UNIVERSITY OF CALIFORNIA
RIVERSIDE

Micro-Colonial Fungi Across Different Microcosms: Observing Micro-Colonial Fungi in Their Niches

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Microbiology

by

Tania Kurbessoian

December 2022

Dissertation Committee:
Dr. Jason E. Stajich, Chairperson
Dr. Patrick Degnan
Dr. Katherine Borkovich

The Dissertation of Tania Kurbessoian is approved:

_____

_____

_____

Committee Chairperson

University of California Riverside

Acknowledgments:

I can only use my words to describe my sincerest thanks and gratitude to the individuals who made this experience possible. In no order of importance, I would like to thank the everyday folks who took the time to provide me with opportunities to better myself. You all helped instill a love of mycology, the confidence in myself, and the opportunity to develop bonds that will last a lifetime.

Lotus, your kindness, friendship, and guidance got me through the rough bits and I can't wait to see you blossom in academia.

To the Stajich lab members - Derrick, Nat, Jesus, Cassie, Kelsey, Julia, and Mark - thanks for all the laughs, and the snippets of code to get me finished with this degree. I've learned so much and am very enthusiastic about the future of mycology. I wish you all the very best on your journeys.

To my brother, Raffi, who sat with me and helped explain code logic after the numerous attempts and tears I expressed - thank you for your never-ending patience. Love you weirdo.

Haik, you and your family have been my anchor when I wanted to float away from all of this. Your unwavering support and love was the only way I was able to get through this. Thank you, I love you.

To my collaborators, Dr. Deborah Hogan, Daniel Murante, Dr. Nicole Pietrasiak, Dr. Laura Selbmann, and Dr. Claudia Coleine. I have learned so much and I hope to keep collaborating with you in the future.

To my awesome committee: Dr. Patrick Degnan, I believe you're definitely making your mark here on campus. I have always appreciated the dedication and time you put into all our meetings and classes. Your determination to have your students produce high-quality work is very much appreciated. Thank you! Dr. Katherine Borkovich, our fungal mama I am so grateful to have met you and worked with you in different capacities and spaces these past couple of years. I appreciate all you have done for our Microbiology department and all you continue to do for future students. You are a solid and welcome presence to all females interested in pursuing this field. Thank you!

Jason, how do I express my genuine appreciation for who you are as an individual and everything you do? It may have been unusual for a summer first-year student to start in a lab and never rotate, but I have no regrets. Your kindness, optimism, and passion for mycology will always be infectious and I learned so much while being under your guidance. Thank you for everything. I hope we can get some ice cream sometime soon.

Dedication

I dedicate this thesis to the person who means the most to me, my mother.
You are gone now, but my journey to this point in my life was all because of you. Thank you. I'll carry your words with me and make sure to instill them in my being as you did to me. In spite of your limited time here, you sacrificed everything to give me the most and best possible. I love you.

ABSTRACT OF THE DISSERTATION

Micro-Colonial Fungi Across Different Microcosms: Observing Micro-Colonial Fungi in Their Niches

by

Tania Kurbessoian

Doctor of Philosophy, Graduate Program in Microbiology
University of California, Riverside, December 2022
Dr. Jason E. Stajich, Chairperson

MCF are taxonomically classified in Ascomycota phylum of the Fungi and as members of the Eurotiomycetes and Dothideomycetes orders. These two lineages diverged 350 million years ago (Lutzoni et al. 2018), yet MCF from both these clades have developed very similar analogous features. The work described in this thesis encompasses a wide range of MCF species, uncovering their unusual niches, and understanding their genomic structure and evolution. The first chapter describes three new species of MCF isolated from biological soil crusts collected from the deserts of southern California. I have named two of these MCF species, *Coniosporium tulheliwenetii*, and *Taxawa tesnikishii* using Cahuilla language terms for nomenclature in consultation with the tribal community in reference to the traditional lands these were collected from. I also identified a new species *Neophaeococcomyces mojaviensis* collected from Sheephole Pass region of the Mojave Desert. The second chapter describes a population study of 25 strains from Antarctica of the endemic MCF, *Friedmanniomyces endolithicus*, collected over 40 years of research on endolithic fungal communities on the continent. Using whole genome sequencing and

population genomics I found variation in genome ploidy which may contribute to the success and survival of *F. endolithicus* in the extremes of Antarctica. *F. endolithicus* is uniquely endemic to Antarctica and I found that the estimated divergence of the species to other Dothideomycetes fungi is approximately 48 million years, which coincides to the predicted Antarctica split from Pangea as it occupies in the present South Pole. My final chapter studied twenty-three isolates of a slow-growing MCF, *Exophiala dermatitidis*, isolated from a persistent infection on the lungs of a cystic fibrosis patient over three years. Population structure showed three different groups with varying mutational rates, phenotypic filamentation, and chromosomal structures as perhaps consequences of the population persistence and even adaptation to the lung environment. My work has provided evidence that the diverse ecological niches that MCF occupy and their potential importance in environmental microbiology and clinical emphasize these slow growing MCFs are deserving of additional attention and study to understand their role as causative disease agents or pivotal foundational members of extreme habitats.

Table of Contents

List of Figures

Introduction

Chapter 1:

Chapter 2:

Chapter 3:

List of Tables

Introduction

Chapter 1:

Chapter 2:

Chapter 3:

Introduction

Fungi play critical roles in disease (Dohlman et al. 2022; Landry and Embers 2022; Nagpal et al. 2020; Guzzo et al. 2022), climate change (Martínez-García et al. 2017; Paterson and Lima 2017), food resources (Gmoser et al. 2020; Dupont et al. 2017) and biomaterials (Cerimi et al. 2019). It is an exciting time to be in fungal biology as sequencing technology only becomes more affordable (Preston et al. 2020 ; Hibbett et al. 2013) and with optimization for long-read technology allowing for global access use (De Coster et al. 2021). With the opportunity to sequence more fungal genomes, we can begin to ask deeper questions about fungal genome structure and to understand how these complex organisms can survive extreme habitats found on this planet. Black yeasts are one type of fungi that has been found in all kinds of extreme environments and will be the focus of my thesis.

Research on the biology and taxonomy of these enigmatic fungi have used multiple nomenclatures, including black yeasts (Akagi et al. 1958; Hoog 1979; Sandford et al. 1978), hyphomycetes (Cole and Kendrick 1969; Ellis 1971; Hoog 1979), meristematic fungi (Urzì et al. 2000; Sterflinger 1998; Sterflinger 2000; de Hoog et al. 1997; Sterflinger et al. 2005; de Hoog and Hermanides-Nijhof 1977), melanized fungi (Kapoor and Haider 1982; de Hoog 1997; Revankar and Sutton 2010; Dadachova et al. 2007), dark septate endophytes (Jumpponen and Trappe 1998; Jumpponen 2001; Mandyam and Jumpponen 2005), extremophilic fungi (Gunde-Cimerman et al. 2003; Ibrar et al. 2020; Gostincar et al. 2010), and micro-colonial fungi (Wollenzien et al. 1995; Wollenzien et al. 1997; Nai et al. 2013; Staley et al. 1982). It was until recently that the field lacked a single naming convention for the group, but this dissertation will use the term micro-colonial fungi (MCF)

as the general description following the lead of experts in this field to describe this group most accurately (personal communication, Anna Gorbushina). Many MCFs have been isolated from endolithic features, and are thus also described as rock-inhabiting fungi (RIF) (Sterflinger et al. 1997; Gueidan et al. 2011; Nai et al. 2013; Egidi et al. 2014) or lithobiontic/lithophilic fungi (Heinen and Lauwers 1986; Caretta et al.) and even as endolithic fungi or cryptoendolithic fungi (Palmer and Friedmann 1988; Selbmann et al. 2005; Onofri et al. 2015; Coleine et al. 2018). Decades before, the suggestion of lumping the odd isolates into dematiaceous fungi (Carrion 1950; Ellis 1971; Ellis 1976) was very common, and many recent papers still hint at this old manner of fungi nomenclature placement (Afsarian et al. 2012).

MCFs are found within the Ascomycota phylum and are paraphyletically distributed in two distinct classes: Dothideomycetes and Eurotiomycetes (**Figure 1**) (Nai et al. 2013; Gueidan et al. 2011; Teixeira et al. 2017; Selbmann et al. 2020). The specific MCF-containing orders include Chaetothyriales (Eurotiomycetes), Pleosporales, Dothideales, and Capnodiales (Dothideomycetes). It has been postulated that these similar phenotypic features arose convergently millions of years ago (Gueidan et al. 2011). This hypothesis indicates that the ancestor of other orders within these two classes did not have MCF features. Meaning, millions of years ago, the selective pressure to adapt to abiotic stressors by the ancestors that four orders drove a common evolutionary outcome to these distinct phenotypes.

Figure 1. Phylogenetic Tree of Chaetothyriales and Dothideomycetes. A maximum likelihood phylogenetic tree constructed of species from both classes**,** Dothideomycetes**,** and Eurotiomycetes which both contain MCF and non**-**MCF organisms**.**

Phenotypes in MCF

MCFs have three distinct phenotypic features used to identify them. These features include melanized cell walls (whether produced by 1,8-DHN melanin, L-DOPA, the tyrosinase pathway, or any combination of the three) (Fogarty and Tobin 1996; Dixon et al. 1991), mycosporine-like amino acids (Bhatia et al. 2011), and meristematic growth (which gave MCFs one of their earlier names) (de Hoog and Hermanides-Nijhof 1977). Melanin is a secondary metabolite produced by MCFs. The collection of 'albino' strains in the lab and melanized strains of the same fungus suggests that the production of melanin is not essential for the fungus to survive (Bell and Wheeler 1986). Instead, melanin provides the fungus with certain advantages in pivotal circumstances and environments (Butler and Day 1998). Different varieties of melanin are produced in many organisms, including, eumelanin, pheomelanin, neuromelanin, pyomelanin, and allomelanins (Ambrico 2016; Schmaler-Ripcke et al. 2009). MCFs primarily produce eumelanin and pheomelanin, while other non-MCF fungi can produce pyomelanin and allomelanin (Riley 1997). The characterized functional roles of melanin include thermoregulation, photoprotection, metal binding, mechanical projection, energy harvesting, cell development, antioxidant, anti-desiccant, and chemical protection (Cordero and Casadevall 2017; Dadachova and Casadevall 2008; Robinson 2001).

