

https://dspace.org/

Genome assembly and comparative genome analysis of the basidiomycetous yeast Naganishia randhawae isolated from avian guano in South Africa

Tshisekedi, Kalonji Abondance

2021
https://hdl.handle.net/10539/33082

Downloaded from a DSpace repository. https://dspace.org/



Genome assembly and comparative genome analysis of the basidiomycetous yeast *Naganishia randhawae* isolated from avian guano in South Africa.

by

Kalonji Abondance Tshisekedi

(Student number: 1423566)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

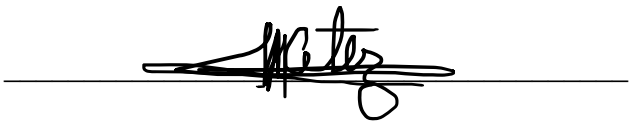
Supervisor: Dr Angela Botes

Co-supervisor: Dr Pieter De Maayer

October 2021

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



(Signature of candidate)

_____08th_____day of__October_2021 at the University of the Witwatersrand, Johannesburg

Abstract

The genus *Naganishia* was recently re-established with the aim of addressing the diversity and heterogeneity of the yeast genus *Cryptococcus*. Many members of this genus have been isolated from harsh terrestrial conditions, including the dry cold valleys of Antarctica and the Atacama Desert. As a result, they have developed adaptive strategies to deal with the stressors associated with cold environments, such as desiccation, reduced availability of water, low availability of nutrients, high levels of ultraviolet (UV) radiation, high or low pH, and high osmotic pressure. Clinical manifestations of *Naganishia* species often include cutaneous lesions, encephalitis, keratitis, onychomycosis, and pneumonia. Members of this genus also have significant biotechnological value due to their capability to produce single cell oils (SCOs) and cold-active enzymes, which can be used in a variety of biotechnological and industrial applications that require low operating temperatures. Despite their potential biotechnological, environmental, and medical importance, the mechanisms underlying the diverse functions of *Naganishia* species remains poorly understood. In this study, we present the first genome sequence of *N. randhawae* eABCC1, which was isolated from avian guano in South Africa. The genome of this strain was further explored using comparative genomic and phylogenomic analysis.

The relevant literature on the taxonomy, ecology, and pathogenicity of *Naganishia* spp. is reviewed in **Chapter 1**. A comprehensive taxonomic history of the genus *Naganishia* is followed by an overview of the different ecological niches, with an emphasis on extreme ecosystems. Finally, biotechnological importance, clinical manifestations and factors that contribute to *Naganishia* pathogenicity are addressed.

In **Chapter 2**, sequencing of *N. randhawae* eABCC1 was performed using the Illumina NovaSeq 6000 platform (paired end read approach 2 X 250 bp) at MR DNA (Texas, USA). Subsequently, *de novo* assembly was undertaken, yielding a draft genome assembly comprised of 386 contigs, which was then structurally and functionally annotated. Two putative laccases were identified in the genomic content of *N. randhawae* eABCC1, which may be related to the development of melanin observed on bird seed agar (BSA). Furthermore, several putative pathogenicity determinants were identified in *N. randhawae* eABCC1.

In **Chapter 3**, a comparative genome analysis of *N. randhawae* eABCC1 with *N. albida* strains JCM2334 and NT2002, as well as *N. vishniacii* ANT03-052, is presented. The pan-genome and its core and accessory elements were analysed at the protein level. These pan-genome fractions were further functionally characterised and categorised according to their Conserved

Orthologous Groups (COGs). Finally, the annotated genomes of the four *Naganishia* strains were screened for metabolic pathways and genes involved in general stress response. Comparative genome analysis of the *Naganishia* pan genome revealed adaptation strategies hinting towards a role in multiple stress resistance.

In **Chapter 4**, the phylogenomic analysis of 161 taxa within the class Tremellomycetes was performed. Single copy orthologues (SCOs) conserved among all the sampled taxa were identified, aligned and concatenated, before the generation of a maximum likelihood (ML) phylogeny. Taxonomic considerations included the proposal of a new order for the family *Phaeotremellaceae* (Phaeotremellaceales **ord. nov.**) and two new species combinations: *Apiotrichum oleaginum* **comb. nov.** and *Naganishia frigoris* **comb. nov.** Overall, findings from this study could pave the way for future phylogenomic studies of fungi at higher taxonomic ranks.

Research outputs

Journal Article

Tshisekedi KA, Habibu A., De Maayer P, **Botes A** (2021). Draft Genome Sequence of the basidiomycetous yeast *Naganishia randhawae* eABCC1 isolated from avian guano in South Africa. *IMA Fungus.* Article submitted on the 14th of Aug 2020 Submission id: IMAF-D-20-00051. (Accepted, 2021).

Conference outputs – Oral presentations

Tshisekedi KA, Habibu A., De Maayer P, **Botes A** (2021). Comparative genome analysis of *Naganishia* yeasts identifies several proteins associated with stress tolerance. South African Society for Microbiology (SASM) Virtual conference. 4th to 6th May 2021.

Conference outputs – Poster presentations.

1. Tshisekedi KA, Habibu A., De Maayer P, **Botes A** (2020). Comparative genomic analysis hints at psychrophilic nature of the novel yeast species *Naganishia randhawae*. FEMS Online Conference on Microbiology 2020, 28th -31st October 2020
2. Tshisekedi KA, Habibu A., De Maayer P, **Botes A** (2019). Draft Genome Sequence of the basidiomycetous yeast *Naganishia randhawae* eABCC1 isolated from avian guano in South Africa. MBRT, Postgraduate Research Day, University of the Witwatersrand, Johannesburg, South Africa, 28th November 2019
3. Tshisekedi KA, **Botes A** (2018). Isolation and identification of pathogenic Cryptococci from environmental samples collected in South Africa. MBRT, Postgraduate Research Day, University of the Witwatersrand, Johannesburg, South Africa, 29th November 2018

I dedicate this work to my sister Jouissance Kalonji, none of this would have been possible without your constant support.

Acknowledgements

I would like to express my gratitude to God for blessing me with this opportunity.

I am indebted to my supervisor and co-supervisor, Dr Angela Botes and Dr Pieter De Maayer. This work would not have been possible without the relentless encouragement, guidance, and assistance. Your wealth of knowledge and constructive feedback has been extremely beneficial to me. I am deeply grateful that you accepted me as a student and have continued to believe in me through the years.

I would like to express my heartfelt gratitude to Dr. Habibu. Thank you for providing computational resources as well as your valuable time.

Thank you to the University of the Witwatersrand for the Postgraduate Merit Award, which enabled me to complete this project.

Members of The Botes Lab. Thank you so much for all your assistance.

To all my friends, particularly Bronwyn Mol and Ruvesh Pillay, thank you for making Wits a home for me and inspiring me to work hard.

Lastly, my family deserves endless gratitude. To my father, I'm glad that your passion for science has rubbed off on me, I'm thankful that this day has arrived, and you are here with us. To my mother thank you for your love and prayers. To my siblings every single one of you bring light into my life.

Table of contents

Chapter 1 The genus *Naganishia*: from polyextremophiles to opportunistic pathogens. 15
1.1 Introduction 16
1.2 Taxonomic history of the genus *Naganishia* 16
1.3 Ecology of the genus *Naganishia* 17
1.4 Metabolic ability of *Naganishia* species 19
1.5 Biotechnological potential of *Naganishia* species 20
1.6 Pathogenicity of the genus *Naganishia* 21
1.7 Genomic studies in the genus *Naganishia* 23
1.8 *Naganishia randhawae* 24
1.9 Impact statement 25
1.10 Aims and objectives 26
1.11 References 27
Chapter 2 Sequencing and assembly of the genome of *Naganishia randhawae* eABCC1 39
2.1 Abstract 40
2.2 Introduction 40
2.3 Materials and methods 42
2.3.1 Sample processing and yeast isolation 42
2.3.2 Genomic DNA extraction 42
2.3.3 Strain identification 42
2.3.4 Phylogenetic analysis 43
2.3.5 Genome sequencing and assembly 44
2.3.6 Gene prediction and annotation 47
2.3.7 Pathogenicity genes in the genome of *Naganishia randhawae* eABCC1 48
2.4 Results and discussion 48
2.4.1 Motivation for genome sequencing and assembly 48
2.4.2 Taxonomic placement of *Naganishia randhawae* eABCC1 49
2.4.3 *Naganishia randhawae* eABCC1 genome metrics 50
2.4.4 Genome annotation of *Naganishia randhawae* eABCC1 51
2.4.5 Pathogen-Host Interaction (PHI) analysis of *Naganishia randhawae* eABCC1 53
2.5 Conclusion 57
2.6 References 58
Chapter 3 Comparative genomic analysis of *Naganishia* reveals a range of molecular features involved in stress tolerance and adaptation 68
3.1 Abstract 69

3.2 Introduction 69
3.3 Materials and methods 71
3.3.1 Genome sequences, structural and functional annotation 71
3.3.2 Pan-genome analysis 71
3.3.3 Analysis of orthologous gene clusters 71
3.3.4 Functional analysis of predicted coding genes using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) 73
3.4 Results and discussion 73
3.4.1 General genome metrics 73
3.4.2 Pan-genome analysis of four *Naganishia* strains 74
3.4.3 Clusters of Orthologous (COG) analysis of the core and pan-genome using eggNOG 76
3.4.4 Orthovenn2 analysis identifies a gene ontological cluster with UV-damage excision repair function in the genome of four *Naganishia* strains 77
3.4.5 Survey of proteins involved in stress tolerance in the genome of four *Naganishia* strains 86
3.4.6 Survey of polysaccharide capsule proteins in the genome of four *Naganishia* strains 89
3.4.7 Trehalose metabolism 89
3.5 Conclusion 91
3.6 References 92
Chapter 4 Re-evaluating the taxonomic status of the fungal class Tremellomycetes: A phylogenomic perspective 101
4.1 Abstract 102
4.2 Introduction 102
4.3 Materials and methods 104
4.3.1 Genome selection and Structural annotation 104
4.3.2 Single Copy Orthologues selection 105
4.3.3 Single Copy Orthologues extraction, quality check and phylogenomic tree construction 105
4.4 Results and discussion 106
4.4.1 Genomic characteristics of the Tremellomycetes 106
4.4.2 General tree topology of the Tremellomycetes 107
4.4.3 Phylogenomic analyses of the genus *Cryptococcus* reveals poor delineation between strains 120
4.4.4 Members of the family *Phaeotremellaceae* do not cluster within the Tremellales 121

4.4.5 Phylogenomic analysis of the order Trichosporonales is congruent with a previous phylogenomic study 122
4.4.6 Phylogenomic analysis places *Cryptococcus* sp. Mo29 in the genus *Naganishia* 123
4.5 Conclusion 123
4.6 References 124
Chapter 5 Summary 130
Appendix 133

List of figures

Figure 2. 1 *Naganishia randhawae* eABCC1 producing a melanin-like brown pigment 49

Figure 2. 2 Maximum Likelihood (ML) phylogeny showing the taxonomic placement of *Naganishia randhawae* eABCC1. 50

Figure 2. 3 Distribution of Clusters of Orthologous Groups. 52

Figure 3. 1 The core, accessory, and unique proteins of the compared *Naganishia* strains……. 76

Figure 3. 2 Proportion of proteins in the genome of four *Naganishia* genomes and their fractions assigned to Clusters of Orthologous Groups. 78

Figure 3. 3 Proportion of proteins in the unique fraction of each *Naganishia* genomes and their Clusters of Orthologous Groups. 79

Figure 3. 4 Venn diagram showing the distribution in the four *Naganishia* genomes using Orthovenn2 (Xu *et al.* 2019). 81

Figure 4. 1 Pie chart showing the environmental sources of 161 Tremellomycetes taxa analyzed in this study.
……..107

Figure 4. 2 Phylogenomic analysis of 161members of the class Tremellomycetes.. 120

List of tables

Table 1. 1 List of Tremellomycetes within the genus *Naganishia*. 18

Table 2. 1 List of fungal strains used in this study……..44

Table 2. 2 Whole-genome assembly metrics of *Naganishia randhawae* eABCC1.....51

Table 2. 3Orthologues identified in *Naganishia randhawae* eABCC1 using the Pathogen-Host Interaction (PHI) database (Winnenburg *et al.*, 2008). 54

Table 3. 1 Genome content and metadata of the strains analysed in this study……...75

Table 3. 2 Proteins involved in carotenoid and mycosporine synthesis 83

Table 3. 3 Putative laccases identified in the four *Naganishia* genomes 82

Table 3. 4 Proteins involved in heat shock response in four *Naganishia* strains.....87

Table 3. 5 Putative copper amine oxidase found in the genome of *Naganishia randhawae* eABCC1 89

Table 3. 6 Proteins involved in capsule synthesis in the genome of four *Naganishia* strains .90

Table 3. 7 Proteins involved in trehalose metabolism in the genome of four *Naganishia* strains 90

Table 4. 1 Genomic features of 161 Tremellomycetes and outgroup strains……...108

List of abbreviations

6–4PPs: Pyrimidine-pyrimidone 6–4 photoproducts

ACT: Actin

AFTOL: Assembling the Fungal Tree of Life

AIDS: acquired immunodeficiency syndrome

AMB: Amphotericin B

ATP: Adenosine triphosphate

BSA: Bird seed agar

BUSCO: Benchmarking Universal Single-Copy Orthologs

CAOs: Copper amine oxidases

CAZymes: Carbohydrate-Active enZymes

CNS: Central nervous system

COG: Cluster of orthologous groups

CoQ: Coenzyme Q

CPDs: Cyclobutane pyrimidine dimers

CYTB: Cytochrome b

DHN: 1,8-dihydroxynaphthalene

DOE: Department of energy

DOPA: dihydroxy-phenolic acid

FGI: Fungal Genome Initiative

GO: Gene Ontology

HGT: Horizontal gene transfer

HIV: Human immunodeficiency virus infection

HOG: High osmolarity glycerol

HSP40: Heat shock protein 40

HSPs: Heat shock proteins

ITS: Internal transcribed spacer regions

JGI: Joint Genome Institute

KEGG: Kyoto Encyclopaedia of Genes and Genomes

L-dopa: L-3,4-dihydroxyphenylalanine

LSU: Large subunit

m.a.s.l: Meter above sea level

MAPK: Mitogen-activated protein kinase

MCO: Multicopper oxidase

MGG: Mycosporine– glutaminol glucoside

ML: Maximum likelihood

MLSA: Multilocus analysis

NCBI: National Center for Biotechnology Information

NGS: Next-generation sequencing

NJ: Neighbour Joining

NTH1: Degradation of neutral trehalase

OGRI: Overall genome relatedness indices

ORFs: open reading frames

PCR: polymerase chain reaction

PHI-base: pathogen-host interactions database

PP: posterior probabilities

rDNA: ribosomal deoxyribonucleic acid

RNA: ribonucleic acid

RPB: RNA polymerase subunits

SCJ1: DnaJ- protein

SCOs: single copy orthologues

SSU: small subunit

TEF1: translation elongation factor 1-alpha

TFs: transcription factors

TPS1: trehalose-6-phosphate synthase

TPS2: trehalose-6-phosphate phosphatase

tRNAs: transfer RNAs

USA: united states of America

UV: ultraviolet

UV-A: ultraviolet A

UV-B: ultraviolet B

UV-C: ultraviolet C

YPD: yeast peptone dextrose

Chapter 1 The genus *Naganishia*: from polyextremophiles to opportunistic pathogens.

1.1 Introduction

Naganishia species are basidiomycetous yeasts that belong to the class Tremellomycetes, order Filobasidiales, family *Filobasidiaceae* (Kurtzman *et al.,* 2011;Liu *et al.,* 2015a). Members of this genus can be found in a variety of environments, including air, arid soil, bird droppings, cheese, decomposing wood, glaciers, and hypersaline waters (Fell *et al.,* 1999; Schmidt *et al.,* 2017; Scorzetti *et al.,* 2000). Several members of the genus *Naganishia* have been linked to infections such as cutaneous lesions, encephalitis, keratitis, onychomycosis, and pneumonia in immunocompromised individuals (Khawcharoenporn *et al.,* 2007).

Naganishia species have an elongated ovoid to spherical cell morphology (Khan *et al.,* 2010; Kurtzman *et al.,* 2011). When cultured on solid media, members of the genus *Naganishia* have a creamy white appearance (Fonseca *et al.,* 1901; Ikeda *et al.,* 2003; Kurtzman *et al.,* 2011). Some species may develop brown colonies when cultured on unique media, such as Niger seed agar, due to the production of the pigment melanin (Yarrow, 1998). Members of the genus reproduce by budding, and sexual reproduction has yet to be observed (Kurtzman *et al.,* 2011; Liu *et al.,* 2015a). From a metabolic perspective, *Naganishia* species produce starch-like compounds and utilize caffeic, ferulic, hydroxybenzoic, L-malic, p-coumaric, protocatechuic and vanillic acid (Fotedar *et al.,* 2018; Liu *et al.,* 2015a). Nitrate is utilized and fermentation has not been observed (Liu *et al.,* 2015a). Their major Coenzyme Q (CoQ) system is CoQ-9 or CoQ-10 (Liu *et al.* 2015a).

1.2 Taxonomic history of the genus *Naganishia*

The genus *Naganishia* has a long taxonomic history that dates back more than 50 years. Goto (1963) proposed the genus in 1963 to accommodate the yeast *Naganishia globosus*. Subsequently this species was reassigned to its synonym *Cryptococcus saitoi,* based on ribosomal deoxyribonucleic acid (rDNA) sequence analysis, resulting in the extinction of the genus (Fonseca *et al.,* 2000).

The majority of the species in the *Naganishia* genus originate from the *C. albidus* clade. The extreme polyphyletic nature of the genus *Cryptococcus* has long presented a challenge to taxonomists (Boekhout *et al.,* 2011; Fell *et al.,* 1999; Scorzetti *et al.,* 2002). Liu and colleagues (2015a, 2015b) proposed a revision of the Tremellomycetes taxonomic framework, which

included the genus *Cryptococcus*. This taxonomic revision was proposed based on a phylogeny that included seven genes: the small rDNA subunit, the large rDNA subunit's D1/D2 domains, the internal transcribed spacer regions (ITS 1 and 2) of rDNA, including 5.8S rDNA, two RNA polymerase II subunits (RPB1 and RPB2), the translation elongation factor 1-α (TEF1), and the mitochondrial gene cytochrome b (CYTB) (Liu *et al.,* 2015b). Consequently, members of the *albidus* clade were assigned to the genus *Naganishia*, and three new species were proposed (Liu *et al.,* 2015a). Recently, two additional novel species have been proposed in the genus namely¸ *N. floricola sp. nov.* (Zhou *et al.,* 2020) and *N. qatarensis sp. nov.* (Fotedar *et al.,* 2018). As such, the genus *Naganishia* is currently comprised of seventeen distinct species (Table 1.1).

1.3 Ecology of the genus *Naganishia*

Members of the genus *Naganishia* have been isolated from a wide range of terrestrial habitats (Costello *et al.,* 2009; Lynch *et al.,* 2012; Pulschen *et al.,* 2015; Schmidt *et al.,* 2017; Solon *et al.,* 2018). *Naganishia albida*, for example, is a pervasive species in terms of geography as well as isolation source. It has been isolated from a variety of settings, including food, clinical, industrial, and cold environments (Kurtzman *et al.,* 2011). *Naganishia cerealis* was isolated from fermented cereals (Passoth *et al.,* 2009), while *N. adeliensis, N. diffluens,* and *N. randhawae* are usually isolated from a variety of plant niches (Kamari *et al.,* 2017; Khan *et al.,* 2010; Setati *et al.,* 2012). Several species including, *Naganishia albida, N. albidosimilis, N. adeliensis, N. friedmannii, N. liquefaciens, N. randhawae,* and *N. uzbekistanensis* have all been isolated from bird droppings (Brito *et al.,* 2019; Rosario *et al.,* 2005; Syakalima *et al.,* 2019; Wu *et al.,* 2012). Birds play a critical role in the spread of microorganisms such as yeasts (Moschetti *et al.,* 2017). By propagating fungal cells from cloacae to the environment, wild birds, especially pigeons, and their excreta serve as a reservoir for pathogenic yeasts such as the well-known yeast pathogen *Cryptococcus neoformans* (Rosario *et al.,* 2005). The factors that favour the survival of yeasts in bird droppings are not well understood.

Table 1.1 List of Tremellomycetes within the genus *Naganishia*.

Taxon	**Basionym**	**Strain code**	**Origin**
N. adeliensis comb. nov.	*Cryptococcus adeliensis*	CBS 8351	Terre Adelie (Antarctica)
N. albida comb. nov.	*Torula albida*	CBS 142	Air (Japan)
N. albidosimilis comb. nov.	*C. albidosimilis*	CBS 7711	Linnaeus Terrace (Antarctica)
N. antarctica comb. nov.	*C. antarcticus*	CBS 7687	Soil (Antarctica)
N. bhutanensis comb. nov.	*C. bhutanensis*	CBS 6294	Forest (Bhutan)
N. cerealis comb. nov.	*C. cerealis*	CBS 10505	Fermented cereals
N. diffluens comb. nov.	*Torulopsis diffluens*	CBS 160	Diseased fingernail (Austria)
N. floricola sp. nov	-	CGMCC 2.5856	Flowers (China)
N. friedmannii comb. nov.	*C. friedmannii*	CBS 7160	Rock (Antarctica)
N. globosa	*N. globosus*	CBS 5106	Blue cheese (Japan)
N. liquefaciens comb. nov.	*Torulopsis liquefaciens*	CBS 968	Sake-moto (Japan)
N. onofrii comb. nov.	*C. onofrii*	DBVPG 5303	Miage Glacier (Italy)
N. qatarensis sp. nov.	-	QCC/Y17/17	Hypersaline environment (Qatar)
N. randhawae comb. nov.	*C. randhawai*	CBS 10160	Tree trunk (India)
N. uzbekistanensis comb. nov.	*C. uzbekistanensis*	CBS 8683	Soil (Uzbekistan)
N. vaughanmartiniae comb. nov.	*C. vaughanmartiniae*	DBVPG4736	Calderone glacier (Italy)
N. vishniacii comb. nov.	*C. vishniacii*	CBS 7110	Soil (Antarctica)

Naganishia species appear to be adapted for survival in some of the most extreme environments on Earth. They dominate microbial communities at some of the highest elevations on earth (> 6000 metres above sea level, m.a.s.l), representing 92 % of the total operational taxonomic unit observed (Lynch *et al.,* 2012a). *Naganishia qatarensis* has been isolated from the hypersaline marine environment in Qatar (Fotedar *et al.,* 2018). Interestingly, nine members of the genus including, *N. adeliensis* (Scorzetti *et al.,* 2000), *N. albida* (Goto *et al.,* 1969), *N. albidosimilis* (Vishniac and Kurtzman, 1992), *N. antarctica* (Vishniac and Kurtzman, 1992), *N. friedmannii* (Vishniac, 1985), *N. globosa* (Turchetti *et al.,* 2008), *N. liquefaciens* (Fonseca *et al.,* 1901), *N. onofrii* (Turchetti *et al.,* 2015), *N. vaughanmartiniae* (Turchetti *et al.,* 2015) and *N. vishniacii* (Vishniac and Hempfling, 1979) have been isolated repeatedly from Antarctica and other cold environments. The mechanisms by which microorganisms respond to cold stress has received considerable attention over the last few years, especially in view of the perceived biotechnological potential of the psychrophilic enzymes produced by these microorganisms. Cold-shock response in yeast involves changes in plasma membrane fluidity, overexpression of genes involved in the transcriptional and translational machinery, and transcriptional activation of typical stress-marker genes such as encoding various heat-shock protein (HSP) and enzymes involved in trehalose synthesis (D'Amico *et al.,* 2002; Inouye and Phadtare, 2014; Kowalski *et al.,* 1995).

To date, however, little is known about how *Naganishia* spp. cope in cold environments. A comparative genomic analysis of the psychrophilic yeast *N. vishniacii* has found that it includes features within its genome that allow it to survive in extremely cold temperatures (Nizovoy *et al.,* 2020). This includes proteins involved in trehalose synthesis, which has mentioned previously, are critical for cell survival in harsh environments including cold temperatures (Hottiger *et al.,* 1994).

1.4 Metabolic ability of *Naganishia* species

Naganishia members can assimilate a diverse range of carbon and nitrogen sources. Their ability to assimilate compounds found in natural habitats (e.g., mono- and oligosaccharides, such as xylose and cellobiose; glycosylated hydroxyl-phenols, such as salicin and arbutin; and nitrogen compounds, such as nitrates) may imply that their ability to colonise a variety of ecosystems cannot be ignored. Additionally, ecophysiological studies of *Naganishia* isolated from high altitudes and extreme habitats indicate that several of them are metabolically

versatile in terms of the variety of organic compounds they can utilise for growth (Schmidt *et al.* 2017). For instance, a strain of *N. friedmannii* isolated from Volcán Llullaillaco metabolised 74 % of the 27 organic compounds commonly used to classify yeast, whereas a strain isolated from Antarctica metabolised only 30 % of the same 27 organic compounds (Barnett *et al.*, 1990; Vimercati *et al.*, 2016). The metabolic versatility of the Llullaillaco isolate may indicate that it is an opportunistic organism, as defined by Polz colleagues (2006), that is, one capable of "exploiting spatially and temporally variable resources". Numerous substrates used by *Naganishia* species are derived from plants (e.g., cellobiose, arbutin, and salicin), indicating that they can benefit from Aeolian-transported plant material, which is likely the primary source of organic matter in extreme elevation ecosystems (Lynch *et al.* 2012b; Vimercati *et al.* 2016; Schmidt *et al.* 2017).

1.5 Biotechnological potential of *Naganishia* species

Members of the genus *Naganishia* and their associated biomolecules have been analysed for their use in various biotechnological applications. *Naganishia adeliensis* and *N. liquefaciens*, for example, have the ability to accumulate large amounts of lipids (Selvakumar *et al.*, 2019; Selvakumar and Sivashanmugam, 2018). Lipid production in yeast is regarded as an alternate feedstock for biodiesel production, assisting in the battle against climate change and the development of sustainable practices (Luque *et al.*, 2010). The use of yeasts has several advantages over using animal or plant sources, such as not being constrained by climatic conditions or geographical location. Additionally, they often have a shorter processing time and can be used on a wider range of substrates, including industrial waste (Luque *et al.*, 2010; Ward and Singh, 2005).

Temperature is a determinant of biochemical reactions, as postulated by the Arrhenius law, which describes the exponential dependence of reaction rate on temperature (Goldanskii, 1976). As a result, at low temperatures, reaction rates are considerably lowered (Longwell and Weiss, 2002). Psychrophilic microorganisms produce cold-adapted enzymes with high specific activities at low temperatures, often up to an order of magnitude higher than those found in mesophiles (Feller and Gerday, 1997; Georlette *et al.*, 2004). The most widely accepted explanation for this cold adaptation is the activity–stability–flexibility relationship, which implies that psychrophilic enzymes increase their structural flexibility in order to compensate for the freezing effect of cold environments (Johns and Somero, 2004). This increased

flexibility improves the degree of complementarity between the catalytic site and the substrate, lowering activation energies and increasing substrate turnover rates (D'Amico *et al.,* 2006; Rodrigues and Tiedje, 2008). Furthermore, the improved catalytic efficiency and decreased thermal stability properties make them suitable for a wide range of industrial applications (Feller and Gerday, 1997). Unsurprisingly, industry has shown great interest in enzymes produced by psychrotolerant microorganisms (De Maayer *et al.* 2014). At present, this avenue of research is still unexplored when considering members of the genus *Naganishia* capable of growth in cold environments.

1.6 Pathogenicity of the genus *Naganishia*

Our understanding of the evolution of non-pathogenic microbial species towards pathogenicity is continuously expanding. Ecological and evolutionary theory has shed light on important pathogen characteristics, such as plasticity and generalism (Brown *et al.,* 2012). Opportunistic pathogens are regarded as generalists, as they can survive in and adapt to different environments (Brown *et al.,* 2012; Ranjan and Poças-Fonseca, 2019). This has largely been attributed to their dual-use virulence factors, which play a role in both environmental survival and virulence (Casadevall and Pirofski, 1999; Köhler *et al.,* 2017; Steenbergen *et al.,* 2003).

Although there have been cases with immunological competent individuals, the majority of opportunistic yeast infections occur in people whose immune systems are suppressed due to an underlying condition, such as the human immunodeficiency virus (HIV) infection, acquired immunodeficiency syndrome (AIDS) and cancer (Banerjee, 2009). The clinical manifestations range from benign, localised lesions that are either temporary or persistent, to widespread, and occasionally fatal systemic infection (Banerjee, 2009; Heitman *et al.,* 2010; Rodrigues *et al.,* 1999). Since the HIV/AIDS epidemic, there has been an upsurge in opportunistic yeast infections (Cuomo *et al.,* 2018; Farrer and Fisher, 2017; Limper *et al.,* 2017). *Candida* spp., *Cryptococcus* spp., and *Trichosporon* spp. are the most commonly reported pathogens (Gupta *et al.,* 2004; May *et al.,* 2016; Ramos *et al.,* 2004). However, there are increasing case reports of seemingly benign yeast species suchnon-*albicans Candida* spp., non-*neoformans Cryptococcus* spp., and even *Saccharomyces cerevisiae* causing disease in humans (Taei *et al.* 2019; Salazar-Leal *et al.* 2019; Huang *et al.* 2020; Wombwell *et al.* 2021).

Members of the genus *Naganishia* are also frequently associated with clinical cases, especially individuals with impaired cell-mediated immunity or invasive devices (Aghaei Gharehbolagh *et al.,* 2017; Lee *et al.,* 2004a; Ragupathi and Reyna, 2015). *Naganishia albida* is responsible for the majority of infections in the genus, which primarily includes cases of cutaneous lesions, encephalitis, keratitis, and onychomycosis (Sugita *et al.* 2003; Serda Kantarcioğlu *et al.* 2007; Ekhtiari *et al.* 2017; Aghaei Gharehbolagh *et al.* 2017). Severe infections, such as fungemia, central nervous system (CNS) infections, and pneumonia, have also been reported (Choe *et al.,* 2019; Narayan *et al.,* 2000; Ragupathi and Reyna, 2015; Rimek *et al.,* 2004).

In recent years, there have also been reports of infections caused by other members of the genus. The presence of *N. adeliensis* was reported in an HIV patient with meningitis (Rimek *et al.,* 2004). This species has also been isolated from a biopsy of a patient with pulmonary disease and the oral cavity of a patient with acute myeloid leukaemia (Rimek *et al.,* 2004; Tintelnot and Losert, 2005). *Naganishia uzbekistanensis* was isolated from the blood of an acquired immunodeficiency syndrome AIDS patient (Kordossis *et al.,* 1998; Sugita *et al.,* 2001) and also from the bone-marrow of lymphoma patients (Powel *et al.,* 2012). *Naganishia diffluens* is associated with subcutaneous infections (Serda Kantarcioğlu *et al.,* 2007); *Naganishia friedmannii* has been confirmed as an etiologic agent of onychomycosis (Ekhtiari *et al.* 2017); and *N. liquefaceins* was present in skin samples from patients with atopic dermatitis (Sugita *et al.,* 2003).

The specific mechanisms underlying the pathogenicity of *Naganishia* species are poorly understood. However, as with the well-studied yeast pathogen *Cryptococcus neoformans,* the occurrence of a polysaccharide capsule, expression of the laccase enzyme, and melanin production have been identified as putative pathogenicity or virulence factors (Ikeda *et al.,* 2002; Park *et al.,* 2017; Vishniac, 2006). The polysaccharide capsule is known to contribute to the virulence of many yeast pathogens by inhibiting effective phagocytosis and yeast clearance by alveolar macrophages (Bose *et al.,* 2003). Laccase is a melanin-producing enzyme, that forms part of a large group of enzymes known as the multi-copper or blue oxidases (Crowe and Olsson, 2001). Melanin is synthesized from exogenous catecholamines such as dihydroxy-phenolic acid (DOPA) (Williamson, 1997) and has been shown to inhibit phagocytosis and increase resistance to antifungal agents such as amphotericin B (AMB) (Nosanchuk and Casadevall, 2003). These virulence features are known to play a number of roles in yeasts and are generally regarded as being “dual-use” (Nosanchuk and Casadevall, 2003). Given the harsh

conditions in which *Naganishia* species reside, it is likely that they evolved to help *Naganishia* overcome the stresses associated with their natural environment (Nosanchuk and Casadevall, 2003). Having said that, their potential role in virulence should not be overlooked.

The method and duration of treatment for opportunistic fungal infections is determined by the location of the infection, the host immune system's, and the severity of the infection (Bernal-Martinez *et al.,* 2010). Care guidelines for *Naganishia*-related infections are limited at the present time due to a low patient population and a dearth of clinical trial data (Khawcharoenporn *et al.* 2007). Amphotericin B (AMB) is the most commonly prescribed medication for fungal infections (Bernal-Martinez *et al.,* 2010; Park *et al.,* 2017). Itraconazole has been used as a substitute for AMB in patients with leukopenia due to the adverse effects of AMB (Bernal-Martinez *et al.,* 2010). Fluconazole has been shown to be effective in treating skin infections (Choe *et al.,* 2019).

1.7 Genomic studies in the genus *Naganishia*

The genetic content of an organism determines its biological function. Since the publication of the *Haemophilus influenzae* genome sequence in 1995 (Fleischmann *et al.,* 1995), genomic studies have made significant contributions to our understanding of microbial biology (Sharma, 2016). Fungi have a large, diverse genome that contains genes allowing them to thrive in a variety of environments, invade animal and plant cells, and participate in both aquatic and terrestrial nutrient cycling (Stajich, 2017).

The advancement of Next-Generation Sequencing (NGS) technologies, bioinformatic algorithms, and the smaller size of fungal genomes in comparison to other eukaryotes have facilitated fungal genome sequencing and analysis. Comparative genome analysis, for example, has become one of the most effective tools for elucidating factors that control species adaptation and diversification in various ecosystems (Naranjo-Ortiz and Gabaldón, 2019; Vries *et al.,* 2018). Similarly, fungal omics are now widely used to identify fungal proteins with biotechnological applications. The sequencing of fungal genomes provides access to the full complement of fungal genome-encoded proteins. This could be used to model the fundamental abilities of fungal strains and to better understand the factors that contribute to their biotechnological or medical relevance.

The genomes of five *Naganishia* taxa have been deposited in the NCBI database as of April 2021 (Jenuth, 2000), namely *N. albida* (five genomes), *N. adeliensis* (one genome), *N. liquefaciens* (one genome), *N. randhawae* (one genome), and *N. vishniacii* (one genome) (one genome). This represents only 30 % of the genus' total number of taxa. To fully exploit the rich diversity of *Naganishia* on a genome-wide scale, it is necessary to sequence the genomes of more taxa. The genomic information could be useful in understanding the potential biotechnological potential of *Naganishia* members, as well as aspects related to pathogenicity and adaptation to their environmental niches, especially given that many members of the genus are found in extreme environments.

The comparison of genomes of closely related species holds the key to revealing genetic differences underlying different adaptation strategies, and it can serve as the foundation for further research into the evolutionary history (Aliyu *et al.,* 2020a; Maayer *et al.,* 2019) of the species in question. Comparative genomics of Tremellomycetes species has shed light on the mechanisms underlying the pathogenicity of numerous human pathogens and established the biotechnological relevance of numerous strains (Aliyu *et al.,* 2020b; Shen *et al.,* 2016; Takashima *et al.,* 2019). Available genome reports have already shed light on the genomic content of these individual taxa (Shen *et al.,* 2016; Takashima *et al.,* 2019). The genomic analysis of *N. albida* strains ATCC 10672 and *NT2002*, for example, revealed the presence of genes involved in microbial lipid production (Vajpeyi and Chandran, 2016; Yong *et al.,* 2016). The genomic composition of *N. vishniacii* ANT05-03 reveals the presence of genes that promote growth in extremely cold environments (Nizovoy *et al.,* 2020). This finding corroborates previous research indicating that *N. vishniacii* is a niche-specific species, having been isolated from Antarctic soil only (Vishniac, 2006).

Given the importance of Tremellomycetes in biotechnology, ecology, and pathogenesis (Aliyu *et al.,* 2020b; Pakshir *et al.,* 2018; Schmidt *et al.,* 2017), comparative genome analysis across orders and genera is critical in answering important scientific questions about the adaptation and differentiation of Tremellomycetes species.

1.8 *Naganishia randhawae*

Although there is literature available on some of the species within the genus *Naganishia*, particularly *N. albida*, little is known about the relatively novel species, *N. randhawae*. The type strain (CBS 10160) was originally isolated from a tree trunk in India (Khan *et al.,* 2010).

Its classification was based on sequence analysis of the D1/D2 domains of the 26S rRNA gene and the internally transcribed spacer regions ITS1 and ITS2. The sequencing results indicated that the species clusters with *N. albida* and shares the highest sequence identity (> 98 % nucleotide identity) with *N. friedmannii* and *N. cerealis* for analysed taxonomic markers (Khan *et al.,* 2010). Phenotypic characteristics differ with these taxa, in that *N. randhawae* CBS 10160 grows at 37 °C, while *N. friedmannii* and *N. cerealis* do not. Furthermore, unlike *N. friedmannii*, *N. randhawae* CBS 10160 assimilates lactose, raffinose, L-rhamnose and myo-inositol. Based on these differences it was proposed as a novel species within the genus (Khan *et al.,* 2010). Following the work Khan and colleagues (Khan *et al.,* 2010), strains of *N. randhawae* have been isolated in avian guano (Wu *et al.,* 2012) and among the fungal diversity in vineyards in Stellenbosch South Africa (Setati *et al.,* 2012).

Little is currently known about the biology of *N. randhawae*. As a result, there is a need to investigate the properties of this yeast at the genomic level to fill this gap in the literature and identify potential biotechnological applications.

1.9 Impact statement

Birds are important in the spread of microorganisms such as filamentous fungi and yeasts (Baroni *et al.,* 2006; Brilhante *et al.,* 2010; Caicedo *et al.,* 1999; Chryssanthou *et al.,* 2011; Elhariri and Refai, 2015; López-Martínez and Castañón-Olivares, 1995; Lugarini *et al.,* 2008; Mancianti *et al.,* 2002; Marinho *et al.,* 2010; Mendes *et al.,* 2014; Rosario *et al.,* 2010) Due to the high nitrogen content, avian guano in particular harbours a rich diversity of fungi (Elhariri and Refai, 2015; Mendes *et al.,* 2014; Silva and Paula, 1963), including several pathogens from the genera *Aspergillus*, *Candida*, *Cryptococcus*, *Histoplasma*, *Penicillium*, *Rhodotorula* and *Trichosporon* (Elhariri and Refai, 2015; Mendes *et al.,* 2014; Santos *et al.,* 2009; Skillman *et al.,* 2013).

Sub-Saharan Africa is home to two thirds (67 %) of all people living with HIV (UNAIDS, 2018). As such, the reservoir of pathogens found in avian guano poses a great infection risk these immunocompromised individuals. Cryptococcosis alone is responsible for 44 % of all HIV-related deaths in South Africa (GERMS-SA, 2015). More research into the prevalence of pathogenic fungi in the South African environment, as well as the impact this may have on our endemic HIV/AIDS population, would be extremely beneficial. The isolation and identification of yeasts from avian guano in South Africa was done for this purpose.

