in red and VNBb isolates appear in green. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: Namibian isolate. Bootstrap values at each node were averaging 100%. 135

4.15 Population Structure of Environmental *Cryptococcus neoformans* (n=24) - (a) The VNB clade is composed of two sub-populations, VNBa and VNBb. Two isolates Ze6 and Ze10 from the VNBb sub-population are sharing some genetic content with the VNBa population indicating that they might represent hybrids. (b) When taking linkage disequilibrum into account, the two VNB populations can still be observed. Using the linked model, some isolates are not fully included into any population (Ze10, Ze6, Ze1, Ze7, Ze21 and Ze23) . Two isolates, Ze20 and Ze2, are sharing genetic materials with both sub-populations. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow. 136

4.16 Mantel Tests for the VNB Molecular Type (n=24) using 10,000 replications. . . 137

4.17 Predicted distribution of the two VNB sub-population - Environmental Niche Model Distribution obtained with Maxent using WorldClim environmental data. 139

4.18 Recombination analysis, using LDhat interval for the VNB sub-populations - The peaks represents the recombination hotspots across the genome. 140

4.19 Population decay of Linkage Disequilibrium within the VNB isolates - The rate of decay of r^2 in the two VNB sub-populations are similar (VNBa is in green, and VNBb appears in blue), in black is the LD decay for the VNB population(n=24). 141

4.20 Bayescan Output comparing VNB sub-population - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the bayescan output showing signs (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}). . 142

4.21 Population-level gene content analysis (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two VNB sub-populations. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0). 143

4.22 Frequency-based Analysis in the VNBa clade. The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green. 144

4.23 Frequency-based Analysis in the VNBb clade. The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green. 144

4.24 Population Genetics analyses of VNB isolates across the genome. The position of significant virulence-associated genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's *D*, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of LD'; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's *D* ; The position of virulent genes under selection is highlighted in blue and gene undergoing strong selection are indicated in green. 151

4.25 Phylogenic Analysis of Environmental *Cryptococcus neoformans* VNI Isolates (n=26) - VNIa isolates are in purple and VNIb isolates appear in blue. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: namibian isolate. Bootstrap values at each node were averaging 100%. 152

4.26 Population Structure of Environmental *Cryptococcus neoformans* (n=26) - (a) The population structure under the unlinked model reveals two isolated sub-population. Only SAe10 shares genomic regions with the isolates of the two populations. (b) Under the linked model, SAe10 shares genetic materials with every "recipient" genome. To a lesser extent, SAe6, BTe4 and BTe2 are also sharing genomic content with all isolates. The VNIa sub-population is still visible. The VNIb sub-population is further subdivided using this model, indicating possible more complex sub-division. 153

4.27 Mantel Tests for the VNI Molecular Type (n=26) using 10,000 replications . . . 154

4.28 Predicted distribution of the two VNI sub-populations - Environmental Niche Model Distribution obtained with Maxent using WorldClim environmental data. 156

4.29 Summary of loci under selection identified using LFMM over a window of 10,000 bases. - Virulence genes appears in red and genes identified in multiple environmental variables are labelled in black. Legend - from the outside to inside: (a) Cryptococcal genes with a known-virulence function identified in the literature; (b) SNPs density histogram and Heat map - (c) Precipitation of Wettest Quarter (bio16) ;(d) Precipitation of Wettest Month (bio13) ; (e) Temperature Seasonality (bio4) ; (f) Mean Temperature of Warmest Quarter (bio10) ; (g) Precipitation of Driest Quarter (bio17) ; (h) Annual Precipitation (bio12) ; (i) Annual Mean Temperature (bio1) ; (j) Mean Temperature of Wettest Quarter (bio8) ; (k) Temperature Annual Range (bio7) ; (l) Precipitation of Driest Month (bio14); Environmental variables from temperature (red) and precipitation (blue). Virulent genes are highlighted in red and genes under selection are in black. Circled regions highlight where virulence-associated genes were found for each variables. 158

4.30 Recombination analysis, using LDhat interval - The peaks represent the recombination hotspots across the genome. 159

4.31 Population decay of Linkage Disequilibrium within the VNI isolates - The rate of decay of r^2 in the VNIa sub-population (green) is slower than in the VNIb sub-population (blue). The black dots represent the LD decay for the VNI population (n=24). 160

4.32 Bayescan Output comparing VNI sub-population - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the bayescan output showing signs (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}). . 161

4.33 Population-level gene content analysis (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two VNI sub-populations. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0). 161

4.34 Frequency-based analysis in VNIa population (Vilella et al., 2005) - The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green. 162

4.35 Frequency-based analysis in VNIb population (Vilella et al., 2005) - The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green. 162

4.36 Population Genetics analyses of VNB isolates across the genome. The position of significant virulence-associated genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's *D*, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of LD'; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue. 164

5.1 (a) Two amoeba cells containing one cryptococcal cell stained with Calcofluor white. Fungal cell quenched with Trypan Blue was not counted. (b) Cryptococcal cell phagocytosed by *A. castellanii*. 182

5.2 Population Structure of the VNB Molecular Clade - Population assignment was performed using Admixture (Alexander et al., 2009) - Cross-validation revealed the presence of K=2 populations. 184

5.3 Phylogenetic Analyses of VNB genetic clusters: The VNBa population can be decomposed in three genetic clusters. The VNBb population appears as one cluster (cluster 2). Environmental isolates (green) are present in three genetic clusters, only cluster 3 is composed strictly of clinical (red) isolates. Each isolates is described as followed - Ze: Zambian environmental isolates, Zc - Zambian clinical isolates, SAe: South African environmental isolates, SAc: South African clinical isolates, NAe: Namibian environmental isolates, BT: isolates of Botswana. Mating type (a or alpha) is provided. 185

5.4 Population Structure of Clinical and Environmental *Cryptococcus neoformans* (n=44) generated using fineStructure (Lawson et al., 2012) (a) On the unlinked model, the two VNB population can be clearly observed, and on model taking into account linkage disequilibrium (b) the four clusters can be observed. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow. 186

5.5 VNB Cluster 1 environmental (n=8) and clinical (n=8) isolates using a BayeScan approach - A total of 128 loci were found to have higher F_{ST} than average. These loci could be mapped to 62 genes. (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST})). 189

5.6 Gene differences in the VNB cluster 1 (n=16) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content which is shared or differs between the clinical and environmental isolates. High homology appears in red and low BLAT gene score appears green (complete homology = 1, absence of homology = 0). 190

5.7 VNB Cluster 3 clinical (n=7) and VNB Cluster 1 environmental (n=7) isolates using a BayeScan approach - 13,444 SNPs have higher F_{ST} than average (highlighted in green), accounting for 2.4% all polymorphisms 192

5.8 Gene differences in the VNB cluster 3 (n=9) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify homologous or non-homologous gene content between the two clusters. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0). 193

5.9 Chromosomal regions and genes with duplications in each cluster - Each region represent a 10-kb segment which was found in multiple copies among at least 40% of individuals in each cluster. 195

5.10 VNBa virulence-associated genes identified using gene-based methods 197

5.11 Genome-Wide Associations Study using the VNB isolates (n=44) (a) -log(p) value for each loci over the genomes. The red and blue line indicate the significance level after applying a Bonferroni correction at 0.05 and 1 respectively. (b) Q-Q plot of GWAS p value (c) The phylogeny of the different genetic clusters used in the analyses (n=4). 202

5.12 Population Genetics analyses of VNB isolates cluster 1 (environmental and clinical isolates) across the genome over a 10kb non-overlapping sliding window. The position of significant virulence-associated genes (in the 0.1% quantile) is highlighted in red on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of F_{ST} statistic with VNB cluster 1 (environmental isolates); (d) Line plot and heat map of the average coverage (e) line plot of the LDhat rho statistic (f) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue. 203

5.13 Population Genetics analyses of VNB clinical isolates (cluster 3) across the genome over a 10kb non-overlapping sliding window. The position of significant virulence-associated genes (in the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of F_{ST} statistic between the two populations; (d) Line plot and heat map of the average coverage (e) line plot of the LDhat rho statistic (f) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue. 204

5.14 Population Structure of Clinical and Environmental *Cryptococcus neoformans* VNI (n=122) (Lawson et al., 2012) - Under both unlinked (a) and linked (b) model, the eight VNI genetic clusters can be seen. Low levels of sharing between the clusters can be observed. The greatest amount of sharing between genomes appear in purple and the lowest in yellow. 206

5.15 Phylogenetic relationship of the eight genetic clusters identified in *Cng* VNI isolates (n=160). African isolates appear in red and Asian isolates are in green. Almost all bootstrap values were averaging 100% at each node. Values below 90% are indicated on the tree. 208

5.16 Recombination segments among VNI clusters - Recombination events were identify using BradNextGen (Marttinen et al., 2012) and the figure was generated using Clicos-FS (Cheong et al., 2015). Cluster 4 appears to be the most recombinogenic cluster, accounting for 49.4% of all recombination events. 211

5.17 Extended Bayesian Skyline Plot of the eight different VNI genetic clusters - Variation of population size through time using a common axis of 1 million generations. The y-axis is represented using a logarithmic scale. (Heled and Drummond, 2008) 213

5.18 Genome-Wide Associations Study using the VNI isolates (n=145) (a) -log(p) value for each loci over the genomes. The red and blue line indicate the significance level after applying a Bonferroni correction at 0.05 and 1 respectively. (b) Q-Q plot of GWAS p value (c) The phylogeny of the different genetic clusters used in the analyses (n=8). 214

5.19 Cell counts at 25°C and 37°C comparing Environmental and Clinical Isolates - *Cryptococcus* isolates were grown to log phase for 12 hours and counted. 5.10^5 cells were inoculated into YPD media and incubated for 24 hours at 25°C and 37°C. After incubation, cells were counted and the ratio of cryptococcal cells was calculated. The standard error for each group was calculated and the Wilcoxon signed-rank test was performed between each group to assess the significant differences. (a) differences between clinical (n=12), in red and environmental isolates in blue (n=15) ; (b) differences between clinical VNI (n=10), in red and environmental isolates in blue (n=13) (p <0.002)) ; (c) differences between Environmental *Cng* isolates, with VNB in red (n=15) and VNI in blue (=13); (d) differences between clinical *Cng* isolates, with VNB in red (n=12) and VNI in blue (=10). 216

5.20 Cell counts at 25°C and 37°C comparing Environmental and Clinical VNB Isolates - *Cryptococcus* isolates were grown to log phase for 12 hours and counted. 5.10^5 cells were inoculated into YPD media and incubated for 24 hours at 25°C and 37°C. After the incubated, cells were counted and the ratio of cryptococcal cells was calculated. The standard error for each group was calculated and the Wilcoxon signed-rank test was performed between each group to assess the significant differences. (a) Comparison within the genetic Cluster 1 of VNB - Clinical isolates (n=5) were compared to Environmental isolates (n=5) ; (b) Comparison between environmental isolates of Cluster 1 (n=10) and clinical isolates of cluster 3 (n=7). 217

5.21 Laccase production differences between environmental and clinical isolates. To account for interexperimental variation, laccase activity is expressed as the ratio of the median of three repeats to the positive control (H99) as explained in (Sabiiti et al., 2014). No difference was observed between the different groups. Within the two molecular type investigated, VNI and VNB, the mean for the standardised absorbance of laccase was higher in environmental isolates. 218

5.22 Capsule size differences between environmental and clinical isolates - The capsule size in (μm) (a) and capsule to ratio measure gave similar results. Clinical VNB isolates had slightly larger capsule ($\text{mean}_{\text{CLINICAL}}$= 5.00 μm, mean_{ENV}= 4.84 μm, p <0.070). This pattern was significantly observed in the VNI molecular type ($\text{mean}_{\text{CLINICAL}}$= 4.88 μm, mean_{ENV}= 4.55 μm,p <0.016). 219

5.23 Capsule production between MAT α/a and clinical and environmental VNB isolates - (a) The capsule to ratio was significantly higher in MAT α VNB isolates compared to MAT a VNB isolates (mean_{α}= 4.94 μm, mean_{a}= 4.80 μm, p <0.013) (b) When comparing environmental and clinical isolates from the genetic cluster 1, no significance could be established ($\text{mean}_{\text{CLINICAL}}$= 4.95 μm, mean_{ENV}= 4.79 μm, p <0.257) but clinical isolates seem to have slightly larger capsules. 219

5.24 Amoeba Uptake - To account for interexperimental variation, the phagocytosis index is expressed as the ratio of the median of three repeats to the positive control (H99). Each isolate was run three times and 30,000 cells were counted for each run. 220

S1 Chao1 α-diversity metric across season: dry season (red) and rainy season (blue) 297

S2 OTUs in Miombo Woodlands 298

S3 OTUs in Miombo Woodlands at the Order Level 298

S4 OTUs in Mopane Woodlands 299

S5 OTUs in Mopane Woodlands at the Phylum Level 299

S6 OTUs in Southern Miombo Woodlands 300

S7 OTUs in Southern Mopane Woodlands at the Phylum Level 300

S8 Cng phylogeny based on MLST - MLST loci were extracted in silico. VNB isolates appear in green and VNI isolates are in red. 303

S9 LFMM Analyses for VNB isolates (n=24) for Max Temperature of Warmest Month (Bio05) variable 309

S10 Tajima's *D* over the *Cng* genome for each VNBa sub-population - Value in the upper 0.1% quantile were considered for the analysis of balancing selection . . . 318

S11 Copy Number Variation - VNB environmental (left) and clinical (right) isolates cluster 1 (n=8) - Average coverage over a 10kb non-overlapping sliding window . 319

S12 Copy Number Variation - VNB environmental isolates cluster 2 (n=17) - Average coverage over a 10kb non-overlapping sliding window 320

S13 Copy Number Variation - VNB clinical isolates cluster 3 (n=9) - Average coverage over a 10kb non-overlapping sliding window 322

S14 Copy Number Variation - VNB clinical isolates cluster 4 (n=3) - Average coverage over a 10kb non-overlapping sliding window 323

S15 Admixture Results - Admixture results confirms the presence of eight different clusters. The cross-validation plot indicates that K=8 is the preferable number of genetic clusters. 323

S16 Recombination Analysis using BradNextGen over Chromosome 12 324

S17 LD decay within the eight different VNI genetic clusters 324

Chapter 1

Introduction

1.1 General Background

The *Cryptococcus* genus is highly diverse and is composed of more than 50 species. Of these, two are opportunistic pathogens that can be fatal to humans. *Cryptococcus neoformans* (*Cn*) infects primarily immunocompromised people and comprises 99% of the global burden of disease. *C. gattii* (*Cg*) accounts for 1% of *Cryptococcus* cases and infects principally immunocompetent people (Chayakulkeeree, 2006). Globally distributed, this encapsulated yeast causes cryptococcosis, a life-threatening disease, which first invades the lungs and then spreads to the nervous system. Originally isolated in peach juice in 1894 by the Italian scientist Francesco Sanfelice, the occurrence of the basidiomycetous fungus increased in the human population in the second part of the twentieth century as a consequence of the HIV/AIDS epidemic. In the 1980s, the immunosuppressed population significantly expanded and was correlated with an increased number of cryptococcal cases. As a consequence, cryptococcosis has become one of the most important opportunistic mycoses (Selik et al., 1997).

In the early days of HIV/AIDS pandemic, 5 to 10% of AIDS patients were reported with cryptococcal infection according the Center for Disease Control (CDC) (CDC, 2012). Current estimates rank cryptococcosis among the most deadly infectious diseases worldwide with numbers comparable to those of tuberculosis (Park et al., 2009). In 2009, over 1 million cases per year were estimated resulting in over 650,000 death/year.

Following the adoption of antiretroviral therapy (ART) in 1996, the development of diagnostic tools and effective treatments, the prevalence of cryptococcal meningitis (CM) dropped significantly in developed countries. First-line treatment includes an induction period of two weeks of amphotericin B and flucytosine followed by consolidation therapy of eight to ten weeks of fluconazole,then maintenance therapy with fluconazole until immune function recovery occurs (Bicanic and Harrison, 2004). However, the incidence of the disease remains high in resource limited settings where access to health care is still difficult. Nowadays, sub-Saharan Africa represents the centre of cryptococcal burden with over 700,000 cases and a 90-day fatality rate of 70% (Park et al., 2009). The disease has a broad spectrum ranging from asymptomatic pulmonary lesions to life-threatening meningoencephalitis. Infected individuals acquire a primary pulmonary infection by inhaling airborne spores. Then, the yeasts disseminate within macrophages across the meninges, resulting in replication in the brain and cerebro-spinal fluid causing a meningoencephalitis, neurological complications, and death (Lin et al., 2006; May et al., 2015). Cryptococcal meningitis accounts for 30% mortality in AIDS patients where highly active antiretroviral therapy (HAART) is unavailable and represents in these regions the most common infection of the central nervous system.

A question of interest is how this environmental yeast acquired its virulence and has become an important human pathogen. There are more than 1.5 million fungal species worldwide (Mora et al., 2011). Of these, only a minority of 150-300 species are a threat to human and, of these, only 10 to 15 appear as common pathogens of humans (Chow and Casadevall, 2011). Many studies have investigated the mechanisms behind the pathogenicity of *Cng* in humans and the selection pressures that have led to its evolution and maintenance in the environment. Current knowledge suggests that selection by amoeba, nematodes and other invertebrate predators in the environment have led to the evolution of *Cryptococcus*'pathogenicity (Casadevall et al., 2003; Steenbergen et al., 2001). However the entire picture surrounding the extent of pathogenicity in *Cryptococcus spp.* remains largely unknown. Variations in virulence between different lineages and genotypes of *Cng* have also been observed (Beale et al., 2015). The widespread emergence of cryptococcal cases worldwide may also reflect the ongoing global rise of fungi which many scientists explain as a consequence of the increase of travel and the globalisation of trade (Fisher et al., 2012); such processes are thought to have led to the emergence of *C. gattii* across the western freeboard of Canada and the USA (Fraser et al., 2005).

Recent research suggests that *Cryptococcus neoformans* var. *grubii* (*Cng*), the lineage that accounts for 90% of CM cases globally, emerged out of Africa sometime within the last 5,000 years (Litvintseva et al., 2011; Simwami, 2011). The ecology of *Cng* largely understudied and the mechanisms behind the capacity of the organisms to spread and invade new continents while maintaining its lethal features, remain unknown. Within the framework of this thesis, the use of next-generation genome sequencing technologies will be applied to *Cryptococcus* 19Mb haplotypes in order to give an accurate description of the population structure and natural ecology of the pathogen in the region where the burden of disease is highest - southern Africa. Genome-wide analysis will then allow me to investigate the epidemiological and ecological processes that drive genotype diversification (Illnait-Zaragozi et al., 2010). Environmental haplotypes will be compared with clinical *Cryptococcus* genomes in order to determine whether the genetic makeup of *Cn* isolated in patients is linked to specific environmental subpopulations. Finally, a functional analysis of *Cryptococcus* fitness will be undertaken using in vitro phenotypic assays of well-studied *Cryptococcus* virulence factors. These analyses should then reveal the potential contribution of genotype to the phenotypic differences observed between lineages and populations of *Cryptococcus*. This understanding of the molecular epidemiology of *Cryptococcus neoformans* may have significant public health implication both in terms of understanding risk-factors for infection and virulence, and may eventually help in reducing the global burden of this disease.

1.2 History and Taxonomy of *Cryptococcus*

1.2.1 History of the *Cryptococcus* Nomenclature

Over a century ago, in 1894, the German pathologists Busse and Buschke observed the first case of CM in a 31 years old female presenting a tibial lesion. The isolated microorganism, resembling a blastomycete fungus, was named *Saccharomyces hominis* (Busse, 1894). The following year, the Italian scientist, Sanfelice, who isolated the organism from fermented fruit juice (and named it *Saccharomyces neoformans*), demonstrated the pathogenicity of the yeast by inoculating his culture into animals in the laboratory. By the end of the 19th century, after the second case of CM described by the Pasteurian Curtis in 1896, three important observations had been made regarding *C. neoformans*: 1) the microorganism has a pathogenic potential in

humans, 2) it is present in the environment and 3) has the ability to cause disease if inoculated into an animal in the laboratory.

In 1901, the French scientist, Vuillemin, reported critical differences between these cultures and known characteristics of the *Saccharomyces* genus. Importantly, these organisms were not able to ferment sugar and developed ascospores; therefore he placed these species under the name of *Cryptococcus* (Ordóñez and Castañeda, 2001). Over the next fifty years, the nomenclature of *Cryptococcus* evolved from *Saccharomyces neoformans* to *Blastomyces neoformans*, *Cryptococcus hominis*, and *Torula histolytica* (Bovers et al., 2008). In 1950, *C. neoformans* was classified into three serotypes based on agglutination and precipitation tests (serotype A, B and C). The capsule of the yeast played an important role in its serospecificity (Neill et al., 1950). Serotype D was later described and named by Wilson et al. (1968).

1.2.2 Molecular-based Typing

In 1982, Kwon-Chung et al. proposed to consider *C. neoformans* (serotype A and D) and *C. bacillisporus* (serotype B and C) as two different varieties, namely *C. neoformans* var. *neoformans* and *C. neoformans* var. *gattii* (Kwon-Chung et al., 1982). By the end of the 1990s, the development of genotyping methods applied to the different serotypes allowed the improvement of the classification of the different species of the *Cryptococcus* genus. In 1999, a third variety was identify based on DNA fingerprinting analyses, namely *C. neoformans* var. *grubii* (Franzot et al., 1999). Using amplified fragment length polymorphism (AFLP) genotyping methods, hybridization events were identified (Serotype AD hybrids), which were suspected to have significant consequences on the evolution of *C. neoformans* (Boekhout et al., 2001). Based on physiological data and despite some successful mating experiments, Kwon-Chung proposed to consider the variety gattii as a distinct species from *C. neoformans* (Kwon-Chung et al., 2002).

Both *C. neoformans* (serotype A, D and AD) and *C. gattii* (serotype B and C) belong to the clade of *Filobasidiella* of the Tremellales. This order of fungi is within the Phylum Basidiomycota and is often known as 'jelly fungi'as their jelly-like fruiting bodies are widely found growing on rotten wood. Genotypic analysis led to the further division of the different varieties of *Cryptococcus* (Figure 1.1) (Meyer et al., 2003). Three genetic subpopulations have been

described within *Cryptococcus neoformans* var. *grubii*: molecular types VNI, VNII and VNB. VNI *Cng* accounts for 95% of cryptococcal infections and is globally distributed. The spread of this lineage is thought to be mainly through the propagation of clonally-derived progeny owing to the scarcity of the MAT a mating type, which is usually absent or at a frequency of below 1% (Simwami, 2011). VNII is also found worldwide, however its occurrence in patients is rare suggesting that it is less virulent than VNI. The recently recognized lineage VNB has been found in southern Africa and presents a genetically diverse population structure showing higher rates of recombination, owing to the presence of balanced frequencies of both MAT a and α, in comparison to other lineages (Litvintseva R, Vilgalys, R, et al., 2006). *C. neoformans* var. *neoformans* is the VNIV molecular type but recent work proposed to consider the molecular type as a separate species (Hagen et al., 2015). *C. gattii* has been divided into VGI, VGII, VGIII and VGIV (Bovers et al., 2008). A consensus regarding the molecular epidemiology of the two pathogenic species of the genus was reached in 2009 and the VN system was confirmed based on multi-locus sequencing typing (MLST) using seven loci (Meyer et al., 2009). However, the full picture regarding *Cryptococcus* nomenclature is unlikely to be complete. In 2000, a genome project was initiated to sequence the 14 haploid chromosomes and around 6,900 genes of *Cryptococcus neoformans* isolate H99 (VNI) (Loftus et al., 2005). This genome sequence has revealed new insights into the genetic makeup of the pathogen, finding 11 gene families that are unique to *C. neoformans* and which may potentially explain its pathogenicity.

1.3 Reproduction and Infectious Cycle

The *Cryptococcus neoformans*/*gattii*-species complex can reproduce either asexually or sexually. Clonal propagation by budding is the most common mechanism of reproduction. Sexual reproduction is ongoing but remains a rare event. However, sex presents multiple evolutionary advantages as it can lead to an increase of genetic diversity in natural populations and to provide the potential to adapt and colonise new environments (Goddard et al., 2005). Moreover, this mechanism is relevant in the context of infection, because sexual reproduction results in the production of *Cryptococcus* basidiospores which are thought to play an important role in the infection process via their aerial dispersal and inhalation.

1.3.1 The reproduction of the *Cryptococcus*-species complex

The mating system is controlled by a single locus, MAT. With two distinct alleles, MAT a and MAT α, this heterothallic mating system requires physical contacts between two opposite haplotypes to reproduce sexually. Other fungi which have the ability to switch mating types using recombination are known to have a homothallic system. Approximately 20 genes are present in the 100-kb *Cryptococcus* MAT locus (Morrow and Fraser, 2009). The latter encodes for pheromones production, pheromone receptors as well as the pheromone-responsive MAPK (Mitogen-Activated Protein Kinase) cascade which leads to meiosis and sporulation (Bovers et al., 2008). Fraser et al. (2004) suggested that the bipolar mating system resulted from the fusion of two previously unlinked sex-determined regions. Subsequently, chromosomal translocation led to a tripolar intermediate state which collapsed via gene conversion into the current bipolar system. The bipolar mating system is observed in other phyla such as ascomycete and zygomycete and is believed to have an ancestral origin within the Kingdom of Fungi (Metin et al., 2010; Kües et al., 2011).

Under nitrogen starvation, *Cryptococcus neoformans* releases pheromones in its environment (Park et al., 2010). Pheromones bind to the heterotrimeric G-protein-coupled receptor of the other mating type which activates the MAPK cascade producing cell fusion and the development of ascospores (Morrow and Fraser, 2009). The MF a pheromone produced by MAT a yeasts causes the MAT α type to form a conjugation tube. In response to the stimulation, MAT α haploid cells release MF α which causes MAT a cells to swell. The enlarged yeast is then able to fuse with the MAT α conjugation tube (Davidson and Moore, Heitman J., 2000).

During reproduction, the fusion of the two cells results in the formation of heterokaryon which then develops into dikaryon hyphae. Karyogamy occurs within the basidium located at the tip of the hyphae resulting in a diploid state (Figure 1.2). After meiosis, haploid basidiospores are produced and form chains by budding from the basidium surface (Kwon-Chung, 1975).

The vast majority of environmental and clinical isolates are MAT α haplotype (98%) (Nielsen et al., 2007). In this case, when a population is principally clonal, how can the sexual reproductive potential of the yeast be maintained? Lin et al. (2006) have shown that the MAT α locus was among the genomic regions which contribute to same-sex mating (MAT α - MAT α) in *C. neoformans*. Therefore, haplotype α cells can also undergo sexual reproduction via haploid

monokaryotic fruiting (Figure 1.2c). This mechanism is inefficient compared to 'normal'(i.e. MAT α - MAT a) mating, and may be rare in nature, but could provide a long-term advantage for the alpha cells as it can reproduce in the absence of the opposite mating type. Without the need of a mating partner, monokaryotic fruiting also produces new recombinant genotypes after diploidisation and meiosis. This process is upregulated when MF a is detected, even in the presence of only a single MAT a haplotype (Wickes et al., 1996). This inbreeding mechanism is thought to have facilitated the global emergence of the α mating type and is consistent with the clonal population structure observed in the environment (Bui et al., 2008).

Dimorphism is relatively rare in the kingdom of Fungi. The alternation between filamentous and yeast forms appears to be an important feature shared by some pathogenic fungi, including *Penicillium marneffei*, *Blastomyces dermatitidis* or *Histoplasma capsulatum* (Lin, 2009) and is often associated with switch at high (37°C) temperatures from hyphal to yeast forms. Over a century, mycologists have focused on the morphological characteristics of the organism but the relationship between fungal dimorphism and the associated virulence remains unclear. In the presence of amoeba, *C. neoformans* are seen to switch from a yeast to a pseudohyphal form. Pseudohyphal cells seem to be protected against amoeba but are unable to cause disease in animal models (Magditch et al., 2012). Another intriguing mechanism is the observed change in chromosome number, known as aneploidy or chromosomal copy number variations (chromosomal copy number variations (CCNV)), especially during the infection process (Morrow and Fraser, 2013). Change in ploidy may represent an adaptive strategy to hostile environments by leading to the rapid generation of genetically-encoded phenotypic diversity. Indeed, CCNV indubitably lead to changes in gene expression and affects protein function (Kronstad et al., 2011). Whole genome analyses have the potential to investigate these changes in ploidy.

1.4 Infection cycle

1.4.1 Human infection and dissemination

Respiratory infections from *C. neoformans* or *C. gattii* usually start by the inhalation of infectious spores. Yeasts and basidiospores have been shown to have the capacity to infect mice (Casadevall and Perfect, 1998). Serological studies demonstrated that most humans have been

exposed to *C. neoformans* before they reach the age of five owing to the ubiquity of this yeast in the environment (Goldman et al., 2001). Mitospores are thought to penetrate the lungs and reach the alveoli. During the primary infection of the airways, yeasts are more likely to display a reduced capsule. Farhi (1970) established that the *Cryptococcus* capsule is reduced when the saprophyte is in soil. Once inside the hosts, the yeast changes its morphology, to escape host immune response, by increasing the size of the polysaccharide capsule. A fraction of cryptococcal cells differentiate into 'titan'cell when exposed to the host pulmonary environment (Okagaki and Nielsen, 2012). This change of morphology potentially allows the enlarged cells (50 to 100 μm) to resist phagocytosis and may facilitate onwards dissemination and infection (Crabtree et al., 2012).

In general, the mammalian immune system controls infection by *Cryptococcus* and either clears fungal cells or restricts them within granulomas, where they remain in a dormant state (Lin, 2009). If host immunity is compromised, cells contained within the granuloma may become reactivated and lead to systemic infection. In the case of *Cryptococcus neoformans*, any organ can be involved, for instance the bones, the skin or the urinary tract (Lin, 2009). Inside its host, *Cryptococcus neoformans* is phagocytosed. In a normal immunocompetent individual, the fungus is killed by alveolar macrophages which represent 95% of the broncho-alveolar cell population (Gordon and Read, 2002). Antigen presentation triggers a Th1 type inflammatory response, leading to the formation of granulomata which contain cryptococcal cells (McQuiston and Williamson, 2012). However, in immunocompromised individuals, macrophages are unprimed and the fungus can survive macrophage uptake, then replicate within the phagocytic cells. To survive inside macrophages, the intracellular pathogen has acquired multiple strategies. First, the polysaccharide capsule increases *Cn* fitness during infection by inhibiting the ingestion of the yeast by macrophages (Vecchiarelli, 2000). Antiphagocytic proteins such as App1 (Stano et al., 2009) and Gat201 from the GATA family transcription factor are also expressed by the fungus (Chun et al., 2011). Produced by the enzyme laccase, melanin protects the cell against oxidative burst in the phagolysosome. Moreover, the basidiomycete can escape phagocytes in a non-lytic mechanism, spreading from cell-to-cell. Suppressing the filamentous form in favour of yeast form seems required for the infection, these phenomena have been observed in other pathogenic fungi (Lin et al., 2005).

To cause cryptococcal meningitis, *C. neoformans* crosses the blood-brain barrier (BBB) and

enters the central nervous system (CNS) causing an infection of the meninges and brain (May et al., 2015). Composed of endothelial cells closely joined together by tight junctions, the BBB is normally impenetrable for most pathogens. However, *Cryptococcus* can enter the CNS using different strategies. The most well documented is the 'Trojan horse'scenario, whereby the yeast 'hides'within macrophages to enter the CNS and infect the brain parenchyma (Charlier et al., 2005). However other means have been described: paracellular traversal mechanism, whereby the fungus crosses the BBB between endothelial cells if tight junctions are damaged, and transcytosis, whereby *Cn* is directly internalised by microvascular brain cells (Shi et al., 2010; Sabiiti and May, 2012). Two non-lytic processes have also been described: cryptococcal cells can spread between macrophages through cell-to-cell transfer, or the fungus can be expelled from macrophages without disrupting the phagocytic cells and therefore avoiding host inflammatory responses (Ma et al., 2007).

1.4.2 Non-Human Infection

Cryptococcus sp. is an environmental saphrophyte, therefore all clinical isolates originated in the environment. Although the yeast is an important human pathogen, human infection is an accidental dead-end event and no onwards transmission occurs. Macrophage evasion and other defence mechanisms are likely to have evolved in response to environmental predation such as amoebae, such as *Acanthomoeba castellani* (Steenbergen et al., 2001). *Cryptococcus* infection is also common in non-human animals (Duncan et al., 2005). A wide range of domestic and wild animals, including marine mammals, have been found to carry the pathogen. A typical example is cats, where cryptococcosis is the most commonly-occurring fungal disease (Lavely and Lipsitz, 2005). The complete life cycle of the pathogenic yeast is still debated (Figure 1.3). *Columba livia*, the common pigeon, appears to act as a mechanistic carrier for *C. neoformans*. Filion et al. (2006) demonstrated that pigeon guano provides the favourable conditions for *Cryptococcus* survival and mating. A widely-held hypothesis regarding the global spread of *Cng* is that these avian species are responsible for the dissemination of the fungus (Littman and Borok, 1968).

1.5 Management of Cryptococcal Meningitis

Meningitis is the most common clinical manifestation of cryptococcosis. Once inside the brain, *Cryptococcus neoformans* and *C. gattii* cause meningoencephalitis, a highly fatal disease if left untreated (Casadevall and Perfect, 1998). When the fungus invades the host, the colonisation of the airways can result in either, no disease, an asymptomatic infection, a pulmonary disease or a disseminated infection, depending on a wide range of factors.

1.5.1 Clinical Manifestation

Clinical symptoms are non-specific and include headache, fever, malaise and altered mental status. Infected patients can present either increased intracranial pressure (ICP), which is the most common complication and leads to visual impairment or hearing loss (Day, 2004). Other syndromes include Cushings syndrome, pneumonia or dementia (Micol et al., 2010).

Before the implementation of antiretroviral therapy (ART), the mortality of cryptococcal meningitis six months after the infection varied between 88 and 100% (Mwaba et al., 2001). *C. neoformans* is responsible for 99% of cases while *C. gattii* is associated with only 1% of cases, however primarily infects immunocompetent people. These two aetiological agents present the same severity even if the clinical outcome and the underlying complications can be different (Steele et al., 2010). The mortality of the disease remains high mainly because antifungal drugs are inadequate and complications associated with raised intracranial pressure are difficult to manage (van der Horst et al., 1997). Further, 15 to 30% of immunosuppressed cases suffer immune reconstitution inflammatory syndrome (IRIS) one month after the beginning of ART (Bicanic et al., 2009a), leading to complications and elevated rates of mortality.

Cohorts from sub-Saharan African revealed that the incidence of cryptococcal meningitis in HIV-infected patients varied between 13 and 44% (Bicanic and Harrison, 2004). In some African regions, where HIV prevalence is the highest, CM is the leading cause of meningitis, with an incidence higher that seen for *Streptococcus pneumoniae* and *Neisseria meningitidis* (Hakim et al., 2000). Generally, the disease is associated with profound immunosuppression (CD4 count $<$100 cells/μl) and CM appears more acute in severely immunocompromised patients. Relapse is a common feature associated with high mortality and is often linked to discontinued antifungal

therapy or failure of immune reconstitution. Several groups of HIV-negative individuals are also predisposed to CM infection, such as people with chronic organ failure, organ transplant or patients on steroids. Since the adoption of highly active antiretroviral therapy, the incidence of cryptococcal meningitis has dropped in developed countries. For instance, in Atlanta, USA, the annual incidence dropped from 66 cases per 1,000 to 7 cases per 1,000 from 1992 to 2000 (Mirza et al., 2003). Antiretroviral therapy allows an improvement of the long-term prognosis but infected individuals must first survive the acute phase. Fungicidal drug regimens need therefore to be clearly defined and management of increased intracranial pressure must be improved.

Asia and Africa remain highly affected by the disease. In Uganda, the burden of CM was estimated at 10.3 cases per 100 person years in patients presenting low CD4 counts (<200 cells/μl) (French et al., 2002). The high incidence observed in these parts of the world seems to be the result of host-pathogen interactions: the dosage and variety of infectious particles, the level of exposure or the host immune status. Given the high fatality rate in the first three months of the disease, efficient diagnosis is central in order to reduce the mortality of CM.

1.5.2 Diagnostic Tests

Examination of the cerebrospinal fluid (CSF) via lumbar puncture (LP), using either microscopy with India Ink stain or cryptococcal polysaccharide antigen detection, are the most common methods employed to diagnose cryptococcal meningitis. In developed countries, brain imaging is recommended, but this tool is unavailable in resource-limited settings (Jackson and Hosseinipour, 2010). With less than $1 per test and a high sensitivity, CSF examination is considered as the gold standard. However, the procedure requires an invasive lumbar-puncture and requires experienced technicians as well as basic laboratory sterile culture facilities. India ink reveals cryptococcal cells present in the CSF with 76% sensitivity (Saha et al., 2008). A characteristic staining pattern ('halo') can be seen around the yeast as the capsule repels the ink. Fungal burden in colony-forming units (CFUs) can be evaluated from CSF examination and can give information regarding disease prognosis (Bicanic et al., 2009b). CRAG detection is more demanding regarding laboratory infrastructure and its cost ranges from $5 in South Africa to $16 in Uganda (Meya et al., 2010).

The development of new diagnostic tools could extremely benefit clinicians. A quicker diagnostic test would reduce treatment cost and the adverse effects of antifungal drugs (Desalermos et al., 2012). A recently-developed lateral flow immunoassay (LFA) has shown comparable results to the EIA with agreement ranging from 90 to 96% (Lindsley et al., 2011). The implementation of LFA using immunochromatographic test strips is a simple and inexpensive way to diagnose CM in resource-limited environments. With a period of incubation of 10 minutes at room temperature, the *Cryptococcus*-specific monoclonal antibodies coated on the strip reveals the presence of the fungus with high sensitivity in one drop of CSF or serum (Vanselow et al., 2012). This LFA can be used as a point-of-care test by untrained personnel in resource-limited areas since it does not require a cold chain for transport and is being implemented in South Africa, Uganda, Rwanda and Mozambique. The estimated cost is around $2, depending of the country's GDP (Roy and Chiller, 2011).

1.5.3 Treatment

In 1997, Van der Horst et al. published their findings and established the basis for antifungal therapy in immunocompromised patients with cryptococcal infection (van der Horst et al., 1997). The treatment is divided in three phases: induction, consolidation and maintenance (Jackson and Hosseinipour, 2010). The early initiation of the treatment is central. Without appropriate therapy, HIV-infected individuals with CM have mortality rates reaching 100% within two weeks after arrival at heath centre (French et al., 2002).

First-line induction therapy includes two weeks'treatment with amphotericin B (AmB) (0.7 1mg/kg/day) and flucytosine (100mg/kg/day) followed by maintenance therapy, 8 to 10 weeks of fluconazole (400mg/day) the maintenance with 200mg/day until immune function recovery (WHO, 2011). In resource-limited settings, AmB is rarely available. Expensive, its application requires intravenous administration which necessitates appropriate facilities (Perfect et al., 2010). Flucytosine which is administrated orally is recommended as an adjunct to AmB. However, flucytosine is only available in Africa through clinical trials as it is also expensive and cannot be used as a single therapy (Sorrell and Chen, 2010). An alternative strategy where flucytosine is not available is an induction therapy of AmBd (1mg/kg/day) for two weeks followed by a high dosage of fluconazole (800 mg/day orally for eight weeks) (Perfect et al., 2010). Since 2000, fluconazole is available for free in many African regions thanks to the Pfizer

Diflucan® Partnership Program (Vanselow et al., 2012). Orally administrated, fluconazole appears as the only replacement regimen against CM.

Early diagnosis and treatment appears to be the best strategy to prevent death associated with CM in resource-limited settings. Applicable on a large-scale, this method requires screening HIV-infected individuals and treating patients with low CD4 count (below 100 cells/μl). Using fluconazole as monotherapy, Meya et al. (2010) applied this strategy in Uganda and found that 71% of HIV-patients who were seropositive for cryptococcal antigen were alive after 30 months. Symptomatic relapse of CM under fluconazole therapy has been observed due to both fluconazole resistance and immune reconstitution syndrome (IRIS) (Bicanic et al., 2006).

An important problem is that many areas in low income countries are deprived of health care and no doctors are often available in rural clinics. The search for an optimal therapy in these settings is still undergoing (Loyse et al., 2013). Public health authorities and NGO's involved in the fight against HIV/AIDS need to continue to make diagnostic tests more available and the possible implementation of LFA at point-of-care is promising. However, access to antifungal drugs needs to be improved and the implementation of large-scale programmes could be very beneficial to improving the survival of at-risk patients.

1.6 Epidemiology and Ecology

Before the HIV pandemic reached Europe, fatal cases of cryptococcal meningitis were rarely described. Within Africa, cases which originated from the Congo River basin were increasingly reported in young adults during the 1950s. These cryptococcal cases appear as further evidence that the AIDS epidemic might have been present in the region during this period (Molez, 1998). In the 1980s, the rise of cryptococcal cases followed the global emergence of HIV/AIDS. Nowadays, the prevalence of CM in central Africa is between 8 to 36% and ranges from 4 to 6% across West Africa (Park et al., 2009).

1.6.1 Distribution of *Cryptococcus*-species complex

Cryptococcus neoformans var. *grubii*

Accounting for 99% of cases, *Cryptococcus neoformans* var. *grubii* has been isolated worldwide. The fungus is endemic in many parts of Africa, the Mediterranean region, India, Southeast Asia, the USA, Australia, Mexico and Brazil (Severo et al., 2009). Recent reports state that *C. neoformans* may have originated in sub-Saharan Africa and may have been disseminated globally through avian migrations (Litvintseva et al., 2011). Non-African *Cng* presents a clonal profile indicating that *Cryptococcus* has been occasionally dispersed throughout the world. The global dissemination of *Cng* is thought to have occurred within the past 5,000 years (Simwami, 2011; Litvintseva et al., 2011).

In a broad epidemiological survey, Meyer et al. (2009) characterised 356 clinical isolates from five different continents. The majority of isolates were *Cng* (96%), with 78% belonging to VNI group compared to 18% for VNII. The AD hybrid (VNIII) represented 1% of isolates and 3% for *C. neoformans* var. *neoformans* (VNIV). This molecular type was more common in Europe and another study later found that VNIV was more prevalent that *Cng* in India (Gokulshankar et al., 2004).

In several African areas, *Cryptococcus neoformans* var. *grubii* is the first or second most common cause of meningitis. French et al. (2002) found that 17% of hospital deaths in Entebbe, Uganda, were due to cryptococcal meningitis. In a South African cohort, the mortality rate among 1,792 HIV patients found that 55% of death were linked to CM (Corbett et al., 2002). In 2001, a study from Lusaka, Zambia, revealed the critical virulence of CM. The mortality reached 100% among 230 adults and 56% of them were on antifungal therapy (Mwaba et al., 2001). *Cryptococcus neoformans* var. *grubii* is also (more rarely) found associated with immunocompetent individuals and a study investigating 15 medical centres in the USA found that 22% of CM patients were HIV-negative (Pappas et al., 2001).

Cryptococcus neoformans var. *neoformans*

Mainly found in Europe, recent studies have found the pathogen to be responsible of 30% of cryptococcal meningitis in HIV patients.

Cryptococcus gattii

C. gattii occurs primarily in hosts with intact immune systems (Pukkila-Worley and Mylonakis, 2008). Found mostly in subtropical areas, the species have been isolated in Europe in temperate and Mediterranean areas (Dromer et al., 1996). VGIII is most widespread molecular type and has been recovered in the South America, India and the USA (Meyer et al., 2003). In Africa, Livintseva et al. (2005) genotyped 176 clinical isolates and found a 13% prevalence of the molecular type VGIV. VGIV type has been previously found in South Africa, Mexico and India but its occurrence is particularly rare (Bicanic and Harrison, 2004).

Cryptococcus gattii shows broad variations in virulence amongst genotypes, some of which have been shown to emerge and cause outbreaks of disease. Since 1999, an outbreak of *C. gattii* (VGII) was detected on Vancouver Island, Canada, infecting predominantly immunocompetent patients (Fraser et al., 2005). The outbreak lineage spread recently and has been found in the Pacific Northwest, USA (MacDougall et al., 2007). The rapid propagation of this genotype illustrates the risk of emergence of fungal diseases due to the globalisation of microbes (Wolfe et al., 2007). Similarly, large-scale dissemination by clonal propagation have been reported for instance with the agent of leprosy, *Mycobacterium leprae* (Monot et al., 2005).

Commonly isolated in Australian Eucalyptus, *C. gattii* was possibly introduced in Vancouver as a consequence of international trade in timber products (Byrnes et al., 2010). Between 1999 and 2003, the infection rate on the island ranged between 8.5 to 37 cases per million inhabitants whereas in endemic places such as Australia, the rate averages 0.94 cases/million people/year (Chen et al., 2000). The hypervirulent strains are genetically identical and possess the MAT α mating type. The role of clonal or sexual propagation in the emergence of this outbreak strains remain largely unknown (Hoang et al., 2004; Voelz et al., 2013).

1.6.2 Ecology of the *Cryptococcus neoformans/gattii*-species complex

The analysis of the environmental population structure of *C. neoformans* has significantly improved our understanding of not only the epidemiology of the pathogen, but also of its ecology. Multiple studies have investigated the different ecological niches of *C. neoformans* and

C. gattii (Granados and Castañeda, 2005a; Mak, 2007; Mak et al., 2015). Coalescence analyses estimated that the two species have diverged from a common ancestor 30 to 40 million years ago (Xu et al., 2001). They seem to have distinct niches but the overlap of the sister species has been previously described (Nielsen et al., 2007).

Ecology of *C. neoformans*

Cryptococcus species, in common with other species in the Tremellaceae, have been found in soil, decaying wood, fruits and vegetables and various tree species (Cornelissen et al., 2003). Zoonotic transmission from bird excreta to humans has been previously demonstrated in a few cases (Lagrou et al., 2005). Various ecological studies support the fact that *Cng* is able to undergo sexual reproduction in pigeon excreta (Ensley et al., 1979; Haag-Wackernagel and Moch, 2004). Urban dwellings with an accumulation of pigeon guano appear as a potentially important reservoir; indeed the feral pigeon mainly inhabits urban centres. Soil contaminated with pigeon droppings appears as a natural rich medium ideal for fungal growth (Littman and Borok, 1968) and, although birds do not develop cryptococcosis, pigeons may act as a mechanistic vector which could have led to the global spread of *Cng* (Litvintseva et al., 2011). Indeed, avian temperatures are generally too high for fungal development with temperatures reaching 41-42°C in the gastrointestinal tract. However, cryptococcal yeasts have been isolated from bird's cloaca, feathers and skin which are at a lower temperature than the birds core and may allow continued infestation of the bird by the fungus. Bird macrophages have been found to suppress cryptococcal growth. Nevertheless, cryptococcis are still able to escape macrophages by vomocytosis and therefore spread over long distances following bird migration (Johnston et al., 2016).

Cryptococcus neoformans is present in arboreal environments and has been frequently isolated in different tree species (i.e.: *Eucalyptus sp.*; *Adansonia digitata*; *Colophospermum mopane*). Decaying wood is a rich organic source and appears as a suitable ecological niche for microorganisms. The enzyme laccase, produced by wood-rotting fungi, is associated in the degradation of lignin (Hiremath et al., 2008). Increased sampling efforts of various tree species in these regions where the flora differ considerably could help to fully understand the ecology of the species. Basidiospores are smaller than cryptococcal cells and are easily dispersed by the wind to other niches and potentially to humans. Previous studies have shown that *Cryptococcus* can

complete its life cycle in plants and wind dispersal has been recorded in fungi (Xue et al., 2007). In rural regions of southern Africa where rates of HIV infection remain high, trees and birds are in close proximity to human habitation and likely lead to the inhalation of basidiospores by individuals suffering from AIDS.

Interestingly, Litvintseva et al. (2011) recently identified an arboreal reservoir in Mopane trees (*C. mopane*). The prevalence in these trees reached 30% indicating that *C. mopane* might play a role in the local transmission of *C. neoformans*. From this ecological niche, genotypes that were primarily from the VNB lineages were principally isolated (Litvintseva R, Vilgalys, R, et al., 2006). VNB is now known to be highly genetically diverse and seems to be geographically restricted since it has only been isolated from this ecoregion. In Southern Africa, where *C. mopane* is endemic, mopane timber plays an important role in the local economy (Mashabane et al., 2001). Additionally, Mopane trees are culturally relevant, for instance, Mopane worms, the larvae of the *Gonimbrasia belina* moth, are considered as a food delicacy owing to their high protein content (Vogel, 2010).

Ecology of *C. gattii*

C. neoformans is thought to have evolved in distinct niches from *C. gattii*, but the two species could possibly interact in nature (Lin et al., 2006). In British Columbia, two genotypes of *C. gattii* were identified. The hypervirulent lineage has the MATα mating type and is unique to the Pacific Northwest. The epidemic lineage has been found to be more virulent than the minor lineage which appears identical to isolates recovered in Australian recombining populations. Fraser et al. (2005) suggested that same-sex reproduction could have promoted the pathogen expansion into new geographical niches, perhaps by promoting spore production.

The distribution of *C. gattii* worldwide is likely to be underestimated because isolation of the fungus at the global level remains patchy. Originally found in *Eucalyptus camaldulensis*, the fungus has now been isolated in more than 50 tree species and is strongly associated with decaying wood (Ellis and Pfeiffer, 1990). An association with trees would restrict the global movement of *C. gattii* compared to potential dispersal of *Cng* via avian migration. The spread of *C. gattii* would therefore be limited to tree planting and tree export (Bovers, 2007).

Australian Eucalyptus trees seem to represent an important ecological niche for *C. gattii* and

are believed to be the source of the Vancouver outbreak (Fraser et al., 2005). The reservoir of *C. gattii* remains to be determined and it is possible that the former serotype B (VGI and VGII) is associated with a wide range of ecological niches (Springer and Chaturvedi, 2010). For instance, the species was isolated in almond tree (*Terminalia catappa*) in Columbia (Callejas et al., 1998). In southern Africa, diverse *Eucalyptus* are present including the species *E. sideroxylon*, *E. tereticornis* and *E. camaldulensis* (Litvintseva et al., 2005) and might represent a potential reservoir of this species.

1.7 *Cryptococcus* in the Genomic Era

Due to rapid advancements in sequencing technologies, whole-genome sequencing (WGS) is increasingly affordable and offers scientists the opportunity to study organisms'genetics at the finest possible resolution. Recent publications on fungi have been able to use WGS to discriminate between separate outbreaks of fungal infection (Litvintseva et al., 2014). Environmental pathogens have evolved for millions of years and have developed multiple strategies to maintain infection. Positive Darwinian selection is thought to be responsible for evolutionary adaptation among fungal species (Hutter et al., 2006; Aguileta et al., 2009). Predicting the species response to environmental alteration allows us to take conservation measures and help in ecosystem restoration. Investigating the connection between the genotype and the phenotype in wild populations of fungi has been a long-term goal in evolutionary biology (Ellison et al., 2011). Next-Generation Sequencing (NGS) technologies allow to precisely define the boundaries of populations and to investigate the genetic underpinnings of important phenotypes.

To date, *Cryptococcus*'epidemiology has been extensively studied using multilocus sequence typing (MLST). The virulence associated with each lineage is known to vary (Beale et al., 2015) and it is hoped that genome sequencing will help us to understand the relative contribution of *Cryptococcus* genotypes to the disease process (May et al., 2015). The *Cryptococcus neoformans* var. *grubii* genome was released in 2005 (Loftus et al., 2005). Composed of 14 chromosomes, the 19 Mb haploid genome is known to be composed of more than 6,900 genes (Ormerod and Fraser, 2013).

WGS technologies allow us to study genomic architecture and to identify genes or genomic regions which have been targeted by natural selection (Nielsen, 2005). Genome-wide scans of

genetic diversity have been increasingly used in the past 15 years to link genotypes with phenotypes (Haasl and Payseur, 2016). For example, using site-frequency spectrum, Nygaard et al. (2010) identified loci under selection in *Plasmodium* species involved in host-parasite interaction. Using Tajima's statistics, Stukenbrock et al.(2007) detected significant departure from the haplotypes expected frequency distribution of the host-specific toxin SnToxA secreted by the wheat pathogen *Pyrenophora tritici-repentis*. Figure 1.4 portrays a methodological framework used to study adaptive selection in environmental populations, aiming to link genomic data to environmental factors and phenotypic differences (Lepais and Bacles, 2014). A complete description of the framework used in the thesis is included in section 4.1.2.

1.8 How do Virulent Fungal Pathogens Emerge?

Virulence is the relative (ie. compared against other genotypes of the species in question) capacity for a microorganism to cause disease in a host. *Cryptococcus neoformans* and *C. gattii* are opportunistic pathogens and human infections are accidental. Therefore, the human host is not the primary niche of the pathogen and does not contribute to the overall survival of the species. The arboreal fungus does not require an animal host to complete its replication. Soil and decaying matter is a complex ecological niche and the pathogenicity of the *Cryptococcus neoformans/gattii* species complex may have originated from environmental selective pressures. Infection of macrophages is similar to infection observed in amoeba and Steenbergen et al. (2001) hypothesized that the pathogenicity of the yeast in mammalian hosts evolved as a defence mechanism from environmental predators.

1.8.1 *Cryptococcus* fitness

Fitness describes the ability of an organism to survive and reproduce in a particular environment. While many species of *Cryptococcus* are present in the environment, only two can invade and cause diseases in human. The *Cryptococcus neoformans/gattii* species complex is associated with different ecological niches which lead to potential fitness variations among *Cryptococcus* populations. During infection, the expression of virulence traits is central in determining the fitness of the pathogen. Panepinto et al. (2006) defined pathogenic fitness in the context of

Cryptococcus neoformans as 'the capability to successfully propagate within the stressful environment of the host, causing disease by expression of virulence traits that damage the host.'The nature and the variety of the ecological niches occupied by the pathogen are still unknown but are excepted to be complex and to include bacteria, animals, plants or even other fungi (Steenbergen and Casadevall, 2003). Each entity will impose a selective force upon cryptococcal cells in the environment, resulting in a range of genotypes that have been exposed to heterogeneous selective forces. These diverse selective regimes, and the resulting diversity in traits that are important for survival, are then expected to translate into variation in fitness amongst genotypes during human infection. Research in the study of host-pathogen interactions between cryptococcal cells and potential hosts using Next-Generation Sequencing (NGS) technologies are just beginning and will shed light on the evolution of cryptococcal pathogenicity.

1.8.2 How does an environmental yeast become a human pathogen?

The origin of virulence in pathogenic fungi is largely unknown and the interactions between the fungus and invertebrates remain abstract. Observations indicates that *C. neoformans* infection of amoeba resembles macrophage invasion (Steenbergen et al., 2001). For instance, *Cng* is able to escape from both amoeba and macrophages using non-lytic process (Chrisman Alvarez M, Casadevall A, 2010). Virulence in murine models are enhanced by serial passage in amoeba such as *Hartmanella vermiformis*, *Willaertia magna* or *Acanthamoeba castellanii* and exposure to amoeba leads to increases in the *Cryptococcus* hyphal form (Malliaris et al., 2004). Lastly, virulence factors observed in cryptococcal infection in mammals are similar to those observed in amoeba (Casadevall, 2012).

Non-vertebrates such as amoeba represent simple and inexpensive models to study the pathogenesis of *Cryptococcus*. Normally, amoeba feed on bacteria but there is an important diversity in this group of organism and some species seem to specialise in feeding on fungi (Coûteaux and Darbyshire, 1998). Known as mycophageous amoeba, the latter were discovered preying on *Cryptococcus* by accident in 1930 (Castellani, 1930), but the evidence of amoeba feeding on yeast was only established in the 1960s (Nero et al., 1964). The two species co-exist in many soil/woody environments and must interact in many different ways across microecological niches where they co-occur.

In the first study of the interactions between amoeba and *C. neoformans*, Steenbergen et al. (2001) showed that acapsular strains were killed in the presence of *A. castellanii*. The morphological switch from yeast to a pseudohyphal form, described by Magditch et al. (2012) seems to enable the survival of the fungal cell in amoeba. Melanisation and the enlargement of capsule size also promoted cell survival. The study showed that the fungal cell developed an intracellular survival strategy and was able to replicate inside a phagocytic vacuole. Malliaris et al. (2004) continued the investigation further and noted that despite variations between capsule structures, the different pathogenic *Cryptococcus* species had the ability to exploit amoeba for growth.

1.8.3 Virulence factors in *C. neoformans*

In the last decade, multiple virulence factors have been identified that are important in mammalian infection (Table 1.1), each meeting the following requirements: 1) the property is associated with the pathogen, 2) the inactivation of the gene leads to a decrease in the observed virulence, and 3) the restoration of the gene results in restoration of the virulence (Steenbergen and Casadevall, 2003).

Polysaccharide Capsule

The polysaccharide capsule is an important virulence factor. The main component is glucuronoxylomannan (GXM) which allows the capsule to protect the cell against dehydration (Aksenov et al., 1973). The polysaccharide capsule has multiple functions both in the environment and within the hosts 1.1 Interestingly, GXM is currently been targeted for the development of a cryptococcal vaccine (Chow and Casadevall, 2011). Capsule enlargement provides protection against microbicidal molecules released by the host during infection (Zaragoza et al., 2008). At least four genes (*CAP10*, *CAP59*, *CAP60* and *CAP64*) have been identified and are required for virulence in mouse model (Kwon-Chung, 1998).

Melanin Production

Produced by laccase through the polymerisation of phenolic compounds, melanin is a critical virulence factor (Steenbergen and Casadevall, 2003). Melanised cryptococcal cells are less susceptible to antimicrobial activities and such as nitrogen and oxygen oxidants (Williamson, 1997). The fungal cells are resistant to high temperature, UV radiation, oxidant and heavy metals (Williamson, 1997). Laccase protects the fungus from oxidative damage of alveolar macrophages (Liu et al., 1999). Also associated with lignin degradation, the enzyme is thought to enhance the saprophyte ability to colonise decaying wood (Lazera et al., 1996).

Thermotolerance

To infect human, fungi must have the ability to grow at 37°C. *C. neoformans* and *C. gattii* are the only Tremellales that are heat tolerant and can grow at temperatures above 30°C (Bovers et al., 2008). Only 270 species of fungi over the 1.5 million estimated are able to cause disease in human (Hawksworth, 2001). Most of them are well adapted to growth in high temperatures, which is essential for mammalian infection. Mannitol production is associated with heat protection and resistance to osmotic stress. Previous studies revealed that the two thermotolerant species displayed elevated transcript levels of heat shock proteins and translation machinery when grown at 25 and 35°C respectively (Steen et al., 2002).

Other Virulence Factors

Phospholipase can have nutritional roles and be involved in survival in the environment. The enzyme, encoded by PLB1 gene, helps in the degradation of cell membranes. The activity of PLB1 has been investigated in murine models by Ghannoum (2000) and a significant correlation was observed with virulence in mice. Many other enzymes are involved in *C. neoformans* virulence. For example, urease, a virulence factor in many bacteria, is present in the majority of clinical *C. neoformans*. Proteinases released by fungal cells can degrade host cell components and are thought to aid in escaping phagocytosis (Chen et al., 1996). Studies in murine models established the virulent character of the alpha mating type over a-mating type (Casadevall et al., 2003). Both in nature and in patients, the alpha mating type is present 30 times more commonly than MAT a.

Table 1.1: Virulence factors for *C. neoformans* - adapted from (Casadevall et al., 2003).

	Function	
Virulence Factors	In the Environment	In Pathogenesis
Capsule	Prevents desiccation (Aksenov et al., 1973) Protection against amoeboid predators (Steenbergen et al., 2001)	Immunomodulation (Chow and Casadevall, 2011) Antiphagocytic (Kozel et al., 1988) Intracellular agressin (Feldmesser et al., 2000)
Laccase	Lignin degradation (Lazera et al., 1996)	Interference with oxidative burst (Liu et al., 1999)
Melanin	Ultraviolet shielding (Wang and Casadevall, 1994) Heat and cold tolerance (Rosas and Casadevall, 1997) Reduced susceptibility to enzymatic degradation (Rosas and Casadevall, 2001) Protection against heavy metals (Garcia-Rivera and Casadevall, 2001) Protection against amoeboid predators	Resistance to oxidative killing (Wang and Casadevall, 1994) Antiphagocytic (Wang and Casadevall, 1994) Immunomodulator (Huffnagle et al., 1995) Resistance to microbicidal peptides (Casadevall et al., 2000) Antifungal-drug resistance (van Duin et al., 2002)
Phospholipase	Nutritional function (Ghannoum, 2000) Protection against amoeboid predators	Intracellular growth (Cox et al., 2000)
Proteases	Nutritional function (Chen et al., 1996)	Tissue damage (Chen et al., 1996)
Urease	Nitrogen scavenging (Cox et al., 2000)	Intracellular growth (Cox et al., 2000)
	Other virulent functions	
Phenotypic switching	Generation of strain diversity to survive environmental stress	Immune evasion (Goldman et al., 1998)
Mating type	Sexual reproduction (Wickes et al., 1996)	Virulence factor regulation (Wickes et al., 1996)
Calcineurin and cAMP signalling	Development and reproduction (D'Souza and Heitman, 2001)	Virulence factor regulation (D'Souza and Heitman, 2001)
Superoxide dismutase	Protection against oxygen-derived oxidants?	Intracellular growth (Cox et al., 2000)

1.9 Thesis Aims and Objectives

Aim Explore the population genetic structure of environmental *Cryptococcus neoformans* var. *grubii* isolates in relation to the population genetic structure of clinical infections across southern Africa.

Specific Objectives:

1. Conduct fieldwork to gather *Cryptococcus neoformans* isolates and to survey the fungal diversity associated with the pathogen in the environment.
2. Using whole-genome sequencing techniques, sequence and align environmental genomes to the reference strain H99 in order to identify the molecular types and genome-diversity present in the environment.
3. Explore the population structure of environmental isolates. Identify potential sub-populations associated with particular ecological niches in various ecoregions and investigate signatures of selection within environmental genomes.
4. Compare environmental population structure with clinical population structure to examine whether the genetic makeup of *Cn* infecting patients is linked to specific environmental subpopulations. Elucidate the importance of sexual reproduction in fungal propagation and potential routes of human transmission.
5. Gain insights into the epidemiology of *Cryptococcus neoformans* across southern Africa.
6. Investigate the potential fitness variation between genetically and geographically different isolates.

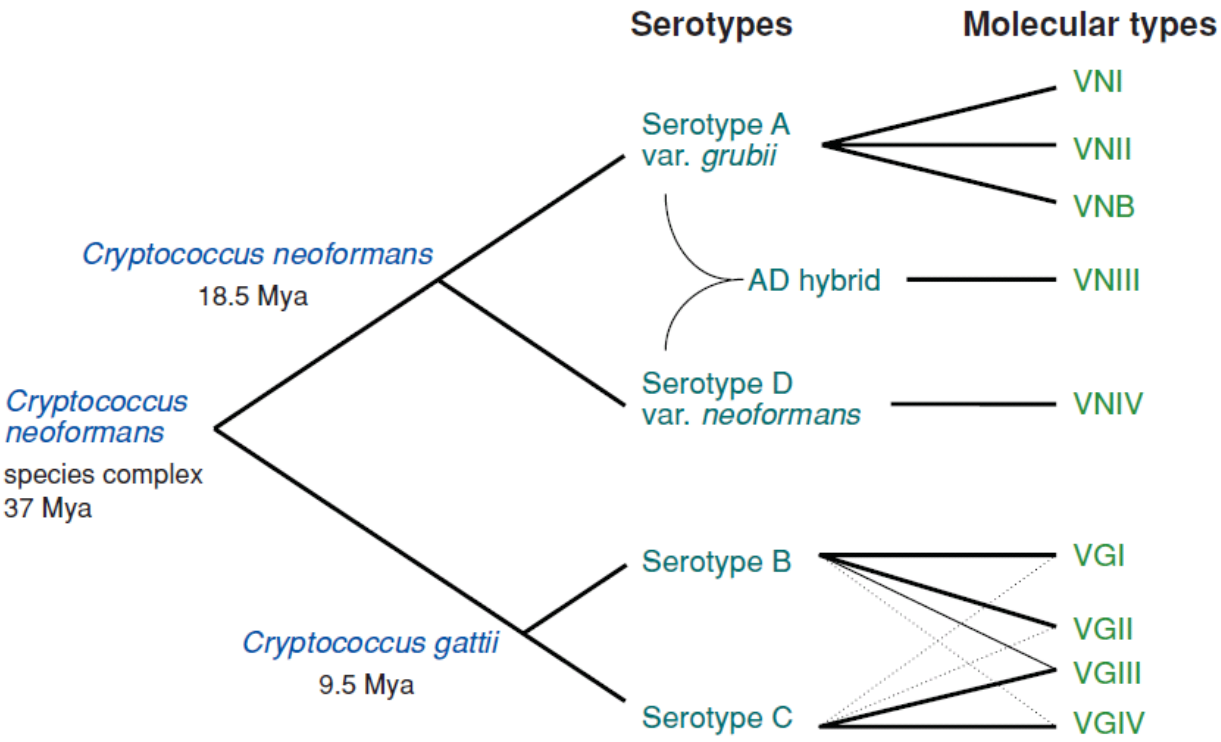


Figure 1.1: The *C. neoformans* species complex - According to Xu et al. (2001), *Cryptococcus neoformans* and *C. gattii* diverged 37 Mya. The serotypes A and D diverged 18.5 Mya, whereas the division between Serotype B and C is estimated at 9.5 Mya. VNI remains the genotype the most widely spread and responsible for most of the burden of cryptococcal meningitis (Lin et al., 2006). Recently two sub-populations of *Cng* have been described in southern Africa by Litvintseva (2011): an endemic arboreal population which appears highly genetically diverse and an avian population linked to pigeon excreta.

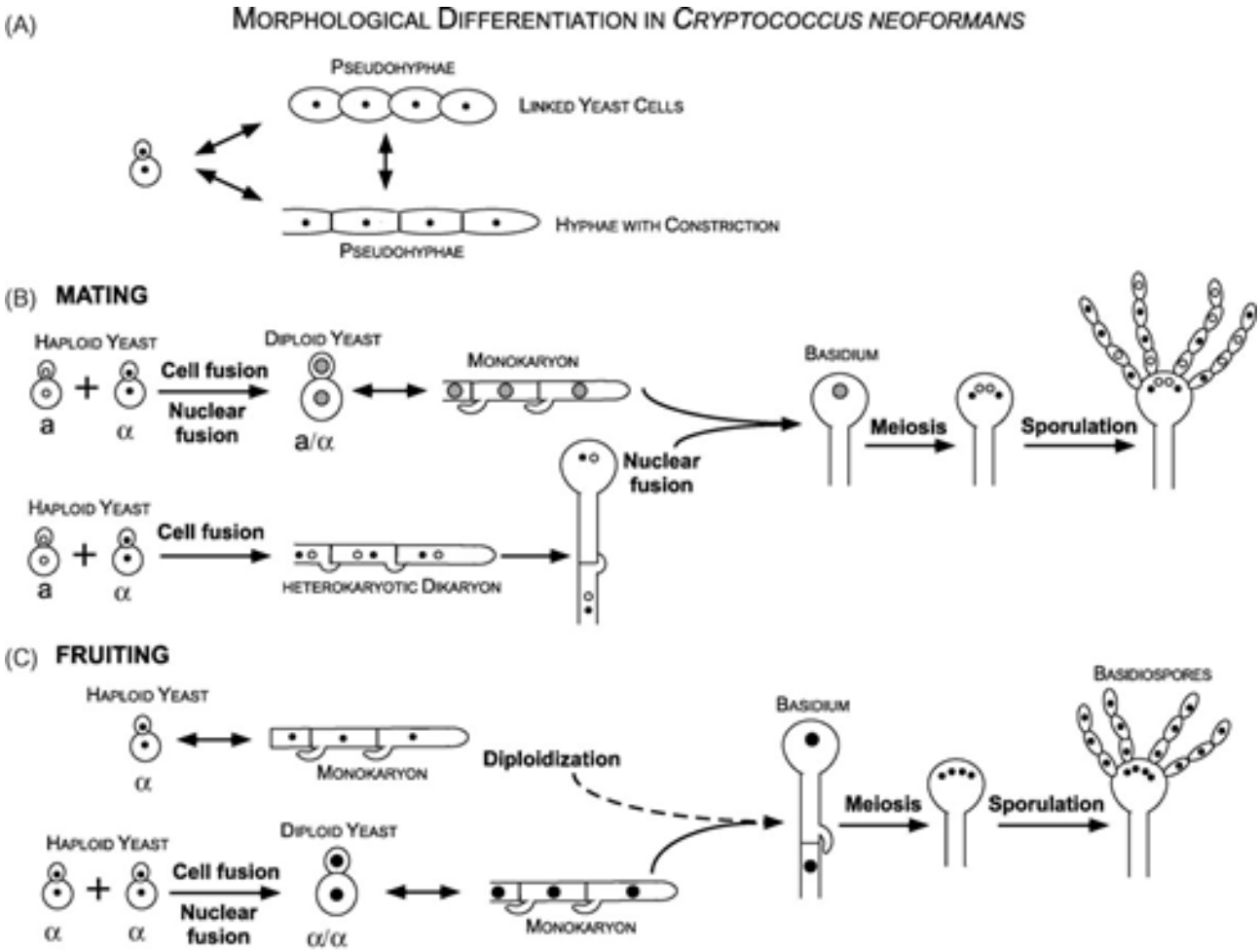


Figure 1.2: *C. neoformans* reproduction (A) Pseudohyphal differentiation. (B) Mating of the haplotype of opposite mating type. Under nutrient starvation, cell secrets pheromones which lead to cell fusion. The resulting dikaryon starts filamentous growth. Blastosportes can bud from the hyphae and undergo mitosis to form yeast cells. In the basidium, the two nuclei fuse and meiosis occurs leading to the production of four chains of basidiospores (by mitosis), which bud out of the basidium. (C) Monokaryotic fruiting cells of one mating type can undergo either cell fusion and nuclei fusion or endoreplication. Blastospores are also formed and haploid basidiospores can be produce in four chains (Lin, 2009)

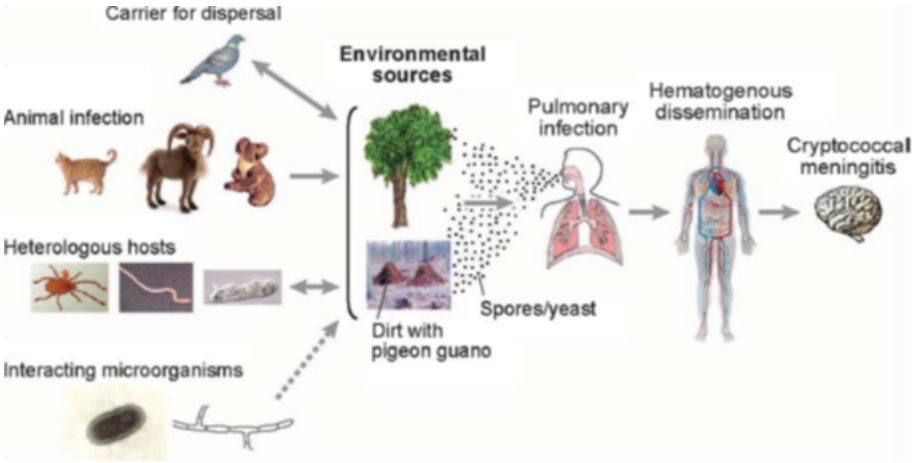


Figure 1.3: Life-cycle of *Cryptococcus neoformans* var. *grubii*. Pigeons (*Columba livia*) act as mechanical transporters in the dissemination of *Cng*. The infection is acquired by inhaling spores from the environment. The basidiomycete fungal yeast is ubiquitous in the environment and is thought to be associated with many environmental niches. Individuals with a suppressed immune system, such as HIV-positive population, can hardly control the dissemination of the fungal pathogen and are at high risk cryptococcal meningitis. (Lin et al., 2006)

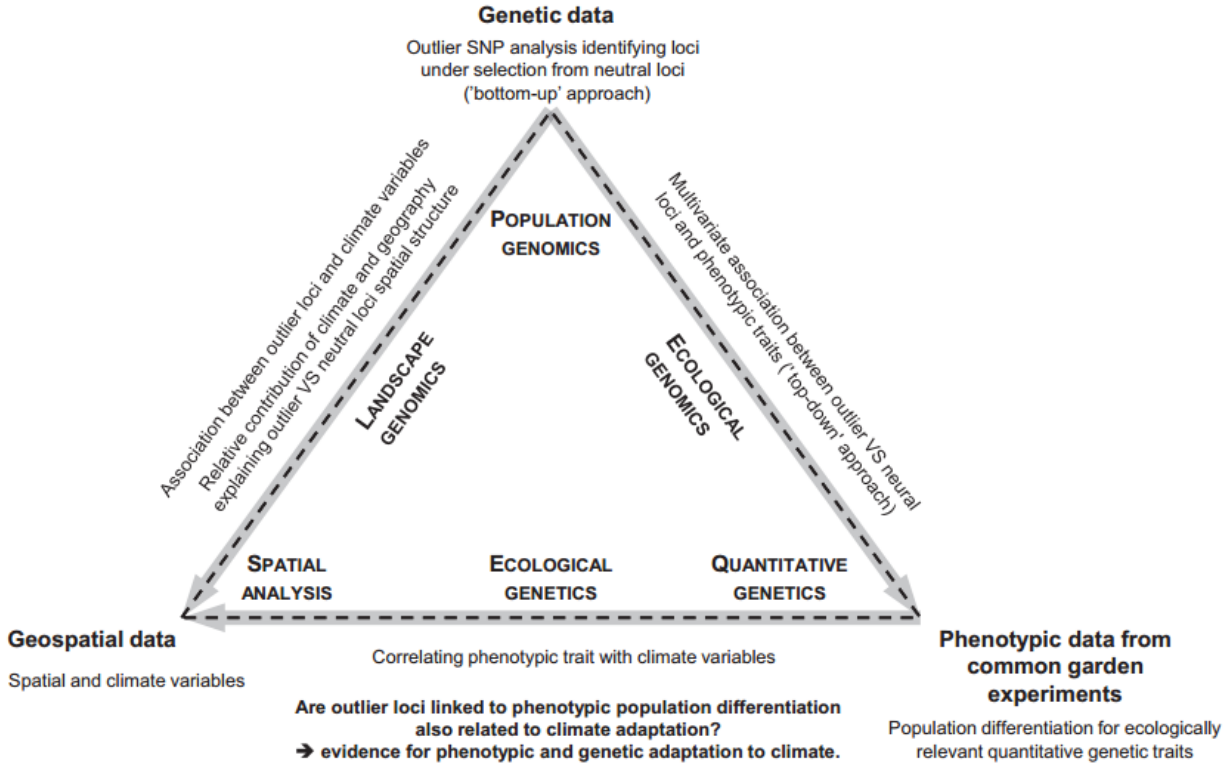


Figure 1.4: Methodological framework to study genetic variations linking genetic, phenotype and environmental factors (Lepais and Bacles, 2014)

Chapter 2

Environmental Sampling

2.1 Introduction

In 1894, the first clinical case of cryptococcosis was discovered by Busse (1894); it then took over half a century for the first environmental isolation of the pathogenic yeast to be reported (Emmons, 1951). Since Emmons'discovery, the pathogen has been recovered from the environment on multiple occasions. Several molecular typing studies have defined environmentally-occurring genotypes and lineages on different continents: North (Litvintseva et al., 2005; Kidd et al., 2007) and South America (Granados and Castañeda, 2005a), Europe (Viviani et al., 2001), Asia (Kuroki et al., 2004); (Randhawa et al., 2006) and Africa (Kangogo et al., 2014; Litvintseva et al., 2005).

Cryptococcus-positive samples have been collected from a wide range of environments including pigeon dropping (Littman and Borok, 1968), tree bark (Chakrabarti et al., 1997), soil (Ajello, 1958) or fruits (Lopez-Martinez and Castanon-Olivares, 1995). Once again, Chester Emmons was the pioneer who first identified in 1955 pigeon guano as an important reservoir for the pathogen (Emmons, 1955). The first isolation in decaying wood dates back to 1986 (Bauwens et al., 1986). The same paper also mentioned the first environmental isolation from the African continent.

C. neoformans and *C. gattii* are known to have distinct ecological niches (Kwon-Chung et al., 2002). The prevalence of the two sister species has been found to differ within tree species in a

given locality and across different biomes (Randhawa et al., 2008). Recent advances supplied evidence of a possible niche specialisation within *Cryptococcus neoformans* var. *grubii* (*Cng*) at a continental scale (Casadevall, 2008). Here, the molecular type VNI seems to be primarily associated with pigeon guano, whereas the VNB (African) molecular type has been frequently isolated in decayed wood.

This study aims to recover environmental *Cryptococcus neoformans* isolates to describe the genetic diversity of the environmental populations of *C. neoformans* and to identify potential routes of infection. The current work focuses primarily on Zambia and South Africa since the clinical isolates present in the study (Chapter 5) were collected as part of MRC funded clinical trials carried by St Georges University of London (SGUL) in Cape Town and Lusaka (Zambian) hospitals.

A recent study isolated 77 environmental samples of *Cryptococcus* from Eastern Botswana (Chen et al., 2015) confirming previous investigations which found *Cng* to be endemic in the Mopane ecoregion; the study found a thirty percent prevalence across this ecoregion (Litvintseva et al., 2011). The low altitude Mopane tree belt, dominated by *Colophospermum mopane*, is present across low-lying regions of Zambia and might play a role in local cryptococcal infection within Zambia. The present study investigates this potential ecological niche in this country and expands investigations to other Zambian ecoregions.

In South Africa, most of the patients involved in the clinical trial lived in Cape Town townships. Accordingly I visited and sampled the urban environment of Cape Town. Additionally, in order to uncover population structure, multiple locations across the country were surveyed with a particular interest on the Eastern Cape rural environment where an important number of people who inhabit Cape Town townships are originally from. This fieldwork aimed to identify whether a particular genetically isolated sub-set is responsible for the majority of the Cape Town *Cryptococcus* infections.

The prevalence of both *C. gattii* and *C. neoformans* is influenced by seasonality: Randhawa et al. (2011) reported lower prevalence during the rainy season across a seven year retrospective survey in northern India. Multiple experiments have shown that *C. neoformans* grows and reproduces preferably under dry conditions (Caicedo et al., 1999; Granados and Castañeda, 2005a). A previous Asian study reported a decrease of *C. neoformans* var. *grubii* prevalence from 24% to 4% between the dry and rainy season (Kuroki et al., 2004). The study also

described a decrease in *C. gattii* prevalence during the Asian winter. Taking these findings into account in the current study, sampling was performed in both the African rainy and the dry season.

The fieldwork aims to:

1. Recover cryptococcal isolates from the environment to explore the genetic diversity of environmental populations of *Cryptococcus*.
2. Discover new ecological niches of the *Cryptococcus*-species complex and expand the current knowledge on the seasonal variations of *Cryptococcus neoformans* prevalence.
3. Investigate the possible link between environmental populations and clinical populations infecting Zambian and South African patients

2.2 Methods

2.2.1 Environmental Sampling

To determine the prevalence of *Cryptococcus neoformans* in Zambia, samples were collected during both the rainy and dry season. The first sampling trip took place in January 2013 at the beginning of the Zambian rainy season and a total of 583 samples were collected from soil, tree bark, pigeon guano and insects. Fieldwork during the Zambian dry season occurred in September 2013 and 773 samples were collected. Sampling in South African dry season was performed in September 2014 and 421 samples were obtained.

Tree species were recorded and tree tags were used with GPS coordinates in order to be able to return to each site. A total of three ecoregions were investigated throughout Zambia namely Central Miombo woodlands (CMW), Southern Miombo woodlands (SMW) and Zambezian and Mopane woodlands (ZMW) (White, 1983).

Samples were collected in using Transwab® Amies swabs (MWE™ - MW170) and sterilized 30-ml screw-capped glass bottle. Amies liquid transport swabs were taken from tree bark. Pairs of samples from bark and soil associated with each tree were also collected. Samples were

collected and processed according to previously established protocols (Randhawa et al., 2005; Litvintseva et al., 2011) and the samples were kept at 4°C until they had been processed on Niger-seed agar.

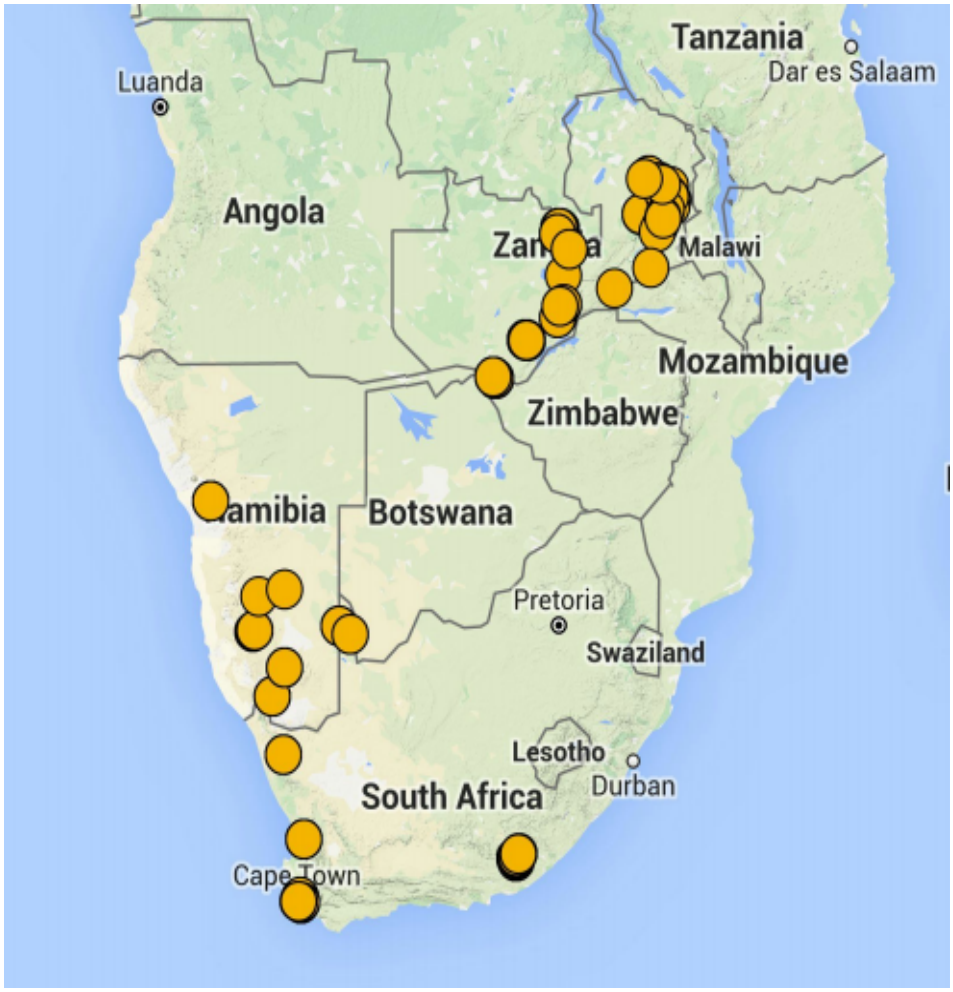


Figure 2.1: Sampling sites across southern Africa (n=1,777). Each site was marked and GPS coordinates were taken.

2.2.2 Isolation and identification of *C. neoformans*

Preparation of Niger-seed Agar Plates

Niger-Seed (NS) medium is broadly used to identify both *Cryptococcus gattii* and *C. neoformans* isolates as they produce melanin on the medium and can be distinguished from other brown yeast colonies from their characteristic chocolate dark brown colour (Staib, 1987) (Figure 2.2). To obtain the NS-medium, 70 g of Niger seeds (*Guizotia abyssinica*) were ground in a coffee-grinder and added to a litre of distilled water. The mixture was then autoclaved for 15 min.

After allowing the solution to cool off, the solution was then filtered through using a triple layer of cheesecloth. 1 g of Glucose (Sigma-Aldrich), 1 g of KH2PO4 (Sigma-Aldrich), 0.78 g of Creatinine (Sigma-Aldrich) and 15 g of Agar (Sigma-Aldrich) were added to the Niger-seed extract. Distilled water was added accordingly to obtained 1 L of Niger-Seed extract solution and the medium was autoclaved a second time for 30 min. Before dispensing the mixture onto culture plates, 0.05 g of chloramphenicol (Sigma-Aldrich) was resuspended in 1 ml of 95% Ethanol and added to the Niger Seed (NS) medium to avoid bacterial growth.

Isolation

The tubes containing soil or animal droppings were weighed to 0.5 g and distilled in 10 ml of distilled water. The samples were then further diluted to 1/10 and 1/100. Swabs were suspended in 1 ml of distilled water and diluted to 1/10. 80 μl of both the undiluted and diluted samples were spread onto NS plates and each dilution was duplicated.

NS plates were incubated at 30°C for 48 hours. To avoid contamination with filamentous fungi, individual yeast colony resembling *C. neoformans* (dark-brown colonies) were sub-cultured onto another Niger-seed agar plate and on SDS-agar. For each sample, a total of five individual dark-brown colonies were isolated and processed for DNA extraction.

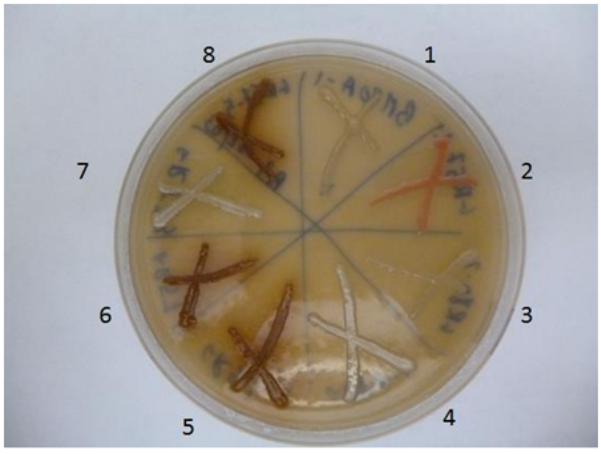


Figure 2.2: Niger seed plate showing characteristic growth of *Cryptococcus neoformans*. , Individual colonies were picked and spread onto a new Niger-seed agar after 48 hours to prevent filamentous growth. After a 48 hours incubation period, the plates were stored at 4C and dark brown colonies (number 5-6-8) were cryopreserved at -80C.

2.2.3 Differentiation between *C. neoformans* and *C. gattii*

DNA Extraction

During genomic DNA extraction, all colonies were cultured at 250 rpm in a 50 ml falcon tube for 60 hours at 37°C into 5 ml of Yeast Protein Digest (YPD) media supplemented with 0.5M of NaCl in order to reduce capsule size. The extraction process was performed using MasterPure Yeast DNA purification kit (Epicentre, UK).

After checking for contamination, the tubes were then spun at 2,000 rpm for 5 min. The supernatant was discarded and 300 μl of Yeast Cell Lysis Solution were added to the pellet. An additional 1 μl of RNaseA was added to the solution. The cells were then vortexed and transferred to a tube filled at one third of silica beads (1.0 mm). The beat beating step was performed using Mini-bead beater-16 (Biospec Products) with two rounds of 45 seconds with two minutes on ice in between. The lysate were transferred into new 1.5 tubes and spun at 14,000 rpm for two minutes to pellet the debris. The supernatant was then transferred into a new tube and incubated at 65°C for 15 minutes. Samples were then placed on ice for 15 minutes before adding 150 μl of Protein Precipitation Solution to each sample. The samples were mixed immediately by inversion. The samples were vortexed by pulse vortexting for ten seconds and the tubes spun at 14,000 rpm for two minutes. The supernatant was transferred into a fresh tube and 500 μl of isopropanol were added to the pellet. The samples were mixed by inversion and spun at 14,000 rpm for 10 minutes.

Supernatant was discarded and the pellet was washed with 500 μl of 70% ethanol. Samples were spun for two minutes at 14,000 rpm and the supernatant was removed. The wash was repeated and an additional spin of 30 seconds was performed to remove the remaining ethanol. Samples were air-dried for 10 minutes with caps opened. Finally 36 μl of EB Buffer were added to the DNA pellet. Samples were placed at 4°C for 30 minutes to allow the resuspension of the DNA and the tubes were stored at -20°C.

Polymerase Chain Reaction

Dark brown colonies were amplified by Polymerase Chain Reaction (PCR) to distinguish between *C. gattii* and *C. neoformans* using two sets of primers: *CAP59*, a capsular gene present

in both *Cryptococcus neoformans* and *C. gattii* species and *SOD1*, which is specific for *Cng* (Meyer et al., 2009) (Table 2.1).

CAP59: Control method: calculated

1. Initial Denaturation 94°C for 3:00
2. Denaturation 94°C for 0:30
3. Annealing: 56°C for 0:30
4. Extension: 72°C for 1:00
5. go to 2, (x 35)
6. Final Elongation: 72°C for 10:00
7. Hold: 4°C for ever

SOD1: Control method: calculated

1. Initial Denaturation: 94°C for 3:00
2. Denaturation: 94°C for 0:45
3. Annealing: 50°C for 1:00
4. Extension: 72°C for 2:00
5. go to 2, (x 35)
6. Final Elongation: 72°C for 10:00
7. Hold: 4°C for ever

Products were then kept at -20°C until DNA precipitation step.

Table 2.1: Primer sequences for the PCR reactions

Locus	Size	Primer	Sequence (5' to 3')	Reference
CAP59	559	*CAP59F*	CTCTACGTCGAGCAAGTCAAG	(Chang and Kwon-Chung, 1
		CAP59R	TCCGCTGCACAAGTGATACCC	
SOD1	700	*SOD1_CN_ALTF*	TCTATTCGAAATGGTCAAGG	(Bolton et al., 1999)
		SOD1_CN_ALTR	CGCAGCTGTTCGTCTGGATA	

Visualisation

Amplification results were visualised on 1.5% agarose (w/v) tris-acetic-acid-EDTA (TAE) gels, stained with 1 μl Gel Red (Biotium Inc, UK). 5 μl of the digested products were loaded onto the gels and electrophoresed at 90 V for 1 hours. In every case, the correct amplification of the DNA fragments was revealed by visualization under UV light. The expected band sizes of *CAP59* and *SOD1* amplicons were 559 and 700 bp respectively. An example of the three colonies of *Cryptococcus*-species complex is shown on Figure 2.3.

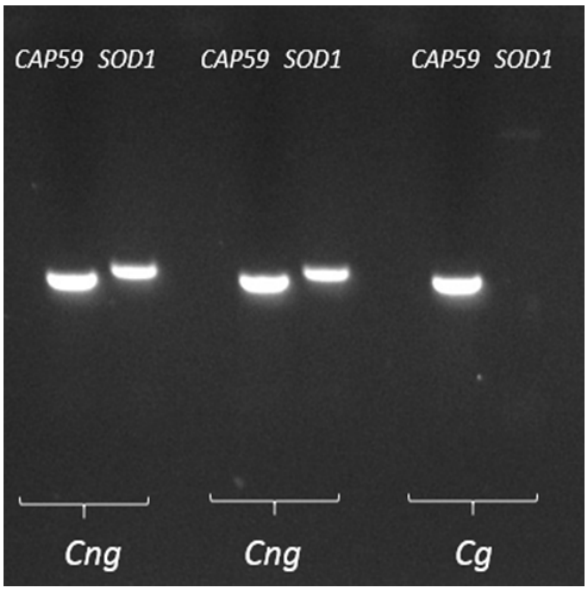


Figure 2.3: PCR using *CAP59* and *SOD1* primers to distinguish *Cryptococcus neoformans* and *C.gattii*. On the three isolated tested the first two amplify both primers meaning that they are *Cng*, whereas the last isolate failed to amplify *SOD1* and is therefore *C. gattii*

Statistics

A Pearsons chi-squared test was used to assess potential differences in the proportions of *C. gattii* and *C. neoformans* in the different ecoregions investigated. The test statistic was implemented in R software (v 3.1.1).

2.3 Results

A total of 1,742 samples were collected throughout Southern Africa (Figure 2.1. Characteristic dark brown colonies (n=234) were isolated and their DNA was extracted. Following amplification of *CAP59* and *SOD1* by Polymerase Chain Reaction, the sister species, *C. gattii* and *C. neoformans*, could be correctly identified. A total of 37 environmental isolates from different samples were positive for *Cryptococcus neoformans* and 38 *C. gattii* were identified accounting for a total of 111 and 123 individual colonies respectively.

Table 2.2: *Cryptococcus*species complex recovery in Zambia

Ecoregion	Total	Soils	Trees	Other	*Cng*	*Cg*
Rainy Season						
Central Miombo Woodlands	314	116	159	39	0 (0.00%)	17 (5.41%)
Southern Miombo Woodlands	175	61	114	0	2 (1.14%)	4 (2.29%)
Zambezian Mopane Woodlands	129	64	65	0	7 (5.43%)	1 (0.78%)
Total	618	241	338	39	9 (1.46%)	22 (3.56%)
Dry Season						
Central Miombo Woodlands	118	45	55	18	2 (2.00%)	15 (12.71%)
Southern Miombo Woodlands	221	126	83	12	1 (0.60%)	1 (0.45%)
Zambezian Mopane Woodlands	364	138	226	0	12 (3.30%)	0 (0.00%)
Total	703	309	364	30	15 (2.38%)	16 (2.28%)

2.3.1 Recovery of *Cryptococcus* species in Zambia

During the rainy season, a total of 618 samples were obtained from three different Zambian ecoregions, Central Miombo woodlands (CMW) (n=314), Southern Miombo Woodlands (SMW) (n=175) and Zambezian Mopane Woodlands (ZMW) (n=129). The species recovery

rate of the *Cryptococcus*-species complex was 5.01% in the rainy season. In the dry season, the recovery rate was 4.41% with a total of 773 samples investigated (Table 2.2, Figure 2.4).

Central Miombo woodlands (CMW)

Covering more than 50% of the country, the Central Miombo woodlands (CMW) represent the most extensive ecoregion in Zambia. The dominant tree type is *Brachystegia sp.* but sporadic grassy wetlands (dambos) can make up to 30% of this ecoregion. The population density in the region is relatively low principally owing to the low-productivity agriculture caused by nutriment-poor soils that are dominated by Kalahari sands (WWF, 2011). Subsistence slash and burn (chitemene) agriculture is practiced by 75% of the population and the Miombo woodlands have been extensively deforested in the recent years. Of the 314 miombo samples sampled in the rainy season, 17 (5.41%) were positive for *Cryptococcus gattii* and no *C. neoformans* could be found. Sampling was performed in the Copperbelt, principally around Kitwe and Ndola, from various sources: *Brachystegia sp.*(n=71, 2 positive), *Julbernardia globiflora* (n=9, 1 positive), *Isoberlinia angolensis* (n=6), *Eucalyptus sp.* (n=11), *Acacia sp.* (n=5), *Ficus sp.* (n=8), Guava tree (n=3), bamboo tree (n=3), mango tree (n=1), and unknown tree species (n=42, 2 positive). Five samples were from animals: chicken (n=2), Red Necked Francolin (n=3). Five isolates were recovered from soil. Cracks in granite kopjes were also investigated; most of them showed the presence of animal droppings from rock Hyrax. Of these cracks (n=34), seven were positive for *Cryptococcus gattii* (20.6%).

In the dry season, 100 samples were obtained in the CMW, yielding two *C. neoformans* and 15 *C. gattii*. The latter was identified in various tree species including: *Brachystegia sp.* (n=52, 4 positive), *Julbernardia globiflora* (n=15, 1 positive), *Isoberlinia angolensis* (n=3), *Eucalyptus* (n=12, 2 positive). Additionally, isolates were found in soil (n=45, 4 positives) and Hyrax middens (n=18, 4 positives). The two *C. neoformans* were also recovered from soil.

Southern Miombo Woodlands

The ecoregion covers the south of Zambia and reaches Zimbabwe in the south and Malawi in the north. The ecoregion covers 15% of the country and spreads from the Southern Province to the south-east of the Eastern Province of Zambia. In the Southern Miombo woodlands (SMW),

Brachystegia sp. predominates alongside the regular occurrence of Mopane (*Colophospermum mopane*) trees.

In the rainy season, samples were collected in Lusaka city centre (n=15), and around the town of Choma (n= 99, 5 positives). In Lusaka, all samples were taken from *Eucalyptus* trees. Around Choma, samples were collected from *Brachystegia sp.* (n=34), *C. mopane* (n=31, 4 positives), *Eucalyptus sp.* (n=3), *Acacia sp.* (n=1) *Bohinia sp.* (n=1) and unknown species (n=29, 1 positive). Three sites were found positive for *C. neoformans* (*C. mopane* n= 3) and four were *C. gattii* positives. (*C. mopane* n=1; soil =3). Interestingly, one Mopane tree appeared to be positive for both *C. neoformans* and *C. gattii*.

In the dry season, 41 samples were collected from Lusaka city centre. Multiple sources were sampled including trees and soil but only one *C. neoformans* isolate was isolated from soil. Additionally, 12 pigeon samples were obtained but none were *C. neoformans* positive. The Choma region was sampled a second time (C. mopane n= 15; unknown species n = 27; soil n= 126), but only one *C. mopane* was positive. Molecular identification found the isolate to be *C. gattii*.

Zambezi and Mopane woodlands (ZMW)

In the ZMW region, *Colophospermum mopane* predominates. The region, which appears to be a potential *C. neoformans* reservoir (Litvintseva R, Vilgalys, R, et al., 2006) covers 15% of Zambia, entering the country from the Botswana border. The Zambezian Mopane woodlands (ZMW) follow a low-elevation area ranging from 200 to 600 m as opposed to the CMW which is contained between 800 to 1,200 m. The ZMW region is thought to have significant evolutionary implications for biodiversity as it represents a changeover zone between tropical and subtropical biomes (WWF, 2011), and remains a center for Zambias National Game Parks, chiefly the Luangwa valley.

In the rainy season, two locations were investigated; in the Kafue strip in the central province (n=23, 1 positive) and near the Botswana border (n=106, 5 positives). All GIS points were recorded near *C. mopane* (n=65). Additionally, soil samples were collected near each tree (n=64). In the Kafue area, one isolate was recovered in soil and was identified as *C. gattii*. Near the Southern border, mostly Mopane trees were sampled (n=43, 3 positives), along with

Acacia sp. (n=2), *Bohinia sp.* (n=1), *Baobab sp.* (n=2), and unknown tree species (n=5). Two of the 56 soil samples were positive. All isolates were *C. neoformans* (n=5).

In the dry season, 12 *C. neoformans* were isolated. A total of 226 trees were sampled, 33 *C. mopane* located near the Botswana border were resampled. Two isolates were identified: one was found on a tree and one was recovered from neighbouring soil. Most of the samples were acquired by sampling the North Luangwa National Park, which is dominated by *C. mopane*. Ten *C. neoformans* were recovered from trees: *C. Mopane* (n=174, 10 positives), *Eucalyptus sp.* (n=17), *Acacia sp.* (n=3), *Bohinia sp.* (n=1), unknown species (n=31). Additionally, two soil samples were positive (n=138, 2 positive). No *C. gattii* was found.

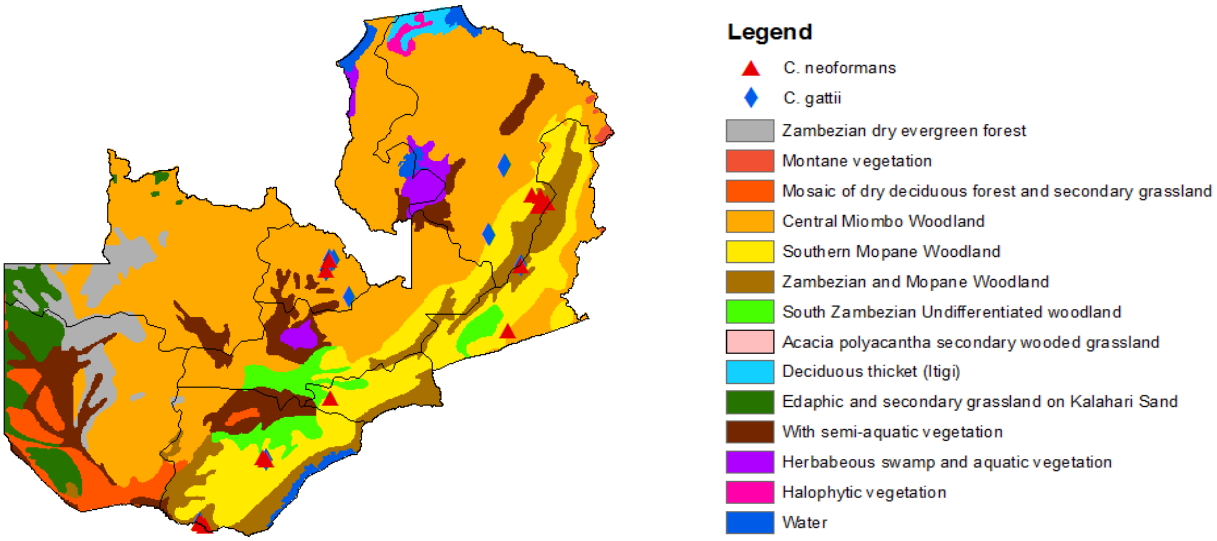


Figure 2.4: Recovery of *Cryptococcus* species-complex from Zambia. The map displays the different Zambian Ecoregions (White 1983). *Cryptococcus neoformans* isolates appear in red and *C. gattii* recovered are in blue.

2.3.2 South Africa

Cape Town

Most of the South African clinical isolates used in Chapter 5 were isolated from patients living in Cape Town. In order to link environmental and clinical isolates, samples were obtained from trees, soil and pigeon guano in different Cape Town areas: Nyanga, Mowbray, Milnerton, Guguletu and Postdam Road. A total of 97 trees were sampled, however these yielded no

positive characteristic brown colonies. 37 samples were obtained from pigeon dropping leading to the isolation of two *C. neoformans* isolates, one in Mowbray and one in Nyanga Table 2.3.

Table 2.3: *Cryptococcus* species complex Recovery in South Africa

Place	Total	Soils	Trees	Pigeon Droppings	*Cng*	Recovery (%)
Cape Town	124	43	44	37	2	1.61%
Eastern Cape, Fort Baufort	95	44	45	6	5	5.26%
Eastern Cape, Rural	55	19	20	16	3	5.45%
Northern Cape	44	16	18	10	1	2.27%
Namibia	103	10	59	34	2	1.94%
Total	421	132	186	103	13	3.09%

Eastern Cape

The region is characterised by a leached, rocky soil dominated by non-grassy forbs such as *Rhus discolor* or *Helichrysum cerastioides*. Samples were recovered from both urban and rural environments.

Urban samples were collected in Fort Beaufort, a town of 25,000 inhabitants located 130 km inland. Fort Beaufort was an appropriate town to sample as it is located in the Eastern Cape and a high number of people living in Cape Town township originate from this region. 89 samples were collected from trees and neighbouring soil (trees n=45, 3 positives; soils n= 44, 2 positives). Six samples were collected from pigeon droppings but no further isolates were recovered. Samples were collected from rural environments around Fort Beaufort (trees n=20, 2 positives; soils n= 19, 1 positive). An additional 16 samples were collected from pigeon excreta but no further isolates were recovered.

Northern Cape and Namibia

Samples were isolated in the Northern Cape up to the Namib Desert. This region experiences an extreme climate with temperature frequently surpassing 40°C (White, 1983). Two *Cryptococcus neoformans* var. *grubii* were isolated in Namibia from pigeon dropping (n=34, 2 positives). One isolate was also recovered in the Northern Cape from an unknown tree species (Table 2.4.

2.4 Discussion

The fieldwork conducted in Zambia and South Africa was successful in recovering isolates of the *Cryptococcus* species-complex. The selective medium, NS agar, developed by Staib (1987) and commonly used to identify this pathogenic yeast allowed the identification and isolation of either *Cryptococcus neoformans* or *C. gattii* from amongst a wide swathe of fungal biodiversity. After DNA extraction, PCR amplification of the *CAP59* and *SOD1* correctly differentiated the two species and 38 *C. gattii* and 37 *C. neoformans* were subsequently identified (Figure 2.5).

Table 2.4: *Cryptococcus*species complex Recovery from the Environment: recovery types were not found to be significantly different using a chi-squared test.

	C. neoformans		*C. gattii*	
	Isolates	%	Isolates	%
C. mopane	18	48.65%	2	5.26%
Brachystegia sp.	1	2.70%	6	15.79%
Soil	10	27.03%	13	34.21%
Pigeon Droppings	2	5.41%	0	0%
Other trees	6	16.22%	6	15.79%
Hyrax droppings	0	0.00%	11	28.95%
Total	37	100.00%	38	100.00%

2.4.1 Trees Species and Decaying Wood

C. neoformans was predominantly recovered from trees, mainly *C. mopane* (48.65%). This species of tree harbours decaying wood in the numerous boles found perforating the tree-trunk and *C. neoformans* has repeatedly been found on decaying wood (Cornelissen et al., 2003). This finding confirms the previous studies which identified *C. mopane* trees as a potential reservoir for *C. neoformans* (Litvintseva et al., 2011, 2005). However, the prevalence found in these two studies reached 30%, whereas in the present study, the *Cng* recovery rate in the Zambezi Mopane Woodlands (ZMW) ecoregion was 5.43% and 3.30% in the rainy and the dry season respectively. It may be that, because we exported the swabs to the UK before attempting recovery (compared to Litvintseva et al. whom undertook their isolations within Botswana), that the delay between swab-sampling and plating out the swab resulted in the deterioration of viable inocula.

C. gattii was not strongly associated with the ZMW (only one isolate recovered) and the species was found mainly to occur in the Central Miombo Woodlands (CMW). Here, 32 of the 38 isolates (84.21%) were found to occur in this ecoregion. The dominant tree type in the Miombo woodlands is *Brachystegia sp.* and six *C. gattii* isolates (15.79%) were recovered from these trees. Two isolates were also found on *Eucalyptus*, a tree species which has been associated with *C. gattii* for decades (Ellis and Pfeiffer, 1990; Pfeiffer and Ellis, 1992).

2.4.2 Soil

Ten *C. neoformans* (27.03%) and 13 *C. gattii* (34.21%) isolates were recovered from soil; this indicates that soil acts a potential ecological niche for the two yeast species. All soil samples were collected from soil surrounding the base of a tree. Previous research, which identified *Cryptococcus neoformans* in soil, emphasised that the soil could be contaminated by pigeon guano (Ajello, 1958; Casadevall and Perfect, 1998; Kidd et al., 2007) or that the pathogen could have originated from tree bark or the trunk hollow (Randhawa et al., 2008). Randhawa et al. reported similar recovery rate from Indian soil (*C. neoformans* 26%; *C. gattii* 24%). Additionally, 5% of their samples were positive for both species. In our study, this scenario occurred only once from a *C. mopane* swab in the SMW ecoregion. However, only five dark-brown colonies were processed per sample and this protocol may not fully reflect the *Cryptococcus* diversity present within each sample.

2.4.3 Pigeon Guano and Other Sources

Four *C. neoformans* isolates (3.48%) were recovered from pigeon guano. Two isolates were collected in small towns in the Namibian desert comforting the hypothesis that bird may act a mechanical carrier of the fungus. Nielsen et al. (2007) showed in vitro that *C. neoformans* could undergo prolific mating in a pigeon guano environment whereas the growth of *C. gattii* is impaired in these conditions. *C. neoformans* has been commonly associated with pigeon guano and found in many countries including Kenya (Kangogo et al., 2014), Iran (Hedayati et al., 2011), India (Khan et al., 1978), Thailand (Kuroki et al., 2004) or Columbia (Granados and Castañeda, 2005a). The identification of the two Namibian isolates is therefore an indicator that the dissemination of *C. neoformans* through avian species occurs.

The incidence of cryptococcosis is high in HIV-AIDS patients and highly fatal (Bicanic et al., 2007). Understanding the route of transmission of *C. neoformans* in urban environments could have important public health implications. Sampling was performed in densely populated areas; however only two isolates were recovered, one from Cape Town city centre and one from Nyanga township. These findings suggest that close proximity with pigeons could be a source for the propagation of viable inocula. Similarly, although the five *C. neoformans* isolates isolated in Fort Beaufort were collected from trees and soil, the omnipresence of the common pigeon, *Colombia livia*, in the town suggests that the majority of trees and surrounding soil were contaminated with pigeon faeces.