There are three main pathways MCFs are known to use for melanin production in MCFs. 1,8-Dihydroxynapthalene (DHN) melanin pathway is the most common among MCFs. Generally, reactive agents acetyl coenzyme-A or malonyl coenzyme-A begins the synthesis of melanin using the pentaketide pathway, and through a series of enzymatic reactions with

polyketide synthases, reductases, laccases, tyrosinases, peroxidases, and catalases produce the final melanin pigment (Bell and Wheeler 1986; Butler and Day 1998; Langfelder et al. 2003; Taborda et al. 2008). L-3,4-dihydroxyphenylalanine (L-DOPA) produces melanin that resembles mammalian melanin (eumelanin), if sulfur is added to the mix it produces the alternative pheomelanin (Ito and Wakamatsu 2003). Generally, reactive agent L-dopa is oxidized to dopaquinone with laccase and tyrosinase enzymes. Dopaquinone is a highly reactive chemical and is quickly and spontaneously converted to cyclodopa and oxidized to dopachrome. The dopachrome is then tautomerized to dihydroxyindoles which polymerize to form melanin (Land et al. 2004; Langfelder et al. 2003; Riley 1997; Taborda et al. 2008). The third melanin pathway uses tyrosine along with homogentistc acid and *p*-hydroxyphenylpyruvate (PHPP) to produce pyomelanin (Coon et al. 1994; Riley 1997; Frases et al. 2007). Other melanin precursors seen in other fungi and organisms include gamma-glutaminyl-4-hydroxybenzene (GHB) (Solano 2014), catechol (Toledo et al. 2017), catecholamines (Liu et al. 1999), and scytalone (Bell et al. 1976; Kubo et al. 1983). It is important to note that in MCFs, the genes in all three pathways for melanin production are not found in gene clusters in close proximity to each other (Teixeira et al. 2017). The pigment melanin is still shrouded in mystery as the crystal structure has yet to be defined (Eisenman and Casadevall 2012), making it difficult for us to understand the properties and how fungi use them effectively.

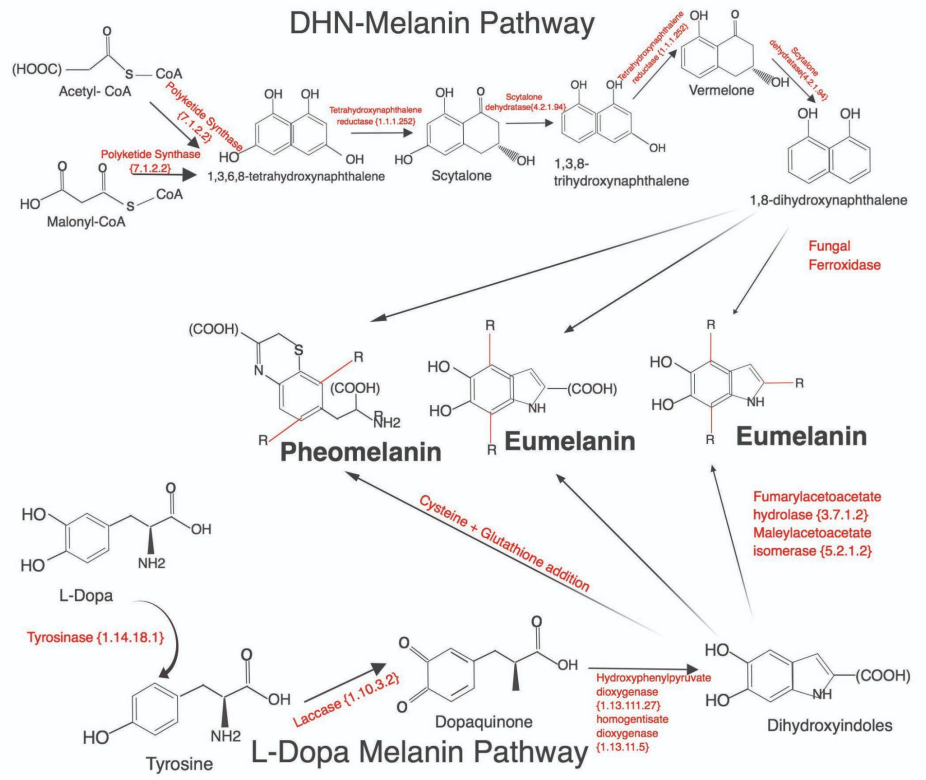


Figure 2. Three melanin synthesis pathways in Ascomycota fungi. The DHN melanin pathway is the most commonly found in MCFs**,** while L-DOPA and Tyrosinase pathways are less observed**.**

Mycosporine-like amino acids (MAAs) are photoprotective pigments produced by many different organisms like marine algae and fungi. Mycosporines are used to describe fungal metabolites with an absorption spectrum of 310-320 nm produced through the shikimic acid pathway (Bhatia et al. 2011). While MAAs are conjugated to the nitrogen base in amino acids increasing its absorption spectrum of 310-360 nm and were first isolated from sporulating mycelia (Leach 1965). These metabolites are highly photostable in a variety of high salinity environments, while also being resistant to abiotic stresses like temperature, UV, and pH (Conde et al. 2000). Additionally, other functions have been hypothesized including, scavenging reactive oxygen species, and protection against hypersaline

environments (Gorbushina et al. 2003; Kogej et al. 2006; Oren and Gunde-Cimerman 2007). Due to their sunscreen and UV protection capabilities, MAAs are considered for industrial applications.

Meristematic growth was first described by de Hoog in 1977 (de Hoog and Hermanides-Nijhof 1977) for fungi that form aggregates of thick-layered, melanized cells which reproduce through isodiametric division. Propagules may disperse through breaking apart or by endogenous conidiogenesis. Some may form blastic conidia from undifferentiated cells and some may even form budding cells (Sterflinger 2006). These variable phenotypic differences are the crux of why we have so many nomenclatures used to describe them (Liu et al. 2022). The observation of variation in cell division phenotypes is a principal reason for so much confusion in this field. As such, closer examination of these processes in different MCFs was required. Mitchison-Field et al. isolated MCFs from ocean water near Cape Cod, including *Hortaea werneckii*, *Knufia petricola*, *Aeurobasidium pullulans*, and *Phaeotheca salicorniae* (Mitchison-Field et al. 2019). These fungi were imaged through differential interference contrast (DIC) microscopy to analyze their growth and division. The MCFs show remarkable plasticity with their cell polarity, modes of cell division, isodiametric and unilateral swelling, and cell size control (Selbmann et al. 2005; Mitchison-Field et al. 2019). Certain MCF divide through a combination of fission and budding (*H. werneckii*), sequential orthogonal septations through meristematic filamentation and conidial chains (*P. salicorniae*), and the production of multiple simultaneous buds (*A. pullulans*). Classic model yeast systems have historically seen budding yeasts and fission yeasts, and this work made an important expansion of

understanding in fungal biology. Fungal biologists hypothesize that these MCF phenotypes may have arisen through convergent evolution, specifically homoplasy, or pleiotropic latent homologies. The latter suggests that similar traits in distant lineages can be generated repeatedly if they can be readily recruited for the same new functionality (Nagy et al. 2018). While homology would suggest similarities due to shared ancestry in related species, the fungal fossil record is not reliable for this theory (Nagy et al. 2018). As *H. werneckii* shows a variation of fission and budding (Zalar et al. 2019), depending on environmental conditions, it raises the possibility of morphological bet-hedging which allows the organism to find what is optimal for the next environmental fluctuation (El Baidouri et al. 2021).

Mold is old

MCFs were probably around as long as agriculture has been a stable feature in human anthropology (Sidney 1799). "Black sooty molds" are commonly found on many fruits, trees, and vegetables for millennia (Dark and Gent 2001). This filamentous mold lands on plants due to "sucking insect" honeydew, a substance excreted as a result of indulging in plant sap (Nelson 2008). As the sooty mold lands and digests the honeydew, it is generally harmless to the plant. Though, if there is an increased volume of soot, the plant may not have access to enough sunlight, thus causing the plant to die or prematurely drop (Nelson 2008). Sooty molds are a collective term used to describe a group of Ascomycete fungi that fall into particular genera including *Cladosporium, Alternaria, Aureobasidium, Antennariella, Limacinula, Scorias, Meliola,* and *Capnodium* (Caplan 1995; Reynolds 1971; Harshberger 1902; Ward 1883; McAlpine and Maiden 1896; Kwon and Park 2003;

Xu et al. 2020; Poursafar et al. 2018; Mirzwa-Mróz et al. 2014). Not all sooty molds are MCFs but all sooty molds fall within the class Dothideomycetes and are critical agricultural concerns when considering the volume of plants destroyed each year from their effects(Wood et al. 1988; Santos et al. 2013).

MCFs in diseases

An example of MCF interaction with humans includes the treatable disease Tinea nigra. Tinea nigra is caused by the MCF *Hortaea werneckii*, introduced through exposure to fungal cells, and can remain on the skin for years without issue to its host (de Hoog and Gerrits van den Ende 1992). They can appear as a brown pigmented patch on the palms or base of feet and sometimes around the neck. As the fungal infection is only superficial, generic antifungal cream can remove the infection. Other more pathogenic, opportunistic, and dangerous cases of MCFs can cause diseases in humans and other wildlife. More than 100 species of MCF can cause phaeohyphomycosis and chromoblastomycosis. MCFs that cause disease include, *Exophiala sp.*, *Alternaria sp.*, *Rhinocladiella mackenzie*, *Cladophialophora sp.*, *Phialophora sp.*, *Fonsecaea sp.*. Cases of these diseases have been identified in a range of animal hosts from mollusks to leopards and humans (Seilern-Moy et al. 2019; Mayer et al. 2022; Seyedmousavi et al. 2013; Manharth et al. 2005; Peel et al. 2021; Lloret et al. 2013; Janovsky et al. 2006; Salgado 2010; Queiroz-Telles et al. 2017). Other well-known disease states of MCFs include opportunistic infections of individuals with cystic fibrosis (CF). Fungal species involved with CF secondary infections include *Scedosporium apiospermum*, *Aspergillus fumigatus*, *Candida albicans* and *Clavispora (Candida) lusitaniae*, and *Exophiala dermatitidis* (Cimon et al. 2000; Bakare et al. 2003;

Ross et al. 2021; Chotirmall et al. 2010; Demers et al. 2018; Kurbessoian et al. 2022). Chapter 3 of this thesis examines a cystic fibrosis patient with persistent MCF *Exophiala dermatitidis* infection. Phenotypic and population genomics of isolates sampled from a 3-year time period to understand within-host evolution.