A yeast (strain eABCC1) was isolated from avian guano in Johannesburg, South Africa, and sequence analysis of the ITS region indicated a nucleotide identity of 99 % with *N. randhawae* CBS 10160 (Khan *et al.*, 2010). However, isolate eABCC1 differed with the type strain in terms of its inability to grow at 37 °C and the production of pigment, likely melanin, when cultured on bird seed agar (Yarrow, 1998). Currently, there is little genomic information to explain or investigate these morphological and physiological differences. In this case, sequencing the entire genome of these species may provide a thorough understanding of the molecular mechanisms underlying these genetic variations.

To date, the genomes of only four species within this genus have been sequenced, namely *N. albida* (six genomes), *N. adeliensis* (one genome), *N. liquefaciens* (one genome) and *N. vishniacii* (one genome) (Vajpeyi and Chandran 2016). The genome sequence of *N. randhawae* will provide valuable genetic information on this species. Furthermore, a comparative analysis of the genome of *N. randhawae* (strain eABCC1) with those of the *Naganishia* species and, on a broader scale, other members of the family *Filobasidiaceae* and the order Filobasidiales will provide deeper insights into the biology of members of the genus, family and order, respectively. The genomic information could be useful in understanding the potential biotechnological potential of members of the genus *Naganishia*, as well as aspects related to their pathogenicity and adaptation to their environmental niches, especially given that many members of the genus are considered psychrophiles or psychrotolerant. Genome-wide phylogenetics (or phylogenomics) have been shown to provide far greater taxonomic resolution than single-gene or multi-gene phylogenies (Kuramae *et al.*, 2006). Thus, phylogenomic approaches will be applied in this study to more resolutely resolve the taxonomy of the class Tremellomycetes.

1.10 Aims and objectives

The aim of this study is to perform a genome assembly and comparative genome analysis of *Naganishia randhawae* eABCC1. This will be accomplished by achieving the following objectives:

1. **Objective 1:** Sequencing, assembly, and annotation of the *N. randhawae* eABCC1 genome
2. **Objective 2:** Comparative genomic analysis of members of the genus *Naganishia*
3. **Objective 3:** Phylogenomic re-evaluation of the class Tremellomycetes

1.11 References

Aliyu, H., Gorte, O., De Maayer, P., Neumann, A. and Ochsenreither, K., 2020. Genomic insights into the lifestyles, functional capacities and oleagenicity of members of the fungal family Trichosporonaceae. *Scientific reports*, *10*(1), pp.1-12.

Aliyu, H., Gorte, O., Zhou, X., Neumann, A. and Ochsenreither, K., 2020. In silico proteomic analysis provides insights into phylogenomics and plant biomass deconstruction potentials of the tremelalles. *Frontiers in bioengineering and biotechnology*, *8*, p.226.

Barnett, J.A., Payne, R.W. and Yarrow, D., 1990. Yeasts: characteristics and identification.

Bernal-Martinez, L., Gomez-Lopez, A., Castelli, M.V., Mesa-Arango, A.C., Zaragoza, O., Rodriguez-Tudela, J.L. and Cuenca-Estrella, M., 2010. Susceptibility profile of clinical isolates of non-*Cryptococcus neoformans/non-Cryptococcus gattii Cryptococcus* species and literature review. *Medical mycology*, *48*(1), pp.90-96.

Boekhout, T., Fonseca, A., Sampaio, J.P., Bandoni, R.J., Fell, J.W. and Kwon-Chung, K.J., 2011. Discussion of teleomorphic and anamorphic basidiomycetous yeasts. In *The Yeasts* (pp. 1339-1372). Elsevier.

Böer, E., Steinborn, G., Florschütz, K., Körner, M., Gellissen, G. and Kunze, G., 2009. Arxula adeninivorans (Blastobotrys adeninivorans)—a dimorphic yeast of great biotechnological potential. In *Yeast biotechnology: diversity and applications* (pp. 615-634). Springer, Dordrecht.

Bose, I., Reese, A.J., Ory, J.J., Janbon, G. and Doering, T.L., 2003. A yeast under cover: the capsule of *Cryptococcus neoformans*. *Eukaryotic cell*, *2*(4), pp.655-663.

Brilhante, R.S.N., Castelo-Branco, D.D.S.C.M., Soares, G.D.P., Astete-Medrano, D.J., Monteiro, A.J., Cordeiro, R.D.A., Sidrim, J.J.C. and Rocha, M.F.G., 2010. Characterization of the gastrointestinal yeast microbiota of cockatiels (Nymphicus hollandicus): a potential hazard to human health. *Journal of medical microbiology*, *59*(6), pp.718-723.

Brown, S.P., Cornforth, D.M. and Mideo, N., 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology*, *20*(7), pp.336-342.

Caicedo, L.D., Alvarez, M.I., Delgado, M. and Cárdenas, A., 1999. Cryptococcus neoformans in bird excreta in the city zoo of Cali, Colombia. *Mycopathologia*, *147*(3), pp.121-124.

Casadevall, A. and Pirofski, L.A., 1999. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infection and immunity*, *67*(8), pp.3703-3713.

Choe, Y.J., Blatt, D.B., Yalcindag, A., Geffert, S.F., Bobenchik, A.M. and Michelow, I.C., 2020. Cryptococcus albidus Fungemia in an immunosuppressed child: case report and systematic literature review. *Journal of the Pediatric Infectious Diseases Society*, *9*(1), pp.100-105.

Chryssanthou, E., Wennberg, H., Bonnedahl, J. and Olsen, B., 2011. Occurrence of yeasts in faecal samples from Antarctic and South American seabirds. *Mycoses*, *54*(6), pp.e811-e815.

Crowe, J.D. and Olsson, S., 2001. Induction of laccase activity in *Rhizoctonia solani* by antagonistic *Pseudomonas fluorescens* strains and a range of chemical treatments. *Applied and Environmental Microbiology*, *67*(5), pp.2088-2094.

Cuomo, C.A., Rhodes, J. and Desjardins, C.A., 2018. Advances in *Cryptococcus* genomics: insights into the evolution of pathogenesis. *Memórias do Instituto Oswaldo Cruz*, *113*.

D'Amico, S., 2002. Claverie P, Collins T, Georlette D, Gratia E, Hoyoux A, Meuwis MA, Feller G, and Gerday C. *Molecular basis of cold adaptation. Philos Trans R Soc Lond B Biol Sci*, *357*, pp.917-925.

D'Amico, S., Marx, J.C., Gerday, C. and Feller, G., 2003. Activity-stability relationships in extremophilic enzymes. *Journal of Biological Chemistry*, *278*(10), pp.7891-7896.

de Assis Baroni, F., Paula, C.R., da Silva, É.G., Viani, F.C., Rivera, I.N., de Oliveira, M.T.B. and Gambale, W., 2006. Cryptococcus neoformans em torres de igrejas da cidade do Rio de Janeiro, RJ, Brasil. *Revista do Instituto de Medicina Tropical de São Paulo*, *48*(2), pp.71-75.

De Maayer, P., Aliyu, H. and Cowan, D.A., 2019. Reorganising the order Bacillales through phylogenomics. *Systematic and applied microbiology*, *42*(2), pp.178-189.

De Maayer, P., Anderson, D., Cary, C. and Cowan, D.A., 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO reports, 15(5), pp.508-517.

De Vries, RP, Tsang, A & Grigoriev, IV 2018, *Fungal Genomics: Methods and Protocols*. Methods in Molecular Biology, Springer.

Duarte, A.W.F., Dos Santos, J.A., Vianna, M.V., Vieira, J.M.F., Mallagutti, V.H., Inforsato, F.J., Wentzel, L.C.P., Lario, L.D., Rodrigues, A., Pagnocca, F.C. and Pessoa Junior, A., 2018.

Cold-adapted enzymes produced by fungi from terrestrial and marine Antarctic environments. *Critical reviews in biotechnology*, *38*(4), pp.600-619.

Ekhtiari, M., Farahyar, S., Falahati, M., Razmjou, E., Ashrafi-Khozani, M., Ghasemi, Z. and Abbasi-Nejat, Z., 2017. The first report of onychomycosis caused by Cryptococcus friedmannii (Naganishia friedmannii) a basidiomycetous yeast. *Medical mycology case reports*, *15*, pp.25-27.

Elhariri, M., Hamza, D., Elhelw, R. and Refai, M., 2015. Lovebirds and cockatiels risk reservoir of *Cryptococcus neoformans*, a potential hazard to human health. *J Vet Sci Med Diagn 4*, *4*, p.2.

Farrer, R.A. and Fisher, M.C., 2017. Describing genomic and epigenomic traits underpinning emerging fungal pathogens. *Advances in genetics*, *100*, pp.73-140.

Fell, J.W., Roeijmans, H. and Boekhout, T., 1999. *Cystofilobasidiales*, a new order of basidiomycetous yeasts. *International Journal of Systematic and Evolutionary Microbiology*, *49*(2), pp.907-913.

Feller, G. and Gerday, C., 1997. Psychrophilic enzymes: molecular basis of cold adaptation. *Cellular and Molecular Life Sciences CMLS*, *53*(10), pp.830-841.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F. and Dougherty, B.A., 37. other authors (1995). *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science*, *269*, pp.496-512.

Fonseca, Á., Boekhout, T. and Fell, J.W., 2011. Cryptococcus vuillemin (1901). In *The yeasts* (pp. 1661-1737). Elsevier.

Fonseca, Á., Scorzetti, G. and Fell, J.W., 1999. Diversity in the yeast Cryptococcus albidus and related species as revealed by ribosomal DNA sequence analysis. *Canadian journal of microbiology*, *46*(1), pp.7-27.

Fotedar, R., Kolecka, A., Boekhout, T., Fell, J.W., Anand, A., Al Malaki, A., Zeyara, A. and Al Marri, M., 2018. Naganishia qatarensis sp. nov., a novel basidiomycetous yeast species from a hypersaline marine environment in Qatar. *International journal of systematic and evolutionary microbiology*, *68*(9), pp.2924-2929.

Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.C., Sonan, G., Feller, G. and Gerday, C., 2004. Some like it cold: biocatalysis at low temperatures. *FEMS microbiology reviews*, *28*(1), pp.25-42.

GERMS-SA G for ER and M disease S in SA (2015) GERMS-SA Annual Report

Gharehbolagh, S.A., Nasimi, M., Afshari, S.A.K., Ghasemi, Z. and Rezaie, S., 2017. First case of superficial infection due to Naganishia albida (formerly Cryptococcus albidus) in Iran: A review of the literature. *Current medical mycology*, *3*(2), p.33.

Goldanskii, V.I., 1976. Chemical reactions at very low temperatures. *Annual Review of Physical Chemistry*, *27*(1), pp.85-126.

Goto, S., 1963. On a new yeast genus Naganishia. *Journal of Fermentation Technology*, *41*, pp.459-462.

Goto, S., Sugiyama, J. and Iizuka, H., 1969. A taxonomic study of Antarctic yeasts. *Mycologia*, *61*(4), pp.748-774.

Gupta, N., Haque, A., Lattif, A.A., Narayan, R.P., Mukhopadhyay, G. and Prasad, R., 2004. Epidemiology and molecular typing of Candida isolates from burn patients. *Mycopathologia*, *158*(4), pp.397-405.

Heitman, J., Kozel, T.R., Kwon-Chung, K.J., Perfect, J.R. and Casadevall, A., 2010. *Cryptococcus: from human pathogen to model yeast*. ASM press.

Hottiger, T., DE Virgilio, C., Hall, M.N., Boller, T. and Wiemken, A., 1994. The role of trehalose synthesis for the acquisition of thermotolerance in yeast: II. Physiological concentrations of trehalose increase the thermal stability of proteins in vitro. *European Journal of Biochemistry*, *219*(1-2), pp.187-193.

Huang, H., Pan, J., Yang, W., Lin, J., Han, Y., Lan, K., Zeng, L., Liang, G. and Liu, J., 2020. First case report of Cryptococcus laurentii knee infection in a previously healthy patient. *BMC Infectious Diseases*, *20*(1), pp.1-8.

Ikeda, R., Sugita, T., Jacobson, E.S. and Shinoda, T., 2002. Laccase and melanization in clinically important Cryptococcus species other than Cryptococcus neoformans. *Journal of Clinical Microbiology*, *40*(4), pp.1214-1218.

Inouye, M. and Phadtare, S., 2014. Cold-shock response and adaptation to near-freezing temperature in cold-adapted yeasts. In *Cold-adapted Yeasts* (pp. 243-257). Springer, Berlin, Heidelberg.

Jenuth, J.P., 2000. The ncbi. In *Bioinformatics methods and protocols* (pp. 301-312). Humana Press, Totowa, NJ.

Johns, G.C. and Somero, G.N., 2004. Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes (Chromis spp.). *Molecular biology and evolution*, *21*(2), pp.314-320.

Khan, Z.U., Ahmad, S., Hagen, F., Fell, J.W., Kowshik, T., Chandy, R. and Boekhout, T., 2010. *Cryptococcus randhawai* sp. nov., a novel anamorphic basidiomycetous yeast isolated from tree trunk hollow of Ficus religiosa (peepal tree) from New Delhi, India. *Antonie Van Leeuwenhoek*, *97*(3), pp.253-259.

Khawcharoenporn, T., Apisarnthanarak, A. and Mundy, L.M., 2007. Non-neoformans cryptococcal infections: a systematic review. *Infection*, *35*(2), pp.51-58.

Köhler, J.R., Hube, B., Puccia, R., Casadevall, A. and Perfect, J.R., 2017. Fungi that infect humans. *Microbiology spectrum*, *5*(3), pp.5-3.

Kordossis, T., Avlami, A., Velegraki, A., Stefanou, I., Georgakopoulos, G., Papalambrou, C. and Legakis, N.J., 1998. First report of Cryptococcus laurentii meningitis and a fatal case of Cryptococcus albidus cryptococcaemia in AIDS patients. *Medical mycology*, *36*(5), pp.335-339.

Kowalski, L.R., Kondo, K. and Inouye, M., 1995. Cold-shock induction of a family of TIP1-related proteins associated with the membrane in Saccharomyces cerevisiae. *Molecular microbiology*, *15*(2), pp.341-353.

Kuramae, E.E., Robert, V., Snel, B., Weiß, M. and Boekhout, T., 2006. Phylogenomics reveal a robust fungal tree of life. *FEMS Yeast Research*, *6*(8), pp.1213-1220.

Kurtzman, C.P., Fell, J.W. and Boekhout, T. eds., 2011. *The yeasts: a taxonomic study*. Elsevier.

Lee, Y.A., Kim, H.J., Lee, T.W., Kim, M.J., Lee, M.H., Lee, J.H. and Ihm, C.G., 2004. First report of *Cryptococcus albidus*-induced disseminated cryptococcosis in a renal transplant recipient. *The Korean journal of internal medicine*, *19*(1), p.53.

Limper, A.H., Adenis, A., Le, T. and Harrison, T.S., 2017. Fungal infections in HIV/AIDS. *The Lancet Infectious Diseases*, *17*(11), pp.e334-e343.

Liu, X.Z., Wang, Q.M., Göker, M., Groenewald, M., Kachalkin, A.V., Lumbsch, H.T., Millanes, A.M., Wedin, M., Yurkov, A.M., Boekhout, T. and Bai, F.Y., 2015. Towards an integrated phylogenetic classification of the Tremellomycetes. *Studies in mycology*, *81*, pp.85-147.

Liu, X.Z., Wang, Q.M., Theelen, B., Groenewald, M., Bai, F.Y. and Boekhout, T., 2015. Phylogeny of tremellomycetous yeasts and related dimorphic and filamentous basidiomycetes reconstructed from multiple gene sequence analyses. *Studies in mycology*, *81*, pp.1-26.

Longwell, J.P. and Weiss, M.A., 1955. High temperature reaction rates in hydrocarbon combustion. *Industrial & Engineering Chemistry*, *47*(8), pp.1634-1643.

Lopez-Martinez, R. and Castanon-Olivares, L.R., 1995. Isolation ofCryptococcus neoformans var. neoformans from bird droppings, fruits and vegetables in Mexico City. *Mycopathologia*, *129*(1), pp.25-28.

Lugarini, C., Goebel, C.S., Condas, L.A.Z., Muro, M.D., de Farias, M.R., Ferreira, F.M. and Vainstein, M.H., 2008. Cryptococcus neoformans isolated from Passerine and Psittacine bird excreta in the state of Paraná, Brazil. *Mycopathologia*, *166*(2), pp.61-69.

Luque, R., Lovett, J.C., Datta, B., Clancy, J., Campelo, J.M. and Romero, A.A., 2010. Biodiesel as feasible petrol fuel replacement: a multidisciplinary overview. *Energy & Environmental Science*, *3*(11), pp.1706-1721.

Lynch, R.C., King, A.J., Farías, M.E., Sowell, P., Vitry, C. and Schmidt, S.K., 2012. The potential for microbial life in the highest-elevation (> 6000 masl) mineral soils of the Atacama region. *Journal of Geophysical Research: Biogeosciences*, *117*(G2).

Mancianti, F., Nardoni, S. and Ceccherelli, R., 2002. Occurrence of yeasts in psittacines droppings from captive birds in Italy. *Mycopathologia*, *153*(3), pp.121-124.

Marinho, M., Táparo, C.V., Silva, B.G., Tencate, L.N. and Perri, S.H.V., 2010. Microbiota from captive passeriformes of the northwest region of São Paulo. *Veterinária e Zootecnia*, *17*(2), pp.288-292.

Martínez, J.L., 2014. Short-sighted evolution of bacterial opportunistic pathogens with an environmental origin. *Frontiers in microbiology*, *5*, p.239.

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

Mendes, J.F., Albano, A.P.N., Coimbra, M.A.A., Ferreira, G.F.D., Gonçalves, C.L., Nascente, P.D.S. and Mello, J.R.B.D., 2014. Fungi isolated from the excreta of wild birds in screening centers in Pelotas, RS, Brazil. *Revista do Instituto de Medicina Tropical de São Paulo*, *56*, pp.525-528.

Naranjo-Ortiz, M.A. and Gabaldón, T., 2019. Fungal evolution: major ecological adaptations and evolutionary transitions. *Biological Reviews*, *94*(4), pp.1443-1476.

Narayan, S., Batta, K., Colloby, P. and Tan, C.Y., 2000. Cutaneous *cryptococcus* infection due to *C. albidus* associated with Sezary syndrome. *British Journal of Dermatology*, *143*(3), pp.632-634.

Nizovoy, P., Bellora, N., Haridas, S., Sun, H., Daum, C., Barry, K., Grigoriev, I.V., Libkind, D., Connell, L.B. and Moliné, M., 2021. Unique genomic traits for cold adaptation in Naganishia vishniacii, a polyextremophile yeast isolated from Antarctica. *FEMS Yeast Research*, *21*(1), p.foaa056.

Nosanchuk, J.D. and Casadevall, A., 2003. The contribution of melanin to microbial pathogenesis. *Cellular microbiology*, *5*(4), pp.203-223.

Pakshir, K., Fakhim, H., Vaezi, A., Meis, J.F., Mahmoodi, M., Zomorodian, K., Javidnia, J., Ansari, S., Hagen, F. and Badali, H., 2018. Molecular epidemiology of environmental Cryptococcus species isolates based on amplified fragment length polymorphism. *Journal de Mycologie MÚdicale*, *28*(4), pp.599-605.

Park, S.S., Lee, H., Park, W.S., Hwang, S.H., Choi, S.I., Choi, M.H., Lee, S.W., Ko, E.J., Choi, Y.J. and Eom, H.S., 2017. A case of disseminated infection with skin manifestation due to non-neoformans and non-gattii Cryptococcus in a patient with refractory acute myeloid leukemia. *Infection & chemotherapy*, *49*(2), pp.142-145.\

Polz, M.F., Hunt, D.E., Preheim, S.P. and Weinreich, D.M., 2006. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), pp.2009-2021.

Powel, M.S., Alizadeh, A.A., Budvytiene, I., Schaenman, J.M. and Banaei, N., 2012. First isolation of *Cryptococcus uzbekistanensis* from an immunocompromised patient with lymphoma. *Journal of clinical microbiology*, *50*(3), pp.1125-1127.

Ragupathi, L. and Reyna, M., 2015. Case report of *Cryptococcus albidus* peritonitis in a peritoneal dialysis patient and a review of the literature. *Peritoneal Dialysis International*, *35*(4), pp.421-427.

Ramos, J.M., Cuenca-Estrella, M., Gutierrez, F., Elia, M. and Rodriguez-Tudela, J.L., 2004. Clinical case of endocarditis due to *Trichosporon inkin* and antifungal susceptibility profile of the organism. *Journal of clinical microbiology*, *42*(5), pp.2341-2344.

Ranjan, K. and Poças-Fonseca, M.J., 2019. Genetic diversity of pathogenic yeasts. *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications*, pp.593-615.

Rimek, D., Haase, G., Lück, A., Casper, J. and Podbielski, A., 2004. First report of a case of meningitis caused by Cryptococcus adeliensis in a patient with acute myeloid leukemia. *Journal of clinical microbiology*, *42*(1), pp.481-483.

Rodrigues, D.F. and Tiedje, J.M., 2008. Coping with our cold planet. *Applied and environmental microbiology*, *74*(6), pp.1677-1686.

Rodrigues, M.L., Alviano, C.S. and Travassos, L.R., 1999. Pathogenicity of Cryptococcus neoformans: virulence factors and immunological mechanisms. *Microbes and infection*, *1*(4), pp.293-301.

Rosario, I., Soro, G., Deniz, S., Ferrer, O., Acosta, F., Padilla, D. and Acosta, B., 2010. Presence of *C. albidus*, *C. laurentii* and *C. uniguttulatus in* crop and droppings of pigeon lofts (Columba livia). *Mycopathologia*, *169*(4), pp.315-319.

Salazar-Leal, J.I., Ramírez-Montelongo, S.M. and Luis, B.A.L., 2019. Clinical significance of nosocomial *Cryptococcus laurentii* in urine: A case series. *Infection Control & Hospital Epidemiology*, *40*(12), pp.1442-1444.

Santos, L.L.D., Ferreira, F.M., Lopes, S.F., Condas, L.A., Muro, M.D. and Lugarini, C., 2009. Research on *Cryptococcus neoformans* and *Candida spp.* in the excretas of captive psittacisittacines and passeriformes. *Arq. ciênc. vet. zool. UNIPAR*, pp.5-9.

Schmidt, S.K., Vimercati, L., Darcy, J.L., Arán, P., Gendron, E.M., Solon, A.J., Porazinska, D. and Dorador, C., 2017. A *Naganishia* in high places: functioning populations or dormant cells from the atmosphere? *Mycology*, *8*(3), pp.153-163.

Scorzetti, G., Fell, J.W., Fonseca, A. and Statzell-Tallman, A., 2002. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. *FEMS yeast research*, *2*(4), pp.495-517.

Scorzetti, G., Petrescu, I., Yarrow, D. and Fell, J.W., 2000. *Cryptococcus adeliensis* sp. nov., a xylanase producing basidiomycetous yeast from Antarctica. *Antonie van Leeuwenhoek*, *77*(2), pp.153-157.

Selvakumar, P. and Sivashanmugam, P., 2018. Study on lipid accumulation in novel oleaginous yeast Naganishia liquefaciens NITTS2 utilizing pre-digested municipal waste activated sludge: a low-cost feedstock for biodiesel production. *Applied biochemistry and biotechnology*, *186*(3), pp.731-749.

Selvakumar, P., Arunagiri, A. and Sivashanmugam, P., 2019. Thermo-sonic assisted enzymatic pre-treatment of sludge biomass as potential feedstock for oleaginous yeast cultivation to produce biodiesel. *Renewable Energy*, *139*, pp.1400-1411.

Serda Kantarcioğlu, A., Boekhout, T., Sybren De Hoog, G., Theelen, B., Yücel, A., Ekmekci, T.R., Fries, B.C., Ikeda, R., Koslu, A. and Altas, K., 2007. Subcutaneous cryptococcosis due to *Cryptococcus diffluens* in a patient with sporotrichoid lesions case report, features of the case isolate and in vitro antifungal susceptibilities. *Medical mycology*, *45*(2), pp.173-181.

Setati, M.E., Jacobson, D., Andong, U.C. and Bauer, F., 2012. The vineyard yeast microbiome, a mixed model microbial map. *PloS one*, *7*(12), p.e52609.

Sharma, K.K., 2016. Fungal genome sequencing: basic biology to biotechnology. *Critical reviews in biotechnology*, *36*(4), pp.743-759.

Shen, Q., Chen, Y., Jin, D., Lin, H., Wang, Q. and Zhao, Y.H., 2016. Comparative genome analysis of the oleaginous yeast Trichosporon fermentans reveals its potential applications in lipid accumulation. *Microbiological research*, *192*, pp.203-210.

Siddiqui, K.S. and Cavicchioli, R., 2006. Cold-adapted enzymes. *Annu. Rev. Biochem.*, *75*, pp.403-433.

Silva, M.E. and Paula, L.A., 1963. Isolation of Cryptococcus neoformans from excrement and nests of pigeons (Columba livia) in Salvador, Bahia (Brazil). *Revista do Instituto de Medicina Tropical de São Paulo*, *5*, pp.9-11.

Skillman, D., Riek, L., Davis, B., Harris, J.R. and Nett, R.J., 2013. Histoplasmosis in a state where it is not known to be endemic—Montana, 2012–2013. *MMWR. Morbidity and mortality weekly report*, *62*(42), p.834.

Stajich, J.E., 2017. Fungal genomes and insights into the evolution of the kingdom. *Microbiology spectrum*, *5*(4), pp.5-4.

Steenbergen, J.N., Nosanchuk, J.D., Malliaris, S.D. and Casadevall, A., 2003. *Cryptococcus neoformans* virulence is enhanced after growth in the genetically malleable host Dictyostelium discoideum. *Infection and immunity*, *71*(9), pp.4862-4872.

Sugita, T., Saito, M., Ito, T., Kato, Y., Tsuboi, R., Takeuchi, S. and Nishikawa, A., 2003. The basidiomycetous yeasts *Cryptococcus diffluens* and *C. liquefaciens* colonize the skin of patients with atopic dermatitis. *Microbiology and immunology*, *47*(12), pp.945-950.

Sugita, T., Takashima, M., Ikeda, R., Nakase, T. and Shinoda, T., 2001. Intraspecies diversity of *Cryptococcus albidus* isolated from humans as revealed by sequences of the internal transcribed spacer regions. *Microbiology and immunology*, *45*(4), pp.291-297.

Taei, M., Chadeganipour, M. and Mohammadi, R., 2019. An alarming rise of non-albicans *Candida* species and uncommon yeasts in the clinical samples; a combination of various molecular techniques for identification of etiologic agents. *BMC research notes*, *12*(1), pp.1-7.

Takashima, M., Manabe, R.I., Nishimura, Y., Endoh, R., Ohkuma, M., Sriswasdi, S., Sugita, T. and Iwasaki, W., 2019. Recognition and delineation of yeast genera based on genomic data: Lessons from Trichosporonales. *Fungal Genetics and Biology*, *130*, pp.31-42.

Tintelnot, K. and Losert, H., 2005. Isolation of *Cryptococcus adeliensis* from clinical samples and the environment in Germany. *Journal of clinical microbiology*, *43*(2), pp.1007-1007.

Turchetti, B., Buzzini, P., Goretti, M., Branda, E., Diolaiuti, G., D'Agata, C., Smiraglia, C. and Vaughan-Martini, A., 2008. Psychrophilic yeasts in glacial environments of Alpine glaciers. *FEMS microbiology ecology*, *63*(1), pp.73-83.

Turchetti, B., Selbmann, L., Blanchette, R.A., Di Mauro, S., Marchegiani, E., Zucconi, L., Arenz, B.E. and Buzzini, P., 2015. Cryptococcusvaughanmartiniae sp. nov. and Cryptococcusonofrii sp. nov.: two new species isolated from worldwide cold environments. *Extremophiles*, *19*(1), pp.149-159.

UNAIDS (2018) UNAIDS data 2018. https://www.unaids.org/en/resources/documents/2018/unaids-data-2018. Accessed 28 Sep 2021

Vimercati Shashwat and Kartik, C., 2016. Draft Genome Sequence of the Oleaginous Yeast *Cryptococcus albidus* var. albidus. *Microbiology Resource Announcements*, *4*(3).

Vishniac HS, Hempfling WP. *Cryptococcus vishniacii* sp. nov., an Antarctic yeast. International Journal of Systematic and Evolutionary Microbiology. 1979 Apr 1;29(2):153-8.

Vishniac, H.S. and Kurtzman, C.P., 1992. Cryptococcus antarcticus sp. nov. and Cryptococcus albidosimilis sp. nov., basidioblastomycetes from Antarctic soils. *International Journal of Systematic and Evolutionary Microbiology*, *42*(4), pp.547-553.

Vishniac, H.S., 1985. *Cryptococcus friedmannii*, a new species of yeast from the Antarctic. *Mycologia*, *77*(1), pp.149-153.

Vishniac, H.S., 2006. A multivariate analysis of soil yeasts isolated from a latitudinal gradient. *Microbial ecology*, *52*(1), pp.90-103.

Ward, O.P. and Singh, A., 2005. Omega-3/6 fatty acids: alternative sources of production. *Process biochemistry*, *40*(12), pp.3627-3652.

Williamson, P.R., 1997. Laccase and melanin in the pathogenesis of Cryptococcus neoformans. *Front. Biosci*, *2*, pp.e99-e107.

Wombwell, E., Bransteitter, B. and Gillen, L.R., Incidence of Saccharomyces cerevisiae Fungemia in Hospitalized Patients Administered Saccharomyces boulardii Probiotic. *Mycoses*.

Wu, Y., Du, P.C., Li, W.G. and Lu, J.X., 2012. Identification and molecular analysis of pathogenic yeasts in droppings of domestic pigeons in Beijing, China. *Mycopathologia*, *174*(3), pp.203-214.

Yarrow, D., 1998. Methods for the isolation, maintenance and identification of yeasts. In *The yeasts* (pp. 77-100). Elsevier.

Yong, X., Yan, Z., Xu, L., Zhou, J., Wu, X., Wu, Y., Li, Y., Chen, Z., Zhou, H., Wei, P. and Jia, H., 2016. Genome sequence of a microbial lipid producing fungus Cryptococcus albidus NT2002. *Journal of biotechnology*, *223*, pp.6-7.

Zhou, Y., Jia, B.S., Zhou, Y.G., Li, A.H. and Xue, L., 2020. Naganishia floricola sp. nov., a novel basidiomycetous yeast species isolated from flowers of Sorbaria sorbifolia. *International Journal of Systematic and Evolutionary Microbiology*, *70*(8), pp.4496-4501.

Chapter 2 Sequencing and assembly of the genome of *Naganishia randhawae* eABCC1

2.1 Abstract

Members of the genus *Naganishia* are mainly isolated from extreme terrestrial environments and have been linked to clinical infections. *Naganishia randhawae* eABCC1 was isolated from avian guano in South Africa. Here we report the draft genome sequence of this isolate, the first genome for this species to be sequenced globally. Library preparation and sequencing were performed using the Illumina NovaSeq 6000 platform (paired-end read approach 2 X 250bp) at MR DNA (https://www.mrdnalab.com/; Texas, USA). The final genome assembly encompasses 20.27 million nucleotides with a G+C content of 51.71 %. Structural and functional annotation using the Funannotate pipeline v1.5.1-93c317b identified 6,943 genes including 6,775 predicted protein-coding sequences and 168 tRNAs. Analysis of the *N. randhawae* eABCC1 genome has revealed that the yeast possesses multiple virulence-related genes, including the melanin producing enzyme, laccase. This represents the first genome for this species. The genomic content of *N. randhawae* eABCC1 indicates that this strain can withstand adverse conditions such as those found in avian guano.

2.2 Introduction

Research in modern molecular biology is primarily driven by genetic information derived through genome sequencing. Genome sequencing was pioneered in 1979, with Roger Staden's work proposing shotgun sequencing for small genomes (Staden, 1979). This technique involved random splicing of the genome of an organism and sequencing using the Sanger chain termination method (Sanger *et al.*, 1977; Staden, 1979). As such, organisms with small genomes served as model organisms in the early stages of whole-genome sequencing.

The sequencing of the complete genome of the yeast *Saccharomyces cerevisiae* S288C in 1996 marked the beginning of the fungal genomic era (Goffeau *et al.*, 1996). This was perceived as an important milestone that influenced our understanding of basic biology, particularly eukaryotic gene function and expression (Galagan *et al.*, 2005a). With the immense biological diversity existing in the fungal kingdom, the genome of *S. cerevisiae* was just the tip of the iceberg. Subsequent sequencing of the genomes of a broad range of fungal taxa such as *Aspergillus* species (Galagan *et al.*, 2005a), *Cryptococcus neoformans* (Loftus *et al.*, 2005), *Neurospora crassa* (Butler *et al.*, 2003) and *Schizosaccharomyces pombe* (Wood *et al.*, 2002)

have provided an essential tool for studying the molecular mechanisms underlying eukaryotic genome evolution.

Sequencing efforts by consortia such as the Fungal Genome Initiative (FGI) (Cuomo and Birren, 2010) and the Fungal Program of the United States Department of Energy (DOE) Joint Genome Institute (JGI) have propelled the field of fungal genomics (Grigoriev *et al.,* 2012). Currently, more than 7,286 fungal genome assemblies are publicly accessible on the National Center for Biotechnology Information (NCBI) database. Next-generation sequencing (NGS) advancements have made large-scale genome sequencing feasible and cost-effective (Sanmiguel, 2011). Sequencing platforms such as Illumina and PacBio are now easily accessible for laboratory experiments (Morozova and Marra, 2008). The available genome data does, however, show a bias towards a few fungal species models that do not accurately represent the vast diversity of the Kingdom Fungi. As such, exploring the broader taxonomic spectrum of Fungi through genome sequencing remains a constant challenge that will require a collective effort from research groups around the world (Gabaldón, 2020).

Member of the genus *Naganishia* belong to the class Tremellomycetes, order Filobasidiales, family *Filobasidiaceae*. They have a global distribution particularly in extreme terrestrial environments characterized by cold temperatures and recurrent diurnal freeze thaw cycles (Costello *et al.,* 2009; Lynch *et al.,* 2012b; Pulschen *et al.,* 2015; Schmidt *et al.,* 2017; Solon *et al.,* 2018). Additionally, many of these species have been linked with human infection. *Naganishia albida* has been implicated in cutaneous lesions, encephalitis, keratitis, onychomycosis, and pneumonia (Burnik *et al.,* 2007; Lee *et al.,* 2004; Ragupathi and Reyna, 2015); *N. diffluens* is associated with subcutaneous infections (Serda Kantarcioğlu *et al.,* 2007); *N. friedmannii* has been confirmed as an etiologic agent of onychomycosis (Ekhtiari *et al.,* 2017) and, *N. uzbekistanensis* has been isolated from the bone-marrow of lymphoma patients (Powel *et al.,* 2012). Despite their wide distribution and medical importance, the mechanism underlying the diverse functions of *Naganishia* species is poorly understood.

In this study, *N. randhawae* strain eABCC1, was isolated from avian guano in South Africa and produced melanin on bird seed agar (BSA) (Yarrow, 1998). Prior to this study, little was known about the genomic content of *N. randhawae*. Here we present the first draft genome sequence of *N. randhawae* strain eABCC1 and highlight important adaptive features encoded on the genome of this taxon. The availability of the genome sequence of strain eABCC1 should

improve our understanding of the biology of this taxon, as well as the broader spectrum of the genus *Naganishia* and the Tremellomycetes as a whole

2.3 Materials and methods

2.3.1 Sample processing and yeast isolation

A yeast strain (eABCC1) was isolated from dry avian guano sampled outside the Holy Trinity church (-26.179315754422458, 28.03078085996224), Johannesburg, South Africa. Briefly, 1 g of avian guano was added to a test tube containing 10ml of 0.9 % (w/v) physiological saline solution (PSS), vortexed at maximum speed for two minutes, and allowed to settle for five minutes. Subsequently, 0.1 ml of the supernatant was plated onto BSA (Yarrow, 1998) in duplicate and plates were incubated at 25 ºC and 30 ºC. Plates were checked daily for the presence of brown-pigmented yeast colonies. Morphologies of identified yeast isolates were observed using a light microscope (Zeiss, Primo star, Germany) at 400x magnification.

2.3.2 Genomic DNA extraction

Strain eABBC1 was inoculated into 5 ml yeast peptone dextrose (YPD) broth (Sherman, 1991) and incubated at 30 ºC for 48 hours on a rotary shaker. Genomic DNA extraction was performed using the Quick-DNA™ Fungal/Bacterial Microprep Kit (ZYMO RESEARCH Kit, California, United States) as per manufacturer's instructions. The genomic DNA was visualized on a 1 % (w/v) agarose gel.

2.3.3 Strain identification

Molecular identification of strain eABBC1 was undertaken on the basis of polymerase chain reaction (PCR) amplification and sequence analysis of the ITS region with the universal primers ITS1 and ITS4 (White *et al.,* 1990). The ITS region of fungal are highly variable sequences that are commonly used as molecular markers for the identification of fungal species (Martin and Rygiewicz, 2005). The PCR amplification was performed using the Taq 5X Master Mix (NEB, Massachusetts, United States) as per the manufacturer's recommendations. The reaction mix was placed in a thermal cycler (Bio-Rad, California, United States) with the following settings: initial denaturation at 95 ºC for 2 minutes; 30 cycles of 95 ºC for 1 minute (denaturation), 52 ºC for 1 minute (annealing), and 68 °C for 1.5 minutes (extension) and a final extension step at 68 °C for 5 minutes. The amplicon was visualized on a 1 % (w/v) agarose

gel. Sanger sequencing was performed at Inqaba Biotech (Pretoria, South Africa) using the Big Dye Terminator Ready Reaction v3.1 cycle sequencing kit (Thermo Scientific, USA). Ambiguous bases were resolved using FinchTV v1.4 (http://www.geospiza.com/Products/finchtv.shtml) and the sequence ends were trimmed in BioEdit v7.0.5.3 (Hall, 1999). This was followed by a BLASTN analysis against the NCBI non redundant database to confirm the identity of the species (Johnson *et al.,* 2008). The ITS nucleotide sequence of *N. randhawae* eABCC1 generated in this study has been deposited in GenBank under the accession number: MT542688.

2.3.4 Phylogenetic analysis

Maximum-likelihood (ML) estimation is a widely used and useful statistical method in phylogenetic analysis (Sullivan, 2005). The ML method produces the tree with the highest likelihood given a specified evolutionary model (Felsenstein, 1981). Regardless of efficiency, ML reconstruction is as simple as calculating likelihood and selecting the best tree. While the ML method provides a theoretical framework, its effectiveness is determined by statistical models, which makes the ML method extremely versatile due to model flexibility (Sullivan, 2005). Various evolutionary information can be inferred smoothly in the statistical framework, making it one of the most reliable tree building algorithms (Goolsby, 2017).