Regarding *C. gattii*, only three studies have been successful in isolating the pathogen in pigeon guano: in Kenya (Kangogo et al., 2014), India (Chowdhary et al., 2012) and Brazil (Lazera et al., 1993). Interestingly in this study, *C. gattii* was strongly associated with hyrax faeces (11 isolates in 52 samples). This finding reveals a new potential animal reservoir for *C. gattii* and stresses how ubiquitous this pathogen is. Additional sampling is required to better understand the evolutionary history of the *Cryptococcus* species-complex, especially in relation to their association with vertebrates.

2.4.4 Variation in Species Geography

In Zambia, the ZMW was the ecoregion which was strongly associated with *C. neoformans* (19/24 isolates), whereas *C. gattii* was strongly associated with the CMW (32/38 isolates) (Table 2.2); each of these species were significantly associated with a different ecoregion ($p<0.001$, calculated in a Pearsons chi-squared test). The Mopane woodlands are drier regions compared to the Miombo woodlands which represent the wettest part of Zambia. The different environmental associations of these *Cryptococcus* species reflects possible ecological specialisation of the two species. Previous research has shown that *C. neoformans* grows better than *C. gattii* in drier conditions (Ruiz and Casas, 1981; Caicedo et al., 1999). Granados et al. (2006) stated that the different geographic distribution of *C. neoformans* var. *grubii* and *C. gattii* could be explained by their different tolerance to climatic conditions. Potential climatic differences between the two Zambian ecoregions will be investigated in the following chapter.

C. neoformans was isolated in Zambia and South Africa, but no *C. gattii* was recovered from

South African samples. However, a previous survey in Cape town was successful in recovering four environmental *C. gattii* isolates from trees and soil (Simwami, 2011). Low incidence (<1%) has been reported in cryptococcosis patients in South Africa (Bicanic et al., 2007) and may reflect the lower presence of *C. gattii* in the environment . The areas investigated in the present study included large urban centre (Cape Town), smaller urban centre (Fort Beaufort) and rural environment. In these densely populated environments, only *C. neoformans* was isolated implying that the species might be more prevalent in close proximity to humans and therefore constitutes a problematic public health issue.

The higher recovery of *Cng* isolates might be explained by the proximity between cryptococcosis patients and bird. These allegations have been published multiple times in the literature (Currie et al., 1994; Garcia-Hermoso et al., 1997). Nosanchuk et al. (2000) are the only group who reported a case of zoonotic transmission. Other studies suggested cases of zoonotic transmission when the same genotype was isolated in patients and pigeon faeces living near the patient home (Franzot et al., 1999). The comparison between clinical and environmental genotypes are discussed in Chapter 5. Importantly, my isolation of environmental *C. neoformans* in Namibia strongly suggests that *Columba livia* plays a central role in the dissemination of cryptococcosis.

2.4.5 Seasonal Variation

Seasonal variation of the *Cryptococcus*-species complex has been little explored (Chowdhary et al., 2012). Abiotic factors such as soil moisture, pH, sun-light, wind or humidity are believed to play an important role in the distribution of fungal species (Kidd et al., 2007). My fieldwork was conducted in areas protected from sunlight in order to maximise the recovery as highlighted by Kidd et al..

In Zambia, the current work found a non-significant increase in the prevalence of *C. neoformans* during the dry season. The prevalence went from 1.46% in the rainy season to 2.38% in the dry season (p=0.476). Two Asian studies reported similar observations. First, Kuroki et al. (2004) reported higher prevalence in the dry season in their study in Thailand. Then, the seven-year retrospective study in Northern India led by Randhawa et al. (2008) also found lower prevalence of *C. neoformans* during the rainy season compared to any other season (p <0.001).

On the other hand, Granados et al. (2005a) found lower prevalence of *C. gattii* during winter. These results were also found in the Thailand study (Kuroki et al., 2004). The present study found a similar pattern, also not significant (p=0.219), with *C. gattii* prevalence reaching 3.56% in the rainy season and 2.28% in the dry season.

The seasonal results should be interpreted with caution. Fieldwork was performed in a limited time frame and over a restricted geographic range with the aim of maximising the number of isolates. The investigated areas remain limited and may not be fully representative of these species ecological niches. Moreover, sampling was performed without any suspicion of colonisation and sampling in the dry and rainy seasons included different areas. A comprehensive study accounting for resampling is needed to fully understand the seasonal variations of *C. neoformans* and *C. gattii*.

2.5 Conclusion

According to Xu et al. (2006a), *C. neoformans* and *C. gattii* speciated around 37 million years ago. The association of the two species to different ecoregions (p<0.001) emphasises that these pathogenic species have evolved to specialise in different environments. The drier conditions of the Mopane woodlands tend to agree with the current trend in *Cryptococcus* research stating that *C. neoformans* distribution is favoured over *C. gattii* in dry conditions (Granados and Castañeda, 2006). The discovery of a potential animal reservoir for *C. gattii* emphasises the lack of knowledge surrounding this pathogenic saprophyte. An increasing sampling effort is required to uncover the full evolutionary history of the *Cryptococcus* species-complex. Both species are able of large-scale colonisation and the effect of global warming might induce a striking change in the distribution of the pathogens (Kidd et al., 2004).

In the next Chapter, the distribution and ecology of *C. neoformans* and C. *gattii* in Zambia will be investigated further. The influences of abiotic factors, including temperature, precipitations or altitude on the soil fungal diversity present in the two Zambian ecoregions where *C. neoformans* and *C. gattii* were isolated will be investigated. This work will help us to understand the effect of seasonality on fungal diversity while describing the biotic factors associated with either of these two pathogenic yeasts.

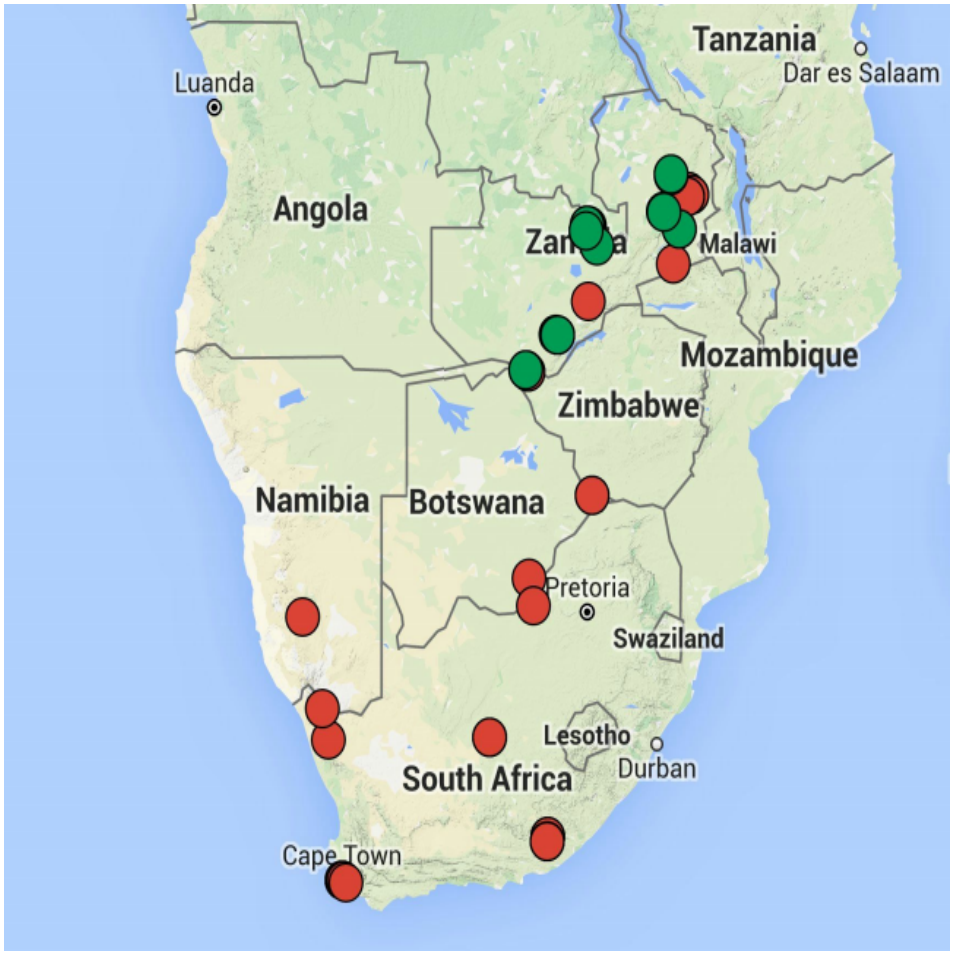


Figure 2.5: Recovery of *Cryptococcus* species-complex in Southern Africa - *Cng* isolates are in red (n=37) and *Cg* (n=38) are in green.

Chapter 3

Influences of Fungal Diversity on *Cryptococcus* Distribution

3.1 Introduction

3.1.1 Background

In the previous chapter, I have shown that *C. neoformans* and *C. gattii* were associated with different ecological regions in Zambia. *C. neoformans* was mostly isolated in the Zambezi Mopane Woodlands (ZMW) whereas *C. gattii* was principally found in the wet Central Miombo Woodlands (CMW). The two species appear to have diverged millions years ago and have the capacity to adapt to new environments and to colonise varied environmental niches (Springer and Chaturvedi, 2010). My study (Chapter 2) suggests the two environmental saprophytes are responding differently to environmental factors. Grandao et al. (2006) showed that the growth of *C. neoformans* var. *grubii* was favoured in dry conditions. The spread of the pathogen may have important public health implications since cryptococcosis is sometimes a fatal disease. A notable example demonstrating the potential of the *Cryptococcus* species-complex to invade new environment is the well-documented Vancouver outbreak (Kidd et al., 2004; Hoang et al., 2004). Emerging since 1999, the disease incidence rose in the region spreading from Vancouver Island to the Pacific Northwest, USA (MacDougall et al., 2007) and in that setting, the virulent VGII *C. gattii* strain was found to infect ostensibly healthy and immunocompetent patients.

Cryptococcus neoformans and *C. gattii* are considered as emerging pathogens. Although this is partly related to these fungi infecting an increasing pool of susceptible HIV/AIDS patients, spatial emergence is occurring. In a globalised world, the constant movement of microbes can lead to the emergence of diseases when an environmental pathogen acquires a selective advantage from environmental pressures which results in a virulent phenotype (Morse, 1995; Wolfe et al., 2007). In the study of environmental diseases, it is important to investigate potential environmental reservoirs and to gain knowledge on the ecology of the microorganism (Hiremath et al., 2008) in order to refine understanding of infection-dynamics. In Zambia, the landscape has changed rapidly over the last decades due intense deforestation, changes in human demography and the development of industrial agriculture (Chidumayo, 2013). The fieldwork described in the previous chapter revealed that Zambia provides an adequate set of spatial scales and ecosystems within which to study the physical and biological factors affecting the *Cryptococcus* species-complex in its natural environment. Hardin (1960) formalised the principle of competitive exclusion; this states that two species which are competing for the same resources, if ecological factors remain constant, cannot coexist with the same population size. The realised niche represents the actual niche where a given species is present, constrained by biotic and abiotic pressures. The fundamental niche is the complete range that a species can potentially occupy (Connell, 1961; Hutchinson, 1959). Therefore, current hypotheses of the worldwide dissemination of *C. neoformans* are that the species gained the ability to sexually reproduce in pigeon guano, which became its realised niche. The pathogen then was able to spread worldwide using bird migratory routes (Casadevall and Perfect, 1998), and thus 'explored'and exploited its fundamental niche. On the other hand, *C. gattii* is thought to be more restricted to its realised niche, which is believed to be sedentary trees (Nielsen et al., 2007).

Interactions with other fungal species or predators can lead to a selective advantage which may help the organism to expand its physical environment. The interaction of *Cryptococcus neoformans* with environment predators will be discussed in Chapter 5. Desnos-Ollivier et al. (2010) described the ecological niches of *C. neoformans* as 'extraordinarily complex', due to their probable inclusion of other fungi, bacteria, protists or viruses. Many studies have discussed the role of ecological interactions within microbial communities and how these relationships can alter species distribution on large geographical scales (Johnson and Stinchcombe, 2007). The previous chapter described the recovery of *C. gattii* and *C. neoformans* in sedentary trees and

my current study will take these observations further by investigating the fungal biodiversity that is associated with the *Cryptococcus* species complex by using a metabarcoding approach.

3.1.2 Importance of understanding fungal biodiversity

In the last decade, advances in metagenomics and genomics have allowed the uncovery of an unexpected amount of genetic diversity which could not have been imagined using solely culture-dependent technics (Xu, 2006b). When studying the ecology of microorganisms, high-throughput metabarcoding has rapidly gained importance. Exploring microbial biogeography allows to uncover the role of microbial dispersion patterns in human health (Koren et al., 2012; Adams et al., 2013) or in agriculture (Peiffer et al., 2013) by revealing the associations between microbial communities and environmental conditions.

Fungi represent an essential element of biodiversity and yet a limited number of studies have explored this kingdom (Desprez-Loustau et al., 2007). Emerging infectious diseases are defined as pathogens expanding their host or geographical range. Around 30% of plant emerging diseases are caused by fungi (Anderson et al., 2004). *Batrachochytrium dendrobatidis* has been introduced outside its endemic range and is now threatening amphibian populations worldwide (Fisher et al., 2012). Fungal diversity studies are revealing large numbers of unknown species and increasing our current knowledge on fungal biogeography and ecology is necessary in order to assess the risk that unknown (or little known) species may pose to human, animal and plant health. To date, a very low number of studies have investigated fungal biodiversity in Africa and recent estimations affirm that the continent has lower fungal richness than other tropical regions (Tedersoo et al., 2014).

The present work attempts to gain insights into the allopatric species-ranges that are observed in the previous chapter by investigating the different taxa and the variation of the abiotic environment associated with either *C. gattii* or *C. neoformans* habitats. *Cryptococcus neoformans* has been isolated on various substraits. Soil accounts for 20 to 25% of *Cryptococcus* recovery in environmental surveys (Chapter 2, Randhawa et al. (2008)). *Cryptococcus* species are ubiquitous in soil environments and have been found to predominate in desert soil, wetter or vegetated soil. Their recurrent recovery in these different environments is believed to be due to the polysaccharide capsule which also allows *Cryptococcus* species to compete and survive

predation by other microorganisms (Vishniac, 2006).

3.1.3 Benefits of Understanding the Ecology of Yeasts in Soil

The 80,000 identified fungal species are divided into seven phylum namely: Chytridiomycota, Zygomycota, Glomeromycota, Basidiomycota, Ascomycota, Rozellacea and Microsporidia (Moncalvo, 2005). In soil, fungi can be differentiated in two groups, the filamentous multicellular organisms and the unicellular yeasts. Yeasts include species of both Ascomycota and Basidiomycota (Barnett, 2004). To date, studies on the ecology of soil yeasts in their natural environments have been neglected (Botha, 2011). Soil constitutes dynamic and highly complex environments, and a better understanding of its structure and diversity is required to understand the fungal interaction with other organisms (Lim et al., 2010). Yeasts, in particular, play a central role in the degradation of organic materials and the dissipation of carbon in soil ecosystems (Cornelissen et al., 2003). Previous work has shown that yeasts are also involved in the maintenance of soil structure by generating polymeric compounds which bind soil particles (Yurkov et al., 2011). In 2010, 1,500 yeast species had been reported, but recent estimations evaluate the number of yeast species to be around 15,000 (Lachance, 2006). The recent advances in high-throughput metabarcoding have not yet been applied to a large extent for yeast ecology, and most fungal diversity studies are based on cultivation-based approaches (Botha, 2011). However, traditional molecular methods which involve the culture of microorganisms in the laboratory are non-quantitative, time-consuming and non-representative of sample diversity. Only a small proportion of organisms can grow on a selective medium and these methods cannot accurately assess microbial communities when compared to a high-throughput approach (Xu, 2006b). My study will compare traditional Sanger-sequencing techniques from yeast isolated on culture plates against a high-throughput Illumina MiSeq-based metabarcoding approach.

3.1.4 Use of Internal transcribed spacer (*ITS*) to Investigate Fungal Diversity

The ribosomal RNA (rRNA) operon contains the 5'external transcribed sequence (5'ETS), 18S rRNA, *ITS1*, 5.8S rRNA, *ITS2*, 28S rRNA and the 3'ETS, as shown on figure 3.2. During the rRNA maturation process, the two ribosomal intergenic spacer (*ITS*) regions are excised

(Christ et al., 2011). The two ribosomal spacers, *ITS1* and *ITS2* have high evolutionary rates and are highly variable. Thus, the *ITS* region provides an ideal target to detect and delineate fungal species, and has been extensively used for the taxonomic profiling of fungi (Lindahl et al., 2013; Xu, 2006b). In 2012, this genetic marker was proposed as the universal genetic barcode for fungi (Schoch et al., 2012; Xu, 2006b). A comparison between *ITS1* and *ITS2* revealed that *ITS2* was more variable and allowed a better assessment of the fungal diversity (Bazzicalupo et al., 2013). Therefore, the current study uses *ITS2* for metabarcoding sequencing, however I will also describe investigation of the fungal soil diversity using culture-dependent *ITS1* Sanger-sequencing.

Aims:

1. Characterise the fungal community composition associated with *C. gattii* and *C. neoformans* environments
2. Understand regional ecology of the two pathogenic species and understand the role of ecology in the allopatric/parapatric speciation previously observed
3. Assess and compare the fungal diversity present in Zambian soil over space and time using *ITS2* Metabarcoding Approach
4. Compare conventional *ITS1* Sanger-sequencing methods with a Metabarcoding *ITS2* approach to uncover fungal diversity

3.2 Methods

3.2.1 Sequencing of the *Internal transcribed spacer 1* (*ITS1*) in Isolated Yeast

Internal transcribed spacer 1 (*ITS1*) Amplification

Environmental samples were inoculated in 80 μl of water in a 96-well plate. Samples were heat shocked at 100°C for 15 min. The PCR reaction was carried out in a final volume 50 μl of using the following reagents, 17.65 μl sterile distilled water, 2.5 μl Qiagen buffer (10X), 1 μl

forward primer *ITS1F* (5- CTTGGTCATTTAGAGGAAGTAA-3) (Gardes and Bruns, 1993) at 10 μM dilution, 1 μl forward primer *ITS4* (5- TCCTCCGCTTATTGATATGC-3) (White et al., 1990) at 10 μM dilution, 1.6 μl dNTPs (2.5 mM) and 0.25 μL Qiagen Taq polymerase. Then, the 48 μl of master mix were added with 2 μl of diluted DNA to each well on a 96-well microtitre PCR plate. Both positive and negative controls were included in each run. The plates were then carefully sealed, vortexed and centrifuged for 10 sec at 1,000 rpm. Using the thermocycler DNA engine, the Polymerase Chain Reaction was performed under the following cycling conditions 94°C for 1 min as an initial denaturation step followed by 35 cycles of 94°c for 1 min, 51°C for 1 min and 72°C for 1 min, with a final elongation step of 72°C for 8 min. Amplicons were then stored at -20°C before clean-up. Correct amplifications were assessed by gel electrophoresis on 1.5% agarose gels, with positive control.

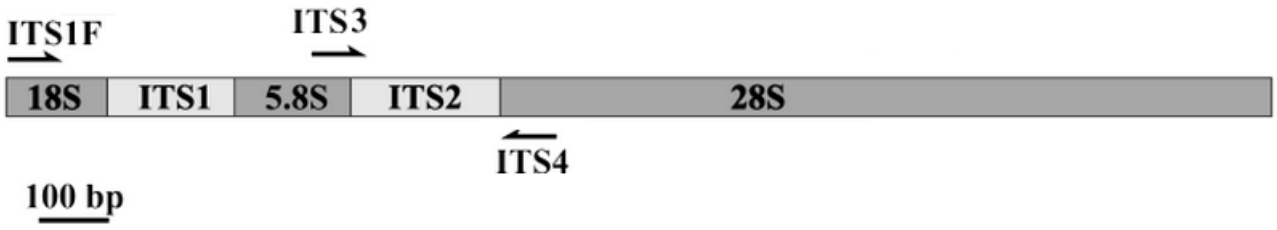


Figure 3.1: Fungal *internal transcribed spacer* (*ITS*) of ribosomal RNA and its associated primers - *ITS3* and *ITS4* were used for the metabarcode of the *ITS2*. *ITS1F* and *ITS4* were used using the Sanger-Sequencing method.

PCR cleanup

In order to remove excess of dNTPs and primers, 60 μl of 20 PEG/2.5 M NaCl were added to each microtitre plate well. Plates were then sealed and vortexed before been spun at 1,000 rpm for ten seconds. Plates were then incubated 4°C overnight. To pellet the PCR products, plates were centrifuged at 3,200 rpm for one hour at 4°C. Supernatant was discarded by inverting the plate onto a folded paper towel. Wrapped in cling film, the inverted plate was centrifuged at 500 rpm for 20 sec. Film was removed and the plate was inverted onto a fresh paper towel and centrifuged for one minute at 500 rpm to remove any remaining PEG/NaCl.

The DNA pellets were washed by adding 150 μl of 70% ethanol. The plates were sealed and centrifuged for 20 min at 3,200 rpm. The supernatant was discarded by inverted centrifugation for one minute at 500 rpm onto a fresh paper towel. Wash in 70% ethanol was repeated and the

supernatant was discarded. The pellets were dried at 37°C for two minutes using a thermocycler with the lid open and the plates left unsealed. 10 μl of deionized water were added to each well. Plates were sealed and vortexed at 1,000 rpm for ten seconds to resuspend the pellets. Plates were incubated for ten minutes at 4°C and the vortexed once again for 10 seconds at 1,000 rpm.

DNA concentrations were measured in order to carry subsequent dilutions and obtain 5 ng/μl of DNA. The sample concentrations were estimated using a NanoDrop 1000 Spectrophotometer (ThermoScientific 2008). The sensor was first washed with 1.2 μl of double distilled water and the blank was established using 1.2 μl of double distilled water. Each measurement was made by pipetting 1.2 μl of the elute sample on the end of the fibber optic cable. The appropriate dilutions were performed and the clean PCR products with the required DNA concentration were kept at -20°C until the next step.

Sequencing reactions

Two master mixes containing each primer, *ITS1f* and *ITS4*, were set up. The 96-well microtitre plate was divided in two, each half containing one primer. Each well contained 1.75 μl of water, 4 μl of the forward or reverse primer (1 pmol/μl), 0.5 μl of Tag-FS (BigDye V1.1) and 1.75 μl of the sequencing buffer. 2 μl of purified PCR products were added to 8 μl of master mix. The plate was then sealed and vortexed. Centrifugation for ten seconds at 1,000 rpm allowed to settle the liquid at the bottom of the well. Plate was then loaded onto a thermocycler using the following program:

Control method: calculated

1. Denaturation: 96°C for 0:10
2. Annealing:50°C for 0:05
3. Extension: 60°C for 2:00
4. go to 1, (x 24)
5. Ramp: 0.10 /s to 4°C

6. Hold: 4°C for ever

Amplicons were then kept at -20°C until DNA precipitation step.

DNA Precipitation

9 μl of 3M NaCl pH 5.2 (4:11) and 51 μl of 95% ethanol were added to each well. Plates were then sealed and vortexed before been spun at 1,000 rpm for 10 seconds. Plates were incubated at -20°C for one hour. The plates were centrifuged at 3,500 rpm at 4°C for one hour. Once centrifuging was completed, the seal was immediately removed and the plates were inverted onto a paper towel to discard the supernatant. Particular attention was paid when handling the plates in order to keep the pellets undisturbed. Inverted plates were then transferred onto a fresh paper towel and spun for one minute at 500 rpm to remove the remaining ethanol. Two consecutive washes were performed using 150 μl of 70% ethanol. The plates were sealed and spun at 3,500 rpm for 30 minutes. Supernatant was discarded by inverting the plate on a paper towel and then centrifuging the inverted plate on a fresh paper towel at 500 rpm for 1 minute. The second spin was performed at 800 rpm for one minute. Finally the plates were sealed and stored at -20°C until ready for loading. Pellet were re-suspended in Hi-Di Formamide and an Applied Biosystems 3730xl DNA analyser (Warrington, UK) was used to determine the DNA sequences of the PCR products.

Cleaning of ABI sequences

Sequences were manually edited manually using CodonCode Aligner (CodonCode Corporation, MA, USA) according to the guideline provided by Taylor & Houston (2011). A 200-bp cut-off was used for trimming the sequences before assembling each contig. Once the consensus sequences were generated, low-quality bases were removed using the Mask Perl script provided on the Fungal Metagenomics Web portal (http:www.borealfungi.uaf.edu). Consensus bases with a phred scores below 20 were changed to Ns. The *ITS1* region was split from the LSU fragment by first aligning the cleaned sequences with Clustalw (Larkin et al., 2007). Then, the alignment was converted to a fasta format and analysed in BioEdit (Hall, 2011). The reverse complement of the *ITS4* primer site (GCATATCAATAAGCGGAGG) was identified

and manually removed with the downstream sequence. Similarly, the forward primer sequence of the *ITS1f* was also removed (CTTGGTCATTTAGAGG). Additional clean-up was performed using TrimSeq (http://imed.med.ucm.es/cgi-bin/emboss.pl) to remove poorer bases located at the beginning and the end of the sequences. The window size was set to 40 with a 5% ambiguity. Finally, the *ITS* sequences were grouped into Operational Taxonomic Unit (OTUs) using the Fungal Metagenomic Web Site, which uses the TGICL program (Pertea et al., 2003) followed by BLAST.

3.2.2 Metabarcoding of *ITS2*

DNA extraction and PCR amplification of the genomic DNA was extracted from 74 soil samples collected in various part of Zambia using the PowerSoil DNA Isolation Kit (MOBIO Laboratories, Inc.). A negative control was used to account for eventual taxa resulting from laboratory contamination. The genomic DNA was then purified and Illumina paired-end libraries were prepared according to the Illumina protocol. The *ITS2* region was amplified using an Illumina MiSeq instrument on a 300-bp paired-end sequencing run.

Environmental variables

A number of environmental data were incorporated into the analysis in order to explore the relationship between environmental factors and the *Cryptococcus* species-complex. Data was selected aiming to measure the operational environment of the saprophyte. Other studies tried to model the fundamental niche of the pathogen (Mak et al., 2015; Mak, 2007). *Cryptococcus* is known to be sensitive to temperature and atmospheric conditions and these variables are investigated using the WorldClim Global Climate Data (Busby, 1991). Additionally, altitudinal influence, Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) will be considered as they are commonly used as predicators in disease mapping (O'Hanlon et al., 2016).

Environmental variables were extracted from WorldClim Global Climate Data (Busby, 1991) and the soil information were found using the Africa Soil Information Service (Hengl et al., 2015). The information for each sample was extracted in R software (v 3.1.1) using the raster (Hijmans et al., 2012) and dismo (Hijmans and van Etten, 2012) packages. The enhanced

vegetation index (EVI) and the Normalized Difference Vegetation Index (NDVI) from 2012 and 2014 were extracted and averaged over this period using MODISTools (Tuck et al., 2014).

Sampling and DNA Extraction

For each sample, between 5 to 10 g of Zambian soil were collected in various ecoregions in Zambian during both the dry and rainy seasons. Collection in the rainy season occurred in January, 2013 and sampling in the dry season was performed in September 2013. Soil was kept at 4°C during field collection and frozen at -20°C in the lab until being processed. DNA was extracted using the PowerSoil® DNA Isolation kit (MOBIO, cat. 12888-100). First, 0.25 g of soil was weighted and transferred to PowerBead Tubes. Tubes were vortexed and 60 μl of the solution C1, which contains lytic reagents, were added to the soil in order to lyse the cells. Tubes were centrifuged at 10,000x g for 30 sec and the supernatant was transferred into a 2 ml collection tube. 250 μl of solution C2 solution were added to the mixture. Solution C2 allows non-DNA organic and inorganic material to precipitate. After vortexting for 5 sec, the solution was incubated for 5 min at 4°C. Tubes were then centrifuged for 1 min at 10,000x g and 600 μl of the supernatant were transferred to a new 2 ml collection tube. 200 μl of solution C3 were added to the soil extract to further pellet non-DNA materials. After vortexing, 750 μl were transferred to another collection tube and 1.2 ml of solution C4 were added. The solution C4 is high in salt content which allows the DNA to bind onto the silica membrane in a spin filter. The soil extract was loaded onto the spin filter 675 μl at a time. The process was repeated three times. The spin filter was washed with 500 μl of the ethanol-based solution C5. The solution helps to remove residual salt, humic acid and other contaminants. Finally, the spin filter were spun at 10,000x g for 1 min to remove any remaining ethanol and 100 μl of the elution buffer was added at the centre of the membrane. The DNA was collected by centrifugation at 10,000x g for 30 sec and stored at -20°C.

Amplicon PCR

Paired-end libraries were prepared for sequencing in a two-step PCR approach using the Illumina 16S Metagenomic Sequencing Library Preparation Protocol. The *ITS2* rDNA region was amplified using *ITS3_KYO2* and *ITS4_KOY3* (Toju et al., 2012). The two primers were paired

with appropriate Illumina adapter overhang nucleotide sequences and a 2-bp linker sequence. The sequence of the two primers with linker and overhang sequences are detailed on figure 3.3.

Amplifications were carried out in at total volume of 25 μl using 2.5 μl (5ng/μl) of DNA, 12.5 μl 2x KAPA HiFi HotStart Ready Mix (ANACHEM, catalogue: KK2602) and 5 μl (1 μM) of each primer.

PCR Conditions for *ITS3_KYO2/ITS4_KOY3*: Control method calculated

1. Initial Denaturation: 95°C for 3:00
2. Denaturation: 95°C for 0:30
3. Annealing: 55°C for 0:30
4. Extension: 72°C for 0:30
5. go to 2, (x 25)
6. Final Elongation: 72°C for 5:00
7. Hold: 4°C for ever

Products were then kept at -20°C until DNA precipitation step. The correct amplification were visualised using Agilent 2200 TapeStation (Agilent Technologies, Inc).

PCR Clean-up

The amplicons were purified using Agencourt AMPure XP magnetic beads according to Illumina protocol. The amplicon PCR plate was centrifuge at 500 rpm for 1 min to collect condensation. The AMPure XP beads were then vortexed for 30 seconds and 20 μl of beads were added to each well. After pipetting the entire volume up and down ten times, the plate was incubated at room temperature for five minutes to allow the DNA to bind onto the beads. Then, the plate was placed onto a magnetic stand for two minutes and the supernatant was discarded. The AMPure XP beads were then washed with 200 μl freshly prepared 80% ethanol. After a 30 seconds incubation, the ethanol was removed. Two washes were performed. The beads were air-dried for 10 minutes to remove the residual ethanol and 52.5 μl of 10 mM Tris pH 8.5

was added to each well to resuspend the DNA. After resuspention, the plate was placed on the magnetic stand for two minutes and 50 μl of the supernatant containing the PCR amplicons were placed into a new 96-well plate. Samples were then kept at -20°C until the Illumina adapters were attached to the amplicons.

Index PCR

The dual indices and Illumina sequencing adapters were attached to the amplicon using the Nextera XT Index Kit. 5 μl of each well containing the cleaned DNA was transferred into a new 96-well plate. Then, 5 μl of the correct Index 1 (N701-N712) and 2 (S501-S508) primers were added to each well. The primers add two 8-bp indices which allow the correct identification of each sample after sequencing. 10 μl of nuclease free H_2O and 25 μl 2x KAPA HiFi HotStart Ready Mix were added to the reaction for a total of 50 μl. The PCR reaction was carried out in a thermal cycler using the following conditions:

1. Initial Denaturation: 95°C for 3:00
2. Denaturation: 95°C for 0:30
3. Annealing: 55°C for 0:30
4. Extension: 72°C for 0:30
5. go to step 2 (x 8)
6. Final Elongation: 72°C for 5:00
7. Hold: 4°C for ever

Finally, two additional PCR Clean-Up were performed as described previously to obtain the final libraries.

Validation of Library

The libraries were validated using Agilent 2200 TapeStation (Figure 3.4) and Kapa qPCR kit (KK4824, Kapa Biosciences). The TapeStation 2200 with D1000 Screen Tapes allows to

obtain the accurate mean fragment length of each library. The quantitative PCR (qPCR) using the Kapa qPCR Universal kit for Illumina Libraries provides an accurate quantification of the enriched library fragments. The information of these two assays is subsequently used to determine the dilution needed for normalising each library to the same concentration and the pooling of each library.

TapeStation 2200 with D1000 Screen Tapes

3 μl of D1000 sample buffer were added to each tube with 1 μl of DNA library. The tubes were then spun for 1 minute in picofuge. The mean band size was estimated by defining a region on the electropherogram (Figure 3.4).

Kapa qPCR

For Kapa qPCR assay, each library was diluted to 1/10,000. The master mix was prepared by adding 360 μl of nuclease free H2O and 40 μl of ROX High. 16 μl of the reaction mix were added to each well of a 96-well ABI Taqman plate (4306737, Life Tech). Then, 4 μl of the diluted libraries were added to each well in triplicate. The ABI plate was then run onto the 7300 Real Time Cycler with the following conditions: 95°C for 5 min, then 35 cycles of 95°C for 30 sec and 60°C for 45 sec. After analysing the amplification plots (Figure 3.5a) and linear plot (Figure 3.5b) the concentration were estimated and normalised using the average fragment size obtained with the TapeStation.

Amplicons were pooled into one tube and submitted to the Imperial BSR Genomics Facility for 300-bp Illumina paired-end sequencing on an Illumina MiSeq v3 sequencer using 10% PhiX.

3.2.3 Statistical Analysis

Raw Illumina fastq files were merged into paired-end reads using PANDASeq (Masella et al., 2012). Then, the clustering of sequences, their taxonomic assignments and the analyses were performed using the quantitative insights into microbial ecology (QIIME) software v1.8.0 (Caporaso et al., 2010). The clustering and assignment of reads into operational taxonomic units (OTUs) were achieved using the OTUs-picking workflow in QIIME. The reads were grouped using USEARCH and UCLUST, the algorithms divide sequences into clusters using a 97% threshold of pairwise identity and a maximum e-value of 0.001 (Edgar, 2010). The OTU table

was rarefied at a sequencing depth of 300 to remove sample heterogeneity. The UNITE fungal ITS database (v.12.9) (Abarenkov et al., 2010) was used to map OTUs to a reference and the identification of extracted *ITS2* OTUs was performed using BLAST.

Alpha and beta-diversity were calculated using QIIME. Diversity indices were first introduced by Whittaker in 1960. Alpha diversity is known as the species richness of a particular community (Whittaker, 1960) and β-diversity is the extent of change of community composition. Rarefaction curves are obtained from alpha diversity measures (observed OTUs and chaos1 estimate) and allowed to assess taxonomic richness of the samples (Magurran, 2013). eta-diversity was estimated using the Bray-Curtis metric. This method allows to calculate pairwise distances between fungal communities. Principal Coordinate Analysis (PCoA) were computed using the core-analysis framework within QIIME.

To investigate potential differences in microbial β-diversity between ecoregions or seasons, the ANalysis Of Similarity test (ANOSIM) and ADONIS with 999 permutations were used between the different categories based on Bray-Curtis dissimilarity. The conjunction of the two tests was used to assess the potential variations between groups (ecoregions or seasons).

ANOSIM is based on a standardised rank correlation analysis between two matrices. The test is commonly used in community ecology. This test is primarily used to compare the variation in species composition and/or abundance among sampling units. The significance is achieved through permutations. ANOSIM examined the variation in species composition among different grouping factors. Samples are assigned to groups and ANOSIM test whether there are significant differences between these groups (Clarke, 1993). A p-value of 0.001 indicates significant differences between the groups at $\alpha = 0.05$. An R coefficient superior to 0.25 implies that groups are different with some overlap. If the R coefficient exceeds 0.5 the groups are considered different (Fierer et al., 2010). ANOSIM appears as a modified version of the Mantel test but uses a model matrix coding for each group (or treatment levels).

ADONIS is similar to ANOSIM and is widely used in analysis of ecological community data. It is common practice in community ecology to combine these two statistics. The ADONIS method is nonparametric and assesses the significance between sample groups based on a distance matrix. The analysis is similar to ANOVA. The statistical significance is achieved by partitioning the sum of squares of the data set based on permutations and using Bray-Curtis matrices. Then, the method computes a R^2 coefficient which transcribes the percentage of variation explained

by the categories and gives the statistical significance (Anderson, 2001).

One-way ANOVA (ANalysis Of Variance) was used to identify taxa which differed between sample groups. A best variables rank correlation test (BEST) was then performed to rank the relative importance of environmental conditions on β-diversity. The BEST analysis identifies subsets of variables whose Euclidian distances are maximally correlated with the Bray-Curtis matrix. The correlation is computed using a Spearmans rank correlation coefficient (Spearman, 1904). Multivariate analysis of variance (MANOVA) also known as permutation ANOVA was performed within the two main ecoregions (CMW and ZMW) with 999 permutation. This method is similar to ADONIS as it partitions sums of squares. The method was used to assess the influences of the climatic factors within each ecoregion by returning a pseudo-F value and a p-value.

Results were finally visualised in GenGIS (Parks et al., 2009) and the UPGMA tree was generated in QIIME. Only OTUs present in more than 0.001% of the total filtered sequences were considered when comparing ecoregions and seasons. In order to highlighted the taxa overlap within the different ecoregions, Venn Diagrams were generated based on OTUs presence and visualised using http:bioinfogp.cnb.csic.estoolsvennyindex.html (Oliveros & Venny n.d.).

3.2.4 Ecological Niche Modelling Maxent

MaxEnt uses maximum entropy to model species geographic distribution (Phillips and Dudík, 2008) based on presence-only data. The software was used to generate the predicted distribution of the two sister species in Zambia. The distribution was modelled using Bioclim layers (Busby, 1991) and altitude. The niche model produced by MaxEnt is an approximation of the species ecological niche within the context of the environmental dimensions that I investigated. The purpose of the software advanced by Phillips et al. (2006) is to make predictions from incomplete information. Maxent calculates the probability distribution of the maximum entropy. The latitude and longitude of the isolates constitutes the sample points and the Bioclim layers the environmental features.

3.3 Results

3.3.1 Fungal Diversity using Metabarcoding *ITS2* Approach

The metabarcoding approach was successful in obtaining operational taxonomic units (OTUs) from Zambian soil. Results revealed that regional origin is associated with the diversity of fungi that I detected despite the high variability of OTUs present across samples. Further, fungal diversity within ecoregions displayed variations that were associated with season.

The MiSeq sequencing run resulted in 8.44 million paired-end reads. Seven samples were removed from the analysis because the sampling depth was below 300 reads. 72,959 reads were clustered into OTUs and assigned to fungi. A total of 11,475 OTUs were called. These were distributed evenly between the 61 samples (Mean 81 OTUs/samples - Table S2). The taxonomic assignment revealed that Ascomycota (34.67%) was the most predominant phylum among the samples. Basidiomycota (11.10%) was the second most abundant phylum, followed by Zygomycota (7.92%), Chytridiomycota (3.63%) and Glomeromycota (1.12%). Additionally, 19 sequences matched the phylum Blastoclodiomycota. 36.19% of the OTUs were assigned to fungi but did not match any phylum and 5.38% of OTUs did not match any sequences in the NCBI database. Reads from the negative control were only assigned at the genus level to *Trichosporon* and others were also assigned to unidentified fungi.

The fungal community structure varied significantly across the three ecoregions investigated (R_{ANOSIM}=0.3704, p <0.001) (Table 3.2). Elevation, latitude and longitude improved regression coefficients indicating that these factors are playing a significant role in the distribution of fungal communities (Table 3.1). These findings were corroborated by ADONIS except for longitude which was not significant (R^2_{ADONIS} = 0.0183, p <0.198).

Analysis of variance (ANOVA) allowed me to identify taxonomic features which underlined varietal patterns. These trends included *Rhynchostoma* and *Chorioactis* along with unknown Chytridiomycota and Ascomycota (Table 3.3). Interestingly, the genus *Chorioactis* has so far only been known to occur in Japan and Texas, USA (Pfister et al., 2008) No Basidiomycota were significant when accounting for multiple sampling.

Table 3.1: ANOSIM and permutation MANOVA of microbial diversity patterns across Zambian Ecoregions

		Bray-Curtis			
		ANOSIM		ADONIS	
Group	Factor	R	P	R	P
Ecoregions	Ecoregions	0.3704	0.001	0.0687	0.001
Ecoregions	Elevation	0.4272	0.001	0.0374	0.001
Ecoregions	Longitude	0.4191	0.001	0.0183	0.198
Ecoregions	Latitude	0.4184	0.001	0.0242	0.006
Ecoregions	Season	0.3744	0.001	0.0687	0.001

Table 3.2: Relative OTUs abundance

Phylum	OTUs/Sample	Relative Abundance
Ascomycota	234.93	34.67%
Basidiomycota	124.82	11.10%
Chytridiomycota	20.29	3.63%
Glomeromycota	18.11	1.12%
Zygomycota	11.68	7.92%
Other	-	5.38%
Unidentified	-	36.19%
Total	11475	100%

Table 3.3: ANOVA - Important Taxonomic Variations between Ecoregions

OTU	P	FDR	Bonferroni	taxonomy
OTU1737432	3.21E-013	1.62E-008	1.62E-008	*Fungi;Ascomycota;unidentified;unidentified;unidentified;unidentified;uncultured*
OTU685087	1.74E-010	4.40E-006	8.80E-006	No blast hit
OTU1903949	1.95E-007	3.05E-003	9.86E-003	*Fungi;Ascomycota;Eurotiomycetes;Incertae sedis;Rhynchostomataceae;Rhynchostoma*
OTU1669117	2.42E-007	3.05E-003	1.22E-002	*Fungi;Ascomycota;Pezizomycetes;Pezizales;Chorioactidaceae;Chorioactis*
OTU756802	4.18E-007	4.22E-003	2.11E-002	*Fungi;Chytridiomycota;unidentified;unidentified;unidentified;unidentified;unidentified sp.*
OTU532415	8.05E-007	5.87E-003	4.06E-002	*Fungi;Chytridiomycota;unidentified;unidentified;unidentified;unidentified;Chytridiomycota sp*
OTU564482	8.84E-007	5.87E-003	4.46E-002	No blast hit
OTU740873	9.31E-007	5.87E-003	4.70E-002	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU87825	1.34E-006	6.77E-003	6.77E-002	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU430390	1.34E-006	6.77E-003	6.77E-002	No blast hit
OTU1836829	1.50E-006	6.87E-003	7.56E-002	No blast hit
OTU1789046	3.12E-006	1.31E-002	1.57E-001	*Fungi;Ascomycota;unidentified;unidentified;unidentified;unidentified;unculturedAscomycota*

3.3.2 Microbial patterns within Ecoregions

The phylum abundance varies within each ecoregion. Generally, the number of sequences per sample was higher in Zambezian Mopane Woodlands (ZMW) samples (ZMW=207 OTUs/sample, CMW=149 OTUs/sample, SMW=53 OTUs/sample, p <0.001) (Table 3.4). In the ZMW, Ascomycota represented 35.61% (n_{OTUs}=3,356) of all OTUs, followed by Basidiomycota (11.15%, n_{OTUs}=1,754)). In the Central Miombo Woodlands (CMW), Ascomycota represented 32.69% (n_{OTUs} =2,392) and Basiodiomycota 11.20% (n_{OTUs}=1,740). The number of unidentified OTUs was slightly higher in the CMW (41.13%) compared to the ZMW (32.14%). Similar abundance was found in the Southern Miombo Woodlands (SMW) with Ascomycota representing 37.67% and Basidiomycota accounting for 9.74% of OTUs.

The fungal compositions of the different ecoregions investigated were compared (Figure 3.6). The rarefaction curves based on Chaos1 α-diversity metric showed a saturation within the three ecoregions indicating that the OTUs identified in the survey accounted for most of the fungal diversity present in Zambian soil. The number OTUs per sample was higher in the ZMW and the CMW and the SMW had a similar amount of fungal diversity. The similarity and differences in the genera detected in each region were assessed based on the relative abundance (RA) of each OTUs. Only genera with a relative abundance above 0.001 were considered for the analysis. A total of 2,148 genera were identified and distributed unevenly across each ecoregion (CMW = 754, ZMW = 1112 and SMW = 282).

Figure 3.6 displays a Venn diagram portraying the differences and similarities between the ecoregions. CMW and ZMW shared 59 genera which were absent in SMW. Although, the number of sequences per samples was subsequently lower in the SMW, the α-diversity observed the region was similar to that observed in the CMW (Fig 3.6b). Interestingly, the genus *Cryptococcus* was only present in the ZMW when considering a relative abundance that was above 0.001; this shows that this genus is more abundant across this ecoregion.

Figure S1 portrays the sample heterogeneity in fungal composition for the three ecoregions. The fungal composition of each sample was highly variable. The full listing of the assigned phylum for each OTUs and their relative abundance are listed in the appendix (Figures S2, S4 & S6) (Average: CMW = 195 OTUs, ZMW = 290 OTUs, SMW = 205 OTUs).

Phylum	ZMW		CMW		SMW	
	OTUs/Sample	%	OTUs/Sample	%	OTUs/Sample	%
Ascomycota	119.86	35.61%	92.00	32.69%	103.75	37.67%
Basidiomycota	62.14	11.15%	49.00	11.20%	60.13	9.74%
Chytridiomycota	9.86	4.33%	8.62	3.21%	8.50	2.27%
Glomeromycota	9.46	0.33%	7.12	2.13%	7.13	0.85%
Zygomycota	5.79	6.54%	4.54	8.11%	5.88	12.70%
Other	-	9.71%	-	1.54%	-	1.25%
Unidentified	-	32.14%	-	41.13%	-	35.51%
Total	207.11	100%	149.75	100%	52.96	100%

Table 3.4: Microbial patterns within Ecoregions

3.3.3 Differences in Fungal Community Structure between the Mopane and Miombo Regions

As described in Chapter 2, *C. gattii* was principally associated with CMW whereas *C. neoformans* was found in the ZMW (p <0.001; the ecoregion that showed a predominance of the genus *Cryptococcus*). The fungal Bray-Curtis distance measure was used to investigate the differences between the two regions in order to understand the role of biotic factors on the two pathogenic species.

Within the different ecoregions, significant differences were observed (Table 3.5). The CMW exhibited significant regional fungal diversity patterns (R_{ANOSIM}=0.4182, p<0.001) when compared to the other regions. A similar pattern was observed for ZMW (R_{ANOSIM}=0.2726, p <0.001). However SMW did not display any significant regional trend using ANOSIM (R_{ANOSIM}=-0.0161, P¡0.545). The result was however significant using the ADONIS approach (R^2_{ADONIS} = 0.0234, p <0.006) but the R^2 value remained small revealing very limited effects. However, strong differences in fungal composition were observed when comparing the fungal composition of each ecoregion to each other (R_{ANOSIM} >0.37 and p <0.001). Differences according to season are detailed on Table S2.

Observed Differences in the Rainy Season

During the rainy season, clear differences could be observed in the fungal community structure of CMW and ZMW (R_{ANOSIM}=0.3704, p <0.001). Figure 3.7 highlights these differences which can clearly be observed. The UPGMA tree (Figure 3.7a) shows a strong clustering of ZMW

Table 3.5: ANOSIM and ADONIS of microbial diversity patterns within Zambian Ecoregions

		Bray-Curtis			
		ANOSIM		ADONIS	
Group	Factor	R	P	R	P
CMW	Ecoregions	0.4184	0.001	0.0428	0.001
ZMW	Ecoregions	0.2726	0.001	0.0416	0.001
SMW	Ecoregions	-0.0161	0.545	0.0234	0.006
ZMW	CMW	0.3704	0.001	0.0687	0.001
ZMW	SMW	0.3748	0.001	0.0687	0.001
SMW	CMW	0.3706	0.001	0.0687	0.001

and CMW samples. The bar plot of the various OTUs identified either regions (Fig. 3.6b) presents two different OTUs profile and PCoA (Figure 3.7c) reveals a clear clustering of the samples per ecoregions.

The one-way ANalysis Of Variance (ANOVA) was then used to investigate the important taxonomic features that potentially explain the observed differences of microbial diversity. However, no taxa could be identified beyond the phylum level and no taxa was significant when accounting for multiple testing (Table 3.6).

Table 3.6: ANOVA - Important Taxa Variations between the Mopane and Miombo Region

OTUs	p <0.01	None significant after Bonferroni
OTU674	0.0030	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU3262	0.0031	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU1546944	0.0036	*unidentified;unidentified;unidentified;unidentified;unidentified*
EU490021	0.0054	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU801	0.0063	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU412154	0.0084	*Glomeromycota;unidentified;unidentified;unidentified;unidentified*
OTU501392	0.0084	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU1459398	0.0086	*unidentified;unidentified;unidentified;unidentified;unidentified*
EU489956	0.0093	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU1521484	0.0096	*unidentified;unidentified;unidentified;unidentified;unidentified*

Observed Differences in the Dry Season

In marked contrast, much less clustering of the OTUs was observed between the two regions in the dry season (Fig. 3.8a) and PCoA did not show a clear clustering of samples from the two ecoregions. The ANOVA results highlighted 27 OTUs (p <0.01) which did however exhibit

strong variation between ZMW and CMW. The *Trichosporon* genus was identified five times using ANOVA analysis (Table 3.7). However, these results were not significant when accounting for multiple testing and *Trichosporon* was present in the negative control and should therefore should discarded from the results.

Table 3.7: ANOVA - Important Taxa Variations between the Mopane and Miombo Region in Dry Season

OTU	p <0.01	Taxonomy
OTU557	0.0016	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedsoil*
OTU78022	0.0025	*Fungi;Ascomycota;Lecanoromycetes;Teloschistales;Physciaceae;Buellia;Buellia*
OTU731946	0.0037	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedectomycorrhizal*
OTU1326976	0.0038	*Fungi;Basidiomycota;Agaricomycetes;Agaricales;Cortinariaceae;Gymnopilus;Gymnopilus*
OTU909605	0.0053	*Fungi;Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon;Trichosporon*
OTU141168	0.0055	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU766080	0.0055	*Fungi;Ascomycota;Dothideomycetes;Pleosporales;Pleosporaceae;Epicoccum;Epicoccum*
OTU1095222	0.0055	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU1589497	0.0055	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU5	0.0059	*Fungi;nidentified;unidentified;unidentified;unidentified;unidentified;unculturedsoil*
OTU1359201	0.0060	*Fungi;Basidiomycota;Microbotryomycetes;Sporidiobolales;Incertae;;*
FR731483	0.0062	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedectomycorrhiza*l
OTU207015	0.0064	*Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;unidentified;unculturedGlomu*s
OTU1005413	0.0068	*Fungi;Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon;Trichosporon*
OTU1734084	0.0068	*Fungi;Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon;Trichospor*on
OTU1951264	0.0071	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU1521484	0.0074	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
DQ871599	0.0075	*Fungi;Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon;Trichosporon*
OTU1384179	0.0076	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU143846	0.0083	*Fungi;Basidiomycota;Agaricomycetes;Polyporales;Polyporaceae;Polyporus;Polyporus*
OTU894276	0.0083	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedbasa*l
OTU1843374	0.0083	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU1886760	0.0083	*Fungi;Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon;Trichospor*on
OTU931918;	0.0090	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*
OTU1881654	0.0096	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*

Taxonomic Differences across distances Example of the Mopane Region

The variation of fungal phyla across space was investigated within the Mopane region. The analysis only investigated samples from the dry season as the samples spread over 900 km within the ZMW. No significant variation could be observed between the samples obtained in the North and South of Zambia (R_{ANOSIM}=-0.008, p=0.499). The soil fungal diversity appeared uniform across space. Figure 3.8 shows the fungal community structure in the North and South of Zambia.

The important taxonomic features which differed across space in the ZMW using ANOVA (Table 3.8) included principally Ascomycota notably *Trichocomaceae*, a family which contains both

Aspergillus and *Penicillium*. No Basidiomycota could be identified further than the Phylum level and these results were not significant when accounting for multiple testing.

Table 3.8: Taxa Variations across space in the Mopane Region (Dry Season)

OTU	p <0.01	Taxnonomy
OTU499419	0.0012	*Fungi;unidentified;unidentified;unidentified;unidentified*
EF669715	0.0017	*Fungi;Ascomycota;Eurotiomycetes;Eurotiales;Trichocomaceae*
OTU1058	0.0020	*Fungi;Chytridiomycota;Chytridiomycetes;Rhizophlyctidales;Rhizophlyctidaceae*
OTU369	0.0024	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
OTU762334	0.0033	*Fungi;Ascomycota;Sordariomycetes;unidentified;unidentified*
OTU961914	0.0033	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU973447	0.0033	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
OTU401840	0.0036	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU1509	0.0047	*Fungi;unidentified;unidentified;unidentified;unidentified*
HM047194	0.0049	*Fungi;Ascomycota;Dothideomycetes;Pleosporales;Pleosporaceae*
OTU106034	0.0050	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU153257	0.0056	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU534914	0.0056	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU31212	0.0061	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU305196	0.0069	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
AY484681	0.0072	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU322107	0.0072	*Fungi;Ascomycota;Sordariomycetes;Hypocreales;Niessliaceae*
OTU389878	0.0072	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU523862	0.0072	*Fungi;Chytridiomycota;Chytridiomycetes;Rhizophlyctidales;Rhizophlyctidaceae*
OTU641395	0.0072	*Fungi;Basidiomycota;unidentified;unidentified;unidentified*
OTU724706	0.0072	*Fungi;Ascomycota;Sordariomycetes;Hypocreales;Nectriaceae*
OTU878496	0.0072	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU724421	0.0075	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU879567	0.0078	*Fungi;Basidiomycota;unidentified;unidentified;unidentified*
AJ557074	0.0079	*Fungi;Ascomycota;Sordariomycetes;Hypocreales;Clavicipitaceae*
OTU307	0.0081	*Fungi;unidentified;unidentified;unidentified;unidentified*
GU055562	0.0089	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
JQ411328	0.0089	*Fungi;Ascomycota;Sordariomycetes;Xylariales;Diatrypaceae*
OTU796265	0.0089	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
EU490166	0.0090	*Fungi;Ascomycota;unidentified;unidentified;unidentified*
OTU86509	0.0091	*Fungi;unidentified;unidentified;unidentified;unidentified*
OTU495308	0.0092	*Fungi;unidentified;unidentified;unidentified;unidentified*
FN547623	0.0097	*Fungi;Glomeromycota;Glomeromycetes;Glomerales;Claroideoglomeraceae*

Soil Fungal Communities associated with *C. neoformans* and *C. gattii*

The presence of the *Cryptococcus* genus using the metabarcoding *ITS2* approach varied between each region (Table 3.9). The abundance of this genus was slightly higher during the rainy season in the ZMW ($RA_{RAINY} = 2.26.10^{-4}$, $RA_{DRY} = 1.80.10^{-4}$). Moreover, the presence of *Cryptococcus* was significantly higher in the ZMW ($RA_{ZMW} = 3.3.10^{-4}$) compared to the CMW ($RA_{CMW} = 7.11.10^{-5}$). Current research in fungal metabarcoding affirms that the *ITS* region

cannot reliably distinguish beyond the genus level. However, the *ITS2* approach identified three *Cryptococcus flavus*, three *C. laurentii*, two *C. magnus*, two *C. nemorosus*, two *C. flavenscens*, one *C. rajasthanensis*, one *C. arrabidensis* and one *C. vishniacii*. But, no pathogenic species of the *Cryptococcus*-species complex could be identified.

In the previous chapter, 11 isolates from the *Cryptococcus*-species complex were detected in soil samples based on culture-based methods (Chapter 2). The study investigated whether a sample which contained either *C. neoformans* or *C. gattii* harbored a particular fungal community. The microbial compositions of the 11 positive samples were compared to non-positive site (n=50) in order to identify potential taxa associated with either *Cg* of *Cng*. However, no significant associations could be detected (Table 3.10)

Table 3.9: Relative *Cryptococcus* Genus Abundance

	Relative Abundance		
Ecoregions	Season	Rainy	Dry
ZMW	3.40^{-4}	3.61^{-4}	3.22^{-4}
CMW	7.11^{-5}	7.39^{-5}	6.78^{-5}
SMW	2.21^{-4}	2.42^{-4}	1.50^{-4}
Total	2.11^{-4}	2.26^{-4}	1.80^{-4}

Table 3.10: ANOSIM and ADONIS in *Cryptococcus*-species Complex Positive samples

		Bray-Curtis			
		ANOSIM		ADONIS	
Group	Factor	R	*p*	R	*p*
Ecoregions	*Cng* or *Cg*	0.0633	0.263	0.0320	0.692
CMW	*Cg*	-0.2944	0.819	0.0354	0.742
ZMW	*Cng*	0.1535	0.128	0.0517	0.339
SMW	*Cng* or *Cg*	0.0185	0.410	0.0931	0.248

3.3.4 Fungal Diversity using Sanger Sequencing *ITS1* Approach

Sequencing of the *ITS1* region of 264 individual yeast colonies using Sanger-sequencing allowed to identify 28 different genera and 61 species when comparing sequences to the NCBIs GenBank. Over 45% of yeast isolated from each ecoregions and sequenced belong to the *Cryptococcus* genus (Figure 3.10) revealing that culture-based methods are a reliable assessment of *Cryptococcus* diversity. The ANOSIM results revealed non-significant associations between each region and

microbial diversity (Table 3.11); the R value approached zero when performing the test at the genus level (R_{ANOSIM}= 0.0113, p <0.004) and at the species level within the *Cryptococcus* genus (R_{ANOSIM}=0.0224, p <0.047).

At the species level, *C. flavescens*, *C terrestris* and *C. humicolus* were the most commonly isolated species in each ecoregion. The Sanger-sequencing method was able to identify members of the *Cryptococcus*-species complex in each ecoregion (3.11). Within the CMW samples, *C. gattii* represented 22% of the sequences. In the ZMW and the SMW, 14% of the NCBI results matched *C. neoformans*. It is worth noting that this approach was biased toward characteristic dark brown colonies resembling the *Cryptococcus*-species complex and that the proportion found in the study do not reflect the true proportions of yeast diversity in the samples.

Table 3.11: ANOSIM and ADONIS in *Cryptococcus*-species Complex Positive samples

		Bray-Curtis			
		ANOSIM		ADONIS	
Group	Factor	R	P	R	P
Ecoregions	*Cng* or *Cg*	0.0633	0.263	0.0320	0.692
Miombo	*Cg*	-0.2944	0.819	0.0354	0.742
Mopane	*Cng*	0.1535	0.128	0.0517	0.339
Southern Miombo	*Cng* or *Cg*	0.0185	0.410	0.0931	0.248

Table 3.12: ANOSIM of microbial diversity using the *ITS1* approach

		ANOSIM			
		Genus Level		*Cryptococcus* species	
Group	Factor	R	P	R	P
Ecoregions	Ecoregions	0.0113	0.004	0.0224	0.047
CMW	Ecoregions	0.0003	0.413	-0.0087	0.756
ZMW	Ecoregions	0.0076	0.124	0.0141	0.149
SMW	Ecoregions	-0.0010	0.532	-0.0234	0.834

The Role of Seasonality on Fungal Communities

As described in the previous chapter, the prevalence of both *C. gattii* and *C. neoformans* is influenced by seasonality. The recovery of the *Cryptococcus*-species complex obtained in the study increased from 1.46% in the rainy season to 2.38% in the dry season. Here, the study looks for potential seasonal variations in the fungal community structure within each ecoregion. Season played a significant role in shaping the microbial diversity in the ZMW

(R_{ANOSIM}=0.3610, $p = 0.019$). The patterns of fungal diversity in the CMW seemed to remain stable across seasons (R_{ANOSIM}=0.1062, p=0.189). (Table 3.13).

Table 3.13: Bray-Curtis Diversity across Season

		Bray-Curtis			
		ANOSIM		ADONIS	
Group	Factor	R	P	R	P
ZMW	Season	0.3610	0.019	0.0893	0.068
CMW	Season	0.1062	0.189	0.0668	0.016

Mopane Region

The number of sequences was significantly higher in the dry season, varying from 145 seq./sample up to 240 seq./sample (Table 3.14). This finding was principally due to the predominance of Ascomycota in the dry season. Accounting for 45.33% of the OTUs in the dry season, the proportion of Ascomycota double in the dry season implicating that this phylum is highly favoured in dry conditions. With respect to the Basidiomycota, the phylum accounted for 10.50% of all OTUs in the rainy season and increased to 13.35% in the dry season.

When investigating the possible presence of key taxa which would be characteristic of the change in microbial diversity across seasons, *Aspergillus peyronelii* was among the sequences identified using ANOVA. Among the 24 most significant results, half of them could not be assigned to a phylum stressing the current lack of knowledge regarding the fungal diversity. However, no taxa remained significant when accounting for multiple testing.

Table 3.14: Microbial patterns within ZMW

Season	Rainy Season		Dry Season	
Phylum	Seq./Sample	Abundance (%)	Seq./Sample	Abundance (%)
Ascomycota	70.22	26.36%	143.37	45.33%
Basidiomycota	49.89	10.50%	67.95	13.35%
Chytridiomycota	9.00	5.96%	10.26	4.06%
Glomeromycota	7.89	0.33%	10.21	0.37%
Zygomycota	5.44	15.86%	5.44	2.52%
Other	-	3.57%	-	0.63%
Unidentified	-	37.42%	-	33.73%
Total	145.33	100.00%	240.26	100.00%

Table 3.15: Feature Importance Taxa within the Mopane Region

OTU	p <0.01	Taxnonomy
OTU52904	0.0007	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus;*
OTU499419	0.0010	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
EF669715	0.0025	*Fungi;Ascomycota;Eurotiomycetes;Eurotiales;Trichocomaceae;Aspergillus;Aspergillus;peyronelii*
AY740163	0.0030	*Fungi;Basidiomycota;Microbotryomycetes;Microbotryales;Microbotryaceae;Sphacelotheca;Sporisorium*
OTU722	0.0038	*Fungi;Ascomycota;Dothideomycetes;Botryosphaeriales;Botryosphaeriaceae;Aplosporella;Aplosporella*
OTU782469	0.0039	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU949683	0.0044	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU193	0.0048	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU859256	0.0053	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU547	0.0060	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedsoil;fungus*
OTU106034	0.0068	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedsoil;fungus*
HM122861	0.0069	*Fungi;Ascomycota;Sordariomycetes;Diaporthales;Valsaceae;Cytospora;Cytospora;cf*
OTU6334	0.0071	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU234407	0.0071	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU25369	0.0071	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedbasal;lineage*
OTU457195	0.0071	*Fungi;Ascomycota;Sordariomycetes;Hypocreales;Nectriaceae;Calonectria;Calonectria;hawksworthii*
OTU762334	0.0071	*Fungi;Ascomycota;Sordariomycetes;unidentified;unidentified;unidentified;Sordariomycetes;s*p
OTU973447	0.0071	*Fungi;Ascomycota;unidentified;unidentified;unidentified;unidentified;unculturedAscomycota*;
OTU535571	0.0080	*Fungi;Chytridiomycota;Chytridiomycetes;Rhizophydiales;unidentified;unidentified;Rhizophydiales;sp*
OTU335	0.0086	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus;*
OTU38062	0.0086	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU973148	0.0086	*Fungi;unidentified;unidentified;unidentified;unidentified;unidentified;unculturedfungus*;
OTU577219	0.0097	*Fungi;Basidiomycota;Exobasidiomycetes;Tilletiales;Tilletiaceae;Tilletia;Tilletia;olida*

Miombo Regions

In the CMW, R_{ANOSIM} coefficient was low (R_{ANOSIM}=0.1062, p <0.189) indicating that the species present were stable across season. The UPGMA tree on Figure 3.13a shows that samples from both seasons clustered together. However, the number of sequences per sampled varied between the two seasons. Similarly to the ZMW region, the number of sequences increased in the dry season (219 seq./sample) compared to the rainy season (135 seq./sample). The proportion of Ascomycota also increased in the dry season and the proportion of Basidiomycota doubled from 8.59% to 16.43% based on the relative OTUs abundance. The number of unidentified fungal sequences was significantly higher in the rainy season reaching 47.47% of all reads (Table 3.16).

Southern Miombo Woodlands

In the SMW, the number of sequences per sample was stable across seasons. However, the abundance of Ascomycota more than doubled in the dry season reaching 56.55%. The pattern was similar to that observed for the two other ecoregions. The Basidiomycota increased during the dry season from 5.79% to 16.31%. Interestingly, the amount of Zygomycota dropped from

Table 3.16: Microbial patterns within CMW

Season	Rainy Season		Dry Season	
Phylum	Seq./Sample	Abundance (%)	Seq./Sample	Abundance (%)
Ascomycota	74.39	29.41%	131.75	39.22%
Basidiomycota	42.83	8.59%	63.38	16.43%
Chytridiomycota	8.06	2.15%	9.88	5.33%
Glomeromycota	3.94	2.88%	8.88	0.63%
Zygomycota	6.33	7.33%	5.63	9.89%
Other	-	2.18%	-	0.03%
Unidentified	-	47.47%	-	28.44%
Total	135.56	100.00	219.50	100.00

Table 3.17: Feature Importance Taxa within the Miombo Region

OTU	p <0.01	Taxnonomy
OTU64754	0.0015	*Ascomycota;Eurotiomycetes;Chaetothyriales;Herpotrichiellaceae;Cladophialophora*
OTU1089	0.0022	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU491	0.0027	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU560021	0.0035	*Ascomycota;Lecanoromycetes;Teloschistales;Physciaceae;Buellia*
OTU409621	0.0038	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU473421	0.0040	*Ascomycota;Eurotiomycetes;Chaetothyriales;unidentified;unidentified*
OTU145428	0.0046	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU105815	0.0050	*Ascomycota;unidentified;unidentified;unidentified;unidentified*
OTU478908	0.0071	*unidentified;unidentified;unidentified;unidentified;unidentifid*
OTU460069	0.0076	*Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon*
OTU722	0.0087	*Ascomycota;Dothideomycetes;Pleosporales;Incerta*
HM047194	0.0088	*Ascomycota;Dothideomycetes;Pleosporales;Pleosporaceae;Epicoccu*m
OTU223007	0.0088	*unidentified;unidentified;unidentified;unidentified;unidentified*
OTU572118	0.0088	*Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon*
OTU38003	0.0092	*Ascomycota;Dothideomycetes;Pleosporales;Pleosporaceae;Alternaria*
DQ87159	0.0095	*Basidiomycota;Tremellomycetes;Tremellales;Trichosporonaceae;Trichosporon*

19.99% in the rainy season to 0.57% in the dry season (Table 3.18).

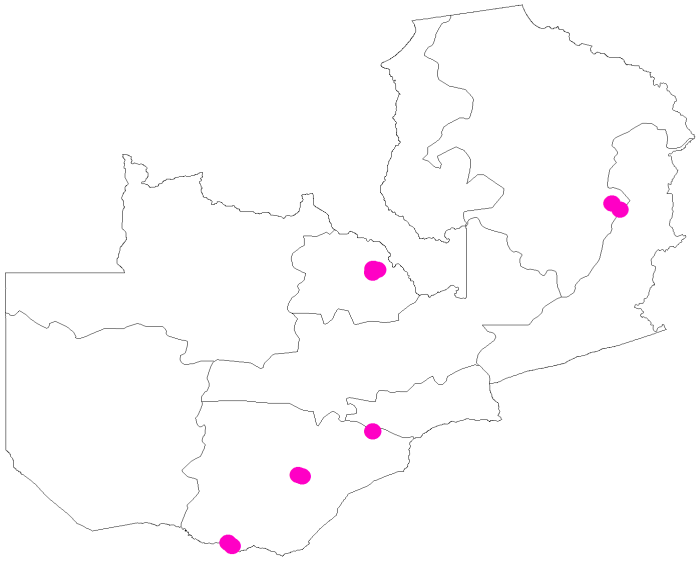


Figure 3.2: Soil Sampling sites used for the Metabarcoding Approach (n = 74)

Primers Name	Illumina Overhangs	Linker Seq.	Primer Seq.
ITS3_KYO2	5'-**TCGTCGGCAGCGTCAGATGTGTATAAGAGAC**AG		GATGAAGAACGYAGYRAA-3'
ITS4_KYO3	5'-**GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC**AG		CTBTTVCCKCTTCACTCG-3'

Figure 3.3: *ITS* Primers with Illumina adapter overhang sequences and linker sequence.

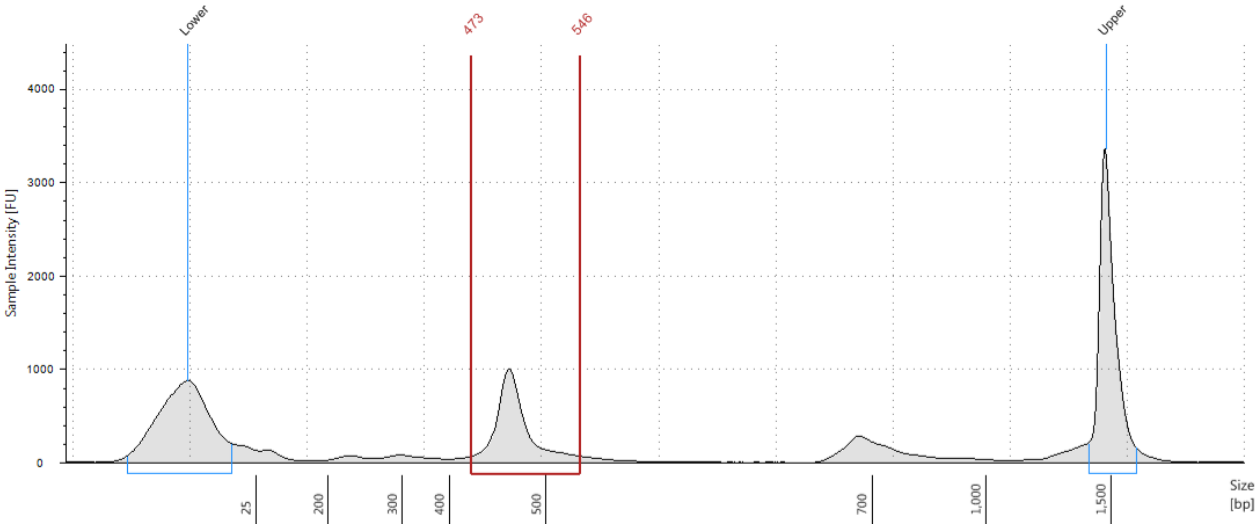
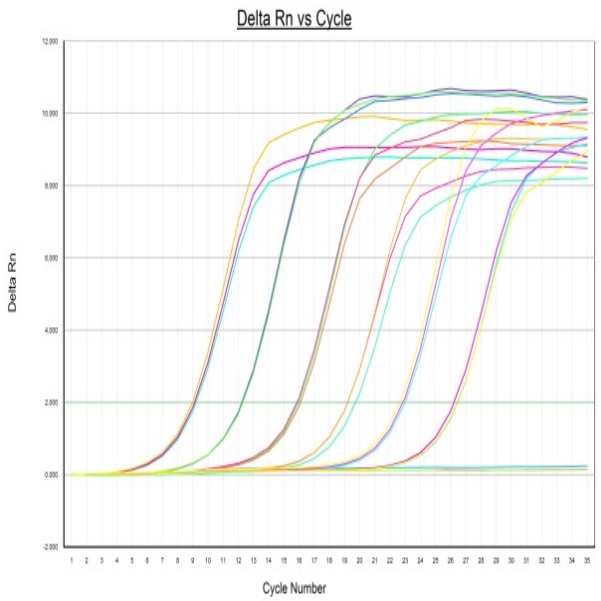
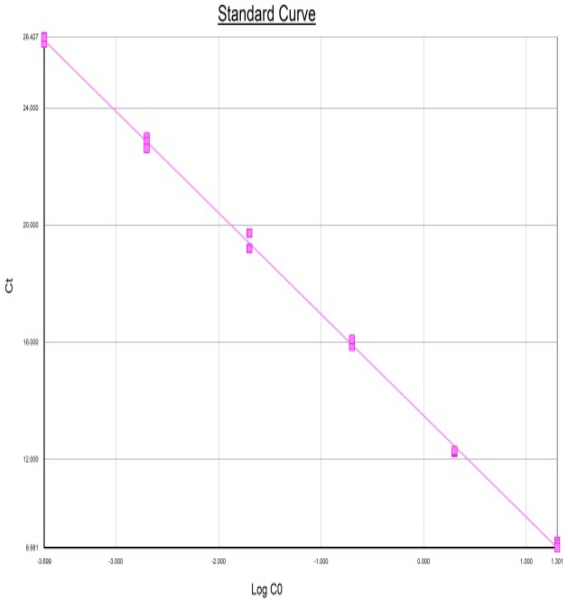


Figure 3.4: TapeStation D1000 electropherogram Final Cleaned Fragment with an average band size of 509 bp, with region plot defined (red bars), and upper and lower reference peaks visible

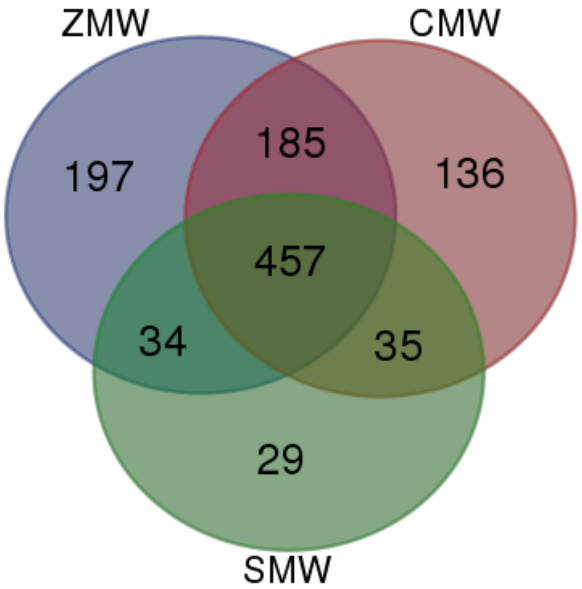


(a) Amplification plots for Kapa Standards.

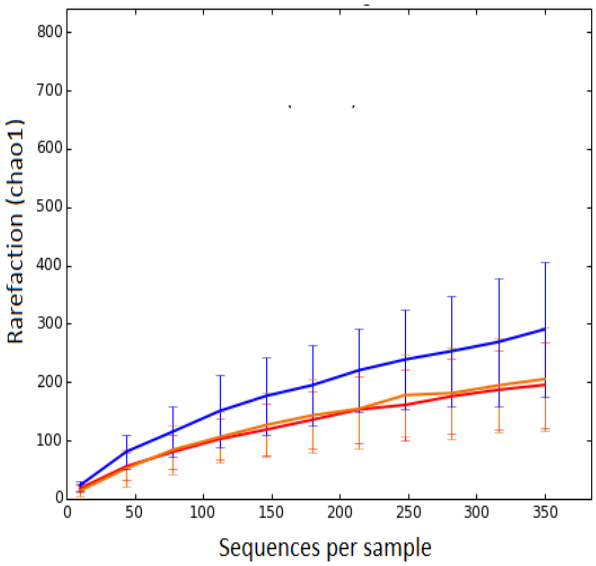


(b) Standard Curve. R=0.99831

Figure 3.5: Analysis of Kapa Real-time PCR

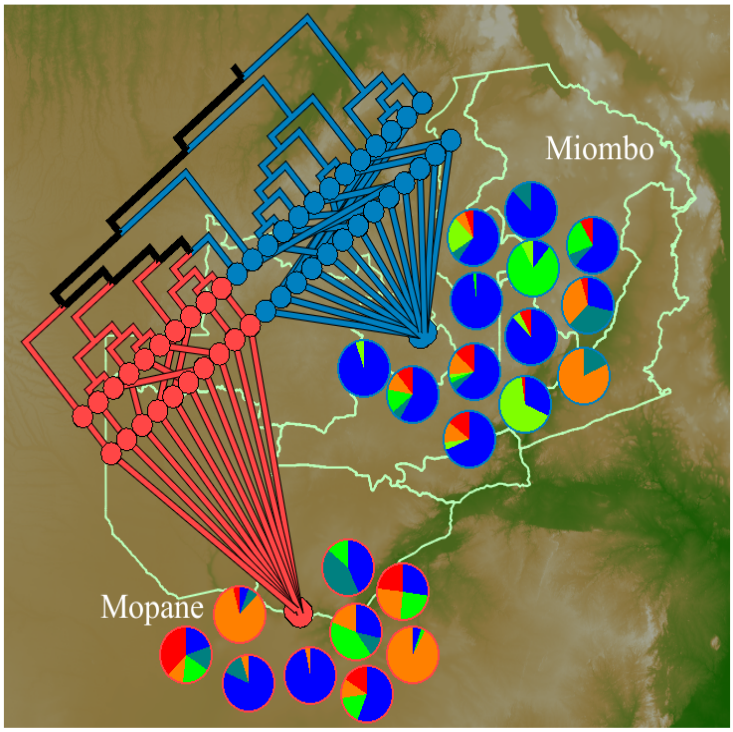


(a) Venn Diagram

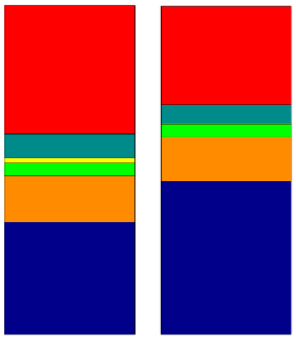


(b) Rarefaction Curve

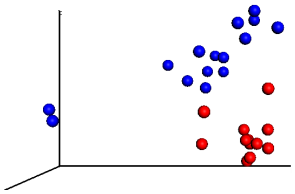
Figure 3.6: Variation between the different Ecoregions (a) OTUs variation between the different Zambian ecoregions,; (b) Chao1 distribution for the three ecoregions (ZMW - blue, CMW - Red, SMW - orange).



(a) Variation between ZMW and CMW

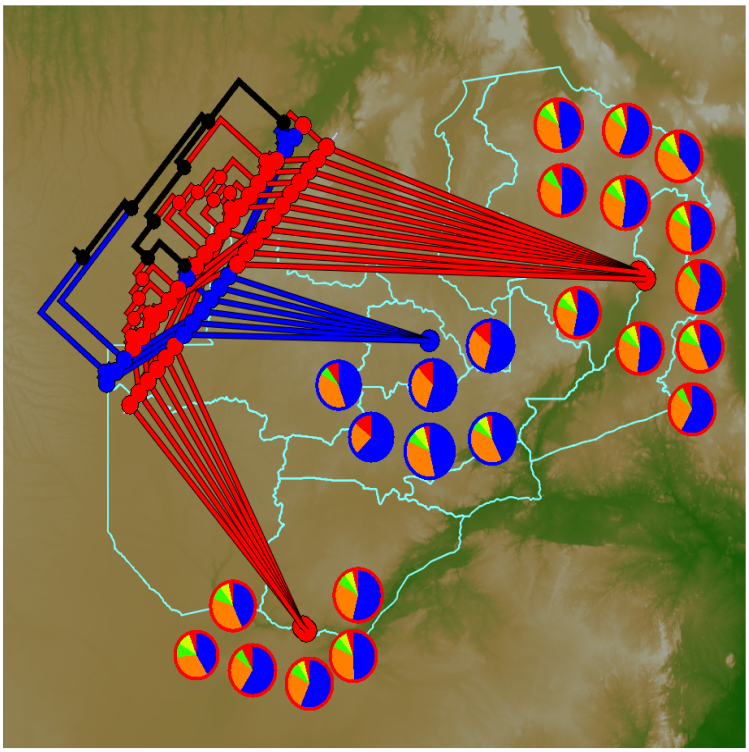


(b) OTUs Profile

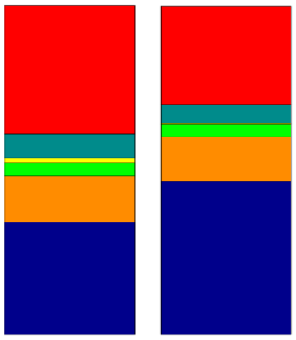


(c) PCoA

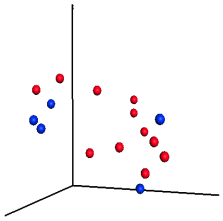
Figure 3.7: Fungal Community Structure Variations within Mopane and Miombo Region in the Rainy Season - (a) UPGMA tree of samples from the CMW (blue) and ZMW (red) showing a clear clustering; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profile for CMW (left) and ZMW (right); (c) PCoA of samples coloured by ecoregions: CMW (blue), ZMW (red).



(a) Variation between ZMW and CMW

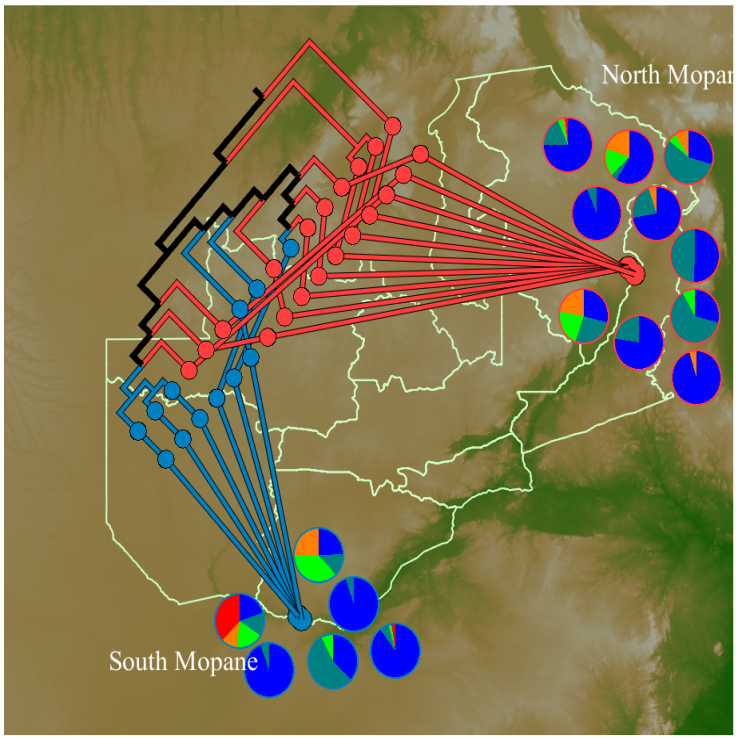


(b) OTUs Profile

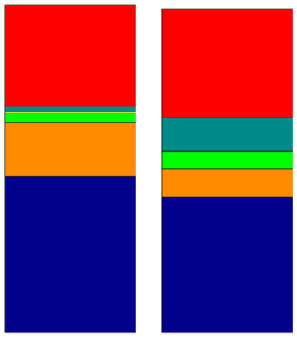


(c) PCoA

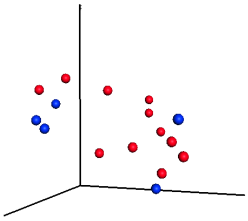
Figure 3.8: Fungal Community Structure Variations within Mopane and Miombo Region in Dry Season - (a) Clustering of samples in each of the ecoregions CMW (blue) ZMW (red); Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profile for CMW (left) and ZMW (right); (c) PCoA of samples coloured by ecoregions: CMW (blue), ZMW (red).



(a) Variation ZMW across space

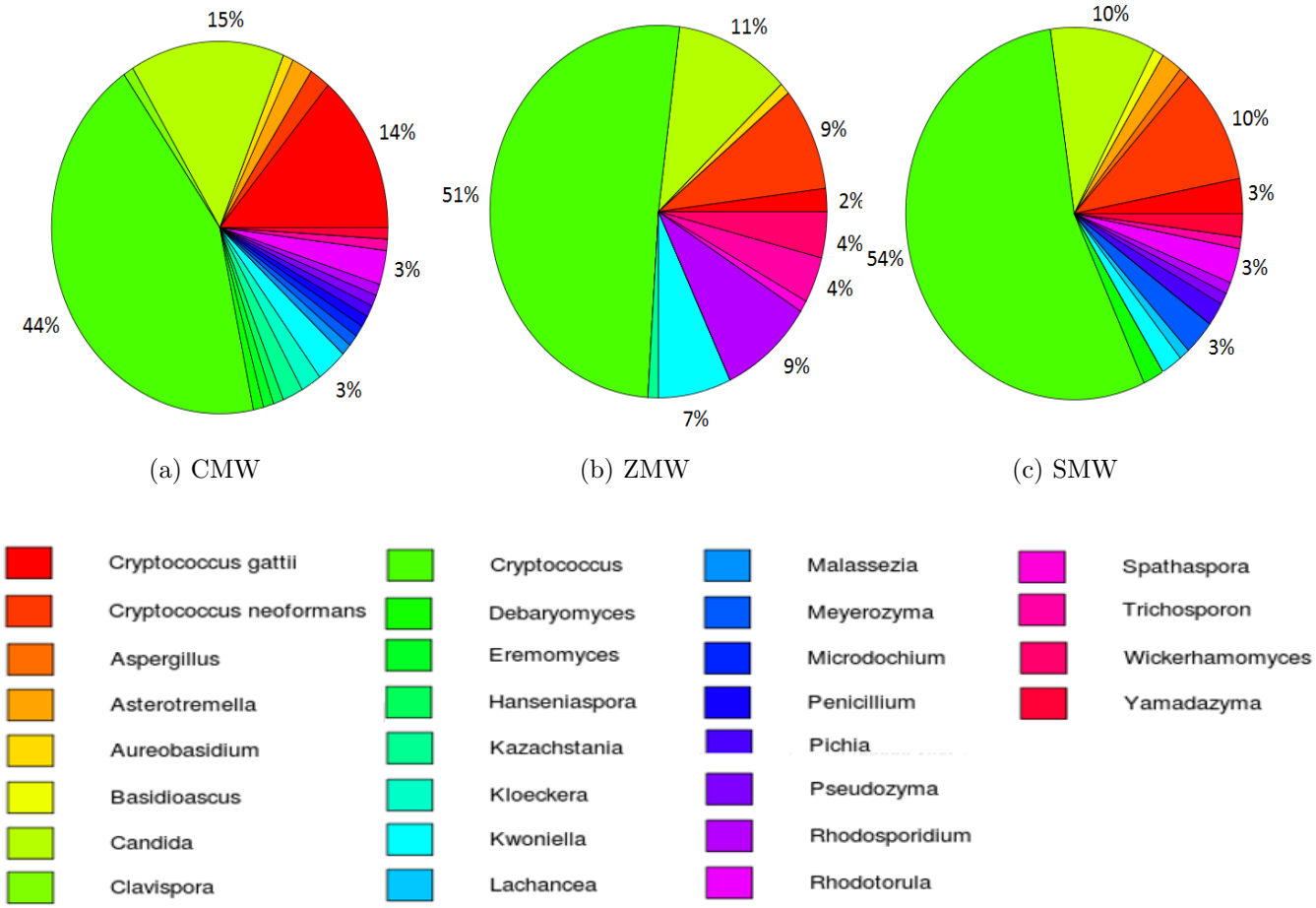


(b) OTUs Profile



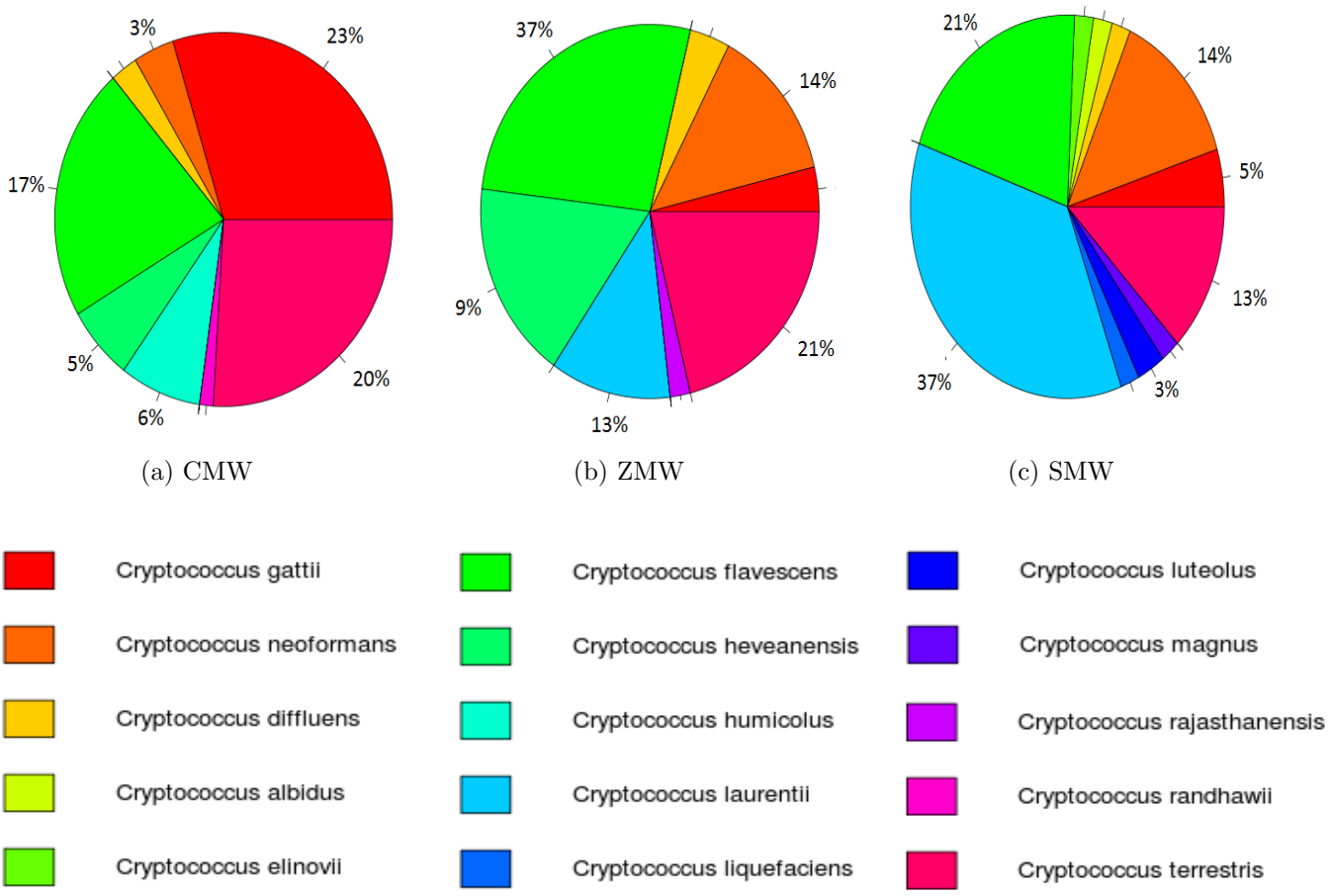
(c) PCoA

Figure 3.9: Fungal Community Structure Variations across Space within the Mopane Region (Dry Season)(a) On the UPGMA, samples appear mixed between the north (red) and the south (blue) of the ZMW; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the north (left) and south (right) show a similar profile; (c) PCoA of samples appear mixed: northern samples (red), southern samples (blue).



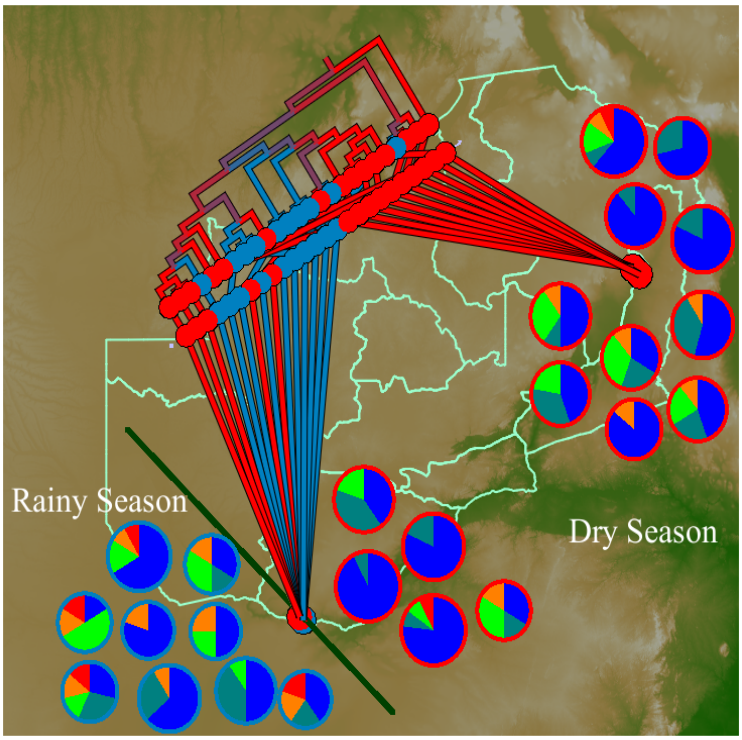
(a) CMW (b) ZMW (c) SMW

Figure 3.10: Distribution of the different Genus identified using culture-based *ITS1* sequencing method

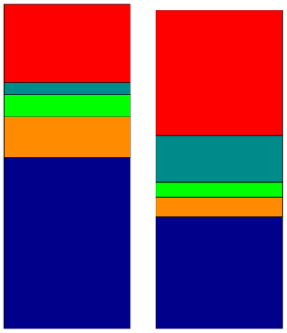


(a) CMW (b) ZMW (c) SMW

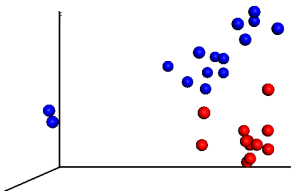
Figure 3.11: Species distribution of the within the *Cryptococcus* Genus identified using culture-based *ITS1* sequencing method



(a) Seasonal Variation in ZMW

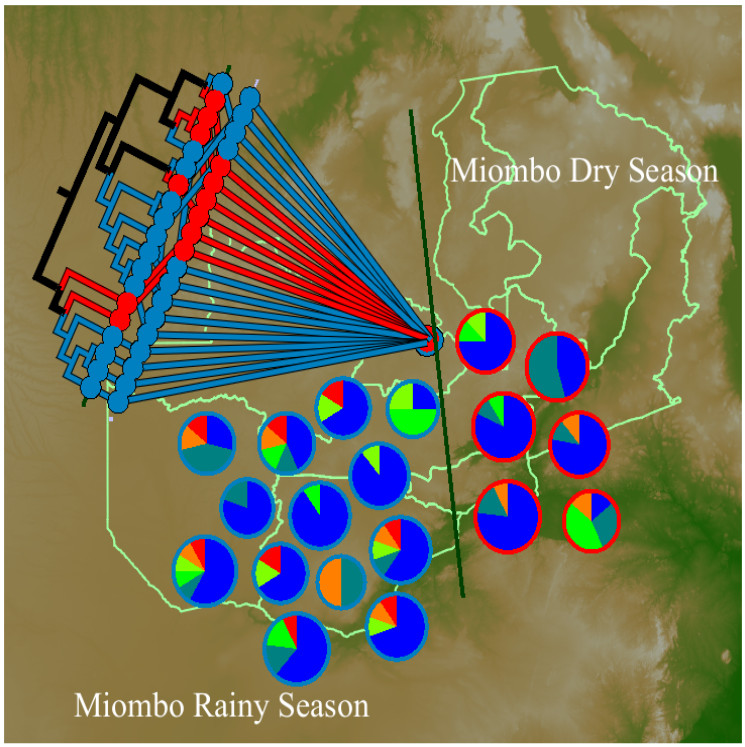


(b) Genus variation

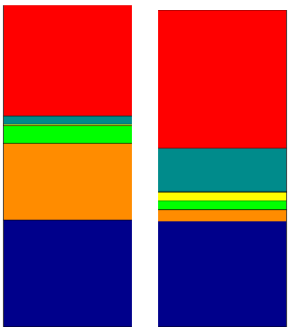


(c) test figure three

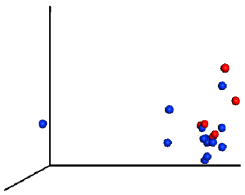
Figure 3.12: Seasonal Variation within the Mopane Region - (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue), *Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA reveals have a mixed profile.



(a) Seasonal Variation in ZMW



(b) OTUs Profile



(c) PcoA

Figure 3.13: Seasonal Variation within the Miombo Region (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA of samples appear mixed.

Table 3.18: Microbial patterns within SMW

Season	Rainy Season		Dry Season	
Phylum	Seq./Sample	Abundance (%)	Seq./Sample	Abundance (%)
Ascomycota	105.60	26.35%	101.00	56.55%
Basidiomycota	66.20	5.79%	50.00	16.31%
Chytridiomycota	10.20	1.60%	5.67	4.40%
Glomeromycota	8.40	1.29%	5.00	0.13%
Zygomycota	6.60	19.99%	4.67	0.57%
Other	-	1.66%	-	0.55%
Unidentified	-	43.32%	-	22.49%
Total	197.00	100.00	166.33	100.00

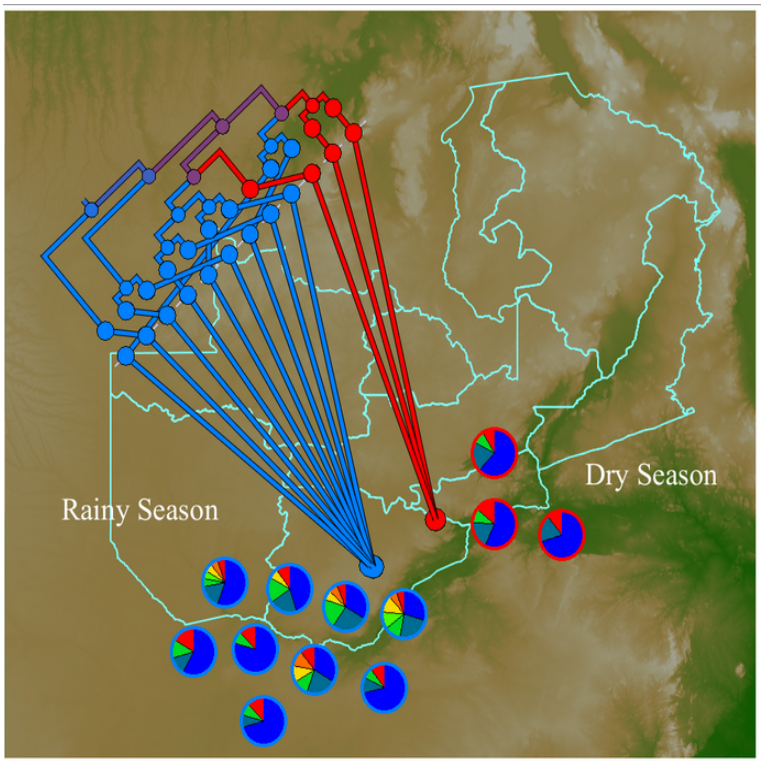
3.3.5 Impacts of Environmental Conditions on Soil Community Structures

Permutational MANOVA and best variables rank correlation test (BEST) Analysis

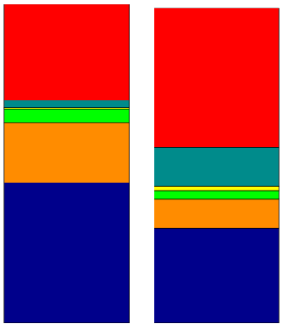
To determine the role of environmental factors on the β-diversity of fungal communities, permutational MANOVA and a best variables rank correlation test (BEST) analyses were performed. The environmental factors investigated were the Bioclim layers (Busby, 1991) available on the WorldClim website (http://www.worldclim.org/bioclim). Soil properties information (Hengl et al., 2015) were extracted but spatial predictions were uniform across Zambia and could not be integrated into the analysis.

Between ecoregions, all environmental metrics displayed significant degrees of correlation with respect to fungal diversity, except for soil properties. The MANOVA tests identified Precipitation of Wettest Quarter (bio16) (R^2=2.6363, p<0.001) to be most significant variable to influence fungal communities (Table 3.19). Then, followed Precipitation of Warmest Quarter (bio18) (R^2=1.5652, p <0.001), Precipitation of Driest Month (Bio14) (R^2=1.4642, p<0.001), Temperature Annual Range (Bio7) (R^2=1.4642, p <0.001) and Mean Diurnal Range (Bio2) (R^2=1.4629, p<0.001).

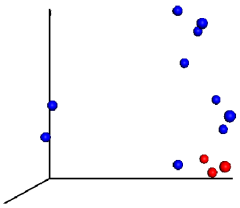
Climatic variables seem to have a more significant influence on β-diversity than geographical variables. In support of this, the MANOVA R^2 coefficient of elevation, latitude and longitude was the lowest (R^2<1.12, p<0.001) but all features appeared significant. BEST analysis de-



(a) Seasonal Variation in ZMW



(b) OTUs Profile



(c) PcoA

Figure 3.14: Seasonal Variation within the SMW - (a) On the UPGMA, samples appear mixed between the dry (red) and rainy (blue) season; Pie-charts represent the proportion of each phylum: *Ascomycota* (blue),*Basidiomycota* (orange), *Chytridiomycota* (green), *Glomeromycota* (yellow), *Zygomycota* (grey), other (red) (b) Genus profiles of the dry (left) and rainy (right); (c) PCoA reveals have a mixed profile.

termines which environmental feature explains the greatest degree of fungal diversity. This method ranked Precipitation of Wettest Month (bio13) as the highest correlation between the diversity distance-matrix. Then Mean Temperature of Warmest Quarter (bio10) and elevation were the second and third most significant factors (Table 3.20).

Within the ZMW, only 'Min Temperature of Coldest Month'(bio6) (R^2=1.1109, p<0.01) and 'Mean Temperature of Wettest Quarter'(bio8) (R^2=1.1659, p=0.042) were significant with MANOVA tests. The BEST analyses identified 'Temperature of Wettest Quarter'(bio8) as most important feature confirming the MANOVA tests. Isothermality (bio3) and 'Mean Temperature of Warmest Quarter'(bio10) were also found to be important contributing factors.

In the CMW, 'Precipitation Seasonality'(bio15) (R^2=1.4653, p=0.003) was the most significant variable in the MANOVA tests. Other features such as 'Annual Mean Temperature'(bio1), 'Precipitation of Wettest Quarter'(bio16), 'Temperature Seasonality'(bio4) and 'Precipitation of Wettest Month'(bio13) were also significant. BEST analysis confirmed the role of 'Precipitation Seasonality'(bio15) as the main feature to explain differences in microbial diversity but the analysis emphasised the role of elevation and latitude on fungal β-diversity.

Environmental Niche Modelling - Maxent

Modelling the environmental niches of the two sister species in Zambia will allow me to determine the potential realised niche of *Cng* and *Cg*. The Maxent output shown in Figure 3.15 revealed that the *C. neoformans* ecological niche is largely associated with the *C. mopane* tree belt (ZMW and SMW), whereas *C. gattii* realised niche seemed to be within the wet, higher altitude, Miombo woodlands (CMW). The environmental variables which gave the highest relative contribution to the model were for *C. neoformans*: Isothermality (bio3) accounting for 71.2% and Precipitation Seasonality (bio15) accounting for 22.0%. Regarding C. gattii, the relevant climatic variables were Isothermality (bio3) 26.5%, Precipitation of Wettest Quarter (bio16) 18.4%, Precipitation of Warmest Quarter (bio18) 16.8%, and Mean Temperature of Driest Quarter (bio9) 7.9%. The full listing of environmental variables contributing to the model is in the Table S4.

Table 3.19: Permutational MANOVA of Environmental Effects on Microbial Diversity Patterns Between Regions

	Bray Curtis					
	Ecoregions		Mopane Woodlands		Miombo Woodlands	
	R^2	P	R^2	P	R^2	P
Longitude	1.1185	0.001	1.0811	0.114	1.0320	0.209
Latitude	1.1185	0.003	1.0811	0.112	1.0320	0.187
Elevation	1.1171	0.002	1.0734	0.140	1.0320	0.186
EVI	1.1871	0.010	1.0811	0.110	1.0552	0.180
NDVI	1.165	0.010	0.144	0.080	NA	NA
Bio1	1.3712	0.001	1.1109	0.069	1.1455	0.048
Bio2	1.4629	0.003	1.1109	0.093	1.1119	0.116
Bio3	1.2988	0.001	1.1740	0.128	NA	NA
Bio4	1.3443	0.001	1.1109	0.084	1.1455	0.046
Bio5	1.3996	0.001	1.1109	0.088	0.9954	0.484
Bio6	1.4009	0.001	1.1109	0.010	1.2599	0.008
Bio7	1.4642	0.001	1.1109	0.068	1.2338	0.016
Bio8	1.3217	0.001	1.1659	0.042	0.9954	0.434
Bio9	1.3147	0.001	1.1109	0.086	0.9954	0.477
Bio10	1.3217	0.001	1.1323	0.061	0.9954	0.489
Bio11	1.2988	0.001	1.1109	0.079	0.9954	0.476
Bio12	1.3704	0.001	1.1109	0.083	1.1455	0.057
Bio13	NA	NA	1.1323	0.068	1.1455	0.050
Bio14	1.4642	0.001	1.1323	0.066	NA	NA
Bio15	1.3417	0.001	1.1323	0.062	1.4653	0.003
Bio16	2.6363	0.001	1.1109	0.098	1.1455	0.037
Bio17	1.3579	0.001	1.2019	0.080	NA	NA
Bio18	1.5652	0.001	1.1109	0.089	1.1455	0.066
Bio19	NA	NA	1.2019	0.081	1.0421	0.339

Table 3.20: BEST Analysis within Ecoregions

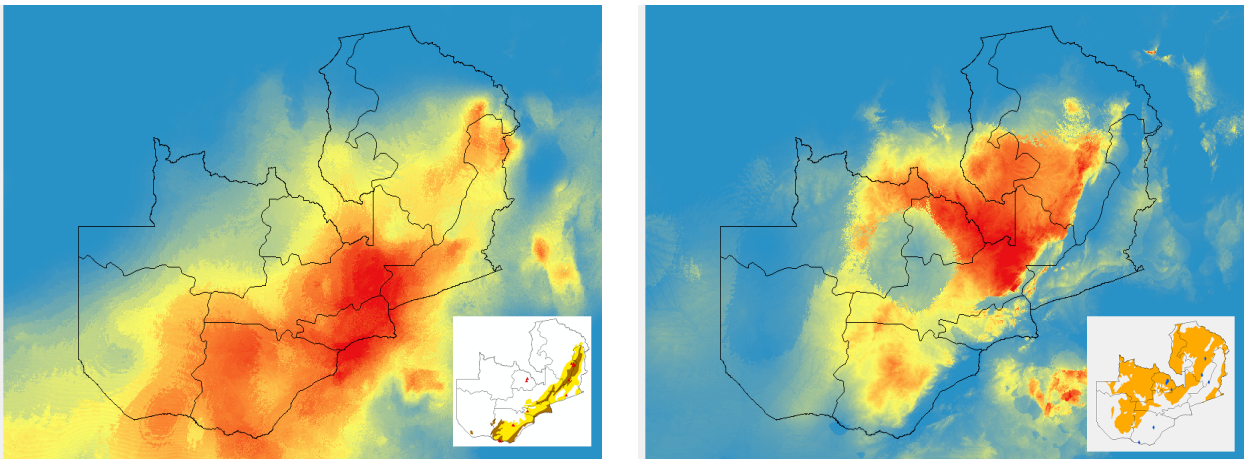
Bray-Curtis Dissimilarity		
Size	p	Ecoregions
1	0.303	bio13
2	0.335	bio10, bio13
3	0.338	Elevation, bio12, bio13
4	0.335	bio1, bio8, bio12, bio13
5	0.331	bio1, bio8, bio10, bio12, bio13
6	0.327	Elevation, bio5, bio8, bio10, bio12, bio13
7	0.324	Elevation, bio1, bio5, bio8, bio10, bio12, bio13
8	0.292	Elevation, bio1, bio5, bio8, bio10, bio11, bio12, bio13
9	0.281	Elevation, bio1, bio5, bio8, bio9, bio10, bio11, bio12, bio13
10	0.266	Elevation, Latitude, bio1, bio3, bio5, bio8, bio10, bio11, bio12, bio13
11	0.256	Elevation, Latitude, bio1, bio3, bio4, bio5, bio8, bio10, bio11, bio12, bio13
12	0.240	Elevation, Latitude, Longitude, bio1, bio3, bio5, bio6, bio8, bio9, bio10, bio11, bio13
13	0.236	Elevation, Latitude, bio1, bio2, bio3, bio5, bio6, bio8, bio9, bio10, bio11, bio12, bio13
14	0.229	Elevation, Latitude, Longitude, bio1, bio2, bio3, bio5, bio6, bio8, bio9, bio10, bio11, bio12, bio13
15	0.197	Elevation, Latitude, bio1, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio11, bio12, bio13
16	0.122	Elevation, Latitude, Longitude, bio1, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio11, bio12, bio13

Table 3.21: BEST Analysis within Mopane Woodlands

Bray-Curtis Dissimilarity		
Size	p	Mopane Woodlands
1	0.234	bio8
2	0.222	bio3, bio8
3	0.222	bio3, bio8, bio10
4	0.218	bio2, bio3, bio4, bio8
5	0.220	bio5, bio7, bio8, bio16, bio18
6	0.221	Latitude, Elevation, bio4, bio5, bio8, bio18
7	0.222	Latitude, Elevation, bio2, bio4, bio5, bio8, bio18
8	0.222	Latitude, Elevation, bio2, bio3, bio4, bio5, bio8, bio18
9	0.222	Latitude, Elevation, bio2, bio4, bio5, bio7, bio10, bio16, bio18
10	0.222	Latitude, Elevation, bio2, bio4, bio5, bio7, bio8, bio10, bio16, bio18
11	0.222	Latitude, Elevation, bio2, bio4, bio5, bio6, bio7, bio8, bio10, bio16, bio18
12	0.221	Latitude, Longitude, Elevation, bio2, bio3, bio4, bio5, bio6, bio8, bio10, bio16, bio18
13	0.221	Latitude, Longitude, Elevation, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio10, bio16, bio18
14	0.217	Latitude, Longitude, Elevation, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio16, bio18
15	0.205	Latitude, Longitude, Elevation, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio15, bio16, bio18
16	0.195	Latitude, Longitude, Elevation, bio1, bio2, bio3, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio15, bio16, bio18

Table 3.22: BEST Analysis within Miombo Woodlands

Bray-Curtis Dissimilarity		
Size	p	Miombo Woodlands
1	0.230	bio15
2	0.225	Latitude, bio16
3	0.234	Latitude, Elevation, bio15
4	0.234	Latitude, Elevation, bio15, bio16
5	0.242	Latitude, Elevation, bio15, bio16, bio18
6	0.238	Latitude, Elevation, bio6, bio15, bio16, bio18
7	0.228	Latitude, Elevation, bio1, bio6, bio15, bio16, bio18
8	0.205	Latitude, Elevation, bio1, bio6, bio7, bio15, bio16, bio18
9	0.188	Latitude, Elevation, bio1, bio5, bio6, bio8, bio15, bio16, bio18
10	0.174	Latitude, Elevation, bio1, bio2, bio4, bio6, bio7, bio9, bio15, bio18
11	0.171	Latitude, Longitude, Elevation, bio1, bio2, bio4, bio5, bio7, bio15, bio16, bio18
12	0.171	Latitude, Longitude, Elevation, bio1, bio2, bio5, bio6, bio7, bio8, bio15, bio16, bio18
13	0.171	Latitude, Longitude, Elevation, bio1, bio2, bio4, bio5, bio6, bio7, bio8, bio15, bio16, bio18
14	0.161	Latitude, Elevation, bio1, bio2, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio15, bio16, bio18
15	0.144	Latitude, Longitude, Elevation, bio1, bio2, bio4, bio5, bio6, bio7, bio8, bio9, bio10, bio15, bio16, bio18



(a) Modelling of *C. neoformans* distribution

(b) Modelling of *C. gattii* distribution

Figure 3.15: Maxent Species Distribution Modelling based on Bioclim Layers - Ecoregion map (bottom-right) portrays the ecoregion associated with each sister species, ZMW and SMW for *Cng* and CMW for *Cg*.