As there has been an increase in human phaeohyphomycosis and chromoblastomycosis cases in the recent century (Revankar et al. 2002; Revankar and Sutton 2010; Matsumoto et al. 1994), scientists from Slovenia have examined man-made extreme environments for possible sources of MCF driving human disease. Environments like dish-washers are interfaces with high heat and moisture contained by rubber seals harboring MCFs like *Exophiala sp.* (Zupančič et al. 2016; Babič et al. 2018; Kulesza et al. 2021; Zalar et al. 2011; Prenafeta-Boldú et al. 2022). Many other instances of MCF harboring man-made environments have also been noted, including humidifiers (Babič et al. 2018; de Hoog et al. 1997; Nishimura and Miyaji 1982), washing machine soap dispensers (Isola et al. 2013), car gasoline tanks (Isola et al. 2013), and sauna facilities (Matos et al. 2002). But to connect the MCF to man-made environments, one must understand the source of these organisms from the natural world. It has been postulated that *Exophiala dermatitidis*, an opportunistic human pathogen, has broad natural reservoirs including wasp nests (Conti-Díaz et al. 1977), healthy bats (Reiss and Mok 1979), lesions on toads (Frank et al. 1970), and rotten wood (Dixon et al. 1980). This suggests human patients receive the infection from environmental exposure (Sudhadham et al. 2011).

MCFs the extreme cosmopolitan fungus

MCFs have been consistently isolated from cosmopolitan locations, like the sooty molds described above, but can also be found in the most extreme environments. The MCF-causing tinea nigra, *Hortaea werneckii* (Dothideomycetes), is mainly collected from warm hypersaline pools of water (Gunde-Cimerman and Plemenitaš 2007; Gunde-Cimerman et al. 2000). It is recommended for individuals to avoid these salterns as they could easily infect the individual with a superficial black yeast infection. The aforementioned rock-inhabiting fungi (RIF) are another example of MCFs found in extreme environments (Liu et al. 2022). RIFs have been isolated globally, from China, the Mediterranean to even Antarctica (Sun et al. 2020; Ruibal et al. 2009; Gonçalves et al. 2019; Onofri et al. 2014; Ruibal et al. 2018; Selbmann et al. 2015; Gromov 1957). RIFs contain slow-growing Ascomycota MCFs and lichen-forming fungi (Eurotiomycetes, Dothideomycetes and Arthoniomycetes) that can be found on the surface of rock features and have been found under a couple of millimeters in more porous rocks (Gueidan et al. 2008; Ruibal et al. 2009). It is important to note that RIF does not necessarily use the rock as a biological or chemical resource but rather as a place of dwelling and propagation (Gorbushina and Krumbein 2000). The secreted enzymes RIFs release into the environment consequently erodes the rocks they inhabit.

A genus of Dothideomycete RIF isolated from the Antarctic were named *Friedmanniomyces*. Currently, only two species have been identified, *F. endolithicus* and *F. simplex* (Onofri et al. 1999; Coleine et al. 2020; Selbmann et al. 2005). *F. endolithicus* is found exclusively and frequently isolated from the endolithic communities found in

Antarctic hyper-arid deserts of Victoria land of McMurdo Dry Valleys (Selbmann et al. 2015). *F. endolithicus* has been tested to resist 400Gy units of radiation (which is 200Gy higher than normal Earth level radiation) and puzzlingly shows improvement in metabolic activity under such conditions (Pacelli et al. 2018). Chapter 2 in this thesis takes 25 isolates of *F. endolithicus* collected from Antarctica by Laura Selbmann's group at University of Tuscia (Viterbo, Italy) and sequenced in collaboration with Claudia Coleine. These isolates are cultured from endolithic samples collected from multiple rocks in Antarctica. Using whole genome sequencing and population genomics I found variation in genome ploidy which may contribute to the success and survival of *F. endolithicus* in the extremes of Antarctica. *F. endolithicus* is uniquely endemic to Antarctica and I found that the estimated divergence of the species to other Dothideomycetes fungi is approximately 48 million years, which coincides to the predicted Antarctica split from Pangea as it occupies in the present South Pole.

On the other side of the globe, MCFs have been collected in the Arctic in Greenland through the efforts of Nina Gunde-Cimerman and her team. Ice is produced in extremely low temperatures, but this also reduces the water activity in the area thus creating another pocket of an extreme environment microorganisms would need to adapt to survive. It has been noticed that with time the Greenland ice sheets (GRIS) have begun to darken. This darkening has been connected to a rise in algal blooms, warming the ice to produce pockets where microorganisms can find refuge from the total extremes and continue to darken the sheets (Perini et al. 2019). Through collection efforts, many isolates of MCFs have been collected from here along with ascomycetous and basidiomycetous yeasts (Gunde-

Cimerman et al. 2003; Perini et al. 2019) including *Cladosporium sp.*, *Oleoguttula mirabilis*, *Aspergillus sp.*, and *Rhodotorula "svalbardensis"*.

MCFs in more complex communities

MCFs are found associated with soils and even in complex relationships, like the one seen between fungus-growing ant species like *Acromyrmex echinatior*, the fungi cultivated by ants *Leucoagaricus gongylphorus*, and fungal pathogens like *Escovopsis weberi* (Schoenian et al. 2011; Batey et al. 2020). Multiple attempts at culturing indicate a variety of genera, including *Phialophora sp.*, *Cladiphialophora sp.*, *Cladosporium sp.*, *Exophiala sp*, *Ochroconis sp., Phaeococcomyces sp.*, and *Penidiella sp.*, suggesting the MCF are not species-specific, highly adaptable and play a major role in this relationship (Little and Currie 2007; Little and Currie 2008).

Biological soil crusts (BSCs) are intriguing and complex consortia found globally in a variety of extremes. BSCs cover up to 70% of dryland ecosystems (Beraldi-Campesi and Retallack 2016) which constitute some of the largest terrestrial biomes, collectively covering 47.2% of Earth's land surface (Lal 2004). BSCs have been found in dry hot deserts (Pombubpa et al. 2020; Warren et al. 2019; Pietrasiak and Johansen 2014; Belnap 2003; Rosentreter and Belnap 2001; Ullmann and Büdel 2001), and even the tundras of British Colombia, Canada (Carr 2022; Carr et al. 2021) and Antarctica (Green and Broady 2001; Jung et al. 2018). BSCs are composed of cyanobacteria, algae, diazotrophic bacteria, lichen, fungi, and bryophytes that colonize and stabilize the soil surfaces against natural forces like wind and water (Warren et al. 2019). BSC's main roles include: 1) mediating

almost all inputs and outputs (gases, nutrients, and water) to and from the strata above and below the surface, and 2) being the zone of high nutrient deposition, transformation, and availability. 3) structuring temporal, spatial, and compositional aspects of the vascular plant community, and 4) facilitating the direct delivery of carbon, nutrients, and water by BSCs from the soil interspace to nearby vascular plants (Belnap 2003). Multiple instances of MCFs have been isolated from BSCs (Carr 2022; Carr et al. 2021), which poses the question of what are their functional roles in these consortia. And are they critical core groups seeing there are many instances of MCFs found within them? Chapter 1 takes two biological soil crust samples from Southern California, and brings into light three MCFs species isolated, sequenced, assembled, annotated, and described as three new species.

Dark septate endophytes (DSEs) are root-inhabiting fungi and are another interesting naming schema, as there are fungi in this group that do not fall within the two MCF orders (Dothideomycetes and Chaetothyriales) but still melanize in this endophytic situation. DSEs can vary as negative (parasitic), neutral (commensal), or even positive (mutualism) symbiotic relationships depending on the host and the abiotic/biotic stressors and function similarly to mycorrhizae. Unlike mycorrhizae, DSE cannot form a complete mantle or Hartig net when colonizing a host and instead invade microsclerotial structures (O'Dell et al. 1993). It seems the term DSEs is liberally used for any melanized septated fungi that associate with plant roots intercellularly or intracellularly (Jumpponen 2001; Jumpponen and Trappe 1998; Mandyam and Jumpponen 2005). Functionally, scientists have been able to determine positive DSE roles through oxidative and water stress (Santos et al. 2017;

Santos et al. 2021). It is interesting to see MCF involvement in a variety of plant-associated systems.

Marine MCFs are involved in carbon incorporation and cycles in the ocean. A study used an MCF *Cladosporium sp.* and DNA isotyping (DNA-SIP) with labeled 13C polysaccharides showing direct assimilation of carbon from phytoplankton (Cunliffe et al. 2017). *Cladosporium* was also detected to release extracellular glucan 1,3-Beta-glucosidase which helps to further digest phytoplankton (Cunliffe et al. 2017). A positive correlation was also detected between the abundance of diatom species and the *Cladosporium* species. On numerous occasions, MCFs have been isolated from the ocean water(Mitchison-Field et al. 2019; Ettinger and Eisen 2021; Ettinger and Eisen 2019). MCF species include *Hortaea werneckii*, *Knufia petricola*, *Aeurobasidium pullulans*, *Phaeotheca salicorniae, Cladosporium sp.,* and *Ramularia sp*..

There are MCFs found co-existing with lichen species and are termed lichenicolous fungi (Muggia and Grube 2018; Chang et al. 2022; Harutyunyan et al. 2008; Selbmann et al. 2014; Selbmann et al. 2013). There is a third group, the Arthoniomycetes which is a sister order to Dothideomycetes that contains fungi that are heavily associated with lichen species. Functionally, the roles of MCFs are still to be determined but a general idea on UV screening and protection of photobionts with excreted extra-cellular melanin (Harutyunyan et al. 2008) and hydration through the production of lipopolysaccharides (LPS). Lichenized features have also been developed with the co-culturing of MCFs with particular photobionts (Gorbushina et al. 2005; Brunauer et al. 2007). MCFs in this

complex association are benefiting the assemblage with the production of melanin and LPS, while in return they receive bio-active carbon from the photobionts in the system (Selbmann et al. 2013) and a slew of stress-protective compounds (Boustie et al. 2011). Photobionts also produce a variety of xanthophylls and carotenoids. While the lichen symbionts produce dibenzofurans (usnic acid) depsides (atranorin), depsidones (lobaric acid), and shikimate derivatives (calycin) (Boustie et al. 2011). The Antarctic lichen, *Ramalina terebrata* produces Ramalin, a novel compound with tested antioxidant function (Paudel et al. 2011). Many lichen-produced compounds are released as extrolites and accumulate as crystals (Boustie et al. 2011).