The taxonomic placement of *N. randhawae* eABCC1 was investigated by constructing a maximum likelihood (ML) phylogeny (Guindon *et al.,* 2010) on the basis of the ITS sequences of strain eABCC1 and the type strains of each *Naganishia* species for which sequences are available on the NCBI nucleotide database (https:// www.ncbi.nlm.nih.gov/ nuccore; Table 2.1). The ITS sequence of *Filobasidium wieringae* CBS1937 (AF444373.1) was included as an outgroup based on its phylogenetic closeness to members of the genus *Naganishia* (Liu *et al.,* 2015a). Sequences were aligned using the M-Coffee webserver (Moretti *et al.,* 2007), before trimming of the unaligned 5′ and 3′ ends. The ML phylogeny was constructed using PhyML-SMS with smart model selection (Akaike Information Criterion) (Guindon *et al.,* 2010; Lefort *et al.,* 2017) and bootstrap support (n = 1,000 replicates).

Table 2.1 List of fungal strains used in this study

Taxon	Strain	Reference No.
Naganishia adeliensis	CBS 8351^{T}	AF145328.2
N. albida	CBS 142 ^{T}	AF145321.1
N. albidosimilis	CBS 7711 ^{T}	AF145325.1
N. antarctica	CBS 7687 ^{T}	AF145326.1
N. bhutanensis	CBS 6294 ^{T}	AF145317.1
N. diffluens	CBS 160 ^{T}	AF145330.1
N. friedmannii	CBS 7160 ^{T}	AF145322.1
N. globosa	CBS 1975 ^{T}	AF444372.1
N. liquefaciens	CBS 968 ^{T}	AF444345.1
N. qatarensis	QCCY17/17 ^{T}	MG852088.1
N. randhawae	CBS 10160 ^{T}	AJ876528.1
N. randhawae	eABCC1	MT542688
N. uzbekistanensis	CBS 8683 ^{T}	AF444339.1
N. vaughanmartiniae	DBVPG4736	KF861792.1
N. vishniacii	CBS 7110 ^{T}	AF145320.1
Outgroup		
Filobasidium wieringae	CBS 1937 ^{T}	AF444373.1

2.3.5 Genome sequencing and assembly

The process of assembling a genome from sequence reads is algorithm-driven and automated (Choudhuri, 2014). However, the choice of assembly method and algorithm are heavily dependent on the species being sequenced, the size of the genome and quality of the raw sequence reads (Dominguez *et al.*, 2018).

For this study, library preparation and sequencing were performed using the Illumina NovaSeq 6000 platform (paired end read approach 2 X 250 base pairs [bp]) at MR DNA (https://www.mrdnalab.com/; Texas, USA). A total of 17,907,138 (paired ends) raw reads were generated with an average read length of 151 nucleotides. Adapter sequences and low quality reads (Phred scores less than 28) were trimmed and removed using the FastQ toolkit v.0.11.8 (Andrews, 2010). Discarding the low-quality reads yielded 8,076,080 reads with a G+C content of 48 %. Genome sequencing using the Illumina platform generates a large number of high-quality short sequence reads (Dominguez *et al.*, 2018). Adapter and multiplex index sequences are typically screened for and removed following base calling on the sequencing instrument.

However, prior to assembly, it is advised to assess the raw sequence data quality (Heydari *et al.,* 2017). Poor quality reads, ambiguous base calling, and contamination are all examples of potential technical errors that may be identified and resolved prior to assembly (Sturm *et al.,* 2016). For example, in this study, a large fraction of the sequencing reads were discarded due to their poor quality, specifically low Phred quality scores and unresolved NNNNNN regions. Assessing the quality of the sequence data is essential, as it may affect downstream applications and potentially result in inaccurate findings (Liao *et al.,* 2019).

The 8,076,080 reads were assembled *de novo* using the SPAdes genome assembler (Bankevich *et al.,* 2012) which is based on the Eulerian *de Bruijn* graph method (Liao *et al.,* 2019). A genome sequence can be assembled using one of two methods: reference mapping or *de novo* assembly. If a reference genome sequence is available the newly acquired sequence reads can be aligned to the reference genome before being assembled in the correct order; this mode of assembly is known as reference mapping and assembly (Conway and Bromage, 2011). When no reference genome sequence exists, the genome must be assembled *de novo*. This approach relies on a hierarchical process that begins with the assembly of paired-end sequence reads into contigs, then into scaffolds, and finally scaffolds are merged into chromosomes (Choudhuri, 2014; Liao *et al.,* 2019). Given that no reference genome sequence is available for *N. randhawae,* the *de novo* genome assembly method was selected.

The quality and completeness of the assembled genome relies on selecting the appropriate mapping and assembly algorithm. Three algorithms, greedy (Huson *et al.,* 2002), Hamiltonian overlap-layout-consensus (OLC) (Idury and Waterman, 1995) and Eulerian *de Bruijn* graph (Compeau *et al.,* 2011) are commonly used. Greedy is a rapid assembly algorithm that joins together sequences that are most similar to one another, based on the maximum amount of sequence overlap possible (Choudhuri, 2014). During this process, some reads may be missed, resulting in gaps that can be filled using paired-end sequencing (Baxter and Harche, 1992; Choudhuri, 2014). A large number of early assemblers, such as Phrap, the TIGR assembler and the CAP, were based on the greedy algorithm.

The overlap-layout-consensus (OLC) algorithm uses reads and overlaps to generate a directed graph based on all pairwise comparisons (Idury and Waterman, 1995). Each sequence is represented as a node in the graph, and an edge is formed between any two nodes whose sequences overlap (Li *et al.,* 2012). The algorithm then attempts to find the Hamiltonian traversal path of the graph that contains all the nodes (sequences) exactly once and merges the

overlapping sequences in the nodes into the genome sequence (Eugene, 2016). Overlap-based approaches have primarily been used for longer reads (greater than 200 bp) (Choudhuri, 2014). Arachne, CABOG (Celera Assembler), Newbler, Minimus, Edena, and MIRA are assemblers that use the OLC algorithm.

Sequence assembly based on significant sequence overlap using the standard Sanger method, works well when there are a sufficient number of sequences read to be assembled (Choudhuri, 2014). Next-generation sequencing, on the other hand, generates hundreds of millions of sequences reads (Behjati and Tarpey, 2013). Using the traditional Sanger approaches makes it nearly impossible to easily assemble a large number of reads. The Eulerian *de Bruijn* graph method is used to solve the problem of scalability in this case, and has been shown to be more efficient as assembly large genomes (Compeau *et al.,* 2011; Liao *et al.,* 2019). The algorithm does not assemble the sequence read directly, but rather breaks it down into smaller sequences known as k-mers (Li *et al.,* 2012). Alignment of these k-mers is performed using (k-1) sequence overlaps. The size of the k-mers varies depending on sequence coverage, read length, and other factors (Alneberg *et al.,* 2014). In contrast to greedy assembly, Eulerian *de Bruijn* graph appears to compensate for missing sequence reads (Compeau *et al.,* 2011). Thus, while computational application of Eulerian *de Bruijn* graph alleviates a number of the challenges associated with *de novo* sequence assembly, it is not a simple and effective process (Choudhuri, 2014). Some assemblers that implement the Eulerian *de Bruijn* graph algorithm are ABySS, ALLPATH, Oases, SOAPdenovo, SPAdes, and Velvet. As such, the Eulerian *de Bruijn* graph was determined to be the most suitable algorithm to undertake the *de novo* assembly of the *N. randhawae* eABCC1 genome.

Finally, continued and further refinement of any genome assembly has become routine practise, as studies have shown that relying on the first assembly alone often results in a large number of scaffolds (Farrer *et al.,* 2009). As such, our first draft genome assembly (10,543 scaffolds) was improved by aligning the assembly output against raw reads using the Integrated Genome Browser v9.0.2 (Nicol *et al.,* 2009) to extend the scaffolds at the 5′ and 3′ ends. Furthermore, a local BLASTN analysis (using BioEdit v7.0.5.3) of the last 100 nucleotides on either end of each scaffold against the assembled genome was performed and overlapping scaffold ends were then merged into a single scaffold.

2.3.6 Gene prediction and annotation

Completeness of the genome was evaluated using the Benchmarking Universal Single-Copy Ortholog (BUSCO) assessment tool v2.0 with the basidiomycota_odb9 reference dataset (Simão *et al.,* 2015). The BUSCO databases is comprised of genes that have evolved within each lineage and are present universally as single-copy orthologs (Waterhouse *et al.,* 2011). This characteristic underpins the evolutionary assumption that they should be present in entire assemblies or gene sets, and only once (Simão *et al.,* 2015). Completeness is assessed in terms of this expected gene content using BUSCO sequence profiles to determine the orthology status of predicted genes (Seppey *et al.,* 2019; Simão *et al.,* 2015). BUSCOs are carefully chosen with precisely adjusted score and length cut-offs to enhance precision and recall, but because gene prediction and orthology assignment are difficult tasks, evaluations may fall short of 100 % correct classification (Waterhouse *et al.,* 2017).

Structural annotation was performed using the Funannotate pipeline v1.5.1-93c317b (Palmer, 2016). This pipeline incorporates structural annotation algorithms such as Augustus v3.2.3 (Stanke *et al.,* 2006) and GeneMark-ES v4.36 (Borodovsky and Lomsadze, 2011), to produce accurate gene models. Gene prediction and annotation in eukaryotes remains problematic due to the variable size of introns and alternative splice variants (Stanke and Waack, 2003). AUGUSTUS was initially developed to address this issue by performing *ab initio* gene predictions (Stanke *et al.* 2004). Based on the Generalised Hidden Markov Model (GHMM), AUGSUTUS defines probability distributions for various genomic sequence sections (Stanke *et al.* 2006). As a result, introns, exons, intergenic regions, and so on correspond to model states, and each state is thought to generate DNA sequences with pre-defined emission probabilities.

GeneMark-ES v2 allowed for a more accurate prediction of protein-coding genes in fungal genomes (Borodovsky and Lomsadze, 2011). The GeneMark-ES v2 algorithm does not require a predetermined training set to estimate GHMM parameters. Rather, the anonymous genomic sequence is used as an input for iterative unsupervised learning (Ter-Hovhannisyan *et al.,* 2008). Because *ab initio* methods are more sensitive than extrinsic methods, these properties make both tools indispensable in genome annotation pipelines.

Spacing differentOnce again various implementations in the Funannotate pipeline were used to perform functional annotation. The eggNOG-mapper v1.0.3.3-g3e22728 (Huerta-Cepas *et al.,* 2019) and Interproscan v5.30-69.0 (Jones *et al.,* 2014) databases were used for functional

domain annotation. This was followed by an HMMer search for Carbohydrate-Active enZymes (CAZYmes) against the dbCAN v7.0 database (Terrapon *et al.*, 2017) and a BLASTP search against the MEROPS database v12.0 to identify proteases (Rawlings *et al.,* 2016). Signal peptides were predicted using SignalP v4.1 (Armenteros *et al.,* 2019), and secondary metabolites were discovered using antiSMASH v4.1.0 (Blin *et al.,* 2019). The tRNA genes were found using tRNAscan-SE v1.23 (Lowe and Eddy, 1997, and the fungal transcription factors were predicted using the Fungal Transcription Factor Databases (Park *et al.,* 2008). The listed databases are an integral part of the pipelines as they enable annotation of the genome from a variety of different databases that cover various aspects of organism biology and the putative function of predicted proteins. The findings shed light on the diverse biological activities occurring within microorganisms and pave the way for comparative analysis.

2.3.7 Pathogenicity genes in the genome of *Naganishia randhawae* eABCC1

The genus *Naganishia* incorporates important human pathogens such as *N. albida* (Choe *et al.,* 2019), *N. diffluens* (Serda Kantarcioğlu *et al.*, 2007), *N. friedmannii* (Ekhtiari *et al.,* 2017) and *N. uzbekistanensis* (Powel *et al.,* 2012). To determine the presence of pathogenicity-related genes in *N. randhawae* eABCC1, the protein dataset was compared against the pathogen-host interactions database (PHI-base v4.6). PHI-base incorporates curated genes and proteins with strong experimental evidence for a role in pathogen-host interactions (Urban *et al.,* 2015; Winnenburg *et al.,* 2008). The fungal pathogenicity proteins dataset was retrieved from the PHI-database (http://www.phi-database.org) and a BLASTP (Shiryev *et al.,* 2007) search of the predicted proteome was performed. Proteins which shared > 70 % amino acid identity and protein alignment > 70 % of the protein length were considered as orthologues of putative pathogenicity proteins in *N. randhawae* eABCC1.

2.4 Results and discussion

2.4.1 Motivation for genome sequencing and assembly

The yeast strain eABCC1 isolated from avian guano was selected for whole-genome sequencing because melanin was observed when cultured on BSA (Yarrow, 1998) (Figure 2.1). Avian guano is known to harbour several melanin-producing yeasts including the well-known pathogenic cryptococci *Cryptococcus deneoformans* and *C. neoformans* (Moschetti *et al.,* 2017; Wu *et al.,* 2012). Fungal melanins are generally localized in the cell wall where they

participate in a variety of physiological roles including the defence against UV radiation, the sequestering of non-specific compounds and neutralization of radicals, including reactive oxygen species (ROS) (Smith and Casadevall, 2019). As such, melanized fungi are often isolated from extreme environments and are successful pathogens (Cordero and Casadevall, 2017). To our knowledge, this is the first time that melanin production has been reported in a *N. randhawae* strain.

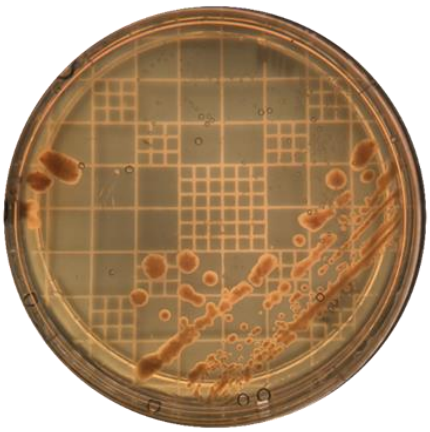


Figure 2. 1 *Naganishia randhawae* eABCC1 produces a melanin-like brown pigment when cultured on bird Seed Agar (Yarrow, 1998).

2.4.2 Taxonomic placement of *Naganishia randhawae* eABCC1

The ITS regions of fungal (rDNA are highly variable sequences that are commonly used as molecular markers for the identification of fungal species (Martin and Rygiewicz, 2005). The ITS region from *N. randhawae* eABCC1 was amplified and sequenced. Comparison of the ITS nucleotide sequence against the NCBI nucleotide database revealed that strain eABCC1 belongs to the species *N. randhawae* based on a 96.22 % nucleotide identity with type strain, *N. randhawae* CBS 10160 (Accession number : AJ876528.1) (Khan *et al.*, 2010), and 99.21 % nucleotide identity with *N. randhawae* strain IWBT-Y823 (Accession number : JQ993379). Phylogenetic analysis of the ITS sequence from the yeast strain eABCC1 confirmed the taxonomic identity as *N. randhawae* as it clustered with *N. randhawae* CBS10160 (Figure 2.2).

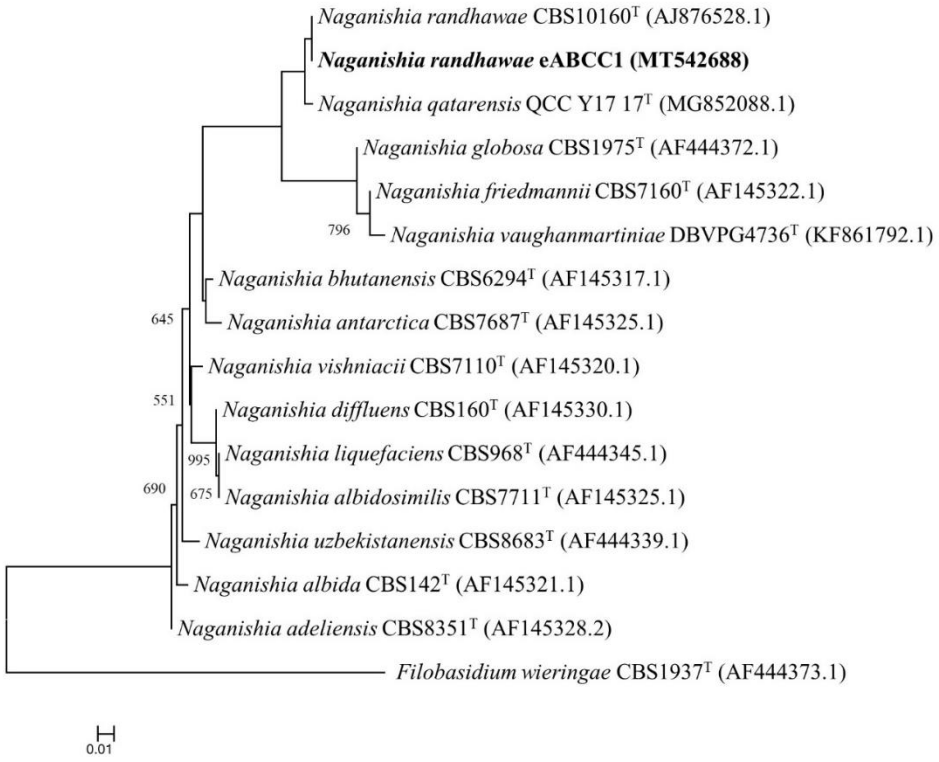


Figure 2.2 Maximum Likelihood (ML) phylogeny showing the taxonomic placement of *Naganishia randhawae* eABCC1 (marked in bold). Type strains are indicated by superscript T. Sequences not generated during this study were obtained from GenBank (accession numbers are shown in parentheses). The trimmed alignment encompassed 621 nucleotide positions. The tree was constructed based on the internal transcribed spacer (ITS) sequences using PhyML-SMS (Guindon *et al.* 2010), with the best fit evolutionary model HKY85 +G. Bootstrap support values (n = 1,000 replicates) greater than 500 are indicated at the nodes. Bars indicate the number of nucleotide substitutions per site.

2.4.3 *Naganishia randhawae* eABCC1 genome metrics

The assembled genome of *N. randhawae* eABCC1 is comprised of 386 scaffolds with a total size of 20,271,596 bp and an average G+C content of 51.71 %. The largest and N_{50} scaffold are 793,440 bp and 191,697 bp in size, respectively. These metrics are in line with other members of the genus *Naganishia* which are discussed in greater detail in Chapter 3 (Bijlani *et al.,* 2020; Han *et al.,* 2020; Nizovoy *et al.,* 2020; Vajpeyi and Chandran, 2016). The genome incorporated 86.8 % of the Basidiomycota BUSCO gene models (86.4 % of these were single copy complete and only 0.4 % were duplicated) (Table 2.2).

Table 2.2 Whole-genome assembly metrics of *Naganishia randhawae* eABCC1 generated using the Funannotate pipeline v1.5.1-93c317b (Palmer, 2016).

Features	Values
Assembly Size (Mb)	20.27
Largest Scaffold (kb)	793,44
Average Scaffold (kb)	52.52
Scaffold	386
Scaffold N50 (kb)	191,7
GC- Content (%)	51.71
Num. Genes	6,943
Num. Proteins	6,775
Num. tRNA	168
BUSCO Values (%)	86.8

2.4.4 Genome annotation of *Naganishia randhawae* eABCC1

Genome annotation with the Funannotate Pipeline v1.5.1-93c317b (Palmer, 2016) predicted a total of 6,943 genes including 6,775 protein-coding sequences and 168 tRNAs.

Functional annotation of the protein coding regions using eggNOG revealed that 4,076 (~60 %) of the 6,775 predicted proteins were assigned to functional categories (Huerta-Cepas *et al.,* 2019) (Figure 2.3). Most (27 %) of the functionally annotated proteins (1,510) belong to the cluster of orthologous groups (COG) category "function unknown" (S). Proteins with known functions predominantly belong to the COG categories posttranslational modification protein turnover chaperones (8 %) and carbohydrate transport and metabolism (7 %). *Naganishia randhawae* eABCC1 has a COG distribution similar to what has been reported in both bacteria and fungi (Liu *et al.,* 2012; Wang *et al.,* 2018).

2.4.4.1 The genome of *Naganishia randhawae* eABCC1 harbours laccases that may be linked to melanin production

Fungal melanins can be synthesized via the 1,8-dihydroxynaphthalene (DHN) or L-3,4-dihydroxyphenylalanine (L-dopa) pathway (Eisenman and Casadevall, 2012). Laccases are the largest subclass of the multicopper oxidase (MCO) family and are the main enzyme of the L-DOPA melanin production pathway (Mansur *et al.*, 1998; Thurston, 1994). They play a

crucial role by facilitating the oxidation of the hydroxyl groups on catecholamines which eventually leads to the formation of quinone (Smith and Casadevall, 2019). Subsequently, the produced quinone undergoes further oxidation and forms dihydroxyindole, which polymerizes into black or brown polymers (García-Borrón and Sánchez, 2011).

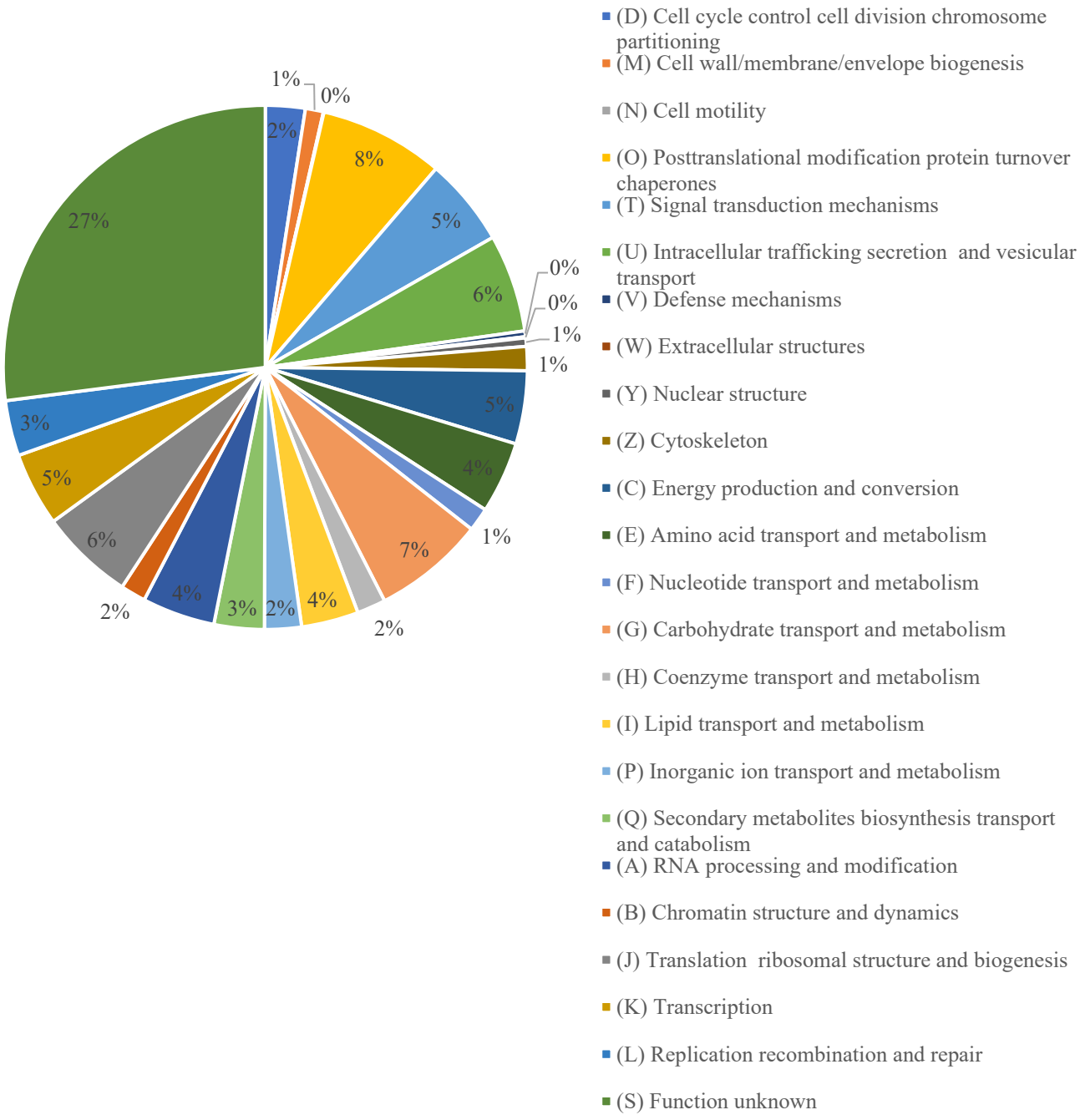


Figure 2.3 Distribution of Clusters of Orthologous Groups (COG: Huerta-Cepas *et al.,* 2019) functional categories and their relative abundances in the genome of *Naganishia randhawae* eABCC1. The pie chart demonstrates the abundance of each subsystem category, and the count of each subsystem feature is shown in as a percentage value.

Two putative laccases (Locus tags: EABCC1_001420 and EABCC1_006080) were identified in the genome of *N. randhawae* eABCC1. These share 56.07 % and 37.15 % amino acid identity, respectively, to the laccase enzyme (KEP53184.1) in the melanin-producing fungal plant pathogen *Rhizoctonia solani* (Shu *et al.,* 2019). Comparison with laccase from *C. neoformans* (OXL07050.1) revealed a 31.80 % and 23.91 % amino acid identity for EABCC1_001420 and EABCC1_006080, respectively.

2.4.5 Pathogen-Host Interaction (PHI) analysis of *Naganishia randhawae* eABCC1

A BLASTP (> 70 % identity) search against the fungal PHI dataset (Winnenburg *et al.,* 2008) identified a total of 64 orthologues across ten fungal pathogens (Table 2.3). The highest number of orthologues found in strain eABCC1 were from the human pathogen *C. neoformans*, of which six orthologues are associated with capsule synthesis (Table 2.3). In a host, the capsule plays multiple protective roles, including inhibiting phagocytosis by macrophages and preventing leukocyte migration to the infection site (Janbon, 2004; Zaragoza *et al.,* 2009). Changes in environmental conditions such as accumulation of carbon dioxide, iron deprivation and increased pH levels can also induce capsule formation (Mylonakis *et al.,* 2002). When guano decays, nutrients become scarce, and uric acid and ammonia levels rise, creating a hostile environment for microorganisms (Crouch *et al.,* 2020). The production of capsule may aid *N. randhawae* in surviving the adverse conditions in bird droppings.

The second largest number of orthologues were identified in the fungal plant pathogen *Fusarium graminearum*, which is responsible for fusarium head blight disease in wheat and the production of mycotoxins that are harmful to animals and humans when ingested (Cambaza, 2018; Malz *et al.,* 2005). In this study, eight out of the sixteen orthologues identified from *F. graminearum* are transcription factors (TFs) (Table 2.3). The deletion of these TFs in *F. graminearum* has been linked with reduced virulence (Son *et al.,* 2011). Regulation of transcription is known to be essential for the growth, development, infection, and pathogenicity of a variety of fungal pathogens (Guo *et al.* 2011; Liu *et al.* 2015).

Table 2.3 Orthologues identified in *Naganishia randhawae* eABCC1 using the Pathogen-Host Interaction (PHI) database (Winnenburg *et al.*, 2008).

PHI Gene Name	**Orthologues**	**Product**	**Species**	**Gene knock-out result**
Glca	PP1	Serine/threonine-protein phosphatase pp1	*A. fumigatus*	lethal
Aspc		Hypothetical protein	*A. fumigatus*	unaffected pathogenicity increased virulence (hypervirulence)
Cyca	CYC1	Iso-1-cytochrome c	*A. fumigatus*	reduced virulence
Rps41	RPS4	40S ribosomal protein S4	*Candida albicans*	reduced virulence
Arf2		Hypothetical protein	*C. albicans*	reduced virulence
Ubp5		Hypothetical protein	*Cryptococcus gattii*	reduced virulence
Sod1	SOD1	Superoxide dismutase [Cu-Zn]	*C. gattii*	reduced virulence loss of pathogenicity
Ugd1	UGD1	UDP-glucose 6-dehydrogenase 1	*C. neoformans*	loss of pathogenicity
Trr1	TRR1	Thioredoxin-disulfide reductase	*C. neoformans*	lethal
Usx1	UXS1	UDP-glucuronic acid decarboxylase 1	*C. neoformans*	loss of pathogenicity
Uxt1		Hypothetical protein	*C. neoformans*	reduced virulence unaffected pathogenicity
Ccn1	CLF1	NineTeen Complex (NTC) component	*C. neoformans*	reduced virulence
Cap59	CAP59	Capsular associated protein	*C. neoformans*	loss of pathogenicity reduced virulence
Met3	MET3	Sulfate adenylyltransferase	*C. neoformans*	loss of pathogenicity
Pka2	PKA2	Serine/threonine protein kinase, AGC	*C. neoformans*	reduced virulence
Lys4	LYS4 1	Mitochondrial Homoaconitase	*C. neoformans*	reduced virulence
Imd1	IMD1	Inosine-5'-monophosphate dehydrogenase	*C. neoformans*	reduced virulence
Vps4	VPS4	Vacuolar protein sorting-associated protein 4	*C. neoformans*	loss of pathogenicity
Pck1	PCK1	Protein kinase C-like 1	*C. neoformans*	reduced virulence
Pmt2	PMT2	Protein O-mannosyltransferase 2	*C. neoformans*	increased virulence (hypervirulence) lethal

PHI Gene Name	Orthologues	Product	Species	Gene knock-out result
Gua1	GUA1	GMP synthase (glutamine-hydrolyzing)	*C. neoformans*	loss of pathogenicity
Vph1	VPH1	H(+)-transporting V0 sector ATPase subunit a	*C. neoformans*	reduced virulence
Pka1	PKA1	Camp-dependent protein kinase catalytic subunit	*C. neoformans*	reduced virulence loss of pathogenicity
Ilv2	ILV2	Acetolactate synthase, mitochondrial	*C. neoformans*	loss of pathogenicity
Nth1	NTH1 2	Alpha, alpha-trehalase nth1	*C. neoformans*	increased virulence (hypervirulence)
Ras1	RAS1	Ras GTPase	*C. neoformans*	reduced virulence loss of pathogenicity
Pho84	PHO84	Inorganic phosphate transporter pho84	*C. neoformans*	reduced virulence
Uut1 (cnag 06230)	DRP1	UAA transporter	*C. neoformans*	loss of pathogenicity
Ade13	ADE13	Adenylosuccinase ade13	*C. neoformans*	reduced virulence
Gzob006	RPT4	26S proteasome subunit rpt4	*Fusarium graminearum*	lethal
Acl1	ACL1	Citrate synthase	*F. graminearum*	loss of pathogenicity
Gzob003		Hypothetical protein	*F. graminearum*	unaffected pathogenicity
Gzob045	RPS23	40S ribosomal protein S23	*F. graminearum*	unaffected pathogenicity
Elp3	ELP3	Elongator subunit	*F. graminearum*	reduced virulence
Fgsg 00677	CKA1	Casein kinase II subunit alpha	*F. graminearum*	lethal
(sc rim15)		Hypothetical protein	*F. graminearum*	reduced virulence
Fgsg 08731	HRR25	Serine/threonine protein kinase	*F. graminearum*	lethal
Gzc2h103		Hypothetical protein	*F. graminearum*	unaffected pathogenicity
Nos1	NUO49	NADH:ubiquinone oxidoreductase 49kD subunit	*F. graminearum*	reduced virulence
Gzc2h048	PYC1	Pyruvate carboxylase	*F. graminearum*	unaffected pathogenicity
(sc yck1/2/3)	YCK1	Casein kinase I	*F. graminearum*	reduced virulence
Gzob024	RPL2A	60S ribosomal protein L2A	*F. graminearum*	unaffected pathogenicity

PHI Gene Name	Orthologues	Product	Species	Gene knock-out result
Gzccaat008	HTA1	Histone H2A	*F. graminearum*	unaffected pathogenicity
(sc sky1)	PKC1	Serine/threonine kinase	*F. graminearum*	reduced virulence
Gzob030	TIF11	Translation initiation factor 1A	*F. graminearum*	lethal
Gbe1	GLC3	Alpha-1,4-glucan branching enzyme	*Fusarium oxysporum*	unaffected pathogenicity
Fvhap2	HAP2	Transcriptional activator	*Fusarium verticillioides*	reduced virulence
Modnm1	DNM1	Dynamin- GTPase protein	*Magnaporthales oryzae*	reduced virulence
Cam		Hypothetical protein	*M. oryzae*	reduced virulence loss of pathogenicity
Mgg 04521.6	WC2	Blue light receptor	*M. oryzae*	reduced virulence
Moccp1	CCP1	Heme peroxidase	*M. oryzae*	reduced virulence
Mgg 00383	SAM2	Methionine adenosyltransferase sam2	*M. oryzae*	reduced virulence
Morad6	UBC2	Ubiquitin-conjugating enzyme E2 2	*M. oryzae*	reduced virulence
Moypt7		Hypothetical protein	*M. oryzae*	loss of pathogenicity
Molys20	LYS4 2	Mitochondrial Homoaconitase	*M. oryzae*	reduced virulence
Ubc2	STE50	Protein ste50	*Ustilago maydis*	reduced virulence loss of pathogenicity
Clb2		Hypothetical protein	*U. maydis*	reduced virulence
Upa2	PPH1	Serine/threonine-protein phosphatase PP2A catalytic subunit	*U. maydis*	unaffected pathogenicity
Ip	SDH2	Succinate dehydrogenase complex, subunit B	*Zymoseptoria tritici*	chemistry target: resistance to chemical
Sdi1	SDH1	Succinate dehydrogenase flavoprotein subunit	*Z. tritici*	unaffected pathogenicity

2.5 Conclusion

Here we have presented the draft genome sequence of *N. randhawae* eABCC1. The genomic content of *N. randhawae* eABCC1 revealed the presence of two putative laccases which may be linked to the production of melanin observed on BSA. We also identified the presence of putative pathogenicity determinants using the PHI database. While these characteristics have been linked to host infection, they likely help *N. randhawae* eABBC1 survive in adverse environments such as avian guano. The complete genome sequence of *N. randhawae* eABCC1 will enable a deeper molecular understanding of *N. randhawae* and the genus *Naganishia* as a whole by means of a comparative genome analysis.

2.6 References

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Loman, N.J., Andersson, A.F. and Quince, C., 2013. Concoct: clustering contigs on coverage and composition. arXiv. *arXiv preprint arXiv:1312.4038*, *10*.

Andrews, S., 2010. Babraham bioinformatics-FastQC a quality control tool for high throughput sequence data. *Babraham Inst*.

Anton, B., Sergey, N., Dmitry, A., Alexey, A.G., Mikhail, D., Alexander, S.K., Valery, M.L., Sergey, I.N., Son, P., Andrey, D.P. and Alexey, V.P., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, *19*(5), pp.455-477.

Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), pp.420-423.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), pp.455-477.

Baxter, L.A. and Harche, F., 1992. Note: On the greedy algorithm for optimal assembly. *Naval Research Logistics (NRL)*, *39*(6), pp.833-837.

Behjati, S. and Tarpey, P.S., 2013. What is next generation sequencing?. *Archives of Disease in Childhood-Education and Practice*, *98*(6), pp.236-238.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H. and Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), pp. 81-87.

Borodovsky, M., Lomsadze, A., Ivanov, N. and Mills, R., Eukaryotic gene prediction using GeneMark. hmm. *Current Protocols in Bioinformatics*.

Burnik, C., Altintaş, N.D., Özkaya, G., Serter, T., Selcuk, Z.T., Firat, P., Arikan, S., Cuenca-Estrella, M. and Topeli, A.R.Z.U., 2007. Acute respiratory distress syndrome due to Cryptococcus albidus pneumonia: case report and review of the literature. *Medical mycology*, *45*(5), pp.469-473.

Cambaza, E., 2018. Comprehensive description of *Fusarium graminearum* pigments and related compounds. *Foods*, 7(10), p.165.

Choe, Y.J., Blatt, D.B., Yalcindag, A., Geffert, S.F., Bobenchik, A.M. and Michelow, I.C., 2020. *Cryptococcus albidus* fungemia in an immunosuppressed child: case report and systematic literature review. *Journal of the Pediatric Infectious Diseases Society*, 9(1), pp.100-105.

Choudhuri, S., 2014. Additional bioinformatic analyses involving protein sequences. In *Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools* (pp. 183-207). Academic Press, Cambridge, MA.

Compeau, P.E., Pevzner, P.A. and Tesler, G., 2011. How to apply de Bruijn graphs to genome assembly. *Nature biotechnology*, *29*(11), pp.987-991.

Conway, T.C. and Bromage, A.J., 2011. Succinct data structures for assembling large genomes. *Bioinformatics*, *27*(4), pp.479-486.

Cordero, R.J. and Casadevall, A., 2017. Functions of fungal melanin beyond virulence. *Fungal Biology Reviews*, 31(2), pp.99-112.

Costello, E.K., Halloy, S.R., Reed, S.C., Sowell, P. and Schmidt, S.K., 2009. Fumarole-supported islands of biodiversity within a hyperarid, high-elevation landscape on Socompa Volcano, Puna de Atacama, Andes. *Applied and Environmental Microbiology*, *75*(3), pp.735-747.

Crouch, N.M., Lynch, V.M. and Clarke, J.A., 2020. A re-evaluation of the chemical composition of avian urinary excreta. *Journal of Ornithology*, 161(1), pp.17-24.

Cuomo, C.A. and Birren, B.W., 2010. The fungal genome initiative and lessons learned from genome sequencing. *Methods in Enzymology*, 470, pp.833-855.

Dominguez, V.D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O.V., Amselem, J., Bouri, L., Bocs, S., Klopp, C. and Gibrat, J.F., 2018. Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, *7*.

Eisenman, H.C. and Casadevall, A., 2012. Synthesis and assembly of fungal melanin. *Applied microbiology and Biotechnology*, 93(3), pp.931-940.

Ekhtiari, M., Farahyar, S., Falahati, M., Razmjou, E., Ashrafi-Khozani, M., Ghasemi, Z. and Abbasi-Nejat, Z., 2017. The first report of onychomycosis caused by *Cryptococcus friedmannii* (*Naganishia friedmannii*) a basidiomycetous yeast. *Medical Mycology Case Reports*, 15, pp.25-27.

Farrer, R.A., Kemen, E., Jones, J.D. and Studholme, D.J., 2009. De novo assembly of the *Pseudomonas syringae pv. syringae* B728a genome using Illumina/Solexa short sequence reads. *FEMS Microbiology Letters*, 291(1), pp.103-111.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, *17*(6), pp.368-376.

Gabaldón, T., 2020. Grand Challenges in Fungal Genomics and Evolution. *Frontiers in Fungal Biology*, 1, p.2.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S. and Rehman, B., 2003. The genome sequence of the filamentous fungus Neurospora crassa. *Nature*, *422*(6934), pp.859-868.

Galagan, J.E., Henn, M.R., Ma, L.J., Cuomo, C.A. and Birren, B., 2005. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Research*, 15(12), pp.1620-1631.