3.4 Discussion

In a globalised world, disease ranges are shifting due to changes in land-cover and climate change. It is therefore important to be able to predict the shift of pathogens from their current ranges into new regions as their ecological niches change. To realise this, it is necessary to understand the species-habitat relationships across varied spatial scales. Understanding the microbial ecology of environmental pathogens could then be used to identify locations where diseases can potentially emerge and threaten human health. The present study characterised the biotic and abiotic environments associated with the two sister species *C. neoformans* and *C.*

gattii in Zambia. The country hosts both species in distinct ecoregions and therefore provides a good model to study the ecology of the *Cryptococcus*-species complex.

3.4.1 Illumina Sequencing

In the last decade, the sequencing of genetic markers and the development metagenomics tools has revolutionised fungal ecology (Ihrmark et al., 2012). In this study, 8.44 million paired-end reads were generated on a 2 x 300-bp reads length MiSeq run and 72,959 reads were assigned to fungi. The number of OTUs (n_{OTUs}=11,475) assigned to fungi and their re-partition into the five major fungal phylum were similar to previous fungal metabarcoding studies (Bálint et al., 2014; Meiser et al., 2014).

The *internal transcribed spacer* (*ITS*) ribosomal DNA is currently accepted as the universal barcode for fungi (Schoch et al., 2012). The region presents both advantages and drawbacks. The marker is suitable for the exploration of fungal diversity as it contains two variable regions (*ITS1* and *ITS2*) (Porter and Brian Golding, 2011). The availability of software pipeline, such as QIIME (Caporaso et al., 2010), allows to perform data cleaning, clustering of sequences and taxonomic assignment at high speed and with inexpensive computer power. The UNITE database is expanding and currently contains 38,353 species based on 97% identity (Kõljalg et al., 2013). However, Blackwell et al. (2011) noted that these numbers probably account for 0.2 to 4.5% of the 0.5 to 10 million of estimated fungal species. The 97% sequence similarity used in this study to assign sequences to references has been widely used in fungal metabarcoding (Blaalid et al., 2013; Walker et al., 2008; O'Brien et al., 2005).

Overall, in this survey, 36.19% of OTUs could not be assigned to a phylum and 56.2% of OTUs could not be identified at the genus level. This finding highlights the current lack of knowledge regarding fungal diversity, especially in the scarcely explored soil ecosystems of tropical Africa. Here, increased efforts are required to describe fungal species. Only 1,312 yeast species were listed in the fifth edition of 'The yeasts, a taxonomic study'(Kreger-van Rij, 2013). Soil yeasts have a key function in the regulation of soil ecosystems and its understanding will have significant implications on agricultural practices. However, current knowledge regarding yeast ecology remains limited (Takashima et al., 2012). The presence of *Cryptococcus* varied significantly between the different ecoregions (Table 3.5). However this genus is polyphyletic

and is present in five different orders which makes interpretations on the role of this genus in ecosystem-functioning to be difficult (Gow et al., 1997). A reclassification of the genus based on phylogenetic information would be beneficial to understand the ecological role *Cryptococcus* in microbial communities.

3.4.2 The Use of *ITS* to Assess Fungal Diversity

The *ITS* region is approximately 650-bp long (White et al., 1990) and current Illumina technology only allows a maximum of 300-bp long reads. In this study, we chose to focus on *ITS2* as the literature suggests targeting this region allows a better assessment of fungal diversity (Bazzicalupo et al., 2013). It is worth noting that errors can arise during the metabarcoding protocol. Taq polymerase can introduce chimeras or heteroduplex during PCR amplification and these errors then can overestimate microbial diversity (O'Brien et al., 2005). To account for PCR errors, increasing the number of replicates helps to reduce the number of irregularity due to PCR reactions. Libraries can also be prepared using a single amplification step (Caporaso et al., 2012). Due to financial restriction, these methods could not be implemented in the study. However, KAPA HiFi DNA polymerase was used since its error rate is 100x lower than regular Taq DNA polymerase (KAPA Biosystem). Additionally, laboratory contamination can be a potential confounder in sequence-based analysis (Salter et al., 2014). To address this problem, one sample acted as a negative control and only one taxa, Trichosporon, was identified.

The rarefaction curves among the three ecoregions appeared to reach saturation, which means that the protocol succeeded in revealing most of the diversity present in the samples. However, the rarefaction curves from for several individual samples (Figures S2 S6 & S4) did not reaching saturation. Gans et al. (2005) claimed that a gram of soil could contain over 10 billion of microbial cells. The survey might not have revealed the entire fungal diversity present in Zambian soil, but the metabarcoding approach has proven to be powerful tool, uncovering on average 188.1 OTUs per samples. However, it is worth noting that sampling location might not reflect the true microbial assemblage of these three Zambian ecoregions. Sampling was performed in rural environment but sample collection was not completely random as it was located near roadways and habitats. Moreover, soil composition is likely to vary at a microscale (Yurkov et al., 2011) and more samples would be needed to account for possible soil heterogeneity.

Comparison between culture-based and metabarcoding approaches

Applying a metabarcoding approach to soil biodiversity survey is a new method and is known to be efficient and reliable (Schmidt et al., 2013). However, the *ITS1* culture-based approach uncovered many yeast species which could not be identified using high-throughput approach. Using the metabarcoding approach, only 15 *Cryptococcus* species could be mapped to the species-level and none of them to *C. neoformans* or *C. gattii*. However, the two species were present in 11 soil samples and were identified using the *ITS1* culture-based approach (Chapter 2). The absence of *C. neoformans* reads using metabarcoding was probably due to the polysaccharide capsule which could not be lysed during the DNA extraction protocol.

The culture-based approach identified 61 species, which is relatively small compared to the 188.1 OTUs per sample found with the Illumina MiSeq run. However, the Sanger-sequencing method uncovered 15 different *Cryptococcus* species. Almost half of the genera that were recovered belongs to the *Cryptococcus* genus. Similar results were obtained in a survey of worldwide cold rocky habitat using the same *ITS* technique (Selbmann et al., 2014). The major drawback of culture-based approach is the lack of statistical power to differentiate fugnal patterns among regions (R_{ANOSIM}= 0.0113, p <0.004) owing to the low numbers of taxa that can be screened.

3.4.3 Fungal Diversity Present in Zambia

Investigating the evolution of microorganisms and understanding their ecological interactions within microbial communities have been debated for more than 40 years in microbial ecology (Johnson and Stinchcombe, 2007). High-throughput metabarcoding allows one to investigate whether regional origin defines fungal patterns. The natural history of the *Cryptococcus*-species complex is poorly understood and Zambia provides an ideal environment to study the habitat of these two pathogenic species.

The variation of beta-diversity among fungal communities of different ecoregions was significant (R_{ANOSIM}=0.3704, p <0.001). The results showed that geography plays an important role in shaping fungal communities. Elevation and latitude were of particular interest as their regression coefficient was higher than the regression coefficient based on fungal BrayCurtis dissimilarity patterns among ecoregions (R_{ANOSIM}=0.4272, p <0.001 and R_{ANOSIM}=0.4184, p <0.001

respectively). In general ADONIS results corroborated the results obtained in ANOSIM, except for longitude which was not significant under ADONIS. In the field of microbial community ecology, it is common practice to combine various ecological metrics (Qian, 2009).

ANOVA was performed in order to identify key taxonomic features which could explain the observed regional fungal patterns. However, only two genera could be identify and no Basidiomycota was found once accounting for multiple testing. Other significant taxa could not be assigned to a reference in the database at the genus level.

3.4.4 Fungal Diversity within in Zambian Ecoregions

C. neoformans and *C. gattii* are associated with two distinct ecoregions (p <0.001, Chapter 2). *C. neoformans* is principally associated with *C. mopane* trees which are mainly present in the Zambezian Mopane Woodlands (ZMW) (Figure 2.4). *C gattii* is found in the wetter part of the country in the higher-altitude Central Miombo Woodland dominated by *Brachystegia* and *Julbernardia* species. Across the three ecoregions investigated the proportions of the five dominant fungal phylum were similar. The species richness was higher in the ZMW (Table 3.5). Strong differences could be observed in the fungal composition of CMW and ZMW (R_{ANOSIM}=0.3704, p <0.001) in both the rainy and the dry season (Figures 3.7 & 3.8). This finding suggests that *C. neoformans* is adapted to a different biotic environment than *C. gattii*. In the ZMW, 22.4% (n=197) of identified genera were unique to the ecoregion, 16.2% (n=136) in the CMW (Figure 3.6). However, no taxonomic feature could be identified as key component of the observed variation between the ecoregions (Tables 3.6 & 3.7).

When investigating the potential effect of space on fungal diversity, the variations of the Bray-Curtis metric were uniform (R_{ANOSIM}=-0.008, p=0.499) between Northern and Southern Zambia. This result suggests that the metabarcoding technique was successful in describing the diversity at the ecoregion level despite the potential risk of soil composition heterogeneity. However, the spatial analysis while accounting for seasonality could only be performed in the ZMW and in the dry season due to the reduced number of samples.

The relative abundance of the *Cryptococcus* genus was higher in the ZMW compared to the CMW across both seasons (Table 3.9). The success of the genus in the Mopane region could be link to either physical or biological factors. *C. neoformans* might be more prone to compete

with other members of the genus. However, caution is required when interpreting relative abundance of a particular genus as this measure does not properly describe taxonomic biomass. Filamentous organisms are particularly underrepresented, whereas yeast cells tend to be more abundant (Lindahl et al., 2013).

3.4.5 The Role of Seasonality on Fungal Communities

The rarefaction curves (Figure S1) revealed a similar amount of diversity across seasons. The alternation of the rainy and dry seasons seemed to principally alter the fungal community structure of the mopane region (R_{ANOSIM}=0.3610, p=0.019). In the CMW, ANOSIM indicated that the microbial patterns remained stable (R_{ANOSIM}=0.1062, p=0.189). *C. neoformans* might be better able to compete with other fungal species than *C. gattii* since the community associated with *Cng* significantly changes across season.

In all regions, the proportion of Ascomycota and Basidiomycota increased in the dry season. The *C. neoformans* recovery was also slightly higher in this season. Here, I hypothesise that cryptococcal infections are higher in the dry season when the drier climatic conditions favour *C. neoformans* growth as shown by Nielsen et al. (Nielsen et al., 2007). The relative abundance of the *Cryptococcus* genus decreased in the dry season (Table 3.10) possibly due to the increased competition from other Basidiomycetes. In each ecoregion, the number of unidentified species was higher in the rainy season which could explain the higher proportion observed of Ascomycota and Basidiomycota. Additionally, during the rainy season, fungi are able to grow on rotting wood while during the dry season, desiccated spores are more likely to be dispersed (Velagapudi et al., 2009).

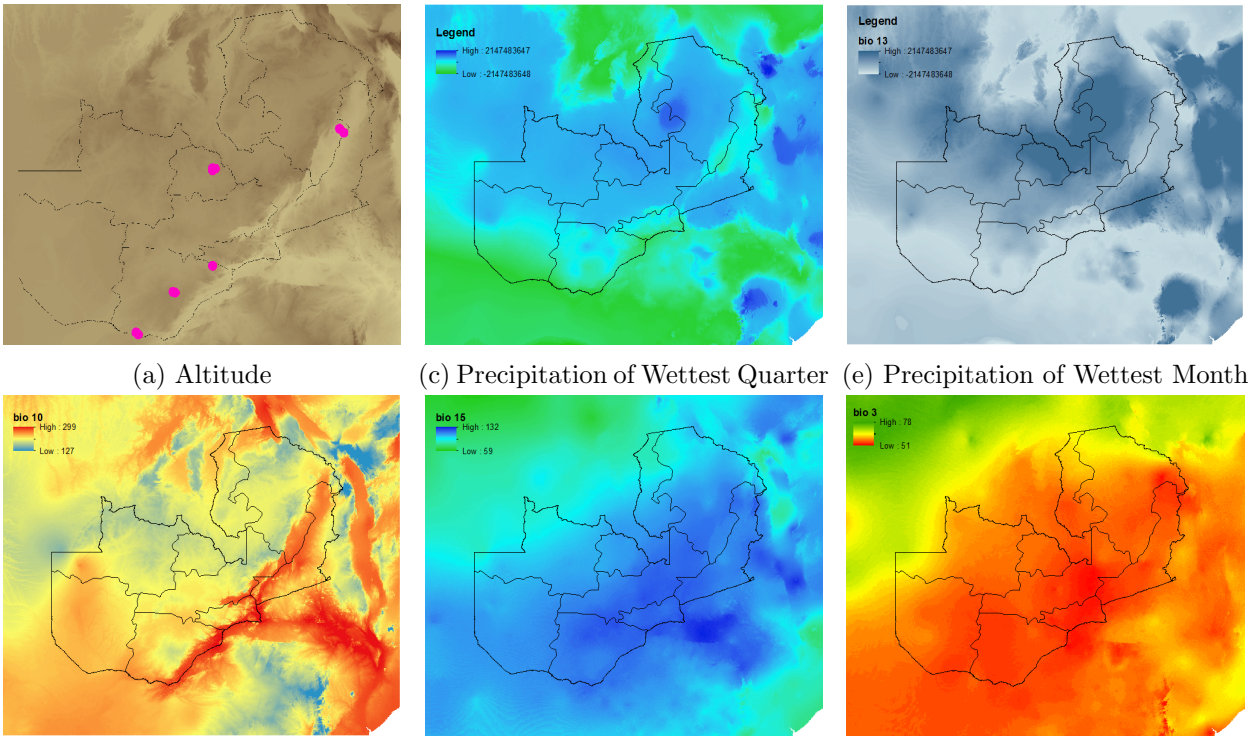
3.4.6 Environmental Conditions and Soil Community Structure

Environmental conditions strongly influence the phenotype of microbial species. They substantially limit their fundamental niche and therefore affect their distribution (Kozak et al., 2008). Nowadays, Martinus Wilhelm's statement 'Everything is everywhere: but the environment selects'(Baas-Becking, 1934) is considered as a central paradigm in microbial ecology. The development of high-throughput technology allows scientists to obtain the statistical power to

study the correlation of microbial assemblage with climatic conditions. Permutational ANOVA and best variables rank correlation test (BEST) have proven to be efficient to study the factors which explain the degree of β-diversity dissimilarity among fungal communities (Bokulich et al., 2014).

Significant climatic factors had a higher MANOVA R^2 coefficient than with geographical variables (Table 3.5) indicating the important role of climate on species distribution. The main climatic factors were Precipitation of Wettest Quarter (bio16) (R^2=2.6363, p <0.001), Precipitation of Warmest Quarter (bio18) (R^2=1.5652, p <0.001) and Precipitation of Driest Month (R^2=1.4629, p <0.001). Precipitation is significantly lower in the Mopane woodlands supporting the hypothesis that *C. neoformans* is more adapted to dry conditions (Granados and Castañeda, 2005a). The BEST analysis also hinted toward the key role of precipitation ranking Precipitation of Wettest Month (bio13) as the principal factor explaining microbial assemblage variability. Generally, the BEST analysis and the MANOVA analysis gave similar results. The second variable of the BEST analysis was Mean Temperature of Warmest Quarter (bio10). Temperatures were more elevated in the ZMW suggesting that *C. neoformans* tolerates better high temperatures. Figure 3.16 shows the six principal variables accounting for differences between CMW and ZMW. All the climatic variables are displayed on Table S4.

A similar approach investigating the correlation of the *Cryptococcus neoformans* and *C. gattii* species complex and climatic factors was performed in Columbia. Granados & Castaeda (2006) recovered *Cryptococcus* from serotype A (*Cng*), B (*Cg*) and C (*Cg*) in various Columbian cities and found that *C. neoformans* var. *grubii* was favoured by high and mean temperature values while *C. gattii* distribution was restricted mostly by extreme temperature conditions. *C. gattii* (serotype B) was highly correlated with humidity. Similar pattern was observed in this study where Precipitation Seasonality (bio15) (R^2=1.4653, p=0.003) was the most significant climatic variable revealed by both MANOVA and BEST tests in the CMW. Soil features such as organic carbon, pH or sand content are known to have a significant impact on community assemblage (Lauber et al., 2009) but the data regarding African soil properties remained limited despite the recent release of a 250m-high resolution map (Hengl et al., 2015).



(a) Altitude (c) Precipitation of Wettest Quarter (e) Precipitation of Wettest Month

(b) Temperature Warmest Quarter (d) Precipitation Seasonality (f) Isothermality

Figure 3.16: Principal physical and bioclimatic variables (a) Altitude (b) Temperature of Warmest Quarter (bio 10)(c) Precipitation of Wettest Quarter (bio 16)(d) Precipitation Seasonality (bio 15) (e) Precipitation of Wettest Month (bio 13)and (f) Isothermality (bio 3). Sample locations are given on figure a.

3.4.7 Ecological niche modelling

Ecological niche modelling (Figure 3.15) confirms the influence of climate on the distribution of the two species. The models reflect the environmental tolerances of the different species investigated (Kearney and Porter, 2004). The Maxent predictions are matching the two principally investigated ecoregions: the CMW for *C. gattii* and the ZMW for *C. neoformans*. *Cng* distribution also included the SMW which acts as a transition area between the low-land mopane woodlands and the central Miombo region. Important variables explaining the Maxent model were identified in the MANOVA and BEST analysis. Isothermality (bio 3) and precipitation seasonality were accounted for 71.2% and 22% of *Cng* distribution respectively. Regarding *C. gattii*, precipitation was a significant factor; five precipitation variables contributed to the model by more than 5% (Table S1).

The distinct distributions of the two sister species suggest that ecology played a role in speciation. Elevation was the one of the most correlated factors. In the study of evolution of environmental pathogen, altitude is known to facilitate the specialisation of pathogenic species in their ecotope (Messenger et al., 2015). The principal reason is that elevation acts as a physical barrier which limit gene flow and the dispersal of populations (Losos and Glor, 2003). However, the natural history of both *C. gattii* and *C. neoformans* remains virtually untouched in southern Africa and additional research using genomic methods should investigate the phylogenetic of the species-complex to understand the possible allopatric or parapatric speciation of this genus. Investigating the factors associated with speciation can have practical implications on emergent diseases. For example, the predictions based on GIS data could be used to estimate the pathogen response to climate change (Thomas et al., 2004).

3.5 Conclusion

My analysis of beta-diversity dissimilarity revealed that the fungal community composition in the different ecoregion varied significantly (R_{ANOSIM}=0.3704, p <0.001). These differences were amplified by both physical factors, such as elevation, or climatic features such as variation in precipitation. The microbial assemblage in the ecoregions associated with either *C. neoformans* or *C. gattii* had different climatic tolerances. ZMW, where *C. neoformans* is present, is

subjected to higher temperature and less precipitation than the Miombo woodlands. The drier conditions of the *C. neoformans* ecological niche could have led to the acquisition of a selective advantage. As hypothesised by Nielsen et al. (Nielsen et al., 2007), this potential advantage may have allowed *C. neoformans* to survive and sexually reproduce in pigeon guano before spreading throughout the world. This survey only accounts for the fungal diversity present in soil ecosystem but the *Cryptococcus* distribution is likely to be explain by bacterial communities (Desnos-Ollivier et al., 2010) and other abotic factors such as wind, soil nutriment concentration or pH (Granados and Castañeda, 2006). Soil is also not the unique niche where the *Cryptococcus*-species complex evolves and the microbial community composition of the principal tree types associated with the pathogen are likely to contribute to a better understanding of *C. neoformans* evolution in its natural environment.

High-throughput metabarcoding allowed to study the influence of regional origin on fungal patterns. These methods will gain interest in the future as research expands and more environments become characterised. For example, the Earth Micro Biome project (http://www.earthmicrobiome.org/) aims to describe the global microbial taxonomic and functional diversity (Gilbert et al., 2010a). In the context of emerging diseases, predictions of the probable emergence of diseases can lead to the surveillance of particular environments. Future work is likely to progress at impressive rate with the development of shotgun metagenomics (Hodkinson and Grice, 2015). These methods are likely to contribute to the understanding of diseases progression mechanisms and ecosystem services.

Finally, the adaptation to different environments is likely to have significant implications on the *Cryptococcus* genome. The following chapter investigates the genetic structure of environmental *Cryptococcus neoformans* populations and the role of the environment in explaining the observed population subdivisions.

Chapter 4

African Environmental *Cryptococcus* populations

4.1 Introduction

To understand the pathogenicity of microbes, it is central to study the evolutionary ecology of wild populations. Recent advances in population genomics allow the investigation of ecologically relevant adaptations and genetic mechanisms, leading to insights into the emergence of virulence phenotypes. In evolutionary biology, linking the phenotype to the genotype has been a long term goal. With the resolution provided by next-generation sequencing, population boundaries can be accurately described allowing signals of natural selection to be determined. Molecular epidemiology aims to identify genome-level differences between populations. Once the diversity of natural population is understood, specific polymorphisms can then be linked to phenotypes.

In this study, *Cryptococcus neoformans* var. *grubii* (*Cng*) has been recovered from various environments in southern Africa (Chapter 2). Then, the fungal community structures associated with the pathogenic species were characterised (Chapter 3). In the present chapter, the genomes of environment *C. neoformans* were sequenced to uncover the population genetic structure of the pathogen and investigate the role of selection in shaping *Cryptococcus neoformans* diversity across Southern Africa.

4.1.1 Whole Genome Sequencing

In *Cryptococcus* research community, multi-locus sequence typing (MLST) methods have been frequently applied to differentiate between the different *Cng* molecular types. A MLST scheme using seven loci was even adopted in 2009 to investigate *C. neoformans* diversity (Meyer et al., 1999). However, this approach has a low discriminatory power and is unable to fully describe the genetic diversity present in the 14 chromosomes which account for over 6,900 genes in the *Cng* genome (Loftus et al., 2005).

Whole-genome sequencing (WGS) has already revolutionised molecular epidemiology. Its high resolution surpasses other typing methods and its cost is becoming increasingly attractive. The Illumina HiSeq 2500 platform introduced in 2012 can generate 120 Gb of data in 27 hours (Shokralla et al., 2012). The application of next-generation sequencing technology is broad, ranging from ancient DNA research (Mouttham et al., 2015), biodiversity studies (Mangrola et al., 2015) or the surveillance of drug-resistant microorganisms (Abdolrasouli et al., 2015). The number of annotated genome references from fungal, bacteria (Hofreuter et al., 2006) and viral (Gardy et al., 2015) isolates is increasing allowing us to identify mutations, structural variations, gene duplications or copy number variations in the population investigated. From these data, the molecular epidemiology of the pathogen can be described, such as the relative contribution of specific genotypes and mutations, to the process of infection.

Using the *Cryptococcus neoformans* isolate H99 (VNI) reference genome, the study will assess the presence of the different genetic lineages (VNI, VNII and VNB See Introduction, section 1.2.2) in the environment. The characterisation of the environmental population structure of *Cng* will expand the knowledge on *Cryptococcus* epidemiology and population genomics analyses will help to detect population-specific alleles and possible hotspots of genetic diversity which might play a role in virulence and pathogenesis.

4.1.2 Investigating Selection

Darwinian selection is a gradual mechanism where biological traits (phenotypes) vary in a population according to the reproductive success of its individuals. Eventually, this gradual process can lead to the niche specialisation of a population, genetic isolation, and the emergence

of new species. Population genomics methods allow to investigate genomic regions which are under selection and which are markers of ongoing adaptation of the organism to its environment (Guillot and Rousset, 2013).

Two methods are available to find genes subjected to selection. First, 'outlier'loci which deviate from genome-wide patterns of diversity are identified. These markers exhibit higher level of genetic differentiation than expected under neutrality (Holderegger and Wagner, 2008). The second approach assesses the correlation between potential genetic markers and environmental variables using logistic regressions (see section 4.1.4).

Selection is thought to be the main mechanism of adaptation (Akey, 2009). It is the non-random propagation of an allele in a population as a result of its phenotypic effect. Alleles enhancing the organism fitness are favoured (positive selection), whereas alleles reducing fitness are gradually removed (purifying or negative selection) (Li et al., 2008). To detect signature of Darwinian selection, several statistical methods are available (reviewed by Vitti et al. (2013)).

Gene-base methods compare the rate of synonymous and nonsynonymous mutations in protein-coding genes (Yang and Bielawski, 2000). The McDonaldKreitman (MK) test is widely used and identifies patterns of selection by comparing the number of silent (dN) and non-silent substitutions (dS) with the number of silent (pS) and non-silent polymorphism (pN) of an outgroup species (Stoletzki and Eyre-Walker, 2011). The ratio of the rates of synonymous and non-synonymous substitution (dN/dS) is also commonly used (Hurst, 2002).

Frequency-based methods such as Tajima's D (Tajima, 1989) search for genomic region undergoing a selective sweep. A sweep leads to reduced level of diversity near the selected locus. Then, new mutations appear in this region causing an increase of rare alleles. Smaller values of Tajima's D are indicative of a surplus of rare alleles and therefore positive selection or recent population expansion (Charlesworth, 2006).

4.1.3 Biogeography

Biogeography studies diversity over time and space. It provides insights in the mechanisms which maintain microorganisms diversity and explain species dispersal (Martiny et al., 2006). In the previous chapter, we found that climatic variables played a significant role in shaping

fungal community structure. The study investigates the role of geography and climatic factors on shaping the genome-wide population genetic diversity of *Cryptococcus*.

Individuals which are closer to each other tend to be more genetically similar principally due to limited dispersal (Wright, 1943). Recent investigations indicate that isolation by distance (IBD) is expected among spore forming fungi since dispersal over short distances is thought to occur at higher frequency than long-range dispersal (Travadon et al., 2011). Mantel tests have been used for decades to study spatial relationships. However, Mantel tests assume a linear relationship between variables (Guillot and Rousset, 2013) and are confounded by spatial autocorrelation (Legendre and Fortin, 2010). This approach is unable to directly assess the importance of the investigated variable. Therefore, a reverse-ecology approach will here be used to disentangle relative impact of evolutionary forces on the genome of *Cryptococcus*. In comparison to forward-ecology which aims at characterising the function of potential candidate genes, reverse-ecology represents an unbiased approach. The latter identifies ecologically relevant genes by examining signatures of adaptation among populations (Ellison et al., 2011). This emerging field is called landscape genomics.

4.1.4 Landscape Genomics

The availability of large sequencing data due to the advances in NGS technologies now allows researchers to uncover candidate regions using hypothesis-generating approaches (Vitti et al., 2013). Through genome scans, genetic variants can be identified using various genome-wide statistical methods. In ecology, the field of landscape genomics is gaining importance. The term landscape genetic was coined in 2003 and it aims to investigate the statistical relationship between landscape and genetic data (Manel et al., 2003). Statistical methods using linear mixed models or Bayesian approaches are been developed to make use of the large quantity of genomic data. Thousands of genetic markers can be investigated in the light of georeferenced samples to gain insights on evolutionary processes such as gene flow, genetic drift or selection (Schwartz et al., 2010). Genes under selection can then be identified and then the effects of environmental variables which drives the differentiation of these genes can be quantified (Bolliger et al., 2014). Finally, the ultimate goal is to associate these genetic patterns to observed fitness differences among the studied populations (Holderegger and Wagner, 2008). The study described here will attempt to understand the nature of the genes involved in genetic differentiation between

Cng lineages and identify genomics regions with outstanding correlation with environmental variables.

Aims:

1. Perform whole-genome sequencing on environmental isolates of *Cryptococcus neoformans* var. *grubii*
2. Investigate the population genetic structure of environmental *C. neoformans* var. *grubii*
3. Identify genes under selection within the different molecular types
4. Investigate the role of the environment on shaping *Cryptococcus neoformans* diversity using a landscape genomics approach

4.2 Methods

4.2.1 Whole-Genome Sequencing of Environmental *Cryptococcus neoformans*

High molecular weight DNA was and DNA libraries were prepared for Illumina HiSeq paired-end sequencing. 50 environmental genomes from isolates of *Cryptococcus* collected across Southern Africa were sequenced with 100x coverage. Reads were aligned to the *Cryptococcus neoformans* var. *grubii* H99 reference using the GATK pipeline and single-nucleotide polymorphisms (SNPs) were identified.

Isolation of *Cryptococcus neoformans* from the environment

As described in Chapter 2, *Cryptococcus neoformans* were recovered from the environment by plating samples from various extracts onto Niger-seed agar. The extraction process and identification are also described in Chapter 2. A total of 50 environmental *Cryptococcus neoformans* var. *grubii* were collected in this study and sequenced. A total of 25 isolates were recovered from Zambia, 11 isolates in South African genomes and two isolates in Namibia. Additionally, eight isolates from South Africa were provided by the University of Stellenbosch and four

isolates from northern South Africa and Botswana were provided by the Broad Institute for a total of 50 environmental *C. neoformans* genomes. The complete details of the samples used in this study are listed in the appendix (Table S5).

Library Preparation and Illumina sequencing High molecular weight DNA was obtained by culturing *Cryptococcus neoformans* isolates into five ml of Yeast Protein Digest (YPD) media (Sigma-Aldrich) supplemented with 0.5M of NaCl for 60 hours at 37°C. DNA was extracted using the MasterPure Yeast DNA Purification kit (Epicentre) as described in chapter 2 (section 2.2.3) and the genomic DNA was resuspended in Buffer EB (Quiagen). DNA concentrations were measured with a Qubit® Broad Range dsDNA Assay (Life Technologies™). Briefly, 2 μl of extracted DNA were added to 198 μl the Qubit® working solution. After vortexing for 3 sec and incubating the tubes at room temperature for two minutes, the DNA concentrations were read from the Qubit® 2.0 Fluorometer (Invitrogen™, Q32866). The genomic DNA was then diluted to 2 ng.μl^{-1}.

Libraries were prepared using the TruSeq® Nano DNA Sample Preparation Kit (Illumina, FC-121-4001). Each time, 24 libraries were prepared for paired-end sequencing on two lanes of Illumina® HiSeq in order to sequence 175 bp fragments. First, 100 ng of gDNA in 52.5 μl were transferred to a Covaris™ microTUBE and the gDNA was sheared using the S2 Ultrasonicator (Covaris™) using the parameters recommended by Illumina®. Then, 50 μl of the fragmented DNA were transferred to a 96-well MicroAmp Optical 96-Well Reaction Plate (Life Technologies) and the clean-up was performed immediately. Sample purification beads were vortexed for 1 min and 80 μl of beads were added to the gDNA. After a 5 min incubation at room temperature to allow the DNA to bind to the beads, the plate was placed onto a magnetic stand for 5 min and the supernatant was discarded. 200 μl of freshly prepared 80% EtOH were added to each well taking the precaution to not disturb the beads. The plate was left for 30 sec to incubate at room temperature and the supernatant was discarded. The wash was repeated a second time. The samples were left to air dry for 5 min to remove the remaining ethanol. 62.5 μl of Resuspension Buffer was then added to each well.

The End Repair step allows to convert into blunt ends the overhangs of the DNA fragments. The 3 to 5 exonuclease and 5 to 3 polymerase fill the gap at 3 and 5 end of the fragments respectively. 40 μl of the End Repair Mix was added to the plate and incubated at 30°C for 30

min. Size selection was performed removing first large DNA fragments and in a second clean-up step removing the smaller fragments to obtain 100 bp DNA fragments. The Purification beads were well-mixed and were then diluted by a factor of 1.6 and 160 μl of the diluted beads were added to each well. The plate was incubated for 5 min to allow the larger fragments of DNA to bind onto the beads and the plate placed onto a magnetic stand for 5 min at room temperature. 250 μl of the supernatant were transferred into another 96-well plate. To remove the smaller DNA fragments, 30 μl of undiluted Purification Beads were added to each well. The plate was incubated at room temperature for 5 min and placed onto a magnetic stand for 5 min. The supernatant was discarded and the DNA was cleaned with two washes of 200 μl of 80% ethanol. The plate were air-dried for 5 min and 17.5 μl of Resuspension buffer were added to each well and transferred into a new plate.

A-tailing was then performed to add a single A nucleotide at the 3 ends of the DNA fragments. This step prevent the ligation of DNA fragments during the ligation reaction and therefore reduces the rate of chimera in the reaction. 12.5 μl of the A-Tailing Mix were added to each well. The plate was centrifuged for 1 min at 500 rpm and incubated at 37°C for 30 min and at 70°C for 5 min.

Adapters were ligated onto the DNA fragments as followed: 2.5 μl of Resuspension Buffer, 2.5 μl of Ligation Mix and 2.5 μl of the appropriate DNA Adapter Index were added to each well; the plate was centrifuged for 1 min at 500 rpm and incubated at 30°C for 10 min. After the incubation, the reaction was stop by adding 5 μl of Stop Ligation Buffer to each well. A clean-up step as described previously was performed before the enrich PCR and 25 μl of DNA fragments were obtained.

The enrichment step amplifies fragments which have adapters on both ends. 5 μl of PCR Primer Cocktail and 20 μl of Enhanced PCR Mix were added to each well. The plate was centrifuged for 1 min at 500 rpm. The PCR was performed in a thermal cycler with each well containing 50 μl. The PCR conditions were 95°C for 3 min, followed by 8 cycles of 98°C for 20 sec, 60°C for 15 sec and 72°C for 30 sec and a final step of 72°C for 5 min. A final clean-up was performed and 30 μl of the final libraries were obtained.

Libraries were quantified using qPCR on an Applied Biosystems 7300 instrument (Life Technologies) using the Kapa library quantification kit (Kapa Biosciences, Boston, MA) and the 2200 TapeStation (Agilent) with D100K ScreenTape assays (Agilent) as described in Chapter

3. The libraries were then normalised to 10 nM and pooled together. The pooled libraries were sent to Medical Research Council Clinical Genomics Centre (Hammersmith, UK), and were sequenced to a read length of 100 bp on an Illumina® HiSeq 2500 sequencer. A total of 12 isolates were sequenced on a lane of flow cell, loading 16 pM of the pooled libraries. All raw reads and linages information regarding each isolates have been submitted to the European Nucleotide Archive.

Alignment and SNP identification The quality of all Illumina WGS raw reads using FastQC (version 0.10.1; Babraham Institute) and aligned to the genome assembly released by the *Cryptococcus neoformans* var. *grubii* H99 Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/). Reads were aligned to the *Cng* H99 genome using the short-read alignment component of Burrows-Wheeler Aligner (BWA) 0.75a aln (Li and Durbin, 2009) using a quality threshold of 15 as explained by Rhodes et al. (2014). The suffix array format generated from FastQs were converted to SAM format using BWA sampe option. SAM files were then converted to BAM files and the BAM files were then sorted and indexed with SAMtools version 0.1.18 (Li and Durbin, 2009). Duplicated reads were marked with Picard (v. 1.72). The resulting BAM files were recalibrated around insertions or deletions (INDELs) using the GATK RealignerTargetCreator and IndelRealigner (McKenna et al., 2010).

The detection of Single-nucleotide polymorphisms (SNPs) and INDELs were called using GATK UnifiedGenotyper version 2.2-2 in haploid mode (Auwera et al., 2013; DePristo et al., 2011). SNPs and INDELs were only called if they appeared in 80% of reads.

4.2.2 Phylogeny

Whole-genome SNPs files were converted to Nexus and Phylip format. RAxML (Stamatakis, 2006) was used to generate Maximum-likelihood trees and the resulting trees were visualised in FigTree version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree).

4.2.3 Population Assignment and Admixture

ChromoPainter and fineStructure Analyses

To first analyse the population structure of environmental *Cryptococcus* populations, we used ChromoPainter and fineStructure (Lawson et al., 2012). ChromoPainter constructs a coancestry matrix based on individual SNPs. Each individual is considered in terms of either being a donor or a recipient of chunks of DNA. The coancestry matrix then records the inferred recombination events between each donor and recipient prior to coalescing with another genome. Two models can be run using ChromoPainter, the first one does not account for linkage disequilibrium, whereas the LD model offers an increased sensitivity to detect population structure as it uses LD between closely positioned SNPs to assign a given region of the genome to its best neighbour.

Using the ChromoPainter similarity matrix, fineStructure performs a model-based Bayesian clustering to identify the most parsimonious population structure. The model assumes that genomes from the same population, which share more DNA regions between each other, will have a similar amount of admixture with isolates from other populations. FineStructure then successively relates the different isolates with a phylogenetic tree using a step-wise approach. At each step, the isolate that is most likely to be merged into a population is the one which has the highest probability using a maximum likelihood approach. FineStructure allows to determine *Cng* population structure within each molecular type. On the fineStructure output, the scale is a gradient going from yellow to blue. The blue colour represents the greatest among of sharing between the donor and recipient genome and yellow the least.

To corroborate the results obtained in fineStructure, population stratification was further investigated using Admixture v1.23. This software uses a maximum likelihood model to estimate population allele frequencies along with ancestry proportions (Alexander et al., 2009). The software can give useful information of recent admixture (recombination) events between populations. An advantage of Admixture is its cross-validation procedure which allows finding the number of K populations that best fits the dataset.

Pairwise population gene flow

BEDASSLE was used to estimate the fixation statistic F_{ST} between each population. Wrights F_{ST} (Wright, 1943) statistic measures the difference of allele frequencies between populations. The F_{ST} values are inflated when population subdivision exists. On the other hand, if for two populations $F_{ST} = 0$, the genetic variation is randomly distributed and no subdivision exist (Cruickshank and Hahn, 2014).

4.2.4 Mating types

The mating type for each environmental *Cng* isolates was determined using PCR analysis. The reaction was performed in a total volume of 25 μl and was composed 0.8 μl of each of the following primers (10 μM), *JOHE7264*, *JOHE7265*, *JOHE7270*, *JOHE7272* (Reimão et al., 2007). 0.5 μl of MgCl2 was added to the reaction mix with 0.5 of dNTPs (10 μM), 2.5 μl of Rxn Buffer, 17.4 μl of H_2O and 0.1 of Taq polymerase. The reaction conditions were as followed: *JOHE7264/JOHE7265*; *JOHE7072/JOHE7272*:

Control method: calculated

1. Initial Denaturation: 94°C for 3:00
2. Denaturation: 94°C for 0:30
3. Annealing: 50°C for 0:30
4. Extension: 72°C for 1:00
5. go to 2, (x 35)
6. Final Elongation: 72°C for 10:00
7. Hold: 4°C for ever

4.2.5 Biogeography: Effect of Distance on Genetic Divergence

Environmental Variables

See Chapter 3, Section 3.2.2

Mantel Tests

Isolation by distance across all individuals was tested using a Mantel test comparing the genetic and the geographic distances between individuals (Manel et al., 2003). Mantel tests assess the association between genetic and geographic distance and detect spatial autocorrelation (Mantel, 1967). Genetic variation was calculated as the Bray-Curtis distances between loci and geographic distances were the Euclidean distances between the sampling localities. Mantel tests and partial Mantel tests were performed using the ecodist package (Goslee and Urban, 2007) in R software using 10,000 permutations. Environmental factors were extracted from Bioclim layers (Busby, 1991). The information for each sample was extracted in R software (v 3.1.1) using the raster (Hijmans et al., 2012) and dismo (Hijmans and van Etten, 2012) packages.

Redundancy analysis (RDA) was performed in R using the package vegan. RDA is an ordination method which examines the variations of how a set of variables is explained by another set. In this study, RDA will investigate how much of the genetic variation is purely attributable to either climate or spatial distances and how much can be explained by a joint effect of both.

Environmental niche modelling

See Chapter 3, section 3.2.4

Landscape Genomics

Partial mantel tests are unable to quantify the relative effects of environmental variables. The test assume a linear relationships between the variables (Guillot and Rousset, 2013) and is unable to account for spatial autocorrelation (Legendre and Fortin, 2010).

Latent Factor Mixed Model (LFMM) is a Bayesian approach used to detect selection in landscape genomics (Frichot et al., 2013). The method investigates the influence of population structure on allele frequencies by introducing unobserved variables as latent factors (Stucki et al., 2014). LFMM provides a way to investigate signatures of local adaptation by identification of high degrees of correlation between polymorphism and environmental variables. The test is based on regression models.

The advantage of this method over other population genomics approaches such as BayEnv (Günther and Coop, 2013) is that potentially significant associations between loci and environmental variables are calculated while accounting for population structure. To detect signatures of selection, the statistical significance threshold was set to 1% and 0.05% before applying Bonferroni correction. A positive false discovery rate of 0.05 and 0.01 was also applied using the qvalue package (Dabney et al., 2004) in R.

4.2.6 Population Genomic Analyses

Identifying candidate genes involved in virulence

Literature searches were performed using the keyword 'virulen*'in the fungiDB database (Stajich et al., 2011) using the *Cryptococcus neoformans* var. *grubii* H99 genome and yielded 317 virulence- associated genes. A subsequent literature search on Web of Science with the keywords Cryptococcus and virulence between 2010 and 2015 allowed to identify another 153 genes with a known virulent function. A total of 470 *Cng* genes associated with virulence were obtained. The complete list of gene is given on table S17.

Population Genomic Summary Statistics

General population statistics, number of segregating sites (S), total number of mutations, number of singletons, nucleotide diversity (π), Waterson's estimator (θ), Tajima's D and LD were estimated using VariScan v.2.0 (Vilella et al., 2005). The nucleotide diversity measures the degree of polymorphism in a population (Nei and Li, 1979). The Waterson estimator estimates the population mutation rate and decreases when the sample size or the recombination rate increase (Watterson, 1975).

Recombination Analysis

The recombination rates that are characteristic of each molecular type was estimated using LDHat (McVean et al., 2004). Recombination can be highly heterogeneous along fungal genomes, even over short distances. The program implements the composite likelihood method of Hudson (Hudson, 2001) using a Bayesian reversible-jump Markov Chain Monte Carlo approach to detect the presence of recombination hotspots. The present of these highly recombining stretches of sequences (hot spots) lead to the fluctuation of the recombination rate. These regions affect the genome disproportionately, and LDHat allows their identification.

4.2.7 Linkage Disequilibrium (LD) decay

Linkage Disequilibrium (LD) is the non-random association of allele frequencies between loci. High LD is expected in clonal organisms where recombination does not act to break up the association between alleles. Therefore, natural selection and demography affect LD patterns. Linkage Disequilibrium is now commonly used to infer population history and the role of selection on populations differentiation (Myers et al., 2005).

LD decay measures the genetic distance where LD decays to half its maximum value. Tibayrenc & Ayala, (2012) noted that this metric can greatly contribute to population genomics. Here, the r^2 value was computed in PLINK (Purcell et al., 2007) using a 4,000 bp window and plotted against physical distance in R. Organisms where the recombination is restricted will have larger LD decay distances compared organisms where recombination is unhampered (Taylor et al., 2015).

4.2.8 BayeScan

BayeScan 2.01 introduced by Foll & Gaggiotti, (2008) uses an outlier approach to identify candidate loci under natural selection. The method uses the allele frequencies that are characteristic of each population and estimates the posterior probabilities of a given locus under a model that includes selection and a neutral model. The programme then determines whether the model that includes selection better fits the data. This approach allows the simultaneous assessment of the influence of both balancing and purifying selection. Loci under balancing selection will

present low F_{ST} values whereas high F_{ST} values reflect patterns of local adaptation (purifying selection) (Excoffier et al., 2009). Using the previously described methods to assign isolates to a population, BayeScan analyses were conducted on both VNI and VNB lineages. Analyses were conducted using the standard parameters. SNPs were then mapped to the reference genome and the genes containing these loci under selection were identified.

4.2.9 Gene Content Comparison

Population assignment provided the framework to compare gene content using a large-scale blast score ratio (LS-BSR) analysis (Sahl et al., 2014). The 50 environmental genomes were compared to identify the acquisition or loss of genes in each *Cng* subtype. Coding regions were predicted for each genome and concatenated using Prodigal (Hyatt et al., 2010). *In-silico* coding DNA sequences (*is*-cDNA) were clustered using USEARCH (Edgar, 2010) using a 0.9 identity cutoff. *Is*-cDNA sequences were aligned using BLASTN. Each query sequence was aligned to each genome with BLASTN and the query bit score was calculated. The BLAT score ratio (BSR) value of each sequence was calculated by dividing the query bit score by the reference bit score. The resulting LS-BSR matrix was visualised in Multiple Experiment Viewer (MeV) (Saeed et al., 2006). After filtering 50% of *is*-cDNA sequences based on variance, the sequences were clustered using Pearson Correlation metric. Then, a t-test using adjusted Bonferroni correction to correct for multiple tests, was performed between each subtype to identify significant sequences. Finally, the significant *is*-cDNA sequences were blasted to identify *Cng* genes which were present in each group.

4.2.10 Identifying Signals of Adaptive Selection

There are different ways to quantify and detect signals of positive selection. In this chapter, two approaches were taken, a frequency-based approach and a gene-based method (Vitti et al., 2013).

Frequency-based analysis

To detect loci under positive selection, a sliding window approach using a 100-bp width was conducted to estimate two statistical parameters, Tajima's D and D', using VariScan v.2.0 (Vilella et al., 2005). A 'trial-and-error'approach was used to identify the appropriate window size (Hutter et al., 2006). Indeed, when the window is too large, the signal is diluted and when the window width is too small, the results lack signal (Thomas et al., 2012).

Tajima's D (Tajima, 1989) is one of the most commonly used parameter in population genetics and has been extensively used to identify loci under selection (Ellison et al., 2011). It compares the nucleotide diversity estimated from the total number of polymorphic sites against the pairwise differences between individuals (Carlson et al., 2005). The statistic allows to identify a surplus of rare alleles and therefore signals of positive selection. In a selective sweep, genetic variants reach a high prevalence along with their linked variants. From this homogeneous background, new mutations arise and are therefore at low frequency (Vitti et al., 2013). Tajima's D <0 suggests a recent selective sweep and population expansion after a bottleneck due to the excess of rare alleles. Tajima's D >0 suggests multiple alleles present at low or high frequencies which is indicative of balancing selection.

In order to identify strong regions undergoing a selective sweep, the D'statistic is used (Lewontin, 1964). Linkage disequilibrium can be used to detect positive selection. Regions with strong LD might have been under selection or else recombination would have caused linkage disequilibrium to break down and reduce the length of the haplotype (Vitti et al., 2013). Lewontin linkage disequilibrium based-statistics D'is the normalised version of D by the maximum linkage disequilibrium observed in alleles frequencies (Lewontin, 1964). D'range from -1 to 1 and |D'| = 1 is known as complete linkage disequilibrium.

Empirical outliers undergoing positive selection are considered as potential candidates if they are detected subsequently with the two statistics described above in the 0.1% quantile. Virulence-associated genes under positive selection will then be investigated.

Gene-based Analysis

To investigate signs of adaptive selection in genes susceptible to favour a subpopulation, two tests based non-synonymous and synonymous mutations were performed, the McDonald-Kreitman test and Tajima's *D*. The McDonald-Kreitman (MK) test identifies genes with an excess of amino acid substitutions. The test compares the ratio of fixed non-synonymous (dN) to fixed synonymous (dS) mutations, MK test = 1 - (dS + pN) / (dN + pS) (McDonald and Kreitman, 1991). The test was performed using *Cng* VNII as an outgroup (pN and pS). The dN/dS ratio (w) was also calculated. To improve test performances, SNPs with an allele frequency below <20% were removed and genes with at least five SNPs were considered as explained by Liti et al., (2009).

In each of the molecular type sub-populations, empirical outliers undergoing positive selection were considered as strong candidates if they were detected subsequently with at least two statistics described above in the 0.5% quantile (MK test and dN/dS) or 0.1% quantile (Tajima's *D*, D').

4.3 Results

4.3.1 WGS and SNP analysis

Whole genome analyses of 50 African environmental *C. neoformans* identified a total of 659,777 single nucleotide polymorphisms (SNPs) (Table S5). A coverage above 100X was achieved for each of the 50 genomes, and each genome was mapped to the 19Mb H99 reference genome (>80% mapping). Ze17 and Ze20 mapped to the reference at 34.68% and 55.35%, respectively. The low coverage is likely to result from the sequencing of mixed environmental colonies.

As displayed on Figure 4.2, the VNI and the VNB molecular types constitute two clearly separated populations. The SNPs present across the two subtypes were determined. This analysis showed that, among the VNI isolates (n=26), 147,430 SNPs were found. The VNB clade harboured a higher genetic diversity with a total of 564,792 SNPs being observed. The two molecular types had 68,794 SNPs in common (Table 4.2 shared SNPs). The VNI molecular type had 78,132 polymorphisms which were only present in this lineage and 509,332 SNPs were

uniquely found in the VNB molecular type (Table S8 unique SNPs). These results present an improvement from MLST based analysis. WGS provides an increased resolution and can identify sub-populations in a more efficient way compared to MLST (Figure S8).

Table 4.1: Population Genetics Statistics of environmental *Cng*

Population	S	η	η E	π	θ	Tajima's D	D'
Cng (n=50)	659,777	1,164,590	183,830	0.359	0.393	-0.310	0.020
VNI (n=26)	147,430	249,486	70,443	0.353	0.443	-0.787	0.149
VNIa (n=13)	69,099	69,152	53,129	0.201	0.321	-1.724*	0.432
VNIb (n=13)	108,792	185,101	57,096	0.474	0.535	-0.520	0.175
VNB (n=24)	564,792	917,389	158,757	0.404	0.420	-0.137	0.125
VNBa (n=9)	357,513	476,773	333,067	0.336	0.442	-1.247*	0.112
VNBb (n=15)	388,858	560,229	109,363	0.391	0.410	-0.199	0.338

Table 4.2: SNPs differences among the different environmental *Cng* Populations

Populations	VNI	VNB	VNBa	VNBb	VNIa	VNIb
VNI	-	-	-	-	-	-
VNB	68,794	-	-	-	-	-
VNBa	61,846	391,392	-	-	-	-
VNBb	62,520	413,671	228,688	-	-	-
VNIa	68,757	43,019	39,930	39,276	-	-
VNIb	111,152	49,398	43,881	45,645	33,106	-

4.3.2 Phylogeny and Population Assignment

The maximum likelihood tree generated on our 50 environmental genome dataset (Figure 4.2) revealed differences within each of the two lineages. The VNB clade has long terminal branches which indicate extensive pairwise genetic distances between the isolates of this particular clade. Two sub-populations can be clearly identified. In the VNI clade, the short branch length observed indicates that the isolates are closely related to each other compared to isolates of the VNB clade.

Each molecular type was recovered from a different environment. VNB isolates were recovered from rural environment (95%), whereas the VNI molecular type (n=26) was found in both rural (50%) and urban environment (50%). VNI Isolates were from Cape Town, South Africa (n=7), Fort Beaufort, South Africa (n=5), Gaborone, Botswana (n=2), Lusaka, Zambia (n=1) and Helmeringhausen, Namibia (n=1). VNB isolates principally inhabit mopane trees,

especially the VNBb sub-population which had 88% of its isolates (n=14) located in Zambia northern mopane forests (Figure 4.5). The VNBa isolates were also associated with arboreal environments and were isolated from Southern Zambia down to South Africa.

To estimate the number of sub-populations in our datasets, an Admixture analysis was performed (Alexander et al., 2009). The latter revealed that the optimal value for the number of sub-populations was K=2 within the VNI clade (VNIa and VNIb) and within the VNB clade (VNBa and VNBb) (Figure 4.2). These four sub-populations can be observed on the phylogenetic tree.

To investigate further the differences across and within the *Cng* lineages, a fineStructure analysis was performed (Lawson et al., 2012) (see Methods) and identified the same population structure found in the Admixture analysis (Figure 4.4). There was a clear separation of the two molecular clades. In the VNB molecular type, Ze10 and Ze6 (VNBb sub-population) shared extensive polymorphisms with the VNBa population and most likely represent admixed (recombinant) individuals. This finding can also be observed with the Admixture analysis (Figure 4.3). In the VNI population, SAe10 has a similar profile and appears as a VNI recombinant individual in both fineStructure and Admixture analysis.

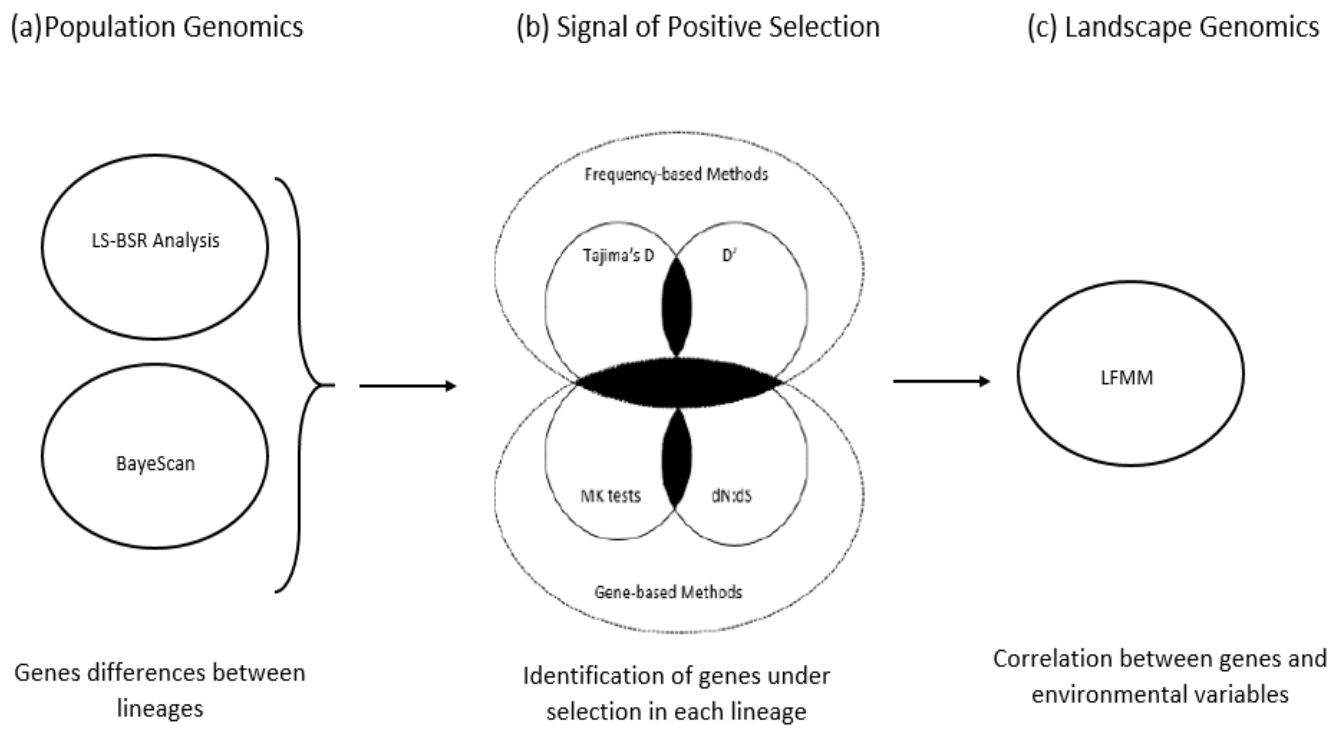
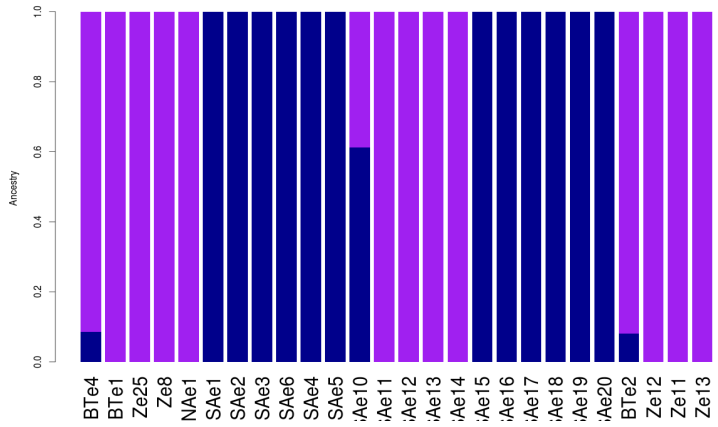


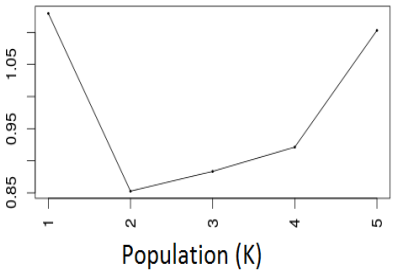
Figure 4.1: Population Genomics Framework - (a) Detection of genes which varies between linages using Bayescan and LS-BSR analysis. (b) Detection of gene under selection. Candidate genes were considered as strong candidates undergoing positive selection if they were detected by at least two statistic parameters, either frequency-based parameters (Tajima's D or D') or gene-based parameters (Mk test or ω). (c) Detection of loci associated with environmental variables using LFMM



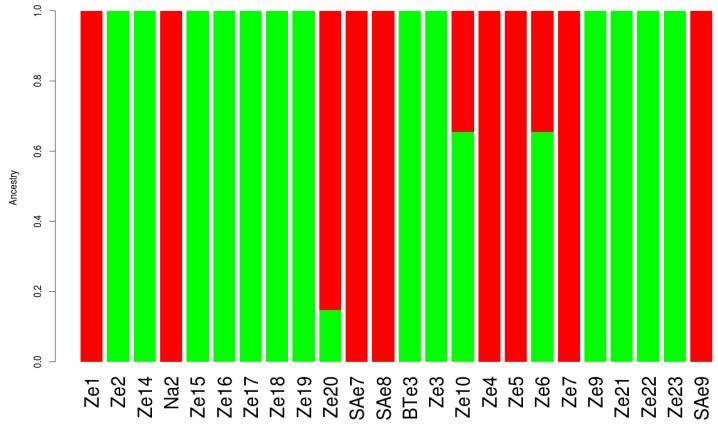
Figure 4.2: Phylogeny using RAxML - Isolates are coloured by populations assigned using *fineStructure* and Admixure. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: namibian isolates, H99 is the VNI reference. Bootstrap values were averaging 100% for each node.



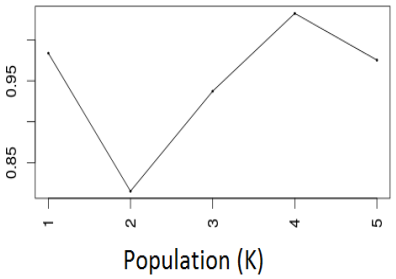
(a) VNI clade



(c) VNI Cross-validation



(b) VNB clade



(d) VNB Cross-validation

Figure 4.3: Admixture Results - In the VNI (a) and VNB (b) clade, two sub-populations can be observed as shown by the admixture cross-validation plot (c and d). Potential hybrids can be observed in both molecular types, SAe10 in VNI and Ze20, Ze10 and Ze7 in the VNB clade.

Pairwise Population Gene flow

The statistic F_{ST} assesses the level of subdivision within a population by comparing the average expected heterozygosity across each locus within subpopulations with the average expected heterozygosity across the entire population. Reasonably high values of F_{ST} were observed between the two VNB sub-populations (F_{ST} = 0.143) and also between the VNI populations (F_{ST} = 0.169). VNB populations had limited gene flow with the two VNI populations (F_{ST} >0.2). The VNBb population showed relatively high levels of subdivision with the VNI clade (F_{ST} >0.180).

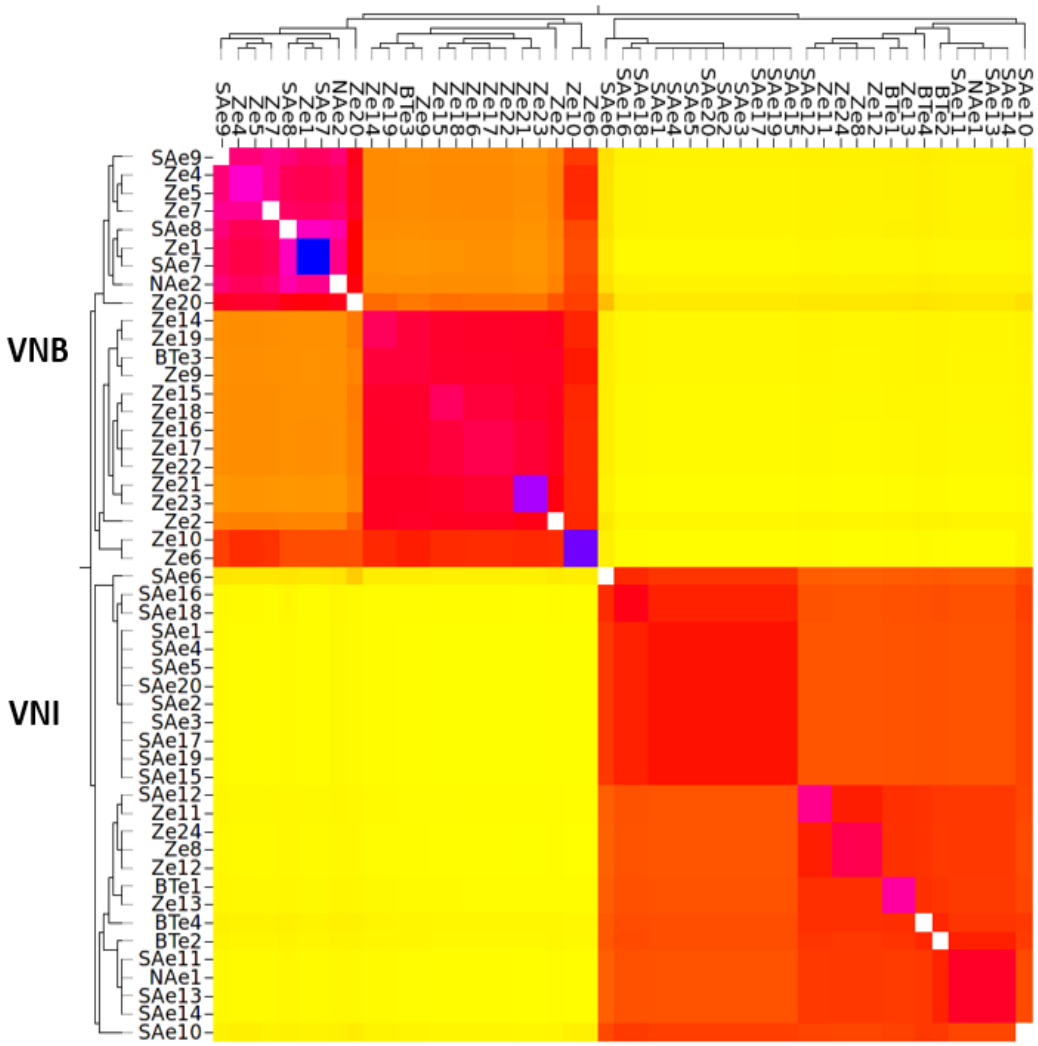


Figure 4.4: FineStructure Analysis (Unlinked Model) - *fineStructure* allow to investigate population structure. On the x-axis, each genome is considered as a recipient, and on the y-axis, the *Cng* isolate is considered a donor of genomic region. The VNB and VNI populations are clearly segregated with very limited sharing of genomic regions between the VNI and VNB populations. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow.

Table 4.3: F_{ST} between populations

F_{ST}	VNIa	VNIb	VNBa	VNBb
VNIa	0.000	-	-	-
VNIb	0.169	0.000	-	-
VNBa	0.208	0.212	0.000	-
VNBb	0.180	0.199	0.143	0.000

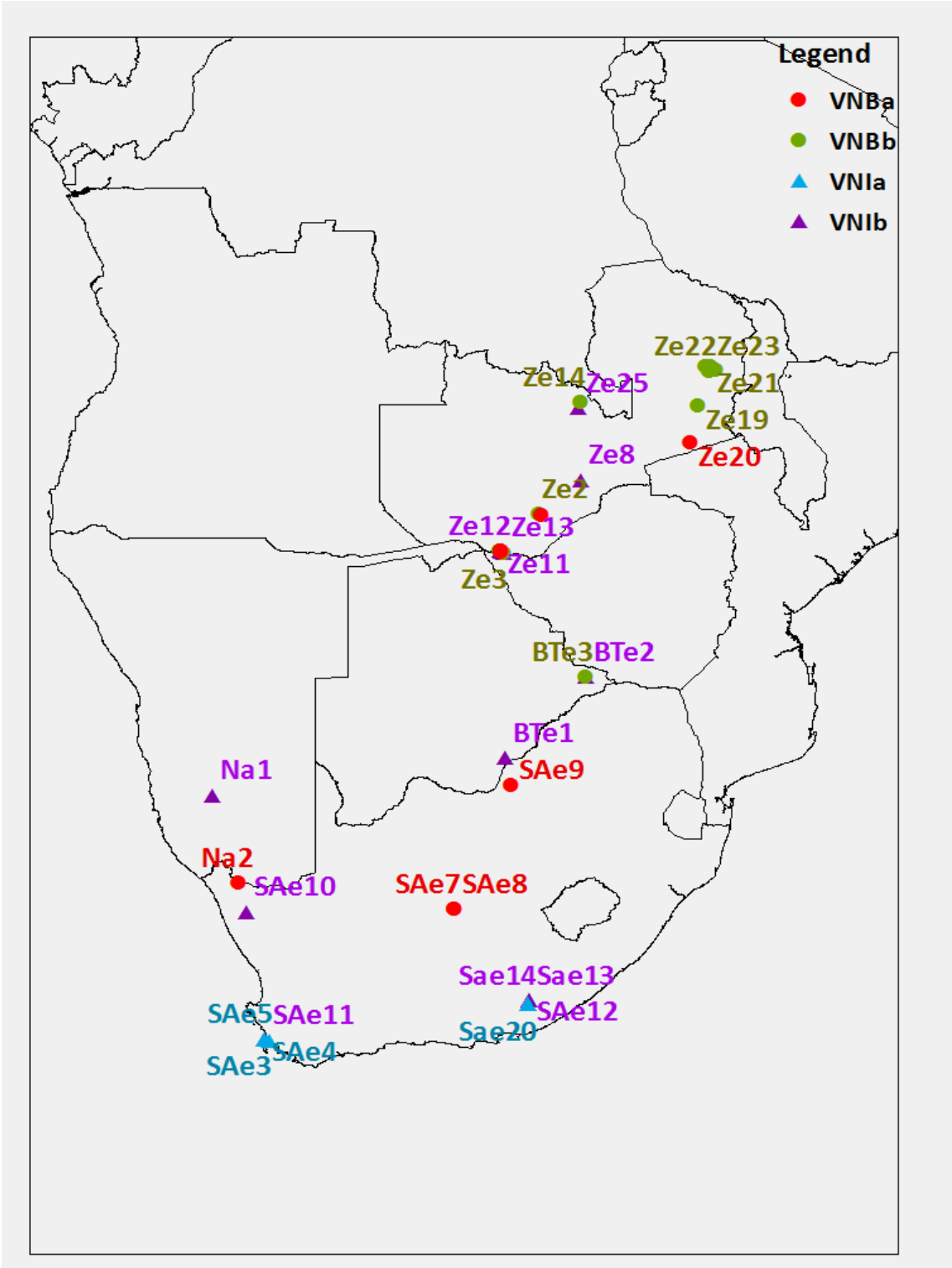


Figure 4.5: Sampling distribution of the different VNI and VNB sub-populations across Southern Africa.

Cng Mating Types

Among the VNB isolates, 14 of the 24 isolates possessed the MAT a locus (58%). In the VNBa clade, seven were positive for MAT a for only two MAT α locus (78% MAT a). In the VNBb lineage, eight isolates were MAT a positive and seven possessed MAT α (53% MAT a). In the VNI clade, only one of the 26 isolates possessed the MAT a locus. The mating type of each isolate is listed on Table S1.

4.3.3 Population Genomic Analyses

The various population genomics measures were used to compare the two molecular types. Genomewide scans were used to identify gene candidates associated with a particular genotype. Then, the study investigated whether these putative genes have known virulent functions.

Population Genetic measures

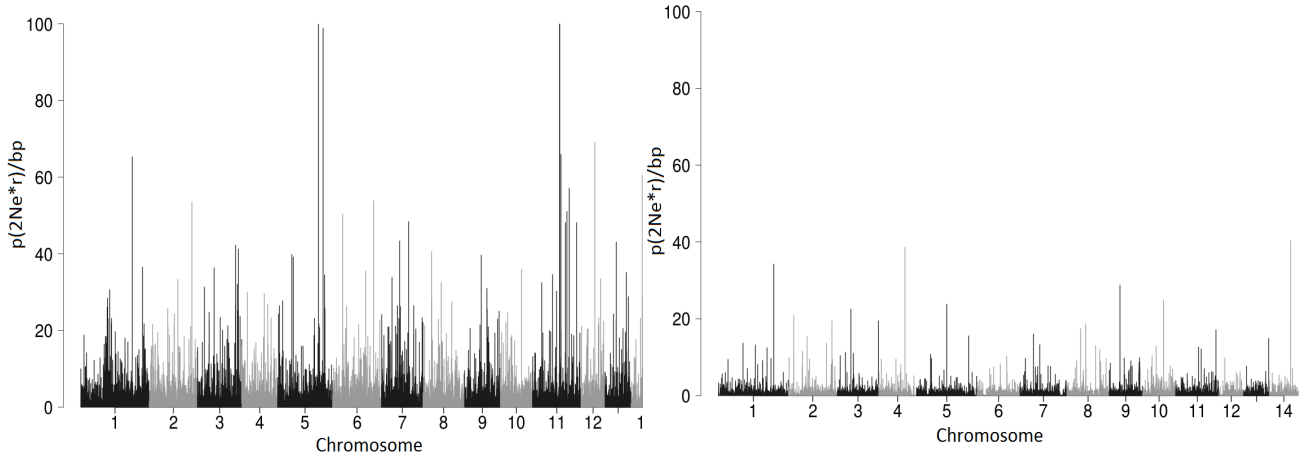
The genetic diversity is higher in the VNB clade as the number of segregating sides (S), number of mutations (η), number of external mutations (ηE), number of nucleotide differences per site (π) are higher in this clade when compared against VNI (Table 4.1). The Watterson's estimate of the population scaled mutation rate (θ) was slightly lower among VNB isolates suggesting an increased sample size or higher recombination rate (Watterson, 1975).

Additionally, the two statistical parameters Tajima's D and D'were computed. Tajima's D value was negative in the VNI clade (Tajima's D_{VNI} = -0.787). Strong negative values are indicative of recent population expansion. Although Tajima's D values for VNI and VNB molecular type were not significantly different from 0, significant negative Tajima's D value were found for the VNIa and VNBa sub-populations. D'in environmental VNI isolates was higher than that seen for VNB lineage indicating strong level of LD in VNI *Cng*.

Recombination Analyses

The recombination rate was significantly higher (calculated using a t-test) in VNB compared to VNI (p_{VNB} = 0.18977/bp^{-1}; p_{VNI} = 0.17521/bp^{-1}; p <0.001). These results corroborate

the phylogenetic observations. VNI molecular type appears more clonal than the VNB isolates which are highly recombinant. However, the VNI clades still presented recombination hotspots (Figure 4.6). The recombination rate affects linkage disequilibrium. The LD decay plots confirmed the higher levels of LD in the VNI clade (Figure 4.7), whereas the VNB clade presented shorter haplotypes indicative of frequent recombination.



(a) VNB isolates (n=24)

(b) VNI isolates (n=26)

Figure 4.6: Recombination analysis, using LDhat interval - The peaks represents the recombination hotspots across the genome.

BayScan

BayeScan 2.01 allows to perform a preliminary examination of selection within the VNB and VNI molecular types (Foll and Gaggiotti, 2008). This approach examines the F_{ST} value of two populations and a common migrant gene pool. BayeScan determines potential outlier markers by estimating the posterior probability at each locus for two models, where selection is or is not included. No significant loci were detected between VNI and VNB molecular types after accounting for multiple testings. Figure 4.8a shows that a high number of loci have elevated -log(P) value. No loci was detected using a false discovery rate of 0.05. However, the volcano plot which compares F_{ST} values to the poserior odds ratio (Figure 4.8b) showed that a high number of loci were under both diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}). A total of 587,464 SNPs differed between the two populations. The high number of differences between the two lineages did not allow to identify any particular locus.

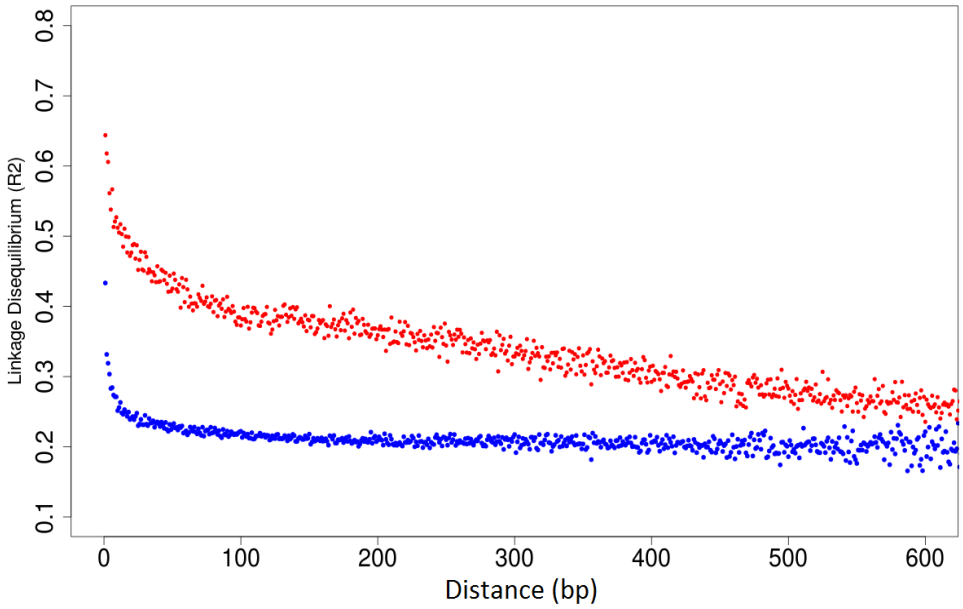


Figure 4.7: Population decay of Linkage Disequilibrium - The rate of decay of r^2 in the VNI population, in red, appears slower than in the VNB population, in blue. Recombination is therefore acting in VNB isolates to break up LD regions.

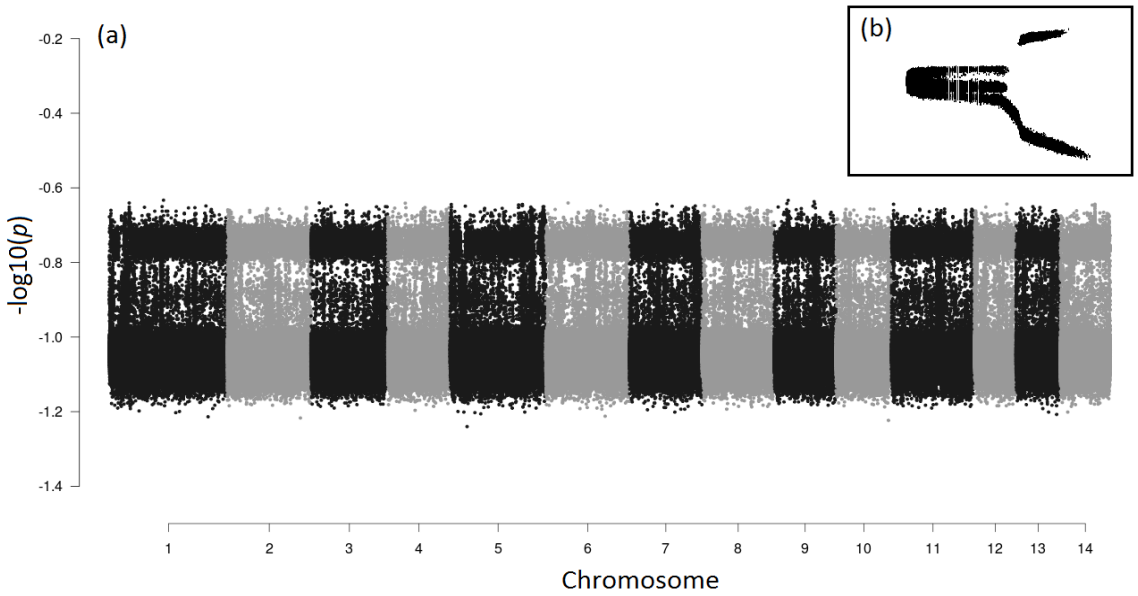


Figure 4.8: Bayescan output comparing VNB and VNI molecular types - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}).

Gene Content

BLAT score ratio (BSR) analyses allows to determine the presence/absence of genes. After filtering 50% of the *is*-cDNA sequences based on the sequences variance, 34,789 sequences were clustered using Pearson Correlation. A *t*-test with adjusted Bonferroni correction between environmental VNI and VNB isolates identified 8,174 significant sequences. *Is*-cDNA sequences were then blasted and 14 genes were significantly associated with the VNB lineage for one gene associated with the VNI clade (CNAG_06650), encoding for a hypothetical protein (Table 4.4). Figure 4.9 is a heat map representing the presence/absence of genes in each molecular type based on sequence similarity. No virulence-associated genes could be identified, although the *fcy3* gene (CNAG_04982) was identified in the VNI clade. This gene was previously found in a transcriptome analysis investigating flucytosine susceptibility in *C. neoformans* (Song et al., 2012). The cytosine deaminase genes are thought to be involved in flucytosine uptake. The highest BSR score in the VNI molecular type was CNAG_07575 (especially within the VNIb clade) which encodes for a hypothetical protein.

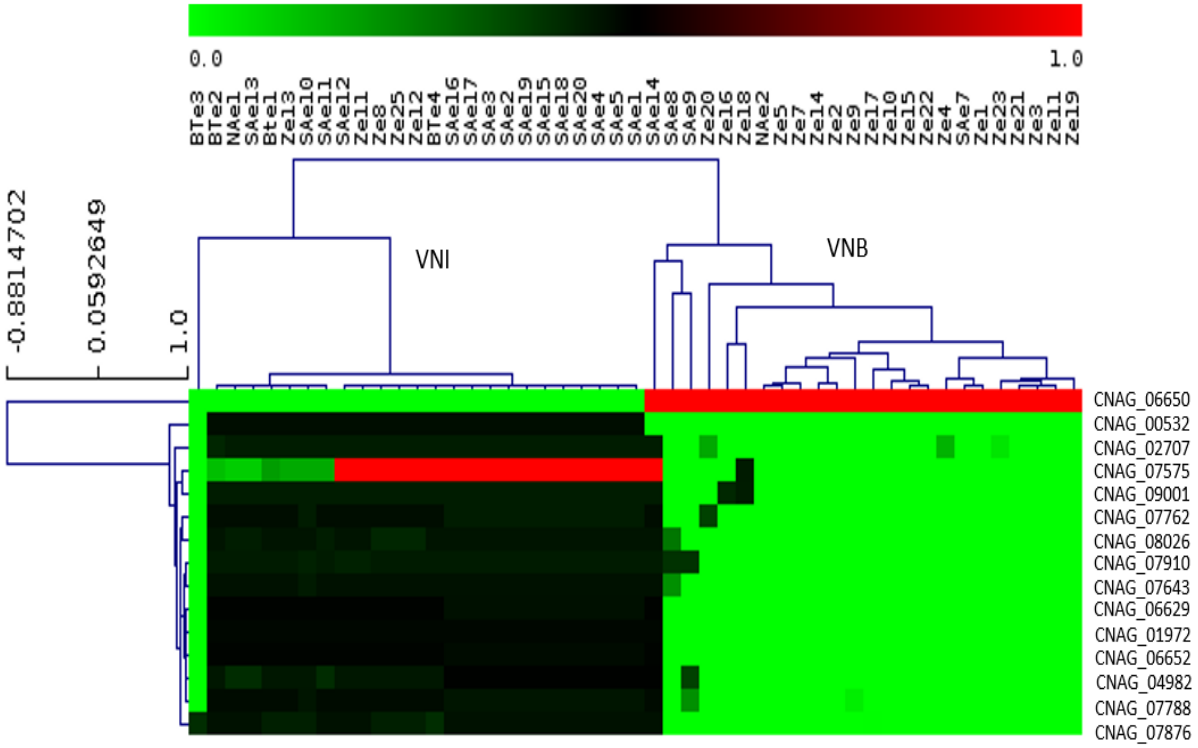


Figure 4.9: Population-level gene content analysis of environmental isolates (n=50) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two molecular types. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0).

Gene	Name	Annotation
CNAG_06650	CNAG_06650	hypothetical protein
CNAG_00532	CNAG_00532	hypothetical protein
CNAG_02707	CNAG_02707	hypothetical protein
CNAG_07575	CNAG_07575	hypothetical protein
CNAG_09001	CNAG_09001	unspecified product
CNAG_07762	CNAG_07762	hypothetical protein
CNAG_08026	CNAG_08026	hypothetical protein
CNAG_07910	CNAG_07910	hypothetical protein
CNAG_07643	CNAG_07643	hypothetical protein
CNAG_06629	CNAG_06629	hypothetical protein
CNAG_01972	TAF10	C2H2 zinc finger protein
CNAG_06652	CNAG_06652	allantoate permease
CNAG_04982	FCY3	putative cytosine-purine permease
CNAG_07788	CNAG_07788	hypothetical protein
CNAG_07876	CNAG_07876	hypothetical protein

Table 4.4: Gene annotation for gene identfied using Gene content analysis

Identifying Genes undergoing Positive Selection

Selection was investigated using two different approaches. Frequency-based methods look for a selected allele and its 'hitchhiker'region. Under positive selection, a beneficial allele acquires a high prevalence rapidly in a population creating a reduction of diversity in the surrounding genomic region (Vitti et al., 2013). Gene based methods allow to search for lineage-specific accelerations of selection which can indicate an excess of amino acid substitutions (Chamary et al., 2006).

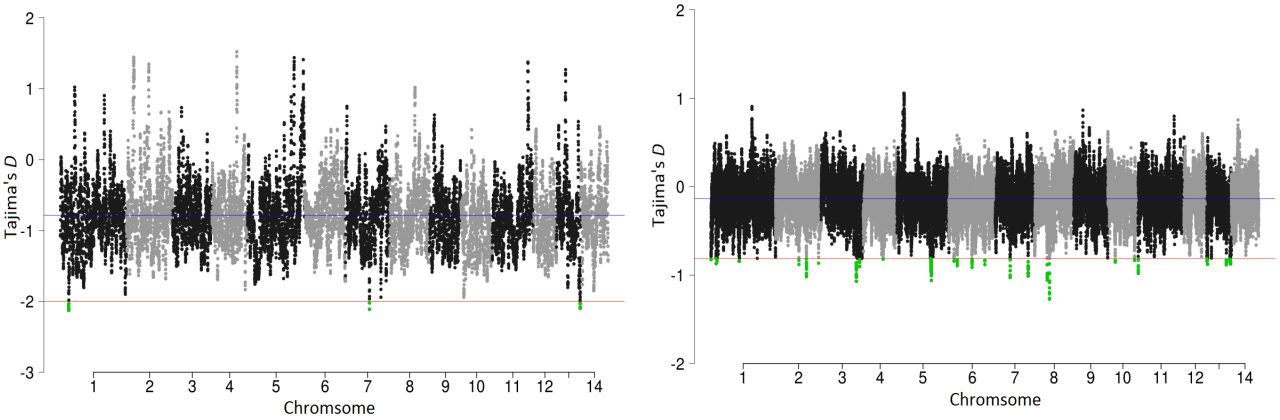
Frequency-based methods To identify candidate loci under selection, a sliding window approach was conducted using two population genetic statistics: Tajima's D (Figure 4.10) and D'(Figure 4.11). For each parameter, the empirical outliers within the 0.1% quantile were considered for the analysis. Little overlap was observed between the significant regions identified by the two statistical parameters.

Within the VNI molecular type, no overlap was observed between the two parameters. Tajima's D statistic identified three genes, none of which possessed a known virulence-associated function. Lewontin's D'statistic allowed the detection of 64 genes under directional selection, with seven virulence-associated genes (Table 4.5). These putative virulence genes included three myol-inositol transporters which are thought to be involved in the utilisation of myo-inositol

as a carbon source in the human central nervous system and are key components of *Cng* pathogenicity (Xue et al., 2010).

In the VNB clade, Tajima's *D* identified 40 genes with significant negative values and D'identified 158 genes. A total of nine genes overlapped (Table S9) and can therefore be considered as strong gene candidates undergoing strong directional selection. These genes were located on chromosome five and contained the virulence-associated gene CNAG_01106. The latter was reported by Erickson et al., (2001) and encodes for the vacuolar-ATPase subunit (VPH1). Tajima's *D* analysis identified one additional virulence genes: the capsular gene *cap59*. D'parameter identified seven virulence genes listed on Table 4.6.

Between the two molecular types, seven genes under selection overlapped including the virulence gene CNAG_06813. The MAT α specific gene encodes for the O-glucosyltransferase and plays a role in virulence by promoting haploid fruiting (Lengeler et al., 2002).

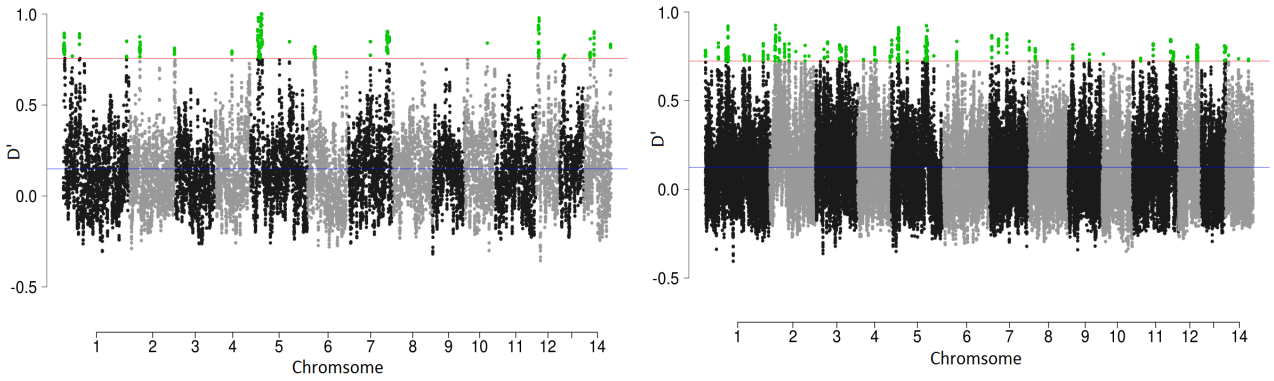


(a) VNI isolates (n=26)

(b) VNB isolates (n=24)

Figure 4.10: Variation of Tajima's *D* across the environmental *Cng* (Vilella et al., 2005) - Tajima's *D* values were calculated over a 100 bp non-overlapping sliding window. Negative value in the 0.1% quantile are highlighted in green.

Gene-based methods In order to investigate the signature of positive selection in the two molecular clades, we applied two gene-based methods; MK and dN/dS (ω) tests (Table 4.7). On average, 6,693 genes per VNB isolate were considered in the dN/dS analysis and 6,538 for the MK tests (see Methods 4.2.10). A total of 418 genes had high MK values including 27 virulence-associated genes (Table 4.8). Of the 6,472 genes considered in the dN/dS analysis, 240 were present in at least 40% of the VNB isolates and had ω value in the 0.5% quantile. Nine



(a) VNI isolates (n=26)

(b) VNB isolates (n=24)

Figure 4.11: Variation of LD across the environmental *Cng* (Vilella et al., 2005) - D'was calculated over a 100 bp non-overlapping window. Loci in the 0.1% quantile are highlighted in green.

Gene	Chr.	Start	End	Product	Literature
Tajima's *D*					
CNAG_00721	1	1,878,927	1,881,469	alpha-1,3-mannosyltransferase (*CAP59*)	(Chang and Kwon-Chung, 1994)
CNAG_01106	5	1,193,915	1,197,594	vacuolar-ATPase subunit (*VPH1*)	(Erickson et al., 2001)
D'					
CNAG_00306	1	787,724	789,157	hypothetical protein	(Ding et al., 2013)
CNAG_00986	5	1,502,869	1,504,532	urm1 activating enzyme (UBA4)	(Liu et al., 2008)
CNAG_01103	5	1,203,225	1,204,453	splicing factor 3A subunit 2	(Erickson et al., 2001)
CNAG_01106	5	1,193,915	1,197,594	vacuolar-ATPase subunit (VPH1)	(Erickson et al., 2001)
CNAG_02409	6	396,236	398,026	protein of unknown function (LIV5)	(Liu et al., 2008)
CNAG_02702	3	964,483	968,105	putative voltage-gated chloride channel (CLC1)	(Li et al., 2012)
CNAG_03637	2	402,338	405,710	ATP-dependent DNA helicase (YKU80)	(Liu et al., 2008)
CNAG_03760	2	727,412	730,668	Rad50-interacting protein 1 homolog (RINT1)	(Liu et al., 2008)
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)	(Lengeler et al., 2002)

Table 4.5: Significant Virulence-associated Genes in the 0.1% quantile in VNB isolates (n=24)

Gene	Chr.	Start	End	Product	Literature
D'					
CNAG_00097	1	268,618	271,508	myo-inositol transporter (ITR1)	(Xue et al., 2010)
CNAG_04093	2	1,601,566	1,604,626	putative transcription factor (YRM103)	(Liu et al., 2008)
CNAG_05377	14	137,967	140,909	myo-inositol transporter (ITR3)	(Xue et al., 2010)
CNAG_05662	14	891,933	894,463	myo-inositol transporter (ITR4)	(Xue et al., 2010)
CNAG_05939	7	1,266,326	1,268,709	hypothetical protein	
CNAG_06301	12	133,480	137,052	agc family protein kinase (SCH9)	(Ko et al., 2009)
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)	(Lengeler et al., 2002)

Table 4.6: Significant Virulence-associated Genes in the 0.1% quantile in VNI isolates (n=26)

of these genes had known virulence functions. Additionally, 34 genes were detected using both tests and therefore appeared as strong gene candidates. None of them had a known virulence function.

In the VNI clade, 1,945 and 3,606 genes were considered for the dN/dS and MK test analysis respectively. Using the MK test, 39 genes were under selection including two virulence genes (Table 4.8). In the dN/dS analysis, 12 genes were found but none was found in the literature to be virulent.

Table 4.7: Direction of Selection in VNB and VNI isolates

	MK tests		ω	
Clades	Mean	Genes	Mean	Genes
VNB	0.021	6,535	0.746	6,693
VNBa	0.019	6,533	0.754	6,693
VNBb	0.022	6,546	0.734	6,698
VNI	-0.034	1,945	1.187	3,606
VNIa	-0.040	713	1.174	1,716
VNIb	0.031	2,790	1.192	4,916

Table 4.8: Virulence-associtaed genes under selection in the VNB and VNI clade

VNB	value	Chr	Start	End	Product	Reference
MK (n=27)	0.023					
CNAG_00654	0.262	1	1,705,106	1,706,138	sulfiredoxin (*SRX1*)	(Ko et al., 2009)
CNAG_00697	0.317	1	1,817,479	1,819,731	UDP-glucose epimerase (*UGE1*)	(Liu et al., 2008)
CNAG_00795	0.217	1	2,097,928	2,099,353	hyphal flocculin (*CFL1*)	(Wang et al., 2012)
CNAG_03772	0.343	2	772,402	775,011	high-affinity glucose transporter (*HXS1*)	(Liu et al., 2013)
CNAG_04090	0.387	2	1,592,779	1,596,069	activating transcription factor (ATF1)	(Missall and Lodge, 2005)
CNAG_05081	0.376	4	391,625	393,846	phosphodiesterase (*PDE1*)	(Hicks et al., 2005)
CNAG_05015	0.360	4	229,871	232,836	catalase 4 (*CAT4*)	(Giles et al., 2006)
CNAG_01242	0.372	5	829,002	831,879	CCAAT -binding transcription factor (*HAPX*)	(Jung et al., 2010)
CNAG_00867	0.447	5	1,782,123	1,784,803	myo-inositol transporter (*ITR3A*)	(Xue et al., 2010).
CNAG_07408	0.380	5	230,211	233,167	ste/ste20/paka protein kinase (STE20alpha	(Wang et al., 2002).
CNAG_02315	0.462	6	631,777	633,561	ubiquinol-cytochrome c reductase, iron-sulfur subunit	
CNAG_02812	0.409	6	714,279	716,015	ornithine carbamoyltransferase (ARG3)	(Idnurm et al., 2009)
CNAG_02049	0.395	6	1,331,959	1,334,268	proline oxidase (*PUT1*)	(Lee et al., 2013)
CNAG_05882	0.333	7	1,105,565	1,108,397	class e vacuolar protein-sorting machinery protein (*HSE102*)	(Liu et al., 2008)
CNAG_03223	0.459	8	383,400	384,978	hypothetical protein	
CNAG_04585	0.350	10	963,306	964,931	hypothetical protein	
CNAG_01681	0.333	11	591,464	594,607	cytosine permease (*FCY2*)	(Song et al., 2012)
CNAG_01690	0.276	11	618,826	622,217	MFS transporter	
CNAG_01872	0.339	11	1,144,865	1,146,031	copper transporter	
CNAG_06064	0.334	12	207,653	208,840	protein tyrosine phosphatase (*PTP1*)	(Lee et al., 2014)
ω (n=9)	0.852					
CNAG_00306	1.240	1	787,724	789,157	hypothetical protein	(Ding et al., 2013)
CNAG_01019	1.250	5	1,416,547	1,417,779	copper/zinc superoxide dismutase (SOD1)	(Narasipura et al., 2003).
CNAG_01251	0.413	5	797,058	804,089	e3 ubiquitin-protein ligase (*UFD4*)	(Liu et al., 2008)
CNAG_01611	1.827	11	412,788	415,621	infection related protein of unknown function (*LIV8*)	(Liu et al., 2008)
CNAG_02802	1.148	3	735,903	737,905	inositol/phosphatidylinositol kinase	
CNAG_03932	0.501	2	1,213,102	1,216,148	cell cycle checkpoint protein (RAD17)	(Idnurm et al., 2009)
CNAG_04714	0.500	10	608,465	611,057	MFS multidrug transporter protein	
CNAG_06031	2.625	12	120,792	123,948	putative beta-glucan synthase (*KRE63*)	(Gilbert et al., 2010b)
CNAG_07724	0.956	8	721,568	725,406	copper-sensing transcription factor (CUF1)	
VNI	value	Chr	Start	End	Product	Ref
MK (n=27)	-0.034					
CNAG_01841		11	1,052,774	1,054,169	gata family transcription factor (*GLN3*)	(Liu et al., 2008)
CNAG_07422		5	466,747	468,297	delayed-type hypersensitivity antigen (*DHA1*)	(Mandel et al., 2000)

Overlap of Genes using the frequency-based and gene-based parameters

To consider a gene as a strong candidate undergoing positive selection, the putative gene needs to be identified with a minimum of two statistical parameters.

In the VNB lineage, nine virulence-associated genes were found using frequency-based methods, including the virulent gene CNAG_01106 which was identified with both Tajima's D and D'. Using gene-based methods, 34 genes were identified with both the MK test and dN/dS, but none of them had a known virulence function. An additional 11 genes were identified when considering frequency-based parameters and gene-based parameters together including CNAG_00306 (Table S11). The latter encodes for metallothionein protein (CMT2) which plays a role in the response to elevated level of Cu (Ding et al., 2013).

No overlap was observed in the VNI clade. However, when comparing genes under selection in both lineages, 12 genes were found in each linage. Of these 12 genes showing signs of positive selection, two had known virulence functions: CNAG_04093 encoding for a putative transcription factor (Liu et al., 2008) and CNAG_06813 a MAT α specific gene described previously. Figure 4.12 and 4.13 summarise the population genomics analysis for each molecular type, highlighting

virulence genes undergoing selection.

4.3.4 Intra-specific variations in the VNB clade

Phylogeny and Population Structure

As previously described, two sub-populations could be detected in the VNB clade (Figure 4.14). The fineStructure analysis revealed that the two populations are sharing important amounts of genetic material (Figure 4.15). Under the linked model (Figure 4.15b), Z20 (VNBa) shared significant numbers of genomic regions with the VNBb clade. The two populations overlapped in southern Zambia (Figure 2.4) indicating that they are not strictly geographically separated. Additionally, the F_{ST} value (F_{ST}=0.143) was the lowest compared to other *Cng* sub-populations.

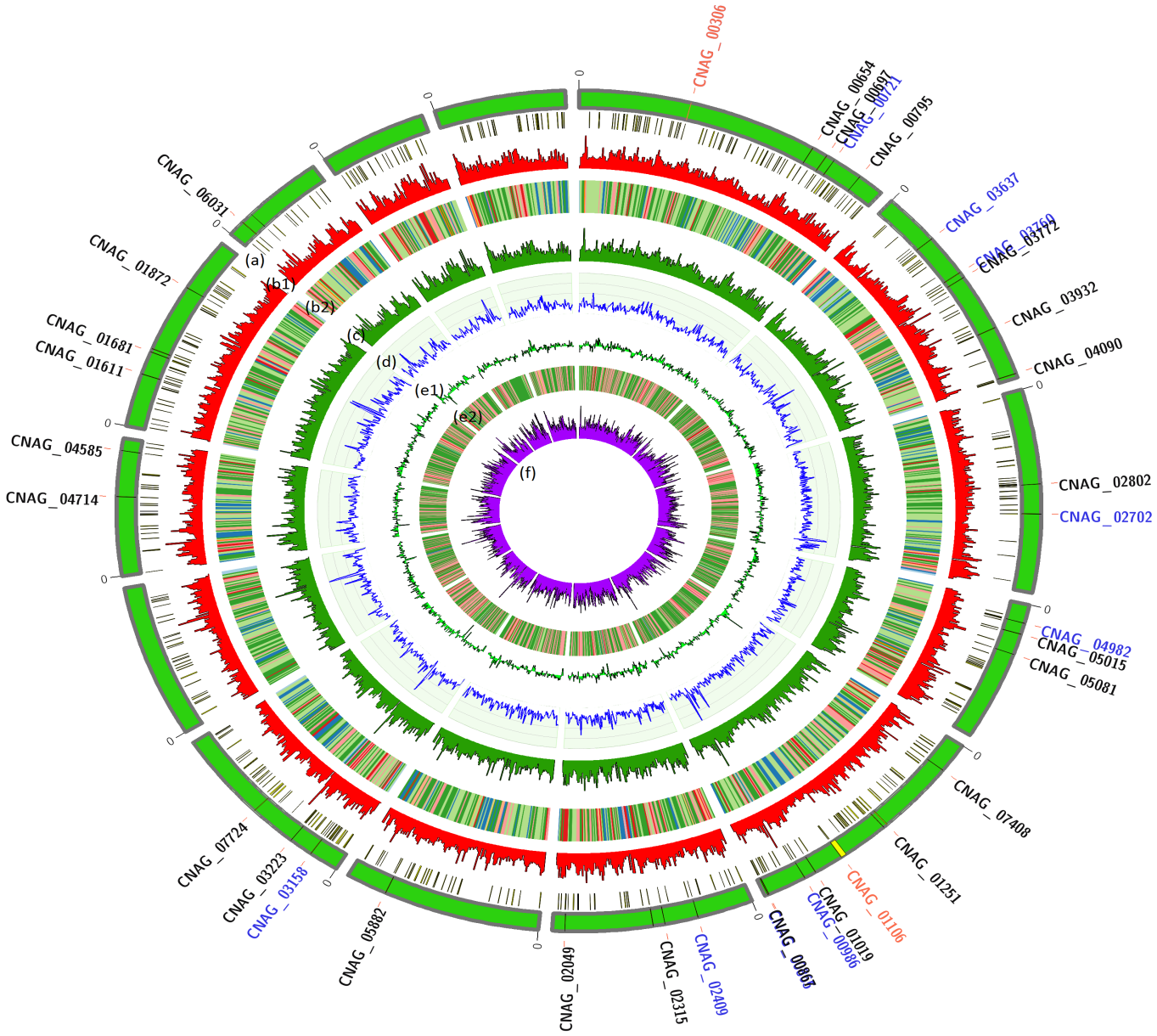


Figure 4.12: Population Genetics analyses of environmental VNB isolates across the genome. In each chromosome, the value of each statistic was averaged over a non-overlapping 10-Kb window - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n = 470) ; (b) SNPs density histogram and heatmap ; (c) Histogram of F_{ST} statistics ; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; (f) histogram of D'; The position of significant virulence genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. Genes highlighted blue are virulence genes under positive selection identified using frequency-based methods (Tajima's D and D'), genes identified using gene-based methods are in black; Strong gene candidates (identified using two statistical parameters) are in red. The band in yellow on chromosome five (containing the virulence-associated CNAG_01106) is a region which was detected under strong selection forces using frequency-based methods.

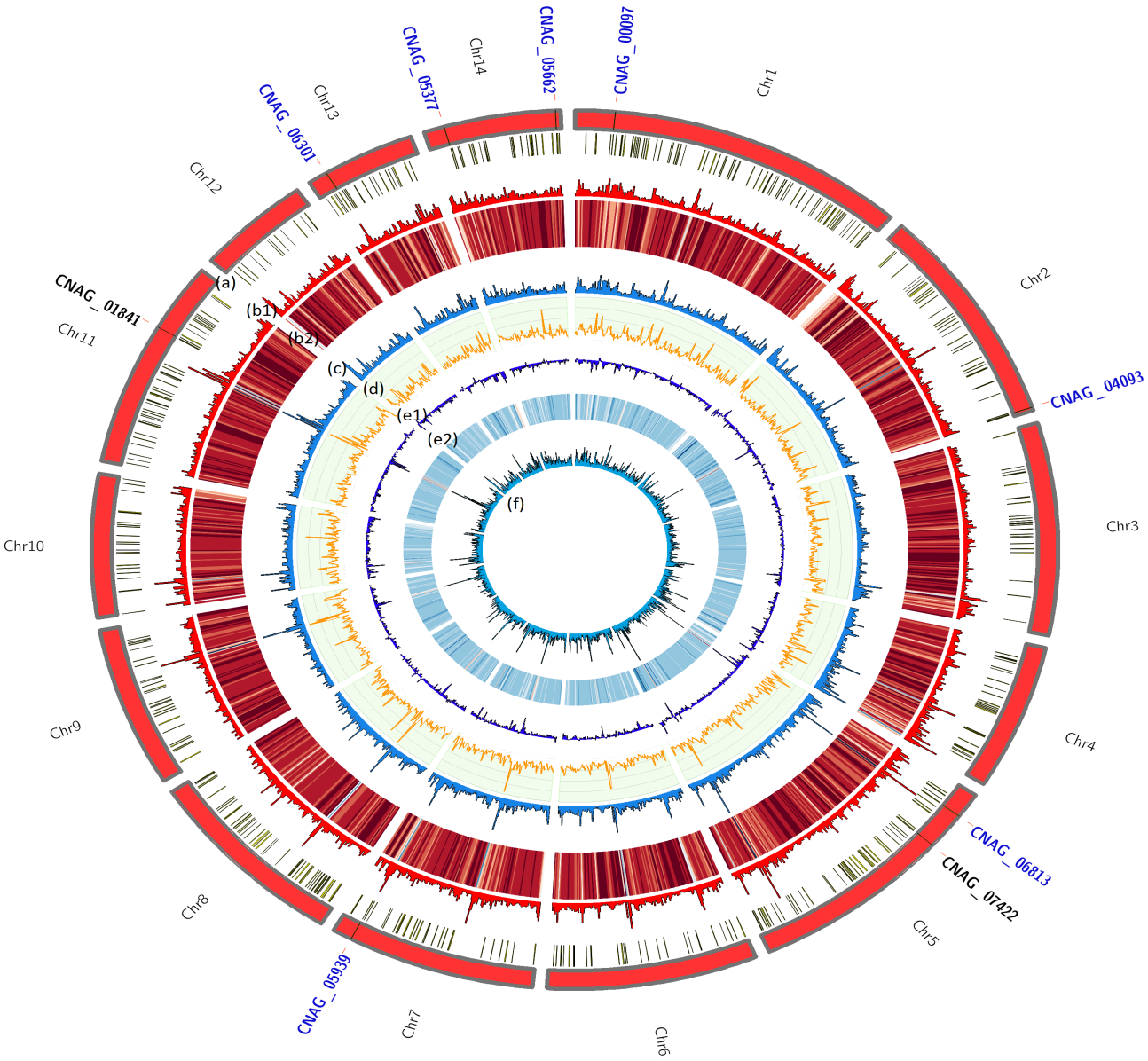


Figure 4.13: Population Genetics analyses of environmental VNI isolates across the genome. In each chromosome, the value of each statistic was averaged over a non-overlapping 10-Kb window. - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n = 470) ;(b) SNPs density histogram and heatmap ; (c) Histogram of F_{ST} statistics ; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; (f) histogram of D'; The position of significant virulent genes (at the 0.1% quantile) is highlighted on chromosome ideograms. Genes highlighted blue are virulent genes under positive selection identified using frequency-based methods (Tajima's D and D'), genes identified using gene-based methods are in black; Strong gene candidates (identified using two parameters) are in red.