Recalcitrant pollutants, like hydrocarbon-contaminated soils, are another resource where a high concentration of MCFs has been detected (Yazdanparast et al. 2017; Seyedmousavi et al. 2011; De Hoog et al. 2006; Isola et al. 2021; Mohammadian et al. 2017; Isola et al. 2013). Oil spills have substantial detrimental effects on marine and terrestrial wildlife, and finding a means to effectively remove the contaminants has been a concern for some time (Bik et al. 2012). On soil, hydrocarbons spread horizontally, partition into the groundwater, and intercalate into soil particles making it impossible to separate through natural means from ingestion to inhaling vapor from the contaminants (Plohl et al. 2002). There are public concerns that hydrocarbon-contaminated water along with soils as they can be toxic and carcinogenic (Field et al. 1992; Clemente et al. 2001; Cerniglia and Sutherland 2001). Microbial bioremediation has been mostly explored with bacteria (Bloomfield and Doolin 2017; Cerniglia and Sutherland 2001; Sarkar et al. 2020), but efforts to utilize fungi in remediation practices has expanded (Akhtar and Mannan 2020; Singh 2006). There have

been assessments of the bacteria associated with hydrocarbon contamination, but insufficient work has characterized the fungal species and their biochemical pathways to break down these complex phenolic organic molecules. A comparison of water samples collected before and after a large oil spill indicates an immense shift from nematode dominant to the fungi-dominant eukaryotic presence (Brannock et al. 2017; Bik et al. 2012). This suggests while some eukaryotes are more sensitive to hydrocarbon contamination, some fungi taxa could tolerate and metabolize the hydrocarbons (Brannock et al. 2017; Bik et al. 2012). The fungal taxa that dominate these hydrocarbon communities include *Cladosporium sp.* and *Alternaria sp.* While these species are highly abundant in the post-spill samples, they were rare taxa in pre-spill fungal assemblages and increased dramatically post-spill. *Cladosporium sp.*, an MCF, has been tested to use hydrocarbon compounds and thrive in polluted environments seemingly intolerable for other fungi (Cofone et al. 1973; Okpokwasili and Okorie 1988). While *Alternaria sp.* demonstrates lignocellulose-degrading enzymes which have been implicated in the breakdown of hydrocarbons (Mtui and Nakamura 2008; Kiiskinen et al. 2004; Atalla et al. 2010). It is interesting to think about MCFs that are in these extreme environments but still surrounded by other fungal species. As we spend more time understanding the function of MCFs it could bring to attention how vital they could be in these special niches.

MCFs have also been isolated from important historical art features, geological sites, and marble sculptures (Wollenzien et al. 1997; Sterflinger et al. 1997; Sterflinger and Krumbein 1997; Mohammadi and Krumbein 2008; Cappitelli et al. 2007; Trovão et al. 2021) and with the production of patina (Krumbein 2001; Gorbushina et al. 1993; Dakal

and Cameotra 2012). *Knufia petricola* (syn. *Sarcinomyces petricola*), an MCF isolated from Mediterranean marble features (Wollenzien et al. 1997; Marvasi et al. 2012), is non-pathogenic and belongs to the Chaetothyriales order (Tsuneda et al. 2011). It was also recently the first MCF to be genetically modified to use the CRISPR/*Cas9*-mediated gene editing and replacement, and gene silencing using RNAi to allow for proper genetic modification and assessment (Voigt et al. 2020). This is an important start to further enable the genetics and molecular biology of a non-pathogenic MCF. Another important historical art feature to consider is painting degradation caused by microbial interactions. Biodegradation or biodeterioration is defined as the irreversible loss of a feature of art destroyed by biological agents (Krumbein et al. 1991). Fungi are among the most harmful etiological agents associated with biodegradation (Urzì and De Leo 2000; Krumbein et al. 1991). MCFs are included in this as their traits (meristematic growth and melanization) enable them to adapt to high-stress environments and sculptures and paintings are sometimes exposed to a lot of stress (Pangallo et al. 2015). In Pangallo et all, MCFs collected from epoxy resin included *Pseudotaeniolina sp.*, *Phaeococcomyces sp.*, and *Catenulostroma sp.*. Important cultural heritage has been maintained on cave and limestone collections, especially in Portugal, which has been tested to have *Hortaea sp.*, *Sarcinomyces sp.*, *Coniosporium sp.*, *Capnobotryella sp.*, *Exophiala sp.*, *Knufia sp.*, and *Trimmatostroma sp.*. Even larger historical features like Egyptian pyramids are home to MCFs like *Hortaea sp.*, and *Pseudotaeniolina sp*. (Rizk et al. 2021; Rizk and Magdy 2022)

MCF genomics

The first MCF to be sequenced was *Exophiala dermatitidis*, through the Broad Institute (Chen et al. 2014). This strain also had proteomics (Tesei et al. 2015; Schultzhaus et al. 2020) and transcriptomics (Poyntner et al. 2016; Poyntner et al. 2018; Malo et al. 2021) completed and was a pivotal genome to help build later genomes with. The MCF that probably has had the most extensive comparative and population genomics assessment is probably the hyper-saline living *Hortaea werneckii,* as GenBank has over 70+ strains available that have been collected from a variety of different extremes (de Cock 1994; Romeo et al. 2020). Those associated with seawater are more stress tolerant than medical strains collected from non-water-related environments (Marchetta et al. 2018). As MCFs have been collectively described as haploid, *H. werneckii* was also considered so. But it wasn't until Sinha and Gostincar (Sinha et al. 2017), concluded an endoreduplication but later it was found that the diploid variants were produced through two haploid hybridizations that occurred between two divergent strains (Gostinčar et al. 2018; Gostinčar et al. 2021). Genome duplication and high heterozygosity (~4.6%) were observed when a comparative phylogenomic analysis was performed (Romeo et al. 2020). It is still unclear how MCF diploids are found but are not produced through recombination. Generally, the size of MCF genomes range from 25.8Mb from *Capronia sp.,* to 43Mb of *Cladophialophora sp.* with generally high GC content for fungi (49-54%) (Moreno et al. 2018). MCFs also tend to have larger genomes with expanded and duplicated protein families that are associated with nutrient acquisition and protection against stressors (Teixeira et al. 2017). These protein families include CYP P450s, aldehyde

dehydrogenases, sugar transporters, and membrane transporter proteins (Grigoriev et al. 2014). Expansion in these proteins would allow the MCF a higher chance of adaptability and survival (Moreno et al. 2018). With duplicated gene families, there is an increased chance of different selective pressures to act upon them, thus enhancing protection from stressful environments. It is important to understand the location of where duplicated genes occur for MCFs. When comparing to other fungi like *Aspergillus fumigatus* (Fedorova et al. 2008), *Candida albicans* (De Las Peñas et al. 2003), and *Fusarium oxysporum* (Ma et al. 2010), paralogous genes or duplicated genes appear in telomeric regions or mobile pathogenicity chromosomes. There remain many gaps in knowledge concerning MCFs.

How to isolate MCFs

Several protocols have been developed to efficiently isolate and enrich cultures of MCFs, including a new protocol developed and present in **Appendix 1**. Other developed protocols include, 1) pre-incubation in an acidic medium at high temperatures to favor thermophilic strains (Sudhadham et al. 2008), 2) using mineral oil to select for hydrophobic black propagules (Satow et al. 2008), and 3) selection using volatile aromatic hydrocarbons like toluene and styrene (Isola et al. 2013).

Bioproducts aplenty

Melanin is one of the important secondary metabolite products produced by MCFs, it has also garnered the attention of companies who want to produce mass quantities of it. There are currently many biotech companies looking into utilizing the complex properties of melanin for the US government, and space astronauts (www.melatech.space). Melanin has

radiotropic and radioprotective qualities which could be of great use in the space industry for radiation shield designs (Vasileiou and Summerer 2021). MCFs have been collected from around the Chernobyl Nuclear Power Plant in Ukraine, indicating radiotopism (Zhdanova et al. 2000). Once MCFs are understood more completely, there could be an effective means of helping digest sources of radiation using these MCFs. Though a concern is how to properly store or contain the likely radioactively accumulated MCFs once they have metabolized the radiation sources (Barnett et al. 2001). Melanin has excellent thermo-tolerance properties which could be useful for high altitudes to generate and maintain warmth. Melanin also has a positive effect on fungal cell walls, allowing for high tolerance in acidic or hyper-saline environments. This could be a very useful shield or bio-material to protect against these extremes. Another consideration could be in regards to volatile organic compounds (VOCs), as MCFs function as effective bio-filters from VOCs produced from construction materials and cleaning chemicals that could lead to sick building syndrome (SBS) (Prenafeta-Boldú et al. 2019).

Another creative outlet for fungal melanin is as a natural source of fabric or substrate dye (Shahid et al. 2013; Liu et al. 2020; Sun et al. 2011; Amal et al. 2011). Having new sources of clean and non-reactive dye could be of great benefit to humanity. A fun and interesting resource, MCFs *Epicoccum granulatum* and *Phialophora parasitica*, isolated from infected *Aquilaria sp.* trees in Eastern and South Asia, produce the most sought-after smoky agarwood or commonly known as oud (Du et al. 2022). First-grade oud is one of the most expensive materials found on earth, costing about $100,000/kg (Akter et al. 2013).

This product is highly sought after in the perfume industry and cracking this pathway could become incredibly profitable (Chhipa et al. 2017).

MCFs, when found in different environments have been noted to range in roles including opportunistic pathogens, commensal species and even positive protective forces. The large breadth of knowledge accumulated over time gives us a glimpse on how much more we still need to study to understand this group.

References

Akhtar N, Mannan MA-U. 2020. Mycoremediation: Expunging environmental pollutants. Biotechnol Rep (Amst). 26:e00452.

Amal, Abeer, Samia. Selection of Pigment (Melanin) production in Streptomyces and their application in Printing and Dyeing of Wool Fabrics. J Adv Pharm Technol Res.

Atalla, Zeinab, Eman, Amani. Screening of some marine-derived fungal isolates for lignin degrading enzymes (LDEs) production. Agric Biol Journal North American.