García-Borrón, J.C. and Sánchez, M.C.O., 2011. Biosynthesis of melanins. Melanins and Melanosomes: Biosynthesis, Structure, Physiological and Pathological Functions. *John Wiley & Sons*.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. and Louis, E.J., 1996. Life with 6000 genes. *Science*, 274(5287), pp.546-567.

Goolsby, E.W., 2017. Rapid maximum likelihood ancestral state reconstruction of continuous characters: A rerooting-free algorithm. *Ecology and evolution*, *7*(8), pp.2791-2797.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A. and Otillar, R., 2012. The genome portal of the department of energy joint genome institute. *Nucleic Acids Research*, 40(D1), pp. 26-32.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A. and Otillar, R., 2012. The genome portal of the department of energy joint genome institute. *Nucleic acids research*, *40*(D1), pp.D26-D32.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), pp.307-321.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, *59*(3), pp.307-321.

Guo, M., Chen, Y., Du, Y., Dong, Y., Guo, W., Zhai, S., Zhang, H., Dong, S., Zhang, Z., Wang, Y. and Wang, P., 2011. The bZIP transcription factor MoAP1 mediates the oxidative stress response and is critical for pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog*, 7(2), p.e1001302.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symp. Ser.* (Vol. 41, pp. 95-98).

Hall, T.A., 1999, January. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *In Nucleic Acids Symposium Series* (Vol. 41, No. 41, pp. 95-98.

Heydari, M., Miclotte, G., Demeester, P., Van de Peer, Y. and Fostier, J., 2017. Evaluation of the impact of Illumina error correction tools on de novo genome assembly. *BMC bioinformatics*, *18*(1), pp.1-13.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. and von Mering, C., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47, pp.309-314.

Huson, D.H., Reinert, K. and Myers, E.W., 2002. The greedy path-merging algorithm for contig scaffolding. *Journal of the ACM (JACM)*, *49*(5), pp.603-615.

Idury, R. and Waterman, M.S., 1995. A new algorithm for shotgun sequencing. *J. Comp. Biol*, *2*, pp.291-306.

Janbon, G., 2004. *Cryptococcus neoformans* capsule biosynthesis and regulation. *FEMS Yeast Research*, 4(8), pp.765-771.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L., 2008. NCBI BLAST: a better web interface. *Nucleic acids research*, *36*(suppl_2), pp.W5-W9.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. and Pesseat, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), pp.1236-1240.

Khan, Z.U., Ahmad, S., Hagen, F., Fell, J.W., Kowshik, T., Chandy, R. and Boekhout, T., 2010. *Cryptococcus randhawai* sp. nov., a novel anamorphic basidiomycetous yeast isolated from tree trunk hollow of *Ficus religiosa* (peepal tree) from New Delhi, India. *Antonie Van Leeuwenhoek*, 97(3), pp.253-259.

Lee, Y.A., Kim, H.J., Lee, T.W., Kim, M.J., Lee, M.H., Lee, J.H. and Ihm, C.G., 2004. First report of Cryptococcus albidus-induced disseminated cryptococcosis in a renal transplant recipient. *The Korean journal of internal medicine*, *19*(1), p.53.

Lefort, V., Longueville, J.E. and Gascuel, O., 2017. SMS: smart model selection in PhyML. *Molecular Biology and Evolution*, 34(9), pp.2422-2424.

Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B. and Yang, B., 2012. Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. *Briefings in functional genomics*, *11*(1), pp.25-37.

Liao, X., Li, M., Zou, Y., Wu, F.X. and Wang, J., 2019. Current challenges and solutions of de novo assembly. *Quantitative Biology*, 7(2), pp.90-109.

Liu, D., Gong, J., Dai, W., Kang, X., Huang, Z., Zhang, H.M., Liu, W., Liu, L., Ma, J., Xia, Z. and Chen, Y., 2012. The genome of *Ganderma lucidum* provide insights into triterpense biosynthesis and wood degradation. *PloS One*, 7(5), p.e36146.

Liu, H., Zhang, S., Ma, J., Dai, Y., Li, C., Lyu, X., Wang, C. and Xu, J.R., 2015. Two Cdc2 kinase genes with distinct functions in vegetative and infectious hyphae in *Fusarium graminearum*. *PLoS Pathogens*, 11(6), p.e1004913.

Liu, X.Z., Wang, Q.M., Göker, M., Groenewald, M., Kachalkin, A.V., Lumbsch, H.T., Millanes, A.M., Wedin, M., Yurkov, A.M., Boekhout, T. and Bai, F.Y., 2015. Towards an integrated phylogenetic classification of the Tremellomycetes. *Studies in mycology*, *81*, pp.85-147.

Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A. and Allen, J.E., 2005. The genome of the

basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*, 307(5713), pp.1321-1324.

Lynch, R.C., King, A.J., Farías, M.E., Sowell, P., Vitry, C. and Schmidt, S.K., 2012. The potential for microbial life in the highest-elevation (> 6000 masl) mineral soils of the Atacama region. *Journal of Geophysical Research: Biogeosciences*, *117*(G2).

Malz, S., Grell, M.N., Thrane, C., Maier, F.J., Rosager, P., Felk, A., Albertsen, K.S., Salomon, S., Bohn, L., Schäfer, W. and Giese, H., 2005. Identification of a gene cluster responsible for the biosynthesis of aurofusarin in the *Fusarium graminearum* species complex. *Fungal Genetics and Biology*, 42(5), pp.420-433.

Mansur, M., Suárez, T. and González, A.E., 1998. Differential gene expression in the laccase gene family from Basidiomycete I-62 (CECT 20197). *Applied and Environmental Microbiology*, 64(2), pp.771-774.

Martin, K.J. and Rygiewicz, P.T., 2005. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiology*, 5(1), pp.1-11.

Moretti, S., Armougom, F., Wallace, I.M., Higgins, D.G., Jongeneel, C.V. and Notredame, C., 2007. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Research*, 35(suppl_2), pp. 645-648.

Morozova, O. and Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5), pp.255-264.

Moschetti, G., Alfonzo, A. and Francesca, N., 2017. Yeasts in birds. In Yeasts in natural ecosystems: diversity (pp. 435-454). *Springer*, Cham.

Myers Jr, E.W., 2016. A history of DNA sequence assembly. *It-Information Technology*, *58*(3), pp.126-132.

Mylonakis, E., Ausubel, F.M., Perfect, J.R., Heitman, J. and Calderwood, S.B., 2002. Killing of *Caenorhabditis elegans* by *Cryptococcus neoformans* as a model of yeast pathogenesis. *Proceedings of the National Academy of Sciences*, 99(24), pp.15675-15680.

Nicol, J.W., Helt, G.A., Blanchard Jr, S.G., Raja, A. and Loraine, A.E., 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, 25(20), pp.2730-2731.

Palmer JM., 2016. Funannotate: a fungal genome annotation and comparative genomics pipeline. Available at: https://github.com/nextgenusfs/funannotate.

Park, J., Park, J., Jang, S., Kim, S., Kong, S., Choi, J., Ahn, K., Kim, J., Lee, S., Kim, S. and Park, B., 2008. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics*, 24(7), pp.1024-1025.

Powel, M.S., Alizadeh, A.A., Budvytiene, I., Schaenman, J.M. and Banaei, N., 2012. First isolation of *Cryptococcus uzbekistanensis* from an immunocompromised patient with lymphoma. *Journal of Clinical Microbiology*, 50(3), pp.1125-1127.

Pulschen, A.A., Rodrigues, F., Duarte, R.T., Araujo, G.G., Santiago, I.F., Paulino-Lima, I.G., Rosa, C.A., Kato, M.J., Pellizari, V.H. and Galante, D., 2015. UV-resistant yeasts isolated from a high-altitude volcanic area on the Atacama Desert as eukaryotic models for astrobiology. *Microbiologyopen*, *4*(4), pp.574-588.

Ragupathi, L. and Reyna, M., 2015. Case report of Cryptococcus albidus peritonitis in a peritoneal dialysis patient and a review of the literature. *Peritoneal Dialysis International*, *35*(4), pp.421-427.

Rawlings, N.D., Barrett, A.J. and Finn, R., 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44(D1), pp. 343-D350.

Sanger, F., Nicklen, S. and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), pp.5463-5467.

SanMiguel, P., 2011. Next-generation sequencing and potential applications in fungal genomics. *Fungal Genomics*, pp. 51-60. Humana Press.

Schmidt, S.K., Vimercati, L., Darcy, J.L., Arán, P., Gendron, E.M., Solon, A.J., Porazinska, D. and Dorador, C., 2017. A Naganishia in high places: functioning populations or dormant cells from the atmosphere?. *Mycology*, *8*(3), pp.153-163.

Seppey, M., Manni, M. and Zdobnov, E.M., 2019. BUSCO: assessing genome assembly and annotation completeness. In *Gene prediction* (pp. 227-245). Humana, New York, NY.

Serda Kantarcioğlu, A., Boekhout, T., Sybren De Hoog, G., Theelen, B., Yücel, A., Ekmekci, T.R., Fries, B.C., Ikeda, R., Koslu, A. and Altas, K., 2007. Subcutaneous cryptococcosis due

to *Cryptococcus diffluens* in a patient with sporotrichoid lesions case report, features of the case isolate and in vitro antifungal susceptibilities. *Medical Mycology*, 45(2), pp.173-181.

Sherman, F., 1991. [1] Getting started with yeast. *Methods in enzymology*, *194*, pp.3-21.

Shiryev, S.A., Papadopoulos, J.S., Schäffer, A.A. and Agarwala, R., 2007. Improved BLAST searches using longer words for protein seeding. *Bioinformatics*, 23(21), pp.2949-2951.

Shu, C., Zhao, M., Anderson, J.P., Garg, G., Singh, K.B., Zheng, W., Wang, C., Yang, M. and Zhou, E., 2019. Transcriptome analysis reveals molecular mechanisms of sclerotial development in the rice sheath blight pathogen *Rhizoctonia solani* AG1-IA. *Functional & Integrative Genomics*, 19(5), pp.743-758.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), pp.3210-3212.

Smith, D.F. and Casadevall, A., 2019. The role of melanin in fungal pathogenesis for animal hosts. *Fungal Physiology and Immunopathogenesis*, pp.1-30.

Solon, A.J., Vimercati, L., Darcy, J.L., Arán, P., Porazinska, D., Dorador, C., Farías, M.E. and Schmidt, S.K., 2018. Microbial communities of high-elevation fumaroles, penitentes, and dry tephra “soils” of the Puna de Atacama volcanic zone. *Microbial ecology*, *76*(2), pp.340-351.

Son, H., Seo, Y.S., Min, K., Park, A.R., Lee, J., Jin, J.M., Lin, Y., Cao, P., Hong, S.Y., Kim, E.K. and Lee, S.H., 2011. A phenome-based functional analysis of transcription factors in the cereal head blight fungus, *Fusarium graminearum*. *PLoS Pathog*, 7(10), p.e1002310.

Son, H., Seo, Y.S., Min, K., Park, A.R., Lee, J., Jin, J.M., Lin, Y., Cao, P., Hong, S.Y., Kim, E.K. and Lee, S.H., 2011. A phenome-based functional analysis of transcription factors in the cereal head blight fungus, Fusarium graminearum. *PLoS pathogens*, *7*(10), p.e1002310.

Staden, R., 1979. A strategy of DNA sequencing employing computer programs. *Nucleic acids research*, 6(7), pp.2601-2610.

Stanke, M., Tzvetkova, A. and Morgenstern, B., 2006. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biology*, 7(1), pp.1-8.

Sturm, M., Schroeder, C. and Bauer, P., 2016. SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC bioinformatics*, *17*(1), pp.1-7.

Sullivan, J., 2005. Maximum-likelihood methods for phylogeny estimation. *Methods in enzymology*, *395*, pp.757-779.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N. and Rao, B.S., 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4(1), pp.1-14.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O. and Borodovsky, M., 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome research*, *18*(12), pp.1979-1990.

Terrapon, N., Lombard, V., Drula, E., Coutinho, P.M. and Henrissat, B., 2017. The CAZy database/the carbohydrate-active enzyme (CAZy) database: principles and usage guidelines. In *A Practical Guide to Using Glycomics databases* pp. 117-131.

Thurston, C.F., 1994. The structure and function of fungal laccases. *Microbiology*, 140(1), pp.19-26.

Urban, M., Irvine, A.G., Cuzick, A. and Hammond-Kosack, K.E., 2015. Using the pathogen-host interactions database (PHI-base) to investigate plant pathogen genomes and genes implicated in virulence. *Frontiers in Plant Science*, 6, pp.605.

Wang, Y., Zeng, X. and Liu, W., 2018. De novo transcriptomic analysis during Lentinula edodes fruiting body growth. *Gene*, 641, pp.326-334.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. and Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, *35*(3), pp.543-548.

Waterhouse, R.M., Zdobnov, E.M. and Kriventseva, E.V., 2011. Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome biology and evolution*, *3*, pp.75-86.

White, T.J., Bruns, T., Lee, S.J.W.T. and Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: a guide to methods and applications*, 18(1), pp.315-322.

Williamson, P.R., 1997. Laccase and melanin in the pathogenesis of *Cryptococcus neoformans*. *Front. Biosci*, 2, pp. e99-e107.

Winnenburg, R., Urban, M., Beacham, A., Baldwin, T.K., Holland, S., Lindeberg, M., Hansen, H., Rawlings, C., Hammond-Kosack, K.E. and Köhler, J., 2007. PHI-base update: additions to the pathogen–host interaction database. *Nucleic Acids Research*, 36(suppl_1), pp. D572-D576.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S. and Basham, D., 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 415(6874), pp.871-880.

Wu, Y., Du, P.C., Li, W.G. and Lu, J.X., 2012. Identification and molecular analysis of pathogenic yeasts in droppings of domestic pigeons in Beijing, China. *Mycopathologia*, 174(3), pp.203-214.

Yarrow, D., 1998. Methods for the isolation, maintenance and identification of yeasts. *The yeasts* (pp. 77-100). Elsevier.

Zaragoza, O., Rodrigues, M.L., De Jesus, M., Frases, S., Dadachova, E. and Casadevall, A., 2009. The capsule of the fungal pathogen *Cryptococcus neoformans*. *Advances in Applied Microbiology*, 68, pp.133-216.

Zhu, X. and Williamson, P.R., 2004. Role of laccase in the biology and virulence of *Cryptococcus neoformans*. *FEMS Yeast Research*, 5(1), pp.1-10.

Chapter 3 Comparative genomic analysis of *Naganishia* reveals a range of molecular features involved in stress tolerance and adaptation

3.1 Abstract

Members of the genus *Naganishia* have been isolated from diverse ecosystems including avian guano, desert soil, glaciers, and hypersaline waters. These environments are characterised by detrimental conditions that could affect their survival. As such, they have acquired adaptive strategies to overcome the stress associated with such unfavourable conditions. To gain insight into the adaptive mechanisms utilised in response to environmental stress we have performed a comparative genome analysis of four *Naganishia* strains, *N. albida* strain JCM2334 and NT2002, *N. randhawae* eABCC1 and *N. vishniacii* ANT03-052, with *Filobasidium wieringae* JCM11695 as the outgroup. From the annotated genomes, a broad range of proteins that are associated with stress tolerance were identified, including heat shock proteins and enzymes involved in trehalose biosynthesis. Orthovenn2 analysis revealed a total of 831 orthologous clusters shared among the *Naganishia* species. Gene ontology annotation of the largest number of proteins under a cluster shared by all *Naganishia* species but not the outgroup taxon *F. wieringae* was assigned to the function "UV-damage excision repair". This could constitute a protective mechanism against UV related damages for *Naganishia* species. Similarly, proteins associated with melanisation and photoprotection were also found, implying additional protection during solar irradiation exposure. These genomic features may contribute to the survival of *Naganishia* in a wide variety of environments.

3.2 Introduction

The kingdom Fungi has been at the forefront of eukaryotic genome studies since the sequencing of the first eukaryote genome of *Saccharomyces cerevisiae* in 1996 (Goffeau *et al.,* 1996). Subsequently, the completion of other yeast genomes such as *Neurospora crassa* and *Schizosaccharomyces pombe* marked the beginning of a new era in molecular mycology (Galagan *et al.,* 2003; Schoch *et al.,* 2014). Lessons learnt from the comparative analysis of these early budding yeasts paved the way for efficient genome analysis not only in fungi but in other eukaryotes (Liti and Louis, 2005). As a result of the advances made in Next Generation Sequencing (NGS) technologies, more than 7,896 fungal genome sequences are open to the public and thus provide a wealth of genomic data within the fungal kingdom (Galagan *et al.,* 2005).

Comparative genomics is a valuable tool for determining the genetic similarities and differences between two or more species at the whole genome level (Kawashima, 2019).

Identifying genetic regions that are conserved among different species is essential to understanding the whole genome itself. It reveals genes that are indispensable and highlights genomic signals that regulate gene expression across a wide variety of organisms (Piškur and Langkjær, 2004).

The genus *Naganishia* (class: Tremellomycetes, order: Filobasidiales) has recently been re-established and is currently comprised of seventeen yeast species that have a global distribution (Liu *et al.,* 2015a). Members of the genus have been isolated from habitats such as avian guano (Wu *et al.,* 2012), cold polar and non-polar regions (Schmidt *et al.,* 2017), desert soils (Goto and Sugiyama, 2011) and hypersaline waters (Fotedar *et al.,* 2018). Most habitats from which *Naganishia* species have been isolated have features that restrict microbial growth, suggesting the use of multiple adaption strategies to cope with the harsh conditions.

Cellular responses to external stress are particularly important models for understanding the biological function of microorganisms. Adaptation mechanisms to stress are generally rapid and highly complex, requiring the temporary coordination of a number of regulatory pathways (Lelandais and Devaux, 2010). Cold temperatures, elevated UV radiation, freeze-thaw cycles, high pH, high salinity, low availability of nutrients and low water activity impose enormous stress on *Naganishia* taxa inhabiting extreme environments (Nicholson *et al.,* 2005; Schmidt *et al.,* 2017). The factors that contribute to the survival of *Naganishia* species in these ecosystems are poorly understood.

The first genome sequence of *N. randhawae* eABCC1 isolated from avian guano was presented in Chapter 2. The genomic content of this strain revealed a repertoire of virulence related genes that could confer protection against environmental stress, for instance the melanin producing enzyme, laccase. Comparative genome analysis of *N. vishniacii* ANT05-03 with closely related species revealed features associated with extreme cold tolerance (Nizovoy *et al.*, 2020). Furthermore, the genome sequences of two *N. albida* strain NT2002 and JCM2334 have been sequenced. The former was isolated from dry desert soil in China (Yong *et al.,* 2016), while the latter was obtained from air in Japan. Here we have conducted comparative genomic analyses of these four *Naganishia* isolates which were derived from different environments, to elucidate the genetic determinants underlying their ability to withstand various types of environmental stress.

3.3 Materials and methods

3.3.1 Genome sequences, structural and functional annotation

The draft genome sequences of four *Naganishia* spp., namely *N. albida* JCM2334 (BCHV00000000.1), *N. albida* NT2002 (LLJT00000000.1) (Yong *et al.,* 2016), *N. randhawae* eABCC1 (Gen bank No. JABRPJ000000000.1) and *N. vishniacii* ANT03-052 (JABEVT010000001) (Nizovoy *et al.,* 2020) were included in this study. The draft genome sequence of *Filobasidium wieringae* JCM11695 (BCIV01000001) was also included as an outgroup. Genome completeness was evaluated using Benchmarking Universal Single-Copy Orthologues (BUSCO) v.2.0 with the basidiomycota_odb9 reference dataset (Simão *et al.,* 2015). Structural and functional annotation was performed using the Funannotate pipeline v1.5.1-93c317b (Palmer, 2016). This pipeline incorporates structural annotation algorithms such as Augustus v3.2.3 (Stanke *et al.,* 2006), GeneMark-ES v4.36 (Borodovsky and Lomsadze, 2011) as well as tools such as Repeatmasker v1.332 (Chen, 2004) and RepeatModeler v1.0.11 (Chen, 2004) to produce accurate gene models.

3.3.2 Pan-genome analysis

The Pan-genome determined using Proteinortho v5.16 (Lechner *et al.,* 2011) as part of the Funannotate pipeline. Proteinortho employs a BLAST-based approach to determine sets of orthologous proteins or nucleic acid sequences that generalizes the reciprocal best alignment heuristic. The output from Proteinortho was then fractioned into different segments. Proteins shared by all four strains were classified as the "core genome", proteins present in two or more strains were identified as the "accessory genome" and proteins only present in a single strain were categorised as "unique". COG annotation using eggNOG v3.0 of the core, accessory and unique proteins was retrieved from the Funannotate output file, and the proportion of each fraction was calculated.

3.3.3 Analysis of orthologous gene clusters

Orthologous genes are clusters of genes in different species that originated by vertical descent from a single gene in the last common ancestor (Fitch, 1970). They usually retain their original function throughout evolution and help establish relationships between the gene function, proteome structure and taxonomic classification (Koonin, 2005). Comparative analysis of orthologous clusters makes it possible to identify overlapping clusters across multiple species.

Identification of orthologous groups is advantageous for genome annotation, gene or protein evolution studies, and comparative genomics. Although successful methods for prokaryotic genome analysis have been developed, their application to eukaryotes has proven difficult (Koonin, 2005). This is largely due the fact that eukaryotic genomes frequently contain multiple paralog genes that have arisen through duplication events (Fitch, 1970; Rubin *et al.,* 2000) and as such, sequence information is often incomplete (Doolittle, 1995; Henikoff *et al.,* 1997). In response to these challenges, OrthoMCL was developed as an alternative approach for automated eukaryotic ortholog group identification (Li *et al.,* 2003). To distinguish functional redundancy from divergence, OrthoMCL identifies "recent" paralogs that belong in ortholog groups as within-species BLAST hits that are reciprocally better than between-species hits (Enright *et al.,* 2002).

For this study, the protein sequences from the annotated genomes were submitted to the OrthoVenn2 web server (https://orthovenn2.bioinfotoolkits.net) to identify and compare orthologous clusters (Xu *et al.,* 2019). An all-against-all BLASTP alignment using the OrthoMCL algorithm was performed to identify putative orthology (Östlund *et al.,* 2010). Disjoint clusters of closely related proteins were then generated with the Markov Clustering Algorithm (Dongen, 1998). The e-value cut-off for the all-to-all BLASTP alignmen was set to 0.05 and the inflation value for the generation of orthologous clusters using the Markov Cluster Algorithm was set to 1.5. The e-value describes the probability that a query sequence will find an equal or better match in the database (Kerfeld and Scott, 2011). An e-value of 0.0 indicates that zero sequences should match equally well or better than one another; the closer the e-value is to zero, the stronger the match. The inflation operator is a parameter in the MCL algorithm that influences the strengthening and weakening of currents (van Dongen and Abreu-Goodger, 2012). The inflation value ranges from 1.5 to 2.5 and influences cluster granularity. Xu and colleagues (2019) recommend the lowest inflation value to ensure better clustering quality.

Finally, a BLASTP analysis of protein sequences in each cluster against the non-redundant Uniprot protein database (UniProt Consortium, 2018) was performed to assign the gene ontology (GO) terms for biological process, molecular function, and cellular component categories.

3.3.4 Functional analysis of predicted coding genes using the Kyoto Encyclopaedia of Genes and Genomes (KEGG)

Protein sequences of the four *Naganishia* strains were analysed with KEGG (Ogata *et al.,* 1999) to identify important metabolic pathways involved in general stress response. Briefly, all four genomes were annotated using BlastKOALA (Kanehisa *et al.,* 2016). This is an automatic annotation server for genome and metagenome sequences that assigns KEGG ontology (KO) to the submitted sequences (Kanehisa *et al.,* 2016). The assigned KOs were then used to characterize individual gene functions and reconstruct KEGG pathways using the KEGG mapper tool (Kanehisa and Goto, 2000).

3.4 Results and discussion

3.4.1 General genome metrics

Members of the genus *Naganishia* have genomes ranging in size from 19.69 to 20.84 Mb and a G+C content from 52.81 to 54 %. The genome of *N. albida* NT2002 is larger than the other genomes studied (~20.84 Mb) with 7,379 proteins coding sequences. The genome of *N. vishniacii* ANT03-052 is the smallest in size (~19.69 Mb) and incorporates 6,602 protein-coding sequences (Table 3.1). The precise determinant of the difference in genome size between the species may not be evident, however, the low number of protein-coding sequences observed in the genome of *N. vishniacii* ANT03-052 could be an adaptive mechanism to thrive in unfavourable conditions characterised by limited resources (Schmidt *et al.,* 2017).

The genome sizes of the *Naganishia* strains are on the smaller range of Tremellomycetes genomes and display a similar G+C content higher than the average for the Basidiomycetes (Storck, 1966). Comparative genome analysis of the Tremellales revealed similar results, although in genera such as *Cryptococcus* and *Phaeotremella* the G+C content differed by more than 5 % (Aliyu *et al.,* 2020b).

On average, 162 tRNAs were observed in all studied strains. The largest number of tRNAs are found on the genome of *N. albida* NT2002 (184 tRNAs), while the smallest number was observed in *N. vishniacii* ANT03-052 (154 tRNAs). Typically psychrophilic microorganisms incorporate a larger number tRNAs on their genome as an adaptive mechanism to low temperature (Dutta and Chaudhuri, 2010). The lower number observed in *N. vishniacii* ANT03-052 may be a function of the quality of the genome assembly for this strain. The genome

completeness of the studied *Naganishia* genomes range from 89.2-90.4 %. Although some genomes have been assembled to less than 50 scaffolds the average completeness is ~88.96 %. This may suggest that *Naganishia* species have lost a number of core proteins since their divergence from other taxa within the Basidiomycota or may likewise be a function of the genome assemblies of these strains or the annotation, as AUGUSTUS, as part of the Funannotate pipeline, was trained with the genome of *Cryptococcus neoformans* JEC21. The latter organism is relatively distantly related to members of the genus *Naganishia* but is the closest relative for which an accurate gene model dataset is available.

3.4.2 Pan-genome analysis of four *Naganishia* strains

The term pan-genome refers to the entire set of genes for all strains within a clade. It includes a core genome (genes present in all the compared strains) (Tettelin *et al.,* 2008) and an accessory genome (dispensable genes found in a subset of the strains, and strain-specific genes) (Medini *et al.,* 2005). Analysing conserved genes is useful in determining the long-term evolutionary relatedness among the studied species (Medini *et al.,* 2005). Similarly, the identification of accessory genomic elements is an important tool that could provide insight into the organism's evolution and niche adaptation (Ozer *et al.*, 2014).

The pan-genome of the four *Naganishia* strains is comprised of 16,679 proteins. The core genome includes 3,772 proteins (~ 23 % of the total pan-genome) that are common to all four strains therefore, while the accessory genome comprises 12,907 proteins (Figure 3.1). The core genome comprises approximately 56% of the total proteins in *N. randhawae* eABCC1, 57 % in *N. vishniacii* ANT03052 and 51-52 % in both *N. albida* strains. Approximately 20 % of the proteins in the accessory genome are strain specific. Pan genome analysis reveals that *N. albida* strain JCM23234 and strain NT2002 share 57 % of their total protein number; implying that since their divergence, these strains have acquired a large proportion of unique proteins in their respective habitats.

The pan-genome of the four *Naganishia* shows evidence of gene family expansions and contractions associated with different lifestyles, providing a clue as to which proteins give a taxon a selective advantage in a particular ecological niche. Furthermore, large proportions of strain-specific proteins in the accessory genome is usually indicative of extensive acquisition of genes via horizontal gene transfer (HGT) (McCarthy and Fitzpatrick, 2019).

Table 3.1 Genome content and metadata of the strains analysed in this study.

	***Naganishia albida* NT2002**	***N. albida* JCM2334**	***N. randhawae* eABCC1**	***N. vishniacii* ANT03-052**	***Filobasidium wieringae* JCM11695**
Isolation source location	Dry dessert soil, China	Air, Japan	Avian guano, South Africa	McMurdo, Dry Valleys, Antarctica	Dew-retted flax, Netherlands
Assembly Size (Mb)	20.84	20.69	20.27	19.69	19.78
Largest Scaffold (bp)	1,150,063	1,576,301	793,44	1,839,583	1,972,145
Average Scaffold (bp)	86,845	279,694	52,517	410,165	549,481
Number of Scaffolds	240	74	386	48	36
Scaffold N50 (bp)	331,936	975,819	191,697	1,080,776	951,588
Percent GC (%)	54.00	53.82	51.71	52.81	49.87
Number of Genes	7,563	7,453	6,943	6,756	7,047
Number of Proteins	7,379	7,286	6,775	6,602	6,899
Number of tRNA	184	167	168	154	148
Unique Proteins	528	456	891	752	1,854
BUSCO Values (%)	89.8	90.4	86.8	89.2	88.6

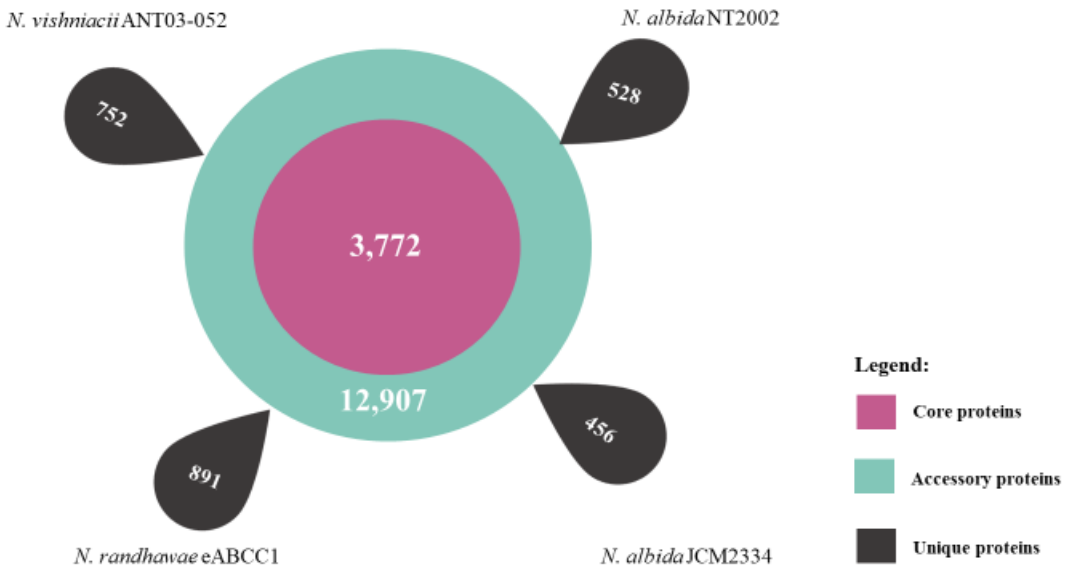


Figure 3.1 The core, accessory, and unique proteins of the compared *Naganishia* strains fractioned using Proteinortho V5.16 data output (Lechner *et al.* 2011). The inner circle (pink) represents the core genome consisting of 3,772 proteins. The outer circle (cyan) represents the accessory genome for all four species = (12,907 proteins), while the outer petals (black) represent the unique proteins associated with each species.

3.4.3 Clusters of Orthologous (COG) analysis of the core and pan-genome using eggNOG

To determine the functional roles of the proteins encoded within the different fractions of the pan-genome of the four *Naganishia* strains, the proteins were classified into their Clusters of Orthologous (COG) proteins using the eggNOG database (Huerta-Cepas *et al.*, 2019). Approximatively 64 % of the proteins could be assigned to COG functional categories. The remaining 35 % did not have orthologues in the eggNOG database. The majority of the open reading frames (ORFs) in a newly sequenced organism encode proteins from homologous families that are conserved across multiple species. However, a portion of the ORFs do not match known DNA coding sequences in other organisms. They usually account for 20-30 % of the genome and are referred to as ORFans (Siew and Fischer, 2003).

Analysis of the proteins COG data revealed that the highest proportion of proteins in the core genome belong to the COG functional categories posttranslational modification (O; 8 %; Figure 2). Post-translational modification is an effective mechanism developed by yeast cells to regulate the function of metabolic enzymes in response environmental disturbances (Tripodi

et al., 2015). This suggests that a high capacity for translation and posttranslational processing may be essential for growth in harsh environments.

The second largest proportion of COG functional categories in the core genome are Intracellular trafficking, secretion, and vesicular transport (U; 7 %) and Translation, ribosomal structure, and biogenesis (J; 7 %; Figure 3.2). These categories are involved in the delivery of newly synthesized proteins to their final destinations in the cell. As such they could play a pivotal roles in the rapid responses to environmental stimuli (Inada and Ueda, 2014).

Less than 35 % of the proteins unique to each strain were assigned to a characterised COG category. The two most abundant COG categories are posttranslational modification (O) and carbohydrate transport and metabolism (G) (Figure 3.3). Proteins involved in carbohydrate metabolism are important for energy production as they participate in glycolysis to produce adenosine triphosphate (ATP) (Flores *et al.,* 2000). Carbohydrate transport and metabolism differ among the *Naganishia* species to allow access to distinct nutrient datasets as they are present in different ecological niches.

3.4.4 Orthovenn2 analysis identifies a gene ontological cluster with UV-damage excision repair function in the genome of four *Naganishia* strains

Orthologous genes are clusters of genes in different species that originated by vertical descent from a single gene in the last common ancestor (Fitch, 1970). They usually retain their original function throughout evolution and help establish relationships between the gene function, proteome structure and taxonomic classification (Koonin, 2005). Comparative analysis of orthologous clusters makes it possible to identify overlapping clusters across multiple species.

Orthovenn2 analysis identified 842 clusters, comprising 3,435 proteins shared among all the *Naganishia* strains (Figure 3.4). Gene ontology annotation of the clusters that were shared revealed that the largest number of proteins under a cluster (17 proteins) was assigned to the function “UV-damage excision repair” (GO:0070914). Ultraviolet radiation is a form of non-ionizing radiation that is emitted by the sun and artificial sources and is classified into three primary types: ultraviolet A (UV-A), ultraviolet B (UV-B), and ultraviolet C (UV-C). These groups are based on the measure of their wavelength (nm) (Bjorn, 2002).

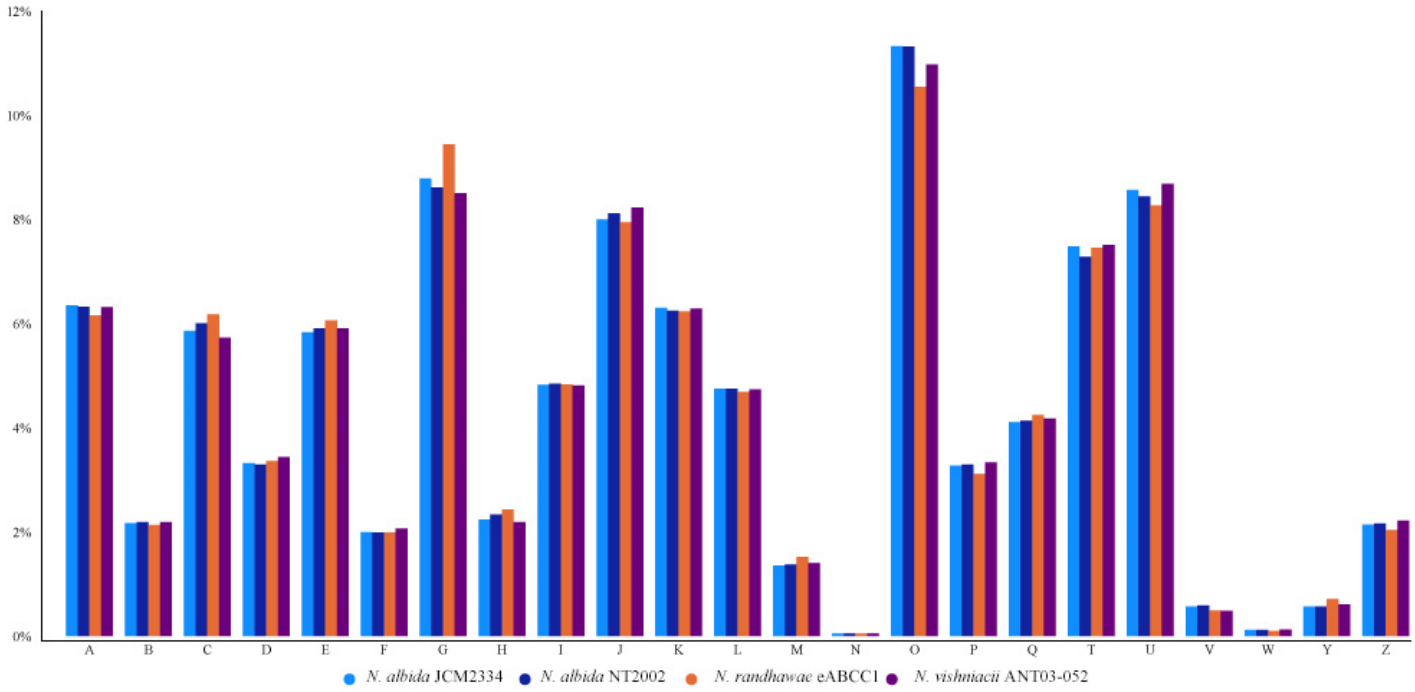


Figure 3.2 Proportion of proteins in the genome of four *Naganishia* genomes and their fractions assigned to Clusters of Orthologous Groups (COG) functional categories (Huerta-Cepas *et al.*, 2019). **A:** RNA processing and modification; **B:** Chromatin structure and dynamics; **C:** Energy production and conversion; **D:** Cell cycle control, cell division, chromosome partitioning; **E:** Amino acid transport and metabolism, **F:** Nucleotide transport and metabolism; **G:** Carbohydrate transport and metabolism; **H:** Coenzyme transport and metabolism; **I:** Lipid transport and metabolism; **J:** Translation, ribosomal structure and biogenesis; **K:** Transcription; **L:** Replication, recombination and repair; **M:** Cell wall/membrane/envelope biogenesis; **N:** Cell motility; **O:** Posttranslational modification, protein turnover, chaperones; **P:** Inorganic ion transport and metabolism; **Q:** Secondary metabolites biosynthesis, transport and catabolism; **T:** Signal transduction mechanisms; **U:** Intracellular trafficking, secretion, and vesicular transport; **V:** Defence mechanisms; **W:** Extracellular structures; **Y:** Nuclear structure; **Z:** Cytoskeleton.