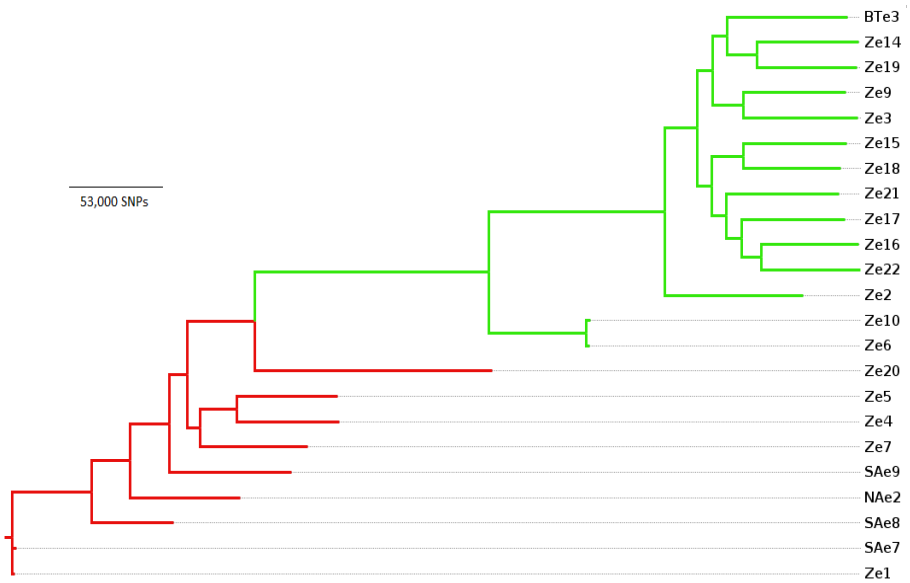
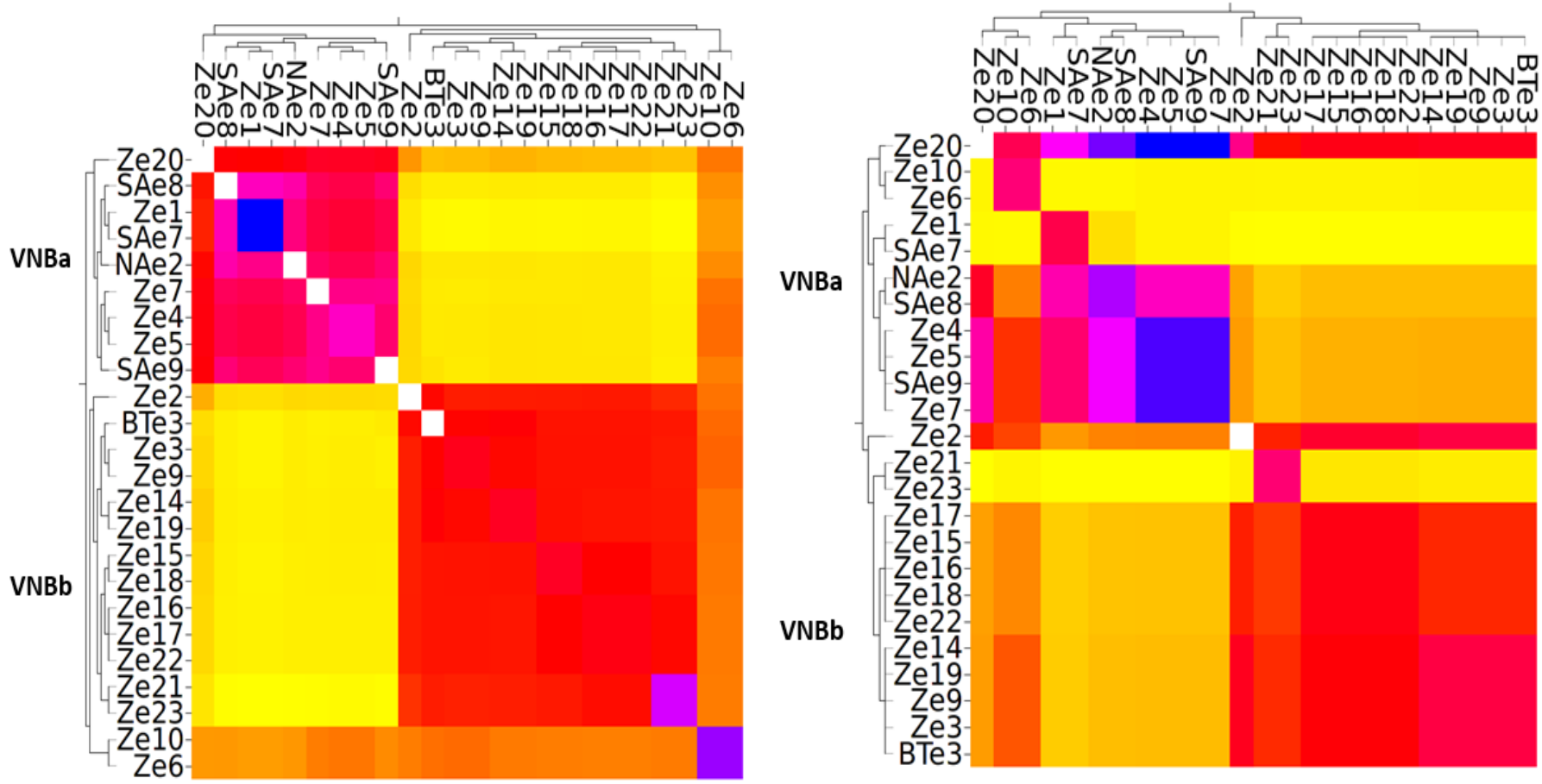


Figure 4.14: Phylogenic analysis of environmental *Cryptococcus neoformans* VNB isolates (n=24) - VNBa isolates are in red and VNBb isolates appear in green. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: Namibian isolate. Bootstrap values at each node were averaging 100%.



(a) FineStructure Analysis (Unlinked Model)

(b) FineStructure Analysis (Linked Model)

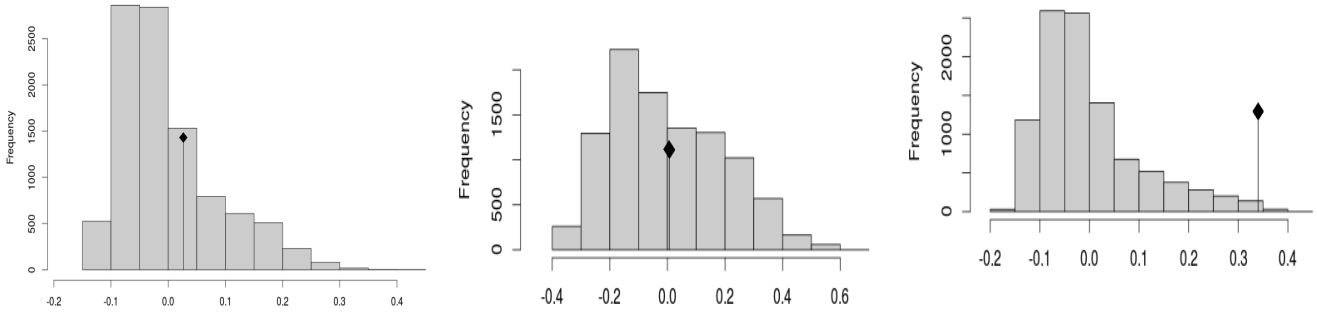
Figure 4.15: Population Structure of Environmental *Cryptococcus neoformans* (n=24) - (a) The VNB clade is composed of two sub-populations, VNBa and VNBb. Two isolates Ze6 and Ze10 from the VNBb sub-population are sharing some genetic content with the VNBa population indicating that they might represent hybrids. (b) When taking linkage disequilibrum into account, the two VNB populations can still be observed. Using the linked model, some isolates are not fully included into any population (Ze10, Ze6, Ze1, Ze7, Ze21 and Ze23) . Two isolates, Ze20 and Ze2, are sharing genetic materials with both sub-populations. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow.

Effect of distance and barriers on genetic divergence

Mantel Test The role of Mantel tests is to show signs of the isolation by distance by assessing the potential correlation between genetic distance and geography. When considering all VNB isolates (Mantel r statistic = 0.0263, p = 0.2890), no signal of isolation by distance could be identified. However, significant isolation by distance was detected in the VNBb clade (Mantel r statistic = 0.340, p = 0.005) (Figure 2.4).

When examining the role of environmental variables on the relation between genetic and geographic distance using partial mantel tests, climatic variables were only observed to have significant impact in the VNBb clade. Within this lineage, 17 variables were significant and four of them improved the r statistic from the initial mantel tests (Table 4.10). These findings highlight the role of precipitation on the distribution of the VNBb lineage as the most significant climatic variables were: Precipitation Seasonality (Bio15), Precipitation of Wettest Quarter (Bio16) and Precipitation of Warmest Quarter (Bio18). The Normalised Difference Vegetation Index (NDVI) was also significantly associated with VNBb population, indication that

Climate seems to play a predominant role in the distribution of VNB isolates. Redundancy analyses (Table 4.11) found that climatic variables explained 75% of the genetic diversity variance. Within the VNB sub-populations, RDA found that 60% of the genetic variations was redundant with climatic variables (explanatory variables).



(a) VNB (n=24)) (b) VNBa (n=9) (c) VNBb (n=15)

Figure 4.16: Mantel Tests for the VNB Molecular Type (n=24) using 10,000 replications.

Table 4.10: Partial Mantel Tests in VNB population

Population	VNB		VNBa		VNBb	
Variables	r	*p*	r	*p*	r	*p*
Mantel test	0.059	0.167	0.228	0.153	0.340	0.005**
Elevation	0.057	0.157	0.226	0.128	0.296	0.042*
Longitude	0.115	0.070	-0.071	0.618	-0.033	0.396
Latitude	-0.072	0.908	0.383	0.153	0.248	0.037*
NDVI	0.078	0.169	0.645	0.279	0.366	0.001***
EVI	0.080	0.144	0.034	0.368	0.308	0.004**
Bio1	0.043	0.232	0.177	0.238	0.326	0.012*
Bio2	0.024	0.289	-0.126	0.731	0.221	0.102
Bio3	-0.069	0.869	0.347	0.063	0.042	0.377
Bio4	-0.130	0.988	0.119	0.157	0.144	0.066
Bio5	0.064	0.157	0.230	0.150	0.284	0.029*
Bio6	-0.014	0.491	-0.055	0.574	0.266	0.074
Bio7	-0.042	0.694	-0.018	0.454	0.042	0.233
Bio8	0.081	0.120	0.220	0.148	0.340	0.006**
Bio9	-0.056	0.840	0.165	0.181	0.280	0.050*
Bio10	0.066	0.146	0.229	0.147	0.339	0.005**
Bio11	-0.015	0.522	0.196	0.205	0.280	0.041*
Bio12	0.043	0.232	-0.220	0.858	0.321	0.001***
Bio13	0.035	0.277	-0.372	0.909	0.336	0.001***
Bio14	-0.029	0.620	0.412	0.117	0.000	NA
Bio15	0.046	0.152	0.685	0.090	0.371	0.001***
Bio16	0.042	0.240	-0.313	0.862	0.345	0.001***
Bio17	-0.011	0.506	0.393	0.148	0.338	0.008**
Bio18	0.068	0.142	0.280	0.048	0.345	0.001***
Bio19	0.031	0.252	0.432	0.125	0.338	0.002**

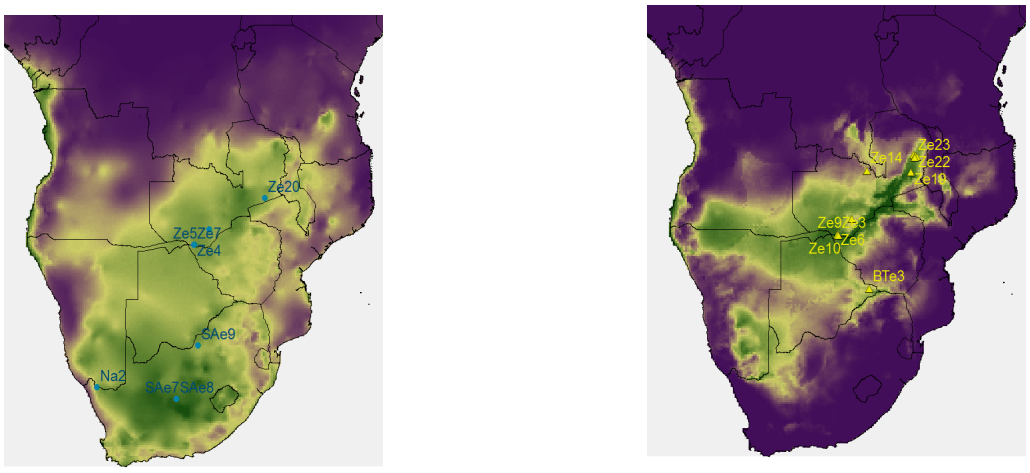
$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

Table 4.11: RDA analysis in VNB populations

	VNB		VNBa		VNBb	
RDA	Inertia	Proportion	Inertia	Proportion	Inertia	Proportion
Total explainable	287,004	1	172,143	1	186,624	1.000
Pure climate	216,691	0.755	118,896	0.691	123,343	0.661
Pure geography	35,688	0.124	0	0.000	11,431	0.061
Joint climate/geography	34,625	0.121	53,247	0.309	51,850	0.278

Environmental niche modelling The environmental niche model of the two VNB sub-populations showed that the predicted geographical occurrence of VNBa clade is larger than the VNBb clade. The VNB lineage seems to be restricted to the mopane tree belt. Isolates of this sub-population were recovered in 92.9% of cases on *C. mopane*.

The VNBa population was also present in South Africa. This finding is the first report of VNB *Cng* outside the mopane tree belt, indicating that the molecular type is present in other regions and potentially on other continents. The VNBb lineage seem to be adapted to the Zambezian mopane woodlands. The niche model matched perfectly this ecoregion, which has a low altitude and clay-rich soils. The MaxEnt model was mostly influenced by Isothermality (bio3) accounting for 98.8% in the VNBa model and 29.5% in the VNBb model. Other climatic variables which seemed to influence the VNBb distribution were Temperature Annual Range (bio07, 20.1%) and Precipitation of Warmest Quarter (bio18, 4.8%).

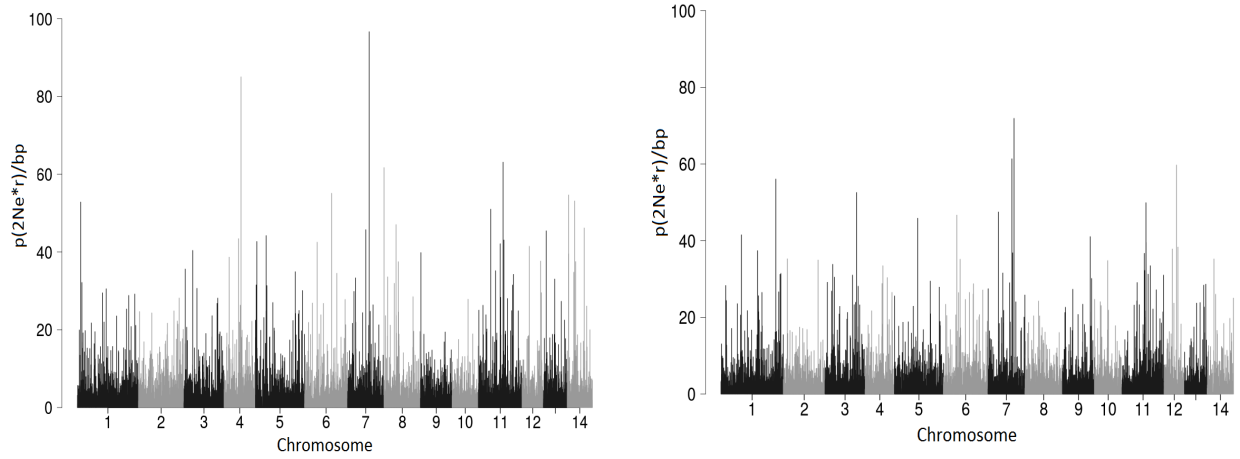


(a) VNBa sub-populations

(b) VNBb sub-population

Figure 4.17: Predicted distribution of the two VNB sub-population - Environmental Niche Model Distribution obtained with Maxent using WorldClim environmental data.

Landscape Genomics Latent Factor Mixed Models (LFMM) analyses revealed no significant locus after accounting for multiple testing. However, due the high number of SNPs in the VNB clade (SNPs = 578,126), loci were only considered as significant when their p value reached $8.649.10^{-8}$. When plotting the LFMM results, no loci were significant but clear correlation could be observed throughout the genome. Figure S9 portrays an example of outlier loci correlated the Max Temperature of Warmest Month (Bio05). When investigating Temperature related variables (Bio01 to Bio11), 65 genes had relatively high p value and of these four genes



(a) VNBa isolates (n=9)

(b) VNBb isolates (n=15)

Figure 4.18: Recombination analysis, using LDhat interval for the VNB sub-populations - The peaks represents the recombination hotspots across the genome.

had known virulence functions (Table 4.12).

Table 4.12: Virulence-associated genes correlated with Environmental Variables in VNB clade

Gene	BioClim	Chr	Start	End	Product	Ref
CNAG_00865	Bio05	5	1,790,294	1,791,427	putative O-acetyltransferase (CAS92)	Moyrand et al. (2007)
CNAG_05395	Bio04	14	198,857	202,798	rab guanyl-nucleotide exchange factor (VAM6)	
CNAG_05934	Bio03	7	1,253,290	1,255,063	translin domain protein (LIV15)	Li et al. (2008)
CNAG_05939	Bio02	7	1,266,326	1,268,709	hypothetical protein	

Population Genomic Analyses

Population Genomics and Recombination Analysis As described previously, VNBa had significant negative Tajima's D value indicating recent population expansion or a recent selective sweep. It is worth noting that this population contained a limited number of isolates (n=9). Recombination analyses found that the VNBa sub-population had a higher recombination rate than the other VNB clade ($p_{\text{VNBa}} = 0.2932/\text{bp}^{-1}$; $p_{\text{VNIBb}} = 0.2512/\text{bp}^{-1}$).

On Figure 4.19, the two sub-populations seem to have similar LD decay profile. Nevertheless, the VNBb subtype displays higher level of linkage disequilibrium than the VNBa sub-population (D=0.338 and 0.112, respectively).

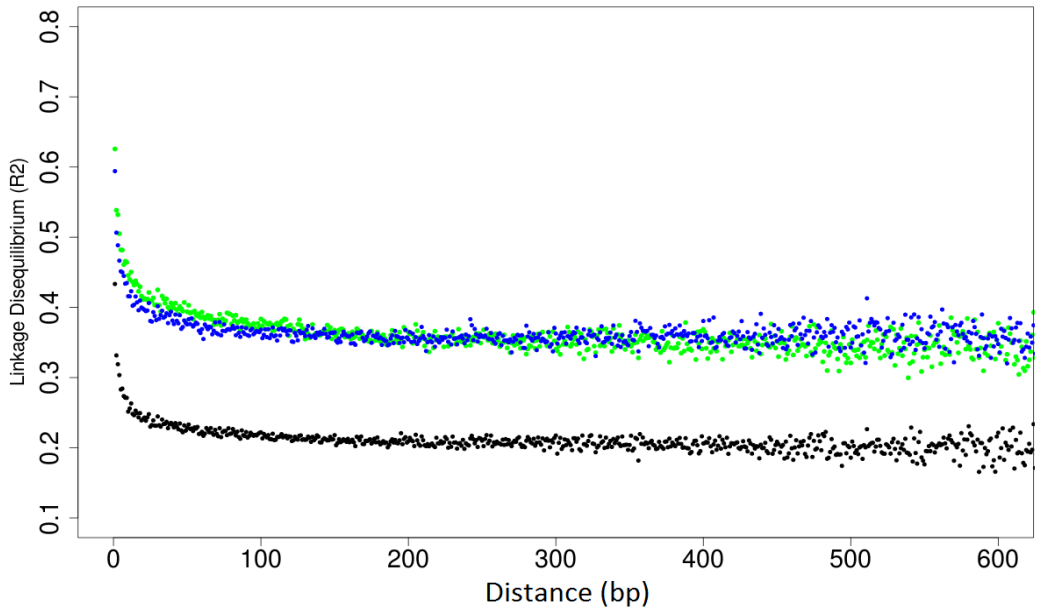


Figure 4.19: Population decay of Linkage Disequilibrium within the VNB isolates - The rate of decay of r^2 in the two VNB sub-populations are similar (VNBa is in green, and VNBb appears in blue), in black is the LD decay for the VNB population(n=24).

Bayescan and F_{ST} Outlier An important number of loci differed between the two sub-populations as shown on the Manhattan plot (Figure 4.20a). None of them were significant when applying a FDR of 0.05. This can be explained by the fact that despites the 228,688 SNPs that the two populations possess in common: the VNBa sub-population has 170,258 unique SNPs and the VNBb isolates have 188,273 unique SNPs (Table S8). However, the volcano plot indicates signs of both diversifying and purifying/balancing selection. A total of 20,529 loci (3.51% of SNPs) had high F_{ST} values indicating diversifying selection.

Gene Content Analysis After applying Bonferroni correction, 1,930 sequences (12%) were significant and the LS-BSR analyses identified 11 genes associated with VNBa and 16 genes associated with VNBb (Figure 4.21). In each clade, one of the genes found had known virulence function.

Among the genes with BSR value close to 1, CNAG_00914 was significantly associated with the VNBa clade (Figure 4.21). The gene encodes for the putative β-glucan synthase (KRE6) and is part of the KRE family which is involve in β-1,6-glucan synthesis. In the study by Gilbert et al., (2010a), H99 *kre6* mutants were avirulent in mice principally due to their inability to maintain cell wall integrity.

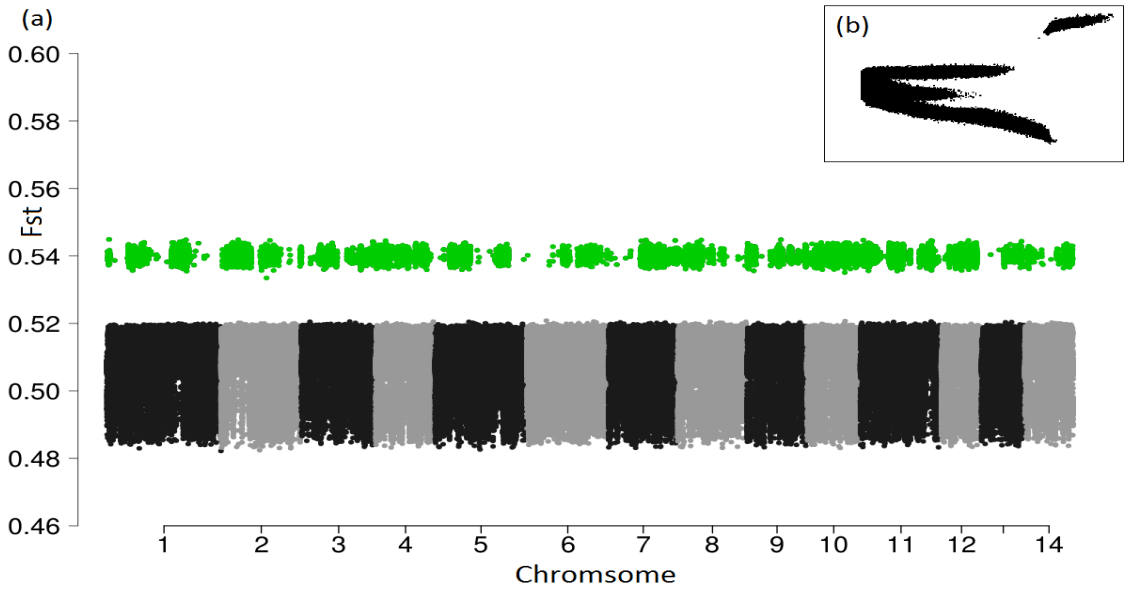


Figure 4.20: Bayescan Output comparing VNB sub-population - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the bayescan output showing signs (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}).

In VNBb, CNAG_01377 which encodes for pab-dependent poly(A)-specific ribonuclease subunit (UBP13) is located on chromosome five. In the context of virulence, Fang et al., (2012) found that *ubp13* mutants showed an increased susceptibility to caspofungin and fluconazole as well as hypomelinization at 37°C.

Population Genetics

Frequency-based methods The VNBa population had a significant negative overall Tajima's D value. When applying genome-wide scan frequency-based methods, 227 and 147 genes were found in the 0.1% quantile using Tajima's D and D'respectively (Figure 4.22). Eight virulence genes were found with Tajima's D and 12 virulence genes using D'(Table 4.14). A total of 19 genes overlapped between the two measures including the two virulent genes CNAG_04510 and CNAG_05377.

Ko et al., (2009) investigated the role of the HOG signalling pathway in response to stress. Using microarray, they found that the VSP4 ATPase (CNAG_04510) decreased by 1.5 fold in HOG signalling mutant compared to the wild type implying its role in cadmium sensitivity. CNAG_05377 encodes for myol-inositol transporter (ITR3) and was also identified in the

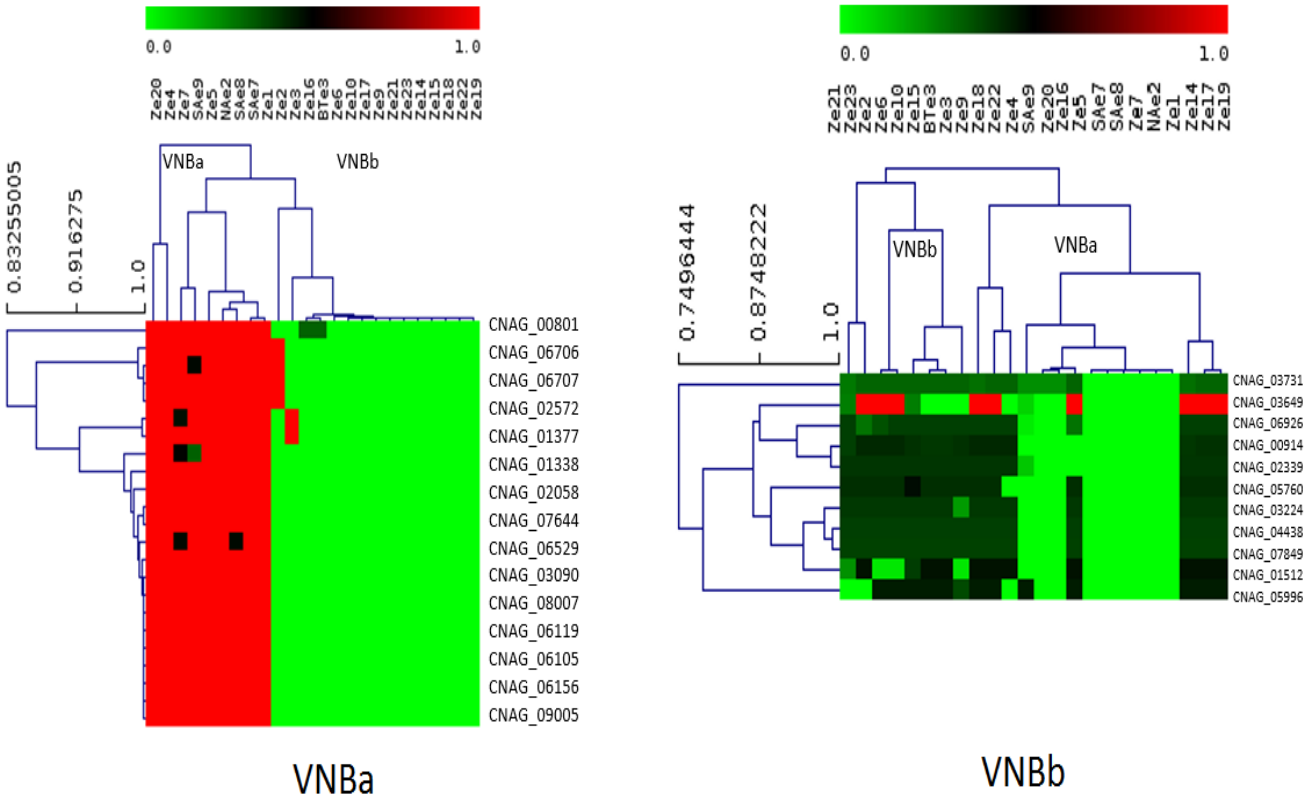
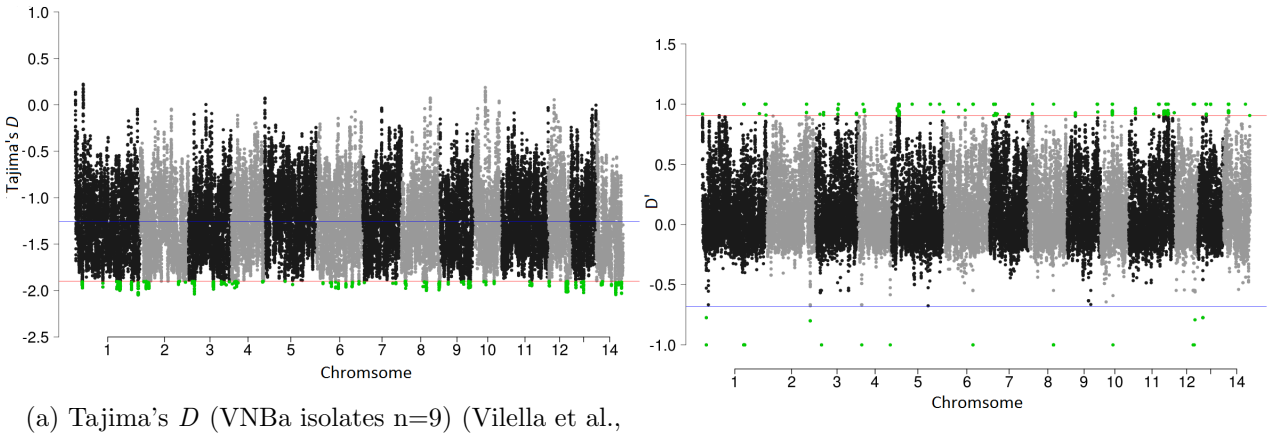


Figure 4.21: Population-level gene content analysis (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two VNB sub-populations. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0).

analysis of the VNI clade.

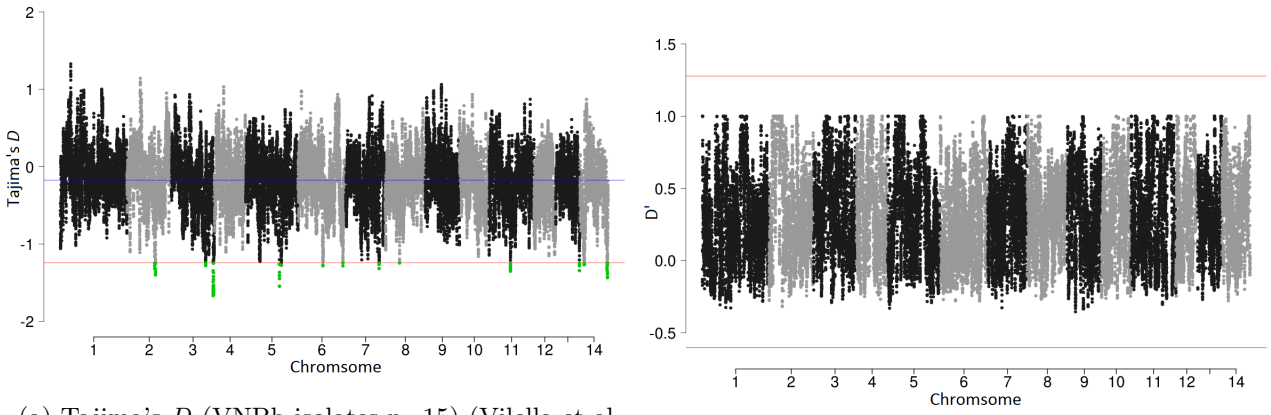
In the VNBb group, only 15 genes were identified using Tajima's *D*. Two of these genes had a known virulence function (Table 4.15) including CNAG_01106 identified in the VNB analysis (section 3.1.4.5.). Lewontin D'parameter was not able to identify any gene.



(a) Tajima's *D* (VNBa isolates n=9) (Vilella et al., 2005)

(b) LD'(VNBa isolates n=9) (Vilella et al., 2005)

Figure 4.22: Frequency-based Analysis in the VNBa clade. The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green.



(a) Tajima's *D* (VNBb isolates n=15) (Vilella et al., 2005)

(b) LD'(VNBb isolates n=15) (Vilella et al., 2005)

Figure 4.23: Frequency-based Analysis in the VNBb clade. The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green.

Gene-based methods A similar number of genes were used in the analysis, 6,693 and 6,698 for VNBa and VNBb respectively. The MK test identified more genes than the dN/dS analysis. In the VNBa groups, 280 genes had a MK value in the 0.5% quantile and were present in at

VNBa	Chr	Start	End	Product	Ref
Tajima's *D*					
CNAG_00097	1	268,618	271,508	myo-inositol transporter (ITR1)	Xue et al. (2010)
CNAG_03406	8	871,061	873,460	putative compass/set1c complex subunit (SPP101)	Liu et al. (2008)
CNAG_04505	9	1,035,860	1,039,071	guanine nucleotide-binding protein subunit alpha (GPA1)	Tolkacheva et al. (1994)
CNAG_03582	8	1,326,783	1,330,431	escrt-III-associated proteolytic activator of rim101 (RIM20)	O'Meara et al. (2010b)
CNAG_03590	8	1,353,408	1,354,983	casein kinase II subunit beta (1576 nt)	Lee et al. (2010)
CNAG_07751	8	1,359,863	1,362,381	siderophore iron transporter MirB	
CNAG_04510	9	1,049,586	1,051,433	AAA-ATPase (VPS4)	Ko et al. (2009)
CNAG_06487	13	645,025	649,238	putative chitin synthase (CHS6)	Banks et al. (2005)
CNAG_05377	14	137,967	140,909	MFS transporter, SP family, solute carrier family 2 (myo-inositol transporter), member 13	Xue et al. (2010)
D'					
CNAG_00772	1	2,027,692	2,029,517	putative UV excision repair protein (RAD23)	Liu et al. (2008)
CNAG_05015	4	229,871	232,836	catalase 4 (CAT4)	Giles et al. (2006)
CNAG_00864	5	1,793,503	1,796,091	myo-inositol transporter (ITR2)	Xue et al. (2010)
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)	Lengeler et al. (2002)
CNAG_06814	5	185,700	188,859	alpha cell-type homeodomain transcription factor (SXI1alpha)	Hull et al. (2002)
CNAG_07408	5	230,211	233,167	ste/ste20/paka protein kinase (STE20alpha)	Wang et al. (2002)
CNAG_06591	7	210,634	216,288	histone deacetylase complex protein (SET302)	Liu et al. (2008)
CNAG_04510	9	1,049,586	1,051,433	AAA-ATPase (VPS4)	Ko et al. (2009)
CNAG_04514	9	1,058,664	1,060,691	mitogen-activated protein kinase (MPK1)	Wang et al. (2011b)
CNAG_01938	11	1,315,183	1,319,444	serine/threonine protein kinase (KIN1)	Liu et al. (2008)
CNAG_05377	14	137,967	140,909	myo-inositol transporter (ITR3)	Xue et al. (2010)
CNAG_05592	14	710,003	713,644	Fan et al. (2007)	

Table 4.14: Virulence-associated genes under selection in the VNBa clade

VNBb	Chr	Start	End	Product	Ref
Tajima's D					
CNAG_01106	5	1,193,915	1,197,594	vacuolar-ATPase subunit (VPH1)	Erickson et al. (2001)
CNAG_01740	11	763,556	765,645	septin (CDC12)	Kruzel et al. (2012)

Table 4.15: Virulence-associated genes under selection in the VNBb clade

least 40% of the isolates compared to 115 genes using the dN/dS analysis. Therefore, the MK test analysis identified 21 genes with a known virulence function compared to only three using the dN/dS methods (Table 4.14).

In the VNBb group, the gene-based methods were more successful to identified putative genes under selection. MK analysis found 287 genes and dN/dS found 153 genes. Virulence genes are listed on Table 4.15.

Overlap between genome-scan methods

Candidate genes are considered as genes undergoing positive selection if they can be identified using at least two parameters. Among the VNBa group, 42 genes were found using both gene-based and frequency methods. Four virulence genes were among the genes identified: CNAG_04510 and CNAG_05377 which are described above and CNAG_05015 and CNAG_07408. These genes appear in red on Figure 4.24a. When comparing the genes under selection in the VNBa clade to the genes identify in the whole VNB dataset, little overlap was found. Only two genes matched: CNAG_07015 and CNAG_06725 encoding for a hypothetical protein and a DNA polymerase delta subunit respectively.

In the VNBb group, four genes can be considered as strong candidate genes undergoing selection. None of them have a known virulence function. However, the virulence gene CNAG_01106 detected using Tajima's D statistic was also identified in the analysis of the VNB molecular type.

Table 4.16: Direction of Selection in VNB isolates

Statistics	VNB	VNBa	VNBb
MK Mean	0.024	0.024	0.024
Dn/Ds Mean	0.846	0.848	0.848
alpha mean	-0.140	-0.141	-0.141
Total Gene	6693	6693	6698

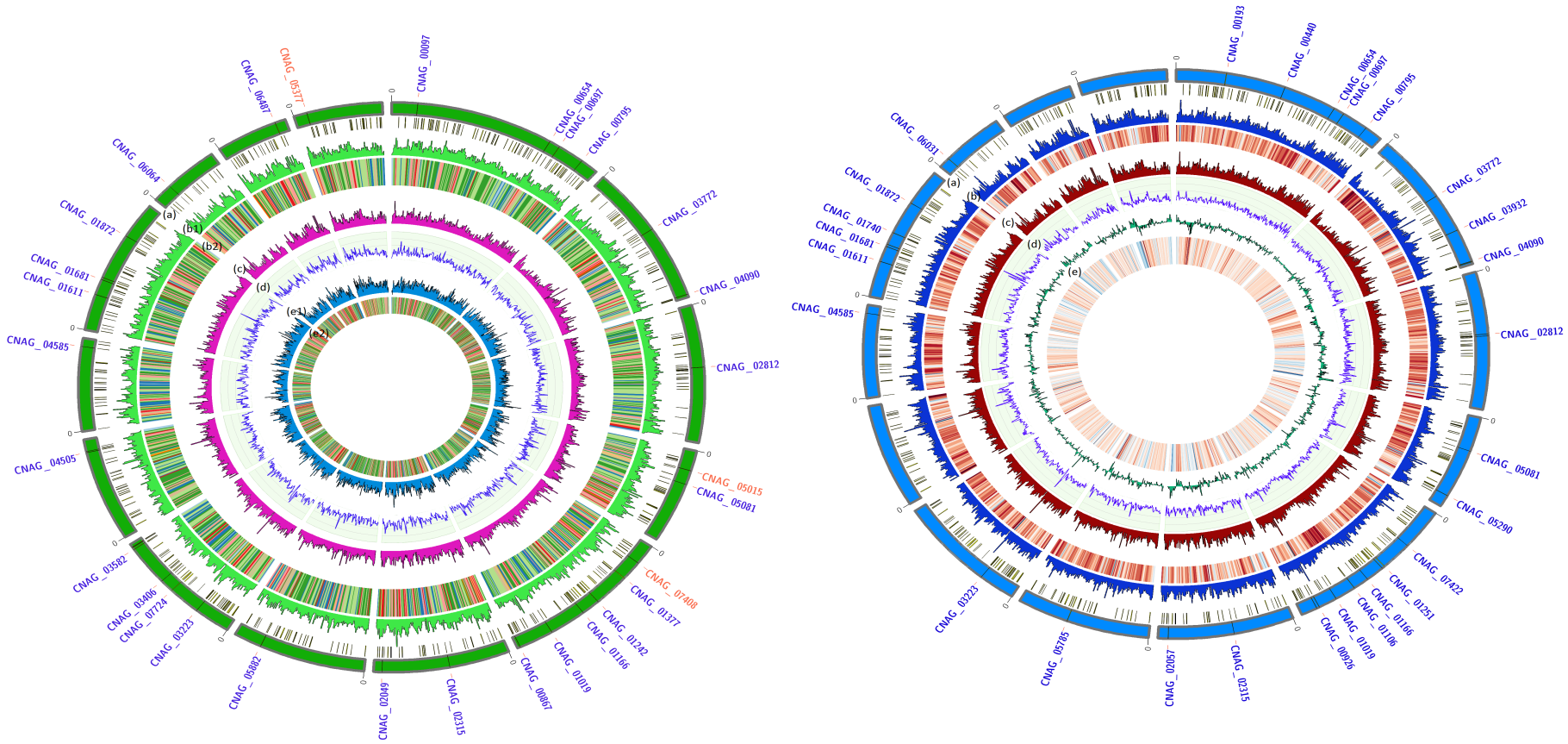
VNBa	value	Chr	Start	End	Product	Ref
MK (n=21)	0.023					
CNAG_00654	0.262	1	1,705,106	1,706,138	sulfiredoxin (SRX1)	Ko et al. (2009)
CNAG_00697	0.317	1	1,817,479	1,819,731	UDP-glucose epimerase (UGE1)	Liu et al. (2008)
CNAG_00795	0.217	1	2,097,928	2,099,353	hyphal flocculin (CFL1)	Wang et al. (2012)
CNAG_03772	0.343	2	772,402	775,011	high-affinity glucose transporter (HXS1)	Liu et al. (2013)
CNAG_04090	0.387	2	1,592,779	1,596,069	activating transcription factor (ATF1)	Missall and Lodge (2005)
CNAG_05081	0.376	4	391,625	393,846	phosphodiesterase (PDE1)	Hicks et al. (2005)
CNAG_05015	0.360	4	229,871	232,836	catalase 4 (CAT4)	Giles et al. (2006)
CNAG_01242	0.372	5	829,002	831,879	CCAAT -binding transcription factor (HAPX)	Jung et al. (2010)
CNAG_00867	0.447	5	1,782,123	1,784,803	myo-inositol transporter (ITR3A)	Xue et al. (2010).
CNAG_07408	0.380	5	230,211	233,167	ste/ste20/paka protein kinase (STE20alpha	Wang et al. (2002).
CNAG_02315	0.462	6	631,777	633,561	ubiquinol-cytochrome c reductase, iron-sulfur subunit	
CNAG_02812	0.409	6	714,279	716,015	ornithine carbamoyltransferase (ARG3)	Idnurm et al. (2009)
CNAG_02049	0.395	6	1,331,959	1,334,268	proline oxidase (PUT1)	Lee et al. (2013)
CNAG_05882	0.333	7	1,105,565	1,108,397	class e vacuolar protein-sorting machinery protein (HSE102)	Liu et al. (2008)
CNAG_03223	0.459	8	383,400	384,978	hypothetical protein	
CNAG_04585	0.350	10	963,306	964,931	hypothetical protein	
CNAG_01681	0.333	11	591,464	594,607	cytosine permease (FCY2)	Song et al. (2012)
CNAG_01690	0.276	11	618,826	622,217	MFS transporter	
CNAG_01872	0.339	11	1,144,865	1,146,031	solute carrier family 31 (copper transporter), member 1	
CNAG_06064	0.334	12	207,653	208,840	protein tyrosine phosphatase (PTP1)	Lee et al. (2014)
ω (n=3)	0.852					
CNAG_01019	4.771	5	1,416,547	1,417,779	copper/zinc superoxide dismutase (SOD1)	Narasipura et al. (2003).
CNAG_01611	2.168	11	412,788	415,621	infection related protein of unknown function (LIV8)	Liu et al. (2008)
CNAG_07724	1.023	8	721,568	725,406	copper-sensing transcription factor (CUF1)	

Table 4.17: Virulence-associated genes under selection in the VNBa clade (gene-based methods)

VNBb	Statistic	Chr	Start	End	Product	Reference
MK (n=26)						
CNAG_00193	0.074	1	521,601	527,076	gata type zinc finger protein (GAT1)	Kmetzsch et al. (2011)
CNAG_00440	0.181	1	1,143,872	1,145,519	cyclin subunit of mediator subcomplex (SSN801)	Wang et al. (2011a)
CNAG_00654	0.364	1	1,705,106	1,706,138	sulfiredoxin (SRX1)	Ko et al. (2009)
CNAG_00697	0.274	1	1,817,479	1,819,731	UDP-glucose epimerase (UGE1)	Moyrand et al. (2007)
CNAG_00795	0.207	1	2,097,928	2,099,353	hyphal flocculin (CFL1)	Wang et al. (2012)
CNAG_00926	0.047	5	1,612,427	1,614,160	alpha-1,3/alpha-1,6-mannosyltransferase	
CNAG_01166	0.194	5	1,026,129	1,027,326	urease accessory protein (URE4)	Singh et al. (2013)
CNAG_01242	0.326	5	829,002	831,879	CCAAT -binding transcription factor (HAPX)	Jung et al. (2010)
CNAG_01681	0.333	11	591,464	594,607	cytosine permease (FCY2)	Song et al. (2012)
CNAG_01690	0.318	11	618,826	622,217	MFS transporter	
CNAG_01872	0.375	11	1,144,865	1,146,031	solute carrier family 31 (copper transporter), member 1	
CNAG_02057	0.354	6	1,315,191	1,316,660	urease accessory protein (URE6)	Singh et al. (2013)
CNAG_02315	0.462	6	631,777	633,561	ubiquinol-cytochrome c reductase, iron-sulfur subunit	
CNAG_02812	0.251	3	714,279	716,015	ornithine carbamoyltransferase (ARG3)	Idnurm et al. (2009)
CNAG_03223	0.304	8	383,400	384,978	hypothetical protein	

CNAG_03772	0.398	2	772,402	775,011	high-affinity glucose transporter (HXS1)	Liu et al. (2013)
CNAG_04090	0.280	2	1,592,779	1,596,069	activating transcription factor (ATF1)	Missall and Lodge (2005)
CNAG_04093	0.292	2	1,601,566	1,604,626	putative transcription factor (YRM103)	Liu et al. (2008)
CNAG_04585	0.338	10	963,306	964,931	hypothetical protein	
CNAG_05081	0.490	4	391,625	393,846	phosphodiesterase (PDE1)	Hicks et al. (2005)
CNAG_05290	0.361	4	946,183	948,334	predicted saga histone acetyltransferase complex component (SPT3)	
CNAG_05785	0.314	7	859,695	864,161	putative transcription factor (STB4)	Liu et al. (2008)
CNAG_07422	0.304	5	466,747	468,297	delayed-type hypersensitivity antigen (DHA1)	Mandel et al. (2000)
ω (n=6)						
CNAG_01019	3.333	5	1,416,547	1,417,779	copper/zinc superoxide dismutase (SOD1)	Narasipura et al. (2003)
CNAG_01251	0.424	5	797,058	804,089	e3 ubiquitin-protein ligase (UFD4)	Liu et al. (2008)
CNAG_01611	1.669	11	412,788	415,621	infection related protein of unknown function (LIV8)	Liu et al. (2008)
CNAG_02802	1.074	3	735,903	737,905	inositol/phosphatidylinositol kinase	
CNAG_03932	0.489	2	1,213,102	1,216,148	cell cycle checkpoint protein (RAD17)	Idnurm et al. (2009)
CNAG_06031	2.600	12	120,792	123,948	putative beta-glucan synthase (KRE63)	Gilbert et al. (2010b)

Table 4.18: Virulence-associated genes under selection in the VNBb clade (gene-based methods)



(a) VNBa

(b) VNBb

Figure 4.24: Population Genetics analyses of VNB isolates across the genome. The position of significant virulence-associated genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of LD'; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue and gene undergoing strong selection are indicated in green.

4.3.5 Environmental VNI Population

Phylogeny and Admixture

On the fineStructure output, the two VNI populations detected in Admixture can be clearly observed (Figure 4.25). However, these observations are more difficult to discern using the linked model (Figure 4.26b). When accounting for linkage disequilibrium, the fineStructure analysis reveals that the VNIb sub-population is also subdivided. The population structure of the VNI molecular type is likely to be more complex than observed with this dataset (n=26). Nevertheless, the VNIa population can be identified on the linked model. On the phylogenetic output (Figure 4.26), this sub-population appears very clonal compared to the VNIb. Finally, Sae6, Sae10, BT3 and BT4 are sharing genetic materials with all the other isolates and might represent potential recombinants.

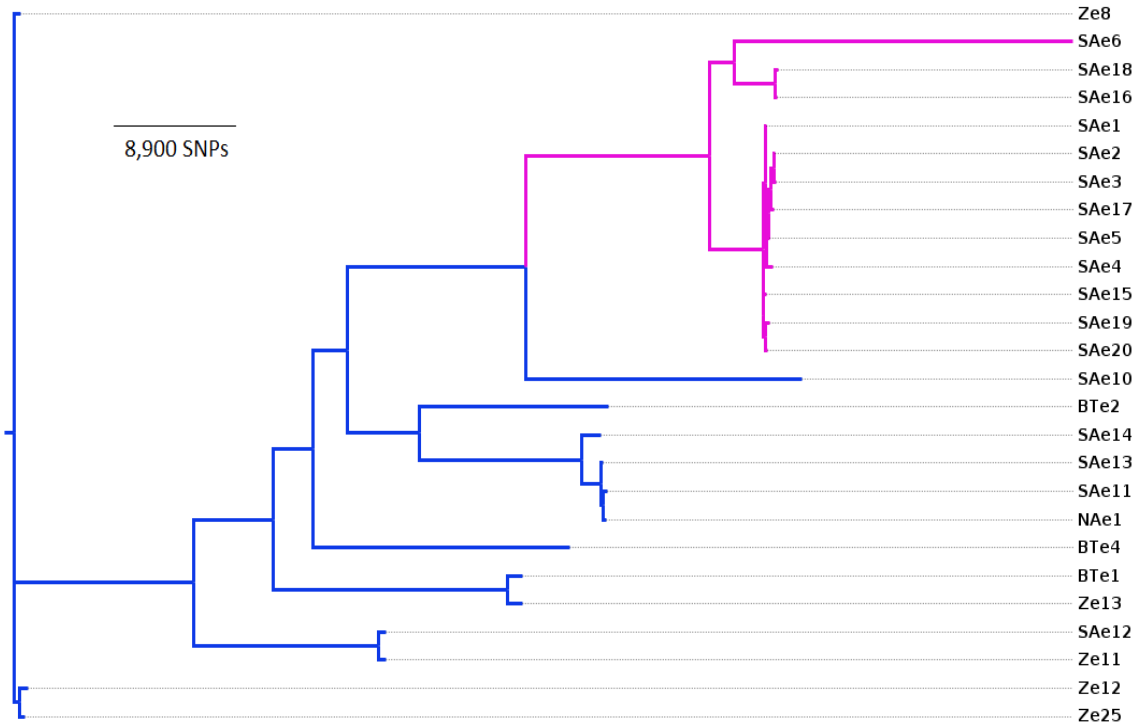
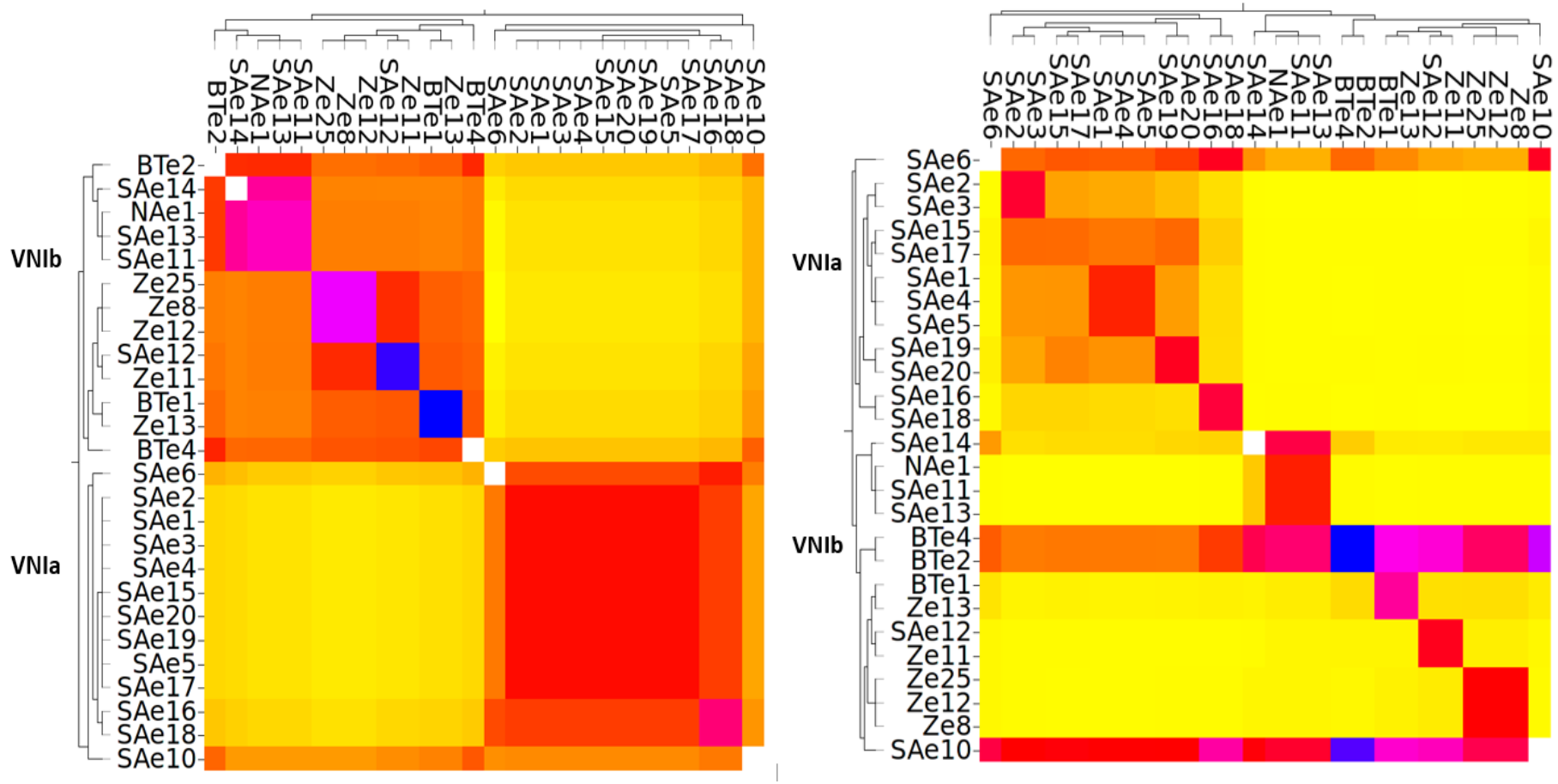


Figure 4.25: Phylogenic Analysis of Environmental *Cryptococcus neoformans* VNI Isolates (n=26) - VNIa isolates are in purple and VNIb isolates appear in blue. Ze: Zambian isolates, SAe: South African isolates, BTe: Botswana isolates and NAe: namibian isolate. Bootstrap values at each node were averaging 100%.



(a) FineStructure Analysis (Unlinked Model)

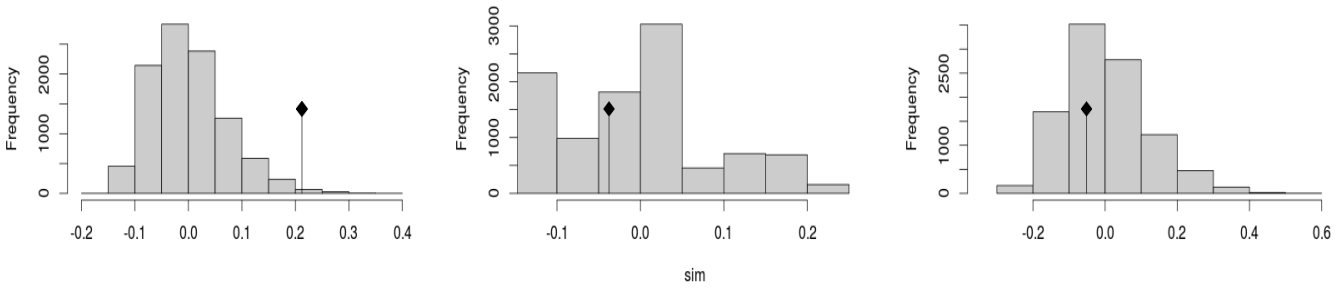
(b) FineStructure Analysis (Linked Model)

Figure 4.26: Population Structure of Environmental *Cryptococcus neoformans* (n=26) - (a) The population structure under the unlinked model reveals two isolated sub-population. Only SAe10 shares genomic regions with the isolates of the two populations. (b) Under the linked model, SAe10 shares genetic materials with every "recipient" genome. To a lesser extent, SAe6, BTe4 and BTe2 are also sharing genomic content with all isolates. The VNIa sub-population is still visible. The VNIb sub-population is further subdivided using this model, indicating possible more complex sub-division.

Effect of distance and barriers on genetic divergence

Isolation by Distance The VNI clade shows signs of isolation by distance (Mantel r statistic = 0.212, p = 0.009). No correlation between geography and genetic distance was detected when investigating each VNI sub-populations separately. Partial mantel tests were performed using climatic variables. Four variables (Latitude, Annual Mean Temperature (Bio1), Mean Temperature of Warmest Quarter (Bio10), and Precipitation Seasonality (Bio15)) improved the r value (Table 4.19) suggesting that these climatic factors are playing a significant role on *Cng* VNI in the environment.

Redundancy analysis (RDA) found that 26.5% of VNI genetic distribution could be explained purely by climate and 34.6% by spatial distance. The influence of climate seems to be reduced in the VNI molecular type compared to the VNB molecular type (where climate explained 75.5% of the variance of genetic distances).



(a) VNI (n=26)) (b) VNIa (n=13) (c) VNIb (n=13)

Figure 4.27: Mantel Tests for the VNI Molecular Type (n=26) using 10,000 replications

Environmental niche modelling The ecological niches of VNI subpopulations using the bioclim layers were model using MaxEnt. We found that the variables contributing the most to the two models (VNIa and VNIb) were Annual Mean Temperature (Bio1 -78.9%) and Isothermality (Bio3 - 66.2%) respectively. These variables were significant when using partial Mantel tests on the VNI datasets. According to the model estimation of the ecological niches, the two inferred subpopulations do not overlap (Figure 4.28). VNIa would be constrained to the south of Southern Africa, whereas VNIb isolates would be located principally inland in South Africa up to Northern Zambia.

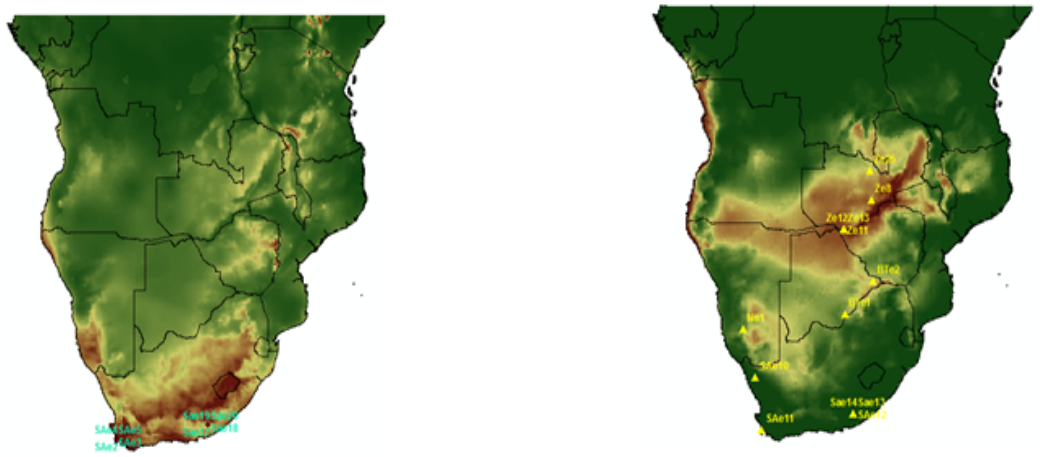
Table 4.19: Partial Mantel Tests in VNI populations

Population	VNI		VNIa		VNIb	
Variables	r	p	r	p	r	p
Mantel test	0.212	0.009**	0.007	0.389	-0.051	0.634
Elevation	0.139	0.053	0.108	0.142	-0.006	0.477
Latitude	0.309	0.002**	-0.179	0.902	0.052	0.313
Longitude	0.023	0.371	0.101	0.135	-0.052	0.633
EVI	0.125	0.069	-0.037	0.565	-0.127	0.846
NDVI	0.124	0.063	-0.403	0.851	-0.116	0.830
Bio1	0.216	0.005**	0.082	0.207	-0.144	0.898
Bio2	0.116	0.055	0.033	0.375	-0.077	0.727
Bio3	0.188	0.009**	0.058	0.280	-0.047	0.629
Bio4	0.191	0.014*	0.014	0.414	-0.047	0.636
Bio5	0.186	0.014*	0.008	0.429	-0.094	0.795
Bio6	0.178	0.016*	0.055	0.356	-0.053	0.645
Bio7	0.181	0.015*	0.048	0.306	-0.058	0.667
Bio8	0.127	0.044*	0.069	0.234	0.044	0.314
Bio9	0.174	0.027*	0.102	0.258	-0.042	0.610
Bio10	0.266	0.001***	0.070	0.268	-0.074	0.727
Bio11	0.147	0.046*	0.132	0.231	-0.038	0.597
Bio12	0.186	0.021*	0.014	0.404	-0.056	0.658
Bio13	0.189	0.022*	0.058	0.292	-0.059	0.670
Bio14	0.177	0.026*	0.022	0.390	-0.057	0.661
Bio15	0.220	0.008**	0.078	0.231	-0.094	0.786
Bio16	0.174	0.015*	0.038	0.346	-0.082	0.747
Bio17	0.158	0.023*	-0.018	0.440	-0.057	0.656
Bio18	0.190	0.009**	0.059	0.243	-0.064	0.684
Bio19	0.165	0.027*	0.021	0.389	-0.054	0.652

$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

Table 4.20: RDA analysis in VNI populations

	VNI		VNIa		VNIb	
RDA	Inertia	Proportion	Inertia	Proportion	Inertia	Proportion
Total explainable	60,928	1	17,671	1	58,064	1.000
Pure climate	16,134	0.265	13,227	0.749	10,227	0.176
Pure geography	21,111	0.346	0	0.000	0	0.000
Joint climate/geography	23,683	0.389	4,444	0.251	47837	0.824



(a) VNIa sub-populations

(b) VNIb sub-population

Figure 4.28: Predicted distribution of the two VNI sub-populations - Environmental Niche Model Distribution obtained with Maxent using WorldClim environmental data.

Landscape Genomics of VNB and VNI

LFMM is a Bayesian approach that detect loci under selection according to environmental variables (Frichot et al., 2013). Several significance thresholds were tested to account for type I errors using false discovery rates (FDR) (Benjamini-Hochberg procedure) and family-wise error rate (FWER) (Bonferroni correction) to account for multiple testing.

No significant SNP was obtained in the VNB clade after correction. However, the VNI clade showed significant associations between the genome and the environment. The numbers of significant SNPs for each environmental variable are listed in Table S14 using different thresholds. We decided to set the statistical significance threshold to 1% before applying Bonferroni correction as described by Stucki (2014).

The 352 significantly associated SNPs identified in 12 variables were mapped to 89 genes (Table S18). Seven genes detected are known to be involved in virulence: CNAG_00150 (Bio4, Bio13, Bio16), CNAG_0149 (Bio13, Bio16), CNAG_03637 (Bio13, Bio16), CNAG_03670 (Bio13, Bio16), CNAG_00546 (Bio4), CNAG_00641 (Bio17), CNAG_02315 (Bio4). Genes identified in more than one variable are shown on Figure 4.29 which displays the $\log(p)$ value over a 10Kb window for each of the twelve environmental variables. The list of genes associated with several climatic variables appears on Table S20.

Four virulent genes were significantly associated with Precipitation in the Wettest Month (Bio13) and in the Wettest Quarter (Bio16) among the respectively 36 and 32 genes identified. In the variable Temperature Seasonality (Bio4), three virulence genes among the 26 genes were detected. The virulence genes obtained using the landscape genomic approach are listed on Table 4.21. Genes under selection present in at least three bioclim variables are listed in Table 4.22. These genes encodes for hypothetical protein except for the virulent gene CNAG_00150 located on chromosome one and two genes located next to each other on chromosome seven, CNAG_06536 and CNAG_06535 encoding for monocarboxylic acid transporter and ribosome biogenesis protein respectively.

Bioclim Variable	Gene	Chr	Product	Ref
Bio13, Bio16	CNAG_00149	1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 4 (931 nt)	
Bio4, Bio13, Bio16	CNAG_00150	1	peptidase	
Bio4	CNAG_00546	1	putative chitin synthase (CHS4)	Banks et al. (2005)
Bio17	CNAG_00641	1	transcription elongation factor SPT5	
Bio4	CNAG_02315	6	ubiquinol-cytochrome c reductase, iron-sulfur subunit	
Bio13, Bio16	CNAG_03637	2	ATP-dependent DNA helicase (YKU80)	Liu et al. (2008)
Bio13, Bio16	CNAG_03670	2	serine/threonine-protein kinase, unfolded protein response (IRE1)	Cheon et al. (2011)

Table 4.21: Virulence-associated genes with significant SNPs after accounting for multiple testing methods

Population Genomic Analyses

Population Genomics and Recombination Analyses The overall Tajima's D value for the VNIa population was significantly negative (Tajima's D = -1.724, p <0.05) and this population showed higher level of LD compared to the VNIb population (D'_{VNIa}=0.432, D'_{VNIb}=0.175). The LD decay plot (Figure 4.30) shows reduced levels of LD decay in the VNIa population. The latter appears more clonal than the VNIb population and its negative Tajima's D value hypothesises a recent selective sweep and/or a population expansion after a recent bottleneck event.

Recombination analysis confirmed the limited recombination among the VNIa clade (p_{VNIa} = 0.11386/bp^{-1}; p_{VNIb} = 0.16919/bp^{-1}; p <0.001). The recombination rate in the two VNI sub-

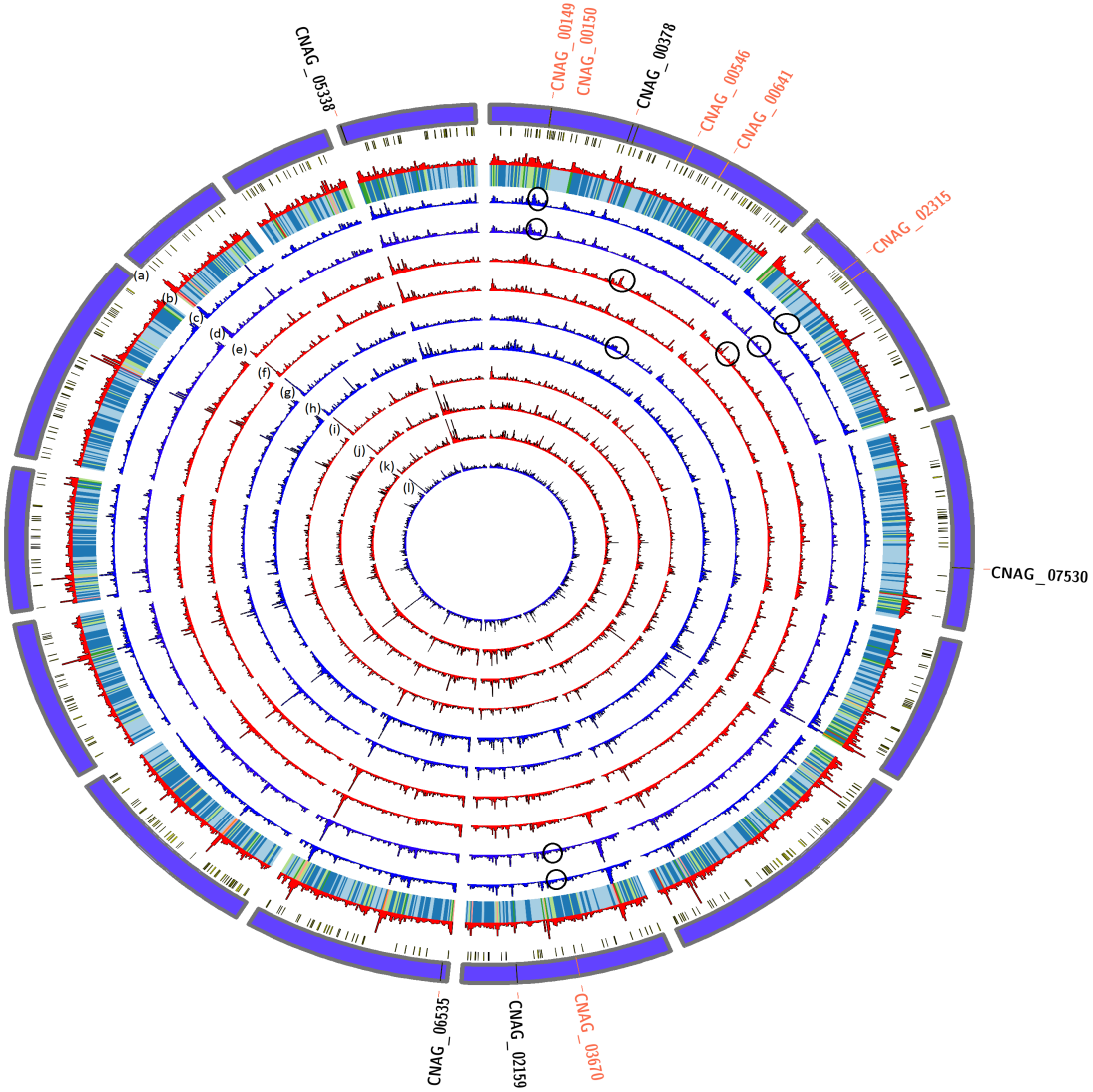
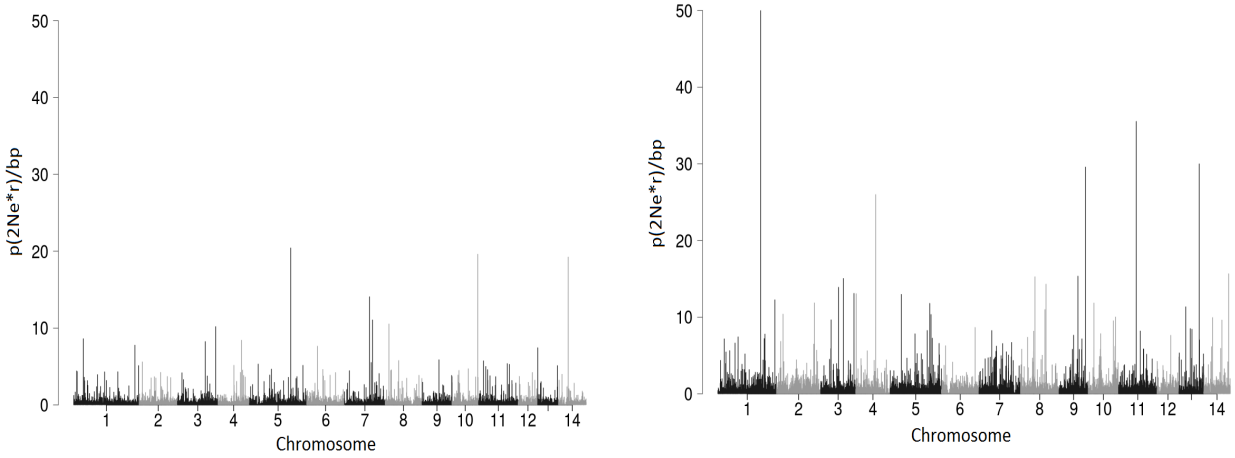


Figure 4.29: Summary of loci under selection identified using LFMM over a window of 10,000 bases. - Virulence genes appears in red and genes identified in multiple environmental variables are labelled in black. Legend - from the outside to inside: (a) Cryptococcal genes with a known-virulence function identified in the literature; (b) SNPs density histogram and Heat map - (c) Precipitation of Wettest Quarter (bio16) ;(d) Precipitation of Wettest Month (bio13) ; (e) Temperature Seasonality (bio4) ; (f) Mean Temperature of Warmest Quarter (bio10) ; (g) Precipitation of Driest Quarter (bio17) ; (h) Annual Precipitation (bio12) ; (i) Annual Mean Temperature (bio1) ; (j) Mean Temperature of Wettest Quarter (bio8) ; (k) Temperature Annual Range (bio7) ; (l) Precipitation of Driest Month (bio14); Environmental variables from temperature (red) and precipitation (blue). Virulent genes are highlighted in red and genes under selection are in black. Circled regions highlight where virulence-associated genes were found for each variables.

Gene	Bioclim Variable	Chr	Start	End	Product
CNAG_00378	Bio1 Bio10 Bio12 Bio13 Bio16 Bio4 Bio5	1	967,715	970,168	hypothetical protein
CNAG_05338	Bio10 Bio13 Bio16 Bio2 Bio4 Bio7 Bio8	14	15,460	16,141	hypothetical protein
CNAG_05339	Bio10 Bio2 Bio4 Bio7 Bio8	14	18,334	19,627	hypothetical protein
CNAG_07358	Bio10 Bio12 Bio13 Bio4 Bio5	1	1,002,298	1,006,448	hypothetical protein
CNAG_02159	Bio13 Bio14 Bio16 Bio17	6	1,049,701	1,051,820	hypothetical protein
CNAG_06536	Bio10 Bio4 Bio7	7	45,550	47,672	monocarboxylic acid transporter
CNAG_06535	Bio10 Bio4 Bio7	7	43,570	45,100	ribosome biogenesis protein UTP30
CNAG_00150	Bio13 Bio16 Bio4	1	411,635	415,418	Peptidase
CNAG_07530	Bio12 Bio13 Bio16	3	1,121,430	1,125,443	hypothetical protein

Table 4.22: Gene containing SNPs under selection in at least 3 climatic variables

populations is significantly lower than in the VNB clades. The LDHat analyses comforted previous findings implying that VNI environmental *Cng* population is more clonal than the VNB molecular type.



(a) VNIa isolates (n=13)

(b) VNIb isolates (n=13)

Figure 4.30: Recombination analysis, using LDhat interval - The peaks represent the recombination hotspots across the genome.

Bayescan and F_{ST} Outlier The two sub-populations, VNIa and VNIb, shared 33,106 SNPs. BayeScan analysis was not able to distinguish significant SNPs using an FDR of 0.05. But as seen in the previous BayeScan analyses, signs of both diversifying and balancing/purifying

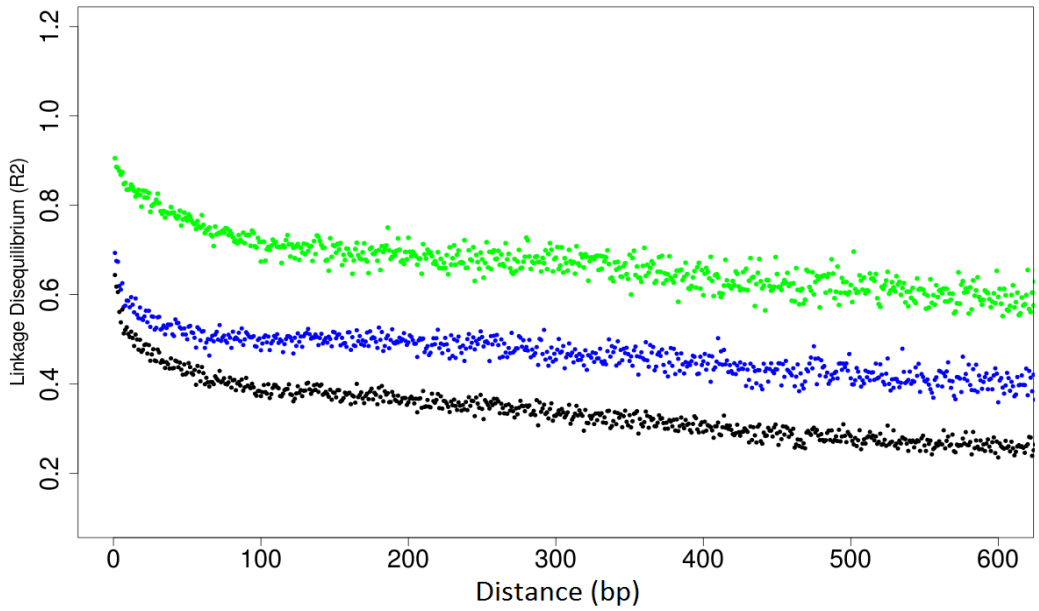


Figure 4.31: Population decay of Linkage Disequilibrium within the VNI isolates - The rate of decay of r^2 in the VNIa sub-population (green) is slower than in the VNIb sub-population (blue). The black dots represent the LD decay for the VNI population (n=24).

selection could be observed with high and low F_{ST} value (as shown on the volcano plot - Figure 4.32b).

Gene Content After filtering the *is*-DNA sequences, 15,474 sequences were analysed and 11% (n=1,709) were considered for the analysis after applying a Bonferroni correction. A total of 24 and 13 genes were significantly associated with VNIa and VNIb respectively. However, none of them had a known virulence function. The details of each gene identified using this method are listed on Table S14 and S16 .

Population Genetics

Frequency-based methods In the VNIa population, no gene could be identified. For the Tajima's *D* value, most SNPs approached a value of D = -2, which is often associated to strong positive selection (Figure 4.34a). Regarding the LD analysis, the D parameter was unable to identify any particular SNP, probably due to the high linkage disequilibrium found in this population (Figure 4.34b).

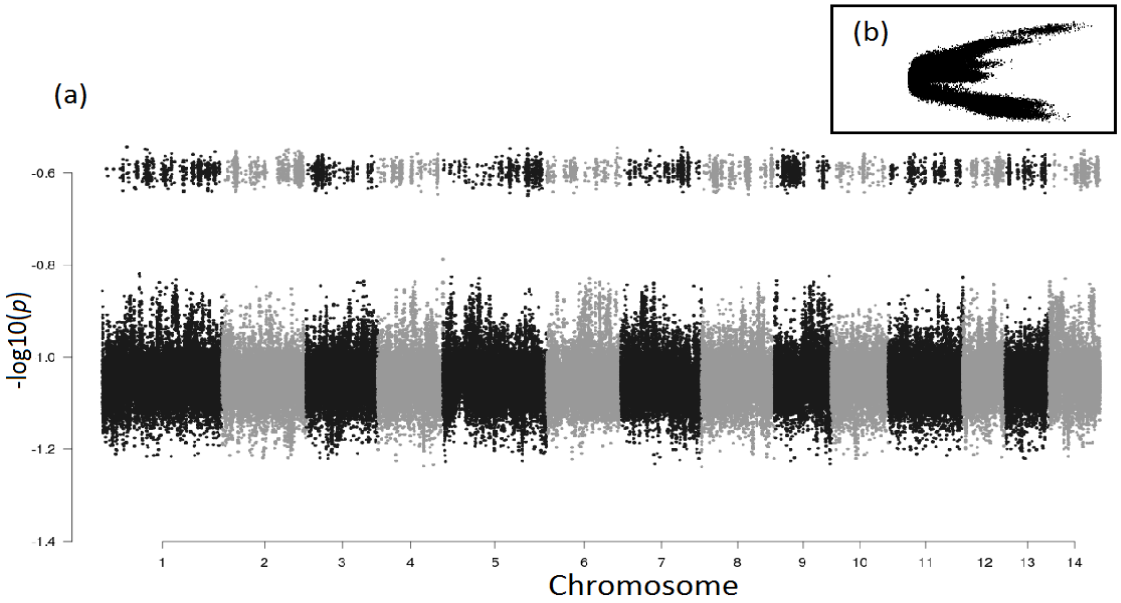


Figure 4.32: Bayescan Output comparing VNI sub-population - (a) Manhattan plot displaying F_{ST} variation across the 14 chromosomes; (b) Volcano plot of the bayescan output showing signs (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST}) and balancing/purifying selection (low F_{ST}).

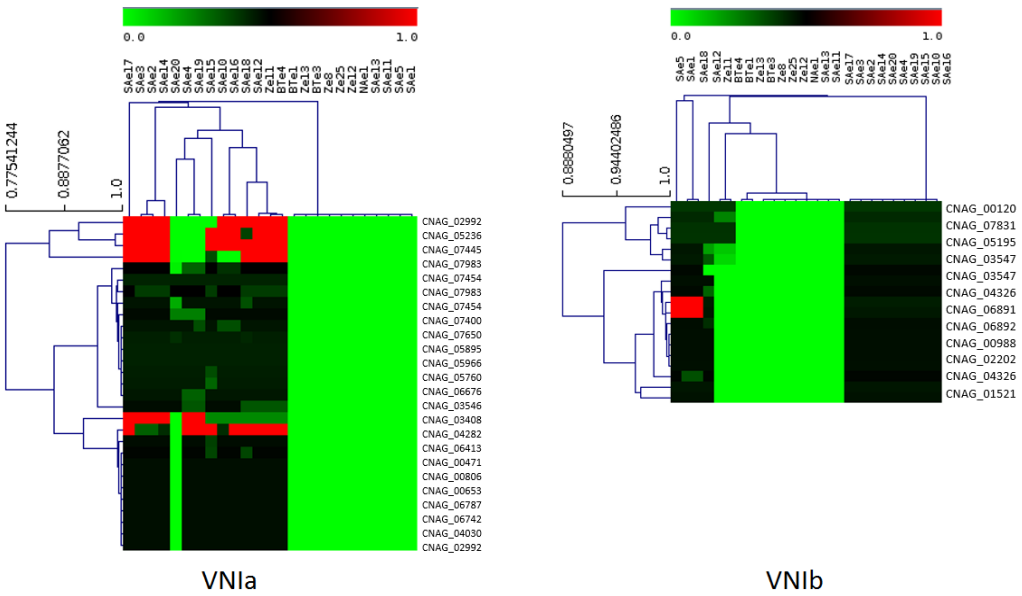
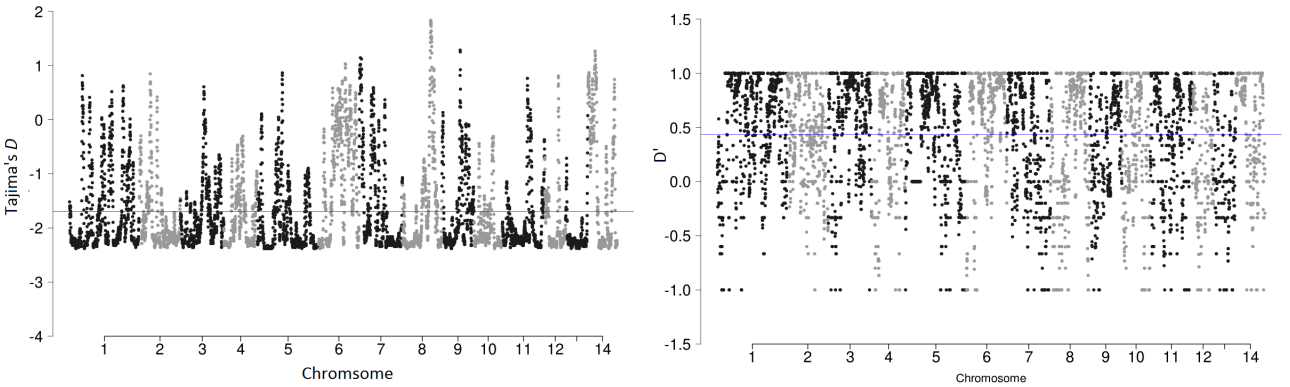


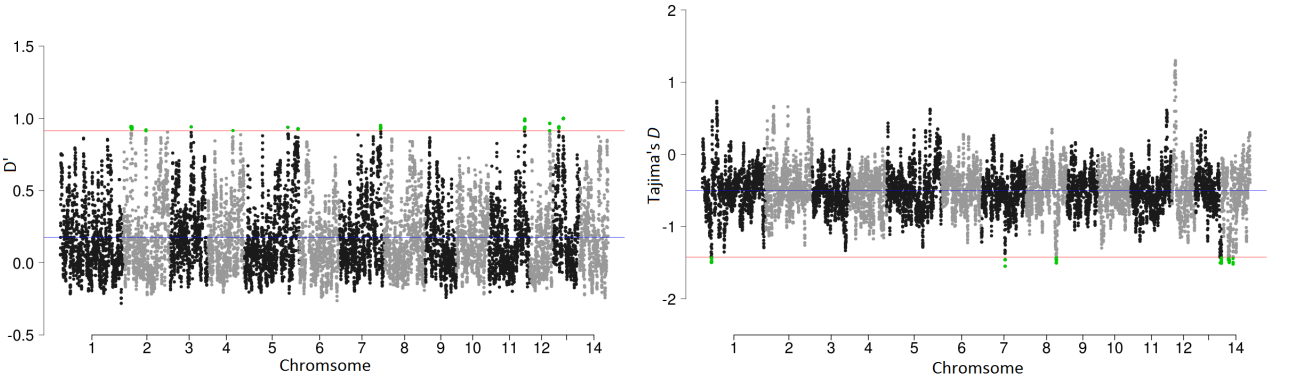
Figure 4.33: Population-level gene content analysis (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content that is shared or differed between the two VNI sub-populations. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0).

In the VNIb population, 27 and 28 genes were associated with Tajima's *D* and D'respectively. Three genes were found by both parameters: CNAG_0637, CNAG_13110 and CNAG_07931.



(a) Tajima's D (VNIa isolates n=13) (b) D'(VNIa isolates n=13)

Figure 4.34: Frequency-based analysis in VNIa population (Vilella et al., 2005) - The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green.



(a) Tajima's *D* (VNIb isolates n=13) (b) D'(VNIb isolates n=13)

Figure 4.35: Frequency-based analysis in VNIb population (Vilella et al., 2005) - The two parameters were estimated using a 100 bp non-overlapping sliding window. Loci in the 0.1% quantile are highlighted in green.

Gene-based methods of positive selection Due to the reduced number of SNPs present in the VNI sub-populations, a limited number of genes met the conditions for the gene-based analyses (a minimum of five SNPs per gene). In the VNIa group, only 1,716 genes were investigated. Eight had significantly elevated MK value and one had elevated ω value. None of these genes were associated with virulence.

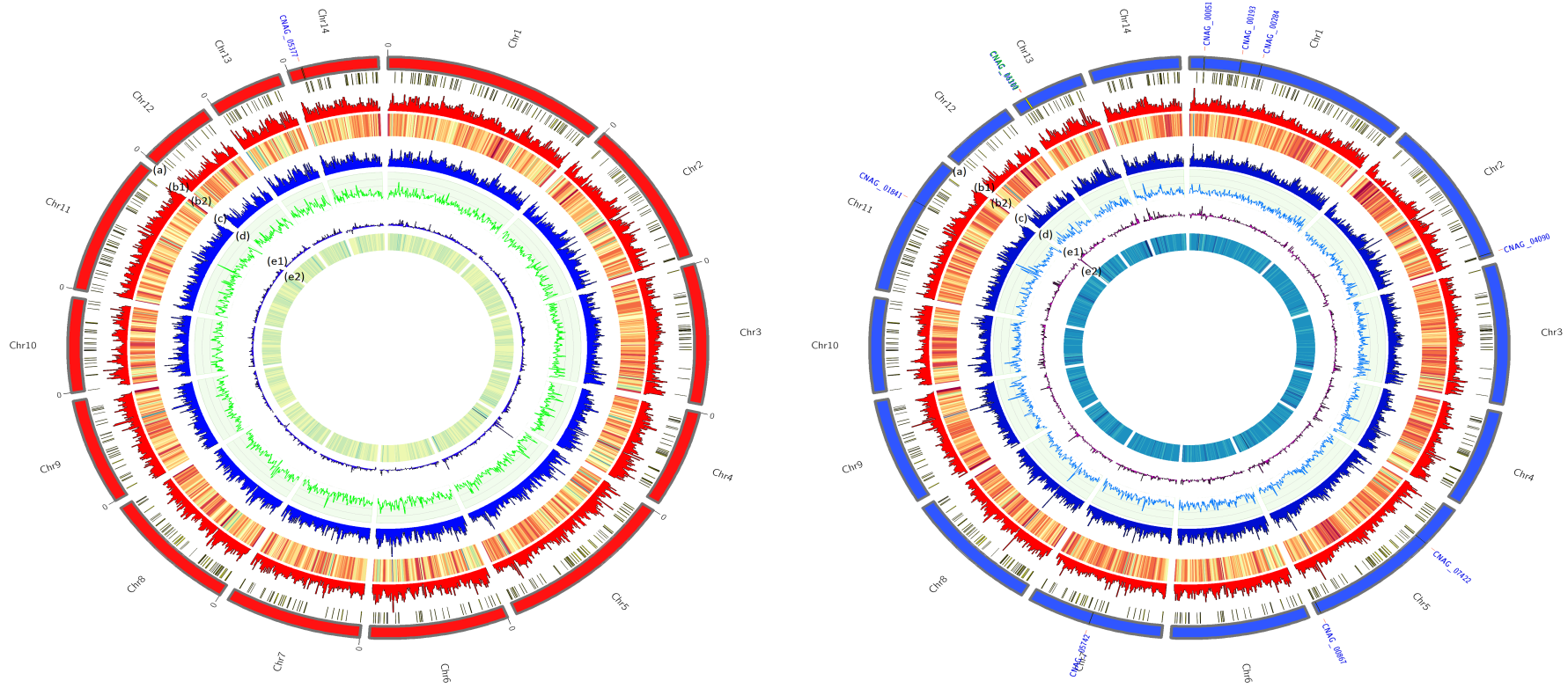
In the VNIb subtype, 4,916 genes were investigated. No gene identified overlapped using both statistics. The MK tests found 68 genes including five genes with a known virulent function

(Table 4.23). The dN/dS analysis found 33 genes including two which are associated with virulence.

Overlap between Frequency-based and Gene-based Analyses No overlap was found between the two approaches in the VNI sub-populations. However, one virulence gene identified in the VNI dataset was also found to be under positive selection in the VNIb dataset: CNAG_06301 which encodes for an agc protein kinase (SCH9). The protein is induced in response to oxidative stress (Ko et al., 2009).

Gene	value	chr	start	end	Product	Ref
Tajima's *D*						
CNAG_05742		7	743,723	746,234	putative site-2 Protease (STP1)	Chun et al. (2007)
CNAG_06301		13	133,480	137,052	agc family protein kinase (SCH9)	Ko et al. (2009)
MK test						
CNAG_00193	0.365	1	521,601	527,076	gata type zinc finger protein (GAT1)	Kmetzsch et al. (2011)
CNAG_04090	0.404	2	1,592,779	1,596,069	activating transcription factor (ATF1)	Missall and Lodge (2005)
CNAG_07422	0.514	5	466,747	468,297	delayed-type hypersensitivity antigen (DHA1)	Mandel et al. (2000)
CNAG_00867	0.321	5	1,782,123	1,784,803	myo-inositol transporter (ITR3A)	Xue et al. (2010)
CNAG_01841	0.275	11	1,052,774	1,054,169	gata family transcription factor (GLN3)	Liu et al. (2008)
ω						
CNAG_00051	0.010	1	148,336	155,003	putative set3c deacetylase complex subunit (SNT1)	Liu et al. (2008)
CNAG_00284	0.049	1	732,808	735,398	efflux protein EncT	

Table 4.23: Virulence-associated genes under selection in the VNIb clade (n=13)



(a) VNIa

(b) VNIb

Figure 4.36: Population Genetics analyses of VNB isolates across the genome. The position of significant virulence-associated genes (at the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of LD'; (d) line plot of the LDhat rho statistic ; (e) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue.

4.4 Discussion

The study was successful at recovering environmental isolates of *Cryptococcus neoformans* var. *grubii* across a broad geographical area (Chapter 2). Subsequently, the genomes of the environmental isolates were sequenced and the *Cng* population structure was investigated. Genes under selection were investigated in the different molecular types and both inter- and intra-lineage analyses were performed. Finally, climatic variables influencing the species evolution were considered.

4.4.1 Population Structure of Environmental Populations

To investigate the population structure of African *Cng* populations, 50 environmental isolates were sequenced using an Illumina HiSeq 2000 platform. Phylogenetic analyses and population structures analyses revealed two sub-populations in each of the two *Cng* molecular types (Figure 4.2). Topologies of phylogenetic trees obtained using different methods (i.e. BEAST (Drummond and Rambaut, 2007)) were largely congruent with RAxML (data not shown). Branch lengths were consistently longer in the VNB clade compared to VNI clade.

The fineStructure analysis revealed the population structure of the environmental dataset and assessed the relatedness between each genome. This analysis showed that the VNI and VNB molecular types are clearly separated. This analysis was nicely corroborated with Admixture and allowed assignment of each isolates to its population. When considering fineStructure and the phylogenetic results together, the VNI population structure seems relatively weak compared to the VNB sub-populations which displayed long branches on the phylogenetic tree (>40,000 SNPs).

Previous research based on MLST datasets hypothesised the presence of a VNB/VNI hybrid population in African environments (Chen et al., 2015; Litvintseva et al., 2011) but the present study using whole-genome sequencing (WGS) showed no evidence of this population. We acknowledge that the complete *Cng* diversity has not been fully sampled in the study given the limited number of isolates sequenced.

4.4.2 Population Genomic Analyses

Recombination

The level of gene flow between the two molecular clades was limited (F_{ST} >0.2). Between the two VNB subpopulations, population subdivisions were more prominent (F_{ST} = 0.163). LDHat analyses identified a high numbers of recombination hotspots within the VNB genomes compared to VNI *Cng*. The recombination in the VNB clade might be explained by the presence of both mating types in the environment leading to sexual recombination. The abundance of the MAT a locus in the VNB molecular type has been observed in previous studies (Chen et al., 2015). Sexual reproduction provides clear benefits as meiosis allows to break up LD regions. Therefore, sexual reproduction increases diversity which improves the efficiency of natural selection (Sun and Heitman, 2015). Clonal expansion via asexual reproduction can also confer an evolutionary advantage. MATα strains are known to be more virulent than MATa isolates (Nielsen et al., 2003). Potential phenotypic differences between the two mating types will be investigate in the following chapter.

Environmental VNI isolates overwhelmingly possessed the MATα locus haplotype (97%). MATα haplotypes are largely dominant throughout the literature (Nielsen et al., 2007). VNI *Cng* constitute a clonal clade, which does not imply a total absence of recombination (Tibayrenc and Ayala, 2012). Indeed low levels of recombination were identified within the VNI isolates (Figures 4.6b & 4.30) and could be the results of same-sex mating (Lin et al., 2006) (See Introduction, section 1.3.1). Additionally, LD decay was lower in the VNB clade (Figure 4.7) indicating lower levels of linkage disequilibrium in this clade. LD is expected in clonal organisms where recombination does not act to break up allele associations. The presence of longer haplotypes in the VNI molecular type, particularly in the VNIa sub-population, could be the results of bottlenecks and founder events (Kim and Nielsen, 2004).

Only two out of the 13 VNIa isolates were found in pigeon guano but all isolates of this sub-population were sampled in urban environments where pigeons are highly present. The reduced diversity observed in this population could be the result of a dissemination by birds such as *Columba livia*. One VNIb isolate had the MAT a locus and was also isolated from pigeon guano. However, the branches on the phylogenetic tree are longer and its population structure was more complex than in VNIa isolates under the fineStructure analysis (linked model - Figure

4.15b). The 13 isolates present in the VNIb lineage are unlikely to account for the whole diversity present in this subpopulation and additional sampling would reveal the presence of other genetic clusters. Additionally, the VNIb sub-population presented more recombination hotspots (Figure 4.30) and a higher recombination rate than VNIa ($p_{\mathrm{VNIa}} = 0.11386/\mathrm{bp}^{-1}$; $p_{\mathrm{VNIb}} = 0.16919/\mathrm{bp}^{-1}$; $p < 0.001$). This finding highlights the complex evolutionary history of the VNI molecular type and challenge the idea of a pan-global genotype (Simwami, 2011).

Genes under selection

To investigate selection acting on the two molecular types, we combined multiple methods using a twofold approach. First, we investigated genome-wide variations using frequency-based methods (Tajima's D and Lewontin D'parameters) and gene-based methods (dN/dS and MK tests). Genes were considered as strong candidates undergoing selection if at least two parameters identified a particular gene as described in the methods. Then, amongst the genes that are putatively undergoing selection, we searched in the literature to determine whether these genes had a known virulence-associated function. These methods have been used to investigate macro and microevolution in a wide range of organisms (Vitti et al., 2013). As stated in many population genomic papers, we acknowledge the inherent difficulty to identify such outlier loci without false positives (Cullingham et al., 2014). However, this 'hypothesis-free'genomic approach represents an unbiased method to uncover genetic region undergoing selection, but additional work is necessary to connect the genetic and the phenotype. An attempt to demonstrate phenotypic differences among the *Cng* molecular types will be undertaken in the following chapter.

Inter-specific variations By searching for lineage-specific accelerations of evolution in VNI and VNB genomes, we can search for biological meaningful variants which are playing a role in the adaptation of the molecular type and might have favoured virulent traits. Methods to identify gene differences (BayeScan and LS-BSR) were non-conclusive, probably due to the high number of genetic differences between the two molecular types (78,132 and 509,332 SNPs unique to VNI and VNB respectively).

The reverse ecology approach identified putative gene candidates involved in virulence in each molecular types (Tables 4.5, 4.6 & 4.8). In the VNB clades, two genes with a known virulence function were found, CNAG_00306 and CNAG_01106. Interestingly, the CNAG_01106 encoding

for vacuolar-ATPase subunit (VPH1) was part of broader region that contained a cluster of five genes under selection (Figure 4.24). A disruption of the VPH1 has been shown to impact key virulent factors (laccase and capsule production, urease expression and the ability to grow at 37°C) (Erickson et al., 2001). Moreover, *vph1* mutant were avirulent in a mouse model and virulence was fully restored in VPH1-reconstituted strains. Phenotypic work should be performed to assess any potential variations of these virulent factors between the molecular types (Chapter 5). The other putative candidate gene CNAG_00306 plays a role in response to elevated Cu (Ding et al., 2013).

The VNI molecular type which is predominantly isolated in patients, displayed a strong negative Tajima's D value (D = -0.787) indicative of a recent selective sweep. However, little overlap was found between the different parameters in the VNI clade. Only two genes could be identified but they had no known virulence-associated functions (Figure 4.13). Candidate genes were difficult to identify in this clade. The number of SNPs was limited due to the clonal population structure of the molecular type. Additionally, isolates were aligned to *Cryptococcus neoformans* var. *grubii* H99 reference Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/) which is a VNI reference and reduced further the number of SNP available for analysis. Interestingly, when comparing gene under selection in both lineages, CNAG_06813 encoding for *CAP1* was found to have strong level of LD and was identified in both VNI and VNB isolates using D statistic. The *CAP1* gene is on MAT locus (Chromosome five) and is specific to MAT α isolates. The latter is a capsular protein and is thought to play a key role in virulence (Lengeler et al., 2002).

Intra-specific variations

VNI Environmental sub-populations In the VNIa sub-population, no gene with a known virulent function was identified using our method. The lack of results is due to the fact that this population had strong linkage disequilibrium and strong negative Tajima's D (D = - 1,724). Additionally, the VNI reference is genetically similar to the VNIa isolates. Only 1,716 could be considered in the gene-based analysis yielding a total of nine putative genes under selection. Within lineage VNIb, three genes with unknown functions were uncovered using frequency-based methods. Seven virulent genes were found with one gene-based test but none of them could be identified using two genomic metrics (Table 4.6).

VNB Environmental sub-populations Greater levels of genetic diversity were observed among the VNB isolates. Using a VNB reference, the genomic clade still harboured high level of genetic diversity (Table S7).From the fineStructure analyses using the model taking into account linkage (Figure 4.15b), we observed important sharing of genetic materials between the two sub-populations VNBa and VNBb. These genetic exchanges could be the result of sexual recombination. Indeed, the overall proportion of MAT a locus within the VNB isolates was 58%. The VNBa isolates had a significant negative Tajima's D value, whereas VNBb population seemed to close to mutation-drift equilibrium (D_{VNBa} = -1,247 and D_{VNBb} = -0.199). However, no conclusion should be drawn regarding VNBa sub-population because the number of VNBa isolates was limited and this population had a strong population structure.

In the VNBa clade, four genes with known virulent functions were identified among the 42 gene candidates. These genes encode for a VSP protein (VSP4), an alpha-specific protein kinase (STE20α), myol-inositol transporter (ITR3) and catalase (CAT4). In the VNBb group, no gene met the criteria described in the methods. 40 genes found with only one parameter were also found in the inter-lineage analysis; three genes encoding for myol-inositol transporter (*itr1*, *itr2*, and *itr3*), *cap1*α, *vph1*, *cat4*, *sod1*. This finding indicates that the analyses was successful in finding gene families positively selected in the environment, despite the caveats of such methods. The performance of outlier detection methods had been previously reviewed (Excoffier et al., 2009; Narum and Hess, 2011; PérezFigueroa et al., 2010) and the use of a consensus of multiple parameters had been described as a robust method to reduce false positive.

4.4.3 Biogegraphy

Biogeography studies the biodiversity over time and space (Martiny et al., 2006). It allows to retrace and quantify patterns of migration based on genetic data and to gain insights on the mechanisms leading to speciation or extinction. Local dispersal is thought to be predominant in fungal organisms (Paape et al., 2013), therefore genetic differentiation is expected to show sign of isolation by distance (Wright, 1943).

Distribution of *Cng* molecular types in Southern Africa

The VNB Molecular Type Significant patterns of spatial genetic correlation were detected between the two genetic VNB clusters. Isolation by distance was significant among the VNBb isolates (Mantel r statistic = 0.340, p <0.005). While Bayesian model did not detect significant loci under selection, 16 climatic variables were significantly correlated with the VNBb genetic diversity using partial Mantel tests (Table 4.10) Regarding the VNBa isolates, no sign of spatial genetic correlation were detected but the number of isolates was limited and this genetic cluster presents a strong population structure. Litvintseva et al. (2005) have previously described geographically isolated clusters among this lineage in Botswana using MLST approaches, however differentiation using WGS data and a broader geographical sampling had not been yet attempted.

The population structure observed within the VNB clade reflects an adaptation to different ecological niches. The two genetic clusters (VNBa and VNBb on Figure 4.17) overlapped geographically and the presence of recombinants (Z1 and Z12) points toward the possibility of the two clades recombining. These results are consistent with our previous estimation of the recombination rate using LDhat, where VNB isolates had a significantly higher recombination rate than the VNI molecular type. RDA analysis reaffirmed the strong influence of climate on the distribution of the VNB isolates (75% of the genetic variance explained solely by climate, Table 4.11).

The environmental niche model of the two VNB clades found that their different geographic locations could be explained by environmental factors, suggesting that these two clades may be adapted to different environments. The predicted ecological niche for the VNBa isolates spread from Zambia to South Africa, whereas VNBb seemed to be contained from Northern Zambia to northern Botswana. The model strongly identified the Zambian Mopane Woodlands as a potential ecological niche for the VNBb population. Indeed, these isolates were principally isolated on *C. mopane* tree (13 for VNBb and only 4 for VNBa; Table S2).

Within the VNBa sub-population, the strong population structure of these isolates and its wide geographic spread indicate that the genetic diversity observed in this study is unlikely to reflect the complete genetic diversity of this clade. Additional sampling is required to fully uncover the diversity of the highly diverse VNB molecular type.

The VNI Molecular Type With only 26 genomes isolates in four countries, the ecological niche of the VNI molecular type has not been fully sampled across southern Africa. The two sub-populations were both isolated in South Africa but VNIa isolates were only present along the coast of the country. On the other hand, VNIb isolates were found throughout southern Africa. VNI isolates showed spatial structuring over this molecular type geographic range (Mantel r statistic = 0.212, $p = 0.009$). Climate also played a significant impact on VNI clade distribution. Climatic factors seemed to be less central to the observed distribution of the molecular type compared to VNB clade. The RDA analysis revealed that only 26% of the VNI distribution could be explained by climatic variables.

VNI *Cng* were isolated in both rural and urban centres, but most patients reporting to hospitals lived in urban settlements. The common pigeon, *Columba livia*, is considered as a potential vector of the fungus and pigeon guano is known to be the primary niche of the VNI molecular type worldwide (Littman and Borok, 1968). Avian birds are present in multiple ecoregions throughout southern Africa. With the recent intensification of urbanisation in Africa, *Columba livia* which settled in these urban environments might have brought environmental *Cng* into these densely populated urban centres and led to an increase of cryptococcosis incidence (Filion et al., 2006). A comparison of environmental genomes with clinical isolates is required to understand whether patients acquired the infection in urban centres or in rural areas (Chapter 5).

Landscape Genomics

When comparing the different landscape genomics methods available, Latent Factor Mixed Model (LFMM) has been found to be the most conservative and among the most robust approach (Stucki et al., 2014). Another advantage of this method is that it takes into account population structure.

Within the VNB molecular type, although landscape genomics analyses were not significant after accounting for multiple testing, it appears clear that some outlier loci were correlated with environmental variables and that climate might help toward the acquisition of virulent traits. For example, when investigating the correlation between VNB isolates and Max Temperature of Warmest Month (Bio5), two non-synonymous mutation in the O-acetyltransferase (*cas92*)

gene (CNAG_00865) on chromosome five were identified (Figure S9). Additional samples are needed to fully uncover the role of natural selection on the adaptation of the different lineages to their ecological niches. Temperature and Precipitation were significant variables but additional variables are likely to contribute to selection.

In the VNI clade, 89 genes which were correlated with an environmental variable, seven had known virulent functions but none of them were identified in the previous analysis of selection. Only one gene was identified using two population genomic parameters, CNAG_00636 encoding for the CDC7 protein kinase.

Of the seven virulent genes found using the landscape genomic approach, one virulent gene was a transcription factor (CNAG_00641) and the six of remaining virulent genes had enzymatic functions. The *yku80* (CNAG_03637) has DNA repair function (Liu et al., 2008) and *ire1* (CNAG_03670) plays a key role in growth at high temperature (Cheon et al., 2011). Both genes had loci correlated with precipitation in the wettest month (Bio13) and quarter (Bio16). However, correlation does not imply causality. The current study is only preliminary but offer potential testable hypothesis regarding the emergence of virulent factors in the environment. The environmental parameters used in this study are unlikely to reflect the main environmental forces driving adaptation in *Cng*. Nevertheless, the study showed that landscape genomic represents a powerful tool to understand evolutionary forces and that this approach can be applied to environmental pathogen like *Cryptococcus neoformans*.

Landscape genetics was coined in the last decade (Manel et al., 2003) and remains a new and exciting field. To understand population dispersal and migration, ecological processes needs to be taken into account. The field provides a powerful alternative (Bolliger et al., 2014) to traditional Mantel tests which exhibit high number of type-I error (Legendre and Fortin, 2010). Research needs to expand the availability of environmental data worldwide to be able to describe more accurately the association between the genome and the environment.

4.5 Conclusion

The sequencing of environmental *Cng* genomes shed lights into the population structure of this organism. The two molecular types had different profiles. The global VNI molecular

type isolated principally in urban settlements appeared as a clonal clade composed of two sub-populations with high levels of linkage disequilibrium. Currently found uniquely in Africa, the VNB molecular type appeared as a naturally evolving organism. The latter was isolated in arboreal ecotope and this African *Cng* molecular type displayed high level of recombination and high genetic diversity. Genomic analyses and the investigation of selection found that the two molecular types were under different evolutionary pressures. The two *Cng* clades have been separated for millions of years and have adapted to different ecological niches leading to different strategies to maintain infection and avoid predation. The reverse-ecology approach identified putative genes under selection which may take part in virulence. However, because fungi are free-living organisms causing opportunist infections in human, it is difficult to identify specific traits which contribute to pathogenesis (Moran et al., 2011).

The validation of the identified genes as genes associated pathogenesis needs to be tested to investigate potential phenotypic differences among the molecular types. Moreover, the molecular type which predominantly infects HIV/AIDS patients in southern Africa remains unknown. The VNI molecular type is the most commonly isolated clade in patients (Beale et al., 2015). In order to investigate the relative contribution of *Cng* sub-populations to human infections, the next chapter identifies and sequences a panel of clinical isolates from Zambia and South Africa collected as part of a MRC clinical trial in order to understand whether the genetic makeup of *Cng* infecting human is linked to specific environmental genotypes.

Chapter 5

Clinical and Environmental Cryptococcal populations

5.1 Introduction

In the previous chapter, I uncovered and described the population structure of environmental *Cryptococcus neoformans* var. *grubii*. Across southern Africa, this research found two molecular types in the environment, VNI and VNB, which displayed different population genetic structures. The two molecular types were associated with different ecotopes and appeared to be exposed to different evolutionary pressures, being influenced by different spatial and climatic (precipitation and temperature) factors. This chapter will investigate the population genetic structure of clinically-isolated *Cng* and will compare it with the previously described environmental populations. The study aims to understand whether clinical isolates are associated with a particularly disease-causing genotype. Globally, the VNI clade is the most prevalent molecular type that is isolated from patients. We recovered VNI *Cng* from densely populated urbanised areas, which is consistent with the current understanding of *Cryptococcus* epidemiology. However, the presence of environmental VNB *Cng* in *C. mopane* in Zambia represents a potential source of localised, more rural, infection. The chapter will compare environmental isolates against clinical isolates in order to gain insights into *Cryptococcus* epidemiology.