Barnett CL, Beresford NA, Frankland JC, Self PL, Howard BJ, Marriott JVR. 2001. Radiocaesium intake in Great Britain as a consequence of the consumption of wild fungi. Mycologist. 15(3):98–104.

Bik HM, Halanych KM, Sharma J, Thomas WK. 2012. Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. PLoS One. 7(6):e38550.

Bloomfield BP, Doolin B. 2017a. Landfarming: A contested space for the management of waste from oil and gas extraction. Environ Plan A. 49(11):2457–2476.

Boustie J, Tomasi S, Grube M. 2011. Bioactive lichen metabolites: alpine habitats as an untapped source. Phytochem Rev. 10(3):287–307.

Brannock PM, Sharma J, Bik HM, Thomas WK, Halanych KM. 2017. Spatial and temporal variation of intertidal nematodes in the northern Gulf of Mexico after the Deepwater Horizon oil spill. Mar Environ Res. 130:200–212.

Brunauer, Blaha, Hager, Türk. Lichenoid structures in vitro of a cultured lichenicolous fungus. Symbiosis.

Butler MJ, Day AW. 1998. Fungal melanins: a review. Can J Microbiol. 44(12):1115–1136.

Cappitelli F, Nosanchuk JD, Casadevall A, Toniolo L, Brusetti L, Florio S, Principi P, Borin S, Sorlini C. 2007. Synthetic consolidants attacked by melanin-producing fungi: case study of the biodeterioration of Milan (Italy) cathedral marble treated with acrylics. Appl Environ Microbiol. 73(1):271–277.

Cerniglia CE, Sutherland JB. 2001. Bioremediation of polycyclic aromatic hydrocarbons by ligninolytic and non-ligninolytic fungi. Fungi in Bioremediation.:136–187. doi:10.1017/cbo9780511541780.008. http://dx.doi.org/10.1017/cbo9780511541780.008.

Chang R, Cao W, Wang Y, Li S, Li X, Bose T, Si HL. 2022. Melanodevriesia, a new genus of endolichenic oleaginous black yeast recovered from the Inner Mongolia Region of China. Fungal Syst Evol. 9:1–9.

Chhipa H, Chowdhary K, Kaushik N. 2017. Artificial production of agarwood oil in *Aquilaria sp.* by fungi: a review. Phytochem Rev. 16(5):835–860.

Clemente AR, Anazawa TA, Durrant LR. 2001. Biodegradation of polycyclic aromatic hydrocarbons by soil fungi. Braz J Microbiol. 32(4):255–261.

de Cock AW. 1994a. Population biology of *Hortaea werneckii* based on restriction patterns of mitochondrial DNA. Antonie Van Leeuwenhoek. 65(1):21–28.

Cofone L, Walker JD, Cooney JJ. 1973. Utilization of Hydrocarbons by *Cladosporium resinae.* Journal of General Microbiology. 76(1):243–246. doi:10.1099/00221287-76-1-243. http://dx.doi.org/10.1099/00221287-76-1-243.

Coon SL, Kotob S, Jarvis BB, Wang S, Fuqua WC, Weiner RM. 1994. Homogentisic acid is the product of MelA, which mediates melanogenesis in the marine bacterium *Shewanella colwelliana* D. Appl Environ Microbiol. 60(8):3006–3010.

Cunliffe M, Hollingsworth A, Bain C, Sharma V, Taylor JD. 2017. Algal polysaccharide utilisation by saprotrophic planktonic marine fungi. Fungal Ecol. 30:135–138.

Dakal TC, Cameotra SS. 2012. Microbially induced deterioration of architectural heritages: routes and mechanisms involved. Environmental Sciences Europe. 24(1):1–13.

De Hoog GS. 1979. Nomenclatural notes on some black yeast-like hyphomycetes. Taxon. 28(4):347–348.

De Hoog GS, Zeng JS, Harrak MJ, Sutton DA. 2006. *Exophiala xenobiotica* sp. nov., an opportunistic black yeast inhabiting environment rich in hydrocarbons. Antonie Van Leeuwenhoek. 90(3):257–268.

De Las Peñas A, Pan S-J, Castaño I, Alder J, Cregg R, Cormack BP. 2003. Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. Genes Dev. 17(18):2245–2258.

Du T-Y, Dao C-J, Mapook A, Stephenson SL, Elgorban AM, Al-Rejaie S, Suwannarach N, Karunarathna SC, Tibpromma S. 2022. Diversity and Biosynthetic Activities of Agarwood Associated Fungi. Diversity. 14(3):211.

El Baidouri F, Zalar P, James TY, Gladfelter AS, Amend AS. 2021. Evolution and Physiology of Amphibious Yeasts. Annu Rev Microbiol. 75:337–357.

Ettinger CL, Eisen JA. 2019. Characterization of the Mycobiome of the Seagrass, *Zostera marina*, Reveals Putative Associations With Marine Chytrids. Front Microbiol. 10:2476.

Ettinger CL, Eisen JA. 2021. Correction: Fungi, bacteria and oomycota opportunistically isolated from the seagrass, *Zostera marina*. PLoS One. 16(5):e0251536.

Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, et al. 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet. 4(4):e1000046.

Field JA, de Jong E, Feijoo Costa G, de Bont JA. 1992. Biodegradation of polycyclic aromatic hydrocarbons by new isolates of white rot fungi. Appl Environ Microbiol. 58(7):2219–2226.

Gorbushina AA, Beck A, Schulte A. 2005. Microcolonial rock inhabiting fungi and lichen photobionts: evidence for mutualistic interactions. Mycol Res. 109(Pt 11):1288–1296.

Gorbushina AA, Krumbein WE, Hamman CH, Panina L, Soukharjevski S, Wollenzien U. 1993. Role of black fungi in color change and biodeterioration of antique marbles. Geomicrobiol J. 11(3-4):205–221.

Gostincar C, Grube M, de Hoog S, Zalar P, Gunde-Cimerman N. 2010. Extremotolerance in fungi: evolution on the edge. FEMS Microbiol Ecol. 71(1):2–11.

Gostinčar C, Stajich JE, Kejžar A, Sinha S, Nislow C, Lenassi M, Gunde-Cimerman N. 2021. Seven Years at High Salinity—Experimental Evolution of the Extremely Halotolerant Black Yeast *Hortaea werneckii*. Journal of Fungi. 7(9):723.

Gostinčar C, Stajich JE, Zupančič J, Zalar P, Gunde-Cimerman N. 2018. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. BMC Genomics. 19(1):364.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42(Database issue):D699–704.

Gunde-Cimerman N, Sonjak S, Zalar P, Frisvad JC, Diderichsen B, Plemenitaš A. 2003. Extremophilic fungi in arctic ice: a relationship between adaptation to low temperature and water activity. Physics and Chemistry of the Earth, Parts A/B/C. 28(28):1273–1278.

Guzzo GL, Andrews JM, Weyrich LS. 2022. The Neglected Gut Microbiome: Fungi, Protozoa, and Bacteriophages in Inflammatory Bowel Disease. Inflamm Bowel Dis. 28(7):1112–1122.

Harutyunyan S, Muggia L, Grube M. 2008. Black fungi in lichens from seasonally arid habitats. Stud Mycol. 61:83–90.

Husaini A, Roslan HA, Hii KSY, Ang CH. 2008. Biodegradation of aliphatic hydrocarbon by indigenous fungi isolated from used motor oil contaminated sites. World J Microbiol Biotechnol. 24(12):2789–2797.

Ibrar M, Ullah MW, Manan S, Farooq U, Rafiq M, Hasan F. 2020. Fungi from the extremes of life: an untapped treasure for bioactive compounds. Appl Microbiol Biotechnol. 104(7):2777–2801.

Isola D, Scano A, Orrù G, Prenafeta-Boldú FX, Zucconi L. 2021. Hydrocarbon-Contaminated Sites: Is There Something More Than *Exophiala xenobiotica*? New Insights into Black Fungal Diversity Using the Long Cold Incubation Method. J Fungi (Basel). 7(10). doi:10.3390/jof7100817. http://dx.doi.org/10.3390/jof7100817.

Isola D, Selbmann L, de Hoog GS, Fenice M, Onofri S, Prenafeta-Boldú FX, Zucconi L. 2013. Isolation and screening of black fungi as degraders of volatile aromatic hydrocarbons. Mycopathologia. 175(5-6):369–379.

Kiiskinen L-L, Rättö M, Kruus K. 2004. Screening for novel laccase-producing microbes. J Appl Microbiol. 97(3):640–646.

Kogej T, Gostinčar C, Volkmann M, Gorbushina AA, Gunde-Cimerman N. 2006. Mycosporines in Extremophilic Fungi—Novel Complementary Osmolytes? Environ Chem. 3(2):105–110.

Krumbein. Patina and cultural heritage–a geomicrobiologist's perspective. Cult Herit Res a pan-European Chall.

Krumbein WE, Urzì CE, Gehrmann C. 1991. Biocorrosion and biodeterioration of antique and medieval glass. Geomicrobiol J. 9(2-3):139–160.

Landry RL, Embers ME. 2022. Does Dementia Have a Microbial Cause? NeuroSci. 3(2):262–283.

Liu Y, Zhang Y, Yu Z, Qi C, Tang R, Zhao B, Wang H, Han Y. 2020. Microbial dyes: dyeing of poplar veneer with melanin secreted by *Lasiodiplodia theobromae* isolated from wood. Appl Microbiol Biotechnol. 104(8):3367–3377.

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464(7287):367–373.

Marchetta A, Gerrits van den Ende B, Al-Hatmi AMS, Hagen F, Zalar P, Sudhadham M, Gunde-Cimerman N, Urzì C, de Hoog S, De Leo F. 2018. Global Molecular Diversity of the Halotolerant Fungus *Hortaea werneckii*. Life. 8(3). doi:10.3390/life8030031. http://dx.doi.org/10.3390/life8030031.

Marvasi M, Donnarumma F, Frandi A, Mastromei G, Sterflinger K, Tiano P, Perito B. 2012. Black microcolonial fungi as deteriogens of two famous marble statues in Florence, Italy. Int Biodeterior Biodegradation. 68:36–44.

Mitchison-Field LMY, Vargas-Muñiz JM, Stormo BM, Vogt EJD, Van Dierdonck S, Pelletier JF, Ehrlich C, Lew DJ, Field CM, Gladfelter AS. 2019. Unconventional Cell Division Cycles from Marine-Derived Yeasts. Curr Biol. 29(20):3439–3456.e5.