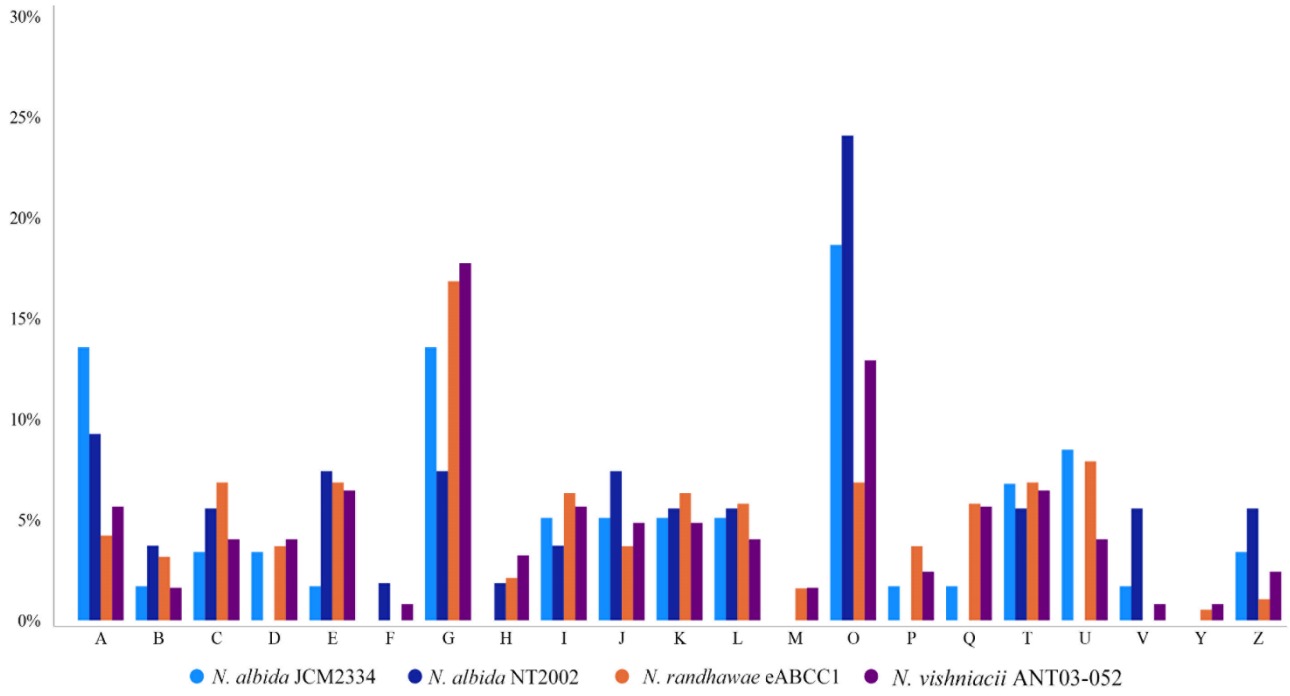


Figure 3.3 Proportion of proteins in the unique fraction of each *Naganishia* genomes and their Clusters of Orthologous Groups (COG) functional categories (Huerta-Cepas *et al.*, 2019). **A:** RNA processing and modification; **B:** Chromatin structure and dynamics; **C:** Energy production and conversion; **D:** Cell cycle control, cell division, chromosome partitioning; **E:** Amino acid transport and metabolism, **F:** Nucleotide transport and metabolism; **G:** Carbohydrate transport and metabolism; **H:** Coenzyme transport and metabolism; **I:** Lipid transport and metabolism; **J:** Translation, ribosomal structure and biogenesis; **K:** Transcription; **L:** Replication, recombination and repair; **M:** Cell wall/membrane/envelope biogenesis; **N:** Cell motility; **O:** Posttranslational modification, protein turnover, chaperones; **P:** Inorganic ion transport and metabolism; **Q:** Secondary metabolites biosynthesis, transport and catabolism; **T:** Signal transduction mechanisms; **U:** Intracellular trafficking, secretion, and vesicular transport; **V:** Defence mechanisms; **Y:** Nuclear structure; **Z:** Cytoskeleton.

When absorbed by living cells, UV radiation produces DNA photoproducts that can have a detrimental effect on yeast cell survival (Eckardt-Schupp *et al.*, 1991). Specifically, its absorption by nucleic acids results in many forms of DNA damage that interfere with DNA replication and transcription (Mojib *et al.*, 2013; Yasui and McCready, 1998). The most common type of UV induced damage are cyclobutane pyrimidine dimers (CPDs) and pyrimidine-pyrimidone 6–4 photoproducts (6–4PPs) (Yasui and McCready, 1998). DNA repair mechanisms are responsible for protecting genetic materials from the harmful effects of solar UV radiation by acting on both CPDs and 6-4PPs (Yasui and McCready, 1998).

Many of the habitats from which *Naganishia* species have been isolated are often marked by high levels of radiation (Mojib *et al.*, 2013). For example, *N. friedmannii* has been isolated from soil in the Atacama desert region of Chile and Argentina which is 6,000 meter above sea level (m.a.s.l) (Lynch *et al.*, 2012a; Schmidt *et al.*, 2017). It is estimated that the level of UV exposure in the Atacama desert is the highest level of any terrestrial environment on earth (Cabrera *et al.*, 1995; Lynch *et al.*, 2012). Indeed, radiation exposure at lower altitudes (1000 m.a.s.l.) in the same region has been documented to destroy *Bacillus* endospores and kill extremophiles such as *Chroococcidiopsis* sp. (Cockell *et al.*, 2008). Considering the high UV levels that *Naganishia* species are subjected to in this area, the predominance of a UV-repair mechanism cluster could confer additional protection against UV-related damage.

3.4.4.1 Survey of proteins involved in photoprotection in the genome of four *Naganishia* strains

Carotenoids and melanins are known as UV protective pigments capable of absorbing UV radiation at different wavelengths, including UV-C (Ellery and Wynn-Williams, 2003; Libkind *et al.*, 2009). In addition, they scavenge free radicals generated by UV radiation, protecting the cell against oxidative damage (Moore *et al.*, 1989; Schroeder and Johnson, 1993).

In chapter 2, we reported the presence of two putative laccase proteins involved in melanin production in *N. randhawae* eABCC1. In this analysis, putative laccases were identified as single copy genes in *N. albida* NT2002 and JCM2334 while the genome of *N. vishniacii* ANT03-052 encoded two predicted laccases (Table 3.2). The presence of melanin has been reported to enhance UV-C resistance in *C. neoformans* (Wang and Casadevall, 1994). UV-C radiation is highly damaging to biomolecules including DNA (Paulino-Lima *et al.*, 2016). Melanin production has only been observed in *N. albida* (strain CRA01 and CRA04) (Pedroso *et al.*, 2007) and *N. randhawae* eABCC1(this study).

A cluster of previously described proteins involved in the synthesis of mycosporine–glutaminol glucoside (MGG) was found in all four *Naganishia* species (Table 3.3). Mycosporine-glutaminol glucosides are water-soluble molecules capable of absorbing UV radiation with a peak at 310 nm (Bernillon *et al.*, 1984). In yeasts MGG acts as a UV-B photoprotective metabolite by protecting against direct damage of DNA and probably against indirect damage via the shikimate pathway (Herrmann and Weaver, 1999; Logemann *et al.*, 2000).

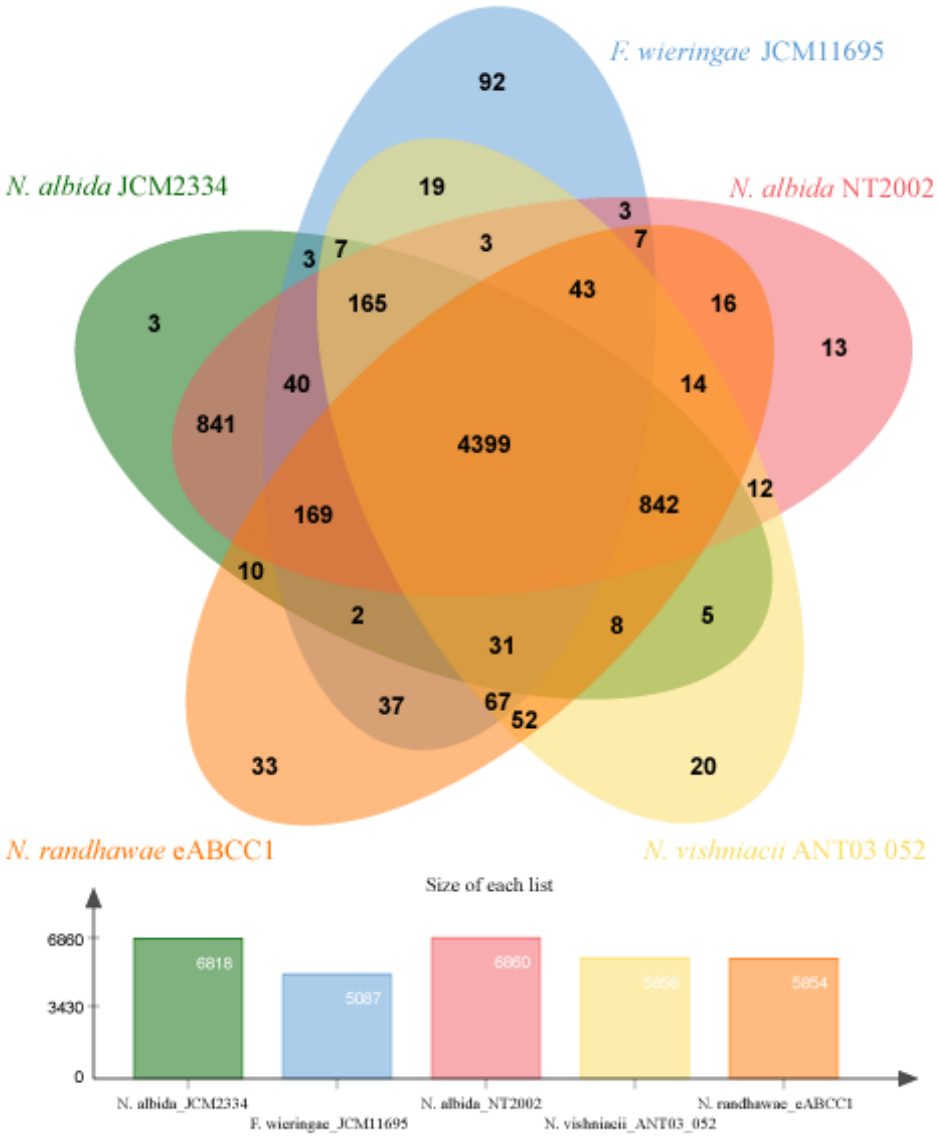


Figure 3.4 Venn diagram showing the distribution of shared and specific clusters of orthologous groups (COG) in the four *Naganishia* genomes and *Filobasidium wieringae* JCM11695. A total of 4,399 shared clusters of orthologous groups were identified using Orthovenn2 (Xu *et al.* 2019).

Comparative genomics of *Phaffia rhodozyma* strains revealed that MGG proteins were not identified in all the studied strains (Bellora *et al.*, 2016). Similarly, Nizovoy and colleagues (2020) reported that genes responsible for the synthesis of both carotenoid pigments and mycosporines were only identified in taxa belonging to the *Filobasidiales*. These findings support the hypothesis that MGG synthesis in yeasts is not essential, but rather an adaptation mechanism in response to high level of radiation.

Table 3.2 Putative laccases identified in the four *Naganishia* genomes identified using BLASTP (Boratyn *et al.,* 2013; Johnson *et al.,* 2008) (e-val < 1e-10). The first column describes the locus in the respective genomes. Column 2-5 BLASTP output is separated per protein name, BLASTP hit of each species, protein ID, amino acid identity and e-value.

Locus tag	**BLASTP hit**	**Subject id**	**% amino acid identity**	**e-value**
***N. albida* NT2002:** NT2002_007068	*Rhizoctonia solani* AG-1 IB	CCO34462.1	36.19	5E-123
***N. albida* JCM2334:** JCM2334_002085	*Rhizoctonia solani* AG-1 IB	CCO34462.1	36.19	2E-122
***N. randhawae* eABCC1:** EABCC1_001420 EABCC1_006080				
	Hysterangium stoloniferum HS.BST	KAF8523282.1	55.62	2E-108
	Rhizoctonia solani AG-1 IB	CEL60344.1	37.15	3E-129
***N. vishniacii* ANT03-052:** ANT03-052_002318 ANT03-052_002317				
	Colletotrichum higginsianum IMI 349063	XP_018159711.1	37.64	2E-28
	Auricularia auricula-judae	AHZ58329.1	36.72	1E-71

Table 3.3 Proteins involved in carotenoid and mycosporine synthesis (Bellora *et al.*, 2016; Nizovoy *et al.*, 2020). Proteins involved in the synthesis of carotenoid and mycosporine-glutaminol glucoside (MGG) were identified in the predicted proteins using BLASTP (e-val < 1e-10). The first column describes the locus in the respective genomes. Column 3-7 BLASTP (Boratyn *et al.*, 2013; Johnson *et al.*, 2008) output is separated per protein name, BLASTP hit of each species, protein ID, amino acid identity and e-value. The colour green represents proteins involved in MGG synthesis and yellow represents proteins involved in carotenoid synthesis.

Species	**Locus tag**	**Protein Name**	**BLASTP hit**	**Subject ID**	**% amino acid identity**	**e-value**
	ANT03-052_000061	ATP-grasp	*Planoprotostelium fungivorum Gg-2016a*	PRP76264.1	32.67%	9,00E-63
	ANT03-052_000062	2-epi-5-epi-valiolone synthase	*Heliocybe sulcata* OMC1185	TFK51961.1	57.96%	4,00E-176
	ANT03-052_000063	O-methyltransferase (O-MT)	*Sphaerulina musiva SO2202*	XP_016758125.1	39.74%	2,00E-54
***N. vishniacii* ANT03-052**	ANT03-052_000698	Farnesyl pyrophosphate synthetase	*Ustilago trichophora RK089*	SPO22926.1	73.45%	0.0
	ANT03-052_000743	Carotenoid-oxygenases	*Kockovaella imperatae CBS 7554*	XP_021873673.1	54.57%	0.0
	ANT03-052_001931	phytoene-beta carotene synthase	*Naematelia encephala CBS 8207*	ORY28280.1	41.48%	6,00E-45
	ANT03-052_002551	phytoene dehydrogenase	*Botrytis cinerea BcDW1*	EMR86132.1	43.25%	2,00E-139
	ANT03-052_003443	isopentenyl diphosphate isomerase	*Kwoniella heveanensis CBS 569*	OCF40338.1	69.41%	8,00E-119
	ANT03-052_005325	NADP-cytochrome P450 reductase	*Saitozyma sp. JCM 24511*	GFZ49232.1	67.75%	0.0
	ANT03-052_006327	farnesyltransferase	*Cryptococcus wingfieldii CBS 7118*	XP_019031513.1	54.61%	1,00E-86
	EABCC1_003600	ATP-grasp	*Planoprotostelium fungivorum Gg-2016a*	PRP76264.1	33.07%	5,00E-62

Species	Locus tag	Protein Name	BLASTP hit	Subject ID	% amino acid identity	e-value
	EABCC1_003599	2-epi-5-epi-valiolone synthase	*Heliocybe sulcata* OMC1185	TFK51961.1	57.48%	1,00E-178
	EABCC1_003598	O-methyltransferase (O-MT)	*Sphaerulina musiva SO2202*	XP_016758125.1	40.07%	1,00E-54
***N. randhawae* eABCC1**	EABCC1_000640	Carotenoid-oxygenases	Kockovaella imperatae CBS 7554	XP_021873673.1	51.75%	0.0
	EABCC1_001146	isopentenyl diphosphate isomerase	*Kwoniella heveanensis CBS 569*	OCF40338.1	68.42%	1,00E-124
	EABCC1_003919	phytoene-beta carotene synthase	*Kockovaella imperatae CBS 7554*	XP_021868037.1	39.26%	4,00E-109
	EABCC1_004836	phytoene dehydrogenase	*Lentithecium fluviatile CBS 122367*	KAF2679432.1	48.83%	2,00E-120
	EABCC1_005346	NADP-cytochrome P450 reductase	*Saitozyma sp. JCM 24511*	GFZ49232.1	65.57%	0.0
	EABCC1_006416	Farnesyl pyrophosphate synthetase	*Ustilago trichophora RK089*	SPO22926.1	73.73%	0.0
	EABCC1_006673	farnesyltransferase	*Kwoniella mangroviensis CBS 8886*	OCF72693.1	51.37%	1,00E-155
	JCM2334_000535	ATP-grasp	*Planoprotostelium fungivorum Gg-2016a*	PRP76264.1	33.79%	4,00E-69
	JCM2334_000533	O-methyltransferase (O-MT)	*Sphaerulina musiva SO2202*	XP_016758125.1	39.16%	2,00E-54
	JCM2334_000534	2-epi-5-epi-valiolone synthase	*Heliocybe sulcata* OMC1185	TFK51961.1	59.23%	6,00E-178
	JCM2334_000762	Carotenoid-oxygenases	*Kockovaella imperatae CBS 7554*	XP_021873673.1	54.48%	0.0
	JCM2334_000876	Farnesyl pyrophosphate synthetase	*Ustilago trichophora RK089*	SPO22926.1	70.74%	0.0
***N. albida* JCM2334**	JCM2334_003292	phytoene dehydrogenase	*Cladonia uncialis subsp. Uncialis 1753*	ANM86665.1	51.50%	0.0

Species	Locus tag	Protein Name	BLASTP hit	Subject ID	% amino acid identity	e-value
	JCM2334_003309	isopentenyl diphosphate isomerase	*Kwoniella heveanensis CBS 569*	OCF40338.1	72.05%	1,00E-123
	JCM2334_004826	phytoene-beta carotene synthase	*Kockovaella imperatae CBS 7554*	XP_021868037.1	42.58%	3,00E-112
	JCM2334_005450	NADP-cytochrome P450 reductase	*Saitozyma sp. JCM 24511*	GFZ49232.1	67.49%	0.0
	JCM2334_006971	farnesyltransferase	*Kwoniella mangroviensis CBS 8886*	OCF72693.1	52.66%	1,00E-164
	NT2002_000334	ATP-grasp	*Planoprotostelium fungivorum Gg-2016a*	PRP76264.1	33.79%	4,00E-69
	NT2002_000332	O-methyltransferase (O-MT)	*Sphaerulina musiva SO2202*	XP_016758125.1	39.16%	2,00E-54
	NT2002_000333	2-epi-5-epi-valiolone synthase	*Heliocybe sulcata* OMC1185	TFK51961.1	59.23%	6,00E-178
***N. albida* NT2002**	NT2002_003118	phytoene-beta carotene synthase	*Kockovaella imperatae CBS 7554*	XP_021868037.1	37.08%	3,00E-131
	NT2002_003679	phytoene dehydrogenase	*Cladonia uncialis subsp. Uncialis 1753*	ANM86665.1	50.17%	0.0
	NT2002_003697	isopentenyl diphosphate isomerase	*Kwoniella heveanensis CBS 569*	OCF40338.1	72.05%	1,00E-123
	NT2002_005357	farnesyltransferase	*Kwoniella mangroviensis CBS 8886*	OCF72693.1	52.66%	8,00E-165
	NT2002_005999	NADP-cytochrome P450 reductase	*Saitozyma sp. JCM 24511*	GFZ49232.1	67.49%	0.0
	NT2002_006501	Farnesyl pyrophosphate synthetase	*Ustilago trichophora RK089*	SPO22926.1	70.74%	0.0
	NT2002_006623	Carotenoid-oxygenases	*Kockovaella imperatae CBS 7554*	XP_021873673.1	56.60%	0.0

3.4.5 Survey of proteins involved in stress tolerance in the genome of four *Naganishia* strains

Yeast cells have evolved to withstand sudden and sometimes severe variations in their external environment (Gasch, 2003). In nature, they must contend with adverse conditions such as desiccation, high or low pH, nutrient scarcity, osmolarity and temperature changes (Gasch, 2003; Mager and Ferreira, 1993) that can disrupt internal cellular environments (Jamieson, 1998). As such yeast cells must rapidly adapt their internal environment to respond to sudden changes and preserve cellular integrity (Pérez and Cansado, 2010). Stress responses are of particular importance to members of the genus *Naganishia* whose environment is highly variable and factors such as temperature, osmolarity or nutrient availability are far from being constant (Gasch, 2003). The genomes of the four *Naganishia* strains employed in this study were evaluated for the presence of well-known molecular determinants of stress response.

3.4.5.1 Heat shock response

Heat-shock proteins are involved in several biological processes, such as transcription, translation and post-translational modification, protein folding, and protein aggregation and disaggregation (Feder and Hofmann, 1999; Parsell and Lindquist, 1993). There are two mechanisms by which heat shock response is induced in fungi, namely temperature stress and other stresses including pH, oxidative stress, osmotic stress and starvation (Tereshina, 2005).

Genome mining of the four *Naganishia* strains identified multiple copies of proteins involved in the heat shock response including DnaJ- protein (SCJ1), various heat shock proteins (HSPs), the stress-responsive transcription factor HSF1 and the type I transmembrane sorting receptor PEP1 (Table 3.4). The genome of *N. randhawae* eABCC1 encoded two additional copies of DnaJ (also known as heat shock protein 40 (HSP40; Table 3.4). Heat shock protein 40 acts as a co-chaperone by regulating adenosine triphosphate (ATP)-dependent polypeptide binding to HSP70 proteins (Langer *et al.,* 1992; Schröder *et al.,* 1993) and has been found to be essential for growth at temperatures above 42 °C in *Escherichia coli* (Schröder *et al.,* 1993). Interestingly, the genomes of psychrophilic bacterial species *Psychromonas ingrahamii* and *Alteromonas* sp. SN2 encode four copies of DnaJ (Math *et al.,* 2012; Riley *et al.,* 2008). Multiple copies of the DnaJ gene suggest that *N. randhawae* eABCC1 may be able to tolerate a range of temperatures.

Table 3.4 Proteins involved in heat shock response in four *Naganishia* strains.

Locus tag	Name	Product
***Naganishia albida* NT2002**		
NT2002_000528	HSF1	stress-responsive transcription factor hsf1
NT2002_003849	HSP10	10 kDa heat shock protein
NT2002_003850	HSP60	chaperonin
NT2002_000859	HSP70_1	Hsp70 chaperone
NT2002_003198	HSP70_2	Hsp70 chaperone
NT2002_006989	HSP70_3	Hsp70 chaperone
NT2002_001608	HSP78	chaperone ATPase hsp78
NT2002_002618	HSP90	Heat shock protein 90
NT2002_004727	SCJ1	DnaJ- protein scj1
NT2002_001154	PEP1_1	Type I transmembrane sorting receptor
NT2002_002125	PEP1_2	Type I transmembrane sorting receptor
***N. albida* JCM2334**		
JCM2334_004297	HSF1	stress-responsive transcription factor hsf1
JCM2334_006695	HSP10	10 kDa heat shock protein
JCM2334_006694	HSP60	chaperonin
JCM2334_000643	HSP70_1	Hsp70 chaperone
JCM2334_002007	HSP70_2	Hsp70 chaperone
JCM2334_002226	HSP70_3	Hsp70 chaperone
JCM2334_006226	SCJ1	DnaJ- protein scj1
JCM2334_002929	PEP1_1	Type I transmembrane sorting receptor
JCM2334_006431	PEP1_2	Type I transmembrane sorting receptor
***N. randhawae* eABCC1**		
EABCC1_000358	HSF1_1	stress-responsive transcription factor hsf1
EABCC1_003168	HSF1_2	stress-responsive transcription factor hsf1
EABCC1_006118	HSP10	10 kDa heat shock protein
EABCC1_006119	HSP60	chaperonin
EABCC1_001716	HSP70_1	Hsp70 chaperone
EABCC1_005133	HSP70_2	Hsp70 chaperone
EABCC1_006339	HSP70_3	Hsp70 chaperone
EABCC1_004240	HSP78	chaperone ATPase hsp78
EABCC1_003748	HSP90	Heat shock protein 90
EABCC1_000492	SCJ1_1	DnaJ- protein scj1
EABCC1_002040	SCJ1_2	DnaJ- protein scj1
EABCC1_002041	SCJ1_3	DnaJ- protein scj1
EABCC1_000192	PEP1_1	Type I transmembrane sorting receptor
EABCC1_000193	PEP1_2	Type I transmembrane sorting receptor

Locus tag	Name	Product
***N. vishniacii* ANT03-052**		
ANT03-052_004868	HSF1	stress-responsive transcription factor hsf1
ANT03-052_002693	HSP10	10 kDa heat shock protein
ANT03-052_002694	HSP60	chaperonin
ANT03-052_000924	HSP70_1	Hsp70 chaperone
ANT03-052_002340	HSP70_2	Hsp70 chaperone
ANT03-052_006186	HSP70_3	Hsp70 chaperone
ANT03-052_000566	HSP78	chaperone ATPase hsp78
ANT03-052_001172	HSP90	Heat shock protein 90
ANT03-052_005951	SCJ1	DnaJ- protein scj1
ANT03-052_003131	PEP1_1	Type I transmembrane sorting receptor
ANT03-052_004439	PEP1_2	Type I transmembrane sorting receptor

3.4.5.2 Osmotic stress response

The two aspects to consider when discussing the osmotic stress response in yeast are their ability to survive sudden changes in and withstand low water activity (Estruch, 2000). In yeasts, this response is achieved via the high osmolarity glycerol (HOG) pathway, a branch of the mitogen-activated protein kinase (MAPK) signal-transduction system. This pathway stimulates a transcriptional response that enables the yeast to adjust to the sudden changes (O'Rourke and Herskowitz, 2004). Using the KEGG mapper tool (Kanehisa and Goto, 2000), we were able to reconstruct the MAPK pathway and identify the relevant proteins in all the studied strains as expected.

Additionally, five putative copper amine oxidases were identified in the genome of *N. randhawae* eABCC1 (Table 3.5). Copper amine oxidases (CAOs) catalyze the oxidative deamination of polyamines, such as b-alanine, resulting in the formation of ammonia and hydrogen peroxide (Cai and Klinman, 1994). Although commonly found in filamentous fungi such as *Aspergillus niger* which hosts six copies (Frébort *et al.,* 1999), their specific function in fungi is not well established (Cona *et al.,* 2006). Researchers have suggested that CAOs may participate in the cellular regulation of H_2O_2 (Cona *et al.,* 2006).

Table 3.5 Putative copper amine oxidases found the genome of *Naganishia randhawae* eABCC1 using BLASTP (Boratyn *et al.,* 2013; Johnson *et al.,* 2008) (e-value < 1e-10). The first column describes the locus in the respective genomes. Column 2-4 BLASTP output is separated per protein name, BLASTP hit of each species, protein ID, amino acid identity and e-value.

Locus tag	**BLASTP hit**	**Subject id**	**% amino acid identity**	**e-value**
EABCC1_000003	*Trichophaea hybrida UTF0779*	KAF8544436.1	44.02	0.0
EABCC1_002072	*Rhizoctonia solani* 123E	KEP53626.1	51.31	0.0
EABCC1_002618	*Kockovaella imperatae CBS 7554*	XP_021872853.1	48.13	0.0
EABCC1_000735	*Saitozyma sp.* JCM 24511	GFZ48473.1	66.53	0.0
EABCC1_006557	*Saitozyma sp.* JCM 24511	GFZ48473.1	65.17	5E-28

3.4.6 Survey of polysaccharide capsule proteins in the genome of four *Naganishia* strains

The adaptation of microorganisms to their environment is often linked with the acquisition of unique attributes that help improve survival in specific ecologic niches. This includes morphological changes and the development of specialized structures such as capsules. In the environment, microbial capsules play an important role in providing resistance to stress factors such as dehydration (Aksenov *et al.,* 1973). The ability of members of the genus *Naganishia* to dominate xeric environments has been attributed to their polysaccharide capsule (Vishniac, 2006). In Chapter 2, proteins involved in capsule synthesis were identified using the PHI-base dataset. Similarly, genome mining of the other *Naganishia* genomes identified these proteins in all strains (Table 3.6).

3.4.7 Trehalose metabolism

Yeast cells accumulate trehalose in response to heat shock and to prevent dehydration when exposed to freezing temperatures (Diniz-Mendes *et al.,* 1999). Proteins involved in trehalose synthesis; trehalose-6-phosphate synthase (TPS1) and trehalose-6-phosphate phosphatase (TPS2) (De Virgilio *et al.,* 1993; Vuorio *et al.,* 1993); and the degradation of neutral trehalase (NTH1) (Estruch, 2000) were identified in all four *Naganishia* species. Two putative copies of TPS1 were found in all species except *N. randhawae* eABCC1, while two putative copies of TPS2 were identified only in *N. randhawae* eABCC1 (Table 3.7). Research has shown that cell trehalose levels correlate with cell survival under adverse conditions, such as during cold temperature, desiccation and starvation (Hottiger *et al.,* 1994).

Table 3.6 Proteins involved in capsule synthesis in the genome of four *Naganishia* strains.

Protein	**Locus tag**
Capsule associated protein (CAP1)	NT2002_007040, JCM2334_002059, EABCC1_004730, ANT03-052_002327
Capsular associated protein (CAP5)	NT2002_000688, JCM2334_001150, EABCC1_000889 ANT03-052_001370
Capsular associated protein (CAP60)	NT2002_006532, JCM2334_000844, EABCC1_005077 ANT03-052_000876
Capsular associated protein (CAP64)	NT2002_006084, JCM2334_001072, EABCC1_001634 ANT03-052_001470
UDP-glucuronic acid decarboxylase 1 (UXS1)	NT2002_005388, JCM2334_005536, EABCC1_000694 ANT03-052_005468
GDP-mannose transporter into the lumen of the Golgi (VRG4)	NT2002_004009, JCM2334_006941, EABCC1_006696 ANT03-052_006375
UDP-galactose transporter (HUT1)	NT2002_000578, JCM2334_002431, EABCC1_004114 ANT03-052_002596
Mannose-6-phosphate isomerase (PMI1)	NT2002_000578, JCM2334_002431, EABCC1_002577 ANT03-052_005229

Table 3.7 Proteins involved in trehalose metabolism in the genome of four *Naganishia* strains.

Protein	**Locus tag**
Trehalose-6-P synthase/phosphatase complex synthase subunit (TPS1)	NT2002_001035, NT2002_001036, JCM2334_006553 JCM2334_006554, EABCC1_003964, ANT03052_002897, ANT03-052_002898
Threalose-6-phosphate phosphatase (TPS2)	NT2002_003311, JCM2334_001920, EABCC1_003032 EABCC1_003033, ANT03-052_002456
Alpha, alpha-trehalase (NTH1)	NT2002_004213, JCM2334_006817, EABCC1_001781 EABCC1_006531, ANT03-052_002705

3.5 Conclusion

Comparative genomic analysis of four *Naganishia* strains was performed to study the adaptive strategies that might aid the survival of members of the genus in diverse ecosystems. The pan-genome analysis suggests that members of the genus *Naganishia* have evolved to proliferate in a multitude of ecological niches. This can be seen in the difference in number of proteins between the core and accessory genome. The broad range of photoprotective mechanisms identified in all four genomes provides possible answers to the molecular basis of its ecological success in some of the most extreme ecosystem on earth. Survey of proteins involved in general stress tolerance identified as expected multiple strategies that enables members of the genus *Naganishia* to cope with various environmental stressors. Findings from this study expand our knowledge of the evolution and adaptation of an important Tremellomycetes genus. While orthologues to a wide range of known molecular factors involved in stress tolerance were identified on the genomes of all four *Naganishia* taxa, further functional characterisation is required to determine their exact role in members of this genus. Furthermore, the genomes incorporate a number of gene models coding for proteins with no or hypothetical function prediction. Further experiments employing assays where the *Naganishia* taxa are exposed to single and multiple stressors, coupled with transcriptomics, will shed more light on the stress tolerance and adaptation pathways in this fascinating yeast genus.

3.6 References

Aksenov, S.I., Babyeva, I.P. and Golubev, V.I., 1973. On the mechanism of adaptation of micro-organisms to conditions of extreme low humidity. *Life sciences and space research*, *11*, pp.55-61.

Aliyu, H., Gorte, O., Zhou, X., Neumann, A. and Ochsenreither, K., 2020. In silico proteomic analysis provides insights into phylogenomics and plant biomass deconstruction potentials of the tremelalles. *Frontiers in bioengineering and biotechnology*, *8*, p.226.

Bellora, N., Moliné, M., David-Palma, M., Coelho, M.A., Hittinger, C.T., Sampaio, J.P., Gonçalves, P. and Libkind, D., 2016. Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: Phaffia rhodozyma. *BMC genomics*, *17*(1), pp.1-16.

Bernillon, J., Bouillant, M.L., Pittet, J.L., Favre-Bonvin, J. and Arpin, N., 1984. Mycosporine glutamine and related mycosporines in the fungus Pyronema omphalodes. *Phytochemistry*, *23*(5), pp.1083-1087.

Bhattacharya, S., Esquivel, B.D. and White, T.C., 2018. Overexpression or deletion of ergosterol biosynthesis genes alters doubling time, response to stress agents, and drug susceptibility in Saccharomyces cerevisiae. *MBio*, *9*(4).

Björn, L.O. ed., 2015. *Photobiology: The science of light and life*. Springer.

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y. and Raytselis, Y., 2013. BLAST: a more efficient report with usability improvements. Nucleic acids research, 41(W1), pp. W29-W33.

Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M.A., Werner-Washburne, M., Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L. and Krystofova, S., 2003. The genome sequence of the filamentous fungus Neurospora crassa.

Cai, D. and Klinman, J.P., 1994. Copper amine oxidase: heterologous expression, purification, and characterization of an active enzyme in Saccharomyces cerevisiae. *Biochemistry*, *33*(24), pp.7647-7653.

Cockell, C.S., McKay, C.P., Warren-Rhodes, K. and Horneck, G., 2008. Ultraviolet radiation-induced limitation to epilithic microbial growth in arid deserts–Dosimetric experiments in the

hyperarid core of the Atacama Desert. *Journal of Photochemistry and Photobiology B: Biology*, *90*(2), pp.79-87.

Cona, A., Rea, G., Angelini, R., Federico, R. and Tavladoraki, P., 2006. Reactive oxygen species production in the plant apoplastic environment. *Trends in Plant Science*, *2*(11), pp.80-88.

De Virgilio, C., Bürckert, N., Bell, W., Jenö, P., Boller, T. and Wiemken, A., 1993. Disruption of TPS2, the gene encoding the 100-kDa subunit of the trehalose-6-phosphate synthase/phosphatase complex in Saccharomyces cerevisiae, causes accumulation of trehalose-6-phosphate and loss of trehalose-6-phosphate phosphatase activity. *European Journal of Biochemistry*, *212*(2), pp.315-323.

Diniz-Mendes, L., Bernardes, E., De Araujo, P.S., Panek, A.D. and Paschoalin, V.M.F., 1999. Preservation of frozen yeast cells by trehalose. *Biotechnology and Bioengineering*, *65*(5), pp.572-578.

dos Santos Pedroso, R., da Costa, K.R.C., Ferreira, J.C. and Candido, R.C., 2007. Evaluation of melanin production by *Cryptococcus* species in four different culture media. *Revista da Sociedade Brasileira de Medicina Tropical*, *40*(5).

Dutta, A. and Chaudhuri, K., 2010. Analysis of tRNA composition and folding in psychrophilic, mesophilic and thermophilic genomes: indications for thermal adaptation. *FEMS Microbiology Letters*, *305*(2), pp.100-108.

Eckardt-Schupp, F., Ahne, A., Obermaier, S. and Wendel, S., 1991. UV-inducible repair in yeast. In *Photobiology* pp. 155-162.

Ellery, A. and Wynn-Williams, D., 2003. Why Raman spectroscopy on Mars? - a case of the right tool for the right job. *Astrobiology*, *3*(3), pp.565-579.

Estruch, F., 2000. Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiology Reviews*, *24*(4), pp.469-486.

Fica Cabrera, S.R., Bozzo, S. and Fuenzalida, H., 1995. Variations in UV radiation in Chile.

Fitch, W.M., 1970. Distinguishing homologous from analogous proteins. *Systematic Zoology*, *19*(2), pp.99-113.

Flores, C.L., Rodríguez, C., Petit, T. and Gancedo, C., 2000. Carbohydrate and energy-yielding metabolism in non-conventional yeasts. *FEMS Microbiology Reviews*, *24*(4), pp.507-529.

Fotedar, R., Kolecka, A., Boekhout, T., Fell, J.W., Anand, A., Al Malaki, A., Zeyara, A. and Al Marri, M., 2018. Naganishia qatarensis sp. nov., a novel basidiomycetous yeast species from a hypersaline marine environment in Qatar. *International Journal of Systematic and Evolutionary Microbiology*, *68*(9), pp.2924-2929.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M. and Louis, E.J., 1996. Life with 6000 genes. *Science*, *274*(5287), pp.546-567.

Goto, S. and Sugiyama, J., 1970. Studies on Himalayan yeasts and molds. IV. Several asporogenous yeasts including two new taxa of *Cryptococcus*. *Canadian Journal of Botany*, *48*(12), pp.2097-2101.

Hottiger, T., DE Virgilio, C., Hall, M.N., Boller, T. and Wiemken, A., 1994. The role of trehalose synthesis for the acquisition of thermotolerance in yeast: II. Physiological concentrations of trehalose increase the thermal stability of proteins in vitro. *European Journal of Biochemistry*, *219*(1-2), pp.187-193.

Hu, Z., He, B., Ma, L., Sun, Y., Niu, Y. and Zeng, B., 2017. Recent advances in ergosterol biosynthesis and regulation mechanisms in Saccharomyces cerevisiae. *Indian Journal of Microbiology*, *57*(3), pp.270-277.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. and von Mering, C., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), pp. 309-314.

Inada, N. and Ueda, T., 2014. Membrane trafficking pathways and their roles in plant–microbe interactions. *Plant and Cell Physiology*, *55*(4), pp.672-686.

Jamienson, D.J., 1998. Oxidative stress response in the yeast *Saccharomyces cerevisiae*. *Yeast*, *14*, pp.1511-1527.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic acids research, 36(suppl_2), pp. W5-W9.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), pp.27-30.

Kanehisa, M., 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Science*, *28*(11), pp.1947-1951.

Kanehisa, M., Sato, Y. and Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, *428*(4), pp.726-731.

Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet., 39, pp.309-338.

Langer, T. and Lu, C., 1992. Echols H*et al.*, Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, *356*(6371), pp.683-689.

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F. and Prohaska, S.J., 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics*, *12*(1), pp.1-9.

Lelandais, G. and Devaux, F., 2010. Comparative functional genomics of stress responses in yeasts. *Omics: a journal of integrative biology*, *14*(5), pp.501-515.

Li, L., Stoeckert, C.J. and Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), pp.2178-2189.

Libkind, D., Moliné, M., Sampaio, J.P. and Van Broock, M., 2009. Yeasts from high-altitude lakes: influence of UV radiation. *FEMS Microbiology Ecology*, *69*(3), pp.353-362.

Liti, G. and Louis, E.J., 2005. Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.*, *59*, pp.135-153.

Liu, X.Z. and Wang, Q.M., 2015. G öker. *M., Groenewald, M., Kachalkin, AV, Lumbsch, HT, Millanes, AM, Wedin, M., Yurkov, AM, Boekhout, T. & Bai, F.-Y*, pp.85-147.

Logemann, E., Tavernaro, A., Schulz, W., Somssich, I.E. and Hahlbrock, K., 2000. UV light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. *Proceedings of the National Academy of Sciences*, *97*(4), pp.1903-1907.

Luna-Tapia, A., Peters, B.M., Eberle, K.E., Kerns, M.E., Foster, T.P., Marrero, L., Noverr, M.C., Fidel, P.L. and Palmer, G.E., 2015. ERG2 and ERG24 are required for normal vacuolar physiology as well as Candida albicans pathogenicity in a murine model of disseminated but not vaginal candidiasis. *Eukaryotic cell*, *14*(10), pp.1006-1016.