5.1.1 Epidemiology and Population Structure

Cryptococcal meningitis (CM) accounts for 13% to 40% of AIDS-related mortality in sub-Saharan Africa (Van Wyk et al., 2014). This systemic mycosis presents the highest incidence in immunocompromised patients accounting for one million cases per year (Park et al., 2009). Despite increasing access to antiretroviral therapy (ART) in sub-Saharan Africa, the mortality in HIV/AIDS patients ranges between 20 to 50% in optimal settings (Boulware et al., 2014) reaching 70% in healthcare-deprived areas (Desalermos et al., 2012).

The globally distributed VNI molecular type displays a clonal population structure and is commonly associated with pigeon guano. Compelling evidence shows that the VNB molecular type evolves in arboreal environments. The latter has a high proportion of MAT a mating type indicating that the molecular type is prone to recombination. The presence of the VNB molecular type in Zambia suggests that the genetic profile of clinical isolates in the country might be different than that observed outside Africa where VNI *Cng* predominates.

Recent research suggests that *Cryptococcus neoformans* var. *grubii*, the species which accounts for 99% of disease global burden, emerged 'out of Africa'within recent history (≈5,000 - 0 ybp) to colonise the world (Litvintseva et al., 2011; Simwami et al., 2011). Litvintseva et al., (2011) & Simwami, (2011) reported that the high genetic variability observed in VNB and African VNI strains support the 'out-of-Africa'hypothesis. The present chapter will investigate the evolutionary history of *Cng* in the light of whole-genome data.

5.1.2 Study of virulence in *Cryptococcus neoformans* var. *grubii*

The different molecular types are under different environmental pressures and have markedly different genetic profiles (Chapter 4). The evolution of the different lineages in their local environment might have led to different clinical manifestations, which has the potential to result in enhanced or reduced virulence and hence clinical outcomes. Evidence of lineage-specific phenotypes and clinical outcomes have been recently reported (Beale et al., 2015; Wiesner et al., 2012). Previous studies demonstrated that increased in vitro phagocytosis by murine macrophages was associated with higher fungal burden at CM diagnosis and hence poor patient survival (Sabiiti et al., 2014). Competition with environmental predators such as amoeba

or nematodes may explain *Cng* uptake and survival in macrophages. Cryptococcosis is an opportunistic infection and the current hypothesis presupposes that the saprophyte has acquired its pathogenicity as a defence mechanism against environmental predators (Steenbergen et al., 2001).

An analysis of *Cryptococcus* fitness will be performed to compare clinical and environmental isolates. The project will try to understand whether clinical isolates have phenotypic features which make them more prone to infect patients. These differences could potentially have an important role in the establishment of cryptococcal infection in humans. The study will examine cryptococcal virulence factors (growth at 37°C, laccase activity, capsule production and phagocytosis) and will investigate phenotypic variation among the different *Cryptococcus* genotypes and lineages.

C. neoformans and *C. gattii* are the only two species of the Tremellales which can tolerate temperatures above 30°C (Bovers et al., 2008). The ability to grow at 37°C is essential for infection and temperature-sensitive *Cng* mutants have been found to be avirulent (Perfect, 2006). Phagocytosis is an important factor in the development of cryptococcal meningitis. The 'Trojan horse'scenario, whereby the yeast 'hides'within macrophages to enter the central nervous system is the most documented strategy of *Cng* crossing the blood brain barrier (Charlier et al., 2005). Laccase activity and capsule production are two virulence factors which have been found to promote *Cryptococcus* ability to establish an infection (Lin et al., 2006). Associated with lignin degradation, laccase is thought to enhance the ability of saprophytes to colonise and metabolise decaying wood (Lazera et al., 1996). Moreover the enzyme has been shown to increase resistance to antimicrobial activities and to interfere with oxidative burst during phagocytosis (Williamson, 1997). The polysaccharide capsule protects against amoeboid predators in the environment and provides protection against phagocytosis (Chaka et al., 1995). Capsule size increases during the infection (Feldmesser et al., 2001) and acapsular mutants display reduced virulence in mouse models (Chang and Kwon-Chung, 1994). Capsule production has also been correlated with higher intracranial pressure (ICP) (Robertson et al., 2014). An exhaustive review of *Cryptococcus* virulence factors can be found in the introduction (See Section 1.8).

5.1.3 Genetic factors underpinning virulence

The demonstration of phenotypic effects is the ultimate validation of putative genetic candidates found to be under selection. Previously, we investigated whether local adaptation favoured any virulence genes in each *Cng* sub-populations. Positive selection is thought to be the primary mechanism of adaptation and studying the allele frequency spectrum can be used to detect signals of natural selection (Haasl and Payseur, 2016). The identification of candidate genes may represent meaningful biological variations. However, functional evidence is a requirement to prove whether genetic differences underpin phenotypic differences. The examination of phenotypic differences might explain some of the genetic variation observed between the different *Cryptococcus* genotypes identified in the previous chapter. Additionally, we will use the previously described 'hypothesis-free'genomic approach (Section 4.1.2.) to identify candidate loci that differentiate between clinical and environmental isolates.

Cryptococcus research recently entered the genomic area. Early research has shown that infection could be linked to genomic changes. For instance, the impact of chromosomal copy number variations (CCNV) are thought confer an advantage to adaptation to stress and influence virulence (Fraser et al., 2004). First, the study will investigate whether infection is caused by a particular genotype. Cryptococcosis is an opportunistic infection, therefore, we will then ask whether clinical isolates exhibit any genotypic or phenotypic characteristics which may predispose them to successfully infect the human host.

Aims:

1. Investigate whether genotypes associated with human infections belong to a particular subset of isolates found in the environment.

2. Investigate whether fitness variants occur among genetically and geographically different isolates that may have clinical relevance.

3. Link genotypic differences observed in environmental *Cng* (Chapter 4) to the in vitro phenotypes to understand the evolutionary ecology of wild *C. neoformans*.

4. Identify genetic variants in clinical compared to environmental isolates and see whether they can be related to any potential phenotypic differences.

5. Re-examine the 'out-of-Africa'hypothesis in molecular type VNI in light of whole-genome data.

5.2 Methods

5.2.1 *Cryptococcus neoformans* var. *grubii* Isolates

African Environmental *Cng* isolates (VNI: 26 isolates, VNB: 25 isolates) have been previously described (Chapter 4). Clinical isolates were collected from patients in South Africa (VNI: 61 isolates, VNB: 14 isolates) and Zambia (VNI: 14 isolates, VNB: 4 isolates) as part of MRC-funded clinical trial. Five VNI isolates from Botswana were provided by the Broad Institute. A total of 106 VNI and 18 VNB clinical African isolates were included in the study. Additionally, 29 clinical and 20 environmental VNI isolates from India were given by the University of Delhi (Chowdhary et al., 2012). A total of 155 VNI isolates and 44 VNB isolates were included in the study.

5.2.2 Whole-Genome Sequencing

See Section 4.2.1

5.2.3 Phylogeny

See Section 4.2.2

5.2.4 Population Assignment and Admixture

See Section 4.2.3

5.2.5 Mating Type

See Section 4.2.4

5.2.6 Population Genomic Analyses

See Section 4.2.6

5.2.7 Chromosome Copy Number Variation (CCNV)

To investigate Copy Number Variation (CNV), the Average Read-Depth per base was obtained using SAMTools (Li et al., 2009). The comparison of each isolate was performed over each chromosome using a 10Kb non-overlapping window. The ploidy was visualised in R using the package qqman (Turner, 2014). Circos plots were generated using ClicO FS (Cheong et al., 2015).

Divergence Estimations and historical demographic trends BEAST v.2.3 was used to estimate time of divergence to the most recent common ancestor (TMRCA) between the genetic clusters. Coding-regions were predicted using Prodigal (Hyatt et al., 2010). Then, multi-fasta alignments were generated using mafft v.7 (Katoh and Standley, 2013). Samtools v.1.4 (Li et al., 2009) was used to convert fasta files from the Bam files generated with GATK (See Section 4.2.1). Owing to computational constraints, we subselected Chromosome 12 for a focused analysis. This Chromosome is 774,062 bp-long and was selected to optimise computation time in BEAST.

Recombination is an important evolutionary force in fungal genetics. Therefore, BradNextGen (Marttinen et al., 2012) was used to identify recombination events analysing the dataset using 25 iterations. Significance was estimated using 100 permutations. The software uses a Bayesian modelling approach to detect homologous recombination events in whole-genome datasets. Recombination regions in each VNI genetic cluster on chromosome 12 were detected and were subsequently removed.

Analyses were computed as described by Simwami, (2011). A Hasegawa Kishino-Yano (HKY) model of sequence evolution was assumed and a strict molecular clock model was applied. The same mutation rate of 4.10^{-9} substitutions per nucleotide per year was used using the same calibration of $4.5.10^{6}$ years for the divergence within the *Cng* lineage (Xu et al., 2000). Additionally, another mutation rate was calculated from the formula proposed by Lynch, (2010). The regression for cellular organisms is -0.81 + 0.68log10(G). With a genome size (G) of 18,880,888 bp, the mutation rate was estimated at $4.13.10^{-9}$ mutations per site per generation. Bayesian Markov

Chain Monte Carlo (MCMC) were run for 10^8 iterations logging parameters every 50,000 steps with a 20% burn-in. Each run was analysed in Tracer v.1.6.

The modified multi-fasta alignment for Chromosome 12 (free of recombination sites) was imported into BEAST. Extended Bayesian Skyline Plot (EBSP) (Heled and Drummond, 2008) were then performed for each VNI genetic cluster in BEAST v.2.3. EBSP allows to observed variations of population size through time. The mutation rate obtained from lynch equation ($4.13.10^{-9}$ per site per cell division) was used. Simulations were run for 108 iterations and were visualised in Tracer v.1.6.

5.2.8 Genome-wide association study (GWAS)

Genome-wide association studies were conducted in the program PLINK version 1.07 (Purcell et al., 2007). GWAS compares genetic variants using a case-control setup. For this study, the two groups were clinical and environmental isolates. In each case, population structure was controlled to obtain the lowest possible genomic inflation. Various methods were used to control for population structure: fineStructure clusters, Admixture clusters, phylogenetic clusters.

5.2.9 Phenotypic Characterisation

To assess potential fitness variations among different isolates which may have clinical implications, the study investigated three well-known virulence factors in vitro: thermotolerance, laccase production and polysaccharide capsule expression. An amoeba experimental model was also used to test whether genotypic variation among the *Cryptococcus* lineages was associated with differences in phagocytosis by the amoeba.

Growth at 37°C

The growth rate of cryptococcal cell was assessed after 24 hours. Prior the assay, *Cng* isolates were grown for 16 hours at 30°C at 200 rpm in yeast peptone dextrose (YPD) broth (1% yeast extract, 1% peptone, and 2% glucose; Sigma-Aldrich). Cells were counted using a haemocytometer and an inoculum of 105 cryptococcal cells was added to 5 ml of YPD broth and

incubated at 25°C and 37°C for 24 hours. The growth rate was estimated by counting cryptococcal cells. Each experiment was performed in triplicate and significance was determined using a non-parametric Wilcoxon Signed Rank Test.

Laccase Activity

Laccase production was assessed by incubating 105 cryptococcal cells in 12.5 mL of fresh yeast culture L-2,3-dihydroxphenylalanine (L-DOPA) liquid media (0.7% L-DOPA, 0.9% Vitamin B1, 3.4% Glycine, 4.2% MgSO4, 9.5% Glucose, 12.1%, KH2PO4 and 69.3% Agar) (Sabiiti and May, 2012). Cultures were grown in the dark at 37°C for 18 h then at room temperature for 24h with constant shaking at 250 rpm. Pigments secreted were measured using spectrophotometry at 450 nm. Results are expressed as a ratio and are normalised each run with laccase activity for the H99 reference strain. Each experiment was performed in triplicate and significance was determined using a non-parametric Wilcoxon Signed Rans Tests.

Capsule Expression

To promote capsule induction, *Cng* isolates were incubated for 24 h at 37°C with 250 rpm shaking. Cryptococcal cells were transferred to a minimal medium (10 mM MgSO4, 29.3 mM KH2PO4, 13 mM glycine, 3 mM thiamine) for 48 h (Mauch et al., 2013). Cell stained with India Ink were observed using an inverted microscope and the size of the cell and the capsule were measured for 100 cryptococcal cells per isolate using ImageJ (Collins, 2007) and results expressed as mean. Each experiment was performed in triplicate and significance was determined using a non-parametric Wilcoxon Signed Rank Test.

Amoeba experimental models

To test for potential variations of *Cryptococcus* lineages in their susceptibility to phagocytosis by environmental predators, we used an amoeba experimental model. Cryptococcal strains were first stained with the fungal-specific stain Calcofluor white (Sigma-Aldrich, UK). Then, fungal cells were incubated with *Acanthamoeba castellanii* for two hours. At the end of the phagocytic assay, external *Cryptococcus* were quenched with Trypan Blue (Sigma-Aldrich, UK)

(Chaka et al., 1995) and external cells were discriminated (Figure 5.1). A total of 30,000 cells were counted for each isolate using the ImageStream Imagine Flow Cytometer (Amnis®). Each experiment was performed in triplicate and significance was determined using a non-parametric Wilcoxon Signed Rank Test.

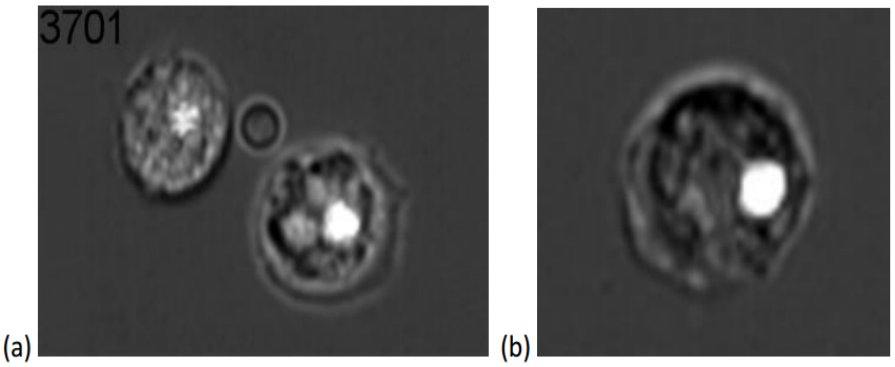


Figure 5.1: (a) Two amoeba cells containing one cryptococcal cell stained with Calcofluor white. Fungal cell quenched with Trypan Blue was not counted. (b) Cryptococcal cell phagocytosed by *A. castellanii*.

5.3 Results

The genetic differences between clinical and environmental isolates in the VNB clade were investigated. Then, using a large dataset of VNI isolates (n=155), we further explored the population structure of the VNI molecular clade. The study aimed to establish whether a particular genotype is causing disease in patients. In vitro phenotypic assays, investigating key *Cryptococcus* virulence factors, were performed to observe any differences between clinical and environmental isolates among VNI and VNB molecular types. Finally, as the current hypothesis postulates that *Cng* originated from Africa (Litvintseva et al., 2011; Simwami, 2011), I reinterpreted the 'out of Africa'hypothesis in light of NGS data.

5.3.1 VNB *Cng*: Clinical vs. Environmental isolates

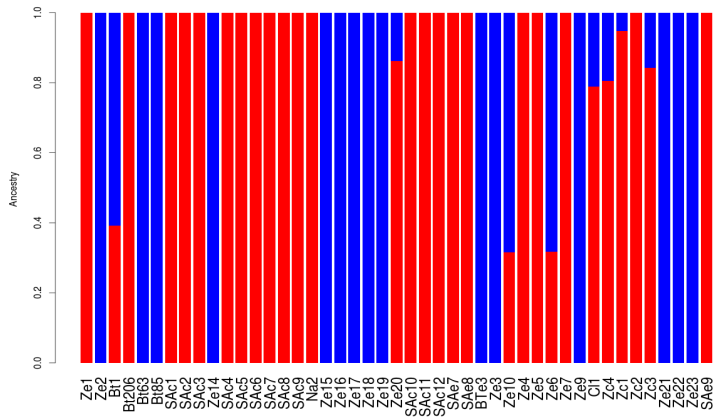
Within the VNB molecular clade, a total of 865,785 SNPs were found which represent nearly 5% of the *Cryptococcus* genome. The substantial amounts of genetic variation present in this clade enabled population genomics analyses to be undertaken in order to investigate the extent to which particular polymorphisms were associated with clinical isolates.

Phylogeny and Population Assignment

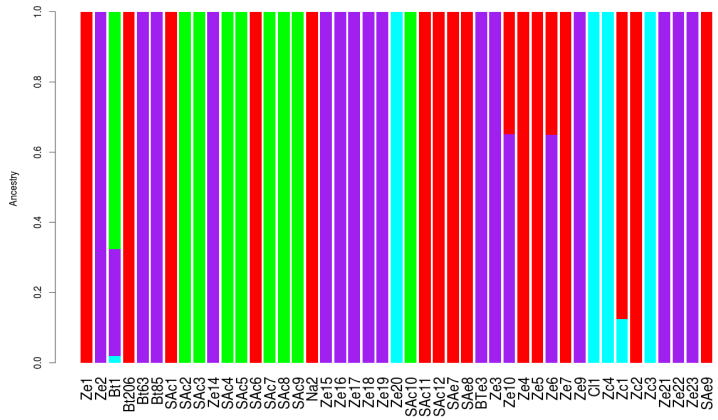
The VNB dataset contains 44 isolates. This consists of the 24 environmental isolates presented in the previous chapter and the 20 isolates collected in clinical trials in Cape Town, South Africa (n=12), Lusaka, Zambia (n=4) and in Botswana (n=4). Admixture analysis identified the two VNB sub-populations (VNBa and VNBb) described previously (Figure 5.2). The analysis also indicated the presence of three genetic clusters within the VNBa sub-population (Cluster 1, 3 and 4). These different genetic clusters are displayed on the maximum likelihood tree (Figure 5.3).

In the fineStructure analyses (Figure 5.4), the four genetic clusters can be clearly observed under both unlinked and linked models. Isolates from cluster 4 are sharing genetic materials with the other clusters. The first cluster shared genetic information with the cluster composed uniquely of clinical isolates (cluster 3). This finding indicates that this clinical cluster might have originated from cluster 1. Under the linked model, Bt1 (cluster 1) is sharing the most information with other isolates. The isolate represents one of the few Botswanan *Cng* of the dataset, suggesting a possible center of origin of the molecular clade in this country.

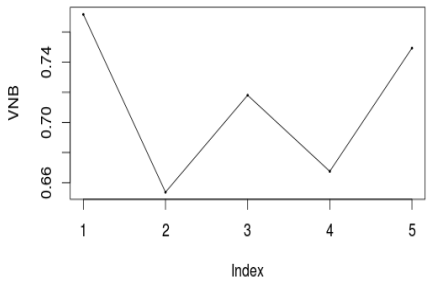
Substantial levels of population subdivision appeared between the VNB genetic clusters (F_{ST} <0.1) (Table 5.1). Each genetic cluster contained clinical isolates, however cluster 3 was uniquely composed of clinical isolates from South Africa. Environmental isolates of cluster 3 must be present in South Africa but the ecological niche of this cluster has clearly not yet been sampled. The proportion of MAT α and a between environmental and clinical isolates were similar in all clusters (Table S25). VNBa was composed of 63.3% (19/30) of clinical isolates, whereas VNBb contained two clinical isolates (13.3%), suggesting that VNBa is more virulent than VNBb (p <0,001, calculated in a t-test).



(a) VNB clade K = 2



(b) VNB clade K = 4



(c) Cross-validation

Figure 5.2: Population Structure of the VNB Molecular Clade - Population assignment was performed using Admixture (Alexander et al., 2009) - Cross-validation revealed the presence of K=2 populations.

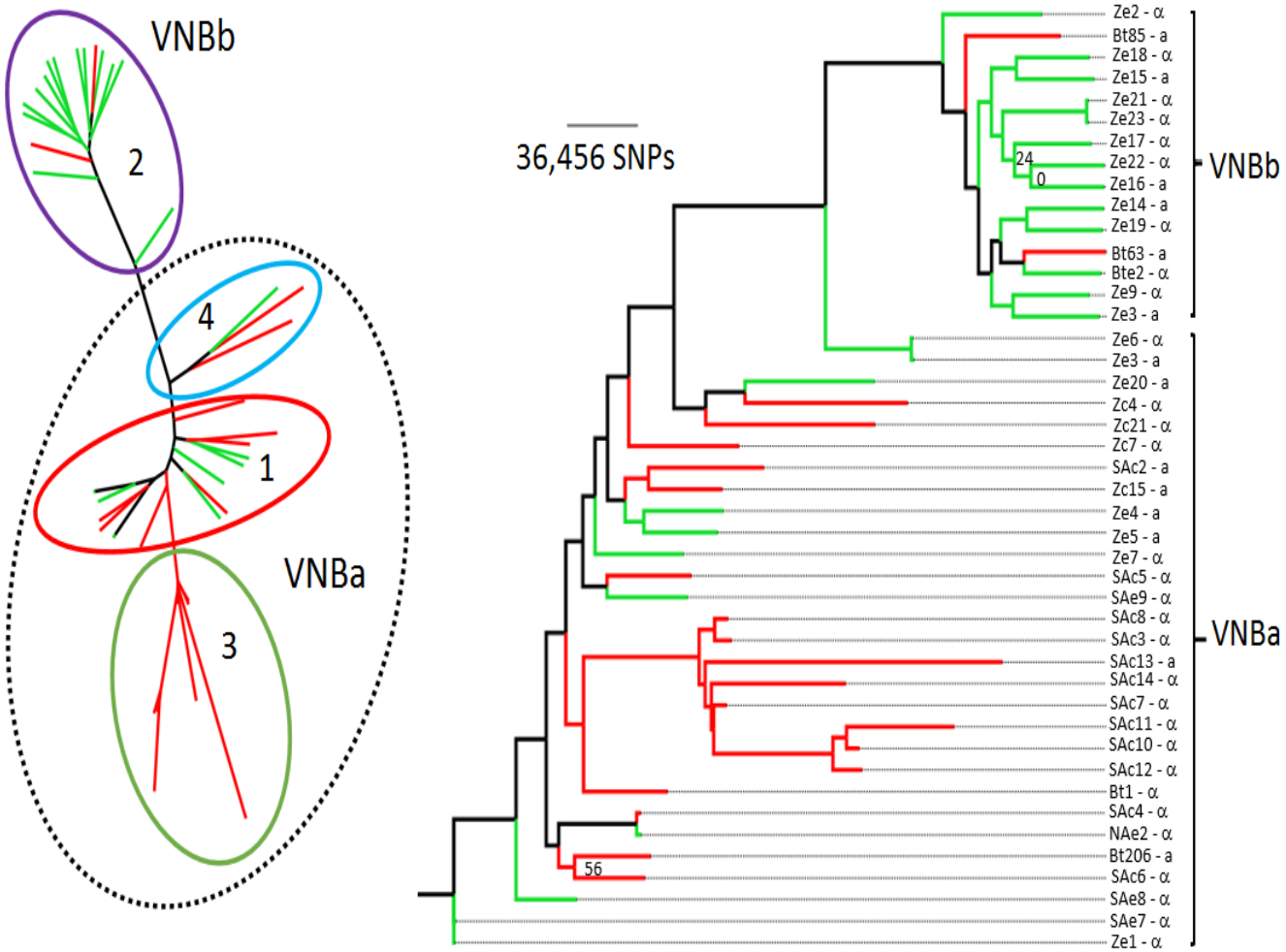
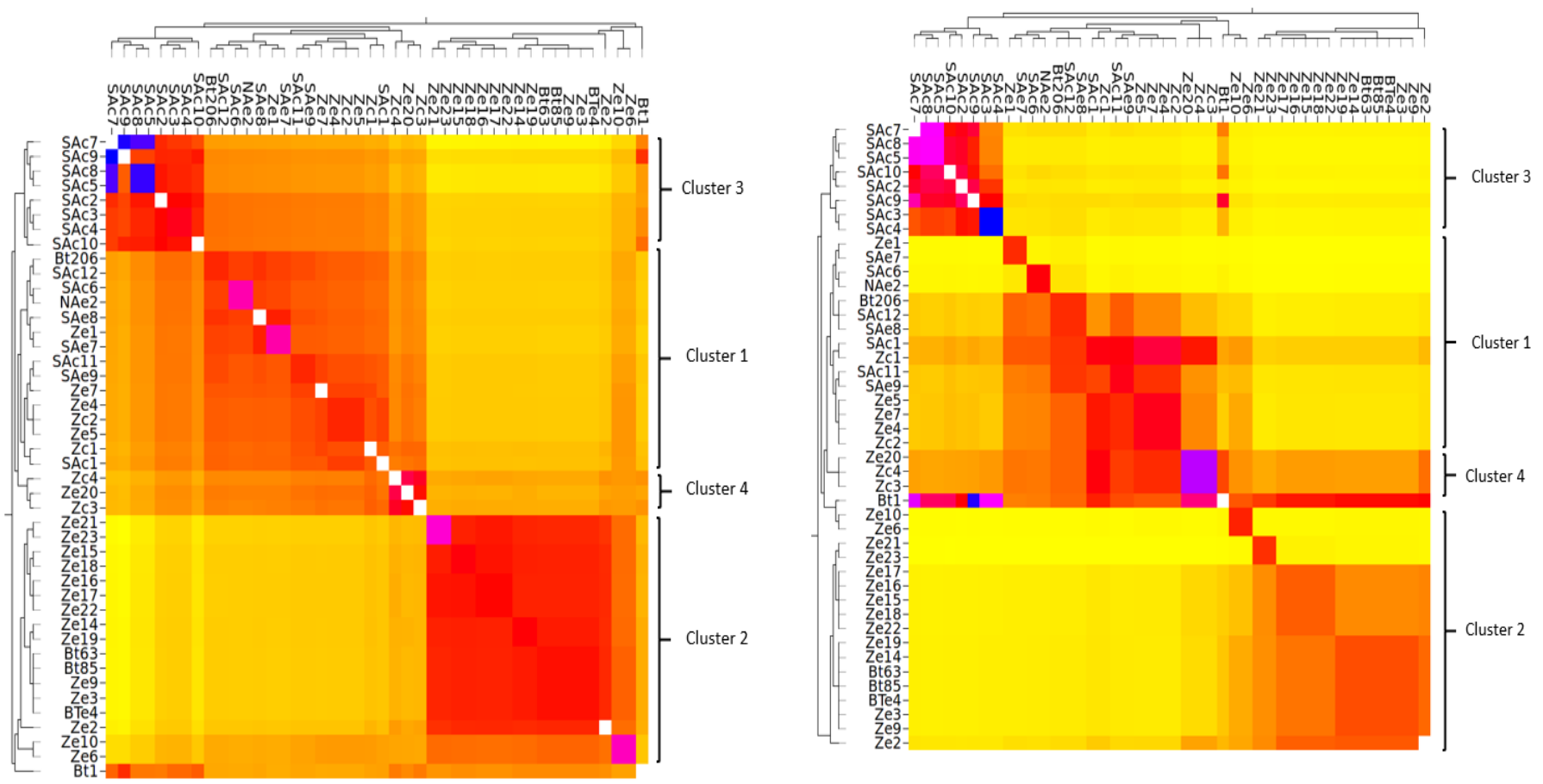


Figure 5.3: Phylogenetic Analyses of VNB genetic clusters: The VNBa population can be decomposed in three genetic clusters. The VNBb population appears as one cluster (cluster 2). Environmental isolates (green) are present in three genetic clusters, only cluster 3 is composed strictly of clinical (red) isolates. Each isolates is described as followed - Ze: Zambian environmental isolates, Zc - Zambian clinical isolates, SAe: South African environmental isolates, SAc: South African clinical isolates, NAe: Namibian environmental isolates, BT: isolates of Botswana. Mating type (a or alpha) is provided.



(a) FineStructure Analysis (Unlinked Model)

(b) FineStructure Analysis (Linked Model)

Figure 5.4: Population Structure of Clinical and Environmental *Cryptococcus neoformans* (n=44) generated using fineStructure (Lawson et al., 2012) (a) On the unlinked model, the two VNB population can be clearly observed, and on model taking into account linkage disequilibrium (b) the four clusters can be observed. The highest amount of shared genome regions between isolates appears in purple and the lowest in yellow.

Table 5.1: F_{ST} statistics among VNB clusters

F_{ST}	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0.000			
Cluster2	0.165	0.000		
Cluster3	0.131	0.161	0.000	
Cluster4	0.115	0.147	0.109	0.000

Population Genomic Analyses

The study aimed to identify loci that are specific to clinical isolates and might explain the success of these isolates in causing infection in humans. To identify genetic variants, the study performed two analyses. First, we compared clinical and environmental isolates within VNB cluster 1, then we also compared the environmental isolates of cluster 1 to isolates of cluster 3. Cluster 2 and 4 were not considered for these analyses to minimise the effect of population structure.

Population Genetic measures A total of 165,059 SNPs were shared between the four genetic clusters. The genetic diversity was higher in cluster 1 and cluster 3 compared to the other genetic clusters with higher numbers of segregating sides (S), mutations (η), external mutations (ηE) and nucleotide differences (Table 5.2). Strong negative Tajima's D values were associated with the two clusters containing clinical isolates suggesting a recent population expansion.

Table 5.2: Population Genetics Statistics of clinical and environmental *Cng* populations

Genetic clusters	S	η	ηE	π	θ	Tajima's D	D'
Cluster1 (n=16)	412,778	415,909	245,519	0.247	0.317	-1.001	-0.055
Cluster2 (n=16)	303,148	305,137	98,703	0.202	0.216	-0.273	0.136
Cluster3 (n=9)	403,355	404,620	258,083	0.308	0.369	-0.872	0.388
Cluster4 (n=4)	238,416	239,339	184,559	0.340	0.345	-0.153	0.005

Genes Associated with Clinical Isolates The population structure remains on the strong limitation of genome-scan analyses as it leads to an increase of false positive (Excoffier et al., 2009). In order to minimise the impact of population structure, clinical isolates of cluster 1

were compared to environmental isolates in cluster 1. Similarly, a comparison of cluster 1 (environmental isolates) and cluster 3 was performed.

VNBa Cluster 1 Environmental vs. Clinical Isolates Tajima's D value was lower in clinical cases (p <2.2.10^{-16}) indicating a surplus of rare alleles within clinical isolates. The Watterson's estimate (θ) was also higher in clinical isolates indicating a potentially higher mutation rate in clinical isolates. The two populations investigated were both composed of eight isolates but the numbers of segregating sites (S), the numbers of mutations and nucleotide differences (π) were higher in the clinical isolates (Table 5.6).

Table 5.3: Population Genetics Statistics of the VNBa *Cng* cluster

Population	S	η	ηE	π	θ	Tajima's D	D'
Cluster1 (n=16)	412,778	415,909	245,519	0.247	0.317	-1.001	-0.055
Environmental (n=8)	200,444	201,327	95,829	0.213	0.218	-0.125	0.060
Clinical (n=8)	375,871	378,241	255,747	0.330	0.388	-0.828	0.050

Bayescan in VNB cluster 1 BayeScan analysis analysed 412,778 SNPs and identified 98 loci with elevated F_{ST} values (Figure 5.5). Although, these loci were not significant when accounting for multiple testing (which is likely due to my relatively small sample size of n = 16), they can clearly be observed on the volcano plot (Figure 5.5b) and are indicative of localised diversifying selection (high F_{ST}). These 98 SNPs were mapped onto 62 genes, which were considered as having been under various selection forces. Five of them encoded for an hypothetical protein (Table 5.4). No loci displayed signs of purifying or balancing selection (low F_{ST} value).

Gene Content Using a large-scale blast score ratio (LS-BSR) analysis (Sahl et al., 2014) (described in Section 4.2.9), we estimated the BSR value for *in-silico* coding DNA (*is*-cDNA) sequences of the two groups. Then, we conducted a t-test while applying a Bonferroni correction to identify *is*-cDNA sequences which significantly varied between clinical and environmental isolates. Twenty-three genes in the clinical group were identified (Figure 5.6) but none of them overlapped with the regions identified using the BayeScan approach. Although most of these genes encoded for hypothetical proteins (Table 5.5), these findings combined with the phylogeny

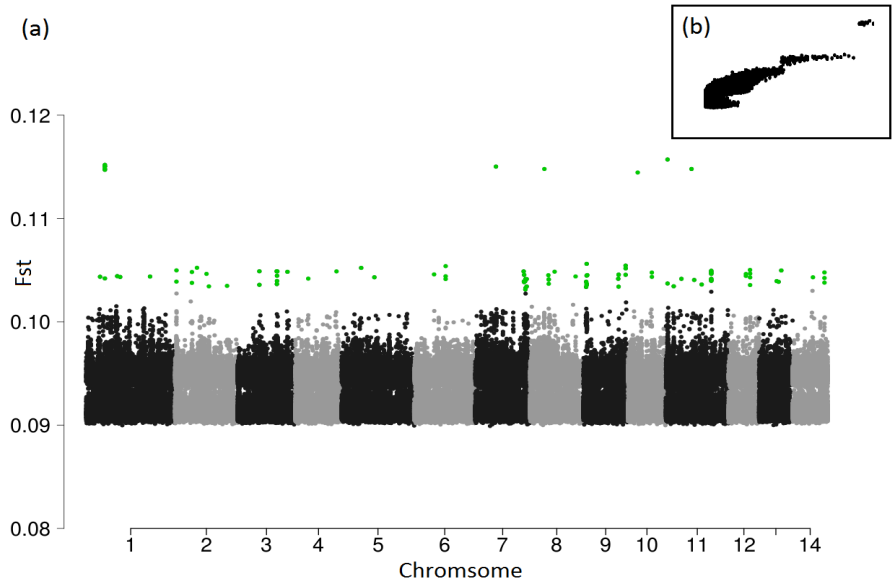


Figure 5.5: VNB Cluster 1 environmental (n=8) and clinical (n=8) isolates using a BayeScan approach - A total of 128 loci were found to have higher F_{ST} than average. These loci could be mapped to 62 genes. (b) Volcano plot of the BayeScan output showing signs diversifying selection (high F_{ST})).

Gene	chr	Start	End	Product	Ref
CNAG_03670	2	492,156	495,969	serine/threonine-protein kinase, unfolded protein response	Cheon et al. (2011)
CNAG_05301	4	981,235	985,235	deoxyuridine 5'-triphosphate nucleotidohydrolase	Liu et al. (2008)
CNAG_04119	9	58,226	63,675	rho guanyl-nucleotide exchange factor	Gerik et al. (2005)
CNAG_01850	11	1,084,661	1,090,444	two-component-like sensor kinase	Bahn et al. (2006)
CNAG_06404	13	402,643	403,473	hypothetical protein	

Table 5.4: Virulence-associated genes identified using BayeScan in the VNBa cluster 1 (n=14).

and BayeScan analysis suggest that clinical cases have a different genetic profile when compared against the environmental isolates.

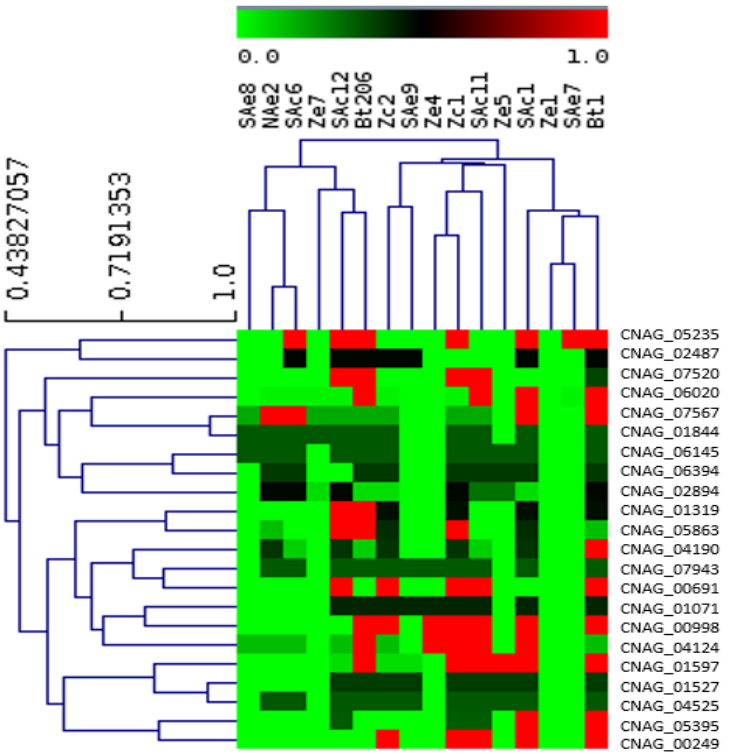


Figure 5.6: Gene differences in the VNB cluster 1 (n=16) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify gene content which is shared or differs between the clinical and environmental isolates. High homology appears in red and low BLAT gene score appears green (complete homology = 1, absence of homology = 0).

Gene	Chr	Start	End	Product
CNAG_00249	1	655,045	657,550	hypothetical protein
CNAG_07520	3	661,885	664,659	endopeptidase
CNAG_07567	3	1,225,709	1,227,025	hypothetical protein
CNAG_05235	4	802,490	804,754	protein BMH2
CNAG_02894	4	512,730	514,064	hypothetical protein
CNAG_01319	5	602,925	605,072	hypothetical protein
CNAG_01071	5	1,279,052	1,281,622	hypothetical protein
CNAG_02487	6	185,796	188,422	3-hydroxy acyl-CoA dehydratase
CNAG_05863	7	1,066,909	1,068,408	hypothetical protein
CNAG_00998	8	1,473,504	1,478,918	hypothetical protein
CNAG_04190	9	243,630	247,286	DNA topoisomerase I
CNAG_04124	9	71,261	73,828	hypothetical protein
CNAG_04525	9	1,088,971	1,091,324	glycosylphosphatidylinositol transamidase
CNAG_01844	11	1,060,621	1,063,361	tetracycline efflux protein
CNAG_01597	11	382,539	383,817	hypothetical protein
CNAG_01527	11	187,193	188,875	hypothetical protein
CNAG_06020	12	99,615	102,068	hypothetical protein
CNAG_06145	12	417,890	421,236	hypothetical protein
CNAG_06394	13	371,364	372,464	hypothetical protein
CNAG_07943	13	649,694	651,292	hypothetical protein
CNAG_05395	14	198,857	202,798	rab guanyl-nucleotide exchange factor (VAM6)

Table 5.5: Genes identified only in the clinical genes of the VNBa cluster 1

VNBa cluster 3 vs. Environmental cluster 1 isolates VNBa clinical isolates in cluster 3 displayed higher levels of genetic diversity than environmental isolates in the closely-related cluster 1 (Table 5.6). The two clusters shared 214,176 SNPs and 189,123 SNPs were unique to cluster 3 compared with 141,516 SNPs unique to environmental cluster 1 (Table S23).

Table 5.6: Population Genetics Statistics of in the VNBa *Cng* - cluster 1 environmental and cluster 3

Genetic clusters	S	η	ηE	π	θ	Tajima's D	D'
Cluster1 (n=16)	412,778	415,909	245,519	0.247	0.317	-1.001	-0.055
Cluster 1 Environmental (n=8)	200,444	201,327	95,829	0.213	0.218	-0.125	0.060
Cluster3 (n=9)	403,355	404,620	258,083	0.308	0.369	-0.872	0.388

BayeScan BayeScan analysis identified signs of purifying or balancing selection (low F_{ST}) as well as putative loci under diversifying selection (high F_{ST}) (Figure 5.7b). However, due to the large number of loci under selection (13,444 SNPs; 2.47% of the total number of SNPs under

analysis), no gene could be identified using this approach. This high number of loci observed could be due to the substantial population structure between the two clusters (F_{ST} >0.130).

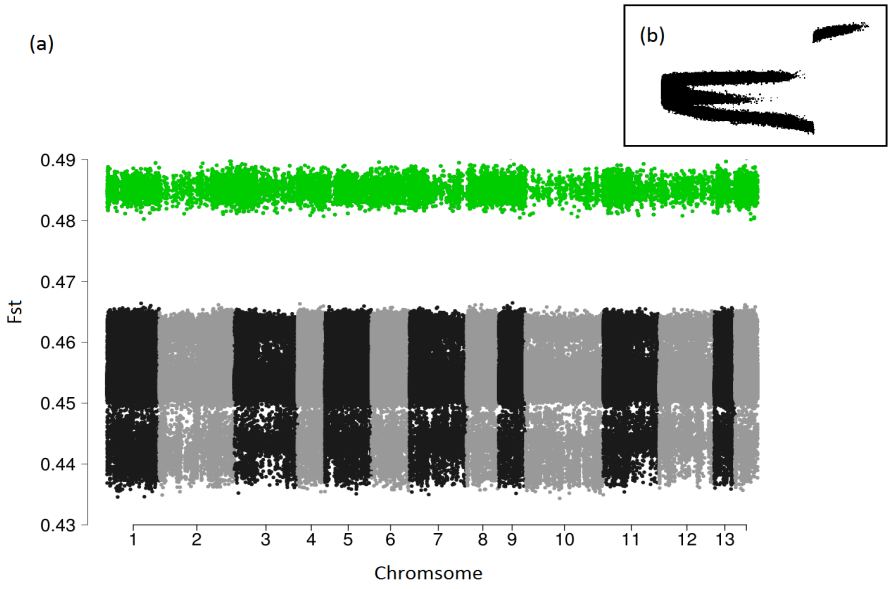


Figure 5.7: VNB Cluster 3 clinical (n=7) and VNB Cluster 1 environmental (n=7) isolates using a BayeScan approach - 13,444 SNPs have higher F_{ST} than average (highlighted in green), accounting for 2.4% all polymorphisms

Gene Content *Is*-cDNA sequences were filtered and a t-test with adjusted Bonferroni correction between the two groups was performed. A total of 40 *is*-cDNA sequences associated with the clinical cluster were identified and blasted. Figure 5.8 displays genes in cluster 3 with a high homology. A total of 11 genes were found including one gene which was found in a previous function study. CNAG_02555 encodes for the SIP402 protein which acts as a Zn2-Cys6 zinc-finger transcription factor (Liu et al., 2008). No gene overlapped with the VNBa cluster 1 analysis.

Chromosome Copy Number Variation (CCNV) Chromosome Copy Number Variations (CCNV) were detected using a 10-Kb non-overlapping sliding window. Recurrent duplicated regions were considered when 40% of isolates in each cluster possessed this duplication. Six 10-Kb regions were commonly found in the four VNBa genetic clusters investigated (cluster 1 environmental, cluster 1 clinical, cluster 2 and cluster 3) (Figure 5.9). Among the clinical isolates of cluster 1, one region was unique to these isolates encoding for five genes. Six 10-Kb regions encoding for 32 genes were uniquely found in cluster 3 (Figure 5.9). No chromosomal

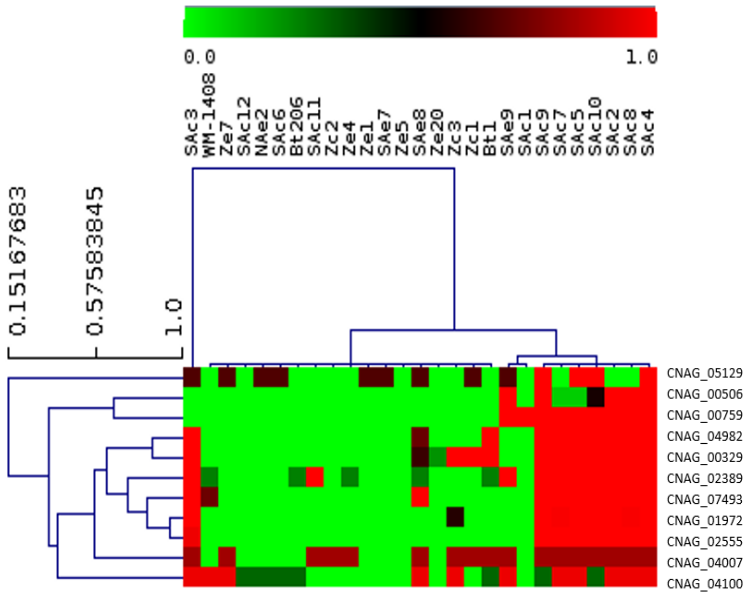


Figure 5.8: Gene differences in the VNB cluster 3 (n=9) (Sahl et al., 2014) - BLAT score ratio analysis allows to identify homologous or non-homologous gene content between the two clusters. High homology appears in red and absence of gene score appears green (complete homology = 1, absence of homology = 0).

Gene	Chr	Start	End	Product
CNAG_00506	1	1,292,586	1,294,809	hypothetical protein
CNAG_00759	1	1,995,419	1,997,498	hypothetical protein
CNAG_00329	1	841,719	843,189	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 8
CNAG_07493	2	1,526,737	1,528,519	hypothetical protein
CNAG_04007	2	1,384,050	1,386,279	septin ring protein
CNAG_05129	4	500,541	502,358	hypothetical protein
CNAG_04982	4	142,196	144,512	putative cytosine-purine permease (FCY3)
CNAG_02389	6	453,490	458,322	AGC protein kinase
CNAG_02555	6	8,039	11,193	Zn2-Cys6 zinc-finger transcription factor (SIP402)
CNAG_04100	9	1,159	5,043	hypothetical protein
CNAG_01972	11	1,406,904	1,411,072	C2H2 zinc finger protein (TAF10)

Table 5.7: Genes identified only in the clinical genes of the VNB cluster 3

duplication was observed in the environmental isolates (cluster 1 environmental and cluster 2, see Figure S12 which displays coverage variations across each genome). In each cluster, no recurrent deletion was found.

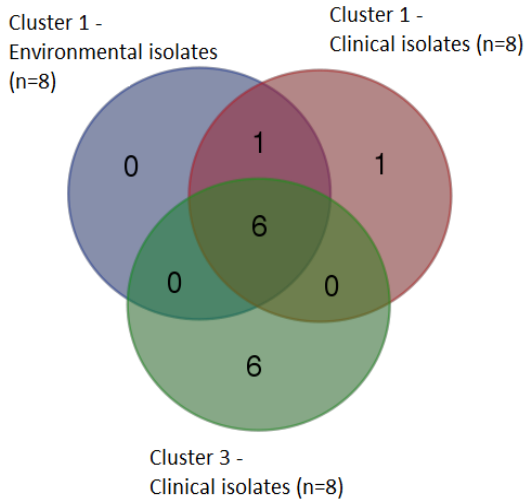
No chromosome was duplicated or deleted in our environmental VNB dataset. However, 70% of clinical VNB isolates had a chromosome duplication. In clinical isolates of cluster 1, ZC07 presented with a duplication of chromosome 14. ZC15 and SAc6 had two copies of chromosome 12. Cluster 3 (Figure S13) had a high number of chromosome duplications and deletions. Every isolate had a CCNV except SAc12 which had a deletion of chromosome 10. Five isolates in the clinical clade had chromosome 1 duplicated, SAc14 had two copies of chromosome 2. Four isolates had the first half of a chromosome duplicated and the second half was depleted. Low Average Read-Depth was observed in chromosome 3 for three isolates suggesting a potential chromosomal deletion. Chromosome 4 was duplicated once (SAc11). SAc8 and SAc9 were diploid for chromosome 5, 6 and 7. SAc11 had a CCNV on chromosome 9. Chromosome 10 was depleted in three isolates and duplicated in 1, while chromosome 12 was duplicated once (SAc7) and deleted once (SAc13). Chromosome 13 was duplicated only in SAc13.

The duplication of chromosome did not show any particular pattern. Previous research affirmed that the duplication of chromosome 12 and 13 in mice played an important role in virulence (Hu et al., 2008). The genes present in the 10Kb-regions investigated were not known to have any virulence-associated function. One virulence gene was found among the 64 genes in cluster 3 (CNAG_03202) and one was identified in clinical isolates of cluster 1 (CNAG_05431).

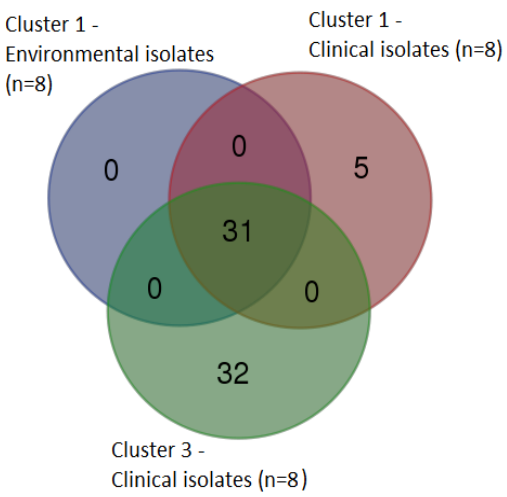
CCNV appears to be a key response and determinant to invade the host. Some regions seem to be naturally present in multiple copies but the duplication of an entire chromosome is likely to be a response to stress since aneuploidy is more commonly observed in clinical isolates.

Investigating signs of selection

Frequency-based methods In this section, we looked for signs of balancing selection and/or regions of sudden population contraction in clinical isolates (genomic regions with elevated Tajima's D value). Generally, studies target regions of short-term selective sweeps (low D value). Using our genome scan approach in our clinical dataset, the majority of 100-bp windows displayed low Tajima's D value (average = -1.001 (Table 5.6, Figure S11)), therefore, this



(a) 10-kb Regions with CCNV



(b) Genes with CCNV

Figure 5.9: Chromosomal regions and genes with duplications in each cluster - Each region represent a 10-kb segment which was found in multiple copies among at least 40% of individuals in each cluster.

section focused on identifying targets of long-term balancing selection (Thomas et al., 2012). Balancing selection is a selective process in which alleles are actively maintained in a population at higher frequencies than expected under the action of genetic drift. Signatures of balancing selection are known to have the potential to highlight polymorphisms which have long-term fitness consequences (Fumagalli et al., 2009). Non-overlapping windows with elevated Tajima's D were investigated in the clinical clusters as originally performed by Hohenlohe et al., (2010).

In the VNBa clinical cluster 1 (n=8), 218 genes were found in 100-bp width regions with elevated Tajima's D. 16 of them encoded for virulence-associated genes (Table 5.8). In cluster 3, due to high number of CCNV events, population genomic statistics were strongly biased. On figure 5.13b, large values can be observed for the different summary statistics on chromosome two, three, seven, ten and twelve. The regions where chromosomes were duplicated were removed and two alpha-specific genes were identified as having been under balancing selection: CNAG_01454, and CNAG_06813 (Table 5.8).

The density of SNPs was very high in regions where chromosomes are duplicated or depleted (figure 5.13). These observations were associated with strong negative and positive Tajima's D (Figure S10c). These observations were not found in the clinical cluster 1 (Figure 5.12), which only contained two CCNVs. Tajima's D values were generally more negative in the clinical clusters indicative of stronger positive and/or purifying selection within these isolates.

VNBa	Chr	Start	End	Product	Ref
Cluster 1 (n=8)					
CNAG_00091	1	253,176	255,431	hypothetical protein (2256 nt)	
CNAG_00097	1	268,618	271,508	myo-inositol transporter (ITR1)	Xue et al. (2010)
CNAG_00864	5	1,793,503	1,796,091	myo-inositol transporter (ITR2)	Xue et al. (2010)
CNAG_00865	5	1,790,294	1,791,427	putative O-acetyltransferase (CAS92)	Moyrand et al. (2007)
CNAG_00867	5	1,782,123	1,784,803	myo-inositol transporter (ITR3A)	Xue et al. (2010)
CNAG_01444	5	295,676	296,599	mRNA polymerase II subunit (RPB4)	Bloom et al. (2013)
CNAG_01445	5	291,364	295,513	autophagy-related protein (APG9)	Lengeler et al. (2002)
CNAG_01446	5	290,468	291,329	heat shock protein, 12Kda (HSP12)	Maeng et al. (2010)
CNAG_01454	5	271,944	275,730	ste-like transcription factor (STE12alpha)	Lengeler et al. (2002)
CNAG_02048	6	1,334,775	1,337,354	proline oxidase (PUT5)	Lee et al. (2013)
CNAG_02049	6	1,331,959	1,334,268	proline oxidase (PUT1)	Lee et al. (2013)
CNAG_02057	6	1,315,191	1,316,660	urease accessory protein (URE6)	Singh et al. (2013)
CNAG_02476	6	213,494	217,758	putative transcription factor (YRM101)	Liu et al. (2008).
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)	Lengeler et al. (2002)
CNAG_06814	5	185,700	188,859	alpha cell-type homeodomain transcription factor (SXI1alpha)	Hull et al. (2002)
CNAG_07408	5	230,211	233,167	ste/ste20/paka protein kinase (STE20alpha)	Wang et al. (2002)
Cluster 3 (n=9)					
CNAG_01454	5	271,944	275,730	ste-like transcription factor (STE12alpha)	Lengeler et al. (2002)
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)	Lengeler et al. (2002)

Table 5.8: Virulence-associated genes under balancing selection in the VNBa clinical isolates (Tajima's D)

Gene-based methods To assess the presence of mutations in virulence genes, we used MK or dN/dS (ω) tests (as described in the Methods 4.2.10). In the VNBa clinical isolates of cluster 1, 187 and 213 genes were found using the MK tests and dN/dS statistic respectively. Genes were considered only when they could be identified in 40% of isolates within the 0.5% quantile. Among the isolates of the VNBa clinical cluster 3, 135 and 193 genes including 12 and five virulence genes were found using the two tests. In environmental cluster, 163 genes including 17 virulence ones were identified using MK tests and 205 genes including six virulence genes using dN/dS statistic (Table 5.9). As shown on Figure 5.10, 12 of the virulence genes found in

the investigated groups overlapped. Interestingly, five genes were uniquely found in the clinical cluster 1 and seven could only be observed in Cluster 3 (Table 5.9).

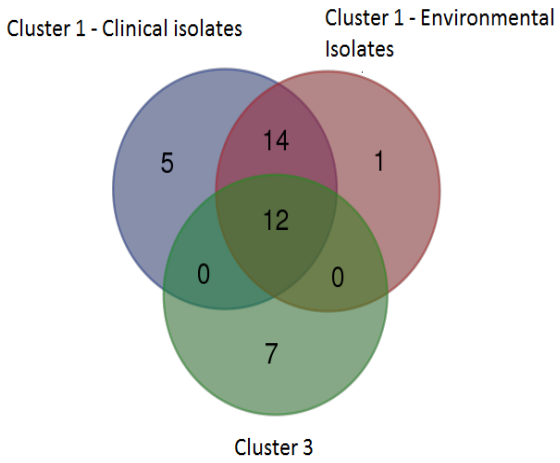


Figure 5.10: VNBa virulence-associated genes identified using gene-based methods

Table 5.9: Virulence-associated Genes under positive selection in VNBa clinical isolates

Gene	Statistic	Chr	Start	End	Product	Literature
Cluster 3 - Clinical		7 genes				
CNAG_07593	ω	11	436,513	439,198	bZip transcription factor (YAP4)	Missall and Lodge (2005)
CNAG_04981	ω	4	139,071	141,901	catalase 1 similar spore-specific catalases (CAT1)	Giles et al. (2006)
CNAG_06396	MK	13	374,705	375,632	hypothetical protein	
CNAG_00865	MK	5	1,790,294	1,791,427	putative O-acetyltransferase (CAS92)	Moyrand et al. (2007)
CNAG_05538	ω	14	571,459	575,067	DNAj-like cochaperone (JJJ1)	Liu et al. (2008)
CNAG_04982	MK	4	142,196	144,512	putative cytosine-purine permease (FCY3)	Song et al. (2012)
CNAG_01475	ω	11	57,180	59,170	sphingolipid C9-methyltransferase (SMT1)	Rhome et al. (2007)
Cluster 1 - Clinical		5 genes				
CNAG_00654	MK	1	1,705,106	1,706,138	sulfiredoxin (SRX1)	Ko et al. (2009)
CNAG_00795	MK	1	2,097,928	2,099,353	hyphal flocculin (CFL1)	Wang et al. (2012)
CNAG_05882	MK	7	1,105,565	1,108,397	class e vacuolar protein-sorting machinery protein (HSE102)	Liu et al. (2008)
CNAG_06064	MK	12	207,653	208,840	protein tyrosine phosphatase (PTP1)	Lee et al. (2014)
CNAG_07724	ω	8	721,568	725,406	copper-sensing transcription factor (CUF1)	Ory et al. (2004)
Cluster 1 -Environmental		1 gene				
CNAG_04714	ω	10	608,465	611,057	MFS multidrug transporter protein	

Overlap between Genes Previously, we described the genes associated with clinical isolates. These putative genes appeared as candidates which may play a role in the ability of *Cryptococcus* to establish infection within human. In order to confirm the role of the genes candidates listed above, gene expression experiments mimicking stressful conditions are required to validate their role in pathogenicity.

In chapter 4, we identified putative virulent gene candidates undergoing recent selective sweeps or under purifying selection pressures. 19 of these genes were also found when investigating clinical isolates (Table 5.11) and can be considered as strong gene candidates which may play a role in pathogenesis. For example, CNAG_04982 (fcγ3) was detected in environmental VNB isolates using the LS-BSR analysis (Figure 4.9) and was also found to be under positive selection using MK test. Three gene candidates (CNAG_06814, CNAG_06813, and CNAG_07408) were located on the MAT α locus, emphasising the likely advantage of α cryptococcal cell in establishing infection. Three myo-inositol transporters were found (*ITR1*, *ITR2* and *ITR3A*) which were located on different chromosomes. Myo-inositol transporters play a role in carbon utilisation in the CNS (Xue et al., 2010) (Table 4.5).

The presence of two capsular related genes, CNAG_06813 and CNAG_00865, indicates possible variations in the induction of the polysaccharide capsule between the different MAT α/a isolates. One gene was involved in response to temperature (CNAG_07408). Wang et al., (2002) found that STE20α mutants became temperature sensitive. These differences between clinical and environmental isolates indicate that clinical isolates might be more adapted to succeed within the human host. The genetic differences might encode for variations which might affect the phenotype of cryptococcal cells. These potential phenotypic differences will be tested 'in vitro'in Section 5.3.3.

Only genes with known virulence functions are listed below (Tables 5.9 & 5.11) but other genes undergoing selection have been uncovered (Table S26). Additional research, such as RNA expression, should be performed to investigate further the role of these gene candidates. A comparison of a panel of environmental and clinical isolates under stressful conditions would help to understand the role of these putative genes under selection.

Table 5.11: Virulence-associated genes associated with Clinical and Environmental VNB isolates

Gene	Chr	Start	End	Product	Description	Ref
CNAG_00654	1	1,705,106	1,706,138	sulfiredoxin (*SRX1*)	Involved in the HOG pathway, SRX1 appears to be a central element in the response to oxidative stress.	(Ko et al., 2009)
CNAG_00795	1	2,097,928	2,099,353	hyphal flocculin (*CFL1*)	CFL1 is cell adhesion property. It is Znf2 downstream factor. Znf2 plays a role in morphological switch. CFL1 controls biofilm formation, cell adhesion, morphogenesis and plays a role in virulence.	(Wang et al., 2012)
CNAG_00097	1	268,618	271,508	myo-inositol transporter (*ITR1*)	Myo-inositol transporter are involved in the utilisation of myo-inositol as a carbon source in the human central nervous system and are key components of *Cng* pathogenicity.	(Xue et al., 2010)
CNAG_00306	1	787,724	789,157	metallothionein protein (*CMT1*)	metallothionein protein help in copper (Cu) acquisition. CNAG_00306 encodes for g cysteine- rich Cu binding and detoxifying protein which plays a role in the response to increase levels of Cu.	(Ding et al., 2013)
CNAG_04982	4	142,196	144,512	putative cytosine-purine permease (*FCY3*)	The cytosine deaminase genes are thought to be involved in flucytosine uptake. The gene was previously found in a transcriptome analysis investigating flucytosine susceptibility in *C. neoformans*.	(Song et al., 2012)
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (*CAP1*α)	The MAT α specific gene encodes for the O-glucosyltransferase and plays a role in virulence by promoting haploid fruiting.	(Lengeler et al., 2002)
CNAG_00867	5	1,782,123	1,784,803	myo-inositol transporter (*ITR3A*)	myo-inositol transporter are involved in the utilisation of myo-inositol as a carbon source in the human central nervous system and are key components of *Cng* pathogenicity.	(Xue et al., 2010)
CNAG_00865	5	1,790,294	1,791,427	putative O-acetyltransferase (*CAS92*)	CAS92 is involved in the acetylation of glucuronoxy-lomannan (GXM) which represents 88% of the capsule mass. The GXM structure helps in the pathophysiology of the infection. The de-O-acetylation of the GXM has been found to increase virulence in *Cn.* var. *neoformans*.	(Janbon et al., 2001) (Moyrand et al., 2007)
CNAG_07408	5	230,211	233,167	ste/ste20 protein kinase	The STE20 alpha-specific protein kinase, also called VSP4 is centrol in *Cryptococcus* virulence and is involved in temperature regulation, differentiation, cytokinesis, mating and virulence. ; ste20a *Cng* mutants were found to temperature-sensitive. Virulence was also attenuated with the survival of 80% of infected mice 37C after infection. A protein defect would also impair capsule production in vivo.	(Wang et al., 2002)
CNAG_00864	5	1,793,503	1,796,091	myo-inositol transporter (ITR2)	myo-inositol transporter are involved in the utilisation of myo-inositol as a carbon source in the human central nervous system and are key components of *Cng* pathogenicity.	(Xue et al., 2010)
CNAG_06814	5	185,700	188,859	SXI1α	SXI1α is a a-specific regulator and control the sexual cycle.	(Hull et al., 2002)
CNAG_02049	6	1,331,959	1,334,268	proline oxidase (PUT1)	PUT1 is a proline oxidase. The degradation of proline is associated with mitochondrial oxidative respiration. Effective proline assimilation is essential for the protection aginst oxidative and nitrosactive stresses.	(Lee et al., 2013)
CNAG_02057	6	1,315,191	1,316,660	urease accessory protein (URE6)	Ure6 is an accessory protein which interacts with URE1 which controls the response to nitrogen. Deletion of Ure6 reduce the invasion of the brain	(Singh et al., 2013)
CNAG_05882	7	1,105,565	1,108,397	vacuolar protein-sorting machinery - HSE102	HSE102 regulate melanine production. The protein is a class E vacuolar protein-sorting machinery protein.	(Liu et al., 2008)
CNAG_05939	7	1,266,326	1,268,709	hypothetical protein	CNAG_05939 encodes for a hypothetical protein, but the latter was observed with a >log2 fold change downregulated expression in the cap10 strain compared to the wild-type KN99a. The disruption of the capsule *CAP10* gene leads to the non-expression of the capsular polysaccharides on the cell surface.	(Tefsen et al., 2014)
CNAG_07724	8	721,568	725,406	copper-sensing transcription factor (*CUF1*)	A necessary protein in the reponse to copper by the regulation of Cuf1p-mediated transcription of a high-affinity copper transporter (Ctr4p).	(Ory et al., 2004)
CNAG_06064	12	207,653	208,840	protein tyrosine phosphatase (PTP1)	Ptp1 is a complementary protein tyrosine phosphatase (PTP) to control some stress responses. It serves as a negative feedback regulator of the mitogen-activated protein kinase (MAPK) signaling cascades controlled by the Hog1 and the transcription factor Atf1.	(Lee et al., 2014)
CNAG_05395	14	198,857	202,798	rab guanyl-nucleotide exchange factor (*VAM6*)	VAM6 has been found to be downregulated in mutants in different two studies. The gene encodes for a homolog of the RabGEF in Saccharomyces cerevisiae. VAM mutants shows small vesicles and mature forms of the vacuolar proteins. Another study found *VAM6* mutants to have poor our CSF survival.	(Liu et al., 2008) (Tseng et al., 2012) (Griffiths et al., 2012)

Genes	Chr	Environmental Populations		Clinical Populations		Product	Reference
CNAG_00654	1	MK Test	VNB (n=24)	MK Test	Clinical Cluster 1 (n=8)	sulfiredoxin (SRX1)	(Ko et al., 2009)
CNAG_00795	1	MK Test	VNB (n=24), VNBa & VNBb	MK Test	Clinical Cluster 1 (n=8)	hyphal flocculin (CFL1)	(Wang et al., 2012)
CNAG_00097	1	Tajima's D	VNBa (n=9)	Tajima's *D*	Clinical Cluster 1 (n=8)	myo-inositol transporter (ITR1)	(Xue et al., 2010)
CNAG_00306	1	D'& *dN/dS*	VNB (n=24)	BayeScan	Clinical Cluster 1 (n=8)	metallothionein protein (CMT1)	(Ding et al., 2013)
CNAG_04982	4	Gene Content	VNB (n=24)	Gene content & MK test	Clinical Cluster 3 (n=8)	putative cytosine-purine permease (FCY3)	(Song et al., 2012)
CNAG_06813	5	Lewontin D'& *dN/dS*	VNB (n=24)	Tajima's D	Clinical Cluster 1 (n=8) & 3 (n=8)	O-glucosyltransferase (CAP1α)	(Lengeler et al., 2002)
CNAG_00867	5	MK Test	VNB (n=24), VNBa & VNBb	Tajima's D	Clinical Cluster 1 (n=8)	myo-inositol transporter (ITR3A)	(Xue et al., 2010)
CNAG_00865	5	LFMM	VNB (n=24)	Tajima's D & MK Test	Clinical Cluster 1 (n=8) & 3 (n=8)	putative O-acetyltransferase (CAS92)	(Moyrand et al., 2007)
CNAG_07408	5	MK Test & Lewontin D'	VNB (n=24), VNBa (n=9)	Tajima's D	Clinical Cluster 1 (n=8)	ste/ste20/paka protein kinase (STE20α)	(Wang et al., 2002)
CNAG_00864	5	Lewontin D'	VNBa (n=9)	Tajima's D	Clinical Cluster 1 (n=8)	myo-inositol transporter (ITR2)	(Xue et al., 2010)
CNAG_06814	5	Lewontin D'	VNBa (n=9)	Tajima's D	Clinical Cluster 1 (n=8)	homeodomain transcription factor (SXI1α)	(Wang et al., 2002)
CNAG_02049	6	MK Test	VNB (n=24), VNBa (n=9)	Tajima's D	Clinical Cluster 1 (n=8)	proline oxidase (PUT1)	(Lee et al., 2013)
CNAG_02057	6	MK Test	VNBb (n=14)	Tajima's D	Clinical Cluster 1 (n=8)	urease accessory protein (URE6)	(Singh et al., 2013)
CNAG_05882	7	MK Test	VNB (n=24), VNBa (n=9)	MK Test	Clinical Cluster 1 (n=8)	vacuolar machinery protein (HSE102)	(Moyrand et al., 2007)
CNAG_05939	7	LFMM	VNB (n=24)	BayeScan	Clinical Cluster 1 (n=8)	hypothetical protein	
CNAG_07724	8	dN/dS	VNBa (n=9)	dN/dS	Clinical Cluster 1 (n=8)	copper-sensing transcription factor (CUF1)	(Ory et al., 2004)
CNAG_06064	12	MK Test	VNB (n=24)	MK Test	Clinical Cluster 1 (n=8)	protein tyrosine phosphatase (PTP1)	(Lee et al., 2014)
CNAG_05395	14	LFMM	VNB (n=24)	BayeScan & Gene Content	Clinical Cluster 1 (n=8)	rab guanyl-nucleotide exchange factor (VAM6)	

Table 5.13: Genes under positive selection in environmental population and of interest in clinical isolates

GWAS Despite the limited number of isolates and the high number of SNPs, we attempted Genome-Wide Association Studies (GWAS) in order to identify loci associated with clinical isolates when compared against environmental isolates for the four genetic clusters. The inflation factor (λ_{GC}) is used as a test statistic to account for potential genomic inflation which can arise from population stratification. Under a completely null distribution, the inflation factor is expected to be equal to 1 (Smith and Newton-Cheh, 2009). The best inflation factor we could obtain was 2.694. The study was taking into account the four different clusters (K=4) but no locus was significant after accounting for multiple testing. The most significant locus was part of hypothetical protein on chromosome 3 with an adjusted p value of $6.404.10^{-3}$ (Figure 5.11). I therefore conclude that additional VNB isolates are required to obtain successful results using a GWAS approach.

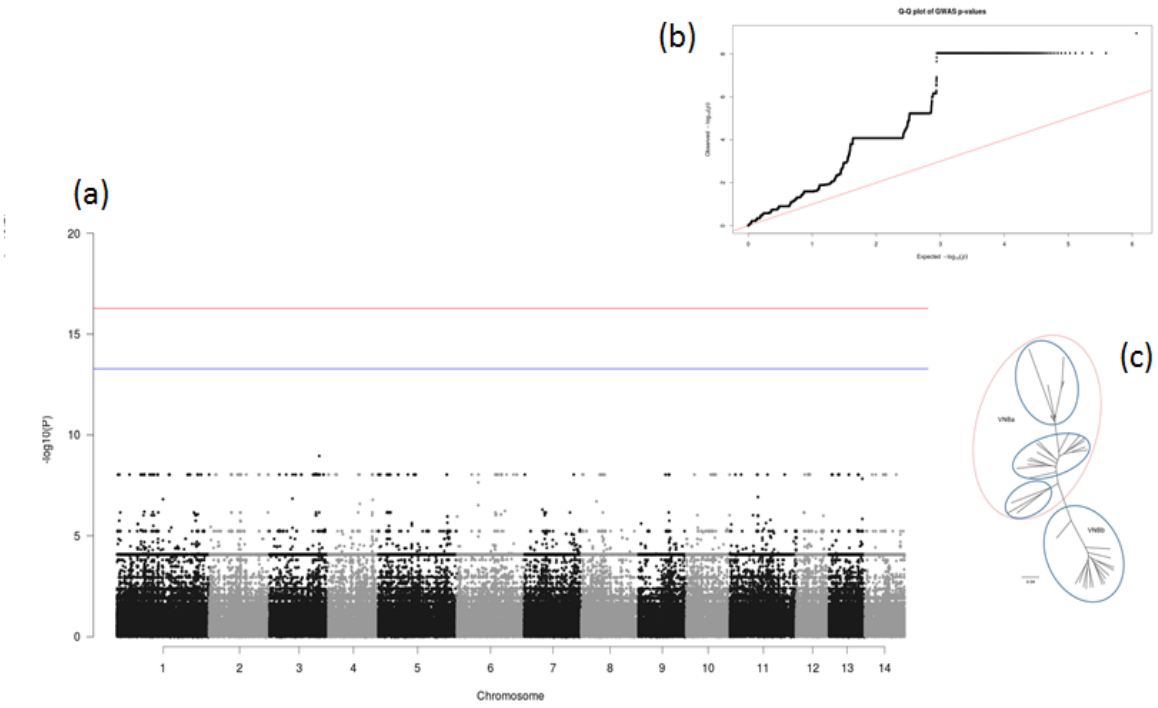
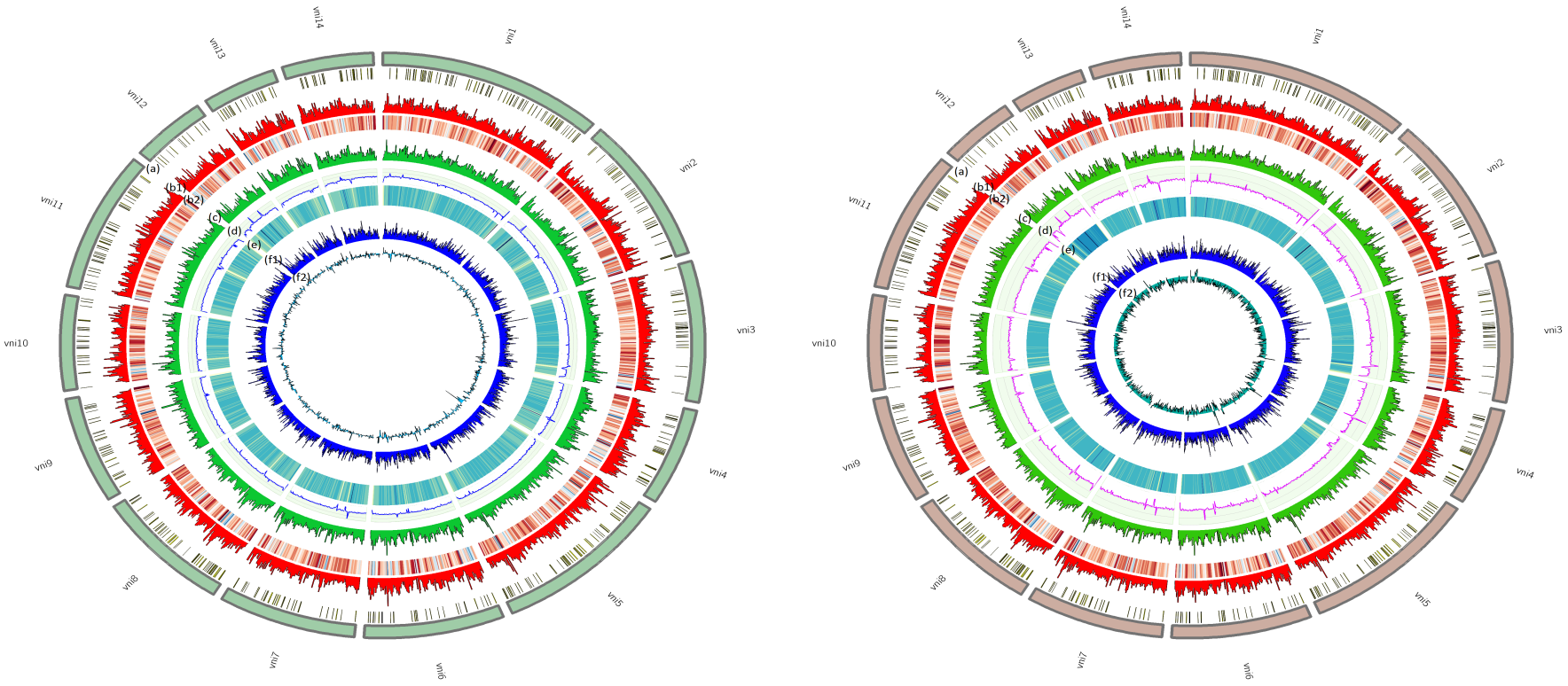


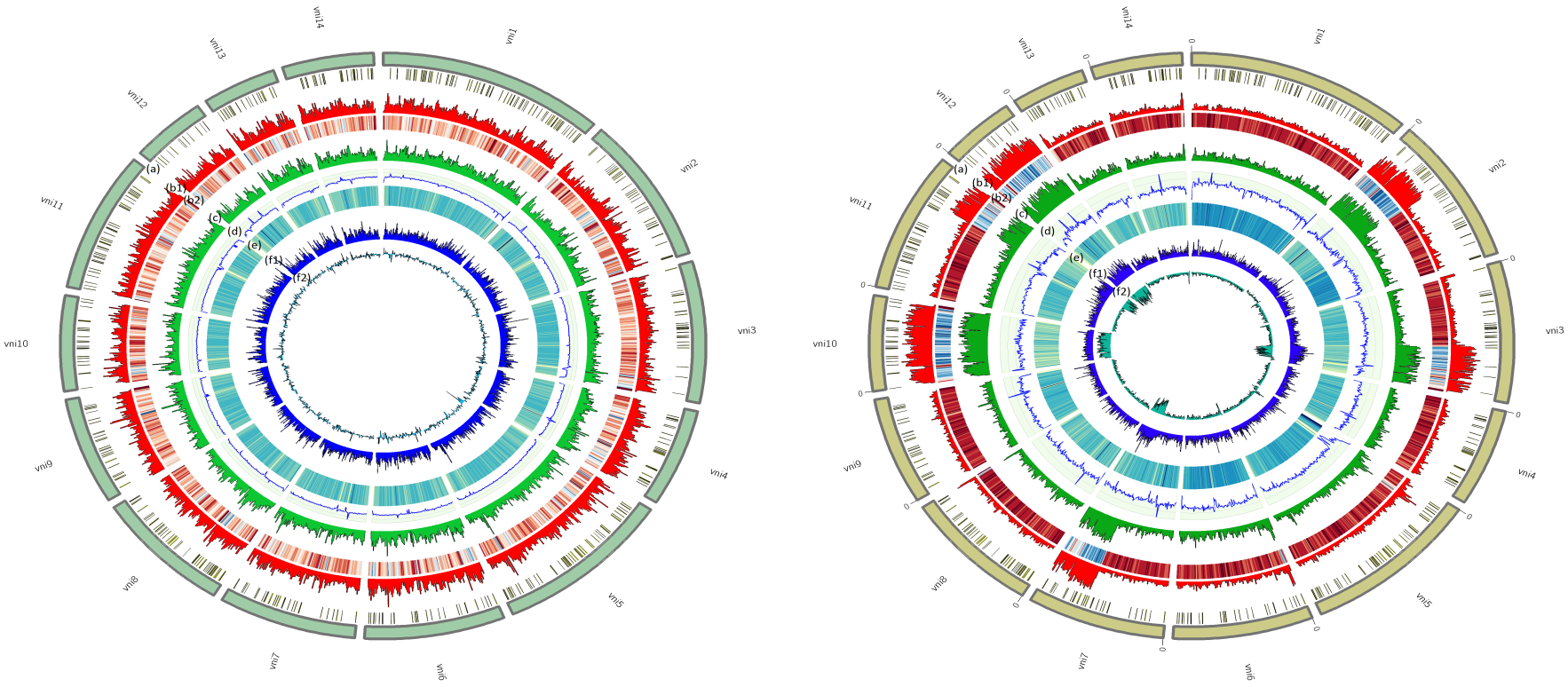
Figure 5.11: Genome-Wide Associations Study using the VNB isolates (n=44) (a) -log(p) value for each loci over the genomes. The red and blue line indicate the significance level after applying a Bonferroni correction at 0.05 and 1 respectively. (b) Q-Q plot of GWAS p value (c) The phylogeny of the different genetic clusters used in the analyses (n=4).



(a) Environmental Isolates (n=8)

(b) Clinical Isolates (n=8)

Figure 5.12: Population Genetics analyses of VNB isolates cluster 1 (environmental and clinical isolates) across the genome over a 10kb non-overlapping sliding window. The position of significant virulence-associated genes (in the 0.1% quantile) is highlighted in red on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of F_{ST} statistic with VNB cluster 1 (environmental isolates); (d) Line plot and heat map of the average coverage (e) line plot of the LDhat rho statistic (f) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue.



(a) VNBa cluster 1 environmental Isolates (n=8)

(b) VNBa cluster 3 clinical Isolates (n=8)

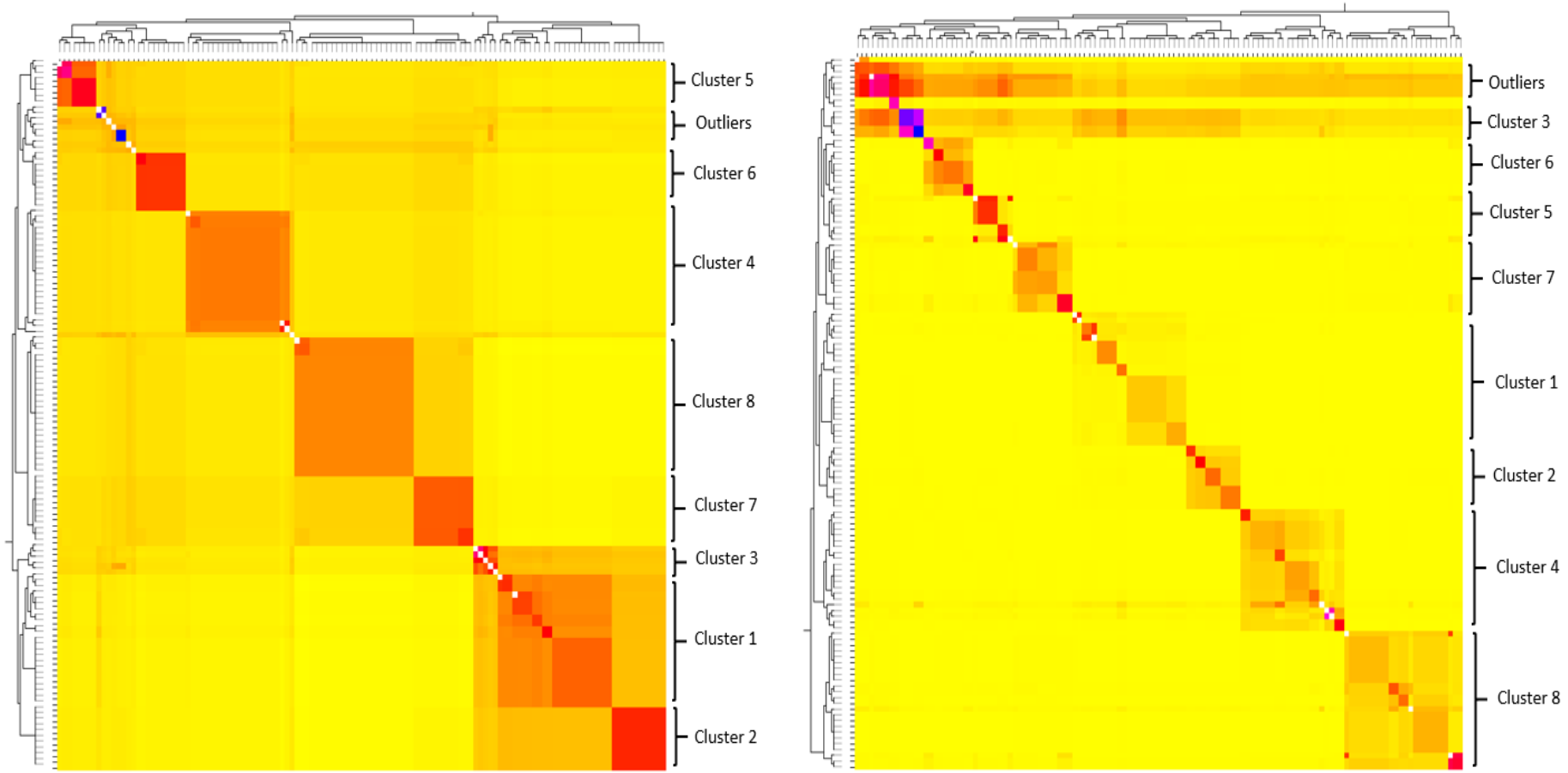
Figure 5.13: Population Genetics analyses of VNB clinical isolates (cluster 3) across the genome over a 10kb non-overlapping sliding window. The position of significant virulence-associated genes (in the 0.1% quantile) is highlighted on the chromosome ideogram. These genes were identified using Tajima's D, LD', ω and MK statistics - legend: from the outside to the inside - (a) Virulent *Cryptococcus neoformans* genes identified in the literature (n=470), (b) SNPs density histogram and heatmap ; (c) histogram of F_{ST} statistic between the two populations; (d) Line plot and heat map of the average coverage (e) line plot of the LDhat rho statistic (f) heatmap and histogram Tajima's D ; The position of virulent genes under selection is highlighted in blue.

5.3.2 The *Cryptococcus neoformans* VNI Molecular Type

Phylogeny and Population Assignment

In our dataset, we identified 155 VNI genomes. Twenty-six were environmental African genomes described in the previous chapter, 80 were clinical African genomes and 49 VNI genomes were collected in Asia (20 environmental and 29 clinical genomes). After removing outliers, 150 genomes clustered into eight strongly supported genetic clusters (Figure 5.15) (n_{C1}=26, n_{C2}=17, n_{C3}=5, n_{C4}=30, n_{C5}=8, n_{C6}=13, n_{C7}=15, n_{C8}=36). These results were supported by an Admixture analysis (Figure S15). Cluster 1, 2 and 3 were composed strictly of African isolates, whereas cluster 5 contained only Asian isolates. The other genetic clusters were composed of a mixture of isolates from both continents.

On the fineStructure analysis, due to computational limitations only a representative panel of 125 genomes could be included in the analysis (Figure 5.14). Using both the unlinked and linked model, the eight clusters could be identified. In the model taking into account linkage disequilibrium (Figure 5.14b), isolates of cluster 3 seem to be sharing genomic regions with other African clusters indicating that these clusters have a shared evolutionary history within this continent.



(a) FineStructure Analysis (Unlinked Model)

(b) FineStructure Analysis (Linked Model)

Figure 5.14: Population Structure of Clinical and Environmental *Cryptococcus neoformans* VNI (n=122) (Lawson et al., 2012) - Under both unlinked (a) and linked (b) model, the eight VNI genetic clusters can be seen. Low levels of sharing between the clusters can be observed. The greatest amount of sharing between genomes appear in purple and the lowest in yellow.

Genetic diversity

Despite the low number of isolates (n=5), cluster 3 displayed the highest level of genetic diversity (77,603 SNPs). Cluster 7 (n=15) harboured the lowest number of polymorphisms with 8,577 SNPs (Table 5.11). Strong negative Tajima's D value was observed in clusters 2, 4 and 6. This finding reflects a possible recent population expansion of these clusters potentially due to the worldwide dissemination of feral pigeons in the last centuries (Litvintseva et al., 2011). Strong linkage disequilibrium was found in cluster 1, 4 and 5, which is consistent with a lack of recombination and a clonal propagation. LD decay plots supported the lack of recombination present in these VNI clusters (Figure S17).

Table 5.19 displays the number of SNPs found between two clusters that were not present in any other genetic cluster. Cluster 3 shared SNPs with all clusters especially clusters 1, 2 and 4 with 2,161, 1,612 and 1,945 SNPs shared respectively. Cluster 4 shared an important number of SNPs with cluster 5, 6, 7 and 8, indicating a strong population history between these clusters. Interestingly, cluster 7 shared 2,669 SNPs with cluster 8, a number that is significantly higher than found with any other cluster. These results might indicate that these isolates have split only recently. The same observation applies to cluster 1 and 2.

In order to ascertain whether different continents contained different levels of diversity, I compared environmental isolates as these are not likely to have been recently moved through air-travel. A pairwise SNP comparison was performed between environmental isolates from Africa (n=26) and Asia (n=20) (Tables S27 & S28). The average pairwise SNP diversity was slightly higher in Africa (19,468) compared to Asia (14,581) (p <0.01). The higher number of SNPs circulating in African isolates could be explained by the fact that environmental African isolates were present in five genetic clusters whereas Asian isolates were only present in four. Additionally, a pairwise comparison of shared and unique SNPs between clusters is provided on Figures S29 & S30, revealing the evolutionary role of cluster 4 which shared the highest number of SNPs with other genetic clusters.

Pairwise Population Gene Flow

A total of 4,852 SNPs were shared between all eight genetic clusters. Population genetic statistics were computed (Table 4.1) and F_{ST} statistics were summarised in Table 5.16. The

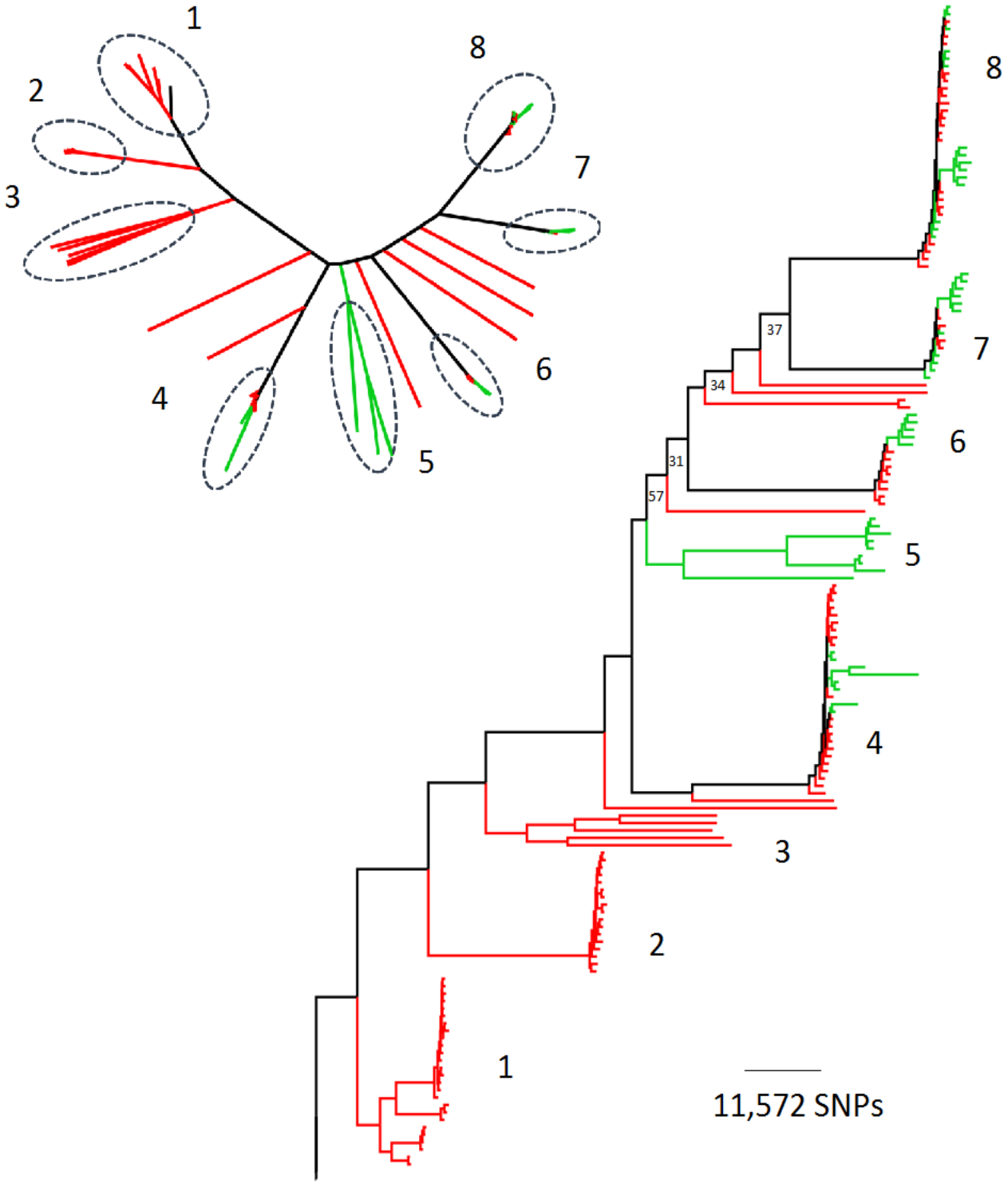


Figure 5.15: Phylogenetic relationship of the eight genetic clusters identified in *Cng* VNI isolates (n=160). African isolates appear in red and Asian isolates are in green. Almost all bootstrap values were averaging 100% at each node. Values below 90% are indicated on the tree.

Table 5.14: Population Genetics Statistics of VNI *Cng* genetic clusters

Clusters	S	η	ηE	π	θ	Tajima's D	D'
C1	27,108	27,364	3,716	0.266	0.257	0.141	0.545
C2	57,632	57,656	53,743	0.102	0.235	-2.426	-0.184
C3	77,603	109,990	81,075	0.559	0.646	-1.027	0.099
C4	60,628	60,734	15,656	0.087	0.172	-1.920	0.562
C5	56,360	56,416	35,576	0.227	0.255	-0.562	0.549
C6	57,207	57,267	50,569	0.197	0.331	-1.923	0.014
C7	8,577	8,628	3,461	0.046	0.049	-0.208	0.011
C8	12,856	12,921	5,261	0.040	0.053	-0.947	-0.174

Table 5.15: SNPs uniquely shared by two clusters

Pairwise SNPs	C1	C2	C3	C4	C5	C6	C7	C8
Cluster 1	-							
Cluster 2	626	-						
Cluster 3	2161	1,612	-					
Cluster 4	189	329	1,945	-				
Cluster 5	261	284	1,490	1,948	-			
Cluster 6	107	142	1,016	1,172	1,161	-		
Cluster 7	35	280	619	1,112	703	1,017	-	
Cluster 8	388	148	761	974	738	921	2,668	-

genetic clusters 3 and 5 displayed long phylogenetic branches on Figure 5.15. These two clusters therefore may represent populations that are undergoing more frequent sexual recombination, and hence evolution of VNI *Cryptococcus* present in Africa and Asia respectively. The F_{ST} value for these two subpopulations was 0.134 which represents the lowest F_{ST} found for the two clusters. Strong population structure was observed between the different genetic clusters. Cluster 4 had limited genetic differentiation with cluster 2 (F_{ST} =0.068), cluster 6 (F_{ST} =0.074), cluster 7 (F_{ST} =0.057) and cluster 8 (F_{ST} =0.100). This cluster had a high genetic diversity with 60,628 SNPs and was composed of environmental isolates from Namibia (n=1), South Africa (n=3) and India (n=1). Clinical isolates were also present in this cluster, from Zambia (n=5), India (n=6), South Africa (n=11). The high genetic diversity and the lack of population differentiation between cluster 4 and the other genetic cluster could indicate that worldwide dissemination of VNI isolates might have originate from this cluster.

Table 5.16: F_{ST} Statistics among VNI clusters

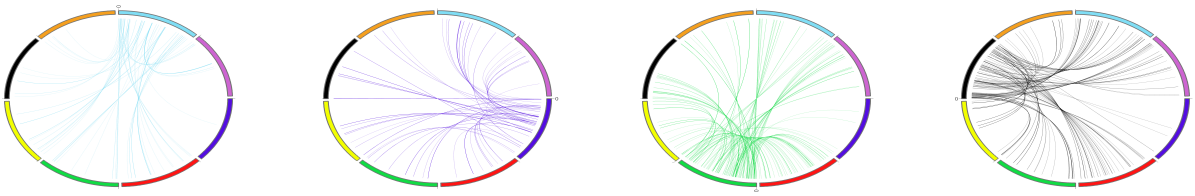
F_{ST}	C1	C2	C3	C4	C5	C6	C7	C8
Cluster 1	-							
Cluster 2	0.176	-						
Cluster 3	0.302	0.382	-					
Cluster 4	0.149	0.068	0.304	-				
Cluster 5	0.249	0.281	0.134	0.204	-			
Cluster 6	0.185	0.204	0.291	0.074	0.213	-		
Cluster 7	0.179	0.183	0.314	0.057	0.230	0.141	-	
Cluster 8	0.233	0.166	0.497	0.100	0.366	0.186	0.145	-

Recombination Analysis in VNI Cng

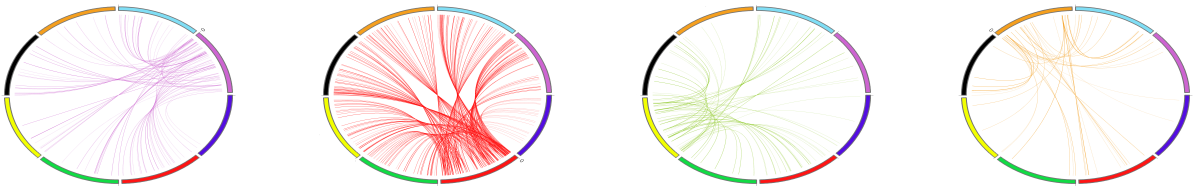
Recombination analysis were performed using BradNextGen (Marttinen et al., 2012) on Chromosome 12 (Figure 5.16). Due to computational limitations, analysis could not be performed on more than one chromosome. Cluster 4 appeared to be recombining more than any other clusters (Table 5.17). 49% (n = 1,474) of the chromosomal segments detected using BradNextGen belonged originally to cluster 4. On Figure 5.16, we observe that this cluster is recombining with all the genetic clusters.

Table 5.17: Recombinating regions among VNI genetic clusters

Cluster	Segments	%
Cluster 1	135	4.5
Cluster 2	233	7.8
Cluster 3	196	6.6
Cluster 4	1475	49.4
Cluster 5	305	10.2
Cluster 6	187	6.3
Cluster 7	313	10.5
Cluster 8	144	4.8



(a) Cluster 1 (c) Cluster 3 (e) Cluster 5 (g) Cluster 7



(b) Cluster 2 (d) Cluster 4 (f) Cluster 6 (h) Cluster 8

Figure 5.16: Recombination segments among VNI clusters - Recombination events were identify using BradNextGen (Marttinen et al., 2012) and the figure was generated using Clicos-FS (Cheong et al., 2015). Cluster 4 appears to be the most recombinogenic cluster, accounting for 49.4% of all recombination events.

Divergence Estimations

The time of divergence from the most common ancestor was estimated in BEAST using the coding sequence of chromosome 12 on a representative panel of 100 *Cng* isolates. Table 5.18 displays the results using the mutation rate estimated using Lynchs formula ($4.13.10^{-9}$ mutations per site per generation) and using a second mutation rate of 4.10^{-9} substitutions per nucleotide per year (Simwami, 2011). In the latter experiment, the divergence within *Cng* lineages was calibrated at 4.10^{6} years as described in year by Simwami, (2011).

The generation time of environmental population is unknown, but in laboratory conditions, *Cryptococcus* generation time oscillates between two to six hours (Granger et al., 1985). If we assume a generation time in the wild ranging from one per day to one generation per month, the emergence of the different molecular types can be estimated (Table 5.18). These estimates vary from previous coalescence analysis where the emergence of VNB isolates was estimated at 4.57 million years (Simwami, 2011). Comparable results were obtained using the same analysis, but the mutation rate of 4.10^{-9} substitutions per nucleotide per year is unlikely to be appropriate.

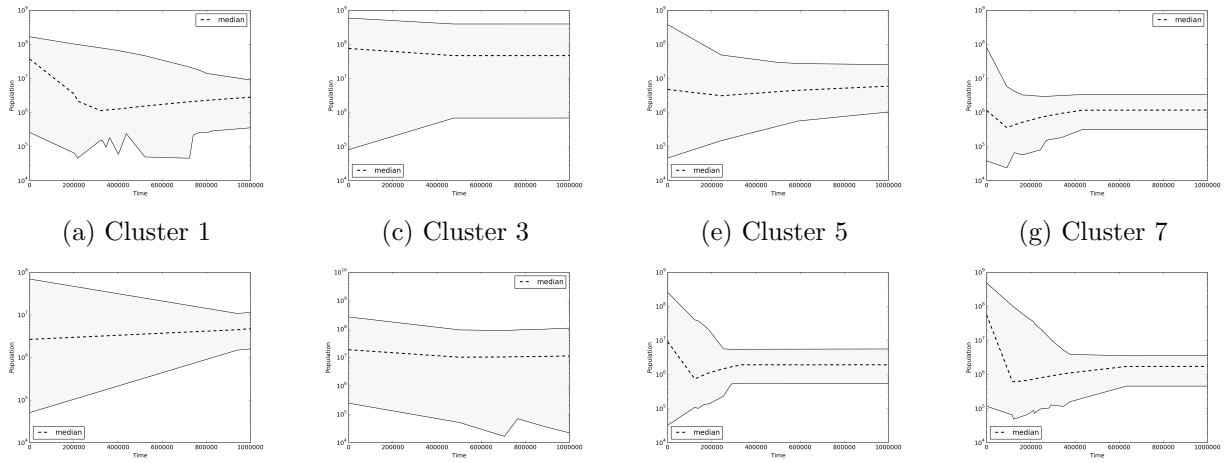
Table 5.18: Divergence Estimations of chromosome 12 using BEAST

Mutation Rate	$4.13.10^{-9}$ mutations per generation			2.10^{-9} mutations per year
	Generation	1 gen./day (years)	1 gen./month (years)	Year
VNII	$1.62.10^{7}$	$4.43.10^{4}$	$1.35.10^{6}$	$7.87.10^{6}$
VNB	$1.33.10^{7}$	$3.64.10^{4}$	$1.11.10^{6}$	$2.84.10^{6}$
VNI	$7.08.10^{6}$	$1.93.10^{4}$	$5.90.10^{5}$	$4.81.10^{6}$
VNI c1	$7.06.10^{6}$	$1.93.10^{4}$	$5.89.10^{5}$	$4.81.10^{6}$
VNI c2	$5.51.10^{5}$	$1.51.10^{3}$	$4.59.10^{4}$	$2.53.10^{5}$
VNI c3	$7.15.10^{5}$	$1.96.10^{3}$	$5.96.10^{4}$	$3.92.10^{5}$
VNI c4	$2.47.10^{6}$	$6.75.10^{3}$	$2.06.10^{5}$	$1.23.10^{6}$
VNI c5	$1.14.10^{6}$	$3.12.10^{3}$	$9.50.10^{4}$	$5.76.10^{5}$
VNI c6	$6.51.10^{5}$	$1.78.10^{3}$	$5.42.10^{4}$	$3.40.10^{5}$
VNI c7	$5.25.10^{5}$	$1.43.10^{3}$	$4.37.10^{4}$	$2.53.10^{5}$
VNI c8	$3.31.10^{5}$	$9.06.10^{2}$	$2.76.10^{4}$	$1.56.10^{5}$

Bayesian Skyline Plot (EBSP)

To further investigate the population history of the VNI molecular type, we performed a Bayesian analysis of changes of population size through time. Recombination regions were

removed and only coding regions were considered. Using the entire chromosome 12, the Extended Bayesian Skyline Plots (EBSP) were generated for each of the VNI cluster. Figure 5.17 portrays these changes in population size with an axis set to 10^6 generations. All the genetic clusters experienced a ten-fold increase in population size 100,000 to 200,000 generations ago. Assuming a generation time of one generation per day or one per month, these changes occurred between 200 to 8,000 years ago, which appears congruent with the hypothesis suggesting that *Cryptococcus neoformans* was globally disseminated alongside the domestication of pigeons within the last five centuries (Litvintseva et al., 2011). These results also support the strong negative Tajima's D value observed previously indicative a recent population expansion. The EBSP model highlights the recent increase in population size of cluster 1, 6, 7 and 8 (Figure 5.17). These findings reveal that the evolutionary history of the VNI molecular type is ambiguous and that different clusters are likely to contribute in different way to the 'neglected epidemic'of *Cryptococcus neoformans* var. *grubii*.



(a) Cluster 1 (c) Cluster 3 (e) Cluster 5 (g) Cluster 7

(b) Cluster 2 (d) Cluster 4 (f) Cluster 6 (h) Cluster 8

Figure 5.17: Extended Bayesian Skyline Plot of the eight different VNI genetic clusters - Variation of population size through time using a common axis of 1 million generations. The y-axis is represented using a logarithmic scale. (Heled and Drummond, 2008)

GWAS

Genome-wide association studies were performed to identify loci which would be specific to clinical *Cng* when compared to environmental isolates. Several analysis using different clustering methods were computed. The present analysis with the lowest genomic inflation factor included the 150 isolates of the eight different VNI sub-populations. However, the genomic

inflation remained elevated (2.30769). Three loci were significant after applying Bonferroni correction (Figure 5.18; Table 5.19). The three loci were encoding for hypothetical protein and only one polymorphism on chromosome 14 was present on a coding region and represented a non-synonymous mutation.

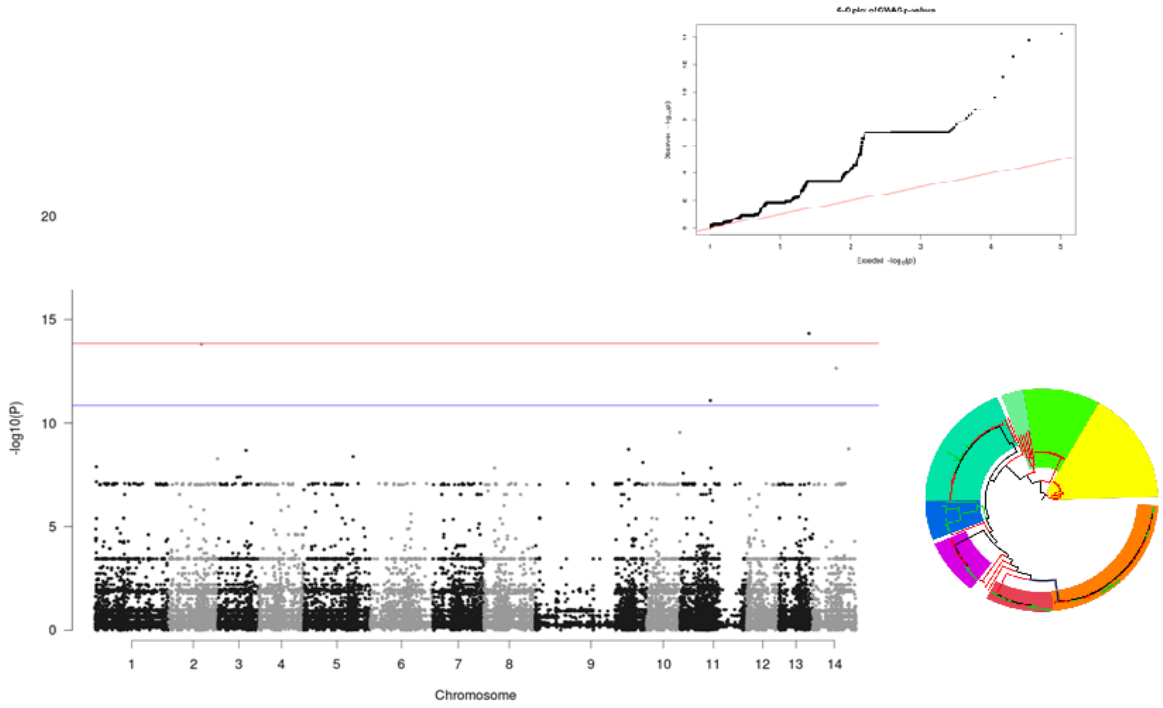


Figure 5.18: Genome-Wide Associations Study using the VNI isolates (n=145) (a) -log(p) value for each loci over the genomes. The red and blue line indicate the significance level after applying a Bonferroni correction at 0.05 and 1 respectively. (b) Q-Q plot of GWAS p value (c) The phylogeny of the different genetic clusters used in the analyses (n=8).

Table 5.19: SNPs uniquely shared by two clusters

SNPs	Bonferroni p-value	Mutation	Gene	Function
13:680100	$2.425e^{-10}$	intron	CNAG_06499	hypothetical protein
2:880473	$8.129e^{-10}$	intergenic	CNAG_03798	hypothetical protein
			CNAG_03805	calcium/proton exchanger
14:205019	$1.163e^{-8}$	NSY	CNAG_05397	hypothetical protein

5.3.3 Phenotype

Multiple genes encoding for various virulence factors were found to be under selection in environmental populations and in clinical *Cryptococcus* populations. To investigate whether cryp-

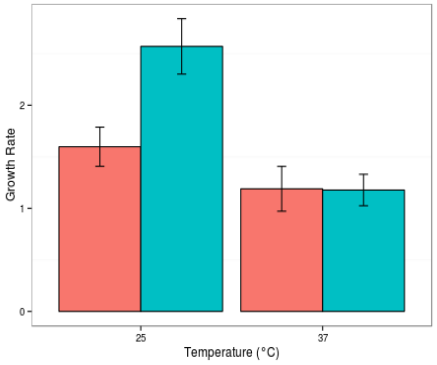
tococcal clinical isolates have a phenotypic advantage enabling them to be successful inside the human host, we investigated four known virulence factors. First, the capacity of cryptococcal strains to grow at 37°C was studied, then the production of laccase and the expression of the cryptococcal capsule were examined. Finally, Steenbergen et al., (2001) have hypothesised that *Cryptococcus neoformans* acquired its pathogenicity by avoiding environmental predators such as amoeba. Therefore, using *Acanthamoeba castellanii* as a model, the phagocytosis capacity of a panel of *Cng* isolates was estimated.

Growth Rate at Different Temperatures

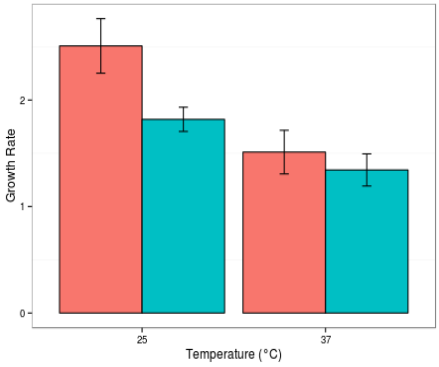
Growth at different temperatures was assessed between the different molecular types. Cryptococcal cells were grown to log phase and incubated for 24 hours. After incubation, cells were counted and the ratio of cryptococcal cells after 24 hours to the inoculum was determined. Environmental VNB isolates replicated better at 25°C than at 37°C (p <0.001). However, no difference was observed between these two temperatures for clinical VNB isolates (Figure 5.19a). In VNI, both environmental and clinical isolates grew better at ambient temperature (p <0.005 and p <0.025, respectively) (Figure 5.19b).

Differences were observed between the two molecular types (Figures 5.19c & 5.19d). Environmental VNB isolates grew better than environmental VNI isolates at 25°C (p <0.012), but clinical VNI isolates replicated faster than clinical VNB isolates at room temperature (p <0.002). VNI *Cng* clinical isolates are better adapted to high temperature than VNB isolates. This finding could be the result of the evolutionary relationship between the VNI molecular type and the feral pigeon.