Mohammadi P, Krumbein WE. 2008. Biodeterioration of ancient stone materials from the Persepolis monuments (Iran). Aerobiologia. 24(1):27–33.

Mohammadian E, Arzanlou M, Babai-Ahari A. 2017. Diversity of culturable fungi inhabiting petroleum-contaminated soils in Southern Iran. Antonie Van Leeuwenhoek. 110(7):903–923.

Moreno LF, Vicente VA, de Hoog S. 2018b. Black yeasts in the omics era: Achievements and challenges. Med Mycol. 56(suppl_1):32–41.

Mtui G, Nakamura Y. 2008. Lignocellulosic enzymes from *Flavodon flavus,* a fungus isolated from Western Indian Ocean off the coast of Dar es Salaam, Tanzania. Afr J Biotechnol. 7(17). doi:10.4314/ajb.v7i17.59228. [accessed 2022 Oct 11]. https://www.ajol.info/index.php/ajb/article/view/59228.

Nagpal R, Neth BJ, Wang S, Mishra SP, Craft S, Yadav H. 2020. Gut mycobiome and its interaction with diet, gut bacteria and alzheimer's disease markers in subjects with mild cognitive impairment: A pilot study. EBioMedicine. 59:102950.

Nagy LG, Kovács GM, Krizsán K. 2018. Complex multicellularity in fungi: evolutionary convergence, single origin, or both? Biol Rev Camb Philos Soc. 93(4):1778–1794.

Noronha P. 2011. Yeast biopaintings: Biofilms as an artistic instrument. Leonardo. 44(1):38–42.

O'dell TE, Massicotte HB, Trappe JM. 1993. Root colonization of *Lupinus latifolius Agardh.* and *Pinus contorta* Dougl. by *Phialocephala fortinii* Wang & Wilcox. New Phytol. 124(1):93–100.

Okpokwasili GC, Okorie BB. 1988. Biodeterioration potentials of microorganisms isolated from car engine lubricating oil. Tribol Int. 21(4):215–220.

Onofri, Pagano, Zucconi, Tosi. 1999. *Friedmanniomyces endolithicus* (Fungi, Hyphomycetes), anam-gen and sp nov, from continental Antarctica. Nova Hedwigia. https://www.schweizerbart.de/content/papers_preview/download/93965.

Oren A, Gunde-Cimerman N. 2007. Mycosporines and mycosporine-like amino acids: UV protectants or multipurpose secondary metabolites? FEMS Microbiol Lett. 269(1):1–10.

Pacelli C, Bryan RA, Onofri S, Selbmann L, Zucconi L, Shuryak I, Dadachova E. 2018. Survival and redox activity of *Friedmanniomyces endolithicus*, an Antarctic endemic black meristematic fungus, after gamma rays exposure. Fungal Biol. 122(12):1222–1227.

Pangallo D, Bučková M, Kraková L, Puškárová A, Šaková N, Grivalský T, Chovanová K, Zemánková M. 2015. Biodeterioration of epoxy resin: a microbial survey through culture-independent and culture-dependent approaches. Environ Microbiol. 17(2):462–479.

Paudel B, Bhattarai HD, Koh HY, Lee SG, Han SJ, Lee HK, Oh H, Shin HW, Yim JH. 2011. Ramalin, a novel nontoxic antioxidant compound from the Antarctic lichen Ramalina terebrata. Phytomedicine. 18(14):1285–1290.

Perini L, Gostinčar C, Anesio AM, Williamson C, Tranter M, Gunde-Cimerman N. 2019. Darkening of the Greenland Ice Sheet: Fungal Abundance and Diversity Are Associated With Algal Bloom. Front Microbiol. 10:557.

Perini L, Gostinčar C, Gunde-Cimerman N. 2019. Fungal and bacterial diversity of Svalbard subglacial ice. Sci Rep. 9(1):20230.

Poyntner C, Blasi B, Arcalis E, Mirastschijski U, Sterflinger K, Tafer H. 2016. The Transcriptome of *Exophiala dermatitidis* during Ex-vivo Skin Model Infection. Front Cell Infect Microbiol. 6:136.

Poyntner C, Mirastschijski U, Sterflinger K, Tafer H. 2018. Transcriptome Study of an *Exophiala dermatitidis* PKS1 Mutant on an ex Vivo Skin Model: Is Melanin Important for Infection? Front Microbiol. 9:1457.

Prenafeta-Boldú FX, Medina-Armijo C, Isola D. 2022. 5 - Black fungi in the built environment—The good, the bad, and the ugly. In: Pacheco-Torgal F, Ivanov V, Falkinham JO, editors. Viruses, Bacteria and Fungi in the Built Environment. Woodhead Publishing. p. 65–99.

Rizk SM, Magdy M. 2022. An indigenous inland genotype of the black yeast *Hortaea werneckii* inhabiting the great pyramid of Giza, Egypt. Front Microbiol. 13. doi:10.3389/fmicb.2022.997495.

Rizk SM, Magdy M, Leo FD, Werner O, Rashed MA-S, Ros RM, Urzì C. 2021. A New Extremotolerant Ecotype of the Fungus *Pseudotaeniolina globosa* Isolated from Djoser Pyramid, Memphis Necropolis, Egypt. J Fungi (Basel). 7(2). doi:10.3390/jof7020104.

Romeo O, Marchetta A, Giosa D, Giuffrè L, Urzì C, De Leo F. 2020. Whole Genome Sequencing and Comparative Genome Analysis of the Halotolerant Deep Sea Black Yeast *Hortaea werneckii*. Life. 10(10). doi:10.3390/life10100229.

Santos M, Cesanelli I, Diánez F, Sánchez-Montesinos B, Moreno-Gavíra A. 2021. Advances in the Role of Dark Septate Endophytes in the Plant Resistance to Abiotic and Biotic Stresses. J Fungi (Basel). 7(11). doi:10.3390/jof7110939. http://dx.doi.org/10.3390/jof7110939.

Santos SGD, Silva PRA da, Garcia AC, Zilli JÉ, Berbara RLL. 2017. Dark septate endophyte decreases stress on rice plants. Braz J Microbiol. 48(2):333–341.

Sarkar J, Saha A, Roy A, Bose H, Pal S, Sar P, Kazy SK. 2020. Development of nitrate stimulated hydrocarbon degrading microbial consortia from refinery sludge as potent bioaugmenting agent for enhanced bioremediation of petroleum contaminated waste. World J Microbiol Biotechnol. 36(10):156.

Satow MM, Attili-Angelis D, de Hoog GS, Angelis DF, Vicente VA. 2008. Selective factors involved in oil flotation isolation of black yeasts from the environment. Stud Mycol. 61:157–163.

Schmaler-Ripcke J, Sugareva V, Gebhardt P, Winkler R, Kniemeyer O, Heinekamp T, Brakhage AA. 2009. Production of pyomelanin, a second type of melanin, via the tyrosine degradation pathway in *Aspergillus fumigatus*. Appl Environ Microbiol. 75(2):493–503.

Schultzhaus ZS, Schultzhaus JN, Romsdahl J, Chen A, Hervey WJ Iv, Leary DH, Wang Z. 2020. Proteomics Reveals Distinct Changes Associated with Increased Gamma Radiation Resistance in the Black Yeast *Exophiala dermatitidis*. Genes. 11(10). doi:10.3390/genes11101128.

Selbmann L, Grube M, Onofri S, Isola D, Zucconi L. 2013. Antarctic epilithic lichens as niches for black meristematic fungi. Biology. 2(2):784–797.

Selbmann L, de Hoog GS, Zucconi L, Isola D, Onofri S. 2014. Black Yeasts in Cold Habitats. In: Buzzini P, Margesin R, editors. Cold-adapted Yeasts: Biodiversity, Adaptation Strategies and Biotechnological Significance. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 173–189.

Selbmann L, Onofri S, Zucconi L, Isola D, Rottigni M, Ghiglione C, Piazza P, Alvaro MC, Schiaparelli S. 2015a. Distributional records of Antarctic fungi based on strains preserved in the Culture Collection of Fungi from Extreme Environments (CCFEE) Mycological Section associated with the Italian National Antarctic Museum (MNA). McCalls. 10:57–71.

Seyedmousavi S, Badali H, Chlebicki A, Zhao J, Prenafeta-Boldú FX, De Hoog GS. 2011. *Exophiala sideris,* a novel black yeast isolated from environments polluted with toxic alkyl benzenes and arsenic. Fungal Biol. 115(10):1030–1037.

Shahid M, Shahid-ul-Islam, Mohammad F. 2013. Recent advancements in natural dye applications: a review. J Clean Prod. 53:310–331.

Shuttleworth, Newman. 2018. Bioremediation in urban pollution mitigation: applications to solid media. Urban Pollution: Science.

Singh H. 2006. Mycoremediation: Fungal Bioremediation. John Wiley & Sons.

Sinha S, Flibotte S, Neira M, Formby S, Plemenitaš A, Cimerman NG, Lenassi M, Gostinčar C, Stajich JE, Nislow C. 2017. Insight into the Recent Genome Duplication of the Halophilic Yeast *Hortaea werneckii*: Combining an Improved Genome with Gene Expression and Chromatin Structure. G3. 7(7):2015–2022.

Sterflinger K, Krumbein WE. 1997. Dematiaceous fungi as a major agent for biopitting on Mediterranean marbles and limestones. Geomicrobiol J. 14(3):219–230.

Sun D, Sun D, Yu X. 2011 Oct 28. Ultrasonic-Assisted Dyeing of Poplar Veneer. WFS.:442–448.

Tesei D, Marzban G, Marchetti-Deschmann M, Tafer H, Arcalis E, Sterflinger K. 2015. Proteome of tolerance fine-tuning in the human pathogen black yeast *Exophiala dermatitidis*. J Proteomics. 128:39–57.

Trovão J, Soares F, Tiago I, Catarino L, Portugal A, Gil F. 2021. A contribution to understand the Portuguese emblematic Ançã limestone bioreceptivity to fungal colonization and biodeterioration. J Cult Herit. 49:305–312.

Tsuneda A, Hambleton S, Currah RS. 2011. The anamorph genus *Knufia* and its phylogenetically allied species in *Coniosporium*, *Sarcinomyces*, and *Phaeococcomyces*. Botany. 89(8):523–536.