Lynch, R.C., King, A.J., Farías, M.E., Sowell, P., Vitry, C. and Schmidt, S.K., 2012. The potential for microbial life in the highest-elevation (> 6000 masl) mineral soils of the Atacama region. *Journal of Geophysical Research: Biogeosciences*, 117.

Mager, W.H. and Ferreira, P.M., 1993. Stress response of yeast. *Biochemical Journal*, 290, p.1.

Math, R.K., Jin, H.M., Kim, J.M., Hahn, Y., Park, W., Madsen, E.L. and Jeon, C.O., 2012. Comparative genomics reveals adaptation by Alteromonas sp. SN2 to marine tidal-flat conditions: cold tolerance and aromatic hydrocarbon metabolism. *PLoS One*, *7*(4), p.e35784.

McCarthy, C.G. and Fitzpatrick, D.A., 2019. Pan-genome analyses of model fungal species. *Microbial Genomics*, *5*(2).

McColl, G., Hoffmann, A.A. and McKechnie, S.W., 1996. Response of two heat shock genes to selection for knockdown heat resistance in Drosophila melanogaster. *Genetics*, *143*(4), pp.1615-1627.

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R., 2005. The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), pp.589-594.

Mojib, N., Farhoomand, A., Andersen, D.T. and Bej, A.K., 2013. UV and cold tolerance of a pigment-producing Antarctic Janthinobacterium sp. Ant5-2. *Extremophiles*, *17*(3), pp.367-378.

Moore, M.M. and Breedveld, M.W., 1989. The role of carotenoids in preventing oxidative damage in the pigmented yeast, Rhodotorula mucilaginosa. *Archives of Biochemistry and Biophysics*, *270*(2), pp.419-431.

Nicholson, W.L., Schuerger, A.C. and Setlow, P., 2005. The solar UV environment and bacterial spore UV resistance: considerations for Earth-to-Mars transport by natural processes and human spaceflight. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *571*(1-2), pp.249-264.

Niinemets, Ü. and Sack, L., 2006. Structural determinants of leaf light-harvesting capacity and photosynthetic potentials. In *Progress in Botany* (pp. 385-419). Springer, Berlin, Heidelberg.

Nizovoy, P., Bellora, N., Haridas, S., Sun, H., Daum, C., Barry, K., Grigoriev, I.V., Libkind, D., Connell, L.B. and Moliné, M., 2021. Unique genomic traits for cold adaptation in Naganishia vishniacii, a polyextremophile yeast isolated from Antarctica. *FEMS Yeast Research*, *21*(1), p. foaa056.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M., 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *27*(1), pp.29-34.

O'Rourke, S.M. and Herskowitz, I., 2004. Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. *Molecular Biology of the Cell*, *15*(2), pp.532-542.

Östlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D.N., Roopra, S., Frings, O. and Sonnhammer, E.L., 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic acids research*, *38*(suppl_1), pp.196-203.

Ozer, E.A., Allen, J.P. and Hauser, A.R., 2014. Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics*, *15*(1), pp.1-17.

Palmer JM., 2016. Funannotate: a fungal genome annotation and comparative genomics pipeline. Available at: https://github.com/nextgenusfs/funannotate.

Parks, L.W. and Casey, W.M., 1995. Physiological implications of sterol biosynthesis in yeast. *Annual Review of Microbiology*, *49*(1), pp.95-116.

Parsell, D.A. and Lindquist, S., 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual Review of Genetics*, *27*, pp.437-49

Patterson, C., 1988. Homology in classical and molecular biology. *Molecular Biology and Evolution*, *5*(6), pp.603-625.

Paulino-Lima, I.G., Fujishima, K., Navarrete, J.U., Galante, D., Rodrigues, F., Azua-Bustos, A. and Rothschild, L.J., 2016. Extremely high UV-C radiation resistant microorganisms from desert environments with different manganese concentrations. *Journal of Photochemistry and Photobiology B: Biology*, *163*, pp.327-336.

Perez, P. and Cansado, J., 2010. Cell integrity signaling and response to stress in fission yeast. *Current Protein and Peptide Science*, *11*(8), pp.680-692.

Piškur, J. and Langkjaer, R.B., 2004. Yeast genome sequencing: the power of comparative genomics. *Molecular Microbiology*, *53*(2), pp.381-389.

Prasad, R., Shah, A.H. and Rawal, M.K., 2016. Antifungals: mechanism of action and drug resistance. *Yeast Membrane Transport*, pp.327-349.

Riley, M., Staley, J.T., Danchin, A., Wang, T.Z., Brettin, T.S., Hauser, L.J., Land, M.L. and Thompson, L.S., 2008. Genomics of an extreme psychrophile, Psychromonas ingrahamii. *BMC Genomics*, *9*(1), pp.1-19.

Schmidt, S.K., Vimercati, L., Darcy, J.L., Arán, P., Gendron, E.M., Solon, A.J., Porazinska, D. and Dorador, C., 2017. A Naganishia in high places: functioning populations or dormant cells from the atmosphere? *Mycology*, *8*(3), pp.153-163.

Schoch, C.L., Robbertse, B., Robert, V., Vu, D., Cardinali, G., Irinyi, L., Meyer, W., Nilsson, R.H., Hughes, K., Miller, A.N. and Kirk, P.M., 2014. Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. *Database*, *2014*.

Schröder, H., Langer, T., Hartl, F.U. and Bukau, B., 1993. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *The EMBO Journal*, *12*(11), pp.4137-4144.

Schroeder, W.A. and Johnson, E.A., 1993. Antioxidant role of carotenoids in Phaffia rhodozyma. *Microbiology*, *139*(5), pp.907-912.

Shah Alam Bhuiyan, M., Eckstein, J., Barbuch, R. and Bard, M., 2007. Synthetically Lethal Interactions Involving Loss of the Yeast ERG24: The Sterol C-14 Reductase Gene. *Lipids*, *42*(1), pp.69-76.

Siew, N. and Fischer, D., 2003. Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins: Structure, Function, and Bioinformatics*, *53*(2), pp.241-251.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single copy orthologs. *Bioinformatics*, *31*(19), pp.3210-3212.

Slater, G.S.C. and Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), pp.1-11.

Stalker, D.M., Hiatt, W.R. and Comai, L., 1985. A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosate. *Journal of Biological Chemistry*, *260*(8), pp.4724-4728.

Storck, R., 1966. Nucleotide Composition of Nucleic acids of fungi II. Deoxyribonucleic acids. *Journal of Bacteriology*, *91*(1), pp.227-230.

Tereshina, V.M., 2005. Thermal resistance in fungi: the role of heat shock proteins and trehalose. *Mikrobiologiia*, *74*(3), pp.293-304.

Tettelin, H., Riley, D., Cattuto, C. and Medini, D., 2008. Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), pp.472-477.

Tripodi, F., Nicastro, R., Reghellin, V. and Coccetti, P., 2015. Post-translational modifications on yeast carbon metabolism: regulatory mechanisms beyond transcriptional control. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1850*(4), pp.620-627.

UniProt Consortium, 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *46*(5), p.2699.

Van Dongen, S.M., 2000. *Graph Clustering by Flow Simulation* (Doctoral dissertation).

Vishniac, H.S., 2006. A multivariate analysis of soil yeasts isolated from a latitudinal gradient. *Microbial Ecology*, *52*(1), pp.90-103.

Vuorio, O.E., Kalkkinen, N. and Londesborough, J., 1993. Cloning of two related genes encoding the 56-kDa and 123-kDa subunits of trehalose synthase from the yeast Saccharomyces cerevisiae. *European Journal of Biochemistry*, *216*(3), pp.849-861.

Wang, Y. and Casadevall, A., 1994. Decreased susceptibility of melanized *Cryptococcus* neoformans to UV light. *Applied and Environmental Microbiology*, *60*(10), pp.3864-3866.

Wu, Y., Du, P.C., Li, W.G. and Lu, J.X., 2012. Identification and molecular analysis of pathogenic yeasts in droppings of domestic pigeons in Beijing, China. *Mycopathologia*, *174*(3), pp.203-214.

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q., Coleman-Derr, D., Xia, Q. and Wang, Y., 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(1), pp.52-58.

Yasui, A. and McCready, S.J., 1998. Alternative repair pathways for UV-induced DNA damage. *Bioessays*, *20*(4), pp.291-297.

Yong, X., Yan, Z., Xu, L., Zhou, J., Wu, X., Wu, Y., Li, Y., Chen, Z., Zhou, H., Wei, P. and Jia, H., 2016. Genome sequence of a microbial lipid producing fungus *Cryptococcus albidus* NT2002. *Journal of Biotechnology*, *223*, pp.6-7.

Chapter 4 Re-evaluating the taxonomic status of the fungal class Tremellomycetes: A phylogenomic perspective

4.1 Abstract

The current taxonomic classification of fungi is based on the combination of morphological, ecological and molecular DNA data. However, these may not reflect the true genotypic and phenotypic delineations of taxa as they often lack the power to resolve the lower taxonomic rank. To resolve this issue, phylogenomics proposes the analysis and comparison of whole genome data. The phylogenomic analysis of 161 taxa within the class Tremellomycetes was performed in this study. A total of 452 single copy orthologues (SCOs) are conserved among the studied taxa and these were used to generate a maximum likelihood (ML) tree. The resultant core genome phylogeny offers clear resolution at both the higher and lower taxonomic level for most taxa, with the exception of members of the genus *Cryptococcus*, where more genus-specific markers may be required for adequate delineation. Taxonomic revision presented in this study includes the proposal of a new order Phaeotremellaceales **ord. nov.** and two new species combinations *Apiotrichum oleaginum* **comb. nov.** and *Naganishia frigoris* **comb. nov.** The generated phylogenomic tree paves the way for a more robust classification of Tremellomycetes at the genome level.

4.2 Introduction

The kingdom Fungi is comprised of an extraordinary number of eukaryotes found in diverse ecosystems (Stajich, 2017). The number of fungal taxa is estimated to be between 1.5 and 7.1 million species (Hibbett, 2006). This rich diversity has posed an obstacle to fungal systematics and classification (Hibbett and Taylor, 2013; Stajich *et al.,* 2009). Traditional classification of fungi was based on the observation of morphological and ecological traits. This method resulted in problematic taxonomy as these diverse fungal groups often display plesiomorphic or convergent morphologies (Wang *et al.,* 2016). Additionally, the lack of fossil records and paleontological information make it very difficult to rely solely on morphology as the method of classification (Mi *et al.,* 2011).

The advent of molecular phylogenetics can be traced back to the early 1990s with the use of markers such as the large subunit (LSU) and small subunit (SSU) of rDNA and the ITS region (Saenz *et al.,* 1994; Swann and Taylor, 1993). The application of DNA sequence data to infer phylogenetic lineages improved taxonomic resolution and scientific reproducibility (Saenz *et al.,* 1994; Swann and Taylor, 1993). However, phylogenetic analysis using single molecular markers revealed that this approach is insufficient for resolving all existing taxonomic

problems in the kingdom Fungi (Bridge *et al.,* 2005). The ITS region, for example, exhibits high sequence variability among closely related organisms (Bridge *et al.,* 2005; Desdevises *et al.,* 2002).

To answer some of the most pertinent questions in fungal evolution, mycologists have resorted to the use of multilocus analysis (MLSA). This technique characterises species using the combined DNA data of multiple housekeeping genes (Gabaldón and Marcet-Houben, 2014; Ren *et al.,* 2016). Commonly employed MLSA markers for fungal classification include rDNA sequences and the sequences of select house-keeping genes such as ACT, rRNA, RPB1, RPB2 and TEF1 (Guarro *et al.* 1999; Liu *et al.* 2015). The MLSA approach facilitated several major taxonomic revisions, this includes new lineages especially at higher taxonomic level (Hibbett *et al.,* 2007; James *et al.,* 2006). For example, the Assembling the Fungal Tree of Life (AFTOL) project generated several phylogenies incorporating comprehensive taxon sampling using the MLSA approach (Lutzoni *et al.,* 2004). One of the major achievements of the AFTOL project was the completion of a six-gene phylogeny of 52 sequenced fungal genomes. Based on this study, Lutzoni and colleagues (2004) identified two major challenges in fungal systematics in the molecular era. The first challenge is to obtain a balanced sample of taxa and genetic markers. The second is to identify and interpret differences in morphological evolution and molecular phylogeny. While transitioning from the use of single genes to the MLSA approach resulted in a substantial increase in phylogenetic information, it became clear that there was still a lack of power to resolve higher level-relationships (Schoch *et al.,* 2014).

The fungal class Tremellomycetes comprises a large number of unicellular and dimorphic taxa forming hyphae and fruiting bodies (Money, 2016). Its current taxonomy consists of five orders, 17 families and 56 genera (Liu *et al.,* 2015a; Takashima *et al.,* 2019). Furthermore, it encompasses more than 370 species, and the list continues to expand as new species are identified around the globe (Weiss *et al.,* 2014). The current taxonomic framework of the Tremellomycetes was established from the work of Liu and colleagues (2015a, 2015b). They used a combination of multiple statistical approaches such as Bayesian Statistics, Maximum Likelihood (ML) and Neighbour Joining (NJ) trees to construct a phylogeny based on seven genes. This included the small rDNA subunit, the D1/D2 domains of the large rDNA subunit, the ITS 1 and 2 of rDNA including 5.8S rDNA, two RPB1 and RPB2, the TEF1 and the CYTB (Liu *et al.* 2015a). This phylogeny was able to lay down a blueprint for a more coherent taxonomic classification and resolved several polyphyletic genera within the Tremellomycetes. The authors also acknowledged that a large number of strains needed to be included in the

study and cautioned taxonomists against including strains solely on the basis of the LSU sequences results (Liu *et al.* 2015b).

With the increasing availability of fungal genomes, recent works have harnessed the information contained in the genomes to develop more robust taxonomic frameworks for several Tremellomycetes taxa. For instance, a series of studies of the order Trichosporonales recently reported misclassifications and proposed two new genera, *Pascua* and *Prillingera* (Aliyu *et al.,* 2020a; Takashima *et al.,* 2019, 2018). Similarly, comparative analysis of the order Tremellales reported that the family *Phaeotremellaceae* delineates in a distant position from other members of the Tremellales (Aliyu *et al.,* 2020a). The taxonomic position of this family was also resolved with caution in the seven gene phylogeny primarily due to the poor results obtained from the Bayesian statistics (Liu *et al.,* 2015a). These phylogenetic incongruities indicate that while the work of Liu and colleagues (2015a) paved the way for a more adequate phylogeny, more work needs to be done to address remaining issues in fungal taxonomy.

In this study, the available genome datasets of fungi in the class Tremellomycetes were examined to establish a phylogeny that will evaluate their current taxonomic position at a genomic level.

4.3 Materials and methods

4.3.1 Genome selection and Structural annotation

The genomes of 219 members of the class Tremellomycetes were retrieved from the NCBI GenBank assembly database (https://www.ncbi.nlm.nih.gov/genbank/). The completeness of the genomes was determined using the Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.3 (Simão *et al.,* 2015). The Basidiomycota odb9 dataset, which comprises of a total of 1,335 single-copy orthologues conserved among the Basidiomycota, was used as reference. Poor quality genomes, incorporating < 70 % of the BUSCOs were discarded based on the BUSCO output. Finally, the genomes of 161 Tremellomycetes were considered for downstream analysis (Table 4.1).

The BUSCO duplicate values of 157 taxa were < 5 %, these taxa were considered to have haploid structures. The genome of four taxa, namely *Cutaneotrichosporon cutaneum* B3, *C. mucoides* JCM 9939^{T}, *Trichosporon ovoides* JCM 9940 and *T. coremiiforme* JCM 2938^{T} revealed duplicates values > 60 % and have been shown to comprise hybrid genomes (Aliyu

et al., 2020a; Takashima *et al.,* 2018). Four taxa, two Agarycomycetes (*Piloderma croceum* F1598 and *Sistotremastrum niveocremeum* HHB9708) and two Dacarymycetes (*Calocera cornea* HHB12733 and *Dacryopinax primogenitus* DJM-731 SS1^{T}) were included as outgroups.

Metadata relating to the assembly metrics and sources of isolation were obtained from the BioProject and BioSample databases from the NCBI (Barrett *et al.,* 2012). All genomes were structurally annotated using the Funannotate pipeline v1.5.1-93c317b (Palmer, 2016) as described in Chapter 3.

4.3.2 Single Copy Orthologues selection

Orthologues were identified among the protein datasets derived from the structural annotation using Orthofinder v2.3.3 (Emms and Kelly, 2015). From the Orthofinder output, SCOs conserved among all the sampled taxa were selected, while those occurring in two copies were retained from the hybrid genomes. Duplicate SCOs identified in the hybrid genomes were separated by comparing each hybrid with its closest relatives using BLASTP and BioEdit v7.0.5.3 (Hall, 1999). Subsequently, the duplicate SCOs from the hybrid genomes were split according to their bit score where the SCOs with the highest bit score was grouped as "Hybrid strain number _1" and the second SCOs grouped as "Hybrid strain number _2".

4.3.3 Single Copy Orthologues extraction, quality check and phylogenomic tree construction

Prior to alignment, the extracted haploid and hybrid SCOs were subjected to a quality control step to repair proteins with incomplete amino acid sequences. First, a BLASTP of the truncated SCOs against the NCBI database (Jenuth, 2000) was completed. Subsequently, the hit result with the highest bit score and lowest e-value were used to conduct a tBlastN and BLASTP analysis using BioEdit (Hall, 1999; Jenuth, 2000) against the genome and proteome of the organisms with truncated SCOs.

The individual SCO datasets were aligned using the M-Coffee implementation of T-coffee v11.00.8cbe486 (Magis *et al.,* 2014; Notredame *et al.,* 2000). The resultant alignments were then concatenated and poorly aligned regions were removed using Gblocks v0.9b (Castresana, 2000; Talavera and Castresana, 2007). Gblocks is a computer programme that identifies and eliminates misaligned positions and divergent regions in a DNA or protein sequence alignment (Talavera and Castresana, 2007). This includes positions that are not homologous or have been

saturated by multiple substitutions; it is therefore convenient to eliminate them prior to performing phylogenetic analysis. Subsequently a ML phylogeny (Guindon *et al.*, 2010) was constructed using IQ-Tree v1.6.7 (Nguyen *et al.*, 2015) with bootstrap support (n = 1,000 replicates) using Fast bootstrap (Kalyaanamoorthy *et al.*, 2017).

The ModelFinder with IQ-Tree v1.6.7 was used to predict the protein evolutionary model. This approach incorporates substitution models from other common model-selection methods to evaluate the best-fit model for the presented data. ModelFinder is faster and more versatile than other model-selection approaches, and it can detect sequence evolution models that other methods cannot. For instance, multimodal distributions of site-to-site rate heterogeneity. In terms of the fit between tree, model, and data, ModelFinder proved to be accurate and frequently outperformed other model-selection methods (Kalyaanamoorthy *et al.*, 2017).

4.4 Results and discussion

4.4.1 Genomic characteristics of the Tremellomycetes

The genomes of 161 taxa belonging to the Cystofilobasidiales (seven strains), Filobasidiales (nine strains), Holtermanniales (two strains), Tremellales (105 strains) and Trichosporonales (38 strains) were incorporated in the analysis. Metadata on the origin of the included Tremellomycetes taxa indicates a broad geographical distribution of species with isolates obtained from decaying wood, humans, soil samples and water bodies (Figure 4.1, Table 4.1). Approximately 37 % of strains in this study are known pathogens or were isolated from clinical settings (Figure 4.1). Pathogenic strains are generally given preference when it comes to microbial genome sequencing due to their medical relevance (Relman, 2011).

The estimated genome size of the Tremellomycetes strains ranged between 15.66Mb (*Cryptococcus depauperatus* CBS 7855) and 42.35 Mb (*T. coremiiforme* JCM 2938, a hybrid genome) with an average genome size of 21.63 Mb. These sizes are in the lower range for reported fungal genomes (8.97-177.57 Mb) and below the 46.48 Mb Basidiomycota average (Mohanta and Bae, 2015). The genome size of *T. coremiiforme* JCM2938 can be linked to the fact it is a hybrid genome (Table 4.1). The G+C content varied from 40.2 to 62.8 % (Table 4.1) which is similar to the reported range for basidiomycetes (44 to 63 %) (Storck, 1966).

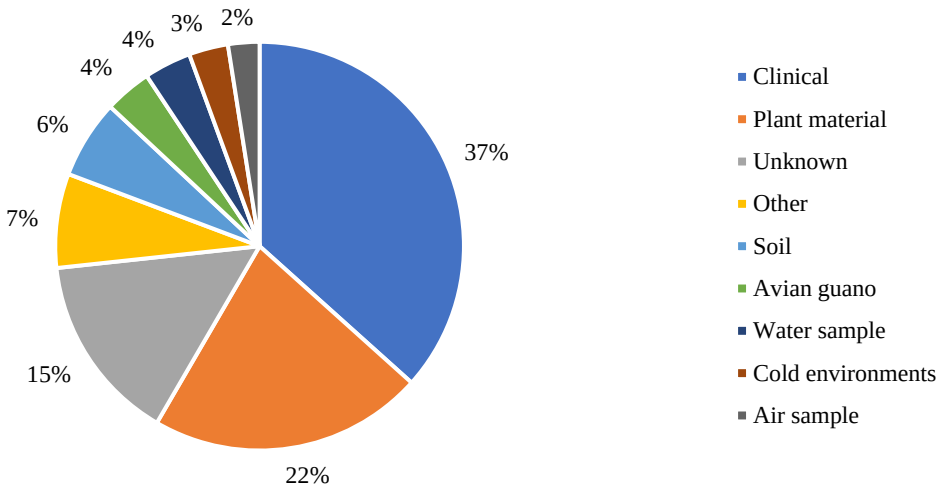


Figure 4.1 Pie chart showing the environmental sources of 161 Tremellomycetes taxa analyzed in this study.

4.4.2 General tree topology of the Tremellomycetes

A total of 452 SCOs were conserved among all the studied strains, including the outgroup taxa. After alignment of the individual SCOs, concatenation and removal of poorly aligned regions, the final alignment comprised of 96,605 amino acid positions. The resultant ML phylogeny showed the clustering of Tremellomycetes into eight distinct clades (Figure 4.2). The largest number of taxa fell within a clade that corresponds to the order Tremellales (~ 65 % of the studied taxa). The first cluster among the Tremellales consisted of members of the *Cryptococcaceae* including the genera *Cryptococcus* and *Kwoniella*. The second cluster within the Tremellales comprised of twelve taxa belonging to the families *Bulleraceae*, *Bulleribasidiaceae*, *Cuniculitremaceae*, *Naemateliaceae*, *Rhynchogastremaceae*, *Tremellaceae*, and *Trimorphomycetaceae*. A single taxon which has been ascribed to the genus *Cryptococcus*, namely *Cryptococcus* Mo.29 (GCA_900079645.1), did not cluster with any of the other *Cryptococcus* spp., nor even in the Tremellales clade, and instead clustered with the genus *Naganishia*. The Trichosporonales was the second largest order in the study (~24 %) forming six main clusters, which included members of the families *Tetragoniomycetaceae* and *Trichosporonaceae*.

Table 4.1 Genomic features of 161 Tremellomycetes and outgroup strains.

Strains	**Isolation source/ locality**	**Assembly Size (bp)**	**GC %**	**# Scaffolds**	**Scaffold N50**	**Comp[¥]**	**Dup[£]**
Apiotrichum akiyoshidainum HP2023	Rhizosphere, Argentina	30,042,643	60.6	1,067	53,521	90.4	4.6
A. brassicae JCM 1599[T]	Cabbage purchased in a market, Japan	23,647,732	56.5	16	2,410,059	95.0	0.2
A. domesticum JCM 9580[T]	Damp and rotten wooden sideboard in a kitchen of the house of a summer-type hypersensitivity pneumonitis patient	24,510,922	58.5	28	3,306,260	91.7	0.4
A. gamsii JCM 9941[T]	Moist humus around roots, Colombia	24,609,388	61.1	29	3,284,570	95.6	1.4
A. gracile JCM 10018	Sour milk, Germany	24,114,851	59.2	17	4,554,691	94.8	0.3
A. laibachii JCM 2947[T]	Soil	30,616,633	59.6	26	2,819,049	95.3	0.4
A. montevideense JCM 9937	Water purification tank, Uruguay	24,872,216	58.2	61	1,972,470	92.0	0.2
A. porosum DSM 27194	Grassland, Germany	25,479,456	59.20	32	1,376,709	95.8	0.6
A. porosum JCM 1458[T]	Exudate of *Taxus baccata*, Germany	25,989,348	59.00	37	3,276,928	96.1	0.7
A. veenhuisii JCM 10691[T]	Buffalo dung, Italy	31,617,680	59.6	35	2,551,142	94.5	0.4
Bullera alba JCM 2954[T]	Air in a dairy, United States	19,417,308	54.3	8	2,864,929	92.6	0.2
Calocera cornea HHB12733	Unknown	33,244,933	56.9	545	169,300		
Cryptococcus amylolentus CBS 6039[T]	Larval frass of buprestid beetle *Enneadesmus forficulus* (Coleoptera, Buprestidae) in *Dombeya rotundifolia* (Malvales, Malvaceae), South Africa	20,254,996	53.40	15	1,545,655	91.8	0.6
C. amylolentus CBS 6273	Unknown	20,280,838	53.20	18	1,529,651	91.4	0.7
C. bacillisporus CA1280	Clinical, USA	17,588,017	48.00	89	737,015	91.9	0.4
C. deneoformans JEC21	Generated from NIH B-3501 and NIH B-3502a-types (progenies of ATCC 28957 ATCC 28958)	19,051,922	48.54	14	1,438,950	91.7	0.6
C. deneoformans XL280	Unknown	19,049,915	48.53	37	1,076,535	92.1	0.5

(*Continued*)

Strains	Isolation source/ locality	Assembly Size (bp)	GC %	# Scaffolds	Scaffold N50	Comp[¥]	Dup[£]
C. depauperatus CBS 7855	Missing, Czech Republic	15,656,276	44.70	41	882,956	82.6	0.3
C. depauperatus CBS 7841^{T}	Missing, Canada	15,797,590	44.40	68	501,291	84.9	0.3
C. deuterogattii 2001/935-1	Unknown	17,566,680	47.90	24	1,248,189	92.5	0.4
C. deuterogattii 99/473	Clinical, Caribbean	17,403,577	47.80	105	520,419	91.5	0.4
C. deuterogattii CA1014	Clinical, USA	17,527,606	47.90	21	1,238,107	92.2	0.4
C. deuterogattii CBS 10090	Clinical, Greece	17,478,748	47.80	24	1,122,164	91.5	0.4
C. deuterogattii CBS 7750	Native rainforest of Northern Brazil	17,479,311	47.70	26	1,125,130	77.2	0.4
C. deuterogattii LA55	Clinical, Brazil	17,360,036	47.80	134	394,791	91.9	0.5
C. deuterogattii MMRL2647	Clinical, Australia	17,485,654	47.90	25	1,214,750	91.9	0.4
C. deuterogattii R265	Clinical, Canada	17,581,311	47.79	15	1,316,417	91.2	0.4
C. deuterogattii R265	Clinical, Canada	17,681,865	47.80	27	1,243,494	86.9	0.7
C. deuterogattii R265	Outbreak strain	17,475,811	47.80	28	1,124,359	91.0	0.6
C. deuterogattii Ram5	Clinical, Australia	17,453,840	47.90	24	915,286	91.8	0.5
C. floricola DSM 27421^{T}	*Echium leucophaeum* Nectar, Spain	20,137,742	53.50	573	119,269	92.3	0.7
C. gattii CA1873	Clinical, USA	17,438,253	48.00	33	761,613	92.6	0.5
C. gattii E566	Eucalyptus tree, Australia	17,996,826	47.90	52	717,019	91.4	0.4
C. gattii EJB2	Clinical, USA	17,671,044	47.90	282	225,095	91.2	0.5
C. gattii NT-10	Clinical, Australia	18,032,565	47.90	226	283,233	91.0	0.4
C. gattii R265	Unknown	17,256,624	47.80	779	71,225	88.9	0.4
C. gattii Ru294	Unknown tree, South Africa	17,860,310	47.90	53	699,094	91.8	0.4
C. gattii VGV MF34	Tree bark, Zambia	17,886,378	47.8	15	1,323,151	92.0	0.4
C. gattii WM276	*Eucalyptus tereticornis* debris, Australia	18,374,760	47.88	14	1,333,124	91.4	0.4
C. neoformans 125.91	Clinical, Tanzania	18,910,907	47.10	37	836,747	90.6	0.3

(*Continued*)

Strains	Isolation source/ locality	Assembly Size (bp)	GC %	# Scaffolds	Scaffold N50	Comp[¥]	Dup[£]
C. neoformans A1-35-8	Pigeon guano, USA	18,819,906	48.20	557	154,753	91.3	2.2
C. neoformans A2-102-5	Pigeon guano, USA	18,668,348	47.40	47	869,879	92.2	0.4
C. neoformans A5-35-17	Pigeon guano, USA	18,811,654	48.21	592	167,878	89.7	2.2
C. neoformans AD1-7a	Clinical, France	18,468,517	48.20	313*	215,408	90.1	0.3
C. neoformans AD1-83a	Mopane, Botswana	18,636,876	48.20	498	148,764	89.8	0.3
C. neoformans AD2-60a	Clinical, France	18,629,733	48.00	168	730,974	91.0	0.3
C. neoformans BK147	Clinical, Vietnam	18,463,082	-	1,201	127,195	90.4	0.3
C. neoformans BK78	Clinical, Vietnam	18,446,098	-	1,130	126,505	90.4	0.4
C. neoformans BK80	Clinical, Vietnam	18,377,618	-	1,111	111,284	90.3	0.3
C. neoformans BMD1338	Clinical, Vietnam	18,442,244	-	1,183	112,773	90.4	0.4
C. neoformans BMD1367	Clinical, Vietnam	18,403,683	-	946	131,475	90.1	0.3
C. neoformans BMD1415	Clinical, Vietnam	18,447,454	-	1,144	116,22	91.0	0.3
C. neoformans BMD1646	Clinical, Vietnam	18,448,816	48.20	1,063	121,739	90.4	0.4
C. neoformans BMD700	Clinical, Vietnam	18,405,114	-	1,052	111,786	91.9	0.4
C. neoformans Br795	Clinical, Brazil	18,930,238	48.20	625	162,775	89.3	0.5
C. neoformans Bt1	Human, Botswana	18,502,290	47.80	52	770,391	92.1	0.3
C. neoformans Bt120	Clinical (CSF/HIV+), Botswana	18,851,366	46.70	31	1,055,601	91.8	0.4
C. neoformans Bt15	Clinical (CSF /HIV+), Botswana	18,726,126	47.40	33	1,109,861	90.3	0.3
C. neoformans Bt63	Clinical (CSF/HIV+), Botswana	18,793,220	47.10	53	775,306	92.0	0.2
C. neoformans Bt85	Clinical (CSF/HIV+), Botswana	19,238,576	46.30	39	948,332	90.5	0.2
C. neoformans C23	Clinical, USA	18,885,614	48.20	107	747,818	91.7	0.4
*C. neoformans C*45	Clinical (sputum/HIV-), USA	18,184,845	48.15	530	93,455	90.5	0.2
C. neoformans C8	Clinical (CSF/HIV+), USA	18,629,901	47.60	158	573,455	90.5	0.4

(*Continued*)

Strains	Isolation source/ locality	Assembly Size (bp)	GC %	# Scaffolds	Scaffold N50	Comp[¥]	Dup[£]
C. neoformans CCTP51	Clinical, Brazil	18,083,839	48.20	565	87,474	89.4	0.3
C. neoformans CHC193	Clinical (CSF/HIV-), China	18,648,543	48.20	519	153,711	89.9	0.3
C. neoformans CM20	Unknown	18,516,783	48.20	588	92,289	90.9	0.4
C. neoformans CM24	Unknown	18,557,031	48.20	511	108,383	91.2	0.3
C. neoformans CM36	Unknown	18,610,299	48.20	465	134,122	91.0	0.4
C. neoformans CM50	Unknown	18,560,305	48.20	511	113,163	91.1	0.3
C. neoformans CM52	Unknown	18,521,151	48.20	497	101,744	91.3	0.3
C. neoformans CM64	Unknown	18,568,655	48.20	568	109,346	90.0	0.3
C. neoformans D17-1	Pigeon guano, South Africa	18,639,132	48.30	537	163,066	89.8	0.4
C. neoformans Gb118	Pigeon guano, Botswana	18,452,313	47.90	68	782,51	91.2	0.3
C. neoformans H99	Clinical, USA	18,916,112	48.20	15	1,422,463	91.4	0.4
C. neoformans H99	Clinical, USA	18,928,706	48.20	14	1,422,457	89.7	0.4
C. neoformans KN99	Laboratory stock	18,914,647	48.20	15	1,422,463	91.3	0.3
C. neoformans MW-RSA1955	Clinical (CSF/HIV+), South Africa	18,512,298	48.20	487	162,169	91.2	0.3
C. neoformans MW-RSA36	Clinical (CSF/HIV+), South Africa	18,447,467	48.20	457	185,051	90.8	0.4
C. neoformans MW-RSA852	Clinical (CSF/HIV+), South Africa	17,986,510	48.20	574*	70,791	89.9	0.2
C. neoformans RCT21	Unknown	18,505,623	48.20	587	97,825	90.1	0.4
C. neoformans RCT54	Unknown	18,606,003	48.20	494	108,421	89.1	0.2
C. neoformans RCT6	Unknown	18,602,782	48.20	588	95,937	90.3	0.5
C. neoformans Th84	Clinical (blood/HIV+), Thailand	18,598,528	47.60	118	616,613	90.4	0.4
C. neoformans Tu259-1	Mopane Tree, Botswana	18,573,228	47.70	120	794,62	89.8	0.3
C. neoformans Tu401-1	Mopane Tree, Botswana	18,574,041	47.30	49	846,909	91.6	0.3

(*Continued*)

Strains	**Isolation source/ locality**	**Assembly Size (bp)**	**GC %**	**# Scaffolds**	**Scaffold N50**	**Comp[¥]**	**Dup[£]**
C. neoformans V2	Human, Brazil	18,368,897	48.10	669	112,025	88.6	0.4
C. neoformans V31	Human, Brazil	18,436,157	48.10	841	108,481	89.2	0.4
C. neoformans WM-1408	Human, Brazil	18,619,184	48.20	898	301,376	91.5	0.4
C. neoformans Ze90-1	Eucalyptus tree, South Africa	18,658,870	48.30	164	464,547	91.0	0.4
C. sp. Mo29	Rainbow hydrothermal site on the Mid-Atlantic Ridge	23,801,231	53.1	687*	165,604	83.9	0.3
C. tetragattii IND107	Clinical, India	17,649,378	48.0	34	803,676	91.6	0.4
C. wingfieldii CBS 7118^{T}	Frass of a scolytid beetle infesting a specimen of *Olea europaea* subsp. *Africana*, South Africa	19,969,799	53.4	83	606,306	92.8	1.0
C. wingfieldii CBS7118^{T}	Frass of a scolytid beetle infesting a specimen of Olea europaea subsp. Africana, South Africa	20,871,786	53.39	15	1,575,190	93.2	1.0
C. deneoformans B-3501A	Unknown	19,699,782	48.5	14	1,461,964	91.7	0.4
T. arboriformis JCM 14201^{T}	Urine of a patient with chronic renal failure, Japan	19,894,493	60.6	28	1,599,305	89.6	0.5
C. curvatum ATCC 20509	Dairy plant	19,908,169	60.70	16	2,509,747	91.0	0.5
C. curvatum JCM 1532^{T}	Sputum of tubercular patient, Netherland	18,637,344	57.90	75	571,59	88.8	0.3
C. curvatum SBUG-Y 855	Sputum of tubercular patient	16,443,618	59.40	354*	62,631	80.2	0.4
C. cutaneum ACCC 20271	Hangzhou oil refinery, China	30,717,177	57.50	21	5,629,136	94.6	0.4
C. cutaneum B3	Soil, China	38,696,417	55.7	592*	116,762	94.3	62.2
C. cutaneum JCM 1462^{T}	Probably from human clinical specimen	23,155,501	62.00	98	744,446	95.9	0.4
C. cyanovorans JCM 31833^{T}	Soil contaminated with cyanide, South Africa	19,941,766	58.0	90	582,111	88.9	0.2

(*Continued*)

Strains	Isolation source/ locality	Assembly Size (bp)	GC %	# Scaffolds	Scaffold N50	Comp[¥]	Dup[£]
C. daszewskae JCM 11166^T	Skin, Germany	17,225,847	61.0	12	2,186,752	87.5	0.2
C. dermatis JCM 11170^T	Human infected skin, Germany	23,337,637	60.0	37	2,690,106	92.7	0.4
C. mucoides JCM 9939^T	Case of meningitis, Belgium	40,783,511	60.1	84	1,668,497	93.5	64.4
C. oleaginosum ATCC 20508	Dairy plant	19,820,908	60.7	8	3,152,035	90.3	0.6
C. oleaginosum IBC0246	Unknown	19,835,558	60.7	180	216,041	93.3	0.6
Dacryopinax primogenitus DJM-731 SS1^T	Unknown	29,503,487	52.0	99*	1,233,089		
Dioszegia aurantiaca JCM 2956^T	Over-wintered nettle stems of Urtica sp., Canada	19,344,119	53.6	52	1,282,637	88.6	0.1
Dioszegia crocea JCM 2961^T	Strawberries *Fragaria abanassa* (Rosales, Rosaceae)	20,595,339	53.2	26	1,954,963	86.8	0.4
Filobasidium wieringae JCM 11695^T	Dew-retted flax, Netherlands	19,781,320	49.9	36	951,588	82.8	0.4
Holtermannia corniformis JCM 1743	Rotten stump, Japan	20,721,755	55.2	129	584,093	89.8	0.7
H. nyarrowii JCM 11471	Carcass of snow petrel *Pogrodama nivea*, Antarctica	17,735,411	52.7	33	1,099,351	82.7	0.2
Kockovaella imperatae NRRL Y-17943	Unknown	17,465,713	52.2	38*	1,071,374	86.1	0.2
Kwoniella bestiolae CBS 10118^T	Unknown	24,360,772	47.3	12	3,422,270	93.5	0.1
K. dejecticola CBS 10117^T	Frass of the litchi fruit borer *Conopomorpha sinensis*, Vietnam	23,862,392	48.4	13	2,116,304	92.9	0.1
K. heveanensis BCC8398	Insect frass, Thailand	25,469,355	51.9	67	709,378	90.7	0.3
K. heveanensis CBS 569^T	Sheet rubber, Indonesia	25,253,281	51.90	242	174,698	90.5	0.1
K. mangroviensis CBS 10435	Seawater, Bahamas	22,653,759	44.90	37	1,966,685	91.5	0.1
K. mangroviensis CBS 8507^T	Seawater	22,654,511	44.9	62	1,048,156	94.4	0.1
K. mangroviensis CBS 8886	Seawater (mangrove), Bahamas	22,872,713	44.90	41	2,035,775	91.1	0.1
K. pini CBS 10737^T	Dead needles of *Pinus sylvestris*, Russia	20,828,584	40.2	18	1,626,097	88.3	0.1