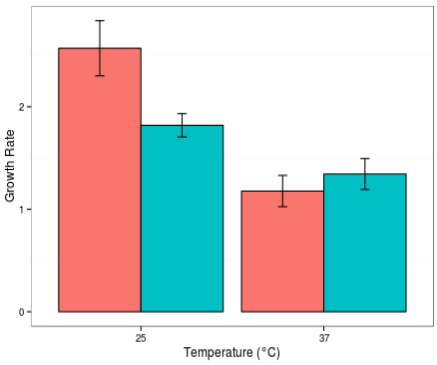
Only a limited number of isolates (n=5) were available to compare clinical and environmental isolates within the VNB genetic cluster 1. However differences were observed and environmental isolates were growing poorly at 37°C compared to clinical isolates (p <0.03). (Figure 5.20a). No significant growth difference was observed within clinical isolates at 25°C and 37°C. No significant differences were observed between environmental isolates of the genetic Cluster 1 and clinical isolates of Cluster 3.



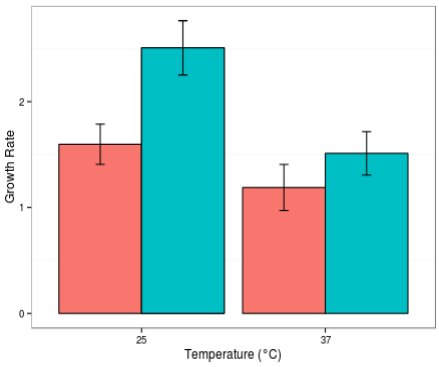
(a) VNB Molecular Type



(b) VNI Molecular Type

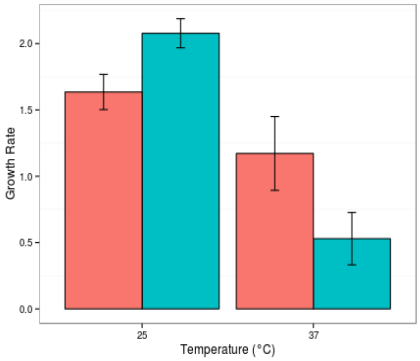


(c) Environmental *Cng*

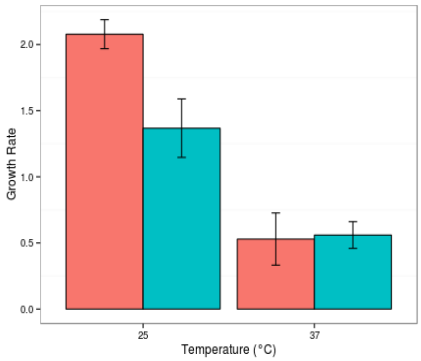


(d) Clinical *Cng*

Figure 5.19: Cell counts at 25°C and 37°C comparing Environmental and Clinical Isolates - *Cryptococcus* isolates were grown to log phase for 12 hours and counted. 5.10^5 cells were inoculated into YPD media and incubated for 24 hours at 25°C and 37°C. After incubation, cells were counted and the ratio of cryptococcal cells was calculated. The standard error for each group was calculated and the Wilcoxon signed-rank test was performed between each group to assess the significant differences. (a) differences between clinical (n=12), in red and environmental isolates in blue (n=15) ; (b) differences between clinical VNI (n=10), in red and environmental isolates in blue (n=13) (p <0.002)) ; (c) differences between Environmental *Cng* isolates, with VNB in red (n=15) and VNI in blue (=13); (d) differences between clinical *Cng* isolates, with VNB in red (n=12) and VNI in blue (=10).



(a) VNB Cluster 1



(b) VNB Cluster 1 and Cluster 3

Figure 5.20: Cell counts at 25°C and 37°C comparing Environmental and Clinical VNB Isolates - *Cryptococcus* isolates were grown to log phase for 12 hours and counted. 5.10^5 cells were inoculated into YPD media and incubated for 24 hours at 25°C and 37°C. After the incubated, cells were counted and the ratio of cryptococcal cells was calculated. The standard error for each group was calculated and the Wilcoxon signed-rank test was performed between each group to assess the significant differences. (a) Comparison within the genetic Cluster 1 of VNB - Clinical isolates (n=5) were compared to Environmental isolates (n=5) ; (b) Comparison between environmental isolates of Cluster 1 (n=10) and clinical isolates of cluster 3 (n=7).

Laccase Production

The production of laccase through the polymerisation of phenolic compounds produces melanin which is a critical virulence factor (Steenbergen and Casadevall, 2003). Melanised cryptococcal cells are less susceptible to antimicrobial activities such as nitrogen and oxidants which are generated by phagocytes (Williamson, 1997). As shown on Figure 5.21, no striking difference was observed between clinical and environmental isolates. The production of laccase was slightly higher in environmental isolates although not significant (p_{VNB} <0.068; p_{VNI} <0.881). The production of laccase did not vary between the molecular types (p <0.648).

In the VNB molecular type, the production of laccase did not differ whether isolates had the MAT α or a mating type (p <0.849). Similarly, no variation was observed between VNBa and VNBb isolates (p <0.246).

Capsule Size

The polysaccharide capsule is an important virulence factor. The main component is glucuronoxylomannan (GXM) which allows the capsule to protect the cell against dehydration (Aksenov et al., 1973). Two measures were used, the mean capsule size (μm) and the capsule

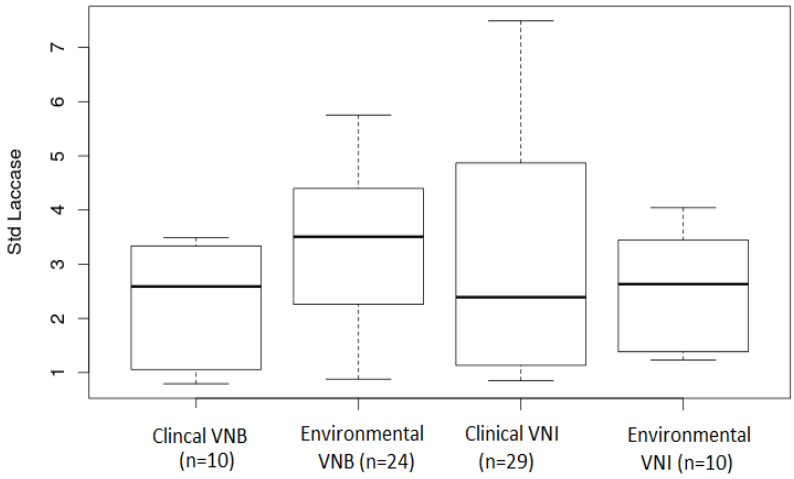
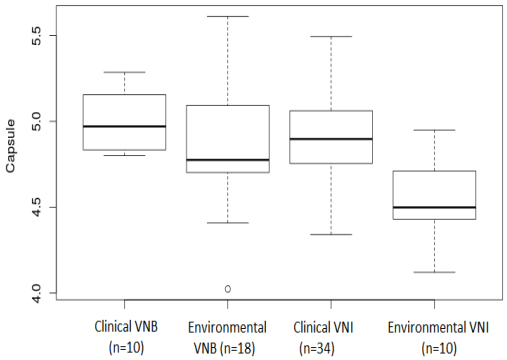


Figure 5.21: Laccase production differences between environmental and clinical isolates. To account for interexperimental variation, laccase activity is expressed as the ratio of the median of three repeats to the positive control (H99) as explained in (Sabiiti et al., 2014). No difference was observed between the different groups. Within the two molecular type investigated, VNI and VNB, the mean for the standardised absorbance of laccase was higher in environmental isolates.

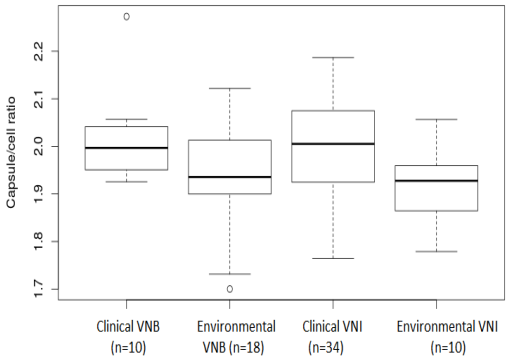
to cell diameter ratio to account for cell size variation among cryptococcal cells.

The capsule size between the two molecular types did not vary (mean $\text{capsule}_{\text{VNI}} = 4.89$ μm; mean $\text{capsule}_{\text{VNB}} = 4.82$ μm). However, the difference was significant when accounting for cell size (p <0.002). No difference was observed when comparing VNI and VNB isolates with some VNII isolates (n=7). The capsule size in *C. gattii* isolates (n=21) was smaller than in *Cng* VNI molecular types (p <0.001) with an average of 4.34 μm. The differences in capsule size between *Cng* and *Cg* might reflect the predominance of *C. neoformans* in cryptococcal infection.

When comparing environmental and clinical isolates within the different molecular types, clinical isolates produced a larger capsule (Figure 5.22). Although, this difference was only significant in VNI (p <0.016), this pattern was clearly visible within the VNB molecular type. MAT α VNB isolates displayed larger capsule than MAT a isolates (p <0.013) (Figure 5.23). This finding emphasises the potential role of the α-specific O-glucosyltransferase (*cap1*α) gene in the production of large capsule. The same trend was observed when investigating clinical and environmental isolates in the VNB genetic cluster 1 alone. No difference was seen between environmental isolates of the VNBa population (in cluster 1) and clinical isolates of cluster 3.

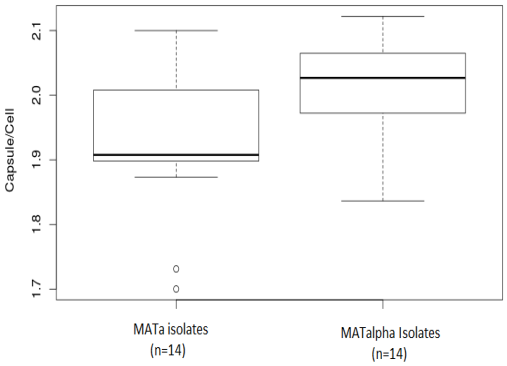


(a) Capsule Size (μm)

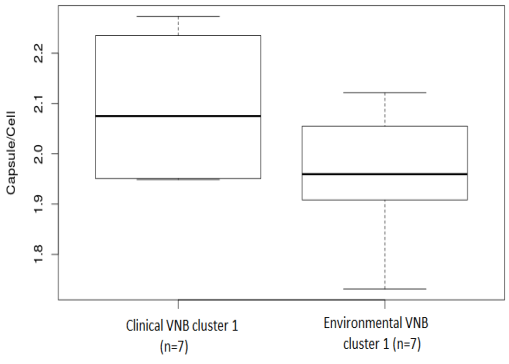


(b) Capsule/diameter Ratio

Figure 5.22: Capsule size differences between environmental and clinical isolates - The capsule size in (μm) (a) and capsule to ratio measure gave similar results. Clinical VNB isolates had slightly larger capsule ($\text{mean}_{\text{CLINICAL}}$= 5.00 μm, mean_{ENV}= 4.84 μm, p <0.070). This pattern was significantly observed in the VNI molecular type ($\text{mean}_{\text{CLINICAL}}$= 4.88 μm, mean_{ENV}= 4.55 μm,p <0.016).



(a) MAT α/a isolates



(b) VNB Cluster 1

Figure 5.23: Capsule production between MAT α/a and clinical and environmental VNB isolates - (a) The capsule to ratio was significantly higher in MAT α VNB isolates compared to MAT a VNB isolates (mean_{α}= 4.94 μm, mean_{a}= 4.80 μm, p <0.013) (b) When comparing environmental and clinical isolates from the genetic cluster 1, no significance could be established ($\text{mean}_{\text{CLINICAL}}$= 4.95 μm, mean_{ENV}= 4.79 μm, p <0.257) but clinical isolates seem to have slightly larger capsules.

Cryptococcus uptake by amoeba was measured after a two hour incubation at 30°C using ImageStream Imagine Flow Cytometer. After staining cryptococcal cells with calcofluor white, the cells which were not phagocytosed were discriminated by quenching outside *Cng* with trypan blue. The use of ImageStream technologies allowed counting of 30,000 amoebic cells per sample and the number of cells containing *Cryptococcus neoformans*. ImageStream technology remains expensive and the number of isolates which could be processed was limited. No significant difference could be determined between clinical and environmental isolates. Nevertheless, a trend was observed and clinical isolates seemed to be more easily phagocytosed than environmental isolates. Previous work showed that cryptococcal strains which are more easily phagocytosed lead to more severe clinical presentation (Sabiiti et al., 2014). Additional work should investigate whether clinical isolates can survive macrophage uptake better than environmental cryptococcal cells and the replication rate inside macrophages should also be examined.

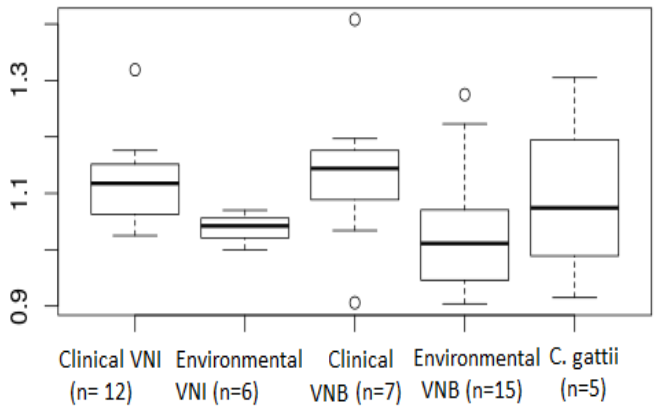


Figure 5.24: Amoeba Uptake - To account for interexperimental variation, the phagocytosis index is expressed as the ratio of the median of three repeats to the positive control (H99). Each isolate was run three times and 30,000 cells were counted for each run.

5.4 Discussion

5.4.1 VNB Molecular Type

The population structure found within the VNB molecular clade confirmed the presence of the two VNB sub-populations described in Chapter 4 and mentioned in Litvintseva, et al. (2006). However, here I show that the VNBa sub-population can be decomposed into three genetic sub-clusters, further highlighting the high genetic diversity that present within VNB. The population structure of this *Cryptococcus neoformans* var. *grubii* molecular type is likely to include other, thus far unsampled clades and genetic diversity. There is also growing evidence that the VNB molecular clade is also present outside Africa in South America (Bovers et al., 2008; Litvintseva et al., 2007) suggesting that further discoveries await. Within the VNBb population, only two clinical cases were found. These clinical cases were from Botswana and other studies have reported the presence of clinical cases from the VNBb clade indicating that this genotype is also virulent and has the potential to infect patients (Chen et al., 2015). No obvious phenotypic differences were found between the two VNB sub-populations. The two populations had a similar amount of mating type a, 28% for VNBa and 40% for VNBb. Low level of linkage disequilibrium (Table 5.2) were found in each cluster, except for cluster 3 which showed signs of LD (D'= 0.388). Low LD among loci is expected in sexual population (Pritchard and Przeworski, 2001) and the presence of both α and a mating types confirms this hypothesis.

The study aimed to identify and characterise potential differences between environmental and clinical isolates. Fungi are free-living and their infection is generally opportunistic, therefore the identification of specific traits which contribute to pathogenesis is difficult (Moran et al., 2011). Previous studies have shown associations between lineage-specific phenotypes and clinical outcome in bacteria and fungi (Enright and Spratt, 1998; Wiesner et al., 2012). The present study was able to identify genes which differed between clinical and environmental isolates. Investigations were only performed on the genetic clusters 1 and 3 to limit the impact of population structure as population structure is known to increase the number of false positives in approaches that use genome scans (Excoffier et al., 2009). Other issues arise when performing genome scans. Demographic history, such as strong bottleneck events or background selection, can be a source of confounding errors (Manel et al., 2015). Careful interpretations are therefore required and expression studies should be performed to assess the role of these

putative genes. This study represents the first comprehensive genome-wide scan for selection in an environmentally-occurring fungal species. Using this method, previous studies have found hypervariable regions involved in the development of virulence in bacterial species (Thomas et al., 2012).

My findings support the hypothesis that clinical isolates are more prone to establish infection in human. Investigations of the differences in vitro phenotypes showed variations between clinical and environmental VNB isolates in their ability to grow at 37°C (p <0.03 - Figure 5.20). Clinical VNI and VNB isolates produced slightly larger capsules than environmental isolates and there was a trend towards them being more readily phagocytosed by amoeba (Figure 5.24). Interestingly, the production of laccase was higher in environmental isolates. Melanin is a key factor in the environment allowing the decay of wood. Nevertheless, laccase production along with macrophage uptake have been associated with lower rate of clearance of infection (Sabiiti et al., 2014). The ability of *Cryptococcus* to survive inside macrophage is likely to have evolved by avoiding predation of environmental predators (Steenbergen and Casadevall, 2003). Increased phagocytic uptake also increases the risk to develop severe cryptococcal disease (Rohatgi et al., 2013). The present findings using our amoeba model conform with this hypothesis, despite the relatively low number of isolates present in our study.

Using a comparative genomic approach, we were able to find differences between environmental and clinical isolates. Several genetic changes were found to be associated with cryptococcal infection of the human host. Using gene-based methods, five genes with a known virulence function were found to be under positive selection in clinical isolates of the genetic cluster 1 and seven when comparing cluster 3 with environmental isolates of the VNBa population (Table 5.9). These results corroborate the phenotypic findings and conform to the hypothesis that there is a genetic basis underlying the successful establishment and infection in the human host. The MAT α locus is likely to play a role in virulence (Metin et al., 2010). Two capsular genes (CNAG_06813 and CNAG_00865) found on the MAT locus were identified using the genome-scan methods and significant differences in capsule production were observed between the two mating types (p <0.013).

The influence of temperature on *Cryptococcus* growth was also clearly observed. Within the VNB genetic cluster 1, clinical isolates replicated faster than environmental isolates (p <003). The *ste20* gene (CNAG_07408) was the only temperature associated gene found to be under

selection in clinical isolates. The latter encodes for the STE20α protein kinase might explain the tolerance to high temperatures. Wang et al., (2002) found that ste20α mutant led to attenuated virulence and increase survival in infected mice. The gene is part of the p21-activated protein kinase family and was the only gene associated with temperature that was identified in this study. Other potential virulence-associated genes have been uncovered (Table S26), but their role in pathogenesis remains to be established.

Gene duplication is an important mechanism for fungal pathogens (Ormerod and Fraser, 2013). The first observation of virulence-associated phenotypes in *Cryptococcus neoformans* was changes in chromosome sizes after murine infections (Fries et al., 1996). Aneuploidy appears as a central mechanism in the infection process. In *Candida glabrata*, chromosome copy number variation (CCNV) was associated with increase glucose uptake (Torres et al., 2007). Additionally, Hu et al. found that CCNV played a role in azole resistance (Hu et al., 2011). Clinical VNB isolates presented frequently with CCNV (70%), whereas environmental isolates did not show any chromosome duplication. Among the VNB isolates, CCNV did not seem to affect any chromosome particularly. Previous research emphasised the role of chromosome 13 duplication (Hu et al., 2011) but in the current study, every chromosome in clinical isolates seemed have to potential to be duplicated (Figure S13). When considering only 10kb-chromosomal regions with a high coverage, only two genes with a known virulence function were found multiple times in clinical isolates. These two genes encodes for an Adenylyl cyclase involved in glucose regulation (CNAG_03202) (Alspaugh et al., 2002) and a zinc transcription factor (CNAG_05431) (O'Meara et al., 2010a). However, the study is unable to affirm whether the genetic features observed within clinical isolates were present prior to infection. The numbers of CCNV in clinical isolates attests that these changes most likely occurred as a consequence of infection. Gene duplication have been recognised as an important force of evolution; duplicated genes might act in a gene dosage response to stressful conditions or duplication might lead to the gene diversification and the evolution of new gene functions (Moran et al., 2011).

5.4.2 The VNI molecular Type

The current understanding of the VNI molecular type is the existence of a pan-global genotype which might have originated from Africa (Simwami, 2011). This hypothesis is currently known as the out of Africa hypothesis. Phylogenetic analyses revealed the presence of eight genetic

subtypes among the VNI molecular types. These findings were supported by fineStructure and admixture analyses. Compared to the VNB molecular type, VNI isolates present a clonal population structure with a total of 198,660 SNPs (865,785 SNPs in VNB). Every genetic cluster contained clinical isolates indicating that any subtype has the potential to cause disease in humans.

The eight genetic clusters seem to have different evolutionary history with strong population differentiation (with F_{ST} values generally above 0.15). Cluster 1, 2 and 3 were strictly composed of African isolates, whereas cluster 5 only contained Asian isolates (Figure 5.15). The presence of clusters containing isolates from the two continents suggests that isolates have been circulating around the world, and for some time. Additionally, cluster 6, 7 and 8 contained environmental isolates from both continents, confirming the global circulation of these genetic clusters. Strictly African clusters are potentially endemic to the continent. From the finestructure analysis, cluster 3 shares a high proportion of genetic material with cluster 1 demonstrating the close phylogenetic relationships between the two clusters. Another possibility is that these clusters might be associated with migratory birds flying principally in the Atlantic flyway and therefore they could be found in Asia. Other clusters containing isolates from the two continents might be associated with avian species present in both Africa and Asia. These birds might therefore be associated with either the black sea Mediterranean flyway or the East Asia/East Africa Flyway (Boere and Stroud, 2006). Sequencing additional isolates from other continents in the future will help to understand the evolutionary history of the molecular types.

From the recombination analysis, we observed that cluster 4 was of particular interest. The cluster was at the origin of 49.2% of recombination events identified using the BradNextGen analysis and shared genomic regions with all other genetic clusters. According to coalescence analyses, clusters 4 and 1 cluster seem to have the oldest tmrca. Interestingly, cluster 4 displayed small population differentiation with cluster 2, 6, 7 and 8 (F_{ST} <0.1) indicating that this cluster might be the source of the global dissemination of the VNI molecular type. This high gene flow between global VNI isolates has been previously highlighted (Litvintseva et al., 2011). Multilocus analyses revealed that global VNI strains were highly clonal compared to African strains. In the context of the VNI molecular type, a similar number of SNPs were observed in each genetic clusters ranging from 77,603 SNPs in cluster 3 to 8,577 SNPs in cluster 7 (Table 5.18). Previous studies using MLST were unsuccessful trying to find spatial clustering between

infected patients (Beale et al., 2015). However, we observed that several genetic clusters occur within the same environment. In our dataset, which included 48 isolates from Cape Town, SA, five of the eight VNI genetic clusters were isolated from Cape Town patients underscoring the highly spatially mixed nature of this lineage. The population structure of the VNI molecular type appears stronger than previously thought. The presence of eight strongly supported genetic clusters highlights the complexity of the evolutionary history of the VNI molecular type. The hypothesis suggesting the existence of pan-global genotype might be due to the lack of resolution provided by multilocus analyses. In light of whole-genome data, we have no evidence that the VNI molecular type originated from Africa. Cluster 1, 2 and 3 might be endemic to the African continent and it might explain the higher genetic diversity observed in Africa compared to Asia (p <0.001). However, recombination analyses indicate that cluster 4 might represent the source of the global dissemination of the molecular type and there is no evidence that this cluster originated from Africa.

Previous coalescence estimates were based on a mutation rate of 2.10^{-9} per nucleotide per year for protein coding genes (Xu et al., 2000). In 2010, Lynch et al. (2010) released a paper highlighting the possibility to scale the mutation rate with the genome size. From his formula, we obtained a mutation rate for *Cng* of $4.13.10^{-9}$ mutations per site per generation. Divergence estimates assuming a generation time averaging one replication per day or per month indicate that the mean tmrca for these genetic clusters varies between 500 to 2,000 years ago suggesting the potential role of humans or migratory birds in the dissemination of the molecular type. A recent ten-fold increase in population size was observed between 200 to 8,000 years in all VNI genetic clusters. These estimates of population size expansion agrees with the hypothesis that *Cryptococcus* was associated with the recent domestication of the feral pigeon, *Columba livia*, by humans. The population size of the clusters 1, 6, 7 and 8 increased significantly more in the past 10,000 years, highlighting the fact that VNI genetic clusters contribute differently to the *Cryptococcus* epidemic, and that there are likely multiple origins of globally-occurring genotypes. Genetic clusters 1, 2 and 3 might represent endemic African clusters and additional sampling efforts are required to uncover the full evolutionary history of the pathogen. Isolates from Europe or South America would give insights on the role of cluster 4 in the worldwide dissemination of *Cryptococcus neoformans*.

Chapter 6

Discussion and Conclusions

6.1 Introduction

Cryptococcal meningitis remains an important fatal disease with an incidence ranging from 0.04 to 12% per year among AIDS-patients (Park et al., 2009). The estimate of 600,000 annual deaths advanced by Park et al. has likely been reduced in the last five years due to the impact of antiretroviral therapy (ART) in many regions worldwide (Perfect and Bicanic, 2014). However, the fungal infection in immunocompromised individuals remain considered as a neglected epidemic (Armstrong-James et al., 2014) responsible for 50% of AIDS related deaths worldwide (Brown et al., 2014).

This PhD thesis investigates the molecular epidemiology of *Cryptococcus neoformans* var. *grubii* (*Cng*) in the light of new generation sequencing technologies. The use of ITS barcoding represents a first attempt to explore the fungal population structure associated with an environmental pathogen. Then, whole-genome sequencing uncovered the population structure of environmental *Cng* and using genome-scan methods, we were able to detect signs of selection within the *Cryptococcus* genome. Virulence-associated genes were detected using frequency-based and gene-based methods. A comparison between environmental and clinical *Cng* attempted to identify genetic features which may play a role in the infection process. The study revealed the intrinsic complexity of the virulence traits associated with this opportunistic pathogen, highlighting the difficulty of pinpointing specific features which contribute to *Cryptococcus* pathogenesis. In vitro phenotypic assays were performed on a panel of *Cng* isolates to find

to phenotypic variations between isolates and to potentially link them to my genetic findings. Finally, new insights were gained regarding the demographic history of *Cng*.

6.2 *Cryptococcus neoformans* and *C. gattii* have distinct ecological niches

The environmental isolation of the pathogenic yeast *Cryptococcus neoformans* (*Cn*) was reported in the early 1950s (Emmons, 1951). The isolation of the saprophyte from the environment has always been problematic yielding low recovery rates (Granados and Castañeda, 2005b; Kuroki et al., 2004; Lazera et al., 2000; Litvintseva et al., 2003; Mseddi et al., 2011; Randhawa et al., 2008). The efficacy of the swabbing technique is limited but to date, it remains the most effective technique (Randhawa et al., 2005). In the present study, the recovery rate average nearly 5% among the 1,742 samples investigated and a total of 37 *C. neoformans* isolates and 38 *C. gattii* were recovered from these field surveys.

6.2.1 *C. neoformans* is adapted to the drier Mopane Woodlands

C. neoformans and *C. gattii* (*Cg*) have been considered as separate species since 2002 (Kwon-Chung et al., 2002). In Zambia, each species was associated with different ecoregions ($p < 0.001$) indicating that the two species are adapted to different environments. *C. gattii* was predominantly found in the central Miombo Woodland (CMW) dominated by *Brachystegia* and *Julbernardia* species, whereas *C. neoformans* was mostly associated with the *Colophospermum mopane* in the Zambezian Mopane Woodlands (ZMW). It is worth noting that 27% of *Cn* and 34% of *Cg* were recovered from soil samples neighbouring trees, showing that the fungus occurs widely in the environments where it was found.

The two species diverged millions of years ago and current research supports the idea that they have acquired new phenotypic differences allowing them to adapt to new environments and colonise new ecological niches (Springer and Chaturvedi, 2010). *C. neoformans* is known to reproduce preferably in dry conditions (Caicedo et al., 1999; Granados and Castañeda, 2005b). Granados et al. (2006) affirmed that the ecological niches variations between *C. neoformans*

var. *grubii* and *C. gattii* could be explained by their different tolerance to climatic conditions. In the ZMW, temperature is higher and the ecoregion is also subjected to a reduced amount of precipitation compared to the wetter (and higher altitude) Miombo Woodlands. The drier conditions of the *C. neoformans* ecological niche could have led to the acquisition of a selective advantage which might explain the higher prevalence of *Cn* in patients.

6.2.2 Our knowledge regarding the fundamental niche of the *Cryptococcus*-species complex remains limited

The study successfully recovered *Cng* from *C. mopane*, a previously described environmental niche (Litvintseva et al., 2003)). However, this work represents the first attempt to recover environmental *Cryptococcus neoformans* in Zambia, a country which experiences some of the highest HIV-infection rates in the world.

In the Miombo Woodlands, *C. gattii* was not particularly associated with any tree type. The evolutionary history of this pathogen in Africa remains to be investigated. Previous work hypothesised that *C. gattii* originated from Australian Eucalyptus (Chen et al., 2000; Ellis and Pfeiffer, 1990). However, a recent comparative genomic study advanced a possible emergence of the pathogen out of South America (Engelthaler et al., 2014). The present study found a strong association between *C. gattii* and Hyrax faeces with 20.6% of samples positive, emphasizing the current lack of knowledge regarding the realised ecological niche of the pathogen. Additional sampling is likely to lead to the discovery of novel niches for the *Cryptococcus*-species complex despite the difficulty of isolating the pathogen in nature.

The prevalence of the two species appear to be influenced by seasonality. The survey investigated both the rainy and dry seasons and the recovery rate went from 1.46% in the rainy season to 2.38% in the dry season. This difference was not significant ($p = 0.476$) however this trend has been observed previously (Chowdhary et al., 2012; Granados and Castañeda, 2005a; Kuroki et al., 2004) and might be due to a possible easier wind propagation of infection propagules during the dry season. However, additional work is required to draw conclusions regarding the influence of seasonality on the *Cryptococcus*-species complex.

6.3 Associations between *Cryptococcus* genetic variations and the environment

The characterisation of the microbial assemblage thanks to high-throughput metabarcoding allowed me to investigate regional patterns of biodiversity. Many worldwide diversity projects, such as the Earth Microbiome project, aim to describe the global microbial taxonomy using barcoding. The understanding of biotic and abiotic features on the species present in different ecological niches is likely to help to predict disease emergence and identify environments where a spillover might occur (Gilbert et al., 2010a). The present work shows that each Zambian ecoregion associated with either *C. gattii* or *C. neoformans* is composed of different assemblages of fungal communities.

6.3.1 The fungal community structure changed between the Zambian ecoregions

The fungal community structure between each ecoregion varied significantly (R_{ANOSIM} >0.350, p <0.001) confirming that fungal species are adapted to different environments. The observed differences were amplified when accounting for altitude, temperature and precipitation levels. When investigating the role of space in the ZMW, the species composition remained stable within each ecoregion. This work represents the first attempt to characterise this pathogens biotic environment. However, with the increased affordability and the development of new technologies like shotgun metagenomics (Hodkinson and Grice, 2015), more sophisticated studies are likely to be published that attempt to reconstruct the ecosystem functions of particular fungal and/or bacterial communities. A similar approach could also be used to investigate the fungal diversity associated with each *Cng* molecular type, however additional (and extensive) sampling efforts would be required to achieve this goal.

C. neoformans is an emerging pathogen and it is important to be able to predict the potential spillover of an environmental pathogen into new ecotope. A notable example of the potential of *Cryptococcus spp.* to emerge is the ongoing outbreak of *C. gattii* in Vancouver (Kidd et al., 2004) where, across the last decade, 70 cases were reported by the Center for Disease Control and Prevention (CDC) with a 20% mortality (Desalermos et al., 2012). A similar scenario for

Cng should not be excluded.

The landscape of Zambia changed significantly in the last decades due increased anthropogenic pressures (Chidumayo, 2013). These changes are likely to affect human-*Cryptococcus* interactions. Metabarcoding studies help to monitor changes in microbial diversity and fungal species could be used as a marker of diversity across this changing Zambian environment (Osono, 2011). Additionally, these types of studies contribute to understand the diversity associated with unexplored ecosystems like African soil. Currently, only 80,000 fungal species have been identified which is likely to represent just a fraction of the six fungal phyla that are present in the environment (Fierer et al., 2007).

6.3.2 Environmental niche model conforms with biodiversity studies

The environmental niche model performed in MaxEnt (Phillips and Dudík, 2008) was based on the widely used Bioclim layers (Busby, 1991). The model matched perfectly the two Zambian ecoregions associated with the two sister species, *Cng* and *Cg*. MaxEnt approximates the species ecological niche using climatic and spatial information. The model reinforces the idea that climate shapes fungal community structure. Elevation was also a crucial factor in explaining *Cryptococcus* distributions; altitude likely acts as a physical barrier to gene flow and may physically limit species dispersal (Losos and Glor, 2003).

When applying the species modelling method to *Cng* molecular types, the VNBb population appeared to be restricted to the Zambezian Mopane Woodlands (Figure 4.17). The VNBa clade was more widely found throughout southern Africa and is likely to be present outside these regions. Regarding VNI isolates, the study only modelled the distribution of two sub-populations which were identified in our environmental dataset. Even if a clear spatial delimitation was observed, the MaxEnt analysis is unlikely to accurately describe the distribution of VNI isolates in southern Africa. In chapter 5, the dataset which included clinical VNI isolates (n=155) was composed of five genetic clusters found in Africa. VNIa contained isolates of clusters 1 and 2 and VNIb contained isolates of clusters 3, 4 6, 7, 8. Additional sampling is needed in order to obtain a clear representation of VNI distribution in the environment, both across African and more widely.

6.3.3 Biogeography: variations in the distribution of *Cng* molecular types in Southern Africa

Partial mantel tests were successful in identifying correlations between individual genomes and environmental variables. However, these methods are known to be confounded by spatial autocorrelation (Guillot and Rousset, 2013) as Mantel tests assume a linear relationship between the investigated variables (Legendre and Fortin, 2010). To overcome this problem, new methods are currently being developed in a Bayesian framework to disentangle the relative effects of a variable. An example of such method is Bedassle (Bradburd et al., 2013). Unfortunately, our environmental dataset contained a limited number of isolates. This method requires a clear repartition of isolates into populations based on the close proximity of phylogenetically similar isolates.

Genetic environment association (GEA) methods aim to characterise empirical patterns of adaptation (Forester et al., 2015). These approaches assess the relationship between environmental variables and genotypic variants. The study investigated the role of space and climatic factors on the patterns of genetic differentiation observed in the *Cryptococcus* genome. Local dispersal is supposed to be predominant in fungal species (Paape et al., 2013). Using whole-genome sequencing, we were able to observed significant patterns of spatial genetic correlation in the VNI and VNB *Cng* molecular types. The field of landscape genomics is still in its early days, but we were successful in identifying virulence-related genes that are associated with environmental variables. The Latent Factor Mixed Model (LFMM) is known to be a robust approach (Stucki et al., 2014), but the impact of population structure on the findings remains debated (Lotterhos and Whitlock, 2015). Understanding the role of the environment in the development of genetic adaptation has been subjected to a longstanding debate among evolutionary biologists (Kawecki and Ebert, 2004). In the near future, the applications of landscape genomics are likely be extended as finer tools for measuring climate are been developed and longer cycles of meteorology patterns become available (Olsen et al., 2011).

6.4 Population genomics of *Cryptococcus neoformans* var. *grubii*

The fineStructure and admixture analyses allowed to assign isolates to different populations. A clear finding was that the two molecular types were completely separated and no introgression could be found. On the other hand, within each molecular type, recombination was observed especially within the recombinogenic VNB isolates.

6.4.1 Signature of selection

Darwinian selection leads to the specialisation of a population and to the eventual emergence of new species. The development of recent population genomics tools allows to study positive selection and investigate adaptation within the genome (Guillot and Rousset, 2013). The advantage of 'hypothesis-free'genomic approaches is the unbiased search for evolutionary accelerations within the genome, resulting in biological variants that are involved in the adaptation of the organism to its environment.

The reverse-ecology approach to identify genetic outliers in this study combined frequency-based methods and gene-based methods. These methods are commonly used to study organisms (Vitti et al., 2013) but the problem of false positives remain a strong drawback (Cullingham et al., 2014).

I reviewed and identified 470 genes identified in functional studies and focus principally on these genes. Within the VNB molecular type, cluster 1 and 3 displayed adaptive features, whereas the VNBb (cluster 2) sub-population seemed to be at genetic equilibrium. The complete distribution of VNB isolates remains to be determined. The high genetic diversity present in this molecular type could either mean that the organism is highly adapted to its ecological niche or that it has the potential to adapt to other environments and cause outbreaks.

The investigations of the VNB genomes were performed using the H99 VNI reference. A VNB reference was also built using SPAdes assembler v2.5 (Bankevich et al., 2012) and congruent phylogenies were obtained (data not shown). However, to describe more accurately the genetic processes that are occurring within the VNB genome, a specific gene annotation for this lineage

is required however is not currently available. Among the VNI isolates, 12 genes displayed signs of positive and/or purifying selection including two virulence-associated genes, CNAG_04093 and CNAG_06813. 89 genes were found with GEA methods (LFMM) including seven genes with a known virulence function. Due to the high genetic diversity present in the VNB clade, genetic variants within the molecular type could be identified. When studying environmental VNB populations, frequency-based methods identified nine genes and 34 genes were found using gene-based methods to be under selection. For instance, the vacuolar-ATPase subunit (VPH1) constituted one of the strong gene candidates (CNAG_01106). CNAG_00306 was also identified using both statistical approaches. The latter encodes for the metallothionein protein (CMT2) and is part of the organisms Cu response (Ding et al., 2013). Among the clinical VNB isolates (Table 5.14 Chapter 5), 19 putative candidates were found using genome-scans methods. These genes were also detected when investigating genes under selection in environmental VNB populations (Chapter 4). The overlap between genes found to be under selection in the environment and in clinical isolates highlights the potential role of these genes in the infection process.

6.4.2 The rise of Genome-scans

The *Cryptococcus* field entered the genomic area in 2005, when Loftus et al. released the 6,900 genes of the *Cng* genome (Loftus et al., 2005). The reference genome provided by the Broad Institute (http://www.broadinstitute.org/) allowed me to physically map Single Nucleotide Polymorphism (SNPs), the marker of choice to perform genome scans. My study used allele frequency spectrum using sliding-based windows methods to study signs of positive selection (Tine et al., 2014). In essence, genome-scans which are commonly applied to population genomics rely on departure from neutrality (Haasl and Payseur, 2015; Manel et al., 2015). However, neutral founder effects can also affect allele frequency randomly and demographic events can mimic patterns of selection during population expansion or isolation by distance (Duforet-Frebourg et al., 2014; Novembre and Di Rienzo, 2009). It is difficult to disentangle background selection from selective sweeps (Karl et al., 2012; Nielsen, 2005). Dealing with population structure to detect selective sweep also remains a source of potential confounding effects (Excoffier et al., 2009). However, statistical analyses to detect positive selection at the population level are been developed at an impressive rate. These approaches are based on the

changes in the site frequency spectrum (Tajima, 1989), the local reduction of diversity (Vatsiou et al., 2015) or extended linkage disequilibrium (Tang et al., 2007). Positive selection acts as the primary mechanism of selection and newly acquired mutations driven to high frequency are likely to be detected using these methods (Stapley et al., 2010).

We observed little overlap between statistical parameters. Using frequency-based methods, we used Tajima's *D*, a commonly applied statistic to detect positive selection (Amambua-Ngwa et al., 2012; Bigham et al., 2010) and Lewontin D', a linkage disequilibrium measure (Lewontin, 1964). Among the possible parameters available, Fu's Fs (Fu, 1997) and Fay and Wu's H (Fay and Wu, 2000) represent two good statistical alternatives. While the former is supposed to be more sensitive than Tajima's *D*, the latter is less subjected to population expansion than Tajima's *D* and allows to distinguish population expansion from purifying selection. Another valid alternative and widely used is XP-EHH, a method which compares haplotype lengths between populations and control for local variations in recombination rate (Sabeti et al., 2007).

The advantage of genome scans is that it allows to prioritise putative loci and in the context of *Cryptococcus neoformans*, detects reliable gene candidates which contribute to the saprophyte success as a pathogen. In the previous chapter, we successfully mapped putative candidate loci to coding regions in the genome. However, genome wide association studies were unsuccessful in this study due to the reduced number of isolates but these methods appear as a promising tool to identify genetic variants involve in various phenotypes (Storz and Wheat, 2010).

6.4.3 The presence of MAT a locus supports the evidence of recombination in the VNB clade

The large mating type (MAT) locus in *Cryptococcus* has been studied for a long time. Located on chromosome five, this large region is thought to be between 103 and 127 kb and controls the bipolar mating-type system responsible for homeodomain genes, pheromones and pheromone receptors (Bovers et al., 2008; Lengeler et al., 2002). VNI isolates predominantly possessed the MAT α locus. However, one VNI isolate (Ze25) found in soil in the Zambian Miombo woodlands had a MAT a locus implying that sexual recombination among VNI *Cng* may occur more often than initially thought (Wickes et al., 1996).

Recombination presents an ecological advantage creating new genetic variants (Hamilton, 1980).

Among environmental VNB isolates, 32% of isolates possessed the MAT a locus. The presence of MAT a therefore indicates the possibility of recombination in the environment. This finding has been previously observed using MLST analysis (Litvintseva R, Vilgalys, R, et al., 2006; Simwami, 2011). LD decay plots were good indicators of linkage disequilibrium and confirmed that recombination was acting to break up long haplotypes within the VNB genome (Figure 4.7). Lower levels of recombination where found among VNI isolates compared to the VNB clade supporting the idea of asexual recombination is predominant among the VNI clade (Litvintseva et al., 2007).

The MAT α locus is known to play a role in virulence (Metin et al., 2010). My study identified two capsular genes which belong to the MAT α locus and significant differences in capsule expression were observed between the two mating types (p <0.013) (Figure 5.23a).

6.4.4 The key role of CCNV in the infection process

There is no doubt that Chromosomal Copy Number Variations (CCNV) play a significant role in the infection process. Previous investigations using flow cytometry advanced that 10% of environmental and clinical strains were diploid (Nielsen et al., 2003). The high number of CCNV found within VNB clinical isolates (58% of isolates compared to 4% for environmental VNB isolates) highlights the role of genomic plasticity in infection. Interestingly, only 17% of VNI clinical isolates had a chromosome duplication compared to 0% of VNI environmental isolates. This finding emphasises the potential of the VNB molecular type to adapt to its environment. Beale et al., (2015) found that infection by the VNB lineage is associated with significantly poorer survival of patients. The ability of VNB isolates to undergo chromosome duplication could be associated with this high mortality, a hypothesis that deserves further scrutiny.

Genome resequencing projects will help to understand the equilibrium between the stability of the *Cryptococcus* genome and its high flexibility (Ormerod and Fraser, 2013). The advantage of deep coverage is that it allows one to assess copy number variation. The DNA sequenced in this project belong to an entire colony of *Cng*. The development of single cell sequencing technologies will help to understand the role of CCNV during the infection and should be explored in the future. Chromosomal duplication can act in two different ways. A duplication

can either lead to a dosage response or promote gene diversification and lead to the evolution of novel gene functions (Moran et al., 2011).

6.5 Aiming to link the genotype to the clinically-relevant phenotype

Linking phenotypic data to genomic findings remains the best way to provide evidence for the genetic basis of phenotypic change (Lepais and Bacles, 2014). Previous work had shown associations between clinical outcome and lineage-specific phenotypes in bacterial and fungal species (Enright and Spratt, 1998; Wiesner et al., 2012). In this study, I have shown using in vitro phenotypic assays that clinical isolates tend to be more prone to infect humans than environmental isolates. They are more easily phagocytosed, produce larger capsules and are thermotolerant. These findings were observable in both VNI and VNB molecular types. Some genetic candidates which were found to be under selection might explain the different phenotypes observed. For instance, the *ste20* gene (CNAG_07408) encoding for the STE20α protein kinase might explain the differences in growth at 37°C. The three capsular genes CNAG_06813, CNAG_5939 and CNAG_00865 might explain the difference in capsule expression.

Gene knockout experiments and expression studies should be able to determine whether these genes are more expressed in clinical isolates compared to environmental isolates. As functional studies continue to uncover the genetic basis of virulence, we will understand better the role of genes under selection. Our genomic findings indicate that carbon utilisation (Yang et al., 2010), iron deprivation (Vartivarian et al., 1993) or tolerance to copper (Raja et al., 2013) are virulence-associated phenotypes which should be investigated. Numerous genes were found to be under strong adaptive forces but were matched to hypothetical protein in the database. The third version of the *Cryptococcus* genome assembly has recently been released and gene functions will continue to be revealed. Future studies should investigate the genetic basis of the mechanisms which allow *C. neoformans* to survive within macrophages as this feature allows the opportunistic pathogen to invade the central nervous system.

To the question why is *C. neoformans* successful as a pathogen? (Coelho et al., 2014), the answer is likely to be due to multiple factors ranging from: its predominance in the environment,

as most people have been exposed to the pathogen before the age of five (Goldman et al., 2001), through to its genomic adaptability in stressful conditions due to CCNV or its virulence factors. However, it also appears that some wild-type *Cng* isolates are inherently more virulent as stressed by Litvintseva & Mitchell, (2012), a hypothesis that resonates with the results of my research.

6.6 Epidemiology of *Cryptococcus neoformans* var. *grubii*

6.6.1 Local infection VNI is an urban acquired-infection and VNB is a rural

Due to the presence of the Mopane tree belt in Zambia, the VNB molecular type might play a significant role in local cryptococcal infection in the country. We analysed 35 clinical Zambian *Cng* genomes from patients however found that only three isolates were lineage VNB. These majority of these patients were admitted into Lusakas hospitals and might have therefore acquired the infection in urban settings where the VNI molecular type predominates (Chapter 4). The MRC control trial where the isolates were collected is currently recruiting patients from Livingston in the south of Zambia and Ndola in the north. These *Cng* isolates might help us to understand the role of the VNB molecular type on local infection as the patients from these regions are more rural. Mopane timber plays a significant role in the local economy as it used for firewood and construction (Litvintseva et al., 2011). Therefore, we can hypothesis that VNB infection reflects infection acquired in rural environments and that the VNI molecular type is predominantly acquired in urban centres.

In South Africa, most of the patients recruited in the clinical trial lived in Cape Town townships. Environmental sampling isolated two *Cng* isolates in Cape Town which were found to belong to the VNI molecular type. Five VNI isolates were also collected in the town of Fort Beaufort on trees contaminated with pigeon faeces. These findings support the hypothesis that the feral pigeon which inhabits urban centres act as a mechanical transporter of VNI *Cng*. A striking example is the recovery of a VNI isolate in Helmeringhausen, a small town in the Namibian

desert. VNI isolates tolerate high temperature better than VNB but their predominance in patients might also be explained by the close proximity between the common pigeon, *Columba livia* and densely populated areas. An ecological studies comparing urban and rural rates of cryptococcosis would help to understand the infection.

6.6.2 Clinical isolates do not belong to a particular genetic cluster

With the development of next-generation sequencing technologies, we have the resolution to investigate phylogenetic relationships at a finer scale. Within the VNI molecular type, isolates from any cluster seem to be able to cause disease in patients (Figure 5.15). Zoonotic transmission has been reported twice (Franzot et al., 1999; Nosanchuk et al., 2000). By sampling extensively urban centres and thanks to the increased resolution offered by NGS technologies, we should be able to investigate local infection in depth and understand possible zoonotic infections.

6.6.3 Inference on the population dynamics and the 'Out of Africa' hypothesis

Previous studies demonstrated that the VNI lineage had arisen from a basal African VNB population with an average time to most recent common ancestor of 4.5 million years and that VNI became globally distributed between 1,600 and 70,500 years ago (Simwami, 2011). Using our mutation rate derived from Lynchs paper, we found that the emergence of *Cng* molecular types are more recent, from 1.6 million to 45,000 years and that VNI became globally distributed between 45,000 to 900 years ago (Table 5.18).

The strong population structure observed between our eight VNI genetic clusters indicates that the current hypothesis of the existence of a pan-global VNI genotype that emerged once from 'Out of Africa'is incorrect and that the story is incomplete. Clusters 1, 2 and 3 might be endemic to the African continent. But cluster 4, which is likely to play a significant role in the global spread of *Cryptococcus neoformans* var. *grubii*. This cluster are recombining with all the other clusters. Cluster 4 was composed of isolates of Africa and Asia and might have originated from another continent. Additional sequencing efforts investigating isolates from Europe and

South America will continue to uncover the complex evolutionary history of this pathogen.

Nielsen et al., (2007) stated that the close interaction between *Cng* and pigeon or other avian species might have led to the dissemination throughout the world. Cryptococcal cells can tolerate high avian temperature and escape avian macrophage cells under unfavourable conditions (Johnston et al., 2016). Our data supports the hypothesis that the recent domestication of *Columbia livia* in the past 200 to 800 years and the expansion of urban centres might have led to an increase in *Cng* population size and contributed to the *Cryptococcus* epidemic. Intense recovery and sequencing of *Cng* isolates from densely populated urban centres will shed lights on role of *Columbia livia* in the dissemination of *Cryptococcus*. Such study would improve our understanding of *Cng* epidemiology and would reveal whether the VNI molecular type is responsible for urban-acquired infections.

6.7 Conclusion

This PhD project recovered *Cryptococcus neoformans* var. *grubii* from the environment and characterised its ecological niche investigating both biotic and abiotic factors. The study represents the first attempt to understand the community structure associated with an environmental pathogen. Fungal pathogens are geographically restricted to their realised environmental niche due to their inability to reproduce outside of it (Nielsen et al., 2007). The constant movement of microbes in our globalised world could lead to the emergence of outbreaks from environmental pathogens and understanding the biotic factors associated with such pathogens will help to foresee a possible emergence (Morse, 1995; Wolfe et al., 2007).

Whole-genome sequencing technologies revealed the complex evolutionary history of the *Cryptococcus*-species complex. The study highlighted genomic regions which were associated with spatial and climatic (temperature and precipitation) factors. Selection was investigated using summary statistics and uncovered genetic variants which appeared to be under selective pressures. Genes with known virulence function were found to be under positive selection in both VNI and VNB molecular types. The two *Cng* molecular types were associated with different environmental settings and were subdivided in several genetic clusters. The presence of clinical isolates in each of these clusters indicates that cryptococcal infection is not associated with any particular genotype. Isolates from each molecular type can cause cryptococcal meningitis

but clinical isolates displayed phenotypic features which indicate that they are more prone to establish infection in human than environmental isolates.

The high proportion of MAT a locus within VNB suggests that recombination frequently occurs within this clade, whereas VNI isolates displayed a clonal profile. The latter is commonly associated with the feral pigeon. Therefore, VNI infections are probably acquired in urban centres while VNB *Cng* are mainly present in *C. mopane* region and other rural settings. In vitro phenotypic assays revealed some phenotypic differences between the two molecular types. When comparing environmental and clinical isolates, clinical isolates produced larger capsules and were more thermotolerant than environmental isolates (p <0.002). Some gene candidates were found which could potentially explained these differences, but expression analyses or gene knockout experiments are required to confirm the role of these putative genes in the infection process. Finally, the study highlighted the complexity of the epidemiology of the VNI molecular types revealing that each cluster is likely to contribute differently to the *Cryptococcus* epidemic.

6.8 Principal findings

1. *C. neoformans* and *C. gattii* were significantly associated with different ecoregions in Zambia. *Cn* was associated in the Mopane woodlands and *Cg* was isolated in the wet Miombo woodlands.

2. The study uncovered a new ecological niche for *C. gattii*. 21% of hyrax faeces were culture- positive indicating gaps in our knowledge regarding this *Cryptococcus spp.* and its association with small mammals.

3. The fungal community structure of the two ecoregion associated with either species of *Cryptococcus* differed. The composition of the fungal community were influenced by seasonality but remains similar over distance within an ecoregion.

4. The different *Cng* molecular types are under different environmental pressures and have highly different genetic profiles

5. Genome-scans successfully identified genomic region under selection which were subsequently mapped to the genome.

6. The two *Cng* molecular types had different recombination profile. Recombination appears frequent within the VNB molecular type but seems restricted among VNI isolates.

7. Clinical cases have a different genetic profile than the environmental isolates. The present study was able to identify genes which differed between clinical and environmental isolates.

8. CCNV appears to play an important role in the *Cng* infection process. VNB clinical isolates were more frequently observed with chromosomal duplication (59%) than VNI isolates (17%). However, CCNV do not seem to target any particular virulence associated genes.

9. Although environmental VNB replicated faster than VNI isolates at ambient temperature, clinical VNI isolates replicated faster than clinical VNB isolates at room temperature. This finding might be the reflection of the long evolutionary relationship between the feral pigeon and the VNI molecular type.

10. A ten-fold increase in population size was observed in each cluster in the past 200 to 8000 years. The VNI molecular type probably did not emerge 'out of Africa'and each VNI clusters probably contribute in a different way to the *Cryptococcus* 'neglected epidemic'.

Bibliography

Abarenkov, Kessy; Henrik Nilsson, R; Larsson, KarlHenrik; Alexander, Ian J; Eberhardt, Ursula; Erland, Susanne; Hø iland, Klaus; Kjø ller, Rasmus; Larsson, Ellen, and Pennanen, Taina. The UNITE database for molecular identification of fungirecent updates and future perspectives. *New Phytologist*, 186(2):281–285, 2010. ISSN 1469-8137.

Abdolrasouli, A; Rhodes, J; Beale, M; Hagen, F; Rogers, T R; Chowdhary, A; Meis, J F, and Fisher, M C. Genomic context of Azole-resistance mutations in Aspergillus fumigatus using whole-genome sequencing. 2015. ISSN 2161-2129.

Adams, Rachel I; Miletto, Marzia; Taylor, John W, and Bruns, Thomas D. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME journal*, 7(7):1262–1273, 2013. ISSN 1751-7362.

Aguileta, Gabriela; Refregier, Guislaine; Yockteng, Roxana; Fournier, Elisabeth, and Giraud, Tatiana. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, 9 (4):656–670, 2009. ISSN 1567-1348.

Ajello, Libero. Occurrence of Cryptococcus neoformans in soils. *American Journal of Epidemiology*, 67(1):72–77, 1958. ISSN 0002-9262.

Akey, Joshua M. Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome research*, 19(5):711–722, 2009. ISSN 1088-9051.

Aksenov, S I; Babyeva, I P, and Golubev, V I. On the mechanism of adaptation of microorganisms to conditions of extreme low humidity. *Life sciences and space research*, 11:55, 1973.

Alexander, David H; Novembre, John, and Lange, Kenneth. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9):1655–1664, 2009. ISSN 1088-9051.

Alspaugh, J Andrew; Pukkila-Worley, Read; Harashima, Toshiaki; Cavallo, Lora M; Funnell, Deanna; Cox, Gary M; Perfect, John R; Kronstad, James W, and Heitman, Joseph. Adenylyl cyclase functions downstream of the Gα protein Gpa1 and controls mating and pathogenicity of Cryptococcus neoformans. *Eukaryotic Cell*, 1(1):75–84, 2002. ISSN 1535-9778.

Amambua-Ngwa, Alfred; Tetteh, Kevin K A; Manske, Magnus; Gomez-Escobar, Natalia; Stewart, Lindsay B; Deerhake, M Elizabeth; Cheeseman, Ian H; Newbold, Christopher I; Holder, Anthony A, and Knuepfer, Ellen. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. 2012. ISSN 1553-7404.

Anderson, Marti J. A new method for nonparametric multivariate analysis of variance. *Austral ecology*, 26(1):32–46, 2001. ISSN 1442-9993.

Anderson, Pamela K; Cunningham, Andrew A; Patel, Nikkita G; Morales, Francisco J; Epstein, Paul R, and Daszak, Peter. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10):535–544, 2004. ISSN 0169-5347.

Armstrong-James, Darius; Meintjes, Graeme, and Brown, Gordon D. A neglected epidemic: fungal infections in HIV/AIDS. *Trends in microbiology*, 22(3):120–127, 2014. ISSN 0966-842X.

Auwera, Geraldine A; Carneiro, Mauricio O; Hartl, Christopher; Poplin, Ryan; del Angel, Guillermo; LevyMoonshine, Ami; Jordan, Tadeusz; Shakir, Khalid; Roazen, David, and Thibault, Joel. From FastQ data to highconfidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, pages 10–11, 2013. ISSN 0471250953.

Baas-Becking, Lourens Gerhard Marinus. *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV, 1934.

Bahn, Yong-Sun; Kojima, Kaihei; Cox, Gary M, and Heitman, Joseph. A unique fungal two-component system regulates stress responses, drug sensitivity, sexual development, and vir-

ulence of Cryptococcus neoformans. *Molecular biology of the cell*, 17(7):3122–3135, 2006. ISSN 1059-1524.

Bálint, Miklós; Schmidt, PhilippAndré; Sharma, Rahul; Thines, Marco, and Schmitt, Imke. An Illumina metabarcoding pipeline for fungi. *Ecology and Evolution*, 2014. ISSN 2045-7758.

Bankevich, Anton; Nurk, Sergey; Antipov, Dmitry; Gurevich, Alexey A; Dvorkin, Mikhail; Kulikov, Alexander S; Lesin, Valery M; Nikolenko, Sergey I; Pham, Son, and Prjibelski, Andrey D. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5):455–477, 2012. ISSN 1066-5277.

Banks, Isaac R; Specht, Charles A; Donlin, Maureen J; Gerik, Kimberly J; Levitz, Stuart M, and Lodge, Jennifer K. A chitin synthase and its regulator protein are critical for chitosan production and growth of the fungal pathogen Cryptococcus neoformans. *Eukaryotic Cell*, 4 (11):1902–1912, 2005. ISSN 1535-9778.

Barnett, James A. A history of research on yeasts 8: taxonomy. *Yeast*, 21(14):1141–1193, 2004. ISSN 1097-0061.

Bauwens, L; Swinne, D; Vroey, Ch, and de Meurichy, W. Isolation of Cryptococcus neoformans van neoformans in the Aviaries of the Antwerp Zoological Gardens/Isolation von Cryptococcus neoformans van neoformans im Vogelhaus des Antwerpener Zoos. *Mycoses*, 29(7): 291–294, 1986. ISSN 1439-0507.

Bazzicalupo, Anna L; Bálint, Miklós, and Schmitt, Imke. Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecology*, 6(1):102–109, 2013. ISSN 1754-5048.

Beale, Mathew A; Sabiiti, Wilber; Robertson, Emma J; Fuentes-Cabrejo, Karen M; OHanlon, Simon J; Jarvis, Joseph N; Loyse, Angela; Meintjes, Graeme; Harrison, Thomas S, and May, Robin C. Genotypic Diversity Is Associated with Clinical Outcome and Phenotype in Cryptococcal Meningitis across Southern Africa. *PLoS Negl Trop Dis*, 9(6):e0003847, 2015. ISSN 1935-2735.

Bicanic, T and Harrison, T S. Cryptococcal meningitis. *British Medical Bulletin*, 72(1):99–118, 2004.

Bicanic, Tihana; Harrison, Thomas; Niepieklo, Alina; Dyakopu, Nontobeko, and Meintjes, Graeme. Symptomatic relapse of HIV-associated cryptococcal meningitis after initial fluconazole monotherapy: the role of fluconazole resistance and immune reconstitution. *Clinical infectious diseases*, 43(8):1069–1070, 2006.

Bicanic, Tihana; Meintjes, Graeme; Wood, Robin; Hayes, Madeleine; Rebe, Kevin; Bekker, Linda-Gail, and Harrison, Thomas. Fungal burden, early fungicidal activity, and outcome in cryptococcal meningitis in antiretroviral-naive or antiretroviral-experienced patients treated with amphotericin B or fluconazole. *Clinical infectious diseases*, 45(1):76–80, 2007.

Bicanic, Tihana; Meintjes, Graeme; Rebe, Kevin; Williams, Anthony; Loyse, Angela; Wood, Robin; Hayes, Madeleine; Jaffar, Shabbar, and Harrison, Thomas. Immune reconstitution inflammatory syndrome in HIV-associated cryptococcal meningitis: a prospective study. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 51(2):130–134, 2009a.

Bicanic, Tihana; Muzoora, Conrad; Brouwer, Annemarie E; Meintjes, Graeme; Longley, Nicky; Taseera, Kabanda; Rebe, Kevin; Loyse, Angela; Jarvis, Joseph, and Bekker, Linda-Gail. Independent association between rate of clearance of infection and clinical outcome of HIV-associated cryptococcal meningitis: analysis of a combined cohort of 262 patients. *Clinical infectious diseases*, 49(5):702–709, 2009b. ISSN 1058-4838.

Bigham, Abigail; Bauchet, Marc; Pinto, Dalila; Mao, Xianyun; Akey, Joshua M; Mei, Rui; Scherer, Stephen W; Julian, Colleen G; Wilson, Megan J, and López Herráez, David. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet*, 6(9):e1001116, 2010.

Blaalid, Rakel; Kumar, Surendra; Nilsson, R Henrik; Abarenkov, Kessy; Kirk, P M, and Kauserud, Hå vard. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular ecology resources*, 13(2):218–224, 2013. ISSN 1755-0998.

Blackwell, Meredith. The Fungi: 1, 2, 3. . . 5.1 million species? *American journal of botany*, 98 (3):426–438, 2011. ISSN 0002-9122.

Bloom, Amanda L M; Solomons, J T; Havel, Virginia E, and Panepinto, John C. Uncoupling of mRNA synthesis and degradation impairs adaptation to host temperature in Cryptococcus neoformans. *Molecular microbiology*, 89(1):65–83, 2013. ISSN 1365-2958.

Boekhout, Teun; Theelen, Bart; Diaz, Mara; Fell, Jack W; Hop, Wim C J; Abeln, Edwin C A; Dromer, Françoise, and Meyer, Wieland. Hybrid genotypes in the pathogenic yeast Cryptococcus neoformans. *Microbiology*, 147(4):891–907, 2001.

Boere, Gerard C and Stroud, David A. The flyway concept: what it is and what it isnt. *Waterbirds around the world*, pages 40–47, 2006.

Bokulich, Nicholas A; Thorngate, John H; Richardson, Paul M, and Mills, David A. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proceedings of the National Academy of Sciences*, 111(1):E139–E148, 2014. ISSN 0027-8424.

Bolliger, Janine; Lander, Tonya, and Balkenhol, Niko. Landscape genetics since 2003: status, challenges and future directions. *Landscape ecology*, 29(3):361–366, 2014. ISSN 0921-2973.

Bolton, Lorna A; Lobetti, R G; Evezard, D N; Picard, Jacqueline A; Nesbit, J W; Van Heerden, J, and Burroughs, R E J. Cryptococcosis in captive cheetah (Acinonyx jubatus): two cases: case report. *Journal of the South African Veterinary Association*, 70(1):p–35, 1999. ISSN 0038-2809.

Botha, Alfred. The importance and ecology of yeasts in soil. *Soil Biology and Biochemistry*, 43 (1):1–8, 2011. ISSN 0038-0717.

Boulware, David R; Meya, David B; Muzoora, Conrad; Rolfes, Melissa A; Huppler Hullsiek, Katherine; Musubire, Abdu; Taseera, Kabanda; Nabeta, Henry W; Schutz, Charlotte, and Williams, Darlisha A. Timing of antiretroviral therapy after diagnosis of cryptococcal meningitis. *New England Journal of Medicine*, 370(26):2487–2498, 2014. ISSN 0028-4793.

Bovers, M; Hagen, F, and Boekhout, T. Diversity of the Cryptococcus neoformans-Cryptococcus gattii species complex. *Revista iberoamericana de micologÃa*, 25(1):1004–1012, 2008.

Bovers, Marjan. Cryptococcus neoformans and Cryptococcus gattii: speciation in progress. 2007.

Bradburd, Gideon S; Ralph, Peter L, and Coop, Graham M. Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, 67(11):3258–3273, 2013. ISSN 1558-5646.

Brown, Jessica C S; Nelson, Justin; VanderSluis, Benjamin; Deshpande, Raamesh; Butts, Arielle; Kagan, Sarah; Polacheck, Itzhack; Krysan, Damian J; Myers, Chad L, and Madhani, Hiten D. Unraveling the Biology of a Fungal Meningitis Pathogen Using Chemical Genetics. *Cell*, 159(5):1168–1187, 2014. ISSN 0092-8674.

Bui, Tien; Lin, Xiaorong; Malik, Richard; Heitman, Joseph, and Carter, Dee. Isolates of Cryptococcus neoformans from infected animals reveal genetic exchange in unisexual, α mating type populations. *Eukaryotic cell*, 7(10):1771–1780, 2008.

Busby, J R. BIOCLIM-a bioclimate analysis and prediction system. *Plant Protection Quarterly (Australia)*, 1991. ISSN 0815-2195.

Busse, O. Uber parasitare zelleinschlusse und ihre zuchtung. . *Zentralbl.Bakteriol.*, 16:175–180, 1894.

Byrnes, Edmond J; Li, Wenjun; Lewit, Yonathan; Ma, Hansong; Voelz, Kerstin; Ren, Ping; Carter, Dee A; Chaturvedi, Vishnu; Bildfell, Robert J, and May, Robin C. Emergence and pathogenicity of highly virulent Cryptococcus gattii genotypes in the northwest United States. *PLoS pathogens*, 6(4):e1000850, 2010.

Caicedo, Luz Dary; Alvarez, María Inés; Delgado, Maritza, and Cárdenas, Alaín. Cryptococcus neoformans in bird excreta in the city zoo of Cali, Colombia. *Mycopathologia*, 147(3):121–124, 1999. ISSN 0301-486X.

Callejas, A; Ordonez, N; Rodriguez, M C, and Castañeda, E. First isolation of Cryptococcus neoformans var. gattii, serotype C, from the environment in Colombia. *Medical Mycology*, 36 (5):341–344, 1998.

Caporaso, J Gregory; Kuczynski, Justin; Stombaugh, Jesse; Bittinger, Kyle; Bushman, Frederic D; Costello, Elizabeth K; Fierer, Noah; Pena, Antonio Gonzalez; Goodrich, Julia K, and Gordon, Jeffrey I. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336, 2010. ISSN 1548-7091.

Caporaso, J Gregory; Lauber, Christian L; Walters, William A; Berg-Lyons, Donna; Huntley, James; Fierer, Noah; Owens, Sarah M; Betley, Jason; Fraser, Louise, and Bauer, Markus. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, 6(8):1621–1624, 2012. ISSN 1751-7362.

Carlson, Christopher S; Thomas, Daryl J; Eberle, Michael A; Swanson, Johanna E; Livingston, Robert J; Rieder, Mark J, and Nickerson, Deborah A. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome research*, 15(11):1553–1565, 2005. ISSN 1088-9051.

Casadevall, A; Steenbergen, J N, and Nosanchuk, J D. 'Ready made' virulence and 'dual use' virulence factors in pathogenic environmental fungi - the Cryptococcus neoformans paradigm. *Current opinion in microbiology*, 6(4):332–337, 2003.

Casadevall, Arturo. Evolution of intracellular pathogens. *Annu. Rev. Microbiol.*, 62:19–33, 2008. ISSN 0066-4227.

Casadevall, Arturo. Amoeba provide insight into the origin of virulence in pathogenic fungi. In *Recent Advances on Model Hosts*, pages 1–10. Springer, 2012. ISBN 1441956379.

Casadevall, Arturo and Perfect, John R. *Cryptococcus neoformans*, volume 595. ASM press Washington, DC, 1998.

Casadevall, Arturo; Rosas, Angel L, and Nosanchuk, Joshua D. Melanin and virulence in Cryptococcus neoformans. *Current opinion in microbiology*, 3(4):354–358, 2000.

Castellani, A. An amoeba growing in cultures of a yeast: fourth note. *J. Trop. Med. Hyg*, 33: 237, 1930.

CDC, . Cryptococcal meningitis: a deadly fungal disease among people living with HIV/AIDS. 2012.

Chaka, Wendy; Scharringa, Jelle; Verheul, A F; Verhoef, Jan; Van Strijp, ANTONIUS GERARDUS, and Hoepelman, Ilja Mohandas. Quantitative analysis of phagocytosis and killing of Cryptococcus neoformans by human peripheral blood mononuclear cells by flow cytometry. *Clinical and diagnostic laboratory immunology*, 2(6):753–759, 1995. ISSN 1556-6811.

Chakrabarti, Arunaloke; Jatana, M; Kumar, P; Chatha, L; Kaushal, A, and Padhye, Arvind A. Isolation of Cryptococcus neoformans var. gattii from Eucalyptus camaldulensis in India. *Journal of Clinical Microbiology*, 35(12):3340–3342, 1997. ISSN 0095-1137.

Chamary, J V; Parmley, Joanna L, and Hurst, Laurence D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, 7(2):98–108, 2006. ISSN 1471-0056.

Chang, Yun C and Kwon-Chung, K J. Complementation of a capsule-deficient mutation of Cryptococcus neoformans restores its virulence. *Molecular and Cellular Biology*, 14(7):4912–4919, 1994. ISSN 0270-7306.

Charlesworth, Deborah. Balancing selection and its effects on sequences in nearby genome regions. 2006. ISSN 1553-7404.

Charlier, Caroline; Chrétien, Fabrice; Baudrimont, Marielle; Mordelet, Elodie; Lortholary, Olivier, and Dromer, Françoise. Capsule Structure Changes Associated with Cryptococcus neoformans Crossing of the Blood-Brain Barrier. *The American journal of pathology*, 166(2): 421–432, 2005.

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

Chen, Lin-Chi; Blank, Erika S, and Casadevall, Arturo. Extracellular proteinase activity of Cryptococcus neoformans. *Clinical and diagnostic laboratory immunology*, 3(5):570–574, 1996.

Chen, Sharon; Sorrell, Tania; Nimmo, Graeme; Speed, Bryan; Currie, Bart; Ellis, David; Marriott, Deborah; Pfeiffer, Tania; Parr, Dinah, and Byth, Karen. Epidemiology and host-and variety-dependent characteristics of infection due to Cryptococcus neoformans in Australia and New Zealand. *Clinical infectious diseases*, 31(2):499–508, 2000.

Chen, Yuan; Litvintseva, Anastasia P; Frazzitta, Aubrey E; Haverkamp, Miriam R; Wang, Liuyang; Fang, Charles; Muthoga, Charles; Mitchell, Thomas G, and Perfect, John R. Comparative Analyses of Clinical and Environmental Populations of Cryptococcus neoformans in Botswana. *Molecular ecology*, 2015. ISSN 1365-294X.

Cheon, Seon Ah; Jung, Kwang-Woo; Chen, Ying-Lien; Heitman, Joseph; Bahn, Yong-Sun, and Kang, Hyun Ah. Unique evolution of the UPR pathway with a novel bZIP transcription factor, Hxl1, for controlling pathogenicity of Cryptococcus neoformans. 2011. ISSN 1553-7374.

Cheong, Wei-Hien; Tan, Yung-Chie; Yap, Soon-Joo, and Ng, Kee-Peng. ClicO FS: an interactive web-based service of Circos. *Bioinformatics*, page btv433, 2015. ISSN 1367-4803.

Chidumayo, E N. Forest degradation and recovery in a miombo woodland landscape in Zambia: 22 years of observations on permanent sample plots. *Forest Ecology and Management*, 291: 154–161, 2013. ISSN 0378-1127.

Chow, Siu-Kei and Casadevall, Arturo. Evaluation of Cryptococcus neoformans galactoxylomannanprotein conjugate as vaccine candidate against murine cryptococcosis. *Vaccine*, 29 (10):1891–1898, 2011.

Chowdhary, Anuradha; Randhawa, Harbans S; Prakash, Anupam, and Meis, Jacques F. Environmental prevalence of Cryptococcus neoformans and Cryptococcus gattii in India: an update. *Critical reviews in microbiology*, 38(1):1–16, 2012. ISSN 1040-841X.

Chrisman Alvarez M, Casadevall A, C J. Phagocytosis and non-lytic phagocytosis of Cryptococcus neoformans by, and from, Acanthamoeba castellanii. *Appl Environ Microbiol*, 76: 6056–6062, 2010.

Christ, Sabina; Wubet, Tesfaye; Theuerl, Susanne; Herold, Nadine, and Buscot, François. Fungal communities in bulk soil and stone compartments of different forest and soil types as revealed by a barcoding ITS rDNA and a functional laccase encoding gene marker. *Soil Biology and Biochemistry*, 43(6):1292–1299, 2011. ISSN 0038-0717.

Chun, Cheryl D; Liu, Oliver W, and Madhani, Hiten D. A link between virulence and homeostatic responses to hypoxia during infection by the human fungal pathogen Cryptococcus neoformans. *PLoS Pathog*, 3(2):e22, 2007.

Chun, Cheryl D; Brown, Jessica, and Madhani, Hiten D. A Major Role for Capsule-Independent Phagocytosis-Inhibitory Mechanisms in Mammalian Infection by Cryptococcus neoformans. *Cell host & microbe*, 9(3):243–251, 2011.

Clarke, K_ R. Non-parametric multivariate analyses of changes in community structure. *Australian journal of ecology*, 18:117, 1993. ISSN 0307-692X.

Coelho, Carolina; Bocca, Anamelia Lorenzetti, and Casadevall, Arturo. The tools for virulence of Cryptococcus neoformans. *Adv. Appl. Microbiol*, 87:1–41, 2014.

Collins, Tony J. ImageJ for microscopy. *Biotechniques*, 43(1 Suppl):25–30, 2007.

Connell, Joseph H. Effects of competition, predation by Thais lapillus, and other factors on natural populations of the barnacle Balanus balanoides. *Ecological Monographs*, 31(1):61–104, 1961. ISSN 0012-9615.

Corbett, Elizabeth L; Churchyard, Gavin J; Charalambos, Salome; Samb, Badara; Moloi, Vicky; Clayton, Tim C; Grant, Alison D; Murray, Jill; Hayes, Richard J, and De Cock, Kevin M. Morbidity and mortality in South African gold miners: impact of untreated disease due to human immunodeficiency virus. *Clinical infectious diseases*, 34(9):1251–1258, 2002.

Cornelissen, S; Botha, A; Conradie, W J, and Wolfaardt, G M. Shifts in community composition provide a mechanism for maintenance of activity of soil yeasts in the presence of elevated copper levels. *Canadian journal of microbiology*, 49(7):425–432, 2003.

Coûteaux, Marie-Madeleine and Darbyshire, John. Functional diversity amongst soil protozoa. *Applied Soil Ecology*, 10(3):229–237, 1998.

Cox, Gary M; Mukherjee, Jean; Cole, Garry T; Casadevall, Arturo, and Perfect, John R. Urease as a virulence factor in experimental cryptococcosis. *Infection and immunity*, 68(2):443–448, 2000.

Crabtree, Juliet N; Okagaki, Laura H; Wiesner, Darin L; Strain, Anna K; Nielsen, Judith N, and Nielsen, Kirsten. Titan cell production enhances the virulence of Cryptococcus neoformans. *Infection and immunity*, 80(11):3776–3785, 2012.

Cruickshank, Tami E and Hahn, Matthew W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13): 3133–3157, 2014. ISSN 1365-294X.

Cullingham, Catherine I; Cooke, Janice E K, and Coltman, David W. Crossspecies outlier detection reveals different evolutionary pressures between sister species. *New Phytologist*, 204(1):215–229, 2014. ISSN 1469-8137.

Currie, Brian P; Freundlich, Lawrence F, and Casadevall, Arturo. Restriction fragment length polymorphism analysis of Cryptococcus neoformans isolates from environmental (pigeon excreta) and clinical sources in New York City. *Journal of clinical microbiology*, 32(5):1188–1192, 1994. ISSN 0095-1137.

Dabney, Alan; Storey, John D, and Warnes, G. Q-value estimation for false discovery rate control. *Medicine*, 344:539–548, 2004.

Davidson, R C and Moore, Heitman J., Odom A R. Characterization of the MF pheromone of the human fungal pathogen Cryptococcus neoformans. *Mol Microbiol*, 38:1017–1026, 2000.

Day, Jeremy N. Cryptococcal meningitis. *Practical Neurology*, 4(5):274–285, 2004.

DePristo, Mark A; Banks, Eric; Poplin, Ryan; Garimella, Kiran V; Maguire, Jared R; Hartl, Christopher; Philippakis, Anthony A; del Angel, Guillermo; Rivas, Manuel A, and Hanna, Matt. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5):491–498, 2011.

Desalermos, Athanasios; Kourkoumpetis, Themistoklis K, and Mylonakis, Eleftherios. Update on the epidemiology and management of cryptococcal meningitis. *Expert Opinion on Pharmacotherapy*, 13(6):783–789, 2012.

Desnos-Ollivier, M; Patel, S; Spaulding, A R; Charlier, C; Garcia-Hermoso, D; Nielsen, K, and Dromer, F. Mixed infections and in vivo evolution in the human fungal pathogen Cryptococcus neoformans. *MBio*, 1(1), 2010.

Desprez-Loustau, Marie-Laure; Robin, Cécile; Buee, Marc; Courtecuisse, Regis; Garbaye, Jean; Suffert, Frédéric; Sache, Ivan, and Rizzo, David M. The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, 22(9):472–480, 2007. ISSN 0169-5347.

Ding, Chen; Festa, Richard A; Chen, Ying-Lien; Espart, Anna; Palacios, Òscar; Espín, Jordi; Capdevila, Mercè; Atrian, Sílvia; Heitman, Joseph, and Thiele, Dennis J. Cryptococcus neoformans copper detoxification machinery is critical for fungal virulence. *Cell host & microbe*, 13(3):265–276, 2013. ISSN 1931-3128.

Dromer, Françoise; Mathoulin, Simone; Dupont, Bertrand; Letenneur, Luc, and Ronin, Olivier. Individual and environmental factors associated with infection due to Cryptococcus neoformans serotype D. *Clinical infectious diseases*, 23(1):91–96, 1996.

Drummond, Alexei J and Rambaut, Andrew. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7(1):214, 2007.

D'Souza, Cletus A and Heitman, Joseph. Conserved cAMP signaling cascades regulate fungal development and virulence. *FEMS microbiology reviews*, 25(3):349–364, 2001.

Duforet-Frebourg, Nicolas; Bazin, Eric, and Blum, Michael G B. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular biology and evolution*, page msu182, 2014. ISSN 0737-4038.

Duncan, C; Stephen, C; Lester, S, and Bartlett, K H. Sub-clinical infection and asymptomatic carriage of Cryptococcus gattii in dogs and cats during an outbreak of cryptococcosis. *Medical Mycology*, 43(6):511–516, 2005.

Edgar, Robert C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461, 2010. ISSN 1367-4803.

Ellis, DAVID H and Pfeiffer, TANIA J. Natural habitat of Cryptococcus neoformans var. gattii. *Journal of Clinical Microbiology*, 28(7):1642–1644, 1990.

Ellison, Christopher E; Hall, Charles; Kowbel, David; Welch, Juliet; Brem, Rachel B; Glass, N L, and Taylor, John W. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*, 108(7):2831–2836, 2011. ISSN 0027-8424.

Emmons, Chester W. Isolation of Cryptococcus neoformans from soil. *Journal of bacteriology*, 62(6):685, 1951.

Emmons, Chester W. Saprophytic sources of Cryptococcus neoformans associated with the pigeon (Columba livia). *American Journal of Epidemiology*, 62(3):227–232, 1955. ISSN 0002-9262.

Engelthaler, David M; Hicks, Nathan D; Gillece, John D; Roe, Chandler C; Schupp, James M; Driebe, Elizabeth M; Gilgado, Felix; Carriconde, Fabian; Trilles, Luciana, and Firacative, Carolina. Cryptococcus gattii in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. *MBio*, 5(4):e01464–14, 2014. ISSN 2150-7511.

Enright, Mark C and Spratt, Brian G. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology*, 144(11):3049–3060, 1998. ISSN 1350-0872.

Ensley, P K; Davis, C E; Anderson, M P, and Fletcher, K C. Cryptococcosis in a male Beccari's crowned pigeon. *Journal of the American Veterinary Medical Association*, 175(9):992, 1979.

Erickson, Todd; Liu, Lide; Gueyikian, Ara; Zhu, Xudong; Gibbons, Jack, and Williamson, Peter R. Multiple virulence factors of Cryptococcus neoformans are dependent on VPH1. *Molecular microbiology*, 42(4):1121–1131, 2001. ISSN 1365-2958.

Excoffier, L; Hofer, T, and Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4):285–298, 2009. ISSN 0018-067X.

Fan, Weihua; Idnurm, Alexander; Breger, Julia; Mylonakis, Eleftherios, and Heitman, Joseph. Eca1, a sarcoplasmic/endoplasmic reticulum Ca2+-ATPase, is involved in stress tolerance and virulence in Cryptococcus neoformans. *Infection and immunity*, 75(7):3394–3405, 2007. ISSN 0019-9567.

Fang, Wei; Price, Michael S; Toffaletti, Dena L; Tenor, Jennifer; Betancourt-Quiroz, Marisol; Price, Jennifer L; Pan, W H; Liao, W Q, and Perfect, John R. Pleiotropic effects of deubiquitinating enzyme Ubp5 on growth and pathogenesis of Cryptococcus neoformans. *PloS one*, 7(6):e38326, 2012. ISSN 1932-6203.

Farhi, Fereshteh; Bulmer, G S, and Tacker, J R. Cryptococcus neoformans IV. The not-so-encapsulated yeast. *Infection and immunity*, 1(6):526–531, 1970.

Fay, Justin C and Wu, Chung-I. Hitchhiking under positive Darwinian selection. *Genetics*, 155 (3):1405–1413, 2000. ISSN 0016-6731.

Feldmesser, Marta; Kress, Yvonne; Novikoff, Phyllis, and Casadevall, Arturo. Cryptococcus neoformans is a facultative intracellular pathogen in murine pulmonary infection. *Infection and immunity*, 68(7):4225–4237, 2000.

Feldmesser, Marta; Kress, Yvonne, and Casadevall, Arturo. Dynamic changes in the morphology of Cryptococcus neoformans during murine pulmonary infection. *Microbiology*, 147(8): 2355–2365, 2001. ISSN 1350-0872.

Fierer, Noah; Breitbart, Mya; Nulton, James; Salamon, Peter; Lozupone, Catherine; Jones, Ryan; Robeson, Michael; Edwards, Robert A; Felts, Ben, and Rayhawk, Steve. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and

viruses in soil. *Applied and environmental microbiology*, 73(21):7059–7066, 2007. ISSN 0099-2240.

Fierer, Noah; Lauber, Christian L; Zhou, Nick; McDonald, Daniel; Costello, Elizabeth K, and Knight, Rob. Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences*, 107(14):6477–6481, 2010. ISSN 0027-8424.

Filion, T; Kidd, S, and Aguirre, K. Isolation of Cryptococcus laurentii from Canada Goose guano in rural upstate New York. *Mycopathologia*, 162(5):363–368, 2006.

Fisher, Matthew C; Henk, Daniel A; Briggs, Cheryl J; Brownstein, John S; Madoff, Lawrence C; McCraw, Sarah L, and Gurr, Sarah J. Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484(7393):186–194, 2012.

Foll, Matthieu and Gaggiotti, Oscar. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180(2): 977–993, 2008. ISSN 0016-6731.

Forester, Brenna R; Jones, Matthew R; Joost, Stéphane; Landguth, Erin L, and Lasky, Jesse R. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular ecology*, 2015. ISSN 1365-294X.

Franzot, S P; Salkin, I F, and Casadevall, A. Cryptococcus neoformans var. grubii: separate varietal status for Cryptococcus neoformans serotype A isolates. *Journal of Clinical Microbiology*, 37(3):838–840, 1999.

Fraser, J A; Giles, S S; Wenink, E C; Geunes-Boyer, S G; Wright, J R; Diezmann, S; Allen, A; Stajich, J E; Dietrich, F S, and Perfect, J R. Same-sex mating and the origin of the Vancouver Island Cryptococcus gattii outbreak. *Nature*, 437(7063):1360–1364, 2005.

Fraser, James A; Diezmann, Stephanie; Subaran, Ryan L; Allen, Andria; Lengeler, Klaus B; Dietrich, Fred S, and Heitman, Joseph. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS biology*, 2(12):e384, 2004.

French, Neil; Gray, Katherine; Watera, Christine; Nakiyingi, Jessica; Lugada, Eric; Moore, Michael; Lalloo, David; Whitworth, James A G, and Gilks, Charles F. Cryptococcal infection in a cohort of HIV-1-infected Ugandan adults. *Aids*, 16(7):1031–1038, 2002.

Frichot, Eric; Schoville, Sean D; Bouchard, Guillaume, and François, Olivier. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular biology and evolution*, 30(7):1687–1699, 2013. ISSN 0737-4038.

Fries, Bettina C; Chen, Fengying; Currie, Brian P, and Casadevall, Arturo. Karyotype instability in Cryptococcus neoformans infection. *Journal of clinical microbiology*, 34(6):1531–1534, 1996. ISSN 0095-1137.

Fu, Yun-Xin. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2):915–925, 1997. ISSN 0016-6731.

Fumagalli, Matteo; Pozzoli, Uberto; Cagliani, Rachele; Comi, Giacomo P; Riva, Stefania; Clerici, Mario; Bresolin, Nereo, and Sironi, Manuela. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *The Journal of experimental medicine*, 206(6):1395–1408, 2009. ISSN 0022-1007.

Gans, Jason; Wolinsky, Murray, and Dunbar, John. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, 309(5739):1387–1390, 2005. ISSN 0036-8075.

Garcia-Hermoso, D; Mathoulin-Pelissier, Simone; Couprie, Bernard; Ronin, Olivier; Dupont, Bertrand, and Dromer, Francoise. DNA typing suggests pigeon droppings as a source of pathogenic Cryptococcus neoformans serotype D. *Journal of clinical microbiology*, 35(10): 2683–2685, 1997. ISSN 0095-1137.

Garcia-Rivera, J and Casadevall, A. Melanization of Cryptococcus neoformans reduces its susceptibility to the antimicrobial effects of silver nitrate. *Medical Mycology*, 39(4):353–357, 2001.

Gardes, Monique and Bruns, Thomas D. ITS primers with enhanced specificity for basidiomycetesapplication to the identification of mycorrhizae and rusts. *Molecular ecology*, 2(2): 113–118, 1993. ISSN 1365-294X.

Gardy, Jennifer L; Naus, Monika; Amlani, Ashraf; Chung, Walter; Kim, Hochan; Tan, Malcolm; Severini, Alberto; Krajden, Mel; Puddicombe, David, and Sahni, Vanita. Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following

the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *Journal of Infectious Diseases*, page jiv271, 2015. ISSN 0022-1899.

Gerik, Kimberly J; Donlin, Maureen J; Soto, Carlos E; Banks, Annette M; Banks, Isaac R; Maligie, Marybeth A; Selitrennikoff, Claude P, and Lodge, Jennifer K. Cell wall integrity is dependent on the PKC1 signal transduction pathway in Cryptococcus neoformans. *Molecular microbiology*, 58(2):393–408, 2005. ISSN 1365-2958.

Ghannoum, Mahmoud A. Potential role of phospholipases in virulence and fungal pathogenesis. *Clinical Microbiology Reviews*, 13(1):122–143, 2000.

Gilbert, Jack A; Meyer, Folker; Antonopoulos, Dion; Balaji, Pavan; Brown, C Titus; Brown, Christopher T; Desai, Narayan; Eisen, Jonathan A; Evers, Dirk, and Field, Dawn. Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Standards in genomic sciences*, 3(3):243, 2010a.

Gilbert, Nicole M; Donlin, Maureen J; Gerik, Kimberly J; Specht, Charles A; Djordjevic, Julianne T; Wilson, Christabel F; Sorrell, Tania C, and Lodge, Jennifer K. KRE genes are required for β1, 6glucan synthesis, maintenance of capsule architecture and cell wall protein anchoring in Cryptococcus neoformans. *Molecular microbiology*, 76(2):517–534, 2010b. ISSN 1365-2958.

Giles, Steven S; Stajich, Jason E; Nichols, Connie; Gerrald, Quincy D; Alspaugh, J Andrew; Dietrich, Fred, and Perfect, John R. The Cryptococcus neoformans catalase gene family and its role in antioxidant defense. *Eukaryotic cell*, 5(9):1447–1459, 2006. ISSN 1535-9778.

Goddard, Matthew R; Godfray, H Charles J, and Burt, Austin. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature*, 434(7033):636–640, 2005.

Gokulshankar, S; Ranganathan, S; Ranjith, M S, and Ranjithsingh, A J A. Prevalence, serotypes and mating patterns of Cryptococcus neoformans in the pellets of different avifauna in Madras, India. *Mycoses*, 47(7):310–314, 2004.

Goldman, David L; Fries, Bettina C; Franzot, Sarah P; Montella, Lya, and Casadevall, Arturo. Phenotypic switching in the human pathogenic fungus Cryptococcus neoformans is associated with changes in virulence and pulmonary inflammatory response in rodents. *Proceedings of the National Academy of Sciences*, 95(25):14967–14972, 1998.

Goldman, David L; Khine, Hnin; Abadi, Jacob; Lindenberg, Dania J; Pirofski, Liise-anne; Niang, Ramata, and Casadevall, Arturo. Serologic evidence for Cryptococcus neoformans infection in early childhood. *Pediatrics*, 107(5):e66–e66, 2001.

Gordon, S B and Read, R C. Macrophage defences against respiratory tract infections The immunology of childhood respiratory infections. *British medical bulletin*, 61(1):45–61, 2002. ISSN 0007-1420.

Goslee, Sarah C and Urban, Dean L. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22(7):1–19, 2007.

Gow, N A R; Robson, G D, and GADD, G M. The fungal colony, 1997.

Granados, D P and Castañeda, E. Isolation and characterization of Cryptococcus neoformans varieties recovered from natural sources in Bogotá, Colombia, and study of ecological conditions in the area. *Microbial ecology*, 49(2):282–290, 2005a. ISSN 0095-3628.

Granados, D P and Castañeda, E. Isolation and characterization of Cryptococcus neoformans varieties recovered from natural sources in Bogotá, Colombia, and study of ecological conditions in the area. *Microbial ecology*, 49(2):282–290, 2005b. ISSN 0095-3628.

Granados, Diana Paola and Castañeda, Elizabeth. Influence of climatic conditions on the isolation of members of the Cryptococcus neoformans species complex from trees in Colombia from 19922004. *FEMS yeast research*, 6(4):636–644, 2006. ISSN 1567-1364.

Granger, Donald L; Perfect, John R, and Durack, DavidT. Virulence of Cryptococcus neoformans. Regulation of capsule synthesis by carbon dioxide. *Journal of Clinical Investigation*, 76(2):508, 1985.

Griffiths, E J; Kretschmer, M, and Kronstad, J W. Aimless mutants of Cryptococcus neoformans: failure to disseminate. *Fungal biology reviews*, 26(2):61–72, 2012. ISSN 1749-4613.

Guillot, Gilles and Rousset, François. Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 4(4):336–344, 2013. ISSN 2041-210X.

Günther, Torsten and Coop, Graham. Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1):205–220, 2013. ISSN 0016-6731.

Haag-Wackernagel, Daniel and Moch, H. Health hazards posed by feral pigeons. *Journal of Infection*, 48(4):307–313, 2004.

Haasl, Ryan J and Payseur, Bret A. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular ecology*, 2015. ISSN 1365-294X.

Haasl, Ryan J and Payseur, Bret A. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular ecology*, 25(1):5–23, 2016. ISSN 1365-294X.

Hagen, Ferry; Khayhan, Kantarawee; Theelen, Bart; Kolecka, Anna; Polacheck, Itzhack; Sionov, Edward; Falk, Rama; Parnmen, Sittiporn; Lumbsch, H Thorsten, and Boekhout, Teun. Recognition of seven species in the Cryptococcus gattii/Cryptococcus neoformans species complex. *Fungal Genetics and Biology*, 2015. ISSN 1087-1845.

Hakim, James G; Gangaidzo, Innocent T; Heyderman, Robert S; Mielke, Jens; Mushangi, Ebbah; Taziwa, Albert; Robertson, Valerie J; Musvaire, Praise, and Mason, Peter R. Impact of HIV infection on meningitis in Harare, Zimbabwe: a prospective study of 406 predominantly adult patients. *Aids*, 14(10):1401–1407, 2000.

Hall, Tom. BioEdit: An important software for molecular biology. *GERF Bull. Biosci*, 2:60–61, 2011.

Hamilton, William D. Sex versus non-sex versus parasite. *Oikos*, pages 282–290, 1980. ISSN 0030-1299.

Hardin, Garrett. The competitive exclusion principle. *science*, 131(3409):1292–1297, 1960. ISSN 0036-8075.

Hawksworth, D L. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological research*, 105(12):1422–1432, 2001.

Hedayati, Mohammad T; Mayahi, Sabah; Fakhar, Mahdi; Shokohi, Tahereh, and Majidi, Mohammad. Cryptococcus neoformans isolation from swallow (Hirundo rustica) excreta in Iran. *Revista do Instituto de Medicina Tropical de São Paulo*, 53(3):125–127, 2011. ISSN 0036-4665.

Heled, Joseph and Drummond, Alexei J. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1):289, 2008. ISSN 1471-2148.

Hengl, Tomislav; Heuvelink, Gerard B M; Kempen, Bas; Leenaars, Johan G B; Walsh, Markus G; Shepherd, Keith D; Sila, Andrew; MacMillan, Robert A; de Jesus, Jorge Mendes, and Tamene, Lulseged. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PloS one*, 10(6):e0125814, 2015. ISSN 1932-6203.

Hicks, Julie K; Bahn, Yong-Sun, and Heitman, Joseph. Pde1 phosphodiesterase modulates cyclic AMP levels through a protein kinase A-mediated negative feedback loop in Cryptococcus neoformans. *Eukaryotic Cell*, 4(12):1971–1981, 2005. ISSN 1535-9778.

Hijmans, Robert J and van Etten, Jacob. raster: Geographic analysis and modeling with raster data. *R package version*, 1:9–92, 2012.

Hijmans, Robert J; Phillips, Steven; Leathwick, John, and Elith, Jane. dismo: Species distribution modeling. *R package version 0.7-17*, 2012.

Hiremath, Sanjay S; Chowdhary, Anuradha; Kowshik, Tusharantak; Randhawa, Harbans S; Sun, Sheng, and Xu, Jianping. Long-distance dispersal and recombination in environmental populations of Cryptococcus neoformans var. grubii from India. *Microbiology*, 154(5):1513–1524, 2008.

Hoang, Linda M N; Maguire, John A; Doyle, Patrick; Fyfe, Murray, and Roscoe, Diane L. Cryptococcus neoformans infections at Vancouver Hospital and Health Sciences Centre (19972002): epidemiology, microbiology and histopathology. *Journal of medical microbiology*, 53(9):935–940, 2004.

Hodkinson, Brendan P and Grice, Elizabeth A. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in wound care*, 4(1):50–58, 2015. ISSN 2162-1918.

Hofreuter, Dirk; Tsai, Jennifer; Watson, Robert O; Novik, Veronica; Altman, Bill; Benitez, Michelle; Clark, Christina; Perbost, Clotilde; Jarvie, Thomas, and Du, Lei. Unique features of a highly pathogenic Campylobacter jejuni strain. *Infection and Immunity*, 74(8):4694–4707, 2006. ISSN 0019-9567.

Hohenlohe, Paul A; Bassham, Susan; Etter, Paul D; Stiffler, Nicholas; Johnson, Eric A, and Cresko, William A. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*, 6(2):e1000862, 2010.

Holderegger, Rolf and Wagner, Helene H. Landscape genetics. *Bioscience*, 58(3):199–207, 2008. ISSN 0006-3568.

Hu, Guanggan; Liu, Iris; Sham, Anita; Stajich, Jason E; Dietrich, Fred S, and Kronstad, James W. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen Cryptococcus neoformans. *Genome biology*, 9(2):R41, 2008. ISSN 1465-6906.

Hu, Guanggan; Wang, Joyce; Choi, Jaehyuk; Jung, Won H; Liu, Iris; Litvintseva, Anastasia P; Bicanic, Tihana; Aurora, Rajeev; Mitchell, Thomas G, and Perfect, John R. Variation in chromosome copy number influences the virulence of Cryptococcus neoformans and occurs in isolates from AIDS patients. *BMC genomics*, 12(1):526, 2011. ISSN 1471-2164.

Hudson, Richard R. Two-locus sampling distributions and their application. *Genetics*, 159(4): 1805–1817, 2001. ISSN 0016-6731.

Huffnagle, Gary B; Chen, Gwo-Hsiao; Curtis, Jeffrey L; McDonald, Roderick A; Strieter, Robert M, and Toews, Galen B. Down-regulation of the afferent phase of T cell-mediated pulmonary inflammation and immunity by a high melanin-producing strain of Cryptococcus neoformans. *The Journal of Immunology*, 155(7):3507–3516, 1995.

Hull, Christina M; Davidson, Robert C, and Heitman, Joseph. Cell identity and sexual development in Cryptococcus neoformans are controlled by the mating-type-specific homeodomain protein Sxi1α. *Genes & development*, 16(23):3046–3060, 2002. ISSN 0890-9369.

Hurst, Laurence D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *TRENDS in Genetics*, 18(9):486–487, 2002. ISSN 0168-9525.

Hutchinson, G Evelyn. Homage to Santa Rosalia or why are there so many kinds of animals? *American naturalist*, pages 145–159, 1959. ISSN 0003-0147.

Hutter, Stephan; Vilella, Albert J, and Rozas, Julio. Genome-wide DNA polymorphism analyses using VariScan. *BMC bioinformatics*, 7(1):409, 2006. ISSN 1471-2105.

Hyatt, Doug; Chen, Gwo-Liang; LoCascio, Philip F; Land, Miriam L; Larimer, Frank W, and Hauser, Loren J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11(1):119, 2010. ISSN 1471-2105.

Idnurm, Alexander; Walton, Felicia J; Floyd, Anna; Reedy, Jennifer L, and Heitman, Joseph. Identification of ENA1 as a virulence gene of the human pathogenic fungus Cryptococcus neoformans through signature-tagged insertional mutagenesis. *Eukaryotic cell*, 8(3):315–326, 2009. ISSN 1535-9778.

Ihrmark, Katarina; Bödeker, Inga T M; Cruz-Martinez, Karelyn; Friberg, Hanna; Kubartova, Ariana; Schenck, Jessica; Strid, Ylva; Stenlid, Jan; Brandström-Durling, Mikael, and Clemmensen, Karina E. New primers to amplify the fungal ITS2 regionevaluation by 454-sequencing of artificial and natural communities. *FEMS microbiology ecology*, 82(3):666–677, 2012. ISSN 1574-6941.

Illnait-Zaragozi, Maria-Teresa; Martínez-Machín, Gerardo F; Fernández-Andreu, Carlos M; Boekhout, Teun; Meis, Jacques F, and Klaassen, Corné H W. Microsatellite typing of clinical and environmental Cryptococcus neoformans var. grubii isolates from Cuba shows multiple genetic lineages. *PLOS ONE*, 5(2):e9124, 2010.

Jackson, Arthur and Hosseinipour, Mina C. Management of cryptococcal meningitis in sub-Saharan Africa. *Current HIV/AIDS Reports*, 7(3):134–142, 2010.

Janbon, Guilhem; Himmelreich, Uwe; Moyrand, Frédérique; Improvisi, Luce, and Dromer, Françoise. Cas1p is a membrane protein necessary for the Oacetylation of the Cryptococcus neoformans capsular polysaccharide. *Molecular microbiology*, 42(2):453–467, 2001. ISSN 1365-2958.

Johnson, Marc T J and Stinchcombe, John R. An emerging synthesis between community ecology and evolutionary biology. *Trends in Ecology & Evolution*, 22(5):250–257, 2007. ISSN 0169-5347.

Johnston, Simon A; Voelz, Kerstin, and May, Robin C. Cryptococcus neoformans Thermotolerance to Avian Body Temperature Is Sufficient For Extracellular Growth But Not Intracellular Survival In Macrophages. *Scientific Reports*, 6: 20977, February 2016. URL `http://dx.doi.org/10.1038/srep20977 10.1038/srep20977 http://www.nature.com/articles/srep20977#supplementary-information`.

Jung, Won Hee; Saikia, Sanjay; Hu, Guanggan; Wang, Joyce; Fung, C K; DSouza, Cletus; White, Rick, and Kronstad, James W. HapX positively and negatively regulates the transcriptional response to iron deprivation in Cryptococcus neoformans. *PLoS Pathog*, 6(11): e1001209–e1001209, 2010.

Kõljalg, Urmas; Nilsson, R Henrik; Abarenkov, Kessy; Tedersoo, Leho; Taylor, Andy F S; Bahram, Mohammad; Bates, Scott T; Bruns, Thomas D; BengtssonPalme, Johan, and Callaghan, Tony M. Towards a unified paradigm for sequencebased identification of fungi. *Molecular Ecology*, 22(21):5271–5277, 2013. ISSN 1365-294X.

Kangogo, M; Boga, H; Wanyoike, W, and Bii, C. Isolation and characterisation of Cryptococcus neoformans and Cryptococcus gattii from environmental sources in Nairobi, Kenya. *East African Medical Journal*, 91(8):281–285, 2014. ISSN 0012-835X.

Karl, Stephen A; Toonen, R J; Grant, W S, and Bowen, B W. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology*, 21(17):4171–4189, 2012. ISSN 1365-294X.

Katoh, Kazutaka and Standley, Daron M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4): 772–780, 2013. ISSN 0737-4038.

Kawecki, Tadeusz J and Ebert, Dieter. Conceptual issues in local adaptation. *Ecology letters*, 7(12):1225–1241, 2004. ISSN 1461-0248.

Kearney, Michael and Porter, Warren P. Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, 85(11):3119–3131, 2004. ISSN 0012-9658.

Khan, Z U; Pal, M; Randhawa, H S, and Sandhu, R S. Carriage of Cryptococcus neoformans in the crops of pigeons. *Journal of medical microbiology*, 11(2):215–218, 1978. ISSN 0022-2615.

Kidd, S E; Hagen, F; Tscharke, R L; Huynh, M; Bartlett, K H; Fyfe, M; Macdougall, L; Boekhout, T; Kwon-Chung, K J, and Meyer, W. A rare genotype of Cryptococcus gattii caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). *Proceedings of the National Academy of Sciences of the United States of America*, 101(49): 17258–17263, 2004. ISSN 0027-8424.

Kidd, Sarah E; Chow, Yat; Mak, Sunny; Bach, Paxton J; Chen, Huiming; Hingston, Adrian O; Kronstad, James W, and Bartlett, Karen H. Characterization of environmental sources of the human and animal pathogen Cryptococcus gattii in British Columbia, Canada, and the Pacific Northwest of the United States. *Applied and environmental microbiology*, 73(5): 1433–1443, 2007. ISSN 0099-2240.

Kim, Yuseob and Nielsen, Rasmus. Linkage disequilibrium as a signature of selective sweeps. *Genetics*, 167(3):1513–1524, 2004. ISSN 0016-6731.

Kmetzsch, Lívia; Staats, Charley Christian; Simon, Elisa; Fonseca, Fernanda L; Oliveira, Débora L; Joffe, Luna S; Rodrigues, Jéssica; Lourenço, Rogério F; Gomes, Suely L, and Nimrichter, Leonardo. The GATA-type transcriptional activator Gat1 regulates nitrogen uptake and metabolism in the human pathogen Cryptococcus neoformans. *Fungal Genetics and Biology*, 48(2):192–199, 2011. ISSN 1087-1845.

Ko, Young-Joon; Yu, Yeong Man; Kim, Gyu-Bum; Lee, Gir-Won; Maeng, Pil Jae; Kim, Sang-soo; Floyd, Anna; Heitman, Joseph, and Bahn, Yong-Sun. Remodeling of global transcription patterns of Cryptococcus neoformans genes mediated by the stress-activated HOG signaling pathways. *Eukaryotic cell*, 8(8):1197–1217, 2009. ISSN 1535-9778.

Koren, Omry; Goodrich, Julia K; Cullender, Tyler C; Spor, Aymé; Laitinen, Kirsi; Bäckhed, Helene Kling; Gonzalez, Antonio; Werner, Jeffrey J; Angenent, Largus T, and Knight, Rob. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*, 150 (3):470–480, 2012. ISSN 0092-8674.

Kozak, Kenneth H; Graham, Catherine H, and Wiens, John J. Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, 23(3):141–148, 2008. ISSN 0169-5347.

Kozel, Thomas R; Pfrommer, Gail S T; Guerlain, Ann S; Highison, Barbara A, and Highison, Gregory J. Role of the capsule in phagocytosis of Cryptococcus neoformans. *Reviews of infectious diseases*, pages S436–S439, 1988.

Kreger-van Rij, Nelly Jeanne Wilhelmina. *The yeasts: a taxonomic study*. Elsevier, 2013. ISBN 1483290174.

Kronstad, J W; Attarian, R; Cadieux, B; Choi, J; D'Souza, C A; Griffiths, E J; Geddes, J M H; Hu, G; Jung, W H, and Kretschmer, M. Expanding fungal pathogenesis: Cryptococcus breaks out of the opportunistic box. *Nature Reviews Microbiology*, 9(3):193–203, 2011.

Kruzel, Emilia K; Giles, Steven S, and Hull, Christina M. Analysis of Cryptococcus neoformans sexual development reveals rewiring of the pheromone-response network by a change in transcription factor identity. *Genetics*, 191(2):435–449, 2012. ISSN 0016-6731.

Kües, Ursula; James, Timothy Y, and Heitman, Joseph. 6 Mating Type in Basidiomycetes: Unipolar, Bipolar, and Tetrapolar Patterns of Sexuality. In *Evolution of fungi and fungal-like organisms*, pages 97–160. Springer, 2011. ISBN 3642199739.

Kuroki, Misuzu; Phichaichumpon, Cheseda; Yasuoka, Akira; Chiranairadul, Piphat; Chosa, Toru; Sirinirund, Petchsri; Miyazaki, Taiga; Kakeya, Hiroshi; Higashiyama, Yasuhito, and Miyazaki, Yoshitsugu. Environmental isolation of Cryptococcus neoformans from an endemic region of HIVassociated cryptococcosis in Thailand. *Yeast*, 21(10):809–812, 2004. ISSN 1097-0061.

Kwon-Chung, K J. A new genus, Filobasidiella, the perfect state of Cryptococcus neoformans. *Mycologia*, 67(6):1197–1200, 1975.

Kwon-Chung, K J. Gene disruption to evaluate the role of fungal candidate virulence genes. *Current opinion in microbiology*, 1(4):381–389, 1998.

Kwon-Chung, K J; Bennett, J E, and Rhodes, J C. Taxonomic studies on Filobasidiella species and their anamorphs. *Antonie van Leeuwenhoek*, 48(1):25–38, 1982.

Kwon-Chung, Kyung J; Boekhout, Teun; Fell, Jack W, and Diaz, Mara. Proposal to conserve the name Cryptococcus gattii against C. hondurianus and C. bacillisporus (Basidiomycota, Hymenomycetes, Tremellomycetidae). *Taxon*, 51(4):804–806, 2002.

Lachance, Marc-André. Yeast biodiversity: how many and how much? In *Biodiversity and ecophysiology of yeasts*, pages 1–9. Springer, 2006. ISBN 3540261001.

Lagrou, Katrien; Van Eldere, Johan; Keuleers, S; Hagen, F; Merckx, Rita; Verhaegen, Jan; Peetermans, W E, and Boekhout, T. Zoonotic transmission of Cryptococcus neoformans from a magpie to an immunocompetent patient. *Journal of internal medicine*, 257(4):385–388, 2005.

Larkin, Mark A; Blackshields, Gordon; Brown, N P; Chenna, R; McGettigan, Paul A; McWilliam, Hamish; Valentin, Franck; Wallace, Iain M; Wilm, Andreas, and Lopez, Rodrigo. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007. ISSN 1367-4803.

Lauber, Christian L; Hamady, Micah; Knight, Rob, and Fierer, Noah. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and environmental microbiology*, 75(15):5111–5120, 2009. ISSN 0099-2240.

Lavely, James and Lipsitz, David. Fungal infections of the central nervous system in the dog and cat. *Clinical techniques in small animal practice*, 20(4):212–219, 2005.

Lawson, Daniel John; Hellenthal, Garrett; Myers, Simon, and Falush, Daniel. Inference of population structure using dense haplotype data. *PLoS genetics*, 8(1):e1002453, 2012. ISSN 1553-7404.

Lazera, M S; Wanke, B, and Nishikawa, M M. Isolation of both varieties of Cryptococcus neoformans from saprophytic sources in the city of Rio de Janeiro, Brazil. *Medical Mycology*, 31(6):449–454, 1993. ISSN 1369-3786.

Lazera, M S; Pires, F D A; Camillo-Coura, L; Nishikawa, M M; Bezerra, C C F; Trilles, L, and Wanke, B. Natural habitat of Cryptococcus neoformans var. neoformans in decaying wood forming hollows in living trees. *Medical Mycology*, 34(2):127–131, 1996.