Urzì, De Leo. 2000. Biodeterioration of cultural heritage in Italy: State of art. ADRIANE workshop.

Vasileiou T, Summerer L. 2021. Correction: A biomimetic approach to shielding from ionizing radiation: The case of melanized fungi. PLoS One. 16(8):e0257068.

Voigt O, Knabe N, Nitsche S, Erdmann EA, Schumacher J, Gorbushina AA. 2020. An advanced genetic toolkit for exploring the biology of the rock-inhabiting black fungus *Knufia petricola.* Sci Rep. 10(1):22021.

Yazdanparast SA, Mohseni S, De Hoog GS. 2017. Consistent high prevalence of *Exophiala dermatitidis*, a neurotropic opportunist, on railway sleepers. J Mycol Med.

Zalar P, Zupančič J, Gostinčar C, Zajc J, de Hoog GS, De Leo F, Azua-Bustos A, Gunde-Cimerman N. 2019. The extremely halotolerant black yeast *Hortaea werneck*ii - a model for intraspecific hybridization in clonal fungi. IMA Fungus. 10:10.

Zhdanova NN, Zakharchenko VA, Vember VV, Nakonechnaya LT. 2000. Fungi from Chernobyl: mycobiota of the inner regions of the containment structures of the damaged nuclear reactor. Mycol Res. 104(12):1421–1426.

Description of new micro-colonial fungi species *Neophaeococcomyces mojavensis*, *Coniosporium tulheliwenetii. Taxawa tesnikishii* cultured from biological soil crusts

Abstract

Black yeasts or Micro-Colonial Fungi (MCFs) are stress-tolerant micro-eukaryotes that specialize in extreme environments. MCFs are paraphyletic and found in the Orders Chaetothyriales (Eurotiomycetes) and Dothideales (Dothidiomycetes). We have isolated and described three new MCFs species from desert biological soil crusts (BSCs) collected from two arid land regions: Joshua Tree National Park (Mojave Desert) and UC Natural Reserve at Boyd Deep Canyon (confluence of Mojave and Sonoran Deserts). BSCs are composite assemblages of cyanobacteria, eukaryotic algae, fungi, lichens, and bryophytes embedded into the surface of desert soils, providing a protective buffer against the harsh desert environment. Our work focused on one type of desert BSC, the cyanolichen crust dominated by: *Collema sp.* Using culture-dependent protocols, three MCFs were axenically isolated from their respective samples along with the extracted DNA. Their genomes were sequenced using Illumina and Nanopore, and finally assembled and annotated using hybrid assembly approaches and established bioinformatics pipelines to conduct final taxonomic phylogenetic analysis and placement. Culture-independent processes were included as an added layer of MCF taxonomic confirmation. We have assessed amplicon data associated with each crust sample to observe for all unknown Chaetothyriales MCF genera and species. The three species described here are unique specimen from desert BSCs, specifically: *Chaetothyriales sp.112,* here named – *"Neophaeococcomyces mojavensis"*, *Cladosporium sp.115* here named – "*Cladosporium*

tulheliwenetii", and *Dothideales sp.119* here named – "*Taxawa tesnikishii"*. We were able to place all three new strains phylogenetically in their respective clades while also taking other unknown Chaetothyriales MCFs from our amplicon data to quantify how much is still unknown in this order. These results provide a better understanding of Chaetothyriales and reveal that there is still much to learn about this order and the MCFs.

Introduction

Biological Soil Crusts

Dryland ecosystems constitute some of the largest terrestrial biomes, collectively covering 47.2% of Earth's land surface (6.15 billion hectares) (Lal 2004). Drylands are expanding due to climate change and population pressures, while also supporting over 38% of the global human population (Lal 2004; Maestre et al. 2012). Biological Soil Crusts (BSCs) cover up to 70% of dryland ecosystems (Rozenstein and Karnieli 2015). They are an ancient feature of dryland ecosystems, with the earliest known fossil records indicating nearly 2.6 billion years ago (Beraldi-Campesi and Retallack 2016). The initial phases of BSC formation in drylands are the stabilization of the soil surface by filamentous cyanobacteria, followed by the colonization of lichens and bryophytes (Weber et al. 2016). Within dryland ecosystems, the carbon and nitrogen dynamics are often described as the "pulse-reserve model" where precipitation events happen sporadically, stimulating biological activity which generates biomass and organic matter. After extended periods of heat and dryness, the biological activity goes into reserve and becomes dormant again (Pietrasiak et al. 2013). BSCs also increase soil fertility by increasing soil stability and

adding fixed atmospheric carbon and nitrogen to the soil underneath, allowing for greater complex microbial and micro-faunal populations (Housman et al. 2007).

Within BSCs is a consortium of cyanobacteria, algae, diazotrophic bacteria, lichen, fungi, and bryophytes that colonize and stabilize the soil surfaces against natural forces like wind and water (Warren et al. 2019). These microorganisms recruit partners to create a mutualistic symbiotic relationship which helps maintain a self-sustaining micro-ecosystem. Phototrophic organisms fix carbon that is later used by surrounding micro-organisms as an energy source (Green et al. 2016). The presence of chemoheterotrophic bacteria and cyanobacteria indicates that BSCs are capable of fixing atmospheric nitrogen to increase levels of fixed nitrogen for nearby living vascular plants (Belnap and Harper 1995; Evans and Ehleringer 1993).BSCs have been shown to reduce surface erodibility as filaments of cyanobacterial sheath material lipopolysaccharides (LPS) entangle surface particles and create a crust that is more resistant to entrainment than the layers below (Lal 2004).

Different ecological sites promote different crust taxa. BSCs are classified into types based on the dominant photosynthetic microorganism in the crust which includes: cyanobacteria, algae, lichens, or mosses to produce Cyano-Lichenized Crust, Light Algal Crusts, Dark Algal Crusts, Lichenized Crust, Rough Moss Crust and Smooth Moss Crust (Pietrasiak 2005).The community types are classified based on dominant taxa and their distinct morphologies (Pietrasiak et al. 2013). BSC's main roles include: 1) mediating almost all inputs and outputs (gasses, nutrients, water) to and from the strata above and below the

surface, 2) being the zone of high nutrient deposition, transformation, and availability, 3) structuring temporal, spatial, and compositional aspects of the vascular plant community, and 4) facilitating the direct delivery of carbon, nutrients, and water by BSCs from the soil interspace to nearby vascular plants.

Micro-colonial Fungi

Black yeasts, meristematic fungi, or micro-colonial fungi, are prevalent names used to describe a special group of melanized fungi. Micro-colonial fungi (MCFs) are ubiquitous and found in the most extreme environments. They are present in widely divergent habitats, such as on the indoor rubbers of dishwashers, in oil deposits, on the surfaces of Mediterranean monuments, on the radioactive walls of the Chernobyl Station, in skin infections in third-world countries, in human superficial and invasive infections, or colonizing lungs of cystic fibrosis patients(de Hoog et al. 2005; Hoog et al. 2000; Vicente et al. 2008; Gümral et al. 2014; Saunte et al. 2012; Thanh and Hien 2019; Mehdiabadi and Schultz ; Little and Currie 2008; de Hoog and Hermanides-Nijhof 1977). MCFs are compelling not only because of where we can discover them, and their unique physiological features but also how little we know about them.

MCFs are identified by three main characteristics: meristematic growth, presence of 1,8-DHN melanin, and membrane-associated carotenoids and intracellular mycosporine-like amino acids (Staley et al. 1982; Gorbushina et al. 2003). These features distinguish them from other fungi, yet phylogenetically they fall into two paraphyletic Ascomycota clades, the Dothideomycetes and Chaetothyriomycetes. Melanin is a highly complex secondary

metabolite that is produced by many organisms through different evolutionary lineages (Siletti et al. 2017; Wheeler 1983). Most importantly in MCFs, it has been used as a virulence factor as a non-specific armor against host immune system to infect its host (Steenbergen and Casadevall 2003; Casadevall et al. 2003; Casadevall et al. 2000; Nosanchuk et al. 2015; Taborda et al. 2008; Gómez and Nosanchuk 2003). Available research indicates melanin has other functional roles, such as metal chelation, photoprotection, protection from heat and cold stress, binding to antifungals to reduce efficacy, radioprotection (physical shielding and quenching of cytotoxic free radicals), protection from desiccation, interacting with a range of electromagnetic radiation frequencies and mechanical-chemical cellular strength (Dadachova et al. 2008; Casadevall et al. 2003; Gorbushina et al. 2003). Understanding poorly studied fungal function in one extreme environment could yield insight into MCF adaptation in other extreme environments.

The objective of this study was to collect MCF from biological soil crust samples from desert drylands. During our collection, we isolated 3 MCF that had not been described previously. Our efforts to converse with indigenous tribes allowed us to provide a new naming scheme for our strains. We also used a new methodology described herein using BUSCO nucleotide and amino acid sequences to build a multi-locus phylogenetic tree, which proved to be similar to a more conventional tree of MCF which used four highly conserved markers.

Materials and Methods

Sampling

Biological soil crusts were collected from two locations in Southern California. The first crust location at the northern edge of Joshua Tree National Park, in the Sheephole Valley Wilderness, Bureau of Land Management. The second crust location was from Boyd Deep Canyon, near the campgrounds in the UC Natural Reserve. Metadata and coordinates can be found in **Table 1**. All materials were collected under permits from the BLM or permission of the UC National Reserves.

Culturing and Isolation

Media for isolation, cultivation, and collection used were Malt Extract Yeast Extract (MEYE) media for nutrient-rich complex media and nutrient-poor, oligotrophic media Glucose Asparagine Agar (GAA). The recipe for both media can be found in our protocols.io and the step-by-step protocol for culturing MCFs (Kurbessoian 2019), also found in **Appendix 1**). The mixing of the sample with the media provides an environment with different layers of dissolved oxygen. Fast-growing fungi require full oxygenation and grow on the surface of the media, while slow-growing MCFs propagate on the bottom of the culture plate, requiring low levels of oxygen. This also allowed a much easier way to culture and select the strains. Once grown, axenic MCFs are isolated by using a flamed loop to pick and place onto clean media.