(*Continued*)

Strains	**Isolation source/ locality**	**Assembly Size (bp)**	**GC %**	**# Scaffolds**	**Scaffold N50**	**Comp[¥]**	**Dup[£]**
K. shandongensis CBS 12478^{T}	Apple orchard soil, China	21,297,554	49.8	132	429,617	95.6	0.1
Mrakia blollopis SK-4	Lake sediment, Antarctica	30,479,820	53.7	167*	1,718,111	87.5	1.6
M. frigida JCM 7857^{T}	Soil, Antarctica	28,616,317	54.00	149	714,753	91.5	2.2
M. frigida Nwmf-AP1	Antarctic soil	28,651,359	54.0	1,539	34,716	82.9	1.0
M. psychrophila NN053900	Alpine glacier soil, China	27,770,102	53.8	1,975*	33,742	83.6	0.9
Naematelia encephala 68-887.2	Unknown	19,786,307	49.3	151*	209,50	81.1	0.2
Naganishia adeliensis IF1SW-F1	International Space Station	19,403,212	53.4	149	506,784		
N. albida JCM 2334^{T}	Air, Japan	20,697,359	53.80	74	975,819	84.3	0.3
N. albida NRRLY-1402^{T}	tubercular lung, Clinical	24,807,186	52.70	834*	114,383	85.7	0.4
N. albida NT2002^{T}	Soil, Chica	20,877,543	54.00	337	331,936	84.4	0.3
N. randhawae eABCC1	Avian guano, South Africa	20,271,596	51.7	386*	191,697	86.8	0.4
Papiliotrema flavescens NRRL Y-50378	Wheat anther, United States	22,790,521	58.5	712	71,416	87.2	0.0
P. laurentii RY1	Food (Kombucha tea); India	19,145,913	56.2	1,152	32,353	87.4	0.5
Pascua guehoae JCM 10690	Soil from a meadow, Netherlands	33,698,914	59.2	35	2,162,681	92.8	0.9
Phaeotremella fagi JCM 13614	Rotten beech *Fagus sylvatica* (*Fagales, Fagaceae*)	22,649,946	42.4	30	1,602,030	90.7	0.1
P. skinneri JCM 9039^{T}	Insect frass beneath the bark of hemlock *Tsuga heterophylla (Pinales, Pinaceae*)	20,788,256	50.2	58	1,003,985	90.2	0.1
Phaffia rhodozyma CBS 7918^{T}	Plant exudate	18,757,695	47.20	345	104,118	76.7	0.5
P. rhodozyma JCM 9681^{T}	Flux of *Betula verrucosa* (Fagales, Betulaceae), Russia	19,108,479	47.00	43	1,120,801	77.8	0.4
P. rhodozyma Xden1	sap on stump of Betula sp. (birch tree), Finland	19,519,467	47.30	775*	109,834	74.3	0.4
Piloderma croceum F 1598	Unknown	59,326,866	46.6	715	529,349		

(*Continued*)

Strains	**Isolation source/ locality**	**Assembly Size (bp)**	**GC %**	**# Scaffolds**	**Scaffold N50**	**Comp[¥]**	**Dup[£]**
Prillingera fragicola JCM 1530^{T}	Strawberry, Japan	20,263,689	58.20	21	1,928,959	93.3	0.6
Saitozyma podzolica DSM 27192	Peat bog soil, Germany	29,888,215	58.80	46*	1,066,819	89.0	0.5
Saitozyma sp. JCM 24511	Soil, Japan	28,180,937	58.5	55	1,281,587	90.9	0.2
Sistotremastrum niveocremeum HHB9708	Unknown	35,361,098	49.2	230	695,997		
Solicoccozyma phenolica JCM 11743	Decaying wood, Portugal	22,263,516	54.90	71	969,091	79.5	0.3
S. terricola JCM 24518	Soil, Japan	24,117,869	49.40	79	1,050,647	76.4	0.4
S. terricola JCM 24523	Soil, Japan	23,185,503	49.40	156	658,505	79.1	0.4
Takashimella koratensis JCM 12878^{T}	Leaf of *Lagerstroemia calyculata* (*Myrtales*, *Lythraceae*), Thailand	25,142,362	54.90	31	1,454,648	92.4	0.4
T. tepidaria JCM 11965^{T}	Water sample from a small stream, Japan	22,370,450	44.70	44	1,186,041	89.7	0.3
Tremella fuciformis TR26	*Annulohypoxylon stygium*	23,635,607	57.0	3,502*	18,448	82.7	0.3
T. mesenterica ATCC 28783	*Alnus rubra*, Canada	27,110,854	41.30	295	180,988	81.3	0.1
T. mesenterica DSM 1558	*Alnus rubra*	28,639,919	46.80	45	1,622,698	80.1	0.2
Trichosporon asahii CBS 2479	*Trichosporia cutis psoriatiformis progressiva*, Japan	24,540,311	59.46	78	1,660,894	93.6	0.4
T. asahii CBS 8904	Maize cobs	25,299,608	59.36	194	3,223,897	93.9	0.7
T. asahii N5_275_008G1	Human faeces, USA	23,418,624	59.60	1,022	32,912	87.5	0.2
T. asahii JCM 2466^{T}	*Trichosporia cutis psoriatiformis progressiva*, Japan	24,687,929	59.4	36	2,256,092	95.4	1.4
T. coremiiforme JCM 2938^{T}	Lesion on farmer head caused by bee sting, Costa Rica	42,353,277	59.6	190	1,468,092	95.0	72.7

(*Continued*)

Strains	**Isolation source/ locality**	**Assembly Size (bp)**	**GC %**	**# Scaffolds**	**Scaffold N50**	**Comp[¥]**	**Dup[£]**
T. faecale JCM 2941^T	Human faeces	24,653,913	60.20	32	3,676,950	94.3	0.7
T. inkin JCM 9195	Human skin	20,339,538	62.70	18	2,739,924	93.4	0.1
T. ovoides JCM 9940	Scalp infected with white piedra	40,322,879	60.20	116	2,824,449	95.8	76.3
Vanrija humicola CBS 4282	Fresh water, Slovakia	22,632,906	62.80	21	1,793,818	94.3	0.4
V. humicola JCM 1457^T	Soil	22,653,840	62.70	10	3,082,120	96.5	0.4
V. humicola UJ1	Unknown	22,628,423	62.70	46	1,340,400	95.2	0.4

[*]Contigs, [¥]Completeness and [£]duplication determined using BUSCO v3.0.3 (Simão et al. 2015) (BUSCO > 70) based on fungi_odb9.

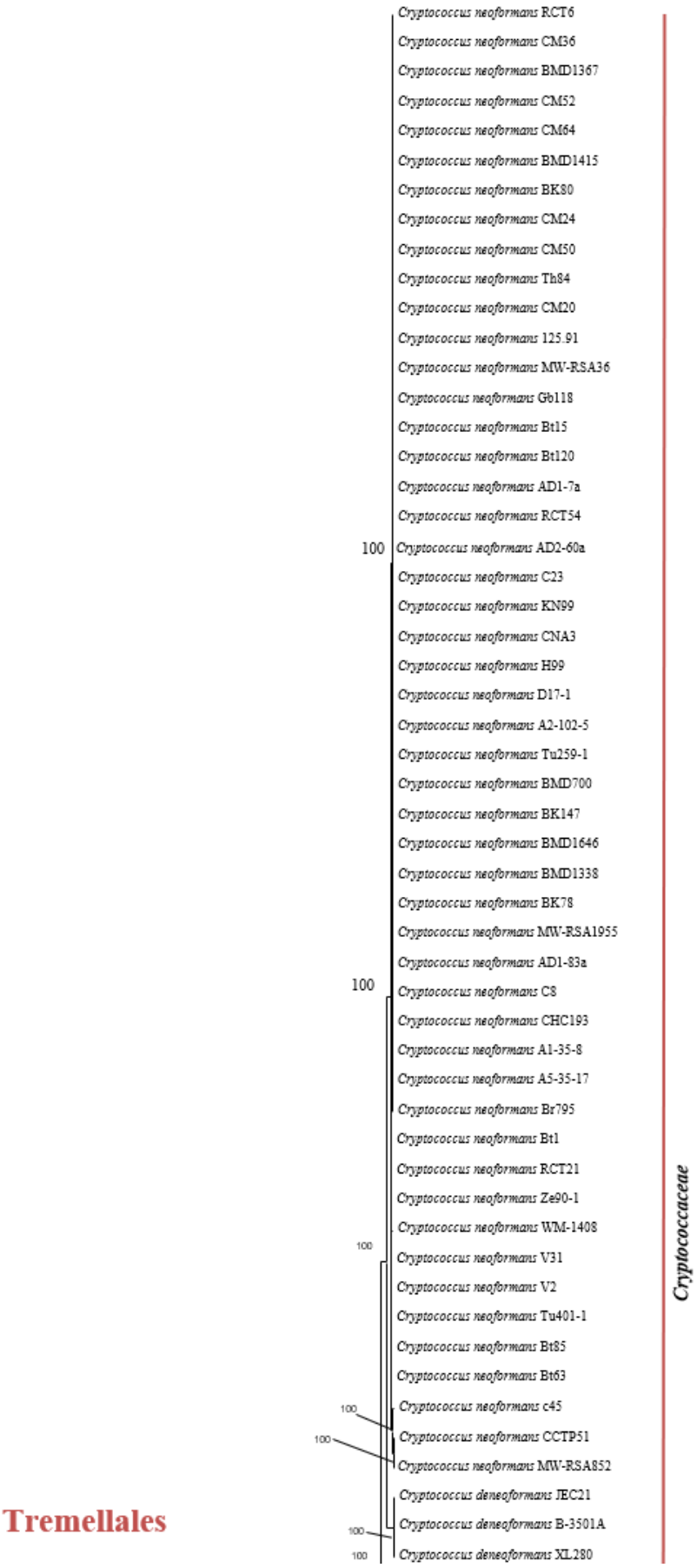


Figure 2 continued

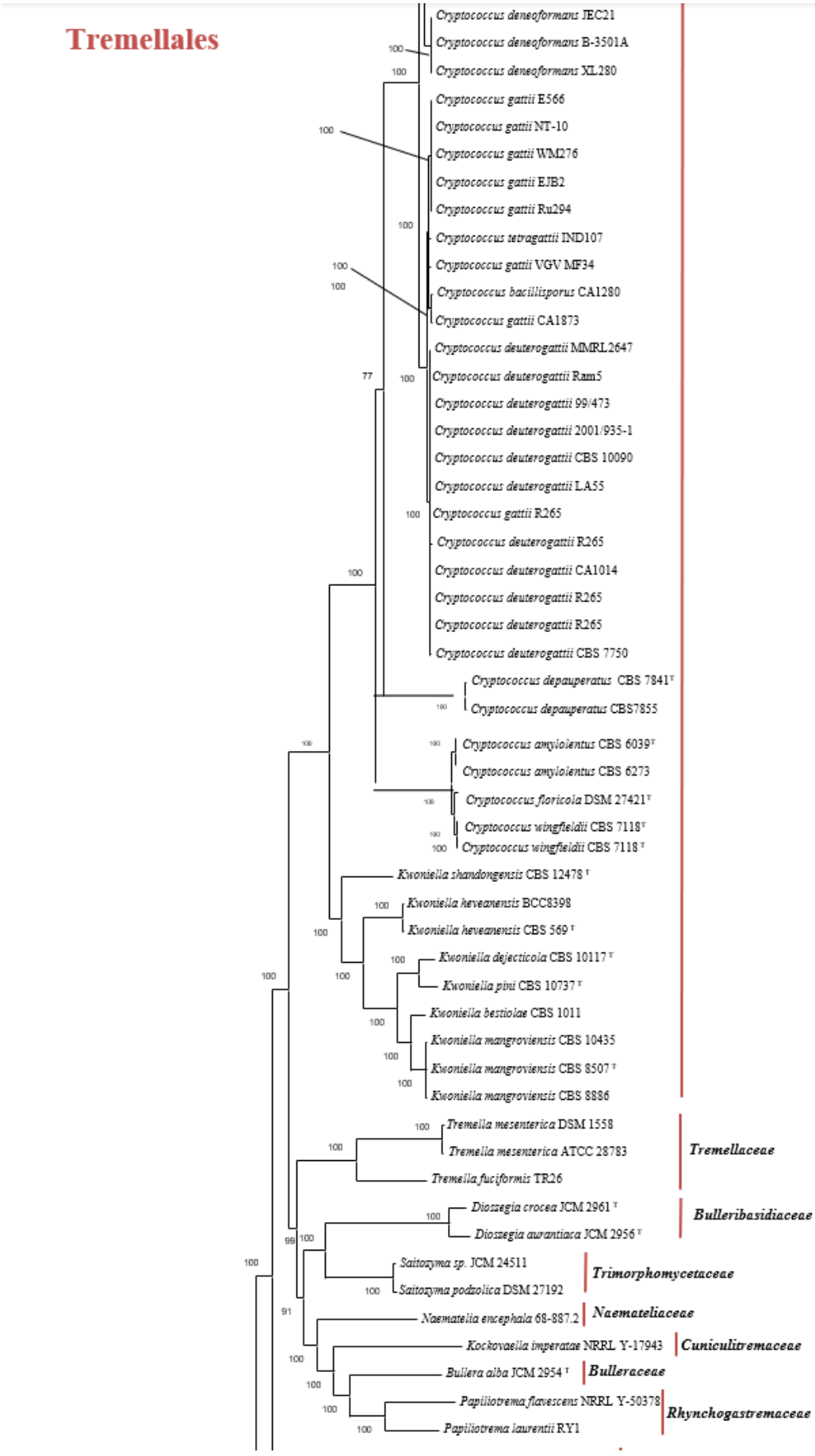


Figure 2 continued

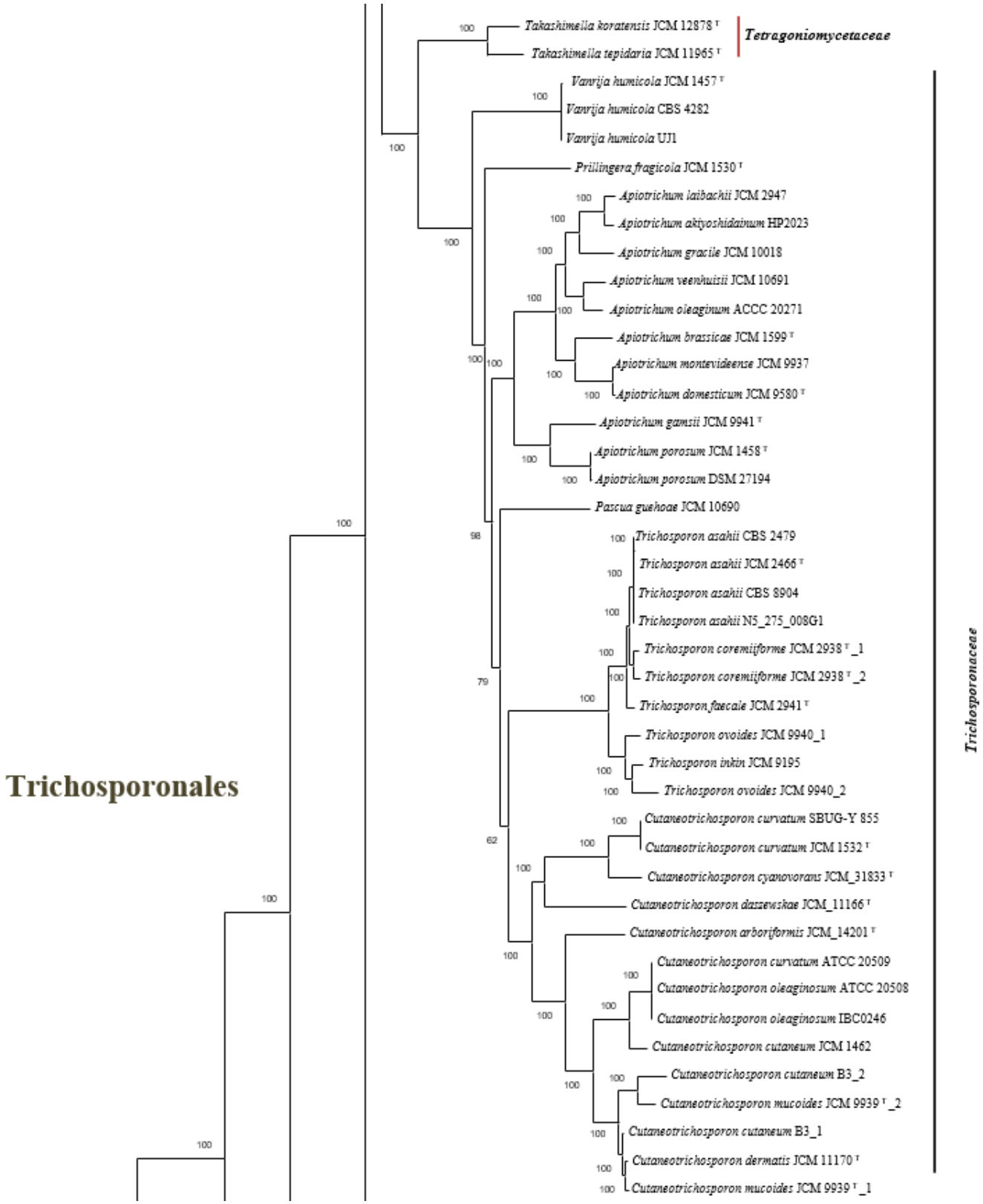


Figure 2 continued

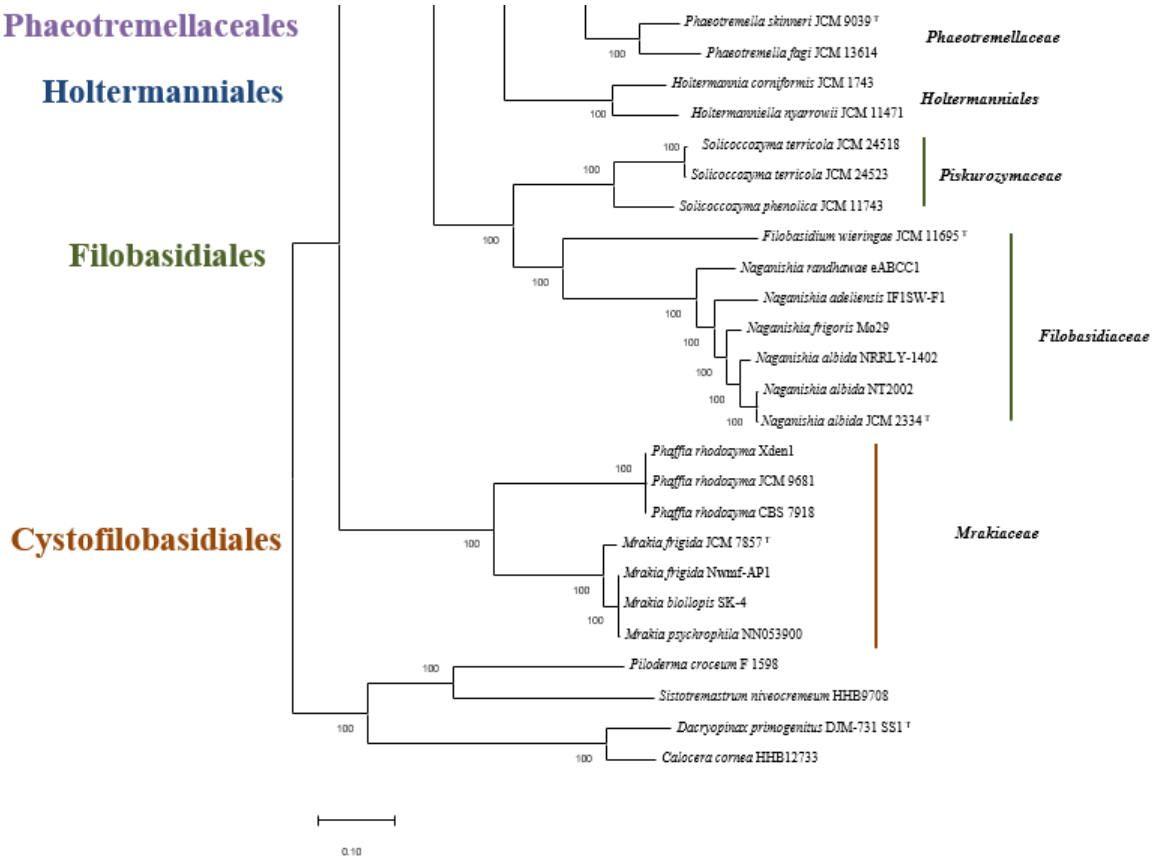


Figure 4.2 Phylogenomic analysis of 161members of the class Tremellomycetes. Maximum likelihood (ML) tree inferred from the concatenated protein alignment (96,605 amino acid) of 452 single copy orthologues (Guindon *et al.* 2010). The phylogeny was generated using IQ-TREE v1.6.7 with confidence values based on 1,000 bootstrap replicates (Nguyen *et al.*, 2015). The labels '_1' and '_2' indicate the two sets of single copy orthologues (SCOs) in the hybrid genomes.

4.4.3 Phylogenomic analyses of the genus *Cryptococcus* reveals poor delineation between strains

The *C. neoformans* / *C. gattii* species complex has been extensively studied due to its medical importance. These strains have been classified using several techniques such as antigen testing, multilocus sequence typing, proteome spectroscopy and whole genome sequencing (Hagen *et al.*, 2015; May *et al.*, 2016; Nnadi *et al.*, 2019). Initially, the genus *Cryptococcus* was represented by a single species, *C. neoformans*, with four distinct serotypes (A-D) (Kwon-Chung, 1976; Kwon-Chung and Bennett, 1978). These serotypes were later separated into varieties; *C. neoformans* var. *gattii* (serotypes B and C*)*, *C. neoformans* var. *grubii* (Serotype A), and *C. neoformans* var. *neoformans* (Serotype D) (Kwon-Chung *et al.*, 1982).

Analysis of global isolates through multilocus sequence typing and whole genome sequencing provided tangible evidence for the subdivision of the species complex into distinct species. As such, *C. neoformans* and *C. deneoformans* have been designated from the *Cryptococcus neoformans* complex; and *C. bacillisporus*, *C. decagattii*, *C. deuterogattii*, *C. gattii*, and *C. tetragattii* from the *C. gattii* complex (Hagen *et al.*, 2015).

In this study, 82 genomes belonging to the genus *Cryptococcus* were included. We observed clear demarcation between *C. neoformans* and *C. deneoformans*. The *C. gattii* species complex however, displays very poor discrimination between the different proposed novel species (Figure 2). For instance, *C. gattii* CA1873 clustered with *C. bacillisporus* CA1280 and *C. gattii* R265 clustered with *C. deuterogattii* strains (Figure 4.2).

These observations are consistent with the reported overall genome relatedness indices (OGRI) of *Cryptococcus* species (Chun and Rainey, 2014). The very high OGRI indices between *C. gattii* and *C. bacillisporus* (ANI%=95.2; dDDH%=61.1 and Kr=0.06); and *C. gattii* and *C. decagattii* (ANI%=94.8; dDDH%=58.7 and Kr=0.05) implies that *C. gattii* strains are in fact closely related (Libkind *et al.*, 2020). These findings strengthen previous claims that the proposed taxonomic revision of the *C. neoformans* / *C. gattii* species complex was premature and will eventually result in nomenclatural instability (Kwon-Chung *et al.*, 2017).

Clear demarcation between the *C. gattii* strains is of the utmost importance as discrimination between these different strains can lead to more accurate clinical diagnoses. The abundance of available *Cryptococcus* genomes could allow for a comprehensive phylogenomic analysis that will examine only the core proteins. Such analysis does not fall within the scope of this study but should be considered for future analysis.

4.4.4 Members of the family *Phaeotremellaceae* do not cluster within the Tremellales

The family *Phaeotremellaceae* was proposed to accommodate species from the *foliacea* clade regrouping former *Tremella* and *Cryptococcus* species (Liu *et al.*, 2015a). The *foliacea* clade was derived from its type strain *Phaeotremella foliacea* (Basionym *Tremella foliacea*), a basidiocarp producing jelly fungus (Roberts, 1995). This clade initially clustered within the polyphyletic genus *Tremella* based on a combination of molecular and morphological evidence (Chen, 1998). The older generic name *Phaeotremella* for the *foliacea* clade was revived based on the work conducted by Liu and colleagues (2015a). In the same study, the *foliacea* clade clustered within the *Tremellales* in the Bayesian tree without posterior probabilities (PP) support (Newton *et al.*, 1999). Maximum likelihood and NJ analyses showed that this clade

branched prior to the Tremellales and Trichosporonales with very poor bootstrap support (Liu *et al.*, 2015b). In a recent phylogenomic analysis of the Tremellales, members of the family *Phaeotremellaceae* (*P. skinneri* JCM 9039^{T} and *P. fagi* JCM 13614) formed a monophyletic clade distinct from the rest of the Tremellales (Aliyu *et al.*, 2020b). Similarly, our phylogenomic tree indicates that *P. skinneri* JCM 9039^{T} and *P. fagi* JCM 13614 form a distinct clade basal to the Tremellales and Trichosporonales (Figure 4.2). We propose the order Phaeotremellaceales **ord. nov.** for taxa belonging to the family *Phaeotremellaceae*.

4.4.5 Phylogenomic analysis of the order Trichosporonales is congruent with a previous phylogenomic study

The order Trichosporonales was introduced to accommodate arthroconidia-producing species based on the D1/D2 domain of the LSU rRNA gene (Fell *et al.*, 2000). Liu and colleagues (2015a) proposed two families, namely the *Trichosporonaceae* and *Tetragoniomycetaceae* comprising of ten genera (Liu *et al.*, 2015b). With the rise in available fungal genomes, the order Trichosporonales has served as a model for a more rigorous fungal phylogenomic framework. For instance, the work of Takashima and colleagues (2018, 2019) reported a genome-based characterization and phylogenetic study of the order Trichosporonales, suggesting the integration of two new genera, *Pascua* and *Prillingera*. *Cutaneotrichosporon cutaneum* ACCC20271 delineates with taxa within the genus *Apiotrichum*. Aliyu and colleagues (2020a) suggested that this strain should be reassigned in the genus *Apiotrichum* based on the single copy orthologue phylogeny of members of the fungal family *Trichosporonaceae*. Therefore, we propose a new combination *Apiotrichum oleaginum* **comb. nov.**

All four hybrid genome included in this study belong to the family *Trichosporonaceae* and have been reported in previous studies (Aliyu *et al.*, 2020a; Takashima *et al.*, 2018). The putative hybrid genome *C. cutaneum* B3_1 clustered with *C. mucoides* JCM 9939T_1, while *C. cutaneum* B3_2 clustered with *C. mucoides* JCM 9939T_2. The remaining hybrid sub genomes formed distinct branches but were still retained within their genus clades (Figure 4.2). These observations are consistent with what was found in previous phylogenomic analysis of the *Trichosporonaceae* implying that they share similar evolutionary histories, with duplication likely having derived subsequent to the evolution of the genus (Aliyu *et al.*, 2020a). Our study also confirms a previous report that *Pascua guehoae* JCM 10690 (*Cutaneotrichosporon guehoae*) and *Prillingera fragicola* JCM1530 (*Vanrija fragicola*) do not cluster with the genera

Cutaneotrichosporon and *Vanrija* respectively, further supporting the suggestion that these represent novel genera (Takashima *et al.*, 2019).

4.4.6 Phylogenomic analysis places *Cryptococcus* sp. Mo29 in the genus *Naganishia*

Our analysis reveals that *Cryptococcu*s sp. Mo29 delineates within the genus *Naganishia*. The genus *Naganishia* was originally proposed to accommodate the yeast *Naganishia globosus* (Goto, 1963). This species was later re-assigned to its synonym *Cryptococcus saitoi*, resulting in the extension of the genus (Fonseca *et al.,* 2000). However, Liu and colleagues re-established the genus in order to resolve the taxonomic classification of the yeast genus *Cryptococcus* (Fell *et al.* 1999; Fonseca *et al.* 2000; Scorzetti *et al.* 2002; Liu *et al.* 2015a). Metadata indicates that *Cryptococcu*s sp. Mo29 was isolated from the Rainbow hydrothermal site of the Mid-Atlantic Ridge (2300m depth). Similarly, several members of the genus *Naganishia* have been isolated from extreme terrestrial environments characterised by cold temperatures (Goto *et al.*, 1969; Schmidt *et al.,* 2017; Vishniac and Hempfling, 1979). We propose a new combination *Naganishia frigoris* **comb. nov.** within the genus *Naganishia*.

4.5 Conclusion

Here we present a phylogenomic framework for the class Tremellomycetes. A total of 161 taxa from five orders were included, with the majority of strains from the Tremellales and Trichosporonales. Congruencies in the findings with other phylogenomic analyses underline the importance of using genome-scale data to address fungal classification. Discrepancies in species clustering within different genera highlight the danger of relying solely on single molecular markers such as rDNA and ITS to propose new taxa. For this reason, we recommend the use of genome scale data for the taxonomic placement of novel species. This genome-level analysis has revealed several taxonomic considerations, including the proposal of a new order for the family *Phaeotremellaceae* (Phaeotremellaceales **ord. nov.**) and two new species combinations: *Apiotrichum oleaginum* **comb. nov.** and *Naganishia frigoris* **comb. nov.** One significant issue observed in these analyses was the poor delineation within the genus *Cryptococcus*. Their positions may be more accurately resolved by performing a thorough analysis of the SCOs shared by members of the genus. The inclusion of multiple species from the poorly represented genera may better resolve the class Tremellomycetes at the family, genus, and species levels. Findings from this study could pave the way for future phylogenomic studies of fungi at higher taxonomic ranks.

4.6 References

Aliyu, H., Gorte, O., De Maayer, P., Neumann, A. and Ochsenreither, K., 2020a. Genomic insights into the lifestyles, functional capacities and oleagenicity of members of the fungal family *Trichosporonaceae*. *Scientific Reports*, *10*(1), pp.1-12.

Aliyu, H., Gorte, O., Zhou, X., Neumann, A. and Ochsenreither, K., 2020b. In silico proteomic analysis provides insights into phylogenomics and plant biomass deconstruction potentials of the tremelalles. *Frontiers in bioengineering and biotechnology*, *8*, p.226.

Barrett, T., Clark, K., Gevorgyan, R., Gorelenkov, V., Gribov, E., Karsch-Mizrachi, I., Kimelman, M., Pruitt, K.D., Resenchuk, S., Tatusova, T. and Yaschenko, E., 2012. BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic acids research*, *40*(D1), pp.57-63.

Bidartondo, M.I., Read, D.J., Trappe, J.M., Merckx, V., Ligrone, R. and Duckett, J.G., 2011. The dawn of symbiosis between plants and fungi. *Biology letters*, *7*(4), pp.574-577.

Bridge, P.D., Spooner, B.M. and Roberts, P.J., 2005. The impact of molecular data in fungal systematics. *Advances in botanical research*, *42*, pp.33-67.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), pp.540-552.

Chen, C.J., 1998. Morphological and molecular studies in the genus Tremella.

Chen, N., 2004. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, *5*(1), pp.4-10.

Chun, J. and Rainey, F.A., 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *International journal of systematic and evolutionary microbiology*, *64*(2), pp.316-324.

Desdevises, Y., Morand, S., Jousson, O. and Legendre, P., 2002. Coevolution between Lamellodiscus (Monogenea: Diplectanidae) and Sparidae (Teleostei): the study of a complex host-parasite system. *Evolution*, *56*(12), pp.2459-2471.

Dornburg, A., Townsend, J.P. and Wang, Z., 2017. Maximizing power in phylogenetics and phylogenomics: a perspective illuminated by fungal big data. *Advances in Genetics*, *100*, pp.1-47.

Emms, D.M. and Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), pp.1-14.

Fell, J.W., Boekhout, T.F. and Fonseca, A., Á., Scorzetti, G. & Statzell-Tallman, A. 2000. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *International Journal of Systematic and Evolutionary*, *50*, pp.1351-1372.

Fonseca, Á., Scorzetti, G. and Fell, J.W., 1999. Diversity in the yeast Cryptococcus albidus and related species as revealed by ribosomal DNA sequence analysis. *Canadian journal of microbiology*, *46*(1), pp.7-27.

Gabaldón, T. and Marcet-Houben, M., 2014. 3 Phylogenomics for the Study of Fungal Biology. In *Fungal Genomics* (pp. 61-79). Springer, Berlin, Heidelberg.

Goto, S., 1963. On a new yeast genus Naganishia. *Journal of Fermentation Technology*, *41*, pp.459-462.

Goto, S., Sugiyama, J. and Iizuka, H., 1969. A taxonomic study of Antarctic yeasts. *Mycologia*, *61*(4), pp.748-774.

Guarro, J., Gené, J. and Stchigel, A.M., 1999. Developments in fungal taxonomy. *Clinical Microbiology Reviews*, *12*(3), pp.454-500.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), pp.307-321.

Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., Falk, R., Parnmen, S., Lumbsch, H.T. and Boekhout, T., 2015. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. *Fungal Genetics and Biology*, *78*, pp.16-48.

Hall, T.A., 1999, January. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series* (Vol. 41, No. 41, pp. 95-98.

Hibbett, D.S. and Taylor, J.W., 2013. Fungal systematics: is a new age of enlightenment at hand. *Nature Reviews Microbiology*, *11*(2), pp.129-133.

Hibbett, D.S., 2006. A phylogenetic overview of the Agaricomycotina. *Mycologia*, *98*(6), pp.917-925.

Hibbett, D.S., 2006. A phylogenetic overview of the Agaricomycotina. *Mycologia*, *98*(6), pp.917-925.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lücking, R. and Lumbsch, H.T., 2007. A higher-level phylogenetic classification of the Fungi. *Mycological Research*, 111(5), pp.509-547.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J. and Lumbsch, H.T., 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, *443*(7113), pp.818-822.

Jenuth, J.P., 2000. The ncbi. In *Bioinformatics Methods and Protocols*, pp. 301-312.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A. and Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, *14*(6), pp.587-589.

Kwon-Chung K. J., Bennett J. E. (1978). Distribution of alpha and alpha mating types of *Cryptococcus neoformans* among natural and clinical isolates. *American Journal of Epidemiology* 108(4): 337-40.

Kwon-Chung, K.J., 1976. A new species of Filobasidiella, the sexual state of *Cryptococcus neoformans* B and C serotypes. *Mycologia*, *68*(4), pp.942-946.

Kwon-Chung, K.J., Bennett, J.E. and Rhodes, J.C., 1982. Taxonomic studies on *Filobasidiella* species and their anamorphs. *Antonie van Leeuwenhoek*, *48*(1), pp.25-38.

Kwon-Chung, K.J., Bennett, J.E., Wickes, B.L., Meyer, W., Cuomo, C.A., Wollenburg, K.R., Bicanic, T.A., Castañeda, E., Chang, Y.C., Chen, J. and Cogliati, M., 2017. The case for adopting the "species complex" nomenclature for the etiologic agents of cryptococcosis. *MSphere*, *2*(1).

Kwon-Chung, K.J., Boekhout, T., Wickes, B.L. and Fell, J.W., 2010. Systematics of the genus *Cryptococcus* and its type species *C. neoformans*. *Cryptococcus: From human pathogen to model yeast*, pp.1-15.

Libkind, D., Čadež, N., Opulente, D.A., Langdon, Q.K., Rosa, C.A., Sampaio, J.P., Gonçalves, P., Hittinger, C.T. and Lachance, M.A., 2020. Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Research*, *20*(6), p. 38-42.

Liu, X.Z., Wang, Q.M., Göker, M., Groenewald, M., Kachalkin, A.V., Lumbsch, H.T., Millanes, A.M., Wedin, M., Yurkov, A.M., Boekhout, T. and Bai, F.Y., 2015b. Towards an integrated phylogenetic classification of the Tremellomycetes. Studies in mycology, 81, pp.85-147.

Liu, X.Z., Wang, Q.M., Theelen, B., Groenewald, M., Bai, F.Y. and Boekhout, T., 2015a. Phylogeny of tremellomycetous yeasts and related dimorphic and filamentous basidiomycetes reconstructed from multiple gene sequence analyses. *Studies in Mycology*, *81*, pp.1-26.

Lutzoni, F., Kauff, F., Cox, C.J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T.Y., Baloch, E. and Grube, M., 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *American Journal of Botany*, *91*(10), pp.1446-1480.

Magis, C., Taly, J.F., Bussotti, G., Chang, J.M., Di Tommaso, P., Erb, I., Espinosa-Carrasco, J. and Notredame, C., 2014. T-Coffee: tree-based consistency objective function for alignment evaluation. In *Multiple Sequence Alignment Methods* (pp. 117-129). Humana Press, Totowa, NJ.

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

Mohanta, T.K. and Bae, H., 2015. The diversity of fungal genome. *Biological Procedures Online*, *17*(1), pp.1-9.

Money, N.P., 2016. Fungal diversity. *The Fungi*, pp. 1-36.

Newton, M.A., Mau, B. and Larget, B., 1999. Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. *Lecture Notes-Monograph Series*, pp.143-162.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. and Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), pp.268-274.

Nnadi, N.E., Giosa, D., Ayanbimpe, G.M., D'Alessandro, E., Cigliano, R.A., Oheri, C.U., Aguiyi, J.C., Enweani, I.B. and Romeo, O., 2019. Whole-Genome Sequencing of an Uncommon *Cryptococcus neoformans* MLST43 Genotype Isolated in Nigeria. *Mycopathologia*, *184*(5), pp.555-557.

Notredame, C., 2000. higgins, DJ, heringa, J. T-Coffee: A novel method for multiple sequence alignments. *J. Mol. Biol*, *302*, pp.205-217.

Palmer JM. 2016. Funannotate: a fungal genome annotation and comparative genomics pipeline. Available at: https://github.com/nextgenusfs/funannotate.

Relman, D.A., 2011. Microbial genomics and infectious diseases. *New England Journal of Medicine*, *365*(4), pp.347-357.

Ren, R., Sun, Y., Zhao, Y., Geiser, D., Ma, H. and Zhou, X., 2016. Phylogenetic resolution of deep eukaryotic and fungal relationships using highly conserved low-copy nuclear genes. *Genome Biology and Evolution*, *8*(9), pp.2683-2701.

Roberts, P., 1995. British Tremella species I: *Tremella aurantia* & *T. mesenterica*. *Mycologist*, *9*(3), pp.110-114.

Saenz, G.S., Taylor, J.W. and Gargas, A., 1994. 18S rRNA gene sequences and supraordinal classification of the *Erysiphales*. *Mycologia*, *86*(2), pp.212-216.

Schmidt, S.K., Vimercati, L., Darcy, J.L., Arán, P., Gendron, E.M., Solon, A.J., Porazinska, D. and Dorador, C., 2017. A *Naganishia* in high places: functioning populations or dormant cells from the atmosphere? *Mycology*, *8*(3), pp.153-163.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W. and Fungal Barcoding Consortium, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), pp.6241-6246.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single copy orthologs. *Bioinformatics*, *31*(19), pp.3210-3212.