Lazera, M S; Cavalcanti, M A Salmito; Londero, A T; Trilles, L; Nishikawa, M M, and Wanke, B. Possible primary ecological niche of Cryptococcus neoformans. *Medical Mycology*, 38(5): 379–383, 2000. ISSN 1460-2709.

Lee, Anthony; Toffaletti, Dena L; Tenor, Jennifer; Soderblom, Erik J; Thompson, J Will; Moseley, M Arthur; Price, Michael, and Perfect, John R. Survival defects of Cryptococcus neoformans mutants exposed to human cerebrospinal fluid result in attenuated virulence in an experimental model of meningitis. *Infection and immunity*, 78(10):4213–4225, 2010. ISSN 0019-9567.

Lee, I Russel; Lui, Edmund Y L; Chow, Eve W L; Arras, Samantha D M; Morrow, Carl A, and Fraser, James A. Reactive oxygen species homeostasis and virulence of the fungal pathogen

Cryptococcus neoformans requires an intact proline catabolism pathway. *Genetics*, 194(2): 421–433, 2013. ISSN 0016-6731.

Lee, Kyung-Tae; Byun, Hyo-Jeong; Jung, Kwang-Woo; Hong, Joohyeon; Cheong, Eunji, and Bahn, Yong-Sun. Distinct and redundant roles of protein tyrosine phosphatases Ptp1 and Ptp2 in governing the differentiation and pathogenicity of Cryptococcus neoformans. *Eukaryotic cell*, 13(6):796–812, 2014. ISSN 1535-9778.

Legendre, Pierre and Fortin, MARIEJosée. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, 10(5):831–844, 2010. ISSN 1755-0998.

Lengeler, Klaus B; Fox, Deborah S; Fraser, James A; Allen, Andria; Forrester, Keri; Dietrich, Fred S, and Heitman, Joseph. Mating-type locus of Cryptococcus neoformans: a step in the evolution of sex chromosomes. *Eukaryotic cell*, 1(5):704–718, 2002.

Lepais, Olivier and Bacles, Cecile F E. Two are better than one: combining landscape genomics and common gardens for detecting local adaptation in forest trees. *Molecular ecology*, 23(19): 4671–4673, 2014. ISSN 1365-294X.

Lewontin, R C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49(1):49, 1964.

Li, Dong; Zhang, Xiaojiao; Li, Zhongming; Yang, Jiao; Pan, Jiao, and Zhu, Xudong. Cryptococcus neoformans Ca2+ homeostasis requires a chloride channel/antiporter Clc1 in JEC21, but not in H99. *FEMS yeast research*, 12(1):69–77, 2012. ISSN 1567-1364.

Li, Heng and Durbin, Richard. Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009. ISSN 1367-4803.

Li, Heng; Handsaker, Bob; Wysoker, Alec; Fennell, Tim; Ruan, Jue; Homer, Nils; Marth, Gabor; Abecasis, Goncalo, and Durbin, Richard. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009. ISSN 1367-4803.

Li, Yong Fuga; Costello, James C; Holloway, Alisha K, and Hahn, Matthew W. Reverse ecology and the power of population genomics. *Evolution*, 62(12):2984–2994, 2008. ISSN 1558-5646.

Lim, Young Woon; Kim, Byung Kwon; Kim, Changmu; Jung, Hack Sung; Kim, Bong-Soo; Lee, Jae-Hak, and Chun, Jongsik. Assessment of soil fungal communities using pyrosequencing. *The Journal of Microbiology*, 48(3):284–289, 2010. ISSN 1225-8873.

Lin, X. Cryptococcus neoformans: Morphogenesis, infection, and evolution. *Infection, Genetics and Evolution*, 9(4):401–416, 2009.

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

Lin, Xiaorong; Huang, Johnny C; Mitchell, Thomas G, and Heitman, Joseph. Virulence attributes and hyphal growth of C. neoformans are quantitative traits and the MATα allele enhances filamentation. *PLoS genetics*, 2(11):e187, 2006.

Lindahl, Björn D; Nilsson, R Henrik; Tedersoo, Leho; Abarenkov, Kessy; Carlsen, Tor; Kjø ller, Rasmus; Kõljalg, Urmas; Pennanen, Taina; Rosendahl, Sø ren, and Stenlid, Jan. Fungal community analysis by highthroughput sequencing of amplified markersa user's guide. *New Phytologist*, 199(1):288–299, 2013. ISSN 1469-8137.

Lindsley, Mark D; Mekha, Nanthawan; Baggett, Henry C; Surinthong, Yupha; Autthateinchai, Rinrapas; Sawatwong, Pongpun; Harris, Julie R; Park, Benjamin J; Chiller, Tom, and Balajee, S Arunmozhi. Evaluation of a newly developed lateral flow immunoassay for the diagnosis of cryptococcosis. *Clinical infectious diseases*, 53(4):321–325, 2011.

Littman, M L and Borok, R. Relation of the pigeon to cryptococcosis: natural carrier state, heat resistance and survival of Cryptococcus neoformans. *Mycopathologia*, 35(3):329–345, 1968.

Litvintseva, ; Carbone, I; Rossouw, J; Thakur, R; Govender, N P, and Mitchell, T G. Evidence that the human pathogenic fungus Cryptococcus neoformans var. grubii may have evolved in Africa. *PLOS ONE*, 6(5):e19688, 2011.

Litvintseva, A P; Kestenbaum, L; Vilgalys, R, and Mitchell, T G. Comparative analysis of environmental and clinical populations of Cryptococcus neoformans. *Journal of Clinical Microbiology*, 43(2):556–564, 2005.

Litvintseva, Anastasia P and Mitchell, Thomas G. Population genetic analyses reveal the African origin and strain variation of Cryptococcus neoformans var. grubii. *PLoS Pathog*, 8 (2):e1002495, 2012.

Litvintseva, Anastasia P; Marra, Robert E; Nielsen, Kirsten; Heitman, Joseph; Vilgalys, Rytas, and Mitchell, Thomas G. Evidence of sexual recombination among Cryptococcus neoformans serotype A isolates in sub-Saharan Africa. *Eukaryotic cell*, 2(6):1162–1168, 2003. ISSN 1535-9778.

Litvintseva, Anastasia P; Lin, Xiaorong; Templeton, Irka; Heitman, Joseph, and Mitchell, Thomas G. Many globally isolated AD hybrid strains of Cryptococcus neoformans originated in Africa. *PLoS Pathog*, 3(8):e114, 2007.

Litvintseva, Anastasia P; Hurst, Steven; Gade, Lalitha; Frace, Michael A; Hilsabeck, Remy; Schupp, James M; Gillece, John D; Roe, Chandler; Smith, David, and Keim, Paul. Whole-genome analysis of Exserohilum rostratum from an outbreak of fungal meningitis and other infections. *Journal of clinical microbiology*, 52(9):3216–3222, 2014. ISSN 0095-1137.

Litvintseva R, Vilgalys, R, et al., AP; Thakur. Multilocus Sequence Typing Reveals Three Genetic Subpopulations of Cryptococcus neoformans var. grubii (Serotype A), Including a Unique Population in Botswana. *GENETICS*, 172(4):2223–2238, 2006.

Liu, Lide; Tewari, Ram P, and Williamson, Peter R. Laccase Protects Cryptococcus neoformans from Antifungal Activity of Alveolar Macrophages. *Infection and immunity*, 67(11):6034–6039, 1999.

Liu, Oliver W; Chun, Cheryl D; Chow, Eric D; Chen, Changbin; Madhani, Hiten D, and Noble, Suzanne M. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. *Cell*, 135(1):174–188, 2008. ISSN 0092-8674.

Liu, Tong-Bao; Wang, Yina; Baker, Gregory M; Fahmy, Hany; Jiang, Linghuo, and Xue, Chaoyang. The Glucose Sensor-Like Protein Hxs1 Is a High-Affinity Glucose Transporter and Required for Virulence in Cryptococcus neoformans. 2013. ISSN 1932-6203.

Loftus, Brendan J; Fung, Eula; Roncaglia, Paola; Rowley, Don; Amedeo, Paolo; Bruno, Dan; Vamathevan, Jessica; Miranda, Molly; Anderson, Iain J, and Fraser, James A. The genome

of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. *Science*, 307 (5713):1321–1324, 2005.

Lopez-Martinez, R and Castanon-Olivares, L R. Isolation ofCryptococcus neoformans var. neoformans from bird droppings, fruits and vegetables in Mexico City. *Mycopathologia*, 129 (1):25–28, 1995. ISSN 0301-486X.

Losos, Jonathan B and Glor, Richard E. Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution*, 18(5):220–227, 2003. ISSN 0169-5347.

Lotterhos, Katie E and Whitlock, Michael C. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology*, 24 (5):1031–1046, 2015. ISSN 1365-294X.

Loyse, Angela; Thangaraj, Harry; Easterbrook, Philippa; Ford, Nathan; Roy, Monika; Chiller, Tom; Govender, Nelesh; Harrison, Thomas S, and Bicanic, Tihana. Cryptococcal meningitis: improving access to essential antifungal medicines in resource-poor countries. *The Lancet Infectious Diseases*, 13(7):629–637, 2013. ISSN 1473-3099.

Lynch, Michael. Evolution of the mutation rate. *Trends in Genetics*, 26(8):345–352, 2010. ISSN 0168-9525.

Ma, Hansong; Croudace, Joanne E; Lammas, David A, and May, Robin C. Direct cell-to-cell spread of a pathogenic yeast. *BMC immunology*, 8(1):15, 2007.

MacDougall, Laura; Kidd, Sarah E; Galanis, Eleni; Mak, Sunny; Leslie, Mira J; Cieslak, Paul R; Kronstad, James W; Morshed, Muhammad G, and Bartlett, Karen H. Spread of Cryptococcus gattii in British Columbia, Canada, and detection in the Pacific Northwest, USA. *Emerging infectious diseases*, 13(1):42, 2007.

Maeng, Shinae; Ko, Young-Joon; Kim, Gyu-Bum; Jung, Kwang-Woo; Floyd, Anna; Heitman, Joseph, and Bahn, Yong-Sun. Comparative transcriptome analysis reveals novel roles of the Ras and cyclic AMP signaling pathways in environmental stress response and antifungal drug sensitivity in Cryptococcus neoformans. *Eukaryotic cell*, 9(3):360–378, 2010. ISSN 1535-9778.

Magditch, Denise A; Liu, Tong-Bao; Xue, Chaoyang, and Idnurm, Alexander. DNA mutations mediate microevolution between host-adapted forms of the pathogenic fungus Cryptococcus neoformans. *PLoS pathogens*, 8(10):e1002936, 2012.

Magurran, Anne E. *Measuring biological diversity*. John Wiley & Sons, 2013. ISBN 1118687922.

Mak, Sunny; Vélez, Nórida; Castañeda, Elizabeth, and Escandón, Patricia. The Fungus among Us: Cryptococcus neoformans and Cryptococcus gattii Ecological Modeling for Colombia. *Journal of Fungi*, 1(3):332–344, 2015.

Mak, Sunny Y. Ecological niche modeling of Cryptococcus gattii in British Columbia, 2007.

Malliaris, Stephanie D; Steenbergen, Judith N, and Casadevall, Arturo. Cryptococcus neoformans var. gattii can exploit Acanthamoeba castellanii for growth. *Medical Mycology*, 42(2): 149–158, 2004.

Mandel, M Alejandra; Grace, Greg G; Orsborn, Kris I; Schafer, Fredda; Murphy, Juneann W; Orbach, Marc J, and Galgiani, John N. The Cryptococcus neoformans GeneDHA1 Encodes an Antigen That Elicits a Delayed-Type Hypersensitivity Reaction in Immune Mice. *Infection and immunity*, 68(11):6196–6201, 2000. ISSN 0019-9567.

Manel, S; Perrier, C; Pratlong, M; AbiRached, L; Paganini, J; Pontarotti, P, and Aurelle, D. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular ecology*, 2015. ISSN 1365-294X.

Manel, Stéphanie; Schwartz, Michael K; Luikart, Gordon, and Taberlet, Pierre. Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution*, 18(4):189–197, 2003. ISSN 0169-5347.

Mangrola, Amitsinh; Dudhagara, Pravin; Koringa, Prakash; Joshi, C G; Parmar, Mansi, and Patel, Rajesh. Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genomics Data*, 4:153–155, 2015. ISSN 2213-5960.

Mantel, Nathan. The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1):209–220, 1967. ISSN 0008-5472.

Martiny, Jennifer B Hughes; Bohannan, Brendan J M; Brown, James H; Colwell, Robert K; Fuhrman, Jed A; Green, Jessica L; Horner-Devine, M Claire; Kane, Matthew; Krumins, Jennifer Adams, and Kuske, Cheryl R. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2):102–112, 2006. ISSN 1740-1526.

Marttinen, Pekka; Hanage, William P; Croucher, Nicholas J; Connor, Thomas R; Harris, Simon R; Bentley, Stephen D, and Corander, Jukka. Detection of recombination events in bacterial genomes from large population samples. *Nucleic acids research*, 40(1):e6–e6, 2012. ISSN 0305-1048.

Masella, Andre P; Bartram, Andrea K; Truszkowski, Jakub M; Brown, Daniel G, and Neufeld, Josh D. PANDAseq: paired-end assembler for illumina sequences. *BMC bioinformatics*, 13 (1):31, 2012. ISSN 1471-2105.

Mashabane, L G; Wessels, D C J, and Potgieter, M J. The utilisation of Colophospermum mopane by the Vatsonga in the Gazankulu region (eastern Northern Province, South Africa). *South African journal of botany*, 67(2):199–205, 2001.

Mauch, Renan Marrichi; Cunha, Vinícius de Oliveira, and Dias, Amanda Latercia Tranches. The copper interference with the melanogenesis of Cryptococcus neoformans. *Revista do Instituto de Medicina Tropical de São Paulo*, 55(2):117–120, 2013. ISSN 0036-4665.

May, Robin C; Stone, Neil R H; Wiesner, Darin L; Bicanic, Tihana, and Nielsen, Kirsten. Cryptococcus: from environmental saprophyte to global pathogen. *Nature Reviews Microbiology*, 2015. ISSN 1740-1526.

McDonald, John H and Kreitman, Martin. Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, 351(6328):652–654, 1991. ISSN 0028-0836.

McKenna, Aaron; Hanna, Matthew; Banks, Eric; Sivachenko, Andrey; Cibulskis, Kristian; Kernytsky, Andrew; Garimella, Kiran; Altshuler, David; Gabriel, Stacey, and Daly, Mark. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9):1297–1303, 2010. ISSN 1088-9051.

McQuiston, Travis J and Williamson, Peter R. Paradoxical roles of alveolar macrophages in the host response to Cryptococcus neoformans. *Journal of Infection and Chemotherapy*, 18 (1):1–9, 2012. ISSN 1341-321X.

McVean, Gilean A T; Myers, Simon R; Hunt, Sarah; Deloukas, Panos; Bentley, David R, and Donnelly, Peter. The fine-scale structure of recombination rate variation in the human genome. *Science*, 304(5670):581–584, 2004. ISSN 0036-8075.

Meiser, Anjuli; Bálint, Miklós, and Schmitt, Imke. Metaanalysis of deepsequenced fungal communities indicates limited taxon sharing between studies and the presence of biogeographic patterns. *New Phytologist*, 201(2):623–635, 2014. ISSN 1469-8137.

Messenger, Louisa A; Garcia, Lineth; Vanhove, Mathieu; Huaranca, Carlos; Bustamante, Marinely; Torrico, Marycruz; Torrico, Faustino; Miles, Michael A, and Llewellyn, Martin S. Ecological host fitting of Trypanosoma cruzi TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. *Molecular ecology*, 24(10): 2406–2422, 2015. ISSN 1365-294X.

Metin, Banu; Findley, Keisha, and Heitman, Joseph. The mating type locus (MAT) and sexual reproduction of Cryptococcus heveanensis: insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS genetics*, 6(5):e1000961, 2010.

Meya, David B; Manabe, Yukari C; Castelnuovo, Barbara; Cook, Bethany A; Elbireer, Ali M; Kambugu, Andrew; Kamya, Moses R; Bohjanen, Paul R, and Boulware, David R. Cost-effectiveness of serum cryptococcal antigen screening to prevent deaths among HIV-infected persons with a CD4+ cell count 100 cells/μL who start HIV therapy in resource-limited settings. *Clinical infectious diseases*, 51(4):448–455, 2010.

Meyer, W; CastaÃeda, A; Jackson, S; Huynh, M, and CastaÃeda, E. Molecular typing of IberoAmerican Cryptococcus neoformans isolates. *Emerging infectious diseases*, 9(2):189, 2003.

Meyer, Wieland; Marszewska, Krystyna; Amirmostofian, Mitra; Igreja, Ricardo Pereira; Hardtke, Claudia; Methling, Katharina; Viviani, Maria Anna; Chindamporn, Ariya; Sukroongreung, Samaniya, and John, Melanie Ann. Molecular typing of global isolates of Cryptococcus neoformans var. neoformans by polymerase chain reaction fingerprinting and randomly amplified polymorphic DNAa pilot study to standardize techniques on which to base a detailed epidemiological survey. *Electrophoresis*, 20(8):1790–1799, 1999.

Meyer, Wieland; Aanensen, David M; Boekhout, Teun; Cogliati, Massimo; Diaz, Mara R; Esposto, Maria Carmela; Fisher, Matthew; Gilgado, Felix; Hagen, Ferry, and Kaocharoen, Sirada. Consensus multi-locus sequence typing scheme for Cryptococcus neoformans and Cryptococcus gattii. *Medical Mycology*, 47(6):561–570, 2009.

Micol, Romain; Tajahmady, Ayden; Lortholary, Olivier; Balkan, Suna; Quillet, Catherine; Dousset, Jean-Philippe; Chanroeun, Hak; Madec, Yoann; Fontanet, Arnaud, and Yazdanpanah, Yazdan. Cost-effectiveness of primary prophylaxis of AIDS associated cryptococcosis in Cambodia. *PLOS ONE*, 5(11):e13856, 2010.

Mirza, Sara A; Phelan, Maureen; Rimland, David; Graviss, Edward; Hamill, Richard; Brandt, Mary E; Gardner, Tracie; Sattah, Matthew; de Leon, Gabriel Ponce, and Baughman, Wendy. The changing epidemiology of cryptococcosis: an update from population-based active surveillance in 2 large metropolitan areas, 19922000. *Clinical infectious diseases*, 36 (6):789–794, 2003.

Missall, Tricia A and Lodge, Jennifer K. Function of the thioredoxin proteins in Cryptococcus neoformans during stress or virulence and regulation by putative transcriptional modulators. *Molecular microbiology*, 57(3):847–858, 2005. ISSN 1365-2958.

Molez, Jean-Francois. The historical question of acquired immunodeficiency syndrome in the 1960s in the Congo River basin area in relation to cryptococcal meningitis. *American Journal of Tropical Medicine and Hygiene*, 58(3):273–276, 1998.

Moncalvo, J M. Molecular systematics: major fungal phylogenetic groups and fungal species concepts. *Evolutionary Genetics of Fungi. Horizon Scientific Press, Norfolk*, 2005.

Monot, Marc; Honoré, Nadine; Garnier, Thierry; Araoz, Romulo; Coppée, Jean-Yves; Lacroix, Céline; Sow, Samba; Spencer, John S; Truman, Richard W, and Williams, Diana L. On the origin of leprosy. *Science*, 308(5724):1040–1042, 2005.

Mora, Camilo; Tittensor, Derek P; Adl, Sina; Simpson, Alastair G B, and Worm, Boris. How many species are there on Earth and in the ocean? *PLoS biology*, 9(8):e1001127, 2011.

Moran, Gary P; Coleman, David C, and Sullivan, Derek J. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. *Eukaryotic cell*, 10(1):34–42, 2011. ISSN 1535-9778.

Morrow, Carl A and Fraser, James A. Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS yeast research*, 9(2):161–177, 2009.

Morrow, Carl A and Fraser, James A. Ploidy variation as an adaptive mechanism in human pathogenic fungi. In *Seminars in cell & developmental biology*. Elsevier, 2013. ISBN 1084-9521.

Morse, Stephen S. Factors in the emergence of infectious diseases. *Emerging infectious diseases*, 1(1):7, 1995.

Mouttham, Nathalie; Klunk, Jennifer; Kuch, Melanie; Fourney, Ron, and Poinar, Hendrik. Surveying the repair of ancient DNA from bones via high-throughput sequencing. *BioTechniques*, 59(1):19, 2015. ISSN 1940-9818.

Moyrand, Frédérique; Fontaine, Thierry, and Janbon, Guilhem. Systematic capsule gene disruption reveals the central role of galactose metabolism on Cryptococcus neoformans virulence. *Molecular microbiology*, 64(3):771–781, 2007. ISSN 1365-2958.

Mseddi, F; Sellami, A; Jarboui, M A; Sellami, H; Makni, F, and Ayadi, A. First environmental isolations of Cryptococcus neoformans and Cryptococcus gattii in Tunisia and review of published studies on environmental isolations in Africa. *Mycopathologia*, 171(5):355–360, 2011.

Mwaba, P; Mwansa, J; Chintu, C; Pobee, J; Scarborough, M; Portsmouth, S, and Zumla, A. Clinical presentation, natural history, and cumulative death rates of 230 adults with primary cryptococcal meningitis in Zambian AIDS patients treated under local conditions. *Postgraduate medical journal*, 77(914):769–773, 2001.

Myers, Simon; Bottolo, Leonardo; Freeman, Colin; McVean, Gil, and Donnelly, Peter. A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 310(5746): 321–324, 2005. ISSN 0036-8075.

Narasipura, Srinivas D; Ault, Jeffrey G; Behr, Melissa J; Chaturvedi, Vishnu, and Chaturvedi, Sudha. Characterization of Cu, Zn superoxide dismutase (SOD1) gene knockout mutant of Cryptococcus neoformans var. gattii: role in biology and virulence. *Molecular microbiology*, 47(6):1681–1694, 2003. ISSN 1365-2958.

Narum, Shawn R and Hess, Jon E. Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11(s1):184–194, 2011. ISSN 1755-0998.

Nei, Masatoshi and Li, Wen-Hsiung. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10): 5269–5273, 1979. ISSN 0027-8424.

Neill, James M; Abrahams, Irving, and Kapros, Charles E. A comparison of the immunogenicity of weakly encapsulated and of strongly encapsulated strains of Cryptococcus neoformans (Torula histolytica). *Journal of Bacteriology*, 59(2):263–275, 1950.

Nero, L C; Tarver, Mae G, and Hedrick, L R. Growth of Acanthamoeba castellani with the yeast Torulopsis famata. *Journal of Bacteriology*, 87(1):220–225, 1964.

Nielsen, K; De Obaldia, A L, and Heitman, J. Cryptococcus neoformans mates on pigeon guano: implications for the realized ecological niche and globalization. *Eukaryotic cell*, 6(6): 949–959, 2007.

Nielsen, Kirsten; Cox, Gary M; Wang, Ping; Toffaletti, Dena L; Perfect, John R, and Heitman, Joseph. Sexual cycle of Cryptococcus neoformans var. grubii and virulence of congenic a and α isolates. *Infection and Immunity*, 71(9):4831–4841, 2003. ISSN 0019-9567.

Nielsen, Rasmus. Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39:197–218, 2005. ISSN 0066-4197.

Nosanchuk, Joshua D; Shoham, Shmuel; Fries, Bettina C; Shapiro, Daniel S; Levitz, Stuart M, and Casadevall, Arturo. Evidence of zoonotic transmission of Cryptococcus neoformans from a pet cockatoo to an immunocompromised patient. *Annals of internal medicine*, 132 (3):205–208, 2000. ISSN 0003-4819.

Novembre, John and Di Rienzo, Anna. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics*, 10(11):745–755, 2009. ISSN 1471-0056.

Nygaard, Sanne; Braunstein, Alexander; Malsen, Gareth; Van Dongen, Stijn; Gardner, Paul P; Krogh, Anders; Otto, Thomas D; Pain, Arnab; Berriman, Matthew, and McAuliffe, Jon. Long-and short-term selective forces on malaria parasite genomes. *PLoS Genet*, 6(9): e1001099, 2010.

O'Brien, Heath E; Parrent, Jeri Lynn; Jackson, Jason A; Moncalvo, Jean-Marc, and Vilgalys, Rytas. Fungal community analysis by large-scale sequencing of environmental samples. *Applied and environmental microbiology*, 71(9):5544–5550, 2005. ISSN 0099-2240.

O'Hanlon, S J; Slater, H C; Cheke, R A; Boatin, B A; Coffeng, L E; Pion, S D; Boussinesq, M; Zouré, H G; Stolk, W A, and Basáñez, M G. Model-Based Geostatistical Mapping of the Prevalence of Onchocerca volvulus in West Africa. *PLoS neglected tropical diseases*, 10(1): e0004328–e0004328, 2016. ISSN 1935-2727.

Okagaki, Laura H and Nielsen, Kirsten. Titan cells confer protection from phagocytosis in Cryptococcus neoformans infections. *Eukaryotic cell*, 11(6):820–826, 2012.

Olsen, L; Choffnes, Eileen R; Relman, David A; Pray, Leslie, and Olsen, L A. Fungal diseases: an emerging threat to human, animal and plant health. Workshop summary. In *Fungal diseases: an emerging threat to human, animal and plant health. Workshop summary*. National Academies Press, 2011. ISBN 030921226X.

O'Meara, Teresa R; Hay, Christie; Price, Michael S; Giles, Steve, and Alspaugh, J Andrew. Cryptococcus neoformans histone acetyltransferase Gcn5 regulates fungal adaptation to the host. *Eukaryotic cell*, 9(8):1193–1202, 2010a. ISSN 1535-9778.

O'Meara, Teresa R; Norton, Diana; Price, Michael S; Hay, Christie; Clements, Meredith F; Nichols, Connie B, and Alspaugh, J Andrew. Interaction of Cryptococcus neoformans Rim101 and protein kinase A regulates capsule. 2010b. ISSN 1553-7374.

Ordóñez, Nelly and Castañeda, Elizabeth. Varieties and serotypes of Cryptococcus neoformans clinical isolates in Colombia. *Revista iberoamericana de micología*, 18(3):128–130, 2001.

Ormerod, Kate L and Fraser, James A. Balancing stability and flexibility within the genome of the pathogen Cryptococcus neoformans. 2013. ISSN 1553-7374.

Ory, Jeramia J; Griffith, Cara L, and Doering, Tamara L. An efficiently regulated promoter system for Cryptococcus neoformans utilizing the CTR4 promoter. *Yeast*, 21(11):919–926, 2004. ISSN 1097-0061.

Osono, Takashi. Diversity and functioning of fungi associated with leaf litter decomposition in Asian forests of different climatic regions. *Fungal Ecology*, 4(6):375–385, 2011. ISSN 1754-5048.

Paape, Timothy; Bataillon, Thomas; Zhou, Peng; JY Kono, Tom; Briskine, Roman; Young, Nevin D, and Tiffin, Peter. Selection, genomewide fitness effects and evolutionary rates in

the model legume Medicago truncatula. *Molecular ecology*, 22(13):3525–3538, 2013. ISSN 1365-294X.

Panepinto, John C and Williamson, Peter R. Intersection of fungal fitness and virulence in Cryptococcus neoformans. *FEMS yeast research*, 6(4):489–498, 2006.

Pappas, Peter G; Perfect, John R; Cloud, Gretchen A; Larsen, Robert A; Pankey, George A; Lancaster, Daniel J; Henderson, Harold; Kauffman, Carol A; Haas, David W, and Saccente, Michael. Cryptococcosis in human immunodeficiency virus-negative patients in the era of effective azole therapy. *Clinical infectious diseases*, 33(5):690–699, 2001.

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

Park, Yoon-Dong; Panepinto, John; Shin, Soowan; Larsen, Peter; Giles, Steven, and Williamson, Peter R. Mating pheromone in Cryptococcus neoformans is regulated by a transcriptional/degradative futile cycle. *Journal of Biological Chemistry*, 285(45):34746–34756, 2010.

Parks, Donovan H; Porter, Michael; Churcher, Sylvia; Wang, Suwen; Blouin, Christian; Whalley, Jacqueline; Brooks, Stephen, and Beiko, Robert G. GenGIS: A geospatial information system for genomic data. *Genome Research*, 19(10):1896–1904, 2009. ISSN 1088-9051.

Peiffer, Jason A; Spor, Aymé; Koren, Omry; Jin, Zhao; Tringe, Susannah Green; Dangl, Jeffery L; Buckler, Edward S, and Ley, Ruth E. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences*, 110(16):6548–6553, 2013. ISSN 0027-8424.

PérezFigueroa, A; GarcíaPereira, M J; Saura, M; RolánAlvarez, E, and Caballero, A. Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, 23(10):2267–2276, 2010. ISSN 1420-9101.

Perfect, John R. Cryptococcus neoformans: the yeast that likes it hot. *FEMS yeast research*, 6(4):463–468, 2006. ISSN 1567-1364.

Perfect, John R and Bicanic, Tihana. Cryptococcosis diagnosis and treatment: what do we know now. *Fungal Genetics and Biology*, 2014. ISSN 1087-1845.

Perfect, John R; Dismukes, William E; Dromer, Francoise; Goldman, David L; Graybill, John R; Hamill, Richard J; Harrison, Thomas S; Larsen, Robert A; Lortholary, Olivier, and Nguyen, Minh-Hong. Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. *Clinical infectious diseases*, 50(3): 291–322, 2010.

Pertea, Geo; Huang, Xiaoqiu; Liang, Feng; Antonescu, Valentin; Sultana, Razvan; Karamycheva, Svetlana; Lee, Yuandan; White, Joseph; Cheung, Foo, and Parvizi, Babak. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19(5):651–652, 2003. ISSN 1367-4803.

Pfeiffer, T J and Ellis, D H. Environmental isolation of Cryptococcus neoformans var. gattii from Eucalyptus tereticornis. *Medical Mycology*, 30(5):407–408, 1992. ISSN 1369-3786.

Pfister, Donald H; Slater, Caroline, and Hansen, Karen. Chorioactidaceae: a new family in the Pezizales (Ascomycota) with four genera. *Mycological research*, 112(5):513–527, 2008. ISSN 0953-7562.

Phillips, Steven J and Dudík, Miroslav. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2):161–175, 2008. ISSN 1600-0587.

Phillips, Steven J; Anderson, Robert P, and Schapire, Robert E. Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3):231–259, 2006. ISSN 0304-3800.

Porter, Teresita M and Brian Golding, G. Are similarityor phylogenybased methods more appropriate for classifying internal transcribed spacer (ITS) metagenomic amplicons? *New Phytologist*, 192(3):775–782, 2011. ISSN 1469-8137.

Pritchard, Jonathan K and Przeworski, Molly. Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics*, 69(1):1–14, 2001. ISSN 0002-9297.

Pukkila-Worley, Read and Mylonakis, Eleftherios. Epidemiology and management of cryptococcal meningitis: developments and challenges. 2008.

Purcell, Shaun; Neale, Benjamin; Todd-Brown, Kathe; Thomas, Lori; Ferreira, Manuel A R; Bender, David; Maller, Julian; Sklar, Pamela; De Bakker, Paul I W, and Daly, Mark J. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007. ISSN 0002-9297.

Qian, Song S. *Environmental and ecological statistics with R.* CRC Press, 2009. ISBN 1420062085.

Raja, Meera R; Waterman, Scott R; Qiu, Jin; Bleher, Reiner; Williamson, Peter R, and O'Halloran, Thomas V. A copper hyperaccumulation phenotype correlates with pathogenesis in Cryptococcus neoformans. *Metallomics*, 5(4):363–371, 2013.

Randhawa, H S; Kowshik, T, and Khan, Z U. Efficacy of swabbing versus a conventional technique for isolation of Cryptococcus neoformans from decayed wood in tree trunk hollows. *Medical Mycology*, 43(1):67–71, 2005.

Randhawa, H S; Kowshik, T; Preeti Sinha, K; Chowdhary, A; Khan, Z U; Yan, Z; Xu, J, and Kumar, A. Distribution of Cryptococcus gattii and Cryptococcus neoformans in decayed trunk wood of Syzygium cumini trees in north-western India. *Medical Mycology*, 44(7):623–630, 2006.

Randhawa, H S; Kowshik, T; Chowdhary, A; Preeti Sinha, K; Khan, Z U; Sun, S, and Xu, J. The expanding host tree species spectrum of Cryptococcus gattii and Cryptococcus neoformans and their isolations from surrounding soil in India. *Medical Mycology*, 46(8):823–833, 2008.

Randhawa, H S; Kowshik, T; Chowdhary, Anuradha; Prakash, Anupam; Khan, Z U, and Xu, Jianping. Seasonal variations in the prevalence of Cryptococcus neoformans var. grubii and Cryptococcus gattii in decayed wood inside trunk hollows of diverse tree species in north-western India: a retrospective study. *Medical mycology*, 49(3):320–323, 2011. ISSN 1369-3786.

Reimão, Juliana Quero; Drummond, Elislene Dias; Terceti, Mateus de Souza; Lyon, Juliana Pereira; Franco, Marília Caixeta, and De Siqueira, Antônio Martins. Isolation of Cryptococcus neoformans from hollows of living trees in the city of Alfenas, MG, Brazil. *Mycoses*, 50(4):261–264, 2007. ISSN 1439-0507.

Rhodes, Johanna; Beale, Mathew A, and Fisher, Matthew C. Illuminating Choices for Library Prep: A Comparison of Library Preparation Methods for Whole Genome Sequencing of Cryptococcus neoformans Using Illumina HiSeq. 2014. ISSN 1932-6203.

Rhome, Ryan; McQuiston, Travis; Kechichian, Talar; Bielawska, Alicja; Hennig, Mirko; Drago, Monica; Morace, Giulia; Luberto, Chiara, and Del Poeta, Maurizio. Biosynthesis and im-

munogenicity of glucosylceramide in Cryptococcus neoformans and other human pathogens. *Eukaryotic cell*, 6(10):1715–1726, 2007. ISSN 1535-9778.

Robertson, Emma J; Najjuka, Grace; Rolfes, Melissa A; Akampurira, Andrew; Jain, Neena; Anantharanjit, Janani; von Hohenberg, Maximilian; Tassieri, Manlio; Carlsson, Allan, and Meya, David B. Cryptococcus neoformans ex vivo capsule size is associated with intracranial pressure and host immune response in HIV-associated cryptococcal meningitis. *Journal of Infectious Diseases*, 209(1):74–82, 2014. ISSN 0022-1899.

Rohatgi, Soma; Gohil, Shruti; Kuniholm, Mark H; Schultz, Hannah; Dufaud, Chad; Armour, Kathryn L; Badri, Sheila; Mailliard, Robbie B, and Pirofski, Liise-anne. Fc gamma receptor 3A polymorphism and risk for HIV-associated cryptococcal disease. *MBio*, 4(5):e00573–13, 2013. ISSN 2150-7511.

Rosas, Angel L and Casadevall, Arturo. Melanization decreases the susceptibility of Cryptococcus neoformans to enzymatic degradation. *Mycopathologia*, 151(2):53–56, 2001.

Rosas, Ángel Luis and Casadevall, Arturo. Melanization affects susceptibility of Cryptococcus neoformans to heat and cold1. *FEMS microbiology letters*, 153(2):265–272, 1997.

Roy, Monika and Chiller, Tom. Preventing deaths from cryptococcal meningitis: from bench to bedside. *Expert review of anti-infective therapy*, 9(9):715–717, 2011.

Ruiz, Rene A and Casas, J Manuel. Culturally relevant and behavioristic counseling for Chicano college students. *Counseling across cultures*, pages 181–202, 1981.

Sabeti, Pardis C; Varilly, Patrick; Fry, Ben; Lohmueller, Jason; Hostetter, Elizabeth; Cotsapas, Chris; Xie, Xiaohui; Byrne, Elizabeth H; McCarroll, Steven A, and Gaudet, Rachelle. Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164):913–918, 2007. ISSN 0028-0836.

Sabiiti, Wilber and May, Robin C. Capsule independent uptake of the fungal pathogen Cryptococcus neoformans into brain microvascular endothelial cells. *PLOS ONE*, 7(4):e35455, 2012.

Sabiiti, Wilber; Robertson, Emma; Beale, Mathew A; Johnston, Simon A; Brouwer, Annemarie E; Loyse, Angela; Jarvis, Joseph N; Gilbert, Andrew S; Fisher, Matthew C, and

Harrison, Thomas S. Efficient phagocytosis and laccase activity affect the outcome of HIV-associated cryptococcosis. *The Journal of clinical investigation*, 124(5):2000, 2014.

Saeed, Alexander I; Bhagabati, Nirmal K; Braisted, John C; Liang, Wei; Sharov, Vasily; Howe, Eleanor A; Li, Jianwei; Thiagarajan, Mathangi; White, Joseph A, and Quackenbush, John. [9] TM4 microarray software suite. *Methods in enzymology*, 411:134–193, 2006. ISSN 0076-6879.

Saha, D C; Xess, Immaculata, and Jain, Neena. Evaluation of conventional & serological methods for rapid diagnosis of cryptococcosis. *Indian Journal of Medical Research*, 127(5): 483, 2008.

Sahl, Jason W; Caporaso, J Gregory; Rasko, David A, and Keim, Paul. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. *PeerJ*, 2:e332, 2014. ISSN 2167-8359.

Salter, S; Cox, M J; Turek, E M; Calus, S T; Cookson, W O; Moffatt, M F; Turner, P; Parkhill, J; Loman, N, and Walker, A W. Reagent contamination can critically impact sequence-based microbiome analyses. bioRxiv. 2014.

Schmidt, Philipp-André; Bálint, Miklós; Greshake, Bastian; Bandow, Cornelia; Römbke, Jörg, and Schmitt, Imke. Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry*, 65:128–132, 2013. ISSN 0038-0717.

Schoch, Conrad L; Seifert, Keith A; Huhndorf, Sabine; Robert, Vincent; Spouge, John L; Levesque, C André; Chen, Wen; Bolchacova, Elena; Voigt, Kerstin, and Crous, Pedro W. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109(16):6241–6246, 2012. ISSN 0027-8424.

Schwartz, Michael K; McKelvey, Kevin S; Cushman, Samuel A, and Luikart, Gordon. Landscape genomics: a brief perspective. In *Spatial complexity, informatics, and wildlife conservation*, pages 165–174. Springer, 2010. ISBN 4431877703.

Selbmann, Laura; Zucconi, Laura; Onofri, Silvano; Cecchini, Clarissa; Isola, Daniela; Turchetti, Benedetta, and Buzzini, Pietro. Taxonomic and phenotypic characterization of yeasts isolated

from worldwide cold rock-associated habitats. *Fungal biology*, 118(1):61–71, 2014. ISSN 1878-6146.

Selik, Richard M; Karon, John M, and Ward, John W. Effect of the human immunodeficiency virus epidemic on mortality from opportunistic infections in the United States in 1993. *Journal of Infectious Diseases*, 176(3):632–636, 1997.

Severo, Cecília Bittencourt; Xavier, Melissa Orzechowski; Gazzoni, Alexandra Flávia, and Severo, Luiz Carlos. Cryptococcosis in children. *Paediatric Respiratory Reviews*, 10(4): 166–171, 2009.

Shi, Meiqing; Li, Shu Shun; Zheng, Chunfu; Jones, Gareth J; Kim, Kwang Sik; Zhou, Hong; Kubes, Paul, and Mody, Christopher H. Real-time imaging of trapping and urease-dependent transmigration of Cryptococcus neoformans in mouse brain. *The Journal of clinical investigation*, 120(5):1683, 2010.

Shokralla, Shadi; Spall, Jennifer L; Gibson, Joel F, and Hajibabaei, Mehrdad. Nextgeneration sequencing technologies for environmental DNA research. *Molecular ecology*, 21(8):1794–1805, 2012. ISSN 1365-294X.

Simwami, S. *The evolution of Cryptococcus neoformans var. grubii in the context of clinical disease using multilocus sequence typing.* PhD thesis, London, 2011.

Simwami, S P; Khayhan, K; Henk, D A; Aanensen, D M; Boekhout, T; Hagen, F; Brouwer, A E; Harrison, T S; Donnelly, C A, and Fisher, M C. Low diversity Cryptococcus neoformans variety grubii multilocus sequence types from Thailand are consistent with an ancestral African origin. *PLoS pathogens*, 7(4):e1001343, 2011.

Singh, Arpita; Panting, Robert J; Varma, Ashok; Saijo, Tomomi; Waldron, Kevin J; Jong, Ambrose; Ngamskulrungroj, Popchai; Chang, Yun C; Rutherford, Julian C, and Kwon-Chung, Kyung J. Factors required for activation of urease as a virulence determinant in Cryptococcus neoformans. *MBio*, 4(3):e00220–13, 2013. ISSN 2150-7511.

Smith, J Gustav and Newton-Cheh, Christopher. Genome-wide association study in humans. *Cardiovascular Genomics: Methods and Protocols*, pages 231–258, 2009. ISSN 1607612461.

Song, Min-Hee; Lee, Jang-Won; Kim, Min Su; Yoon, Ja-Kyung; White, Theodore C; Floyd, Anna; Heitman, Joseph; Strain, Anna K; Nielsen, Judith N, and Nielsen, Kirsten. A

flucytosine-responsive Mbp1/Swi4-like protein, Mbs1, plays pleiotropic roles in antifungal drug resistance, stress response, and virulence of Cryptococcus neoformans. *Eukaryotic cell*, 11(1):53–67, 2012. ISSN 1535-9778.

Sorrell, Tania C and Chen, Sharon C A. Recent advances in management of cryptococcal meningitis: commentary. *F1000 medicine reports*, 2, 2010.

Spearman, Charles. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904. ISSN 0002-9556.

Springer, Deborah J and Chaturvedi, Vishnu. Projecting global occurrence of Cryptococcus gattii. *Emerging infectious diseases*, 16(1):14, 2010.

Staib, F. Cryptococcosis in the acquired immunodeficiency syndrome; mycological-diagnostic and epidemiological observations. . *AIDS-Forschung (AIFO)*, 7:363–382, 1987.

Stajich, Jason E; Harris, Todd; Brunk, Brian P; Brestelli, John; Fischer, Steve; Harb, Omar S; Kissinger, Jessica C; Li, Wei; Nayak, Vishal, and Pinney, Deborah F. FungiDB: an integrated functional genomics database for fungi. *Nucleic acids research*, page gkr918, 2011. ISSN 0305-1048.

Stamatakis, Alexandros. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690, 2006. ISSN 1367-4803.

Stano, Paola; Williams, Virginia; Villani, Maristella; Cymbalyuk, Eugene S; Qureshi, Asfia; Huang, Yuxiang; Morace, Giulia; Luberto, Chiara; Tomlinson, Stephen, and Del Poeta, Maurizio. App1: an antiphagocytic protein that binds to complement receptors 3 and 2. *The Journal of Immunology*, 182(1):84–91, 2009.

Stapley, Jessica; Reger, Julia; Feulner, Philine G D; Smadja, Carole; Galindo, Juan; Ekblom, Robert; Bennison, Clair; Ball, Alexander D; Beckerman, Andrew P, and Slate, Jon. Adaptation genomics: the next generation. *Trends in ecology & evolution*, 25(12):705–712, 2010. ISSN 0169-5347.

Steele, Katherine T; Thakur, Rameshwari; Nthobatsang, Rudo; Steenhoff, Andrew P, and Bisson, Gregory P. In-hospital mortality of HIV-infected cryptococcal meningitis patients

with C. gattii and C. neoformans infection in Gaborone, Botswana. *Medical Mycology*, 48(8): 1112–1115, 2010.

Steen, Barbara R; Lian, Tian; Zuyderduyn, Scott; MacDonald, William Kim; Marra, Marco; Jones, Steven J M, and Kronstad, James W. Temperature-regulated transcription in the pathogenic fungus Cryptococcus neoformans. *Genome research*, 12(9):1386–1400, 2002.

Steenbergen, J N; Shuman, H A, and Casadevall, A. Cryptococcus neoformans interactions with amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in macrophages. *Proceedings of the National Academy of Sciences*, 98(26):15245–15250, 2001.

Steenbergen, Judith N and Casadevall, Arturo. The origin and maintenance of virulence for the human pathogenic fungus Cryptococcus neoformans. *Microbes and infection*, 5(7):667–675, 2003.

Stoletzki, Nina and Eyre-Walker, Adam. Estimation of the neutrality index. *Molecular biology and evolution*, 28(1):63–70, 2011.

Storz, Jay F and Wheat, Christopher W. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, 64(9):2489–2509, 2010. ISSN 1558-5646.

Stucki, Sylvie; Orozco-terWengel, Pablo; Bruford, Michael W; Colli, Licia; Masembe, Charles; Negrini, Riccardo; Taberlet, Pierre, and Joost, Stéphane. High performance computation of landscape genomic models integrating local indices of spatial association. *arXiv preprint arXiv:1405.7658*, 2014.

Stukenbrock, Eva H and McDonald, Bruce A. Geographical variation and positive diversifying selection in the hostspecific toxin SnToxA. *Molecular plant pathology*, 8(3):321–332, 2007. ISSN 1364-3703.

Sun, Sheng and Heitman, Joseph. From two to one: Unipolar sexual reproduction. *Fungal biology reviews*, 29(3):118–125, 2015. ISSN 1749-4613.

Tajima, Fumio. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595, 1989. ISSN 0016-6731.

Takashima, Masako; Sugita, Takashi; Van, Bui Hong; Nakamura, Megumi; Endoh, Rikiya, and Ohkuma, Moriya. Taxonomic richness of yeasts in Japan within subtropical and cool temperate areas. 2012. ISSN 1932-6203.

Tang, Kun; Thornton, Kevin R, and Stoneking, Mark. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol*, 5(7):e171, 2007.

Taylor, D Lee and Houston, Shawn. A bioinformatics pipeline for sequence-based analyses of fungal biodiversity. In *Fungal Genomics*, pages 141–155. Springer, 2011. ISBN 1617790397.

Taylor, John W; Hann-Soden, Christopher; Branco, Sara; Sylvain, Iman, and Ellison, Christopher E. Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences*, 112 (29):8901–8908, 2015. ISSN 0027-8424.

Tedersoo, Leho; Bahram, Mohammad; Põlme, Sergei; Kõljalg, Urmas; Yorou, Nourou S; Wijesundera, Ravi; Ruiz, Luis Villarreal; Vasco-Palacios, Aída M; Thu, Pham Quang, and Suija, Ave. Global diversity and geography of soil fungi. *Science*, 346(6213):1256688, 2014. ISSN 0036-8075.

Tefsen, Boris; Grijpstra, Jan; Ordonez, Soledad; Lammers, Menno; van Die, Irma, and De Cock, Hans. Deletion of the CAP10 gene of Cryptococcus neoformans results in a pleiotropic phenotype with changes in expression of virulence factors. *Research in microbiology*, 165(6): 399–410, 2014. ISSN 0923-2508.

Thomas, Chris D; Cameron, Alison; Green, Rhys E; Bakkenes, Michel; Beaumont, Linda J; Collingham, Yvonne C; Erasmus, Barend F N; De Siqueira, Marinez Ferreira; Grainger, Alan, and Hannah, Lee. Extinction risk from climate change. *Nature*, 427(6970):145–148, 2004. ISSN 0028-0836.

Thomas, Jonathan C; Godfrey, Paul A; Feldgarden, Michael, and Robinson, D Ashley. Candidate targets of balancing selection in the genome of Staphylococcus aureus. *Molecular biology and evolution*, 29(4):1175–1186, 2012. ISSN 0737-4038.

Tibayrenc, Michel and Ayala, Francisco J. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proceedings of the National Academy of Sciences*, 109(48):E3305–E3313, 2012. ISSN 0027-8424.

Tine, Mbaye; Kuhl, Heiner; Gagnaire, Pierre-Alexandre; Louro, Bruno; Desmarais, Erick; Martins, Rute S T; Hecht, Jochen; Knaust, Florian; Belkhir, Khalid, and Klages, Sven. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications*, 5, 2014.

Toju, Hirokazu; Tanabe, Akifumi S; Yamamoto, Satoshi, and Sato, Hirotoshi. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. 2012. ISSN 1932-6203.

Tolkacheva, Tatyana; McNamara, Peter; Piekarz, Eileen, and Courchesne, William. Cloning of a Cryptococcus neoformans gene, GPA1, encoding a G-protein alpha-subunit homolog. *Infection and immunity*, 62(7):2849–2856, 1994. ISSN 0019-9567.

Torres, Eduardo M; Sokolsky, Tanya; Tucker, Cheryl M; Chan, Leon Y; Boselli, Monica; Dunham, Maitreya J, and Amon, Angelika. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, 317(5840):916–924, 2007. ISSN 0036-8075.

Travadon, Renaud; Sache, Ivan; Dutech, Cyril; Stachowiak, Anna; Marquer, Bruno, and Bousset, Lydia. Absence of isolation by distance patterns at the regional scale in the fungal plant pathogen Leptosphaeria maculans. *Fungal biology*, 115(7):649–659, 2011. ISSN 1878-6146.

Tseng, Hsiang-Kuang; Liu, Chang-Pan; Price, Michael S; Jong, Ambrose Y; Chang, Jui-Chih; Toffaletti, Dena L; Betancourt-Quiroz, Marisol; Frazzitta, Aubrey E; Cho, Wen-Long, and Perfect, John R. Identification of genes from the fungal pathogen Cryptococcus neoformans related to transmigration into the central nervous system. *PloS one*, 7(9):e45083, 2012. ISSN 1932-6203.

Tuck, Sean L; Phillips, Helen R P; Hintzen, Rogier E; Scharlemann, Jörn P W; Purvis, Andy, and Hudson, Lawrence N. MODISToolsdownloading and processing MODIS remotely sensed data in R. *Ecology and evolution*, 4(24):4658, 2014.

Turner, Stephen D. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *bioRxiv*, page 5165, 2014.

van der Horst, Charles M; Saag, Michael S; Cloud, Gretchen A; Hamill, Richard J; Graybill, J Richard; Sobel, Jack D; Johnson, Philip C; Tuazon, Carmelita U; Kerkering, Thomas, and Moskovitz, Bruce L. Treatment of cryptococcal meningitis associated with the acquired immunodeficiency syndrome. *New England Journal of Medicine*, 337(1):15–21, 1997.

van Duin, David; Casadevall, Arturo, and Nosanchuk, Joshua D. Melanization of Cryptococcus neoformans and Histoplasma capsulatum reduces their susceptibilities to amphotericin B and caspofungin. *Antimicrobial agents and chemotherapy*, 46(11):3394–3400, 2002.

Van Wyk, Marelize; Govender, Nelesh P; Mitchell, Thomas G, and Litvintseva, Anastasia P. Multilocus sequence typing of serially collected isolates of Cryptococcus from HIV-infected patients in South Africa. *Journal of clinical microbiology*, pages JCM–03177, 2014. ISSN 0095-1137.

Vanselow, Michelle; Brandt, Mary E, and Park, Benjamin J. Diagnosis and management of cryptococcal disease in resource-limited settings. *Current Fungal Infection Reports*, 6(1): 35–40, 2012.

Vartivarian, Shahe E; Anaissie, Elias J; Cowart, Richard E; Sprigg, Helen A; Tingler, Mary J, and Jacobson, Eric S. Regulation of cryptococcal capsular polysaccharide by iron. *Journal of Infectious Diseases*, 167(1):186–190, 1993. ISSN 0022-1899.

Vatsiou, Alexandra I; Bazin, Eric, and Gaggiotti, Oscar E. Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular ecology*, 2015. ISSN 1365-294X.

Vecchiarelli, A. Immunoregulation by capsular components of Cryptococcus neoformans. *Medical Mycology*, 38(6):407–417, 2000.

Velagapudi, Rajesh; Hsueh, Yen-Ping; Geunes-Boyer, Scarlett; Wright, Jo Rae, and Heitman, Joseph. Spores as infectious propagules of Cryptococcus neoformans. *Infection and immunity*, 77(10):4345–4355, 2009. ISSN 0019-9567.

Vilella, Albert J; Blanco-Garcia, Angel; Hutter, Stephan, and Rozas, Julio. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*, 21(11):2791–2793, 2005. ISSN 1367-4803.

Vishniac, Helen S. A multivariate analysis of soil yeasts isolated from a latitudinal gradient. *Microbial ecology*, 52(1):90–103, 2006. ISSN 0095-3628.

Vitti, Joseph J; Grossman, Sharon R, and Sabeti, Pardis C. Detecting natural selection in genomic data. *Annual review of genetics*, 47:97–120, 2013. ISSN 0066-4197.

Viviani, M A; Esposto, M C; Cogliati, M; Montagna, M T, and Wickes, B L. Isolation of a Cryptococcus neoformans serotype A MATa strain from the Italian environment. *Medical mycology*, 39(5):383–386, 2001. ISSN 1460-2709.

Voelz, Kerstin; Ma, Hansong; Phadke, Sujal; Byrnes, Edmond J; Zhu, Pinkuan; Mueller, Olaf; Farrer, Rhys A; Henk, Daniel A; Lewit, Yonathan, and Hsueh, Yen-Ping. Transmission of hypervirulence traits via sexual reproduction within and between lineages of the human fungal pathogen Cryptococcus gattii. *PLoS Genet*, 9(9):e1003771, 2013. ISSN 1553-7404.

Vogel, Gretchen. For more protein, filet of cricket. *Science*, 327(5967):811, 2010.

Walker, John F; Miller Jr, Orson K, and Horton, Jonathan L. Seasonal dynamics of ectomycorrhizal fungus assemblages on oak seedlings in the southeastern Appalachian Mountains. *Mycorrhiza*, 18(3):123–132, 2008. ISSN 0940-6360.

Wang, Lin-Ing; Lin, Yu-Sheng; Liu, Kung-Hung; Jong, Ambrose Y, and Shen, Wei-Chiang. Cryptococcus neoformans mediator protein Ssn8 negatively regulates diverse physiological processes and is required for virulence. *PloS one*, 6(4):e19162–e19162, 2011a. ISSN 1932-6203.

Wang, Linqi; Zhai, Bing, and Lin, Xiaorong. The link between morphotype transition and virulence in Cryptococcus neoformans. *PLoS Pathog*, 8(6):e1002765, 2012.

Wang, Ping; Nichols, Connie B; Lengeler, Klaus B; Cardenas, Maria E; Cox, Gary M; Perfect, John R, and Heitman, Joseph. Mating-type-specific and nonspecific PAK kinases play shared and divergent roles in Cryptococcus neoformans. *Eukaryotic cell*, 1(2):257–272, 2002. ISSN 1535-9778.

Wang, Yina; Liu, Tong-Bao; Patel, Shyam; Jiang, Linghuo, and Xue, Chaoyang. The casein kinase I protein Cck1 regulates multiple signaling pathways and is essential for cell integrity and fungal virulence in Cryptococcus neoformans. *Eukaryotic cell*, 10(11):1455–1464, 2011b. ISSN 1535-9778.

Wang, Yulin and Casadevall, Arturo. Decreased susceptibility of melanized Cryptococcus neoformans to UV light. *Applied and environmental microbiology*, 60(10):3864–3866, 1994.

Watterson, G A. On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2):256–276, 1975. ISSN 0040-5809.

White, F. The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa (3 Plates, Northwestern Africa, North-

eastern Africa, and Southern Africa, 1:5,000,000). *UNESCO, Paris. I*, (SBN: 9231019554), 1983.

White, Th J; Bruns, T; Lee, SJWT, and Taylor, J W. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, 18:315–322, 1990.

Whittaker, Robert Harding. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological monographs*, 30(3):279–338, 1960. ISSN 0012-9615.

WHO, . Rapid advice: Diagnosis, prevention and management of cryptococcal disease in HIV-infected adults, adolescents and children. Technical report, 2011. URL `http://www.who.int/hiv/pub/cryptococcal_disease2011/en/`.

Wickes, Brian L; Mayorga, MARfA E; Edman, Ursula, and Edman, Jeffrey C. Dimorphism and haploid fruiting in Cryptococcus neoformans: association with the alpha-mating type. *Proceedings of the National Academy of Sciences*, 93(14):7327–7331, 1996.

Wiesner, Darin L; Moskalenko, Oleksandr; Corcoran, Jennifer M; McDonald, Tami; Rolfes, Melissa A; Meya, David B; Kajumbula, Henry; Kambugu, Andrew; Bohjanen, Paul R, and Knight, Joseph F. Cryptococcal genotype influences immunologic response and human clinical outcome after meningitis. *MBio*, 3(5):e00196–12, 2012. ISSN 2150-7511.

Williamson, Peter R. Laccase and melanin in the pathogenesis of Cryptococcus neoformans. *Front. Biosci*, 2:e99–e107, 1997.

Wilson, D E; Bennett, J E, and Bailey, Johnsie W. Serologic grouping of Cryptococcus neoformans. In *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, NY)*, volume 127, pages 820–823. Royal Society of Medicine, 1968. ISBN 1535-3702.

Wolfe, Nathan D; Dunavan, Claire Panosian, and Diamond, Jared. Origins of major human infectious diseases. *Nature*, 447(7142):279–283, 2007.

Wright, Sewall. Isolation by distance. *Genetics*, 28(2):114, 1943.

WWF, . Ecoregions of Zambia, 2011. URL `http://www.eoearth.org/article/Ecoregions_of_Zambia_(W`

Xu, J. Fundamentals of fungal molecular population genetic analyses. *Current issues in molecular biology*, 8(2):75, 2006a.

Xu, J. INVITED REVIEW: Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular ecology*, 15(7):1713–1731, 2006b.

Xu, J; Vilgalys, R, and Mitchell, T G. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus Cryptococcus neoformans. *Molecular ecology*, 9(10):1471–1481, 2001.

Xu, Jianping; Vilgalys, Rytas, and Mitchell, Thomas G. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus Cryptococcus neoformans. *Molecular Ecology*, 9(10):1471–1481, 2000. ISSN 1365-294X.

Xue, C; Tada, Y; Dong, X, and Heitman, J. The Human Fungal Pathogen Cryptococcus Can Complete Its Sexual Cycle during a Pathogenic Association with Plants. *Cell host & microbe*, 1(4):263–273, 2007.

Xue, Chaoyang; Liu, Tongbao; Chen, Lydia; Li, Wenjun; Liu, Iris; Kronstad, James W; Seyfang, Andreas, and Heitman, Joseph. Role of an expanded inositol transporter repertoire in Cryptococcus neoformans sexual reproduction and virulence. *MBio*, 1(1):e00084–10, 2010. ISSN 2150-7511.

Yang, Jiao; Li, Dong; Liu, Xiaoguang; Pan, Jiao; Yan, Bing, and Zhu, Xudong. Regulation of virulence factors, carbon utilization and virulence by SNF1 in Cryptococcusneoformans JEC21 and divergent actions of SNF1 between cryptococcal strains. *Fungal Genetics and Biology*, 47(12):994–1000, 2010. ISSN 1087-1845.

Yang, Ziheng and Bielawski, Joseph P. Statistical methods for detecting molecular adaptation. *Trends in ecology & evolution*, 15(12):496–503, 2000. ISSN 0169-5347.

Yurkov, Andrey M; Kemler, Martin, and Begerow, Dominik. Species accumulation curves and incidence-based species richness estimators to appraise the diversity of cultivable yeasts from beech forest soils. *PLoS One*, 6(8):e23671, 2011. ISSN 1932-6203.

Zaragoza, Oscar; Chrisman, Cara J; Castelli, Maria Victoria; Frases, Susana; CuencaEstrella, Manuel; RodríguezTudela, Juan Luis, and Casadevall, Arturo. Capsule enlargement in Cryp-

tococcus neoformans confers resistance to oxidative stress suggesting a mechanism for intracellular survival. *Cellular microbiology*, 10(10):2043–2057, 2008.

Appendix

Isolates	Lineage	Country	MAT	Origin	Date	Environment	Latitude	Longitude	Elevation	Provenence
BTe1	VNI-B	Botswana	α	Pigeon guano	2007	Rural	-24.658	25.912	1118.5	Broad Institute
BTe2	VNI-B	Botswana	α	C. mopane	2007	Rural	-21.981	28.428	693.5	Broad Institute
BTe3	VNB-B	Botswana	α	C. mopane	2007	Rural	-21.981	28.428	693.5	Broad Institute
NAe1	VNI-B	Namibia	α	Pigeon guano	21/09/2014	Urban	-25.889	16.821	1393.0	Present Study
NAe2	VNB-A	Namibia	α	Unkown tree species	28/09/2014	Rural	-28.766	17.626	159.5	Present Study
SAe1	VNI-A	South Africa	α	Unknown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe10	VNI-A	South Africa	α	Pigeon faeces	28/09/2014	Rural	-29.731	17.866	890.3	Present Study
SAe11	VNI-B	South Africa	α	Unknown tree species	19/09/2014	Urban	-33.947	18.469	29.5	Present Study
SAe12	VNI-B	South Africa	α	Unknown tree species	21/09/2014	Rural	-32.648	26.652	532.9	Present Study
SAe13	VNI-B	South Africa	α	Unknown tree species	22/09/2014	Rural	-32.648	26.652	532.9	Present Study
Sae14	VNI-B	South Africa	α	Unknown tree species	19/09/2014	Rural	-32.648	26.652	532.9	Present Study
Sae15	VNI-A	South Africa	α	Pigeon guano	24/09/2014	Urban	-33.993	18.563	25.8	Present Study
Sae16	VNI-A	South Africa	α	Unknown tree species	21/09/2014	Urban	-32.779	26.630	442.7	Present Study
Sae17	VNI-A	South Africa	α	Unknown tree species	21/09/2014	Urban	-32.781	26.625	431.3	Present Study
Sae18	VNI-A	South Africa	α	Unknown tree species	21/09/2014	Rural	-32.781	26.625	431.3	Present Study
Sae19	VNI-A	South Africa	α	Unknown tree species	21/09/2014	Urban	-32.778	26.628	438.7	Present Study
SAe2	VNI-A	South Africa	α	Unkown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
Sae20	VNI-A	South Africa	α	Unknown tree species	21/09/2014	Urban	-32.781	26.625	431.3	Present Study
SAe3	VNI-A	South Africa	α	Unknown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe4	VNI-A	South Africa	α	Unkown tree species	04/03/2010	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe5	VNI-A	South Africa	α	Unknown tree species	04/03/2010	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe6	VNI	South Africa	α	Unknown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe7	VNB-A	South Africa	α	Unknown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe8	VNB-A	South Africa	α	Unknown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe9	VNB-A	South Africa	α	Eucalyptus sp.	2007	Rural	-25.533	26.083	1243.2	Broad Institute
Ze1	VNB-A	Zambia	α	C. mopane	19/01/2013	Rural	-16.650	27.043	1216.4	Present Study
Ze10	VNB-B	Zambia	α	C. mopane	02/09/2013	Rural	-17.833	25.768	903.1	Present Study
Ze11	VNI-B	Zambia	α	C. mopane	02/09/2013	Rural	-17.833	25.768	903.1	Present Study

Ze12	VNI-B	Zambia	α	C. mopane	02/09/2013	Rural	-17.847	25.792	902.3	Present Study
Ze13	VNI-B	Zambia	α	C. mopane	02/09/2013	Rural	-17.892	25.858	894.3	Present Study
Ze14	VNB-B	Zambia	a	Brachestesia sp.	17/09/2013	Rural	-12.927	28.245	1172.9	Present Study
Ze15	VNB-B	Zambia	a	C. mopane	22/09/2013	Rural	-11.732	32.142	742.2	Present Study
Ze16	VNB-B	Zambia	a	C. mopane	22/09/2013	Rural	-11.753	32.306	702.2	Present Study
Ze17	VNB-B	Zambia	α	C. mopane	23/09/2013	Rural	-11.868	32.436	589.9	Present Study
Ze18	VNB-B	Zambia	α	C. mopane	23/09/2013	Rural	-11.868	32.436	589.9	Present Study
Ze19	VNB-B	Zambia	α	C. mopane	25/09/2013	Rural	-13.037	31.918	509.2	Present Study
Ze2	VNB-B	Zambia	α	C. mopane	20/01/2013	Rural	-16.626	26.977	1218.7	Present Study
Ze20	VNB-A	Zambia	a	Brachestesia sp.	25/09/2013	Rural	-14.252	31.679	1167.6	Present Study
Ze21	VNB-B	Zambia	α	C. mopane	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze22	VNB-B	Zambia	α	C. mopane	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze23	VNB-B	Zambia	α	C. mopane	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze24	VNI-B	Zambia	α	Soil	17/09/2013	Rural	-13.121	28.184	28.7	Present Study
Ze3	VNB-B	Zambia	a	C. mopane	21/01/2013	Rural	-17.892	25.858	894.3	Present Study
Ze4	VNB-A	Zambia	a	C. mopane	22/01/2013	Rural	-17.842	25.786	904.4	Present Study
Ze5	VNB-A	Zambia	a	C. mopane	22/01/2013	Rural	-17.853	25.802	898.8	Present Study
Ze6	VNB-B	Zambia	α	Soil	22/01/2013	Rural	-17.834	25.767	899.8	Present Study
Ze7	VNB-A	Zambia	α	Soil	22/01/2013	Rural	-17.842	25.786	904.1	Present Study
Ze8	VNI-B	Zambia	α	Eucalyptus sp.	02/09/2013	Urban	-15.508	28.269	1214.2	Present Study
Ze9	VNB-B	Zambia	α	C. mopane	02/09/2013	Rural	-17.896	25.857	895.4	Present Study
BM17-1	Cg	Zambia		Tree	19/01/2013	SMW	-16.65	27.04	1211.6	Present Study
P17-2	Cg	Zambia		Tree	25/09/2013	ZMW	-13.04	31.92	509.2	Present Study
GR31-1	Cg	Zambia		Tree	15/01/2013	CMW	-12.94	28.32	1216.3	Present Study
GR53-1	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1220.2	Present Study
GR57-4	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1238.2	Present Study
GR65-5	Cg	Zambia		Tree	16/01/2013	CMW	-12.99	28.24	1191.6	Present Study
GR67-5	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1194.5	Present Study
GR67D-7	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1194.5	Present Study
GR70D-2	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1182.0	Present Study
GR79D-3	Cg	Zambia		Tree	17/01/2013	CMW	-12.93	28.24	1161.0	Present Study
GRS53	Cg	Zambia		Soil	16/01/2013	CMW	-12.99	28.24	1220.2	Present Study
GRS54-2	Cg	Zambia		Soil	16/01/2013	CMW	-12.99	28.24	1222.9	Present Study
GRS58-2	Cg	Zambia		Soil	16/01/2013	CMW	-12.99	28.24	1212.8	Present Study
GRS61	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1214.7	Present Study
GRS64-4	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1197.4	Present Study
GR67D-3	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1194.5	Present Study
GRS70-1	Cg	Zambia		Rock cove	16/01/2013	CMW	-12.99	28.24	1182.0	Present Study
GR98-1	Cg	Zambia		Soil	17/01/2013	CMW	-12.93	28.25	1187.9	Present Study
BM08A1	Cg	Zambia		Tree	19/01/2013	SMW	-16.65	27.04	1216.4	Present Study
BMS08A-1	Cg	Zambia		Soil	19/01/2013	SMW	-16.65	27.04	1216.4	Present Study
BM09-13	Cg	Zambia		Soil	19/01/2013	SMW	-16.65	27.04	1213.4	Present Study
Z5-10	Cg	Zambia		Tree	25/11/2012	CMW	-8.46	30.17		Present Study
VS24A1-bis	Cg	Zambia		Tree	02/09/2013	ZMW	-17.83	25.77	903.0	Present Study
BM*08B1	Cg	Zambia		Soil	08/09/2013	CMW	-16.65	27.04	1216.4	Present Study
H5-1	Cg	Zambia		Rock cove	17/09/2013	CMW	-13.12	28.19	1239.3	Present Study
H7-2	Cg	Zambia		Rock cove	17/09/2013	CMW	-13.12	28.19	1250.2	Present Study
H10-2	Cg	Zambia		Rock cove	17/09/2013	CMW	-13.12	28.19	1256.9	Present Study
H11-1	Cg	Zambia		Rock cove	17/09/2013	CMW	-13.12	28.19	1256.9	Present Study
H44-4	Cg	Zambia		Tree	17/09/2013	CMW	-12.93	28.25	1172.9	Present Study
K1-1	Cg	Zambia		Tree	20/09/2013	CMW	-12.45	31.29	1438.9	Present Study
K2-1	Cg	Zambia		Tree	20/09/2013	CMW	-12.45	31.29	1438.9	Present Study
K13-2	Cg	Zambia		Tree	20/09/2013	CMW	-11.19	31.59	1501.7	Present Study
K14-6	Cg	Zambia		Tree	20/09/2013	CMW	-11.17	31.61	1427.1	Present Study
KS14-1	Cg	Zambia		Soil	20/09/2013	CMW	-11.17	31.61	1427.1	Present Study
P17-1	Cg	Zambia		Tree	25/09/2013	CMW	-13.04	31.92	509.2	Present Study
I2-4	Cg	Zambia		Tree bark	18/09/2013	CMW	-13.63	28.62	1259.8	Present Study
G18-1	Cg	Zambia		Tree bark	18/09/2013	CMW	-12.48	31.32	1500.4	Present Study
HS10-3	Cg	Zambia		Rock cove	17/09/2013	CMW	-13.12	28.19		Present Study

Table S1: Environmental Isolates Collected in Southern Africa, 49 *Cryptococcus neoformans* var. *grubii* and 38 *Cryptococcus gattii*

Chapter III

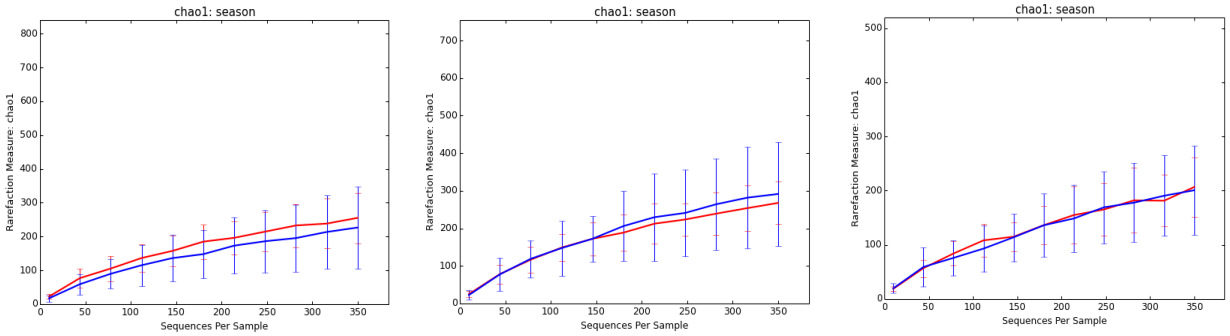
SampleID	Date	Season	Ecoregions	Latitude	Longitude	Elevation	OTUs
BM06	19/01/2013	Rainy	SMW	-16.650	27.042	1221	227
BM07w	19/01/2013	Rainy	SMW	-16.650	27.043	1216	81
BM17	19/01/2013	Rainy	SMW	-16.650	27.043	1216	176
BM7d	19/01/2013	Rainy	SMW	-16.650	27.043	1216	245
BMS03w	19/01/2013	Rainy	SMW	-16.651	27.042	1224	239
BMS08	19/01/2013	Rainy	SMW	-16.650	27.043	1216	235
BMS11	19/01/2013	Rainy	SMW	-16.649	27.044	1215	113
BMS31w	20/01/2013	Rainy	SMW	-16.624	26.974	1213	128
BMS47	20/01/2013	Rainy	SMW	-16.626	26.977	1218	97
BMS53	21/01/2013	Rainy	SMW	-16.642	27.027	1188	128
EucY9	21/01/2013	Rainy	SMW	-16.642	27.027	1188	153
G16	14/01/2013	Rainy	CMW	-12.934	28.260	1189	256
G18	14/01/2013	Rainy	CMW	-12.933	28.259	1188	278
GR33	15/01/2013	Rainy	CMW	-12.945	28.323	1216	173
GRS19	14/01/2013	Rainy	CMW	-12.933	28.259	1186	80
GRS31	15/01/2013	Rainy	CMW	-12.945	28.324	1217	89
GRS34	15/01/2013	Rainy	CMW	-12.944	28.323	1215	133
GRS51	16/01/2013	Rainy	CMW	-12.994	28.240	1219	128
GRS52	16/01/2013	Rainy	CMW	-12.994	28.240	1222	115
GRS58	16/01/2013	Rainy	CMW	-12.995	28.241	1212	164
GRS59	16/01/2013	Rainy	CMW	-12.994	28.241	1214	138
GRS64	16/01/2013	Rainy	CMW	-12.995	28.241	1212	179
GRS69	16/01/2013	Rainy	CMW	-12.993	28.240	1190	113
GRS70	16/01/2013	Rainy	CMW	-12.993	28.240	1181	67
GRS72	16/01/2013	Rainy	CMW	-12.993	28.240	1181	102
GRS88I	17/01/2013	Rainy	CMW	-12.926	28.242	1170	139
GRS92	17/01/2013	Rainy	CMW	-12.927	28.245	1179	222
GRS98	17/01/2013	Rainy	CMW	-12.927	28.246	1187	0

H10	17/09/2013	Dry	CMW	-12.927	28.246	1187	82
H2	17/09/2013	Dry	CMW	-12.927	28.245	1172	340
H23	17/09/2013	Dry	CMW	-12.927	28.245	1172	0
H26	17/09/2013	Dry	CMW	-12.927	28.245	1172	182
H44	17/09/2013	Dry	CMW	-12.927	28.245	1172	297
H7	17/09/2013	Dry	CMW	-12.927	28.245	1172	327
I13	17/09/2013	Dry	CMW	-12.927	28.245	1172	182
I2	17/09/2013	Dry	CMW	-12.927	28.245	1172	43
I3	17/09/2013	Dry	CMW	-12.927	28.245	1172	289
K14	23/01/2013	Rainy	CMW	-15.844	28.239	984	144
K2	23/01/2013	Rainy	CMW	-15.844	28.239	984	177
K7	23/01/2013	Rainy	CMW	-15.843	28.240	983	189
NL13	23/01/2013	Rainy	CMW	-15.843	28.239	986	179
NL21	22/09/2013	Dry	ZMW	-11.753	32.306	702	382
NL24	22/09/2013	Dry	ZMW	-11.753	32.306	702	320
NL36	22/09/2013	Dry	ZMW	-11.753	32.306	702	182
NL38	22/09/2013	Dry	ZMW	-11.753	32.306	702	197
NL57	22/09/2013	Dry	ZMW	-11.753	32.306	702	187
NL65	22/09/2013	Dry	ZMW	-11.753	32.306	702	281
NL9	22/09/2013	Dry	ZMW	-11.753	32.306	702	248
O23	22/09/2013	Dry	ZMW	-11.753	32.306	702	249
O43	23/09/2013	Dry	ZMW	-11.868	32.436	589	206
O57	23/09/2013	Dry	ZMW	-11.868	32.436	589	283
O8	23/09/2013	Dry	ZMW	-11.868	32.436	904	213
V16	23/09/2013	Dry	ZMW	-17.842	25.786	589	155
V18	22/01/2013	Rainy	ZMW	-17.842	25.786	904	165
V18d	22/01/2013	Rainy	ZMW	-17.842	25.786	904	165
V31bis	22/01/2013	Rainy	ZMW	-17.842	25.786	904	159
VF3	23/01/2013	Rainy	ZMW	-17.842	25.786	904	216
VF5A	23/01/2013	Rainy	ZMW	-17.842	25.786	904	13
VF9A	23/01/2013	Rainy	ZMW	-17.842	25.786	904	69
VS17	21/01/2013	Rainy	ZMW	-17.898	25.850	890	171
VS2	21/01/2013	Rainy	ZMW	-17.899	25.856	896	489
VS21	21/01/2013	Rainy	ZMW	-17.899	25.856	896	325

VS21bis	21/01/2013	Rainy	ZMW	-17.899	25.856	896	0
VS24	21/01/2013	Rainy	ZMW	-17.899	25.856	896	132
VS31	21/01/2013	Rainy	ZMW	-17.899	25.856	896	216
VS37A	21/01/2013	Rainy	ZMW	-17.899	25.856	896	132
VS5	21/01/2013	Rainy	ZMW	-17.899	25.856	896	242
W13	28/09/2013	Dry	ZMW	-17.842	25.786	904	113
W14	28/09/2014	Dry	ZMW	-17.842	25.786	904	192
W20	28/09/2015	Dry	ZMW	-17.842	25.786	904	362
W22	28/09/2016	Dry	ZMW	-17.842	25.786	904	163
W29	28/09/2017	Dry	ZMW	-17.842	25.786	904	270
W44	28/09/2018	Dry	ZMW	-17.842	25.786	904	239

Table S2: List of sample used in ITS2 metabarcoding

S2

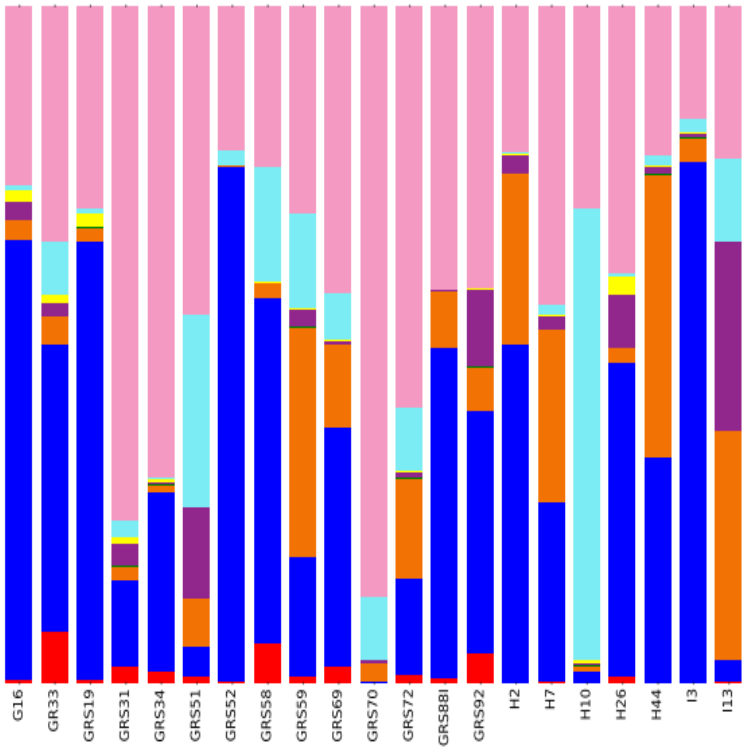


(a) Ecoregion (b) ZMW (c) CMW

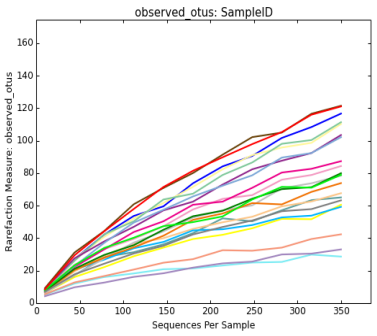
Figure S1: Chao1 α-diversity metric across season: dry season (red) and rainy season (blue)

Table S3: Microbial patterns within Ecoregions per season

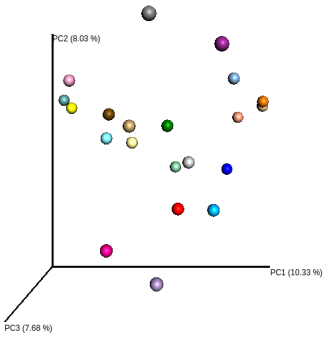
	ZMW				CMW				SMW			
Season	Rainy Season		Dry Season		Rainy Season		Dry Season		Rainy Season		Dry Season	
Phylum	Seq./Sample	%	Seq./Sample	%	Seq./Sample	%	Seq./Sample	%	Seq./Sample	%	Seq./Sample	%
Ascomycota	70.22	26.36%	67.95	33.89%	58.56	29.41%	167.38	50.30%	105.60	26.35%	101.00	56.55%
Basidiomycota	49.89	10.50%	10.26	9.98%	28.17	8.59%	95.88	13.44%	66.20	5.79%	50.00	16.31%
Chytridiomycota	9.00	5.96%	10.21	3.04%	4.39	2.15%	18.13	4.36%	10.20	1.60%	5.67	3.40%
Glomeromycota	7.89	0.33%	5.95	0.28%	3.94	2.88%	14.25	0.51%	8.40	1.29%	5.0	0.13%
Zygomycota	5.44	15.86%	19.00	27.11%	2.50	7.33%	9.13	8.09%	6.60	19.99%	4.67	0.57%
Other	-	3.57%	-	0.47%	-	2.18	-	0.02%	-	1.66%	-	0.55%
Unidentified	-	37.42%	-	25.22%	-	47.47%	-	23.27%	-	43.32%	-	22.49%
Total	145.33	100.00%	143.37	100.00%	137.67	100.00%	221.63	100.00%	197.00	100.00%	166.33	100.00%



(a) Variation at Phylum Levels within samples



(b) Alpna Diversity

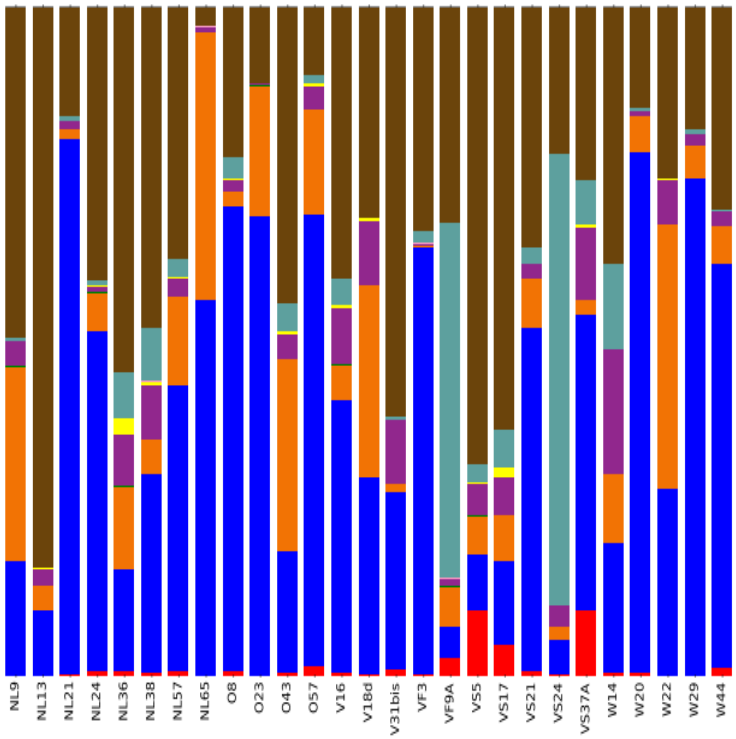


(c) PCoA

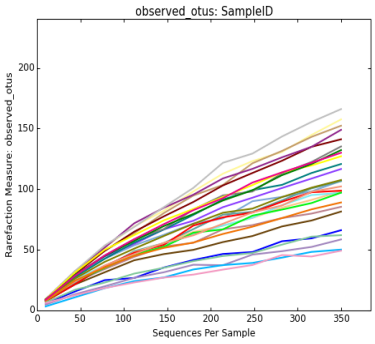
Figure S2: OTUs in Miombo Woodlands

		Total	G16	GR33	GRS19	GRS31	GRS34	GRS51	GRS52	GRS58	GRS59	GRS69	GRS70	GRS72	GRS88I	GRS92	H2	H7	H10	H26	H44	I3	I13
Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	No blast hit;Other	1.5%	0.5%	7.6%	0.6%	2.4%	1.9%	1.0%	0.2%	6.1%	1.1%	2.5%	0.1%	1.3%	0.7%	4.3%	0.1%	0.2%	0.2%	0.9%	0.1%	0.1%	0.3%
	k__Fungi;p__Ascomycota	34.9%	65.1%	42.4%	64.7%	12.9%	26.3%	4.5%	76.1%	50.8%	17.6%	35.3%	0.2%	14.2%	49.0%	36.0%	50.0%	26.7%	1.5%	46.5%	33.5%	77.0%	3.3%
	k__Fungi;p__Basidiomycota	11.1%	2.9%	4.2%	2.0%	2.0%	1.2%	7.0%	0.3%	2.3%	34.0%	12.2%	2.6%	14.9%	8.2%	6.5%	25.3%	25.3%	1.0%	2.1%	41.7%	3.6%	33.7%
	k__Fungi;p__Blastocladiomycota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Fungi;p__Chytridiomycota	3.9%	2.9%	1.9%	0.2%	3.4%	0.4%	13.5%	0.0%	0.1%	2.5%	0.5%	0.5%	0.7%	0.3%	11.4%	2.7%	2.0%	0.2%	8.0%	1.2%	0.5%	28.0%
	k__Fungi;p__Glomeromycota	0.5%	1.5%	1.4%	2.0%	0.8%	0.5%	0.1%	0.0%	0.1%	0.3%	0.1%	0.1%	0.2%	0.0%	0.1%	0.0%	0.1%	0.7%	2.5%	0.1%	0.4%	0.0%
	k__Fungi;p__Zygomycota	8.8%	0.7%	7.8%	0.6%	2.5%	0.3%	28.5%	2.1%	17.1%	14.1%	7.0%	9.4%	9.5%	0.1%	0.1%	0.4%	1.7%	66.7%	0.7%	1.5%	1.8%	12.3%
	k__Fungi;p__unidentified	39.3%	26.4%	34.7%	29.9%	75.9%	69.4%	45.4%	21.3%	23.6%	30.4%	42.2%	87.0%	59.1%	41.7%	41.5%	21.5%	43.9%	29.7%	39.3%	22.0%	16.6%	22.5%

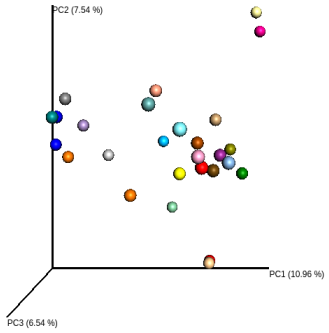
Figure S3: OTUs in Miombo Woodlands at the Order Level



(a) Variation at Phylum Levels within samples



(b) Alpha diversity

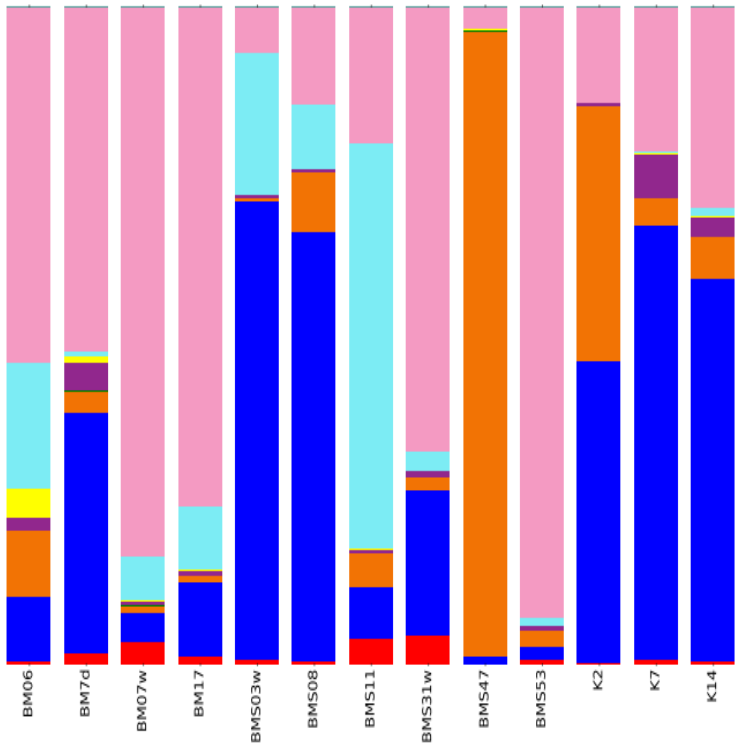


(c) PCoA

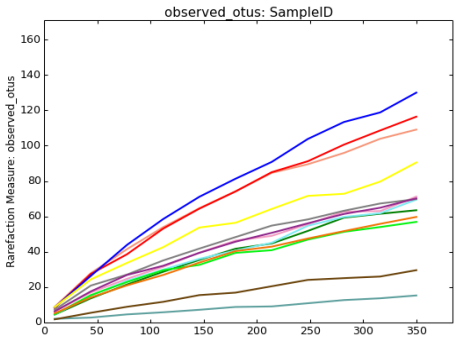
Figure S4: OTUs in Mopane Woodlands

		Total	NL9	NL13	NL21	NL24	NL36	NL38	NL57	NL65	O8	O23	O43	O57	V16	V18d	V31bis	VF3	VF9A	VS5	VS17	VS21	VS24	VS37A	W14	W20	W22	W29	W44
Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	No blast hit;Other	1.5%	0.2%	0.1%	0.4%	0.9%	0.9%	0.5%	0.9%	0.1%	0.8%	0.1%	0.6%	1.6%	0.6%	0.3%	1.1%	0.3%	2.7%	9.8%	4.8%	0.9%	0.4%	10.0%	0.5%	0.6%	0.1%	0.2%	1.3%
	k__Fungi;p__Ascomycota	39.6%	16.9%	9.9%	79.9%	50.7%	15.0%	29.7%	42.6%	56.1%	69.3%	68.6%	18.0%	67.2%	40.8%	29.4%	26.4%	63.5%	4.8%	8.4%	12.4%	51.1%	5.1%	44.0%	19.5%	77.7%	28.0%	74.0%	60.3%
	k__Fungi;p__Basidiomycota	11.5%	29.1%	3.7%	1.4%	5.8%	12.5%	5.1%	13.3%	40.1%	2.2%	19.6%	28.8%	15.8%	5.2%	28.7%	1.4%	0.5%	5.9%	5.7%	6.9%	7.4%	2.0%	2.2%	10.3%	5.3%	39.5%	5.0%	5.7%
	k__Fungi;p__Blastocladiomycota	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Fungi;p__Chytridiomycota	4.5%	3.6%	2.2%	1.1%	0.9%	7.8%	8.0%	2.5%	0.7%	1.8%	0.2%	3.5%	3.3%	8.6%	9.5%	9.4%	0.3%	1.2%	4.8%	5.5%	2.1%	3.0%	10.8%	18.5%	0.8%	6.4%	1.6%	2.1%
	k__Fungi;p__Glomeromycota	0.4%	0.2%	0.3%	0.2%	0.1%	2.5%	0.7%	0.5%	0.0%	0.3%	0.2%	0.6%	0.7%	0.3%	0.4%	0.1%	0.0%	0.1%	0.3%	1.5%	0.1%	0.1%	0.5%	0.1%	0.0%	0.3%	0.0%	0.1%
	k__Fungi;p__Incertae_sedis	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Fungi;p__Neocallimastigomycota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Fungi;p__Zygomycota	6.9%	0.6%	0.2%	0.8%	0.7%	6.9%	8.0%	2.8%	0.1%	3.1%	0.1%	4.2%	1.1%	4.0%	0.2%	0.6%	1.9%	53.2%	2.6%	5.8%	2.4%	67.4%	6.6%	12.6%	0.4%	0.0%	0.8%	0.1%
	k__Fungi;p__unidentified	35.7%	49.3%	83.6%	16.2%	40.9%	54.5%	47.9%	37.5%	2.9%	22.5%	11.3%	44.2%	10.3%	40.6%	31.4%	61.1%	33.4%	32.2%	68.4%	63.0%	36.0%	21.9%	25.8%	38.4%	15.2%	25.6%	18.3%	30.4%
	k__unidentified;p__unidentified	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

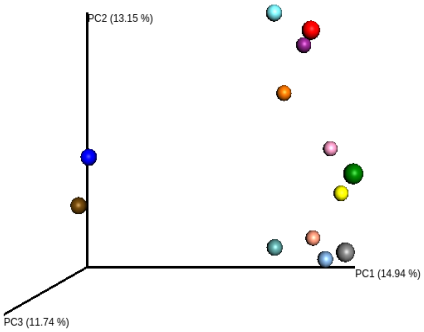
Figure S5: OTUs in Mopane Woodlands at the Phylum Level



(a) Variation at Phylum Levels within samples



(b) Alpha diversity



(c) PCoA

Figure S6: OTUs in Southern Miombo Woodlands

		Total	BM06	BM7d	BM07w	BM17	BMS03w	BMS08	BMS11	BMS31w	BMS47	BMS53	K2	K7	K14
Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	No blast hit;Other	1.5%	0.6%	1.9%	3.4%	1.2%	0.8%	0.5%	4.0%	4.5%	0.2%	0.7%	0.2%	0.8%	0.7%
	k__Fungi;p__Ascomycota	30.7%	9.8%	36.5%	4.6%	11.3%	69.7%	65.2%	7.7%	21.9%	1.0%	2.2%	46.0%	65.8%	57.9%
	k__Fungi;p__Basidiomycota	13.8%	10.1%	3.2%	1.1%	1.0%	0.4%	9.1%	5.2%	2.2%	95.0%	2.3%	38.6%	4.2%	6.5%
	k__Fungi;p__Blastocladiomycota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Fungi;p__Chytridiomycota	1.6%	2.0%	4.4%	0.5%	0.8%	0.4%	0.5%	0.5%	0.8%	0.2%	0.7%	0.5%	6.7%	3.0%
	k__Fungi;p__Glomeromycota	0.5%	4.2%	0.8%	0.4%	0.2%	0.1%	0.0%	0.4%	0.0%	0.1%	0.1%	0.0%	0.0%	0.3%
	k__Fungi;p__Zygomycota	10.5%	19.3%	0.8%	6.7%	9.6%	21.8%	9.9%	61.5%	3.2%	0.2%	1.2%	0.1%	0.5%	1.2%
	k__Fungi;p__unidentified	41.4%	54.0%	52.3%	83.4%	75.8%	6.9%	14.7%	20.6%	67.4%	3.3%	92.9%	14.6%	21.9%	30.5%
	k__unidentified;p__unidentified	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure S7: OTUs in Southern Mopane Woodlands at the Phylum Level

Table S4: Relative contribution of Bioclim layers to the Maxent model (Busby, 1991)

Bioclim Variable		Relative Contribution (%)	
Number	Name	*C. neoformans*	*C. gattii*
bio1	Annual Mean Temperature	0	0
bio2	Mean Diurnal Range	0	0
bio3	Isothermality	71.2	26.5
bio4	Temperature Seasonality	0	0
bio5	Max Temperature of Warmest Month	0.4	0
bio6	Min Temperature of Coldest Month	0	3
bio7	Temperature Annual Range	3.3	0
bio8	Mean Temperature of Wettest Quarter	0	0.2
bio9	Mean Temperature of Driest Quarter	0	7.9
bio10	Mean Temperature of Warmest Quarter	0	0
bio11	Mean Temperature of Coldest Quarter	0	5.1
bio12	Annual Precipitation	0	0
bio13	Precipitation of Wettest Month	0	3.9
bio14	Precipitation of Driest Month	2.6	9.1
bio15	Precipitation Seasonality	22	4.5
bio16	Precipitation of Wettest Quarter	0	18.4
bio17	Precipitation of Driest Quarter	0.5	5.6
bio18	Precipitation of Warmest Quarter	0	16.8
bio19	Precipitation of Coldest Quarter	0	0
Alt	Altitude	0	3.2

Chapter IV

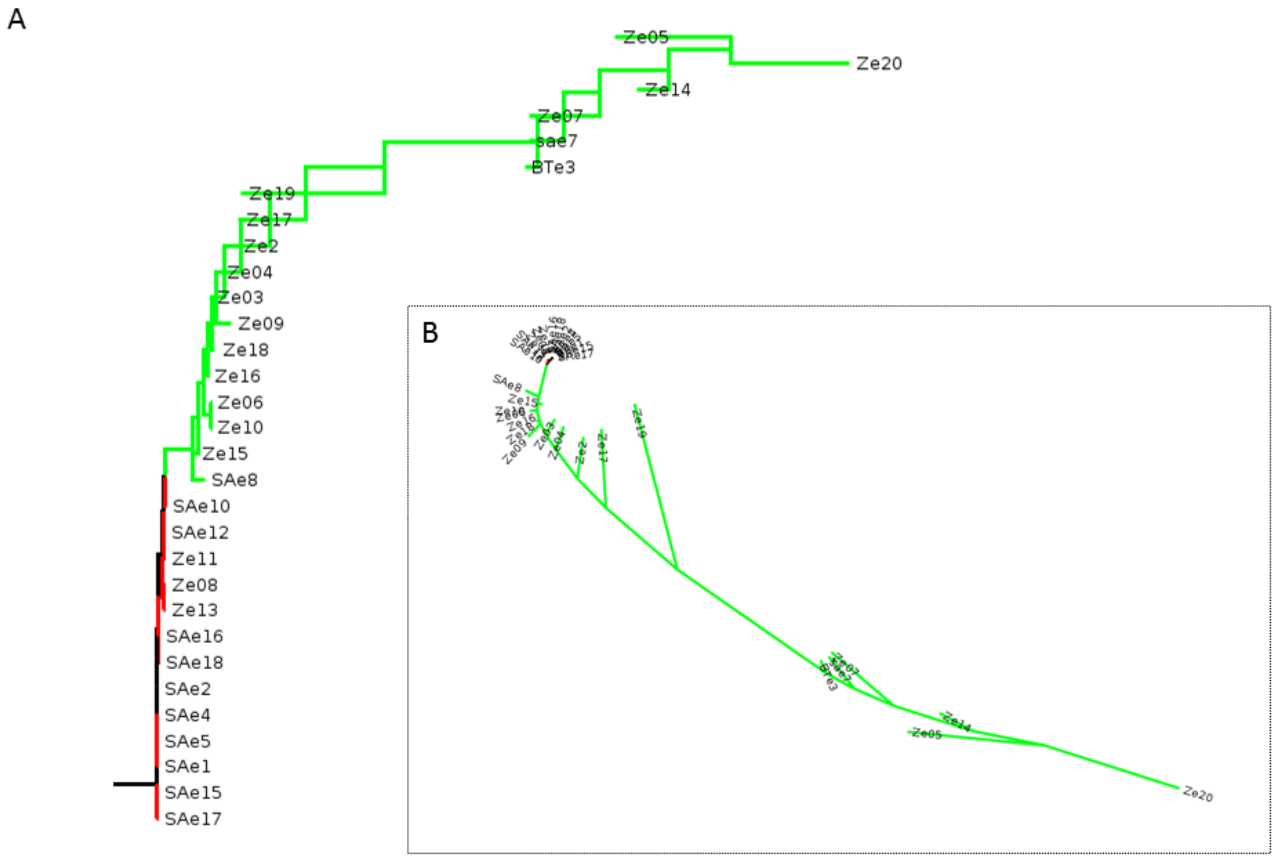


Figure S8: Cng phylogeny based on MLST - MLST loci were extracted in silico. VNB isolates appear in green and VNI isolates are in red.