Stajich, J.E., 2017. Fungal genomes and insights into the evolution of the kingdom. *The Fungal Kingdom*, pp.619-633.

Stajich, J.E., Berbee, M.L., Blackwell, M., Hibbett, D.S., James, T.Y., Spatafora, J.W. and Taylor, J.W., 2009. Primer--the fungi. *Current biology: CB*, *19*(18), p.R840.

Storck, R., 1966. Nucleotide Composition of Nucleic acids of fungi II. Deoxyribonucleic acids. *Journal of Bacteriology*, *91*(1), pp.227-230.

Swann, E.C. and Taylor, J.W., 1993. Higher taxa of basidiomycetes: an 18S rRNA gene perspective. *Mycologia*, *85*(6), pp.923-936.

Takashima, M., Manabe, R.I., Nishimura, Y., Endoh, R., Ohkuma, M., Sriswasdi, S., Sugita, T. and Iwasaki, W., 2019. Recognition and delineation of yeast genera based on genomic data: Lessons from Trichosporonales. *Fungal Genetics and Biology*, *130*, pp.31-42.

Takashima, M., Sriswasdi, S., Manabe, R.I., Ohkuma, M., Sugita, T. and Iwasaki, W., 2018. A Trichosporonales genome tree based on 27 haploid and three evolutionarily conserved 'natural'hybrid genomes. *Yeast*, *35*(1), pp.99-111.

Talavera, G. and Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), pp.564-577.

Vishniac, H.S. and Hempfling, W.P., 1979. *Cryptococcus vishniacii* sp. nov., an Antarctic yeast. *International Journal of Systematic and Evolutionary Microbiology*, *29*(2), pp.153-158.

Vishniac, H.S. and Hempfling, W.P., 1979. Cryptococcus vishniacii sp. nov., an Antarctic yeast. *International Journal of Systematic and Evolutionary Microbiology*, *29*(2), pp.153-158.

Wang, Z., Nilsson, R.H., James, T.Y., Dai, Y. and Townsend, J.P., 2016. Future perspectives and challenges of fungal systematics in the age of big data. In *Biology of Microfungi* (pp. 25-46). Springer, Cham.

Weiss, M., Bauer, R., Sampaio, J.P. and Oberwinkler, F., 2014. 12 Tremellomycetes and Related Groups. *Systematics and Evolution*. pp. 331-355.

Weiss, M., Bauer, R., Sampaio, J.P. and Oberwinkler, F., 2014. 12 Tremellomycetes and related groups. In *Systematics and Evolution* (pp. 331-355). Springer, Berlin, Heidelberg.

Chapter 5 Summary

Summary

Members of the genus *Naganishia* have a global distribution with a predominance in harsh environments. Literature has shown that some taxa within the genus are opportunistic pathogens while others are of great biotechnological values. Nonetheless, little is known about the biology of these species and genomic data available does not cover the wide diversity of the genus.

Naganishia randhawae eABCC1 was isolated from avian guano in Johannesburg, South Africa. Uncharacteristically, this strain failed to grow at 37 °C and produced melanin when cultured on birdseed agar (BSA). To gain insights into the biology of this strain, the genome of *N. randhawae* eABCC1 was sequenced, assembled, and annotated in Chapter 2. The *Naganishia randhawae* eABCC1genome is 19.90 Mb in size, has a G+C content of 51.71 %, codes for 6,775 proteins and incorporates 168 tRNA genes. The genome codes for proteins linked with melanin production (which was observed in this strain) and putative pathogenicity proteins such as those involved in capsule formation. However, these features are not limited to a role in pathogenicity. Given the diverse habits that members of this genus inhabit, capsule formation and melanin production could also play a role in environmental survival.

Comparative genome analysis (Chapter 3), together with proteome mining, revealed that *Naganishia* species, including *N. randhawae* eABCC1, possess several proteins that could play a role in stress tolerance. For instance, a cluster of proteins only shared by *Naganishia* species was assigned to the Gene ontology function: “ultraviolet (UV)-damage excision repair”. These genes include members of the ATP-binding cassette which play a role in the repair of UV-damaged DNA in *Escherichia coli*. Similarly, proteins associated with melanisation and photoprotection were also found, implying additional protection during solar irradiation exposure. The identification of two putative laccases within the genome of *N. randhawae* eABCC1 could explain the production of melanin when this strain is cultured on BSA, however, deletion or metabolomic studies will be needed to confirm laccase activity and melanin biosynthesis.

Lastly, in Chapter 4, a phylogenomic analysis of using the genomes 161 taxa within the class Tremellomycetes was performed. The core genome phylogeny, comprised of 452 single copy orthologous (SCO) proteins, displayed congruence with genome wide phylogeny at lower taxonomic ranks. This analysis offers clear resolution at both the higher and lower taxonomic level for most taxa, with the exception of members of the genus *Cryptococcus*, where more

genus-specific markers may be required for adequate delineation. Taxonomic considerations presented in this study includes the proposal of a new order Phaeotremellaceales **ord. nov.** and two new species combinations *Apiotrichum oleaginum* **comb. nov.** and *Naganishia frigoris* **comb. nov.** Findings from study will expand our knowledge of the evolution and adaptation of members of the genus *Naganishia* as well as pave the way for future phylogenomic studies of fungi at higher taxonomic ranks.

Appendix

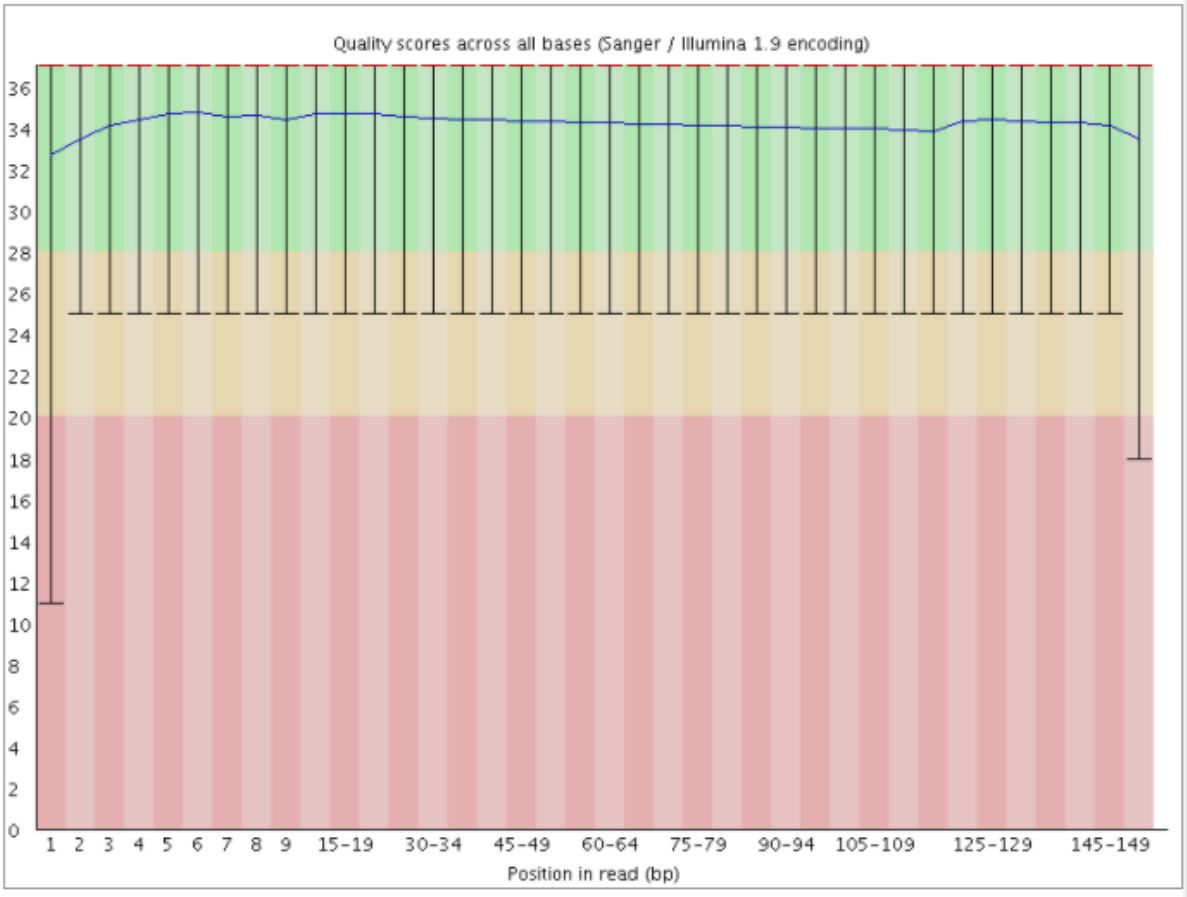
FASTQC report of the *Naganishia randhawae* eABCC1 genome

File 1: T1BOTES_S1_L001_R1_001.fastq.gz

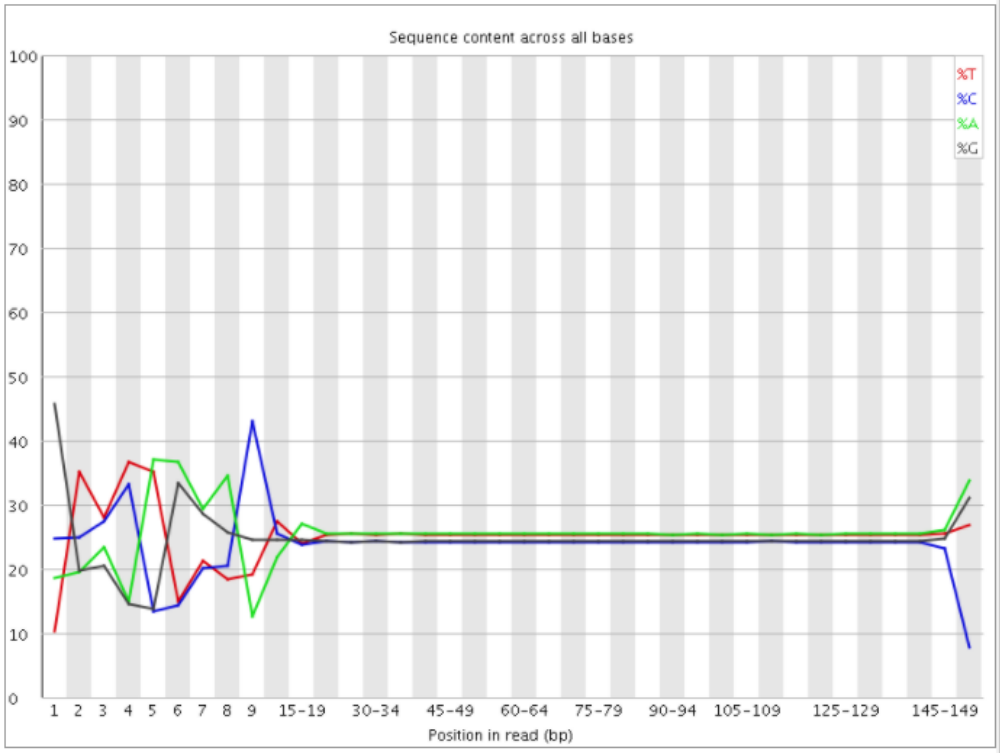
Basic Statistics

Measure	Value
Filename	T1BOTES_S1_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	8076080
Filtered Sequences	0
Sequence length	120-151
%GC	48

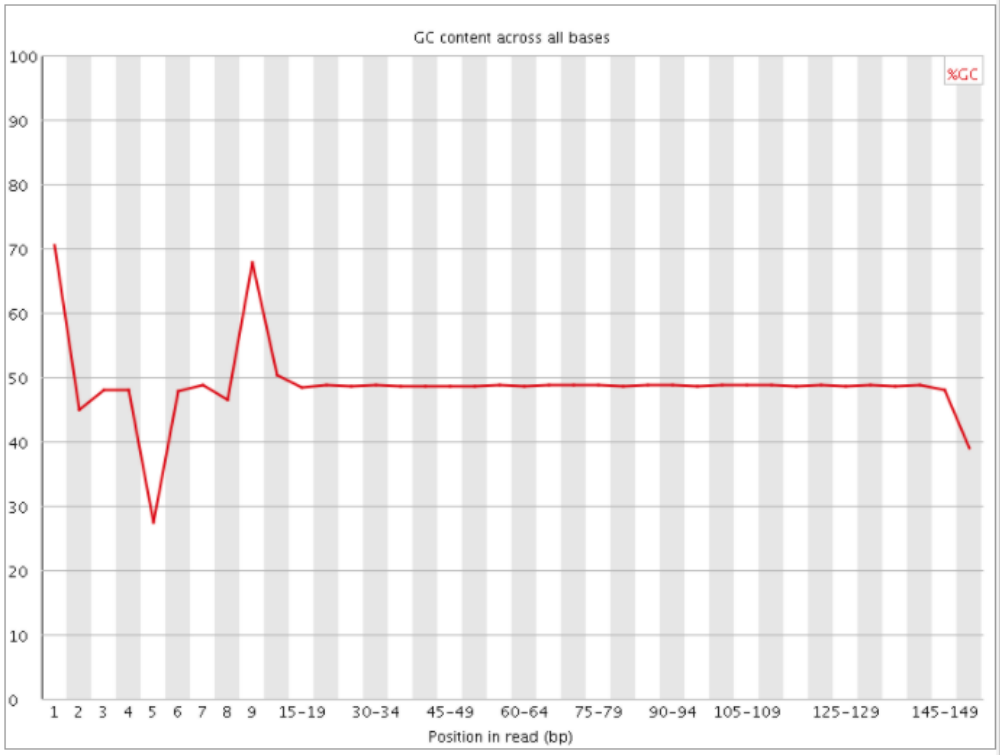
Per Base Sequence Quality



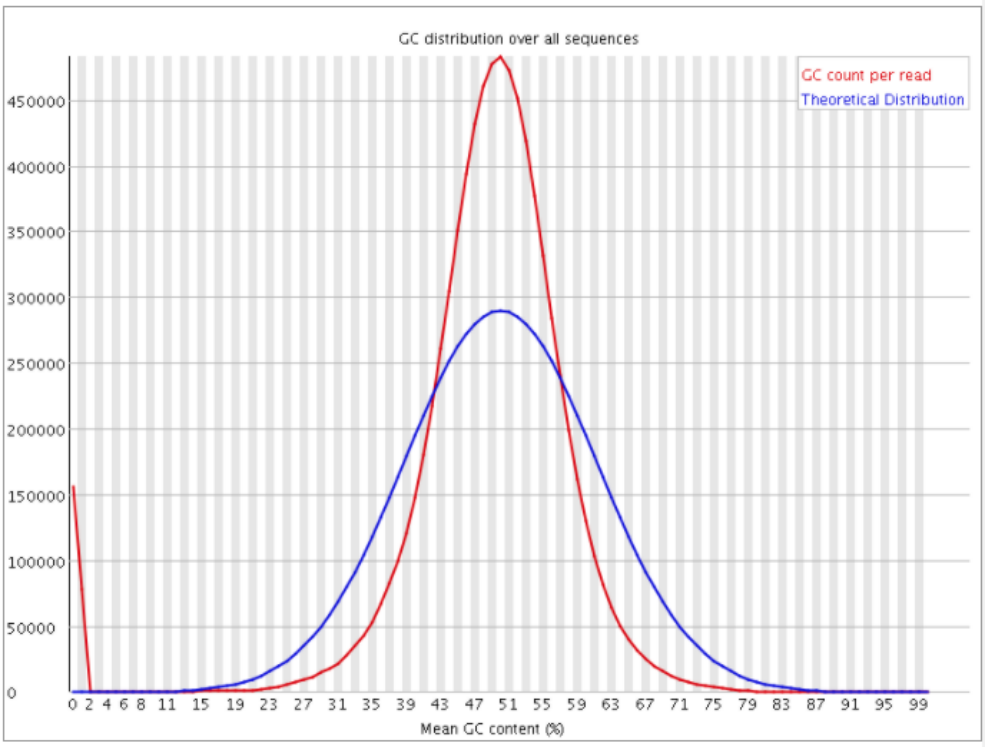
Per Base Sequence Content



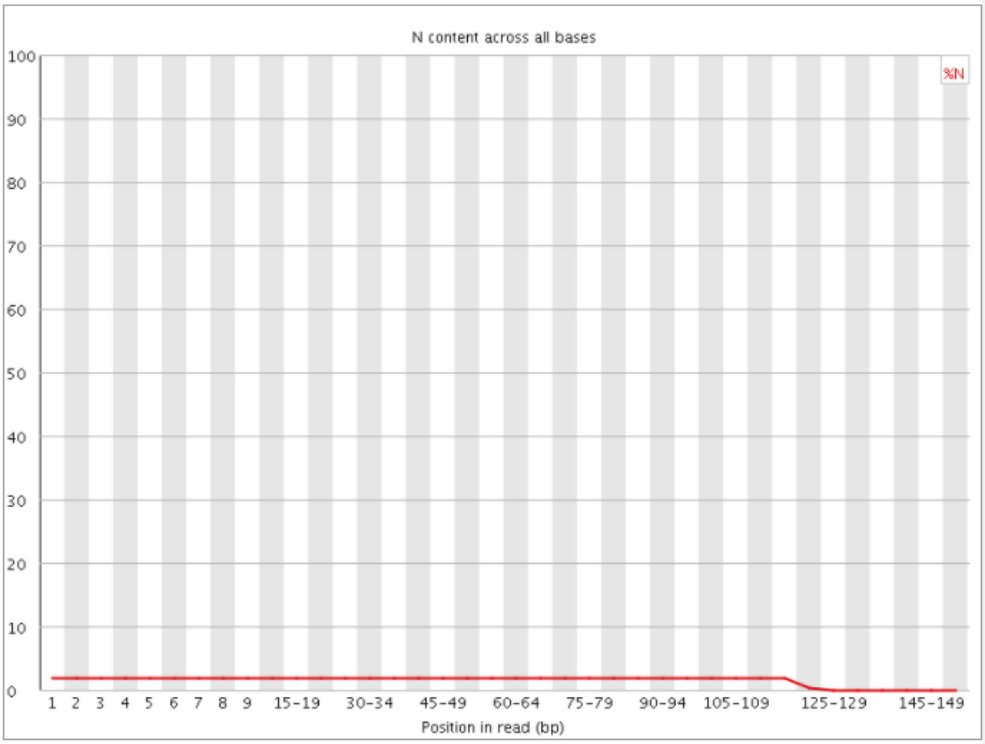
Per Base GC Content



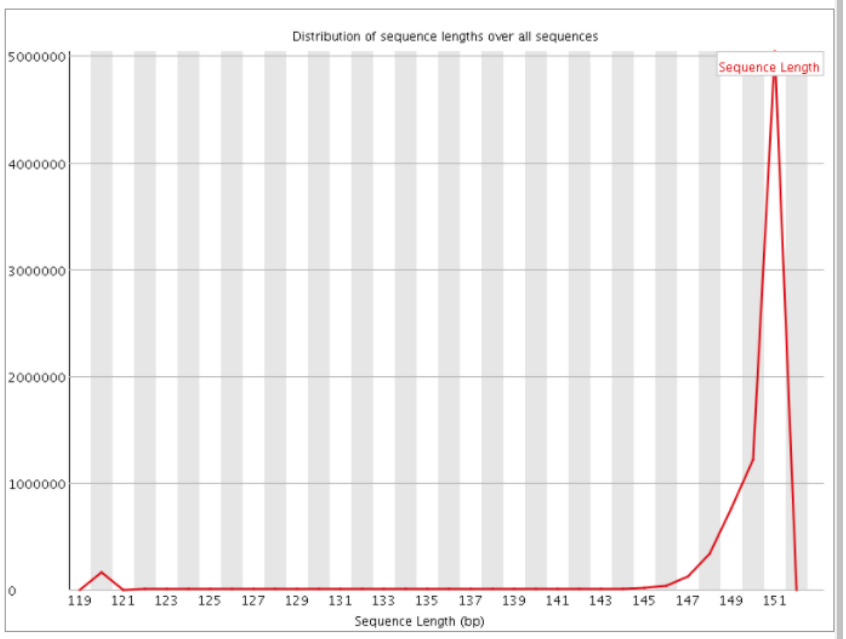
Per Sequence GC Content



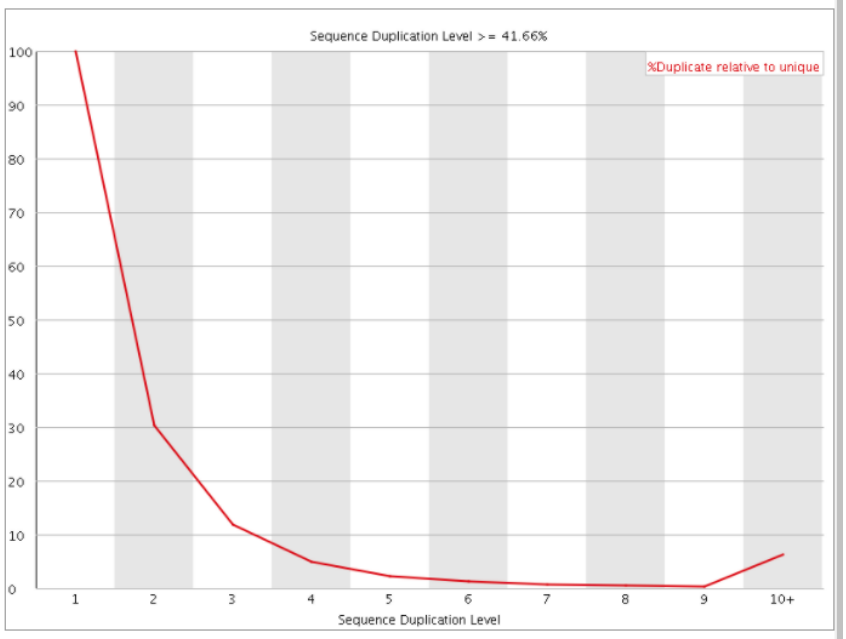
Per Base N Content



Sequence Length Distribution



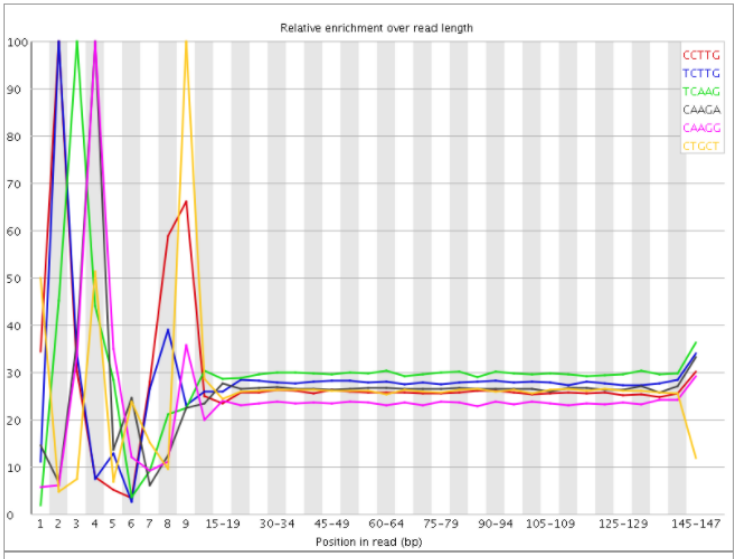
Sequence Duplication Levels



Overrepresented Sequences

Sequence	Count	Percentage
NNNN

Kmer Content



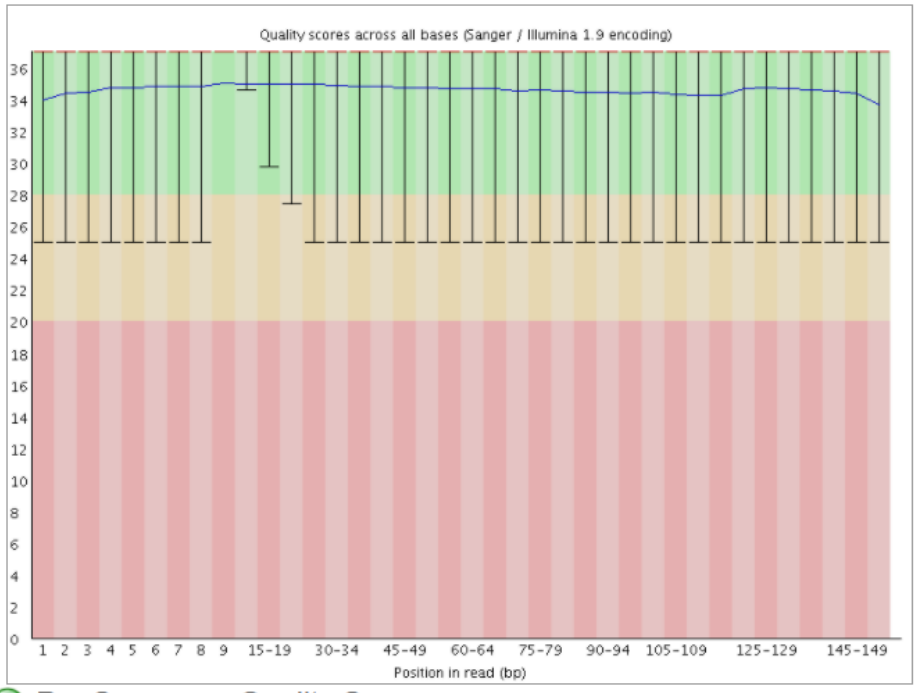
Sequence	Count	Obs/Exp Overall	Obs/Exp Max	Max Obs/Exp Position
CCTTG	1697085	1.5646665	5.907818	2
TCTTG	1774480	1.5567709	5.5590215	2
TCAAG	1775520	1.5391223	5.1344256	3
CAAGA	1773160	1.5278914	5.725337	4
CAAGG	1674255	1.5011288	6.291574	4
CTGCT	1538985	1.4189028	5.402341	9
GTCTT	1419470	1.2453167	5.7004333	1
TCCAG	1354200	1.241075	5.284529	2
ATCCT	1389230	1.2309504	5.1609793	6
CTGGA	1348640	1.2164533	5.27603	4
AAGAC	1381365	1.1902905	5.555944	5
ACAGC	1276305	1.1626972	5.7896514	8
CAGGA	1270005	1.1386803	5.0963187	4
GACAG	1244330	1.1156602	5.6746855	7
GTCCA	1133175	1.0385137	5.9441524	1
GACTG	1140565	1.0287727	5.5951886	7
ACACC	1088340	1.0073782	5.7781286	6
ATACA	1188620	0.9902363	5.160587	6
TGGAC	1093550	0.9863659	5.8335757	5
GTCCT	1067755	0.98444134	5.050607	1
ATACC	1093000	0.96268356	5.0711427	6
TATCC	1085880	0.96216214	5.076469	5
AGGAC	1070010	0.9593657	5.199456	5
GGATA	1112305	0.9489768	5.641695	1
ATACT	1083370	0.90797853	6.1093554	6
ACTGC	980090	0.8982167	5.04377	8
GTACA	1026320	0.88967294	5.7915974	1
TACTG	1003845	0.8754216	5.18167	7
GTATT	1050665	0.871867	5.244422	1
ACTGT	965530	0.8420082	5.17278	8
GTGCA	902780	0.8142942	5.115188	1
GTATA	963555	0.79480296	7.6075687	1
TATAC	941710	0.78925246	6.950825	5
ACACT	877815	0.7731546	5.051904	6
GGGTA	832810	0.73931503	5.0570054	1
TACAC	826975	0.7283762	6.9955854	5
GTGTA	836345	0.71782774	7.2432804	1

File 2: T1BOTES_S1_L001_R2_001.fastq.gz

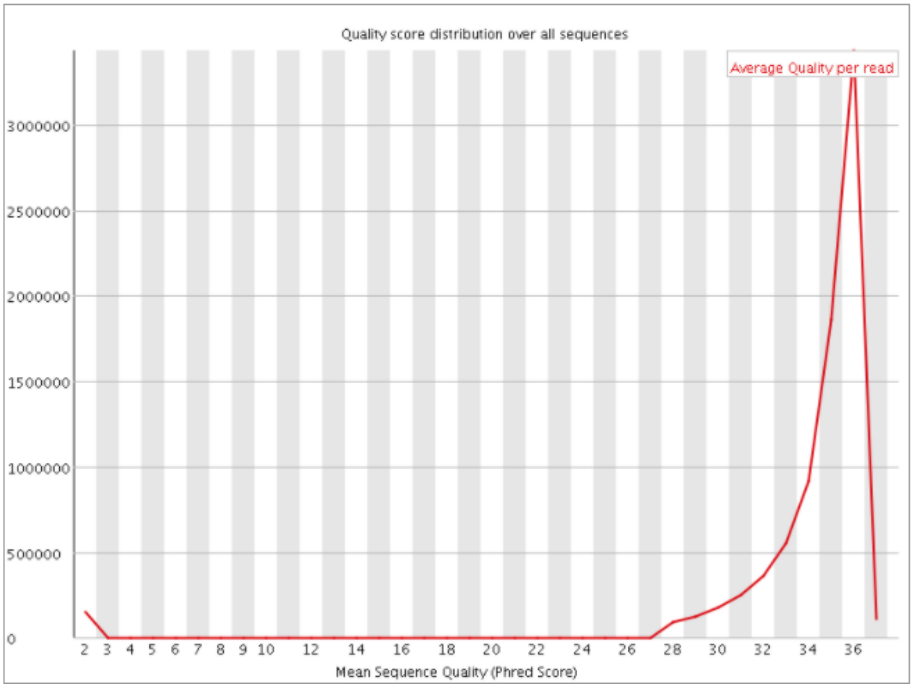
Basic Statistics

Measure	Value
Filename	T1BOTES_S1_L001_R2_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	8076080
Filtered Sequences	0
Sequence length	120-151
%GC	48

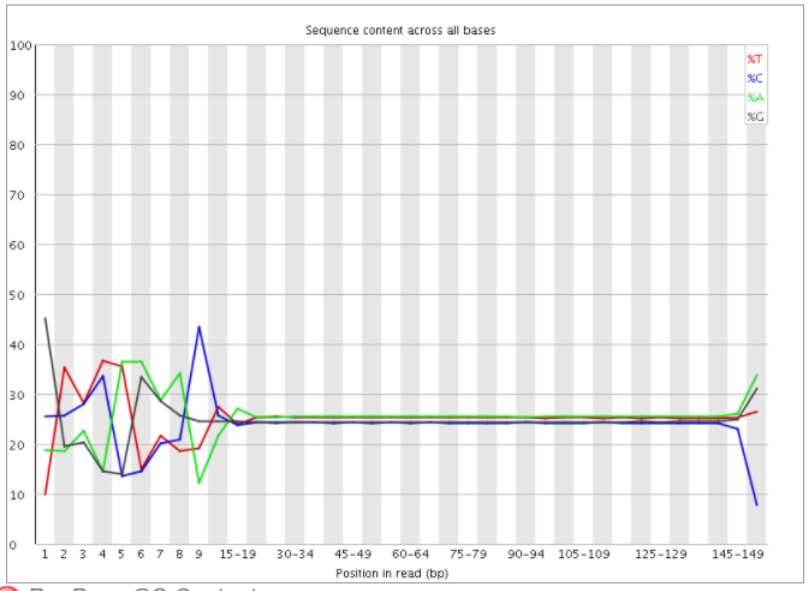
Per Base Sequence Quality



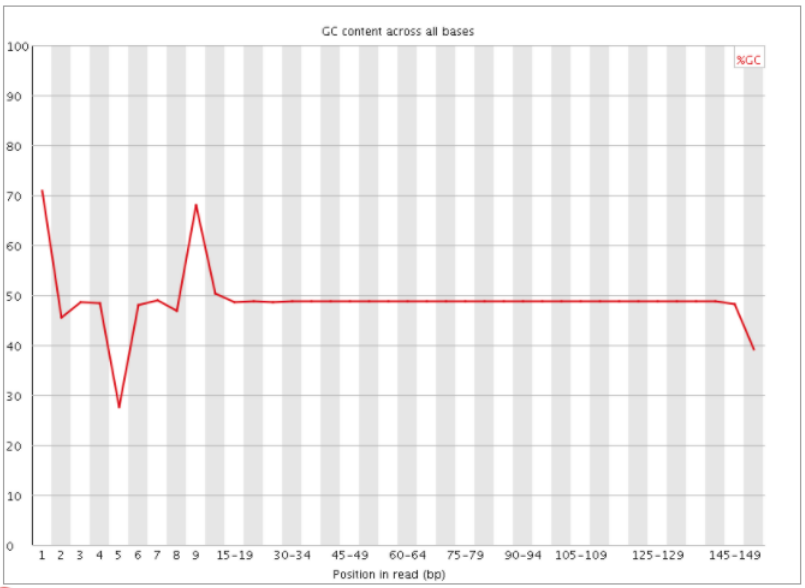
Per Sequence Quality Scores



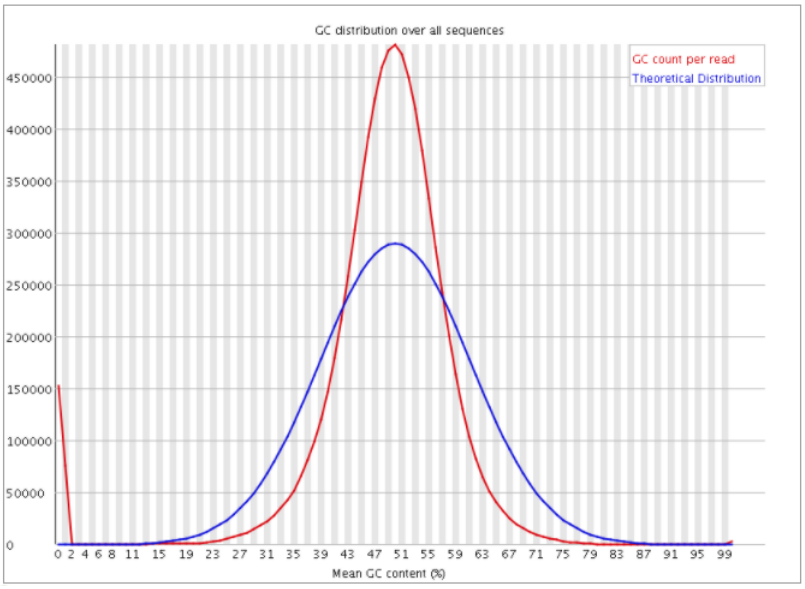
Per Base Sequence Content



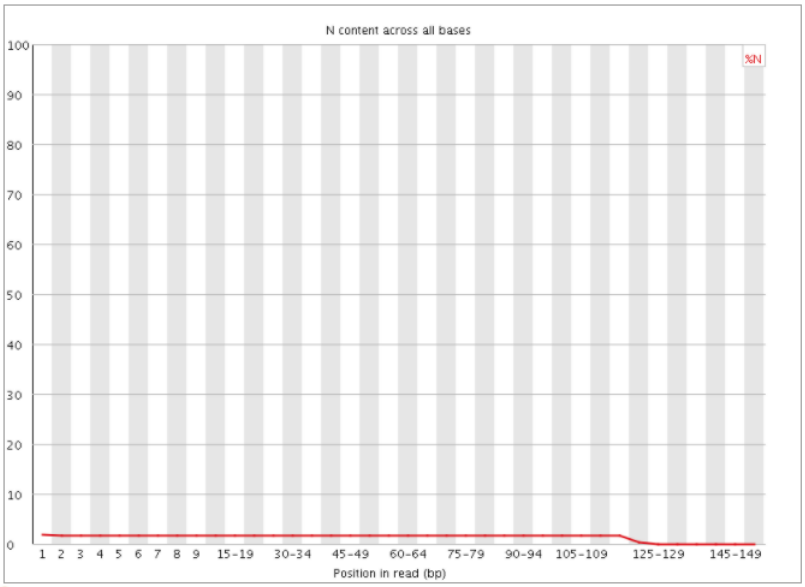
Per Base GC Content



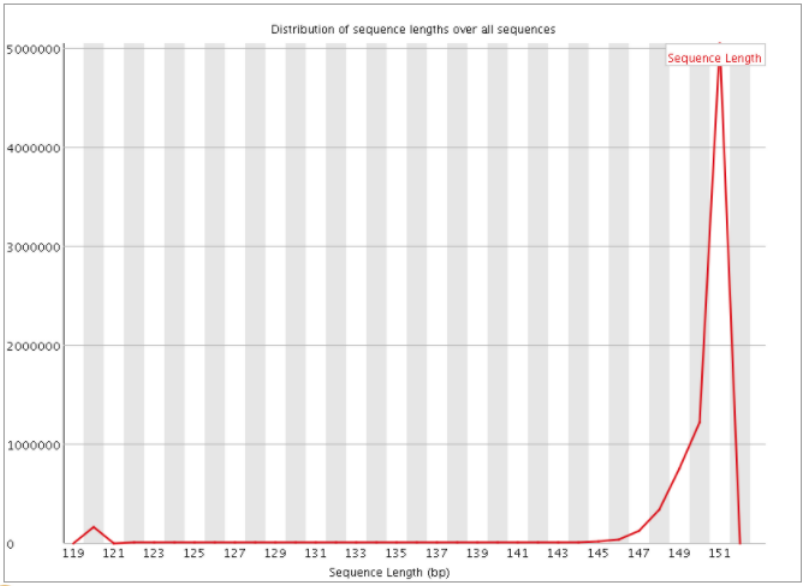
Per Sequence GC Content



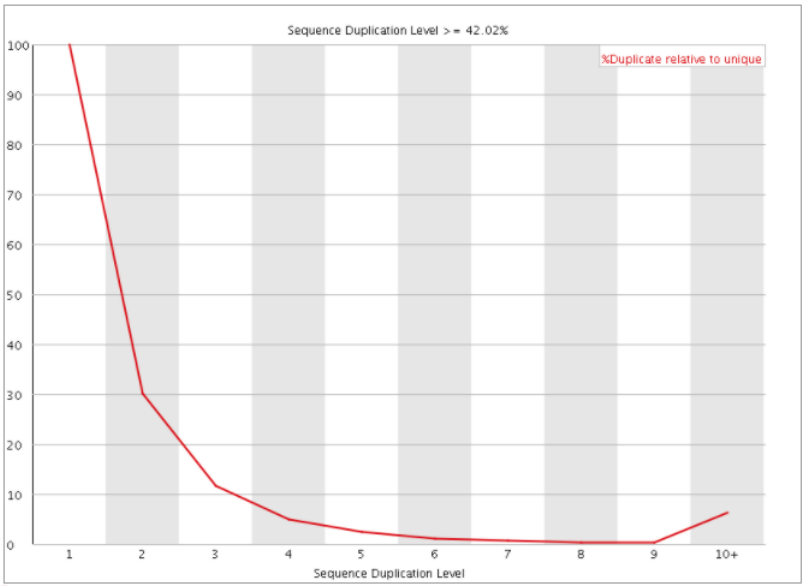
Per Base N Content



Sequence Length Distribution



Sequence Duplication Levels



Overrepresented Sequences

Sequence	Count	Percentage	Possible Source
NNNN