Isolates	Lineage	Country	Origin	Date	Environment	Latitude	Longitude	Elevation	Provenance
BTe1	VNI-B	Botswana	Pigeon guano	2007	Rural	-24.658	25.912	1118.5	Broad Institute
BTe2	VNI-B	Botswana	C. mopane	2007	Rural	-21.981	28.428	693.5	Broad Institute
BTe3	VNB-B	Botswana	C. mopane	2007	Rural	-21.981	28.428	693.5	Broad Institute
NAe1	VNI-B	Namibia	Pigeon guano	21/09/2014	Urban	-25.889	16.821	1393.0	Present Study
NAe2	VNB-A	Namibia	Unkown tree species	28/09/2014	Rural	-28.766	17.626	159.5	Present Study
SAe1	VNI-A	South Africa	Unknown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe10	VNI-A	Namibia	Pigeon faeces	28/09/2014	Rural	-29.731	17.866	890.3	Present Study
SAe11	VNI-B	South Africa	Unknown tree species	19/09/2014	Urban	-33.947	18.469	29.5	Present Study
SAe12	VNI-B	South Africa	Unknown tree species	21/09/2014	Rural	-32.648	26.652	532.9	Present Study
SAe13	VNI-B	South Africa	Unknown tree species	22/09/2014	Rural	-32.648	26.652	532.9	Present Study
Sae14	VNI-B	South Africa	Unknown tree species	19/09/2014	Rural	-32.648	26.652	532.9	Present Study
Sae15	VNI-A	South Africa	Pigeon guano	24/09/2014	Urban	-33.993	18.563	25.8	Present Study
Sae16	VNI-A	South Africa	Unknown tree species	21/09/2014	Urban	-32.779	26.630	442.7	Present Study
Sae17	VNI-A	South Africa	Unknown tree species	21/09/2014	Urban	-32.781	26.625	431.3	Present Study
Sae18	VNI-A	South Africa	Unknown tree species	21/09/2014	Rural	-32.781	26.625	431.3	Present Study
Sae19	VNI-A	South Africa	Unknown tree species	21/09/2014	Urban	-32.778	26.628	438.7	Present Study
SAe2	VNI-A	South Africa	Unkown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
Sae20	VNI-A	South Africa	Unknown tree species	21/09/2014	Urban	-32.781	26.625	431.3	Present Study
SAe3	VNI-A	South Africa	Unknown tree species	27/11/2007	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe4	VNI-A	South Africa	Unkown tree species	04/03/2010	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe5	VNI-A	South Africa	Unknown tree species	04/03/2010	Urban	-33.928	18.417	28.7	Stellenbosch University
SAe6	VNI	South Africa	Unkown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe7	VNB-A	South Africa	Unknown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe8	VNB-A	South Africa	Unknown tree species	28/01/2008	Rural	-29.641	24.324	1160.5	Stellenbosch University
SAe9	VNB-A	South Africa	*Eucalyptus sp.*	2007	Rural	-25.533	26.083	1243.2	Broad Institute
Ze1	VNB-A	Zambia	*C. mopane*	19/01/2013	Rural	-16.650	27.043	1216.4	Present Study
Ze10	VNB-B	Zambia	*C. mopane*	02/09/2013	Rural	-17.833	25.768	903.1	Present Study
Ze11	VNI-B	Zambia	*C. mopane*	02/09/2013	Rural	-17.833	25.768	903.1	Present Study
Ze12	VNI-B	Zambia	*C. mopane*	02/09/2013	Rural	-17.847	25.792	902.3	Present Study
Ze13	VNI-B	Zambia	*C. mopane*	02/09/2013	Rural	-17.892	25.858	894.3	Present Study
Ze14	VNB-B	Zambia	*Brachestesia sp.*	17/09/2013	Rural	-12.927	28.245	1172.9	Present Study
Ze15	VNB-B	Zambia	*C. mopane*	22/09/2013	Rural	-11.732	32.142	742.2	Present Study
Ze16	VNB-B	Zambia	*C. mopane*	22/09/2013	Rural	-11.753	32.306	702.2	Present Study
Ze17	VNB-B	Zambia	*C. mopane*	23/09/2013	Rural	-11.868	32.436	589.9	Present Study
Ze18	VNB-B	Zambia	*C. mopane*	23/09/2013	Rural	-11.868	32.436	589.9	Present Study
Ze19	VNB-B	Zambia	*C. mopane*	25/09/2013	Rural	-13.037	31.918	509.2	Present Study
Ze2	VNB-B	Zambia	*C. mopane*	20/01/2013	Rural	-16.626	26.977	1218.7	Present Study
Ze20	VNB-A	Zambia	*Brachestesia sp.*	25/09/2013	Rural	-14.252	31.679	1167.6	Present Study
Ze21	VNB-B	Zambia	*C. mopane*	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze22	VNB-B	Zambia	*C. mopane*	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze23	VNB-B	Zambia	*C. mopane*	02/08/2014	Rural	-11.923	32.260	636.0	Present Study
Ze24	VNI-B	Zambia	Soil	17/09/2013	Rural	-13.121	28.184	28.7	Present Study
Ze3	VNB-B	Zambia	*C. mopane*	21/01/2013	Rural	-17.892	25.858	894.3	Present Study
Ze4	VNB-A	Zambia	*C. mopane*	22/01/2013	Rural	-17.842	25.786	904.4	Present Study
Ze5	VNB-A	Zambia	*C. mopane*	22/01/2013	Rural	-17.853	25.802	898.8	Present Study
Ze6	VNB-B	Zambia	Soil	22/01/2013	Rural	-17.834	25.767	899.8	Present Study
Ze7	VNB-A	Zambia	Soil	22/01/2013	Rural	-17.842	25.786	904.1	Present Study
Ze8	VNI-B	Zambia	*Eucalyptus sp.*	02/09/2013	Urban	-15.508	28.269	1214.2	Present Study
Ze9	VNB-B	Zambia	*C. mopane*	02/09/2013	Rural	-17.896	25.857	895.4	Present Study

Table S5: Environmental Isolates Collected in Southern Africa (n=49)

Isolates	Lineage	SNPs	Properly paired	QC passed reads	Duplicates	Mapped (%)
BTe1	VNI	49105	78.22%	29617364	966646	23870684 (80.60%)
BTe2	VNI	46717	76.30%	37194096	1426394	29171744 (78.43%)
BTe3	VNB	225752	82.54%	33024982	1082610	27959743 (84.66%)
Na1	VNI	47908	91.56%	39096578	1430370	36199432 (92.59%)
Na2	VNB	228110	88.41%	29128480	1148482	26149952 (89.77%)
SAe1	VNI	13106	92.23%	4718592	0	4405334 (93.36%)
SAe10	VNI	40040	90.58%	40288398	1364352	37026083 (91.90%)
SAe11	VNI	48070	91.64%	44235392	1608217	41019777 (92.73%)
SAe12	VNI	50630	89.87%	40825332	1629607	37152822 (91.00%)
SAe13	VNI	47932	91.06%	40359166	1579998	37160320 (92.07%)
Sae14	VNI	45791	90.40%	41947760	11001126	37922680 (90.40%)
Sae15	VNI	12942	92.77%	4194304	27414	3928712 (93.67%)
Sae16	VNI	11306	92.67%	32686642	1185400	30558949 (93.49%)
Sae17	VNI	13478	90.35%	32843918	1245839	30068215 (91.55%)
Sae18	VNI	11345	93.76%	38112850	1288877	36125411 (94.79%)
Sae19	VNI	12773	92.14%	2621440	12139	2448871 (93.42%)
SAe2	VNI	13553	91.67%	33674154	1159460	31201090 (92.66%)
Sae20	VNI	13099	93.53%	4194304	26976	3965344 (94.54%)
SAe3	VNI	13527	91.98%	31514476	0	29354013 (93.14%)
SAe4	VNI	12970	91.88%	3670016	0	3418387 (93.14%)
SAe5	VNI	13435	92.29%	16252928	0	15198094 (93.51%)
SAe6	VNI	20587	87.09%	524288	0	466867 (89.05%)
SAe7	VNB	227771	88.72%	21744964	560109	19748633 (90.82%)
SAe8	VNB	230235	88.72%	39092750	0	35472028 (90.74%)
SAe9	VNB	228606	78.55%	32055210	0	25923501 (80.87%)
Ze1	VNB	229494	89.10%	24092548	669479	21852463 (90.70%)
Ze10	VNB	228870	85.98%	25609122	0	22536999 (88.00%)
Ze11	VNI	50615	92.59%	38595694	1502825	36230201 (93.87%)
Ze12	VNI	51210	92.69%	33450214	0	31346062 (93.71%)
Ze13	VNI	49291	92.72%	32930040	1059178	30868813 (93.74%)
Ze14	VNB	226095	88.21%	27611248	0	24728703 (89.56%)
Ze15	VNB	228906	88.71%	30085276	0	27151664 (90.25%)
Ze16	VNB	228272	89.11%	36765416	0	33378262 (90.79%)
Ze17	VNB	224979	33.81%	38672106	0	13412041 (34.68%)
Ze18	VNB	226013		37049652	0	
Ze19	VNB	224831	88.23%	36029014	0	32396137 (89.92%)
Ze2	VNB	228474	85.65%	25225122	0	22030443 (87.34%)
Ze20	VNB	233528	54.26%	39159723	0	21675945 (55.35%)
Ze21	VNB	227097	79.95%	40210994	1374069	32775147 (81.51%)
Ze22	VNB	227833	86.17%	18350080	388234	16110005 (87.79%)
Ze23	VNB	227186	85.08%	39476726	1417298	34302743 (86.89%)
Ze24	VNI	51182				
Ze3	VNB	227768	82.80%	24214310	0	20436331 (84.40%)
Ze4	VNB	231762				
Ze5	VNB	229188	74.03%	16625000	0	12589705 (75.73%)
Ze6	VNB	228169	86.53%	16202400	0	14321015 (88.39%)
Ze7	VNB	230349				
Ze8	VNI	51094	90.29%	21109950	0	19435138 (92.07%)
Ze9	VNB	227092	86.00%	45815366	0	40282028 (87.92%)

Table S6: Sequencing Results for Environmental *C. neoformans* Genomes (n=50)

Isolates	Lineage	SNPs (H99 reference)	SNPs (VNB reference)
BTe3	VNB	225752	184297
Na2	VNB	228110	152303
SAe7	VNB	227771	149769
SAe8	VNB	230235	151737
SAe9	VNB	228606	0
Ze1	VNB	229494	149803
Ze2	VNB	228474	177515
Ze3	VNB	227768	184190
Ze4	VNB	231762	149958
Ze5	VNB	229188	149070
Ze6	VNB	228169	171000
Ze7	VNB	230349	150011
Ze9	VNB	227092	184584
Ze10	VNB	228870	164293
Ze15	VNB	228906	183131
Ze16	VNB	228272	181926
Ze17	VNB	224979	182738
Ze18	VNB	226013	184644
Ze19	VNB	224831	183838
Ze20	VNB	233528	131808
Ze21	VNB	227097	183941
Ze22	VNB	227833	182114
Ze23	VNB	227186	183964

Table S7: SNPs found using VNI H99 reference and using a VNB SAe9 reference

Table S8: SNPs unique to each *Cng* Populations

Populations	VNI	VNB	VNBa	VNBb	VNIa	VNIb
VNI	-	78,132	85,080	84,406	78169	35774
VNB	509,332	-	186,734	164,455	535,107	528,728
VNBa	337,100	7,554	-	170,258	359,016	355,065
VNBb	354,441	3,290	188,273	-	377,685	371,316
VNIa	228	25,966	29,055	29,709	-	35,879
VNIa	155	61,909	67,426	65,662	7,8201	-
VNIb	111,152	49,398	43,881	45,645	33,106	-

Gene	chr	location	Product	
CNAG_01117	5	1163211	1,167,713	elongation factor 3 (EF3)
CNAG_01115	5	1169182	1,179,527	hypothetical protein
CNAG_01111	5	1183060	1,184,494	translation initiation factor 3 subunit K
CNAG_01110	5	1184559	1,185,429	co-chaperone
CNAG_01109	5	1185241	1,186,312	hypothetical protein
CNAG_12470	5	1186506	1,187,397	hypothetical RNA
CNAG_12471	5	1186671	1,187,404	hypothetical RNA
CNAG_01106	5	1193915	1,197,594	vacuolar-ATPase subunit (VPH1)
CNAG_01105	5	1197607	1,201,494	hypothetical protein

Table S9: Strong Gene Candidates under selection in the VNB lineage

Gene	Chr	location	Product	
VNI				
CNAG_06650	7	386,543	387,806	hypothetical protein
VNB				
CNAG_00532	1	1,359,861	1,361,600	hypothetical protein
CNAG_02707	3	949,796	951,728	hypothetical protein
CNAG_07575	3	1,338,366	1,340,046	hypothetical protein
CNAG_09001.2	UN	725	1,291	unspecified product
CNAG_07762	5	291,486	292,074	hypothetical protein
CNAG_08026	14	455,445	457,031	hypothetical protein
CNAG_07910	12	545,256	545,815	hypothetical protein
CNAG_07643	6	780,965	782,055	hypothetical protein
CNAG_06629	7	322,289	323,345	hypothetical protein
CNAG_01972	11	1,406,904	1,411,072	C2H2 zinc finger protein (TAF10)
CNAG_06652	7	390,814	393,384	allantoate permease
CNAG_04982	4	142,196	144,512	putative cytosine-purine permease (FCY3)
CNAG_07788	9	1,169,905	1,170,625	hypothetical protein
CNAG_07876	17	339,080	340,980	hypothetical protein

Table S10: Gene identify using the BSR analyses

Gene	chr	location	Product	
CNAG_00105	1	289,668	291,138	pyridoxal reductase
CNAG_00301	1	775,691	777,242	hypothetical protein
CNAG_00306	1	787,724	789,157	hypothetical protein
CNAG_06725	2	222,242	224,597	DNA polymerase delta subunit 2
CNAG_03761	2	730,611	734,436	mRNA 3'-end-processing protein RNA14
CNAG_07493	2	1,526,737	1,528,519	hypothetical protein
CNAG_02776	3	790,914	791,596	hypothetical protein
CNAG_04975	4	128,153	129,219	hypothetical protein
CNAG_05092	4	417,487	417,930	hypothetical protein
CNAG_05165	4	599,392	600,828	hypothetical protein
CNAG_06813	5	189,560	192,241	O-glucosyltransferase (CAP1alpha)
CNAG_07005	5	242,901	246,213	armadillo/beta-catenin repeat protein; encoded on mat locus (BSP2)
CNAG_07004	5	244,073	247,927	putative dihydrolipoyl dehydrogenase (LPD1alpha)
CNAG_07410	5	248,015	251,440	cid1-family polyA polymerase (CID1alpha)
CNAG_07015	5	251,935	255,597	hypothetical protein
CNAG_01456	5	267,859	269,211	hypothetical protein
CNAG_07425	5	591,725	592,967	hypothetical protein
CNAG_01070	5	1,281,798	1,283,750	class II aldolase/adducin family protein
CNAG_01012	5	1,438,275	1,439,699	hypothetical protein
CNAG_02556	6	5,753	7,090	phytanoyl-CoA dioxygenase
CNAG_07624	6	170,971	171,547	hypothetical protein
CNAG_02420	6	369,051	370,925	hypothetical protein
CNAG_02302	6	659,821	661,450	hypothetical protein
CNAG_02143	6	1,084,749	1,088,015	hypothetical protein
CNAG_07648	6	1,245,217	1,246,377	hypothetical protein
CNAG_07667	7	596,703	600,191	HAL protein kinase
CNAG_06590	7	207,289	209,626	hypothetical protein
CNAG_06525	7	10,214	12,607	nuclear protein
CNAG_05757	7	785,586	787,180	hypothetical protein
CNAG_03300	8	601,363	602,822	hypothetical protein
CNAG_03095	8	37,338	38,313	hypothetical protein
CNAG_03284	8	564,445	566,037	rRNA biogenesis protein RRP36
CNAG_03323	8	648,399	649,279	hypothetical protein
CNAG_03555	8	1,267,918	1,269,465	acylglycerone-phosphate reductase
CNAG_04124	9	71,261	73,828	hypothetical protein
CNAG_04457	9	912,988	915,929	hypothetical protein
CNAG_05964	9	1,347,911	1,349,549	hypothetical protein
CNAG_04926	10	8,336	10,243	hypothetical protein
CNAG_07588	11	133,914	134,786	hypothetical protein
CNAG_01900	11	1,205,599	1,207,321	hypothetical protein
CNAG_01945	11	1,334,767	1,335,348	hypothetical protein
CNAG_06199	12	94,466	599,944	nuclear DNA helicase II
CNAG_06200	12	599,694	602,151	hypothetical protein
CNAG_06413	13	418,426	421,014	hypothetical protein
CNAG_06502	13	687,410	691,144	solute carrier family 25 (mitochondrial carnitine/acylcarnitine transporter), member 20/29
CNAG_06507	13	698,252	699,893	hypothetical protein
CNAG_05335	14	8,445	8,863	hypothetical protein
CNAG_05339	14	18,334	19,627	hypothetical protein
CNAG_05421	14	268,721	271,011	hypothetical protein
CNAG_07883	14	482,264	484,723	hypothetical protein
CNAG_05542	14	586,437	587,522	hypothetical protein
CNAG_05612	14	757,536	759,649	hypothetical protein

Table S11: VNB genes under selection

Isolates	Total Gene	MK Mean	MK>0	MK=0	MK<0	TOTAL	Dn/Ds	average w>5	w>5	average w>2	w>2	alpha mean
Ze1	6693	0.02306581	3005	5435	2134	10574	0.8441765	7.402128	47	3.64927	338	-0.1446043
Ze2	6691	0.02228277	2953	5427	2168	10548	0.8344186	7.46875	40	3.62396	309	-0.1450357
Ze14	6668	0.02420823	2933	5360	2137	10430	0.8475175	7.528369	47	3.734747	325	-0.1412365
Ze15	6699	0.02337106	2983	5445	2175	10603	0.848087	7.210317	42	3.653923	338	-0.1345772
Ze16	6676	0.02344228	2960	5425	2166	10551	0.8395259	7.319444	36	3.575951	335	-0.1338455
Ze17	6669	0.02379132	2941	5363	2151	10455	0.8470809	7.117424	44	3.635296	337	-0.131849
Ze18	6673	0.02445167	2969	5347	2110	10426	0.8393405	7.268519	45	3.643771	318	-0.1325726
Ze19	6682	0.02409595	2926	5362	2143	10431	0.8445474	7.22037	45	3.658451	327	-0.1389479
Ze20	6719	0.02126947	3031	5504	2170	10705	0.8447551	7.450794	42	3.584482	342	-0.1402928
SAe7	6698	0.02287324	3004	5438	2138	10580	0.8462231	7.428472	48	3.669294	339	-0.1463916
SAe8	6726	0.02262496	3000	5396	2100	10496	0.8527945	7.125887	47	3.697112	334	-0.1617885
BTe3	6669	0.02424363	2958	5382	2134	10474	0.8334861	7.253133	38	3.596352	316	-0.1244133
Ze3	6684	0.02410982	2950	5404	2146	10500	0.8443452	7.452381	42	3.687321	321	-0.1433074
Ze10	6683	0.02587269	3045	5442	2106	10593	0.8342959	7.288462	39	3.666879	316	-0.1325608
Ze4	6712	0.02323202	3015	5425	2109	10549	0.8655045	8.395425	51	3.985078	332	-0.1435839
Ze5	6710	0.02514551	3037	5440	2085	10562	0.8506016	7.198333	50	3.646352	351	-0.1412176
Ze6	6681	0.0259534	3044	5440	2102	10586	0.8319284	7.262821	39	3.656098	312	-0.1325934
Ze7	6732	0.02456655	3067	5488	2089	10644	0.8570894	7.504762	49	3.751834	344	-0.1343114
Ze9	6688	0.02420138	2978	5375	2132	10485	0.8487895	7.106061	44	3.635163	339	-0.1417223
Na2	6710	0.02410427	3007	5415	2104	10526	0.8527948	7.456439	44	3.721468	331	-0.1556394
Ze21	6688	0.02420138	2978	5375	2132	10485	0.8487895	7.106061	44	3.635163	339	-0.1417223
Ze22	6688	0.02366801	2985	5410	2128	10523	0.848487	7.398148	45	3.669229	337	-0.1411771
Ze23	6687	0.02339097	2952	5394	2159	10505	0.8448984	7.199275	46	3.742042	317	-0.1390209
SAe9	6710	0.02295332	3048	5448	2123	10619	0.8524273	7.428	50	3.751327	340	-0.143468

Table S12: direction of selection VNB

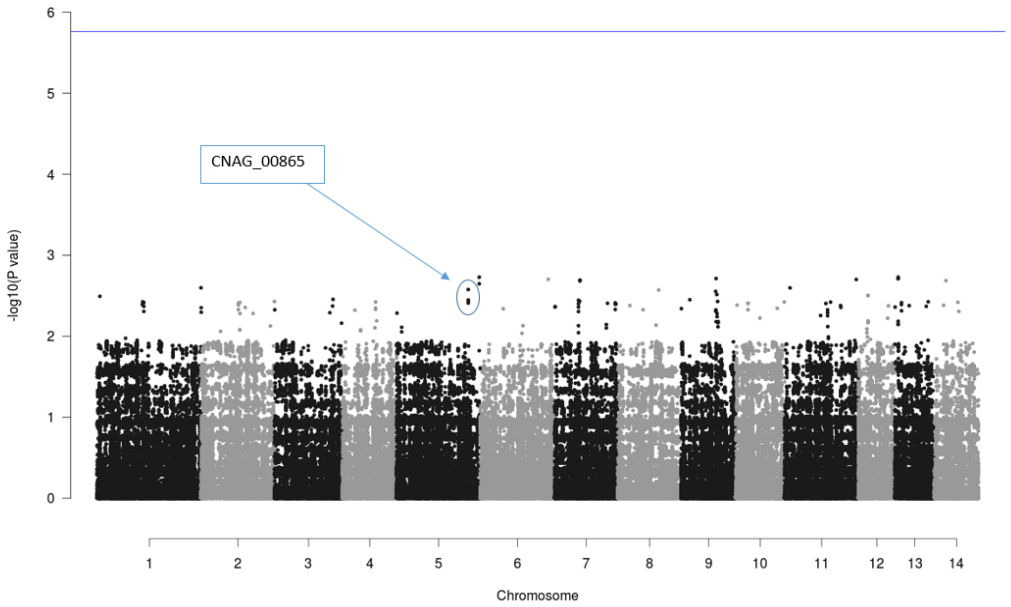


Figure S9: LFMM Analyses for VNB isolates (n=24) for Max Temperature of Warmest Month (Bio05) variable

Gene	Chromosome	Start	End	Product
CNAG_00801	1	2120132	2121489	hypothetical protein centroid_95289
CNAG_06706	2	261724	263626	proline, glutamic acid and leucinerich protein 1 centroid_4117
CNAG_06707	2	260065	261943	hypothetical protein - centroid_488857
CNAG_02572	3	1335789	1336693	hypothetical protein centroid_281459
CNAG_01377	5	450478	454927	PAB dependent poly(A) specific ribonuclease subunit PAN2 - centroid_284162
CNAG_01338	5	561412	565790	xeroderma pigmentosum group C complementing protein centroid_11469
CNAG_02058	6	1312666	1315226	hypothetical protein centroid_15744
CNAG_07644	6	783589	784343	hypothetical protein centroid_701474
CNAG_06529	7	20795	22419	hypothetical protein centroid_351776
CNAG_03090	8	24472	26968	hypothetical protein - centroid_18193
CNAG_08007	11	805522	806023	hypothetical protein centroid_479689
CNAG_06119	12	362838	365291	spermine transporter - centroid_118292
CNAG_06105	12	322686	326574	hypothetical protein - centroid_572437
CNAG_06156	12	450393	454538	hypothetical protein - centroid_724762

Table S13: Gene identified in the VNBa clade using the Gene content analyses

Gene	Chromosome	Start	End	Product
CNAG_00120	1	320587	321863	hypothetical protein centroid_474420
CNAG_07831	4	1058232	1058869	hypothetical protein centroid_453156
CNAG_05195	4	688088	692641	ubiquitin conjugation factor E4 B centroid_167297
CNAG_03547	8	1249924	1251103	hypothetical protein centroid_494371
CNAG_03547	8	1249924	1251103	hypothetical protein centroid_177228
CNAG_04326	9	572191	575234	malate dehydrogenase (oxaloacetate decarboxylating) centroid_179091
CNAG_06891	3	1438098	1440172	hypothetical protein centroid_261469
CNAG_06892	3	1440324	1442110	preconditioning inducible protein centroid_451126
CNAG_00988	5	1495289	1499258	importin subunit beta 1 centroid_170447
CNAG_02202	6	928670	929702	adenylylsulfate kinase centroid_33147
CNAG_04326	9	572191	575234	malate dehydrogenase (oxaloacetate decarboxylating) - centroid_179091
CNAG_01521	11	172980	174759	metallo beta lactamase centroid_245935

Table S14: Gene identified in the VNIa clade using the Gene content analyses

Gene	Chromosome	Start	End	Product
CNAG_03731	2	647676	649939	hypothetical protein centroid_217148
CNAG_03649	2	439385	440695	hypothetical protein centroid_671558
CNAG_06926	3	1535764	1539504	hypothetical protein centroid_403030
CNAG_00914	5	1641933	1644824	glucosidase centroid_178419
CNAG_02339	6	570155	571309	hypothetical protein centroid_175896
CNAG_05760	7	797490	798920	endoplasmic reticulum protein - centroid_684288
CNAG_03224	8	385970	386833	hypothetical protein centroid_533785
CNAG_04438	9	871191	873048	hypothetical protein centroid_688916
CNAG_07849	10	426793	428250	hypothetical protein centroid_416857
CNAG_01512	11	152936	155993	hypothetical protein - centroid_660910
CNAG_05996	12	38200	41583	solute carrier family 32 (vesicular inhibitory amino acid transporter) centroid_541515

Table S15: Gene identified in the VNBb clade using the Gene content analyses

Gene	Chromosome	Start	End	Product
CNAG_03731	2	647676	649939	hypothetical protein - centroid_217148
CNAG_03649	2	439385	440695	hypothetical protein centroid_671558
CNAG_06926	3	1535764	1539504	hypothetical protein centroid_403030
CNAG_00914	5	1641933	1644824	glucosidase centroid_178419
CNAG_02339	6	570155	571309	hypothetical protein - centroid_175896
CNAG_05760	7	797490	798920	endoplasmic reticulum protein - centroid_684288
CNAG_03224	8	385970	386833	hypothetical protein centroid_533785
CNAG_04438	9	871191	873048	hypothetical protein centroid_688916
CNAG_07849	10	426793	428250	hypothetical protein centroid_416857
CNAG_01512	11	152936	155993	hypothetical protein - centroid_660910
CNAG_05996	12	38200	41583	solute carrier family 32 (vesicular inhibitory amino acid transporter) centroid_541515

Table S16: Gene identified in the VNIb clade using the Gene content analyses

Virulence-associated Genes identified from the literature									
CNAG_01166	CNAG_04796	CNAG_01489	CNAG_03627	CNAG_03590	CNAG_00193	CNAG_05449	CNAG_00776	CNAG_07387	CNAG_03695
CNAG_04282	CNAG_05465	CNAG_00762	CNAG_01106	CNAG_02727	CNAG_01984	CNAG_05282	CNAG_04096	CNAG_00165	CNAG_01915
CNAG_02399	CNAG_05144	CNAG_00051	CNAG_06278	CNAG_03637	CNAG_05563	CNAG_02780	CNAG_06487	CNAG_02802	CNAG_02751
CNAG_05139	CNAG_04552	CNAG_05925	CNAG_02702	CNAG_03932	CNAG_00415	CNAG_06064	CNAG_04714	CNAG_02805	CNAG_03345
CNAG_03482	CNAG_00914	CNAG_05105	CNAG_01938	CNAG_01720	CNAG_01251	CNAG_01758	CNAG_02522	CNAG_03677	CNAG_00289
CNAG_07724	CNAG_05791	CNAG_05469	CNAG_07593	CNAG_05218	CNAG_05708	CNAG_03139	CNAG_02436	CNAG_03492	CNAG_06396
CNAG_01537	CNAG_00363	CNAG_00097	CNAG_00306	CNAG_05766	CNAG_03138	CNAG_00895	CNAG_07554	CNAG_00746	CNAG_01234
CNAG_01681	CNAG_05562	CNAG_06765	CNAG_07408	CNAG_00876	CNAG_06834	CNAG_02337	CNAG_04358	CNAG_03694	CNAG_06761
CNAG_00600	CNAG_02867	CNAG_01841	CNAG_02085	CNAG_06031	CNAG_02764	CNAG_01464	CNAG_05847	CNAG_01158	CNAG_06267
CNAG_03465	CNAG_00125	CNAG_04177	CNAG_02221	CNAG_05721	CNAG_03229	CNAG_00556	CNAG_03582	CNAG_00854	CNAG_05290
CNAG_01172	CNAG_01653	CNAG_04158	CNAG_00795	CNAG_00575	CNAG_02049	CNAG_03263	CNAG_04984	CNAG_01446	CNAG_00052
CNAG_02885	CNAG_04804	CNAG_04982	CNAG_05968	CNAG_04093	CNAG_01454	CNAG_02930	CNAG_02503	CNAG_00284	CNAG_05817
CNAG_04514	CNAG_05882	CNAG_02115	CNAG_05392	CNAG_02555	CNAG_06168	CNAG_01845	CNAG_01925	CNAG_05687	CNAG_06821
CNAG_04293	CNAG_02915	CNAG_05838	CNAG_05422	CNAG_03818	CNAG_06591	CNAG_04678	CNAG_02714	CNAG_02958	CNAG_00124
CNAG_02756	CNAG_01019	CNAG_03190	CNAG_01551	CNAG_03894	CNAG_00772	CNAG_04837	CNAG_01341	CNAG_06813	CNAG_00655
CNAG_00622	CNAG_01283	CNAG_04174	CNAG_05653	CNAG_02753	CNAG_06203	CNAG_03116	CNAG_02260	CNAG_06443	CNAG_02978
CNAG_03464	CNAG_06085	CNAG_00441	CNAG_03409	CNAG_05622	CNAG_01173	CNAG_05301	CNAG_00926	CNAG_06346	CNAG_07422
CNAG_00570	CNAG_01156	CNAG_01839	CNAG_02827	CNAG_03621	CNAG_05015	CNAG_02581	CNAG_01230	CNAG_04642	CNAG_01445
CNAG_00721	CNAG_04388	CNAG_06464	CNAG_04990	CNAG_01890	CNAG_01350	CNAG_05595	CNAG_04969	CNAG_05155	CNAG_01872
CNAG_03316	CNAG_02029	CNAG_03398	CNAG_00025	CNAG_05395	CNAG_00828	CNAG_02820	CNAG_00848	CNAG_04510	CNAG_03096
CNAG_02531	CNAG_00181	CNAG_00769	CNAG_04352	CNAG_04761	CNAG_05785	CNAG_01583	CNAG_03644	CNAG_06345	
CNAG_02797	CNAG_06469	CNAG_01278	CNAG_07507	CNAG_04758	CNAG_05742	CNAG_03838	CNAG_02434	CNAG_03322	
CNAG_05540	CNAG_02036	CNAG_05667	CNAG_03366	CNAG_06224	CNAG_02409	CNAG_00701	CNAG_06317	CNAG_00546	
CNAG_06151	CNAG_06134	CNAG_02779	CNAG_04308	CNAG_07464	CNAG_01988	CNAG_03143	CNAG_01590	CNAG_07636	
CNAG_00802	CNAG_02801	CNAG_05592	CNAG_03188	CNAG_00235	CNAG_03664	CNAG_01690	CNAG_01475	CNAG_02205	
CNAG_05381	CNAG_03735	CNAG_07334	CNAG_01243	CNAG_02812	CNAG_00865	CNAG_06500	CNAG_00457	CNAG_06287	
CNAG_00106	CNAG_00888	CNAG_00986	CNAG_06814	CNAG_00383	CNAG_05662	CNAG_07499	CNAG_06404	CNAG_00596	
CNAG_03772	CNAG_04312	CNAG_04090	CNAG_00601	CNAG_02883	CNAG_05590	CNAG_03130	CNAG_02131	CNAG_01909	
CNAG_03158	CNAG_06016	CNAG_00867	CNAG_03498	CNAG_01644	CNAG_06606	CNAG_01037	CNAG_02315	CNAG_05607	
CNAG_04392	CNAG_00678	CNAG_05835	CNAG_03120	CNAG_00531	CNAG_05690	CNAG_03226	CNAG_00744	CNAG_06453	
CNAG_03438	CNAG_04735	CNAG_06524	CNAG_03929	CNAG_04276	CNAG_03213	CNAG_05939	CNAG_04567	CNAG_03782	
CNAG_07770	CNAG_05062	CNAG_03917	CNAG_01740	CNAG_06322	CNAG_05538	CNAG_03981	CNAG_03099	CNAG_05663	
CNAG_05280	CNAG_00996	CNAG_03202	CNAG_00730	CNAG_05427	CNAG_07315	CNAG_00641	CNAG_01331	CNAG_05762	
CNAG_03670	CNAG_00156	CNAG_03765	CNAG_03910	CNAG_02968	CNAG_06976	CNAG_00091	CNAG_04320	CNAG_05650	
CNAG_06242	CNAG_05970	CNAG_04863	CNAG_00396	CNAG_01611	CNAG_00883	CNAG_00734	CNAG_05135	CNAG_03426	
CNAG_00130	CNAG_05934	CNAG_01183	CNAG_03134	CNAG_00613	CNAG_01850	CNAG_05220	CNAG_05148	CNAG_02620	
CNAG_04864	CNAG_06648	CNAG_01444	CNAG_02057	CNAG_00745	CNAG_00440	CNAG_06125	CNAG_05278	CNAG_04818	
CNAG_01232	CNAG_03164	CNAG_00757	CNAG_01697	CNAG_01377	CNAG_03149	CNAG_01743	CNAG_05883	CNAG_04585	
CNAG_00979	CNAG_03760	CNAG_06370	CNAG_07744	CNAG_04371	CNAG_04453	CNAG_02959	CNAG_03577	CNAG_02083	
CNAG_00697	CNAG_04215	CNAG_03171	CNAG_01167	CNAG_00162	CNAG_05534	CNAG_00815	CNAG_07775	CNAG_01737	
CNAG_01607	CNAG_00554	CNAG_05797	CNAG_01905	CNAG_03355	CNAG_07427	CNAG_05081	CNAG_05462	CNAG_02806	
CNAG_01523	CNAG_01580	CNAG_07604	CNAG_04931	CNAG_05886	CNAG_01163	CNAG_03979	CNAG_06835	CNAG_07937	
CNAG_06241	CNAG_05256	CNAG_00460	CNAG_03406	CNAG_04505	CNAG_05377	CNAG_04763	CNAG_07151	CNAG_00508	
CNAG_04640	CNAG_07403	CNAG_02048	CNAG_01208	CNAG_00864	CNAG_01138	CNAG_05761	CNAG_03845	CNAG_06312	
CNAG_01371	CNAG_03648	CNAG_01626	CNAG_03133	CNAG_05222	CNAG_06301	CNAG_06291	CNAG_05431	CNAG_05152	
CNAG_02225	CNAG_03427	CNAG_04693	CNAG_00187	CNAG_01242	CNAG_04243	CNAG_05818	CNAG_04687	CNAG_01919	
CNAG_01678	CNAG_00445	CNAG_00293	CNAG_00150	CNAG_04119	CNAG_00375	CNAG_01103	CNAG_01654	CNAG_07751	
CNAG_02071	CNAG_04981	CNAG_03151	CNAG_04024	CNAG_04521	CNAG_02130	CNAG_02217	CNAG_01195	CNAG_06917	
CNAG_04162	CNAG_04730	CNAG_05303	CNAG_05583	CNAG_04753	CNAG_00268	CNAG_00149	CNAG_00261	CNAG_00411	
CNAG_00654	CNAG_02476	CNAG_03385	CNAG_05023	CNAG_05998	CNAG_05292	CNAG_06574	CNAG_03422	CNAG_03223	

Table S17: *Cryptococcus neoformans* var. *grubii* - Virulence genes (n = 470)

Table S18: Signficant SNPs identified using different multiple testing methods (VNI isolates)

LFMM	pFDR		Bonferroni Correction	
K=2	0.05	0.01	(1/SNPs)	(0.05/SNPs)
Latitude	26822	13185	88	0
Longitude	0	0	0	0
Elevation	0	0	0	0
Bio1	3147	0	1	0
Bio2	0	0	2	0
Bio3	0	0	0	0
Bio4	1330	137	39	0
Bio5	0	0	1	0
Bio6	0	0	0	0
Bio7	0	0	11	0
Bio8	0	0	4	0
Bio9	0	0	0	0
Bio10	1279	79	45	0
Bio11	0	0	0	0
Bio12	766	0	2	0
Bio13	37706	8665	63	0
Bio14	4850	0	6	0
Bio15	0	0	0	0
Bio16	39476	8658	48	0
Bio17	2803	44	43	0
Bio18	0	0	0	0
Bio19	0	0	0	0

Environmental Variable	Number	Genes Names
Bio1 Bio10 Bio12 Bio13 Bio16 Bio4 Bio5	1	CNAG_00378
Bio10 Bio13 Bio16 Bio2 Bio4 Bio7 Bio8	1	CNAG_05338
Bio10 Bio2 Bio4 Bio7 Bio8	1	CNAG_05339
Bio10 Bio12 Bio13 Bio4 Bio5	1	CNAG_07358
Bio13 Bio14 Bio16 Bio17	1	CNAG_02159
Bio10 Bio4 Bio7	2	CNAG_06536 CNAG_06535
Bio13 Bio16 Bio4	1	CNAG_00150
Bio12 Bio13 Bio16	1	CNAG_07530
Bio10 Bio4	3	CNAG_01847 CNAG_04485 CNAG_00534
Bio13 Bio16	23	CNAG_03637 CNAG_06527 CNAG_06193 CNAG_07791 CNAG_07769 CNAG_04699 CNAG_03633 CNAG_04467 CNAG_00174 CNAG_05212 CNAG_00134 CNAG_00136 CNAG_07656 CNAG_07164 CNAG_05202 CNAG_04417 CNAG_03670 CNAG_01339 CNAG_04468 CNAG_00149 CNAG_00135 CNAG_06192 CNAG_07647
Bio14 Bio17	6	CNAG_04466 CNAG_07644 CNAG_02241 CNAG_02160 CNAG_03258 CNAG_03252
Bio4	16	CNAG_04484 CNAG_02587 CNAG_03025 CNAG_02315 CNAG_01813 CNAG_06879 CNAG_03024 CNAG_03671 CNAG_02095 CNAG_02561 CNAG_00546 CNAG_01411 CNAG_03365 CNAG_02096 CNAG_05334 CNAG_05333
Bio10	4	CNAG_05337 CNAG_00535 CNAG_05346 CNAG_07869
Bio13	7	CNAG_02418 CNAG_04682 CNAG_02917 CNAG_07679 CNAG_02097 CNAG_03479 CNAG_02417
Bio16	4	CNAG_00138 CNAG_01814 CNAG_00139 CNAG_07177
Bio17	16	CNAG_05069 CNAG_00642 CNAG_00638 CNAG_05068 CNAG_00640 CNAG_00641 CNAG_07778 CNAG_06229 CNAG_05193 CNAG_05070 CNAG_06230 CNAG_00636 CNAG_00637 CNAG_00635 CNAG_00639 CNAG_00633

Table S19: Genes Detected using a Landscape Genomics Approach

VNIb	Chr	Start	End	Product	Ref
Tajima's *D*					
CNAG_05742	7	743,723	746,234	putative site-2 protease (STP1)	(Chun et al., 2011)
CNAG_06301	13	133,480	137,052	AGC/Akt protein kinase	(Ko et al., 2009)

Table S20: Virulence-associated genes under selection in the VNBb clade

Table S21: Genes under selection in environmental population

Gene under selection in VNB environmental populations (n=54)		
MK test & ω & D'	1	CNAG_05165
Tajima's D & D'	9	CNAG_12470 CNAG_12471 CNAG_01105 CNAG_01110 CNAG_01111 CNAG_01115 CNAG_01109 CNAG_01117 CNAG_01106
ω & Tajima's D	2	CNAG_07920 CNAG_03888
ω & D'	2	CNAG_00306 CNAG_01070
MK test & D'	5	CNAG_06502 CNAG_07015 CNAG_06725 CNAG_00301 CNAG_03761
MK test & ω	33	CNAG_02776 CNAG_03300 CNAG_05339 CNAG_01945 CNAG_06590 CNAG_05092 CNAG_07624 CNAG_02143 CNAG_05421 CNAG_02420 CNAG_00105 CNAG_07883 CNAG_04457 CNAG_05542 CNAG_01012 CNAG_03323 CNAG_04975 CNAG_05964 CNAG_07588 CNAG_03555 CNAG_03095 CNAG_07425 CNAG_06525 CNAG_02346 CNAG_06507 CNAG_04124 CNAG_02302 CNAG_05757 CNAG_07648 CNAG_02556 CNAG_05335 CNAG_05612 CNAG_03284
Gene under Selection in VNBa subpopulations (n=42)		
D'& Tajima's D	1	CNAG_04509
D'& ω	1	CNAG_01945
D'& Tajima's D	18	CNAG_00566 CNAG_06393 CNAG_02194 CNAG_04205 CNAG_03137 CNAG_05001 CNAG_06394 CNAG_13009 CNAG_02395 CNAG_00835 CNAG_12767 CNAG_01827 CNAG_04510 CNAG_03132 CNAG_00565 CNAG_05377 CNAG_13125 CNAG_04511
MK test & Tajima's D	7	CNAG_00562 CNAG_06725 CNAG_04413 CNAG_02117 CNAG_07486 CNAG_06764 CNAG_05550
ω & Tajima's D	3	CNAG_04503 CNAG_03407 CNAG_07930
Gene under Selection in VNBb subpopulations (n=5)		
MK tests & Tajima's D	3	CNAG_03883 CNAG_01741 CNAG_03873
ω & Tajima's D	1	CNAG_03888
LFMM (VNI)	89	CNAG_00378 CNAG_00134 CNAG_00135 CNAG_00136 CNAG_00138 CNAG_00139 CNAG_00149 CNAG_00150 CNAG_00174 CNAG_00378 CNAG_00534 CNAG_00535 CNAG_00546 CNAG_00633 CNAG_00635 CNAG_00636 CNAG_00637 CNAG_00638 CNAG_00639 CNAG_00640 CNAG_00641 CNAG_00642 CNAG_01339 CNAG_01411 CNAG_01813 CNAG_01814 CNAG_01847 CNAG_02095 CNAG_02096 CNAG_02097 CNAG_02159 CNAG_02160 CNAG_02160 CNAG_02241 CNAG_02315 CNAG_02417 CNAG_02418 CNAG_02561 CNAG_02587 CNAG_02917 CNAG_03024 CNAG_03025 CNAG_03252 CNAG_03258 CNAG_03365 CNAG_03479 CNAG_03633 CNAG_03637 CNAG_03670 CNAG_03671 CNAG_04417 CNAG_04466 CNAG_04467 CNAG_04468 CNAG_04484 CNAG_04485 CNAG_04682 CNAG_04699 CNAG_05068 CNAG_05069 CNAG_05070 CNAG_05193 CNAG_05202 CNAG_05212 CNAG_05333 CNAG_05334 CNAG_05337 CNAG_05338 CNAG_05339 CNAG_05346 CNAG_06192 CNAG_06193 CNAG_06229 CNAG_06230 CNAG_06527 CNAG_06535 CNAG_06536 CNAG_06879 CNAG_07164 CNAG_07177 CNAG_07358 CNAG_07530 CNAG_07644 CNAG_07647 CNAG_07656 CNAG_07679 CNAG_07778 CNAG_07791 CNAG_07869

Chapter V

Table S23: SNPs variations in the VNB molecular clade

VNB	Unique SNPs	Shared SNPs			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	126,085	-			
Cluster 2	165,253	232,799	-		
Cluster 3	177,304	219,267	178,916	-	
Cluster 4	96,847	253,180	227,828	193,094	-

Chr	SNPs	Fst	Gene
1	359424	0.104	CNAG_00131
1	479592	0.115	CNAG_00177
1	479593	0.115	CNAG_00177
1	479605	0.115	CNAG_00177
1	479619	0.115	CNAG_00177
1	479828	0.115	CNAG_00177
1	479830	0.104	CNAG_00177
1	809202	0.104	CNAG_12046
1	871315	0.104	CNAG_07354
1	1690438	0.104	CNAG_00649
2	22270	0.104	CNAG_06794
2	22752	0.105	CNAG_06794
2	493677	0.104	CNAG_03670
2	493712	0.105	CNAG_03670
2	637136	0.105	CNAG_12161
2	805135	0.105	CNAG_03783
2	877925	0.103	-
2	1356584	0.103	CNAG_03996
3	598070	0.104	CNAG_02857
3	598236	0.105	CNAG_02857
3	1076189	0.105	CNAG_02665
3	1084259	0.104	CNAG_02663
3	1087381	0.104	CNAG_12292
3	1087447	0.104	CNAG_12292
3	1094783	0.105	CNAG_02662
3	1390585	0.105	-
4	325343	0.104	CNAG_05060
4	981724	0.105	CNAG_05301
5	435459	0.105	CNAG_01385
5	765378	0.104	CNAG_07430
6	453339	0.105	-
6	712494	0.104	CNAG_02280
6	712496	0.104	CNAG_02280
6	712824	0.105	CNAG_02280
7	563146	0.115	-
7	1256111	0.105	CNAG_05935
7	1262925	0.104	CNAG_05937
7	1264347	0.104	CNAG_05938
7	1264429	0.105	CNAG_05940
7	1273737	0.104	CNAG_12703
7	1276917	0.103	CNAG_05941
7	1278950	0.103	CNAG_12704
7	1278963	0.104	CNAG_12704
7	1279421	0.104	CNAG_12704
7	1305184	0.104	CNAG_07691
7	1305276	0.103	CNAG_12707
8	353425	0.115	CNAG_03210

8	412285	0.105	CNAG_03235
8	417323	0.104	CNAG_03239
8	417367	0.104	CNAG_03239
8	417407	0.104	CNAG_03239
8	548091	0.105	CNAG_03275
8	1143604	0.104	CNAG_07739
9	59751	0.103	CNAG_04119
9	61941	0.104	CNAG_04119
9	71421	0.104	-
9	71428	0.104	-
9	71507	0.106	-
9	71598	0.104	-
9	81543	0.105	-
9	939763	0.104	CNAG_04467
9	948380	0.103	CNAG_04471
9	949427	0.105	CNAG_04472
9	950283	0.105	CNAG_04472
9	1114214	0.105	CNAG_04536
9	1114218	0.105	CNAG_04536
9	1114417	0.105	CNAG_04536
10	210367	0.114	CNAG_04702
10	658884	0.105	CNAG_04701
10	660319	0.104	CNAG_04700
11	32300	0.116	CNAG_01467
11	32336	0.104	CNAG_01467
11	186932	0.103	-
11	388395	0.104	-
11	668877	0.115	CNAG_01707
11	729718	0.104	CNAG_01727
11	890283	0.104	-
11	1088842	0.104	CNAG_01850
11	1088848	0.105	CNAG_01850
11	1089337	0.105	CNAG_01850
11	1089376	0.105	CNAG_01850
11	1089388	0.104	CNAG_01850
11	1089391	0.105	CNAG_01850
11	1089406	0.104	CNAG_01850
11	1089634	0.104	CNAG_01850
12	421688	0.104	CNAG_08018
12	421701	0.105	CNAG_08018
12	539874	0.105	CNAG_13075
12	540469	0.105	CNAG_06181
12	540477	0.104	CNAG_06181
12	540478	0.104	CNAG_06181
13	402758	0.104	CNAG_06404
13	431599	0.104	CNAG_06419
13	480505	0.105	CNAG_06435
14	498108	0.104	CNAG_05512
14	819632	0.105	CNAG_05634
14	819634	0.104	CNAG_05634

Table S24: Genes identified in the VNBa population using a BayeScan Approach

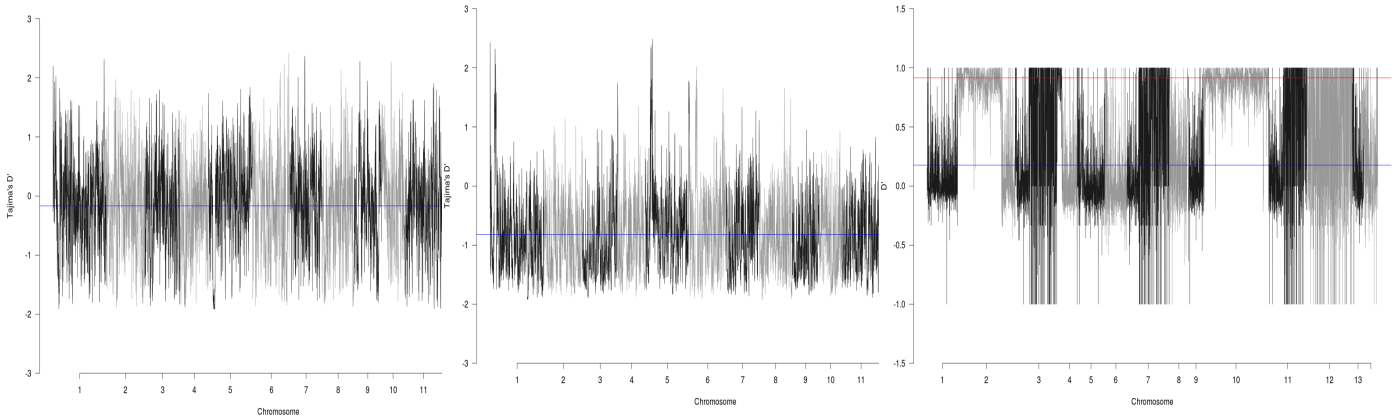
Bayescan in VNBa population

Table S25: Mating Type in Clinical VNB isolates

Clinical isolates	Mating type	Country	VNBa population
SAc1	a	South Africa	VNB-A
SAc2	α	South Africa	VNB-A
SAc3	α	South Africa	VNB-A
SAc4	α	South Africa	VNB-A
SAc5	α	South Africa	VNB-A
SAc6	α	South Africa	VNB-A
SAc7	α	South Africa	VNB-A
SAc8	α	South Africa	VNB-A
SAc9	a	South Africa	VNB-A
SAc10	α	South Africa	VNB-A
SAc11	α	South Africa	VNB-A
SAc12	α	South Africa	VNB-A
SAc13	α	South Africa	VNB-A
SAc14	α	South Africa	VNB-A
Zc07	α	Zambia	VNB-A
Zc15	a	Zambia	VNB-A
Zc21	α	Zambia	VNB-A
Bt1	α	Botswana	VNB-A
Bt206	a	Botswana	VNB-A
Bt63	a	Botswana	VNB-B
Bt85	a	Botswana	VNB-B

Table S26: Gene under selection in Clinical VNBa

Cluster	Product
Cluster 1	
CNAG_00867	myo-inositol transporter (ITR3A)
CNAG_07414	pantoate-beta-alanine ligase (PAN6)
CNAG_06807	hypothetical protein
CNAG_01456	hypothetical protein
CNAG_02056	hypothetical protein
CNAG_07015	hypothetical protein
CNAG_07408	ste/ste20/paka protein kinase (STE20alpha)
CNAG_02049	proline oxidase (PUT1)
CNAG_01436	serine/threonine-protein phosphatase 6 catalytic subunit
CNAG_02055	hypothetical protein
CNAG_00105	pyridoxal reductase
CNAG_00868	hypothetical protein
CNAG_00870	hypothetical protein
Cluster 3	
CNAG_07837	hypothetical protein
CNAG_06977	L-iditol 2-dehydrogenase



(a) Cluster 1 (Environmental Isolates) (n=8)

(b) Cluster 1 (Clinical Isolates) (n=8)

(c) Cluster 3 (Clinical Isolates) (n=9)

Figure S10: Tajima's D over the *Cng* genome for each VNBa sub-population - Value in the upper 0.1% quantile were considered for the analysis of balancing selection

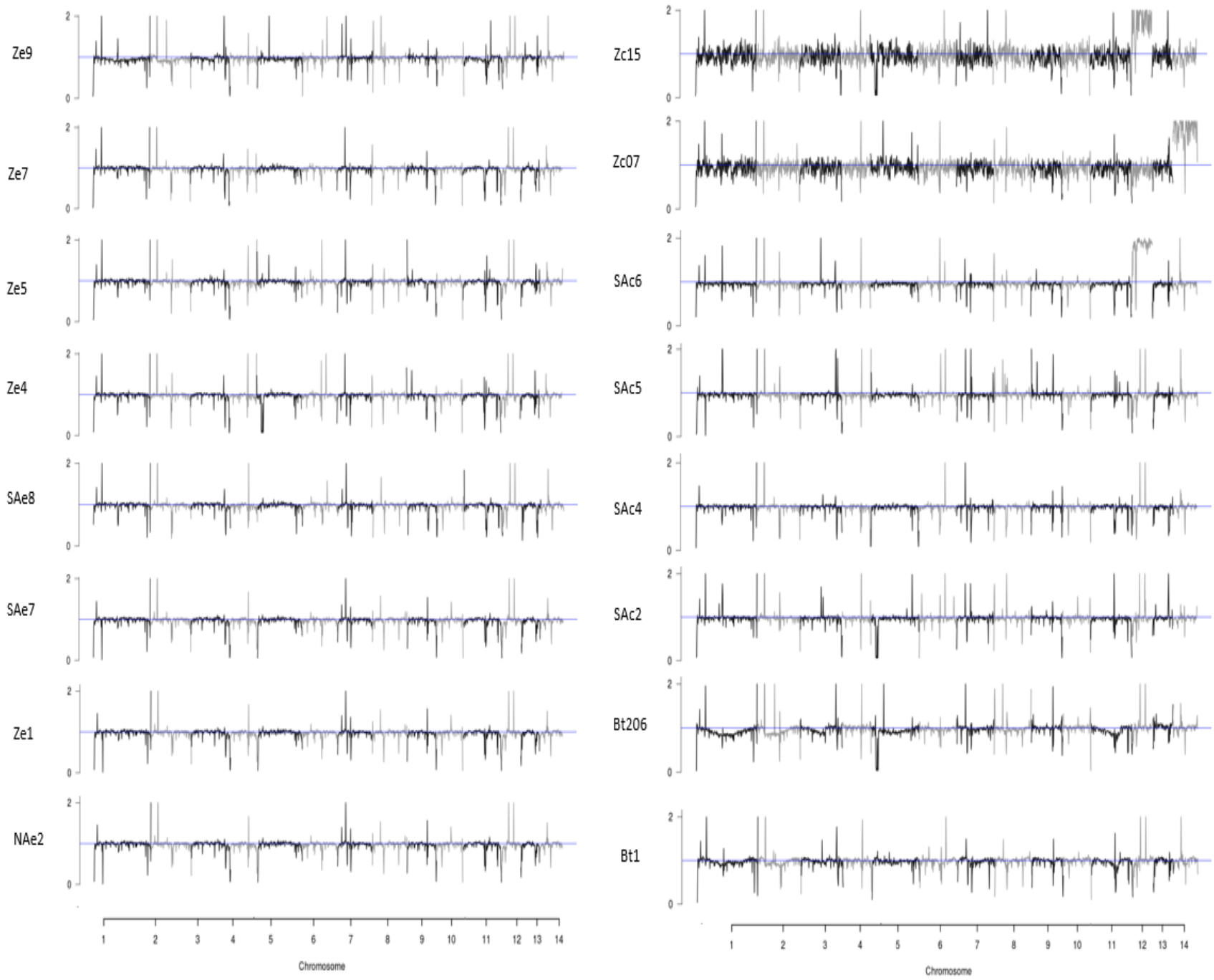


Figure S11: Copy Number Variation - VNB environmental (left) and clinical (right) isolates cluster 1 (n=8) - Average coverage over a

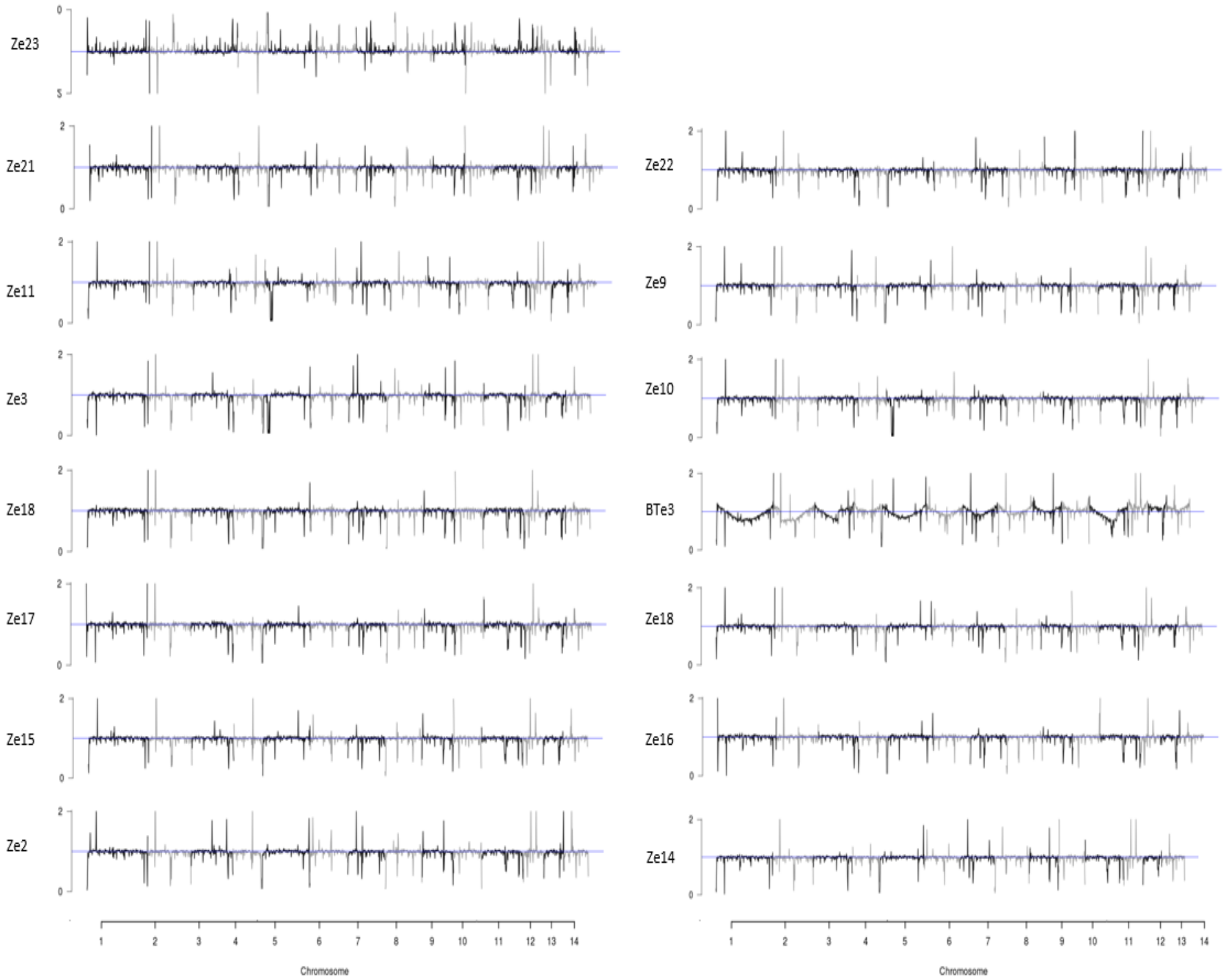


Figure S12: Copy Number Variation - VNB environmental isolates cluster 2 (n=17) - Average coverage over a 10kb non-overlapping

Table S27: Unique SNPs in each VNI genetic cluster

Unique SNPs	C1	C2	C3	C4	C5	C6	C7	C8
Cluster 1	6,859	11,218	10,282	14,589	14,831	16,178	15,897	16,058
Cluster 2	25,199	10,617	16,266	18,903	19,220	21,261	20,221	20,944
Cluster 3	64,939	61,614	30,360	47,190	48,973	52,185	51,513	52,123
Cluster 4	66,153	61,158	44,097	24,804	40,973	44,303	42,519	43,118
Cluster 5	51,848	46,928	31,333	26,426	13,493	30,372	29,979	30,448
Cluster 6	46,344	42,118	27,694	22,905	23,521	9,836	22,784	23,796
Cluster 7	46,597	41,612	27,556	21,655	23,662	23,318	7,465	19,338
Cluster 8	48,562	44,139	29,979	24,058	25,935	26,134	21,142	9,854

Table S28: Shared SNPs in each VNI genetic cluster

Shared SNPs	C1	C2	C3	C4	C5	C6	C7	C8
Cluster 1	-							
Cluster 2	11,218	-						
Cluster 3	16,826	20,151	-					
Cluster 4	12,519	17,514	34,575	-				
Cluster 5	12,277	17,197	32,792	37,699	-			
Cluster 6	10,930	15,156	29,580	34,369	33,753	-		
Cluster 7	11,211	16,196	30,252	36,153	34,146	34,490	-	
Cluster 8	11,050	15,473	29,642	35,554	33,677	33,478	38,470	-

VNI

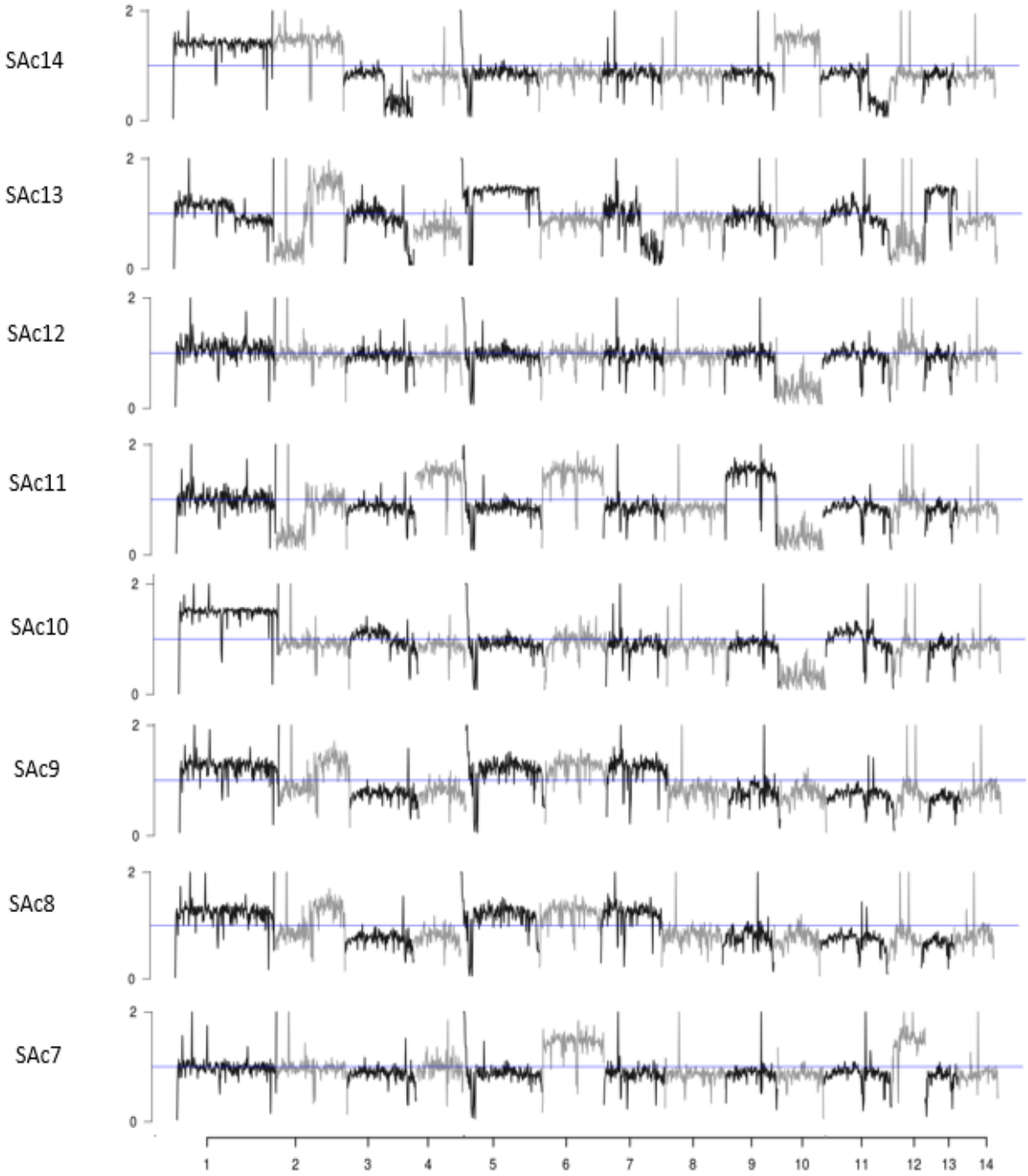


Figure S13: Copy Number Variation - VNB clinical isolates cluster 3 (n=9) - Average coverage over a 10kb non-overlapping sliding window

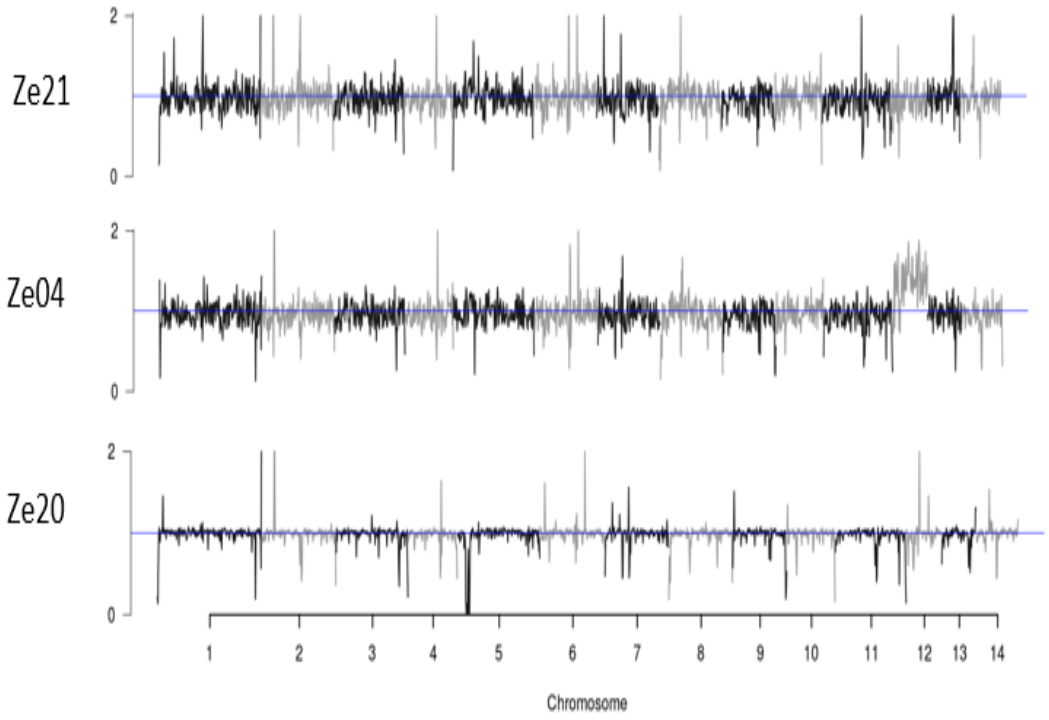
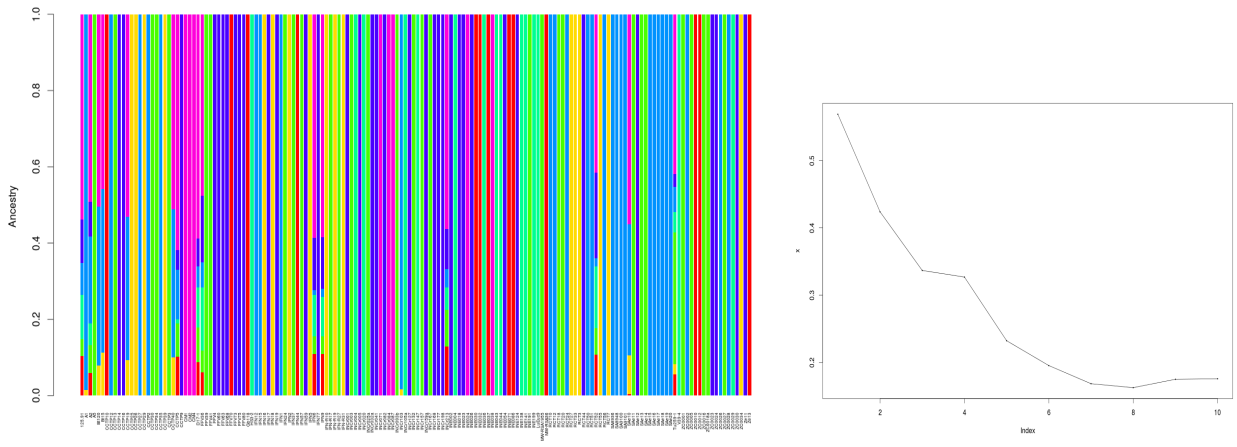


Figure S14: Copy Number Variation - VNB clinical isolates cluster 4 (n=3) - Average coverage over a 10kb non-overlapping sliding window



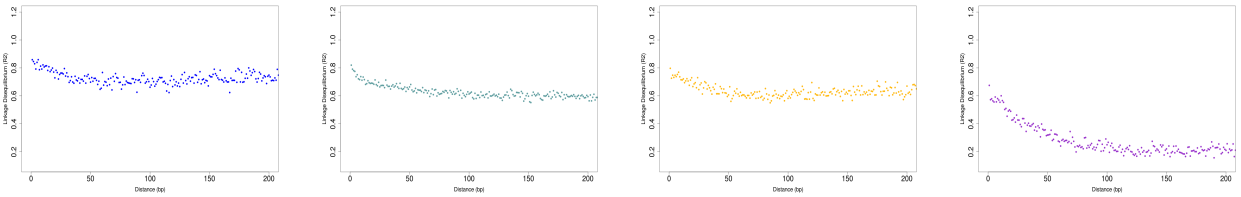
(a) VNI clade (n=160)

(b) VNI Cross-validation

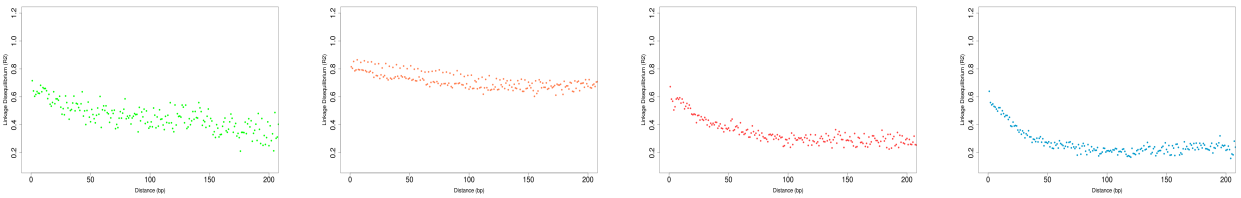
Figure S15: Admixture Results - Admixture results confirms the presence of eight different clusters. The cross-validation plot indicates that K=8 is the preferable number of genetic clusters.



Figure S16: Recombination Analysis using BradNextGen over Chromosome 12



(a) Cluster 1 (c) Cluster 3 (e) Cluster 5 (g) Cluster 7



(b) Cluster 2 (d) Cluster 4 (f) Cluster 6 (h) Cluster 8

Figure S17: LD decay within the eight different VNI genetic clusters

	BTe4	Ze24	Ze8	NA1	SAe2	SAe6	SAe4	SAe5	SAe10	SAe11	SAe12	SAe13	SAe14	SAe15	SAe16	SAe17	SAe18	SAe19	SAe20	BTe2	Ze12	Ze11	Ze13
BTe4		18845	18833	20488	42255	46517	42362	42228	29443	20475	19624	20481	20482	42383	43166	42234	43161	42470	42364	20433	18816	19610	19891
Ze24	21479		797	21045	44699	49128	44843	44675	32115	20998	15802	21044	21035	44867	45803	44683	45763	44956	44812	23300	595	15773	19920
Ze8	21405	735		20960	44611	45092	44775	44611	32036	20920	15753	20948	20951	44788	45730	44598	45685	44886	44748	23192	873	15722	19910
NA1	19848	17771	17748		41716	45964	41866	41713	29306	318	18451	335	338	41874	42614	41705	42592	41993	41853	15785	17729	18475	19776
SAe2	7260	7070	7044	7361		12997	838	406	6545	7339	7269	7361	7358	832	6886	367	6875	1091	745	7714	7081	7266	7345
SAe6	18570	18547	18533	18657	20045		20037	20030	19028	18658	18570	18659	18662	20048	20131	20048	20126	20048	20050	18715	18545	18565	18644
SAe4	6798	6645	6639	6942	269	12420		84	6120	6931	6830	6939	6942	400	6499	253	6493	577	309	7243	6650	6835	6906
SAe5	7129	6942	6940	7254	302	12878	549		6437	7238	7171	7255	7256	714	6779	315	6779	940	606	7588	6971	7163	7239
SAe10	20935	20973	20956	21438	33032	38467	33176	33028		21411	21065	21450	21438	33187	34326	33011	34309	33309	33154	21190	21003	21061	21322
SAe11	19997	17886	17870	480	41856	46127	42017	41859	29441		18613	479	461	42026	42759	41853	42727	42150	42004	15967	17865	18622	19921
SAe12	21706	15250	15263	21173	44346	48599	44476	44352	31655	21173		21199	21192	44473	45287	44331	45247	44562	44468	23232	15235	682	20081
SAe13	19865	17794	17760	359	41740	45990	41887	41738	29342	341	18501		376	41894	42632	41728	41728	42018	41876	15827	17750	18500	19805
SAe14	19865	17784	17762	361	41736	45992	41889	41738	29329	322	18492	375		41895	42628	41725	42602	42019	41879	15827	17743	18508	19819
SAe15	6777	6627	6610	6908	221	12389	358	207	6089	6898	6785	6904	6906		6487	210	6487	504	251	7231	6626	6786	6880
SAe16	5924	5927	5916	6012	4639	10836	4821	4636	5592	5995	5963	6006	6003	4851		4608	236	4959	4811	6071	5931	5969	6034
SAe17	7164	6979	6956	7275	292	12925	747	344	6449	7261	7179	7274	7272	746	6780		6767	1009	672	7631	6987	7177	7256
SAe18	5958	5926	5910	6029	4667	10870	4854	4675	5614	6002	5962	7274	6016	4890	275	4634		4998	4850	6064	5936	5970	6022
SAe19	6695	6547	6539	6858	311	12220	366	264	6042	6853	6705	6859	6861	335	6426	304	6426		144	7158	6554	6703	6827
SAe20	6915	6729	6727	7044	291	12548	424	256	6213	7033	6937	7043	7047	408	6604	293	6604	470		7363	6738	6935	7011
BTe2	18632	18865	18819	14624	40908	44861	41006	40886	27897	14644	19349	14642	14643	41036	41512	40900	41466	41132	41011		18791	19338	19700
Ze12	21500	645	985	21053	44760	49176	44898	44754	32195	21027	15837	21050	21044	44916	45857	44741	45823	45013	44871	23276		15819	19960
Ze11	21677	15206	15217	21182	44328	48579	44466	44329	31636	21167	667	21183	21192	44459	45278	44314	45240	44545	44451	23206	15202		20066
Ze13	20634	18029	18081	21159	43083	47334	43213	43081	30573	21142	18742	21164	21179	43229	44019	43069	43968	43345	43203	22244	18019	18742	

Table S29: Unique SNPs Ciculating in the Environmental African Population (n=23)

	INE002	INE004	INE018	INE023	INE024	INE026	INE030	INE032	INE036	INE037	INE038	INE043	INE044	INE054	INE071	INE086	INE124	INE134	INE138	INE140	INE141	INE191	INE199
INE002		14937	716	691	21071	627	19648	19575	14867	19600	21991	14933	14882	645	21525	19564	384	20766	14817	14819	14771	20770	14743
INE004	15682		15670	15712	21266	15660	19684	19630	729	19639	22472	649	601	15631	19328	19594	15614	20743	935	516	628	20710	602
INE018	739	14948		759	21109	671	19691	19598	14899	19606	22026	14940	14903	426	21546	19578	613	20772	14838	14851	14798	20796	14772
INE023	540	14816	585		20928	397	19525	19447	14742	19459	21850	14824	14755	524	21391	19425	446	20640	14705	14711	14643	20636	14617
INE024	20046	19496	20061	20054		20036	20115	20025	19454	20045	10177	19527	19481	20027	22624	19983	19967	20133	19434	19463	19411	20147	19415
INE026	635	14923	656	556	21069		19652	19574	14866	19592	21982	14891	14839	526	21519	19538	416	20756	14772	14774	14719	20765	14699
INE030	18800	18091	18820	18828	20291	18796		992	18100	1018	21164	18106	18047	18787	9603	858	18738	21178	18031	17997	17988	21171	17980
INE032	18707	18017	18707	18730	20182	18698	972		18030	727	21009	18041	17982	18688	9542	642	18630	21004	17962	17953	17943	21040	17937
INE036	15575	692	15584	15601	21187	15566	19656	19606		19609	22414	730	632	15533	19315	19592	15501	20653	587	576	529	20628	530
INE037	18764	18058	18747	18774	20234	18748	1030	759	18065		21096	18081	18024	18726	9620	642	18684	21077	18025	18000	18002	21111	18001
INE038	19960	19696	19972	19970	9171	19943	19981	19846	19675	19901		19721	19664	19947	22474	19821	19890	20060	19645	19678	19617	20065	19621
INE043	15842	813	15826	15884	21461	15792	19863	19818	931	19826	22661		698	15727	19470	19755	15687	20901	731	534	671	20880	667
INE044	15765	739	15763	15789	21389	15714	19778	19733	807	19743	22578	672		15629	19406	19666	15584	10830	626	499	560	20809	584
INE054	968	15209	726	998	21375	841	19958	19879	15148	19885	22301	15141	15069		21813	19836	591	21002	15020	15000	14944	21008	14923
INE071	11802	8860	11800	11819	13926	11788	728	687	8884	733	14782	8838	8800	11767		640	11724	14487	8795	8726	8754	14466	8735
INE086	18962	18247	18953	18974	20406	18928	1104	908	18282	876	21250	18244	18181	18911	9761		18846	21239	18202	18144	18159	21292	18153
INE124	939	15424	1145	1152	21547	963	20141	20053	15348	20075	22476	15333	15256	823	22002	20003		21180	15155	15120	15079	21199	15063
INE134	18735	17967	18718	18760	19127	18717	19995	19841	17914	19882	20060	17961	17916	18648	22179	19810	18594		17844	17863	17815	1113	17810
INE138	15823	674	15821	15862	21465	15770	19885	19836	885	19867	22682	828	749	15703	19524	19810	15606	20881		566	320	20860	351
INE140	16035	987	16044	16078	21704	15982	20061	20037	1084	20052	22925	841	832	15893	19665	19962	15781	21110	776		648	21082	658
INE141	15931	1043	15935	15954	21596	15871	19996	19971	981	19998	22808	922	837	15781	19637	19921	15684	21006	474	592		20981	377
INE191	18597	17792	18600	18614	18999	18584	19846	19735	17747	19774	19923	17798	17753	18512	22016	19721	18471	971	17681	17693	17648		17636
INE199	15985	1099	15991	16010	21682	15933	20070	20047	1064	20079	22894	1000	943	15842	19700	19997	15750	21083	587	684	459	21051	

Table S30: Unique SNPs Ciculating in the Environmental Indian Population (n=23)

Sample	Latitude	Longitude	Sample	Latitude	Longitude	Sample	Latitude	Longitude	Sample	Latitude	Longitude
LU01	-15.448	28.404	GR91b	-12.927	28.244	BM11Ab	-16.649	27.044	V14B	-17.898	25.850
LUS01	-15.448	28.404	GR92	-12.927	28.245	BM11B	-16.649	27.044	V15	-17.898	25.851
LU02	-15.448	28.404	GR92b	-12.927	28.245	BM12	-16.649	27.044	V15b	-17.898	25.851
LUS02	-15.448	28.404	GR93	-12.927	28.245	BM13A	-16.649	27.044	V16	-17.898	25.850
LU03	-15.448	28.404	GR93b	-12.927	28.245	BM13Ab	-16.649	27.044	V17	-17.898	25.852
LUS03	-15.448	28.404	GR94	-12.927	28.245	BM13B	-16.649	27.044	V18	-17.892	25.858
LU04	-15.448	28.404	GR94b	-12.927	28.245	BM13Bb	-16.649	27.044	V18b	-17.892	25.858
LUS04	-15.448	28.404	GR95	-12.927	28.245	BM14A	-16.649	27.044	V19	-17.892	25.858
LU05	-15.448	28.404	GR95b	-12.927	28.245	BM14Ab	-16.649	27.044	V20A	-17.891	25.859
LUS05	-15.448	28.404	GR96	-12.927	28.245	BM14B	-16.649	27.044	V20B	-17.891	25.859
LU06	-15.448	28.404	GR97	-12.927	28.246	BM14Bb	-16.649	27.044	V21	-17.891	25.859
LUS06	-15.448	28.404	GR98	-12.927	28.246	BM15A	-16.648	27.044	V22	-17.885	25.845
LU07	-15.448	28.404	GR98b	-12.927	28.246	BM15Ab	-16.648	27.044	V22b	-17.885	25.845
LUS07	-15.448	28.404	GR99	-12.929	28.245	BM15B	-16.648	27.044	V23	-17.886	25.845
LU08	-15.448	28.404	GR99b	-12.929	28.245	BM15Bb	-16.648	27.044	V24A	-17.834	25.767
LUS08	-15.448	28.404	GR100	-12.929	28.245	BM16A	-16.648	27.044	V24B	-17.834	25.767
LU09	-15.448	28.405	GR100-DRY	-12.929	28.245	BM16Ab	-16.648	27.044	V25A	-17.834	25.767
LUS09	-15.448	28.405	GR106	-12.929	28.245	BM16B	-16.648	27.044	V25B	-17.834	25.767
LU10	-15.448	28.405	GRPig01	-12.929	28.245	BM16Bb	-16.648	27.044	V26	-17.834	25.767
GR01	-16.651	27.042	GRS01	-16.651	27.042	BM17	-16.650	27.043	V27	-17.833	25.768
GR02	-16.651	27.042	GRS02	-16.651	27.042	BM17b	-16.650	27.043	V28	-17.833	25.768
GR03	-16.651	27.042	GRS03	-16.651	27.042	BM18	-16.625	27.027	V29	-17.835	25.773
GR04	-16.651	27.042	GRS04	-16.651	27.042	BM18b	-16.625	27.027	V29B	-17.835	25.773
GR05	-16.651	27.042	GRS05	-16.651	27.042	BM19	-16.619	27.026	V30	-17.835	25.773
GR06	-16.650	27.042	GRS06	-16.650	27.042	BM20A	-16.619	27.026	V31	-17.842	25.786
GR07	-16.650	27.043	GRS07	-16.650	27.043	BM20Ab	-16.619	27.026	V31-A	-17.842	25.786
GR08	-16.650	27.043	GRS08	-16.650	27.043	BM20B	-16.619	27.026	V32	-17.842	25.786
GR09	-16.650	27.043	GRS09A	-16.650	27.043	BM20Bb	-16.619	27.026	V33	-17.843	25.785
GR10	-16.650	27.043	GRS09B	-16.650	27.043	BM21	-16.619	27.026	V34	-17.843	25.785
GR11	-12.934	28.260	GRS09D	-16.650	27.043	BM21b	-16.619	27.026	V35	-17.843	25.785
GR12	-12.934	28.260	GRS09C	-16.650	27.043	BM22	-16.619	27.026	V35B	-17.843	25.785
GR13	-12.934	28.260	GRS10	-16.650	27.043	BM22b	-16.619	27.026	V36	-17.846	25.793
GR14	-12.933	28.260	GRS11	-12.934	28.260	BM23A	-16.619	27.025	V36b	-17.846	25.793
GR15	-12.933	28.260	GRA11	-12.934	28.260	BM23Ab	-16.619	27.025	V36B	-17.846	25.793
GR16	-12.934	28.260	GRS12	-12.934	28.260	BM23B	-16.619	27.025	V37	-17.847	25.793
GR17	-12.933	28.259	GRS13	-12.934	28.260	BM23Bb	-16.619	27.025	V37B	-17.847	25.793
GR18	-12.933	28.259	GRS14	-12.933	28.260	BM24	-16.619	27.025	V38A	-17.848	25.792
GR19	-12.933	28.259	GRS15	-12.933	28.260	BM25	-16.619	27.024	V38B	-17.848	25.792
GR20	-12.933	28.259	GRS16	-12.934	28.260	BM25b	-16.619	27.024	V39	-17.847	25.792
GR21	-12.930	28.249	GRS17	-12.933	28.259	BM26	-16.626	27.027	V40A	-17.853	25.803
GR22	-12.929	28.248	GRS18	-12.933	28.259	BM27	-16.626	27.027	V40B	-17.853	25.803
GR23	-12.929	28.248	GRS19	-12.933	28.259	BM28	-16.627	27.027	V41	-17.853	25.803
GRA24	-12.929	28.248	GRS20	-12.933	28.259	BM29	-16.624	26.974	V42	-17.853	25.802
GRA25	-12.929	28.248	GRS21	-12.930	28.249	BM30A	-16.624	26.974	V42B	-17.853	25.802
GRA26	-12.929	28.248	GRS22	-12.929	28.248	BM30Ab	-16.624	26.974	V43	-17.853	25.802
GRA27	-12.929	28.248	GRS22B	-12.929	28.248	BM30B	-16.624	26.974	V43B	-17.853	25.802
GR28	-12.944	28.325	GRS23	-12.929	28.248	BM31A	-16.624	26.974	V44A	-17.853	25.802
GR29	-12.944	28.325	GRS24A	-12.929	28.248	BM31B	-16.624	26.974	V44B	-17.853	25.802
GR29b	-12.944	28.325	GRS25A	-12.929	28.248	BM32	-16.624	26.977	V45	-17.853	25.802
GR30	-12.944	28.325	GRS26	-12.929	28.248	BM33	-16.625	26.977	V46	-17.853	25.802
GR31	-12.944	28.325	GRS29	-12.944	28.325	BM34	-16.625	26.977	V47	-17.853	25.802
GR32	-12.945	28.324	GRS229	-12.944	28.325	BM35	-16.625	26.977	V48	-17.853	25.802
GR33	-12.945	28.323	GRS30	-12.945	28.324	BM36A	-16.625	26.977	V49	-17.853	25.802
GR33b	-12.945	28.323	GRS31	-12.945	28.324	BM36Ab	-16.625	26.977	V50	-17.853	25.802
GR34	-12.944	28.323	GRS31-2	-12.945	28.324	BM36B	-16.625	26.977	V51	-17.853	25.802
GR35	-12.944	28.323	GRS32	-12.945	28.324	BM36Bb	-16.625	26.977	V51	-17.853	25.802
GR36	-12.944	28.323	GRS33	-12.945	28.323	BM37A	-16.625	26.977	V53	-17.853	25.802
GR37	-12.944	28.322	GRS34	-12.944	28.323	BM37B	-16.625	26.977	V?2	-17.853	25.802
GR38	-12.944	28.322	GRS35	-12.944	28.323	BM37Bb	-16.625	26.977	VS01	-17.900	25.856
GR39	-12.944	28.322	GRS36	-12.944	28.323	BM38A	-16.625	26.977	VS02	-17.899	25.856

GR40	-12.944	28.322	GRS37	-12.944	28.322	BM38Ab	-16.625	26.977	VS03	-17.900	25.856
GR40b	-12.944	28.322	GRS38	-12.944	28.322	BM38B	-16.625	26.977	VS04	-17.899	25.855
GR41	-12.944	28.322	GRS39	-12.944	28.322	BM39	-16.625	26.977	VS05	-17.899	25.855
GR41b	-12.944	28.322	GRS40	-12.944	28.322	BM40A	-16.626	26.977	VS06	-17.899	25.855
GR42	-12.944	28.322	GRS41	-12.944	28.322	BM40B	-16.626	26.977	VS07	-17.898	25.853
GR43	-12.921	28.259	GRS42	-12.944	28.322	BM41	-16.626	26.977	VS08	-17.899	25.853
GR43 b	-12.921	28.259	GRA42	-12.944	28.322	BM41B	-16.626	26.977	VS08b	-17.899	25.853
GR44	-12.920	28.259	GRS43	-12.921	28.259	BM42	-16.626	26.977	VS09	-17.899	25.853
GR45	-12.919	28.259	GRS44	-12.920	28.259	BM43A	-16.626	26.977	VS10	-17.898	25.850
GR45b	-12.919	28.259	GRS45	-12.919	28.259	BM43B	-16.626	26.977	VS11	-17.898	25.850
GR46	-12.919	28.259	GRS46	-12.919	28.259	BM44	-16.642	27.028	VS13	-17.898	25.850
GR47	-12.919	28.260	GRS47	-12.919	28.260	BM45	-16.642	27.028	VS14	-17.898	25.850
GR47b	-12.919	28.260	GRS48	-12.922	28.256	BM46	-16.641	27.028	VS15	-17.898	25.850
GR48	-12.922	28.256	GRS49	-12.922	28.256	BM47	-16.641	27.028	VS16	-17.898	25.850
GR49	-12.922	28.256	GRS50	-12.922	28.257	BM48	-16.642	27.028	VS17	-17.898	25.850
GR50	-12.922	28.257	GRS51	-12.994	28.240	BM49	-16.642	27.028	VS18	-17.898	25.850
GR50b	-12.922	28.257	GRA51	-12.994	28.240	BM50	-16.642	27.028	VS19	-17.898	25.850
GR50 -9B	-12.922	28.257	GRS52	-12.994	28.240	BM51	-16.642	27.027	VS20	-17.891	25.859
GR51	-12.994	28.240	GRS53	-12.994	28.240	BM52	-16.642	27.028	VS21	-17.891	25.859
GR51 b	-12.994	28.240	GRS54	-12.994	28.241	BM53	-16.642	27.027	VS22	-17.885	25.845
GR52	-12.994	28.240	GRS55	-12.994	28.241	BMS1	-16.651	27.042	VS23	-17.885	25.845
GR52b	-12.994	28.240	GRS56	-12.995	28.241	BMS2	-16.651	27.042	VS24	-17.834	25.767
GR53	-12.994	28.240	GRS57	-12.995	28.241	BMS3	-16.651	27.042	VS24B	-17.834	25.767
GR54	-12.994	28.241	GRS58	-12.995	28.241	BMS4	-16.651	27.042	VS25A	-17.834	25.767
GR54b	-12.994	28.241	GRS59	-12.994	28.241	BMS5	-16.651	27.042	VS25B	-17.834	25.767
GR55	-12.994	28.241	GRS60	-12.994	28.241	BMS6	-16.650	27.042	VS26A	-17.834	25.767
GR56	-12.995	28.241	GRS61	-12.994	28.241	BMS7	-16.650	27.043	VS26B	-17.834	25.767
GR56b	-12.995	28.241	GRS62	-12.994	28.241	BMS8	-16.650	27.043	VS27	-17.833	25.768
GR57	-12.995	28.241	GRS63	-12.993	28.240	BMS9	-16.650	27.043	VS27B	-17.833	25.768
GR57b	-12.995	28.241	GRS64	-12.993	28.240	BMS11	-16.649	27.044	VS28	-17.833	25.768
GR58	-12.995	28.241	GRS65	-12.993	28.240	BMS12	-16.649	27.044	VS28B	-17.833	25.768
GR59	-12.994	28.241	GRS66	-12.993	28.240	BMS13	-16.649	27.044	VS29A	-17.835	25.773
GR60	-12.994	28.241	GRS67	-12.993	28.239	BMS14	-16.649	27.044	VS29B	-17.835	25.773
GR60b	-12.994	28.241	GRS68	-12.993	28.239	BMS15A	-16.648	27.044	VS30A	-17.835	25.773
GR61	-12.994	28.241	GRS69	-12.993	28.240	BMS15B	-16.648	27.044	VS30B	-17.835	25.773
GR62	-12.994	28.241	GRS70	-12.993	28.240	BMS16	-16.648	27.044	VS31A	-17.842	25.786
GR62A	-12.994	28.241	GRS71	-12.955	28.252	BML16	-16.650	27.043	VS31B	-17.842	25.786
GR62B	-12.994	28.241	GRS72	-12.930	28.242	BMs17	-16.650	27.043	VS32A	-17.842	25.786
GR62b	-12.994	28.241	GRS73	-12.930	28.242	BMS18	-16.625	27.027	VS32B	-17.842	25.786
GR63	-12.993	28.240	GRS74	-12.930	28.242	BMS19	-16.619	27.026	VS33A	-17.843	25.785
GR64	-12.993	28.240	GRS75	-12.930	28.242	BMS20A	-16.619	27.026	VS33B	-17.843	25.785
GR65	-12.993	28.240	GRS76	-12.930	28.242	BMS20C	-16.619	27.026	VS34A	-17.843	25.785
GR66	-12.993	28.240	GRS77	-12.930	28.242	BMS21	-16.619	27.026	VS34B	-17.843	25.785
GR67	-12.993	28.239	GRS78	-12.930	28.242	BMS22	-16.619	27.026	VS35A	-17.843	25.785
GR67b	-12.993	28.239	GRS79	-12.929	28.241	BMS23	-16.619	27.025	VS35B	-17.843	25.785
GR68	-12.993	28.239	GRS80	-12.929	28.241	BMS24	-16.619	27.025	VS36A	-17.846	25.793
GR69	-12.993	28.240	GRS81	-12.929	28.241	BMS25	-16.619	27.024	VS36D	-17.846	25.793
GR70	-12.993	28.240	GRS82	-12.929	28.241	BMS26	-16.626	27.027	VS36B	-17.846	25.793
GR70 b	-12.993	28.240	GRS83	-12.929	28.241	BMS27	-16.626	27.027	VS37A	-17.847	25.793
GR71	-12.955	28.252	GRS84	-12.929	28.241	BMS28	-16.627	27.027	VS37B	-17.847	25.793
GR71b	-12.955	28.252	GRS85	-12.929	28.241	BMS29	-16.624	26.974	VS38A	-17.848	25.792
GR72	-12.930	28.242	GRS86	-12.927	28.242	BMS30	-16.624	26.974	VS38B	-17.848	25.792
GR72b	-12.930	28.242	GRS87	-12.926	28.242	BMS31	-16.624	26.974	VS39A	-17.847	25.792
GR73	-12.930	28.242	GRS88	-12.926	28.242	BMS32	-16.624	26.977	VS39B	-17.847	25.792
GR73b	-12.930	28.242	GRS89	-12.926	28.242	BMS33	-16.625	26.977	VS40A	-17.853	25.803
GR74	-12.930	28.242	GRS90	-12.926	28.242	BMS34	-16.625	26.977	VS40B	-17.853	25.803
GR74b	-12.930	28.242	GRS91	-12.927	28.244	BMS35	-16.625	26.977	VS41A	-17.853	25.803
GR75	-12.930	28.242	GRS92	-12.927	28.245	BMS36	-16.625	26.977	VS41B	-17.853	25.803
GR75b	-12.930	28.242	GRS93	-12.927	28.245	BMS37	-16.625	26.977	VS42A	-17.853	25.802
GR76	-12.930	28.242	GRS94	-12.927	28.245	BMS38	-16.625	26.977	VS42B	-17.853	25.802
GR76b	-12.930	28.242	GRS95	-12.927	28.245	BMS39	-16.625	26.977	VS43	-17.853	25.802
GR77	-12.930	28.242	GRS96	-12.927	28.245	BMS40	-16.626	26.977	VS43b	-17.853	25.802
GR77b	-12.930	28.242	GRS97	-12.927	28.246	BMS41	-16.626	26.977	VS44	-17.853	25.802

GR78	-12.930	28.242	GRS98	-12.927	28.246	BMS44	-16.642	27.028	VS45	-17.853	25.802
GR78b	-12.930	28.242	GRS99	-12.929	28.245	BMS45	-16.642	27.028	VF1	Un	Un
GR79	-12.929	28.241	GRS100	-12.929	28.245	BMS46	-16.641	27.028	VF2	Un	Un
GR79b	-12.929	28.241	PIG1 - GR			BMS47	-16.641	27.028	VF3	Un	Un
GR80	-12.929	28.241	BM1	-16.651	27.042	BMS48	-16.642	27.028	VF3A	Un	Un
GR80b	-12.929	28.241	BM1- Dry	-16.651	27.042	BMS49	-16.642	27.028	VF3B	Un	Un
GR81	-12.929	28.241	BM2	-16.651	27.042	BMS50	-16.642	27.028	VF3C	Un	Un
GR81b	-12.929	28.241	BM2- Dry	-16.651	27.042	BMS51	-16.642	27.027	VFS3	Un	Un
GR82	-12.929	28.241	BM3	-16.651	27.042	BMS52	-16.642	27.028	VF4	Un	Un
GR82b	-12.929	28.241	BM3- Dry	-16.651	27.042	BMS53	-16.642	27.027	VF4B	Un	Un
GR83	-12.929	28.241	BM4	-16.651	27.042	V01	-17.900	25.856	VF5	Un	Un
GR83b	-12.929	28.241	BM4b	-16.651	27.042	V02	-17.899	25.856	VFs5B	Un	Un
GR84	-12.929	28.241	BM5	-16.651	27.042	V03	-17.900	25.856	VF6	Un	Un
GR84b	-12.929	28.241	BM5- Dry	-16.651	27.042	V04	-17.899	25.855	VF6B	Un	Un
GR85	-12.929	28.241	BM6	-16.650	27.042	V05	-17.898	25.853	VF7	Un	Un
GR85b	-12.929	28.241	BM6- Dry	-16.650	27.042	V06	-17.899	25.853	VF7B	Un	Un
GR86	-12.927	28.242	BM7	-16.650	27.043	V07	-17.899	25.853	VF8	Un	Un
GR86b	-12.927	28.242	BM7- Dry	-16.650	27.043	V08A	-17.899	25.853	VF8B	Un	Un
GR87	-12.926	28.242	BM8A	-16.650	27.043	V08B	-17.899	25.853	VF9	Un	Un
GR87b	-12.926	28.242	BM8Ab	-16.650	27.043	V08b	-17.899	25.853	VF9B	Un	Un
GR88	-12.926	28.242	BM8B	-16.650	27.043	V09	-17.899	25.853	VF10.	Un	Un
GR88b	-12.926	28.242	BM8Bb	-16.650	27.043	V10	-17.898	25.850	KA1	-15.889	28.198
GR89	-12.926	28.242	BM9A	-16.650	27.043	V10b	-17.898	25.850	KA1S	-15.889	28.198
GR89b	-12.926	28.242	BM9B	-16.650	27.043	V10B	-17.898	25.850	KA2A	-15.843	28.240
GR90	-12.926	28.242	BM9Bb	-16.650	27.043	V11	-17.898	25.850	KA2S	-15.843	28.240
GR90b	-12.926	28.242	BM10	-16.650	27.043	V12	-17.898	25.849	KA2B	-15.843	28.240
GR91	-12.927	28.244	BM10b	-16.650	27.043	V13	-17.898	25.849	KA3	-15.843	28.239
BM03	-16.651	27.042	BM11A	-16.649	27.044	V14A	-17.898	25.850	KA3S	-15.843	28.239
Sample	Latitude	Longitude	Sample	Latitude	Longitude	Sample	Latitude	Longitude	Sample	Latitude	Longitude
KA4	-15.843	28.239	F9	-32.710	26.602	BM04	-16.651	27.042	GR87	-12.926	28.242
KA4b	-15.843	28.239	F10	-32.710	26.602	BM05	-16.651	27.042	GR88	-12.926	28.242
KA5	-15.843	28.239	F11	-32.710	26.602	BM06	-16.650	27.042	GR89	-12.926	28.242
KA5A	-15.843	28.239	F12	-32.710	26.602	BM07	-16.650	27.043	GR90	-12.926	28.242
KA5B	-15.843	28.239	F13	-32.705	26.589	BM08	-16.650	27.043	GR91	-12.927	28.244
KA6	-15.843	28.239	F14	-32.705	26.589	BM09	-16.650	27.043	GR92	-12.927	28.245
KA6b	-15.843	28.239	F15	-32.704	26.585	BM10	-16.650	27.043	GR93-94	-12.927	28.245
KA7	-15.843	28.239	F16	-32.704	26.585	BM11	-16.649	27.044	GR95	-12.927	28.245
KA7A	-15.843	28.239	F17	-32.704	26.585	BM12	-16.649	27.044	GR96	-12.927	28.245
KA8	-15.843	28.239	F18	-32.704	26.585	BM13	-16.649	27.044	GR97	-12.927	28.246
KA9	-15.844	28.239	F19	-32.704	26.585	BM14	-16.649	27.044	GR98	-12.927	28.246
KA9B	-15.844	28.239	F20	-32.704	26.585	BM15-16	-16.648	27.044	GR99	-12.929	28.245
KA10	-15.844	28.239	F21	-32.704	26.585	BM17	-16.650	27.043	H1	-13.121	28.182
KA10B	-15.844	28.239	F22	-32.704	26.585	BM18	-16.625	27.027	H13	-13.118	28.186
A1	-33.946	18.471	F23	-32.704	26.585	BM19	-16.619	27.026	H17	-13.116	28.187
A2	-33.946	18.471	F24	-32.704	26.585	BM20	-16.619	27.026	H2-	-13.122	28.182
A3	-33.946	18.471	F25	-32.703	26.585	BM21	-16.619	27.026	H24	-13.121	28.186
A4	-33.983	18.585	F26	-32.703	26.585	BM22	-16.619	27.026	H3-	-13.121	28.184
A5	-33.978	18.572	F27	-32.703	26.585	BM23	-16.619	27.025	H44	-12.927	28.245
A6	-33.983	18.585	F28	-32.703	26.585	BM24	-16.619	27.025	H47	-12.928	28.245
A7	-33.978	18.572	F29	-32.648	26.652	BM25	-16.619	27.024	H5-	-13.119	28.186
A8	-33.983	18.585	F30	-32.648	26.652	BM26	-16.626	27.027	H52	-12.926	28.246
A9	-33.978	18.572	F31	-32.648	26.652	BM27	-16.626	27.027	H53	-12.927	28.247
A10	-33.977	18.568	F32	-32.648	26.652	BM28	-16.627	27.027	H55	-12.927	28.247
A11	-33.977	18.568	F33	-32.648	26.652	BM29	-16.624	26.974	H56	-12.928	28.247
A12	-33.977	18.568	F34	-32.648	26.652	BM30	-16.624	26.974	H6-	-13.119	28.186
A13	-33.977	18.568	F35	-32.648	26.652	BM31	-16.625	26.977	H7-	-13.119	28.186
A14	-33.977	18.568	F36	-32.648	26.652	BM32	-16.624	26.977	H8-	-13.119	28.186
A15	-33.977	18.568	F37	-32.650	26.652	BM33	-16.625	26.977	H9-	-13.118	28.186
A16	-33.977	18.568	F38	-32.650	26.652	BM34	-16.625	26.977	I1	-13.631	28.623
A17	-33.976	18.564	F39	-32.650	26.652	BM35	-16.625	26.977	I7	-13.631	28.623
A18	-33.976	18.564	F40	-32.650	26.652	BM36	-16.625	26.977	IMPERIAL	42.881	-0.475
A19	-33.982	18.565	F41	-32.650	26.652	BM37	-16.625	26.977	JUNCTION 459	-11.538	32.130

A20	-33.985	18.564	F42	-32.565	26.743	BM38	-16.625	26.977	K1	-12.452	31.293
A21	-33.992	18.563	F43	-32.565	26.743	BM39	-16.625	26.977	K11PIGEON	-11.171	31.601
A22	-33.992	18.563	F44	-32.565	26.743	BM40	-16.626	26.977	K13	-11.191	31.591
A23SCHOOL	-33.993	18.563	F45	-32.565	26.743	BM44	-16.642	27.028	K14	-11.191	31.591
A24	-33.993	18.563	F46	-32.565	26.743	BM45	-16.642	27.028	K3	-11.167	31.606
A25	-33.993	18.563	F47	-32.578	26.723	BM46	-16.641	27.028	K5	-11.164	31.610
A26	-33.993	18.563	F48	-32.578	26.723	BM48	-16.642	27.028	K6	-11.164	31.611
A27	-33.993	18.563	GOST1	-27.916	17.490	BM49	-16.642	27.028	K7	-11.162	31.611
A28	-33.993	18.563	GOST2	-27.916	17.490	BM51	-16.642	27.027	KA05	-15.843	28.239
A29	-33.993	18.563	GOST3	-27.916	17.490	BM52	-16.642	27.028	KA07	-15.843	28.239
A30	-33.993	18.563	GOST4	-27.916	17.490	BM53	-16.642	27.027	KA09	-15.844	28.239
A31	-33.993	18.563	GOST5	-27.916	17.490	C1	-16.648	27.013	KA10	-15.844	28.239
A32	-33.993	18.563	HELINGER1	-25.889	16.821	C11	-16.639	27.045	KA2	-15.843	28.240
A33	-33.993	18.563	HELINGER2	-25.889	16.821	C2	-16.649	27.039	KA3	-15.843	28.239
A34	-33.993	18.563	HELINGER3	-25.889	16.821	CHAMA ROAD	-12.078	32.279	KAF01	-15.889	28.198
A35	-33.993	18.563	HELINGER4	-25.889	16.821	CHIFUNDA TURN	-11.530	32.440	KANUNSHYA GATE	-12.048	32.216
A36	-33.993	18.563	HELINGER5	-25.889	16.821	CHIKWINDA GATE	-12.520	32.010	LU01-3	51.516	-0.172
A37	-33.993	18.563	HELINGER6	-25.889	16.821	G1	-12.446	31.297	LU04	51.516	-0.172
A38CAMPUS	-33.944	18.466	HELINGER7	-25.889	16.821	G2	-12.446	31.297	LU05-8	-15.448	28.404
A39	-33.944	18.466	HELINGER8	-25.889	16.821	G3	-12.446	31.297	LU09	-15.448	28.405
A40	-33.944	18.466	HELINGER9	-25.889	16.821	G4	-12.446	31.297	LUSAKAPOQUIN	-15.398	28.270
A41	-33.944	18.466	HELINGER10	-25.889	16.821	G5	-12.446	31.297	M1	51.561	-0.162
A42	-33.944	18.466	HELINGER11	-25.889	16.821	G6	-12.446	31.297	M3	51.562	-0.163
AIAIS1	-32.152	18.675	HELINGER12	-25.889	16.821	G7	-12.446	31.297	M5	51.565	-0.169
AIAIS2	-32.152	18.675	LL1	-32.152	18.675	G10	-12.450	31.314	M7	51.565	-0.169
AIAIS3	-32.152	18.675	LL2	-32.152	18.675	G11	-12.454	31.315	M8	51.566	-0.168
AIAIS4	-32.152	18.675	LL3	-32.152	18.675	G12	-12.476	31.316	MANO CAMP	-11.365	32.007
AIAIS5	-32.152	18.675	LL4	-32.152	18.675	G19	-12.475	31.300	MUTINONDO W	-12.455	31.291
AIAIS6	-32.152	18.675	LL5	-32.152	18.675	G20	-12.464	31.297	NL1	-11.720	32.126
AIAIS7	-32.152	18.675	LL6	-32.152	18.675	G21	-12.452	31.293	NL10	-11.726	32.134
AIAIS8	-32.152	18.675	LL7	-32.152	18.675	G6-MUTI FALLS	-12.445	31.301	NL11	-11.726	32.134
AIAIS9	-32.152	18.675	LL8	-32.152	18.675	G8	-12.448	31.313	NL13	-11.732	32.142
AIAIS10	-32.152	18.675	LL9	-32.152	18.675	G9	-12.448	31.313	NL2	-11.720	32.126
B1	-33.875	18.511	LL10	-32.152	18.675	GM47	-16.642	27.028	NL21	-11.751	32.173
B3	-33.876	18.511	LL11	-32.152	18.675	GR01	-12.930	28.274	NL23	-11.770	32.208
B4	-33.876	18.510	LL12	-32.152	18.675	GR02	-12.930	28.274	NL3	-11.720	32.126
B5	-33.878	18.508	LL13	-32.152	18.675	GR03	-12.929	28.274	NL32	-11.881	32.197
B6	-33.878	18.508	LL14	-32.152	18.675	GR04	-12.930	28.274	NL38	-11.753	32.306
B7	-33.878	18.508	LL15	-32.152	18.675	GR06	-12.929	28.273	NL4	-11.719	32.126
B8	-33.878	18.508	LL16	-32.152	18.675	GR07	-12.929	28.273	NL43	-11.766	32.338
B9	-33.872	18.510	LL17	-32.152	18.675	GR08	-12.928	28.273	NL46	-11.771	32.344
B10	-33.872	18.510	LM1	-32.152	18.675	GR09	-12.928	28.273	NL49	-11.856	32.433
B11EUCA	-33.832	18.526	LM2	-32.152	18.675	GR10	-12.928	28.274	NL5	-11.719	32.126
B12	-33.832	18.526	LM3	-32.152	18.675	GR100	-12.929	28.244	NL55	-11.855	32.432
B13	-33.832	18.526	LM4	-32.152	18.675	GR106	-14.490	28.405	NL6	-11.719	32.126
B14	-33.832	18.526	LM5	-32.152	18.675	GR11	-12.934	28.260	NL7	-11.719	32.126
B15	-33.832	18.526	LO1	-27.916	17.490	GR12	-12.934	28.260	NL8	-11.721	32.130
B16	-33.832	18.526	LP1	-24.843	16.976	GR13	-12.934	28.260	NL9	-11.721	32.130
B17	-33.832	18.526	LQ1	-24.627	17.961	GR14	-12.933	28.260	O1	-11.861	32.433
B18	-33.832	18.526	MATPLACE1	-33.946	18.470	GR15	-12.933	28.260	O13	-11.887	32.454
B19	-33.820	18.539	MATPLACE2	-33.946	18.470	GR16	-12.934	28.260	O21	-11.974	32.514
B20	-33.823	18.540	MATPLACE3	-33.946	18.470	GR17	-12.933	28.259	O38	-12.184	32.484
B21	-33.823	18.540	MATPLACE4	-33.946	18.470	GR18	-12.933	28.259	O4-	-11.868	32.436
B22	-33.823	18.540	MATPLACE5	-33.946	18.470	GR19	-12.933	28.259	O51	-12.287	32.268
B23	-33.824	18.540	MATPLACE6	-33.946	18.470	GR20	-12.933	28.259	P04	-12.665	32.125
CLINIC01	-33.946	18.471	NB1	-25.769	19.995	GR21	-12.930	28.249	P1	-12.540	32.147
CRASH1	-25.940	16.772	NB2	-25.769	19.995	GR22	-12.929	28.248	P12	-12.887	32.004

CRASH2	-25.940	16.772	NB3	-25.769	19.995	GR23	-12.929	28.248	P15	-13.037	31.918
CRASH3	-25.940	16.772	NB4	-25.769	19.995	GR24	-12.929	28.248	P2-	-12.603	32.135
CRASH4	-25.940	16.772	NB5	-25.769	19.995	GR25	-12.929	28.248	P35	-14.252	31.679
E1	-33.969	18.594	NB6	-25.769	19.995	GR26	-12.929	28.248	P45	-14.920	30.323
E2	-33.969	18.594	NB7	-25.769	19.995	GR27	-12.929	28.248	TO BUF-FALO	-11.459	32.126
E3	-33.969	18.594	NB8	-25.769	19.995	GR28	-12.944	28.325	TO NLNP	-11.260	31.443
E4	-32.778	26.628	NB9	-25.769	19.995	GR29	-12.944	28.325	V01	-17.900	25.856
E5	-32.778	26.628	NB10	-25.769	19.995	GR30	-12.944	28.325	V02	-17.899	25.856
E6	-32.778	26.628	SP1	-21.834	15.189	GR31	-12.944	28.325	V03	-17.900	25.856
E7	-32.779	26.627	SP2	-21.834	15.190	GR32	-12.945	28.324	V04	-17.899	25.855
E8	-32.779	26.627	SP3	-21.834	15.190	GR33	-12.945	28.323	V07	-17.898	25.853
E9	-32.779	26.627	SP4	-21.834	15.190	GR34	-12.944	28.323	V08	-17.899	25.853
E10	-32.779	26.627	SP5	-21.834	15.190	GR35	-12.944	28.323	V10	-17.898	25.850
E11	-32.779	26.626	SP6	-21.834	15.190	GR36	-12.944	28.323	V11	-17.898	25.850
E12	-32.779	26.626	SP7	-21.834	15.190	GR37	-12.944	28.322	V12	-17.898	25.849
E13	-32.779	26.626	SP8	-21.834	15.190	GR38	-12.944	28.322	V13	-17.898	25.849
E14	-32.779	26.626	SP9	-21.834	15.190	GR39	-12.944	28.322	V14	-17.898	25.850
E15	-32.779	26.626	SP10	-21.834	15.190	GR40	-12.944	28.322	V15	-17.898	25.851
E16	-32.779	26.626	SP11	-21.834	15.190	GR41	-12.944	28.322	V16	-17.898	25.850
E17	-32.781	26.622	SP12	-21.834	15.190	GR42	-12.944	28.322	V17	-17.898	25.852
E18	-32.782	26.622	SP13	-21.834	15.190	GR43	-12.921	28.259	V18	-17.892	25.858
E19	-32.782	26.622	SP14	-21.834	15.190	GR44	-12.920	28.259	V19	-17.892	25.858
E20	-32.782	26.622	SP15	-21.834	15.190	GR45	-12.919	28.259	V20	-17.891	25.859
E21	-32.781	26.625	SP16	-21.834	15.190	GR46	-12.919	28.259	V21	-17.891	25.859
E22	-32.781	26.625	SP17	-21.834	15.190	GR47	-12.919	28.260	V22	-17.885	25.845
E23	-32.780	26.628	SP18	-21.834	15.190	GR48	-12.922	28.256	V23	-17.886	25.845
E24	-32.780	26.628	SP19	-21.834	15.190	GR49	-12.922	28.256	V24	-17.834	25.767
E25	-32.780	26.628	SPR1	-26.022	20.405	GR50	-12.922	28.257	V25	-17.834	25.767
E26	-32.780	26.628	SSS1	-29.673	17.900	GR51	-12.994	28.240	V27	-17.833	25.768
E27	-32.779	26.630	SSS2	-29.673	17.900	GR52	-12.994	28.240	V28	-17.833	25.768
E28	-32.779	26.630	SSS3	-29.673	17.900	GR53	-12.994	28.240	V29	-17.835	25.773
E29	-32.779	26.630	SSS4	-29.673	17.900	GR54	-12.994	28.241	V30	-17.835	25.773
E30	-32.778	26.631	SSS5	-29.673	17.900	GR55	-12.994	28.241	V31	-17.842	25.786
E31	-32.778	26.631	SSS6	-29.673	17.900	GR56	-12.995	28.241	V32	-17.842	25.786
E32	-32.778	26.631	SSS7	-29.673	17.900	GR58	-12.995	28.241	V33	-17.843	25.785
E33	-32.778	26.631	SSS8	-29.673	17.900	GR59	-12.994	28.241	V34	-17.843	25.785
E34	-32.778	26.631	SSS9	-29.673	17.900	GR60	-12.994	28.241	V35	-17.843	25.785
E35	-32.777	26.631	SSS10	-29.673	17.900	GR61	-12.994	28.241	V36	-17.846	25.793
E36	-32.777	26.631	SSS11	-29.673	17.900	GR62	-12.994	28.241	V37	-17.847	25.793
E37	-32.777	26.631	TRANS1	-26.022	20.405	GR63	-12.993	28.240	V38	-17.848	25.792
E38	-32.777	26.635	TRANS2	-26.022	20.405	GR64	-12.993	28.240	V39	-17.847	25.792
E39	-32.777	26.635	Z1	-33.946	18.470	GR65-66	-12.993	28.240	V40	-17.853	25.803
E40	-32.777	26.635	Z2	-33.946	18.470	GR67	-12.993	28.239	V41	-17.853	25.803
E41	-32.776	26.636	Z3	-33.946	18.470	GR69	-12.993	28.240	V42	-17.853	25.802
E42	-32.776	26.636	Z4	-33.946	18.470	GR70	-12.993	28.240	V43	-17.853	25.802
E43	-32.776	26.636	Z5	-33.946	18.470	GR71	-12.955	28.252	V44	-17.853	25.802
E44	-32.776	26.636	Z6	-33.947	18.469	GR72	-12.930	28.242	V45	-17.853	25.802
E45	-32.776	26.636	Z7	-33.947	18.469	GR73	-12.930	28.242	VS26	-17.834	25.767
E46	-32.776	26.636	Z8	-33.947	18.469	GR74	-12.930	28.242	W1	-17.834	25.768
E47	-32.776	26.636	Fish River1	-27.043	17.948	GR75	-12.930	28.242	W13	-17.842	25.785
E48	-32.776	26.636	Fish River2	-27.043	17.948	GR76	-12.930	28.242	W18	-17.846	25.791
E49	-32.776	26.636	Fish River3	-27.043	17.948	GR77	-12.930	28.242	W2	-17.833	25.768
E50	-32.776	26.636	Fish River4	-27.043	17.948	GR78	-12.930	28.242	W20	-17.892	25.858
F1	-32.712	26.594	Fish River5	-27.043	17.948	GR79	-12.929	28.241	W33	-17.895	25.857
F2	-32.712	26.594	Fish River6	-27.043	17.948	GR80	-12.929	28.241	W36	-17.896	25.857
F3	-32.712	26.594	Fish River7	-27.043	17.948	GR81	-12.929	28.241	W40	-17.883	25.858

F4	-32.712	26.594	Fish River8	-27.043	17.948	GR82	-12.929	28.241	W41	-17.883	25.858
F5	-32.712	26.594	Fish River9	-27.043	17.948	GR83	-12.929	28.241	W6	-17.834	25.770
F6	-32.712	26.594	Fish River10	-27.043	17.948	GR84	-12.929	28.241	W8	-17.834	25.771
F7	-32.712	26.594	BM01	-16.651	27.042	GR85	-12.929	28.241	Y4	-15.508	28.269
F8	-32.710	26.602	BM02	-16.651	27.042	GR86	-12.927	28.242			

Table S31: Samples Location