Morphological analysis

Strains were grown on Malt Extract Agar (MEA; Oxoid, England), Oatmeal Agar (OA; Oxoid) and Potato dextrose Agar (PDA; Sigma, Germany) incubated at 28℃ for 14 days under UV light. Colonies were photographed with a Nikon C-DSD230 dissecting microscope. Microscopic mounts were made in lactic acid and examined using a light microscope (Nikon Eclipse 80i). Pictures were taken with a digital camera (Nikon Digital Sight, DS-5M) and adjusted in Adobe Photoshop CS.

DNA extraction

Each isolate was grown on MEYE for approximately 1.5 weeks at 22℃. Genomic DNA was extracted from hyphae collected from MEYE plates using the CTAB protocol described in our protocols.io document (Carter-House et al. 2020). Melanin was removed using multiple iterations of phenol:chloroform and chloroform washes of our sample. Genomic DNA was measured by Nanodrop and diluted to ~28ng/µl. DNA extractions were sent to the UCR Genomics core (Riverside, CA) for 2x150bp sequencing on an Illumina NovoSeq 6000. DNA from all three isolates was also extracted and sequenced on Oxford Nanopore (ONT) platform with library preparation and sequencing following the manufacturer's directions (Oxford Nanopore, Oxford United Kingdom). Flow cell versions FAK68780, FAQ05493, FAL40904, and FAL35715 were used along with base-calling on the cluster using guppy (v.3.4.4.+a296cb) (Wick *et al.* 2019).

ITS Sequencing

To identify the isolates, extracted DNA was used as a template for PCR amplification using funga primersl ITS1-F (CTTGGTCATTTAGAGGAAGTAA) and ITS4 (TCCTCCGCTTATTGATATGC) along with the PCR setup. PCR was performed in 25 uL reaction volumes using 12.5 uL of Taq 2X MasterMix (NEB), 9.5uL water, 1uL of 20 picomol ITS1-F, 1uL of 20 picomol ITS4 and 1uL of template DNA. Using Bio-Rad thermal cycler, the samples were passed through denaturation and enzyme activation at 95℃ for 5 minutes, followed by amplification for 35 cycles at the following conditions: 30 seconds at 95℃, 1 minute at 72℃, and 30 seconds at 65℃. A final 5-minute extension at 72℃ completes the protocol. PCR results typically yield between 50-60 ng of amplified DNA.

Genome assembly and annotation

Initial genome assemblies were constructed for the three MCF isolates using Illumina sequencing. Each MCF was also re-sequenced using Oxford Nanopore technology. All genomes were *de novo* assembled with the AAFTF pipeline (v.0.2.3) (Palmer and Stajich 2022) which performs read QC and filtering with BBTools bbduk (*v.38.86)* (Bushnell 2014) followed by SPAdes (*v.3.15.2*) (Bankevich *et al.* 2012) assembly using default parameters, followed by screening to remove short contigs < 200 bp and contamination using NCBI's VecScreen. The BUSCO ascomycota_odb10 database (Manni *et al.* 2021) was used to determine how complete the assembly 3 isolates were. A hybrid assembly of each was generated using MaSUrCa *(v.3.3.4)* (Zimin *et al.* 2013) as the assembler using both Nanopore and Illumina sequencing reads. General default parameters were used

except: CA_PARAMETERS=cgwErrorRate=0.15, NUM_THREADS=16, and JF_SIZE=200000000.

We predicted genes in each near-complete genome assembly with Funannotate (*v1.8.1*) (Palmer and Stajich 2020). A masked genome was created by generating a library of sequence repeats with the RepeatModeler pipeline (Smit and Hubley 2008). These species-specific predicted repeats were combined with fungal repeats in the RepBase (Bao *et al.* 2015) to identify and mask repetitive regions in the genome assembly with RepeatMasker (*v.4-1-1*) (SMIT 2004). To predict genes, *ab initio* gene predictors SNAP (*v.2013_11_29*) (Korf 2004) and AUGUSTUS (*v.3.3.3*) (Stanke *et al.* 2006) were used along with additional gene models by GeneMark.HMM-ES (*v.4.62_lic*) (Brůna *et al.* 2020), and GlimmerHMM (*v.3.0.4*) (Majoros *et al.* 2004) utilize a self-training procedure to optimize *ab initio* predictions. Additional exon evidence to provide hints to gene predictors was generated by DIAMOND BLASTX alignment of SwissprotDB proteins and polished by Exonerate (*v.2.4.0*) (Slater and Birney 2005). Finally, EvidenceModeler (*v.1.1.1*) (Haas *et al.* 2008) generated consensus gene models in Funannotate that were constructed using default evidence weights. Non-protein-coding tRNA genes were predicted by tRNAscan-SE (*v.2.0.9*) (Lowe and Chan 2016).

The annotated genomes were processed with antiSMASH (*v.5.1.1*) (Blin *et al.* 2021) to predict secondary metabolite biosynthesis gene clusters. These annotations were also incorporated into the functional annotation by Funannotate. Putative protein functions were assigned to genes based on sequence similarity to InterProScan5 (*v.5.51-85.0*) (Jones *et al.*

2014), Pfam (*v.35.0*) (Finn *et al.* 2014), Eggnog (*v.2.1.6-d35afda*) (Huerta-Cepas *et al.* 2019), dbCAN2 (*v.9.0*) (Zhang *et al.* 2018) and MEROPS (*v.12.0*) (Rawlings *et al.* 2018) databases relying on NCBI BLAST (*v.2.9.0+)* (Sofi *et al.* 2022) and HMMer (*v.3.3.2*) (Potter *et al.* 2018). Gene Ontology terms were assigned to protein products based on the inferred homology based on these sequence similarity analyses. The final annotation produced by Funannotate was deposited in NCBI as a genome assembly with gene model annotation.

Sequence Read Archive (SRA) files and genome assembly and annotation files can be found under BioProject PRJNA631111.

Mating type determination

The Mating Type (MAT) locus was identified by searching for homologous MAT genes. The identified homologous regions were examined for their conserved synteny of the MAT locus using clinker (Gilchrist and Chooi 2021) and a custom Biopython script (Cock *et al.* 2009) (Kurbessoian 2022) to extract the annotated region of the genome which contained the locus.

Telomeric repeat determination

Identification of telomeric repeat sequences was performed using the FindTelomeres.py script (https://github.com/JanaSperschneider/FindTelomeres). Briefly, this searches for chromosomal assembly with a regular expression pattern for telomeric sequences at each scaffold's 5' and 3' end. Telomere repeat sequences were also predicted using A Telomere

Identification toolkit (tidk) (*v.0.1.5*) "explore" option (https://github.com/tolkit/telomeric-identifier).

Inferring phylogenetic relationships of species

Inference of the species relationship was made through phylogenetic trees containing the three new MCF strains described here and available named species with sequence data. 197 public genomes were downloaded from NCBI's Genbank (Table 1). Using BUSCO (*v5.3.2*) ascomycota_odb10 dataset 1706 conserved markers were searched across the dataset of genomes from the public and my 3 newly completed MCFs. Summation of the completeness, genome size and quality statistics were compiled with a custom Perl script (asm_stats.pl).

Using the BUSCO generated markers found in each genome, used the script busco_to_phyling.py (part of PHYling https://github.com/stajichlab/PHYling_unified) which builds gene models from BUSCO results. The results are Coding Sequence and Protein files from each of the surveyed genomes. These files of proteins and coding sequences provided to PHYling to generate individual alignments of each marker sequence matched against Hidden Markov Models of BUSCO protein markers, aligned, trimmed for phylogenetic processing by clipkit (Steenwyk et al. 2020). Coding sequence alignments are generated by projecting protein alignments onto coding sequences. The individual alignments as well as concetenated alignments are processed for phylogenetic tree reconstruction. PHYling_unified toolset was run to build individual gene trees using either FastTree *(v.2.1.11)* or IQTREE2 (*v.2.2.1*) trees for each BUSCO family for both CDS and protein gene alignments. The gene trees were combined into a single coalescent consensus

cladogram tree with ASTRAL (*v.5.15.5*). RaXML-NG (*v.1.1.0*) was used to estimate branch lengths with the tree topology determined by ASTRAL, using the concatenate alignment of all markers produced by PHYling. The short-hand species prefixes used in the analyses were replaced with full genus, species, and strain identifiers with a PHYling script (rename_tree_nodes.pl). The tree was visualized using the R packages ggtree (*v.3.4.0*) (Yu et al. 2017), treeio (*v.1.20.0*) (Wang et al. 2020), and phytools (*v.1.0-3*) (Revell 2012).

A second multi-gene tree was generated using collected BT2 (Beta-Tubulin), CAM (Calmodulin), ITS (Inter-Transcribed Spacer), and LSU (Large SubUnit) genes from a total of 32 species. These samples were chosen to compare to the three MCF species with closely related species to confirm identification as a novel lineage. The final selection of 32 species was based on iterations of sequence searching and tree building data analyses to determine close relatives. To include genes from genomes, genome assemblies were searched by TBLASTN tool of NCBI-BLAST (*v.2.9.0+*) using *Zymoseptoria tritici* genes (AAS55060.1 for Beta tubulin) and (AEI69650.1 for Calmodulin). The collection of ITS and LSU genes were obtained from NCBI-Genbank based on BLASTN searches. Each gene file contained FASTA formatted sequences from each species or isolate, the sequence were aligned with MUSCLE (*v.5.1*) and trimmed for low quality alignment regions using Clipkit (*v.1.3.0)*. Nucleotide substitution models were determined for each alignment with IQTREE2. The alignments were combined into a single 'supermatrix' alignment with combine_multiseq_aln.py script provided in PHYling. The script also produced a partition file listing the sub-locations of each marker gene alignment in the super-alignment.

IQTree2 was run to find a substitution model fit for and partition merging. To reduce the computational burden that occurs with including a large number of organism ribosomal sequences we used the relaxed hierarchical clustering algorithm (Lanfear et al. 2014) using the tag -m MF+MERGE and running 1000 bootstrapping on the alignment.

Choosing to work with Indigenous Tribes

My choice for naming of the three MCFs isolated from BSCs is rooted in the traditional indigenous land where these were found. *Neophaeococcomyces mojaviensis* was isolated from BSCs collected outside of what is today called Joshua Tree National Park at Sheephole Valley Wilderness, territories ancestral to Yuhaaviatam/Maarenga'yam (Serrano) and Newe Segobia (Western Shoshone) Tribes (Brown and Boyd 1922; Kroeber 1925). The isolates of *Coniosporium tulheliwenetii* and *Taxawa tesnikishii* were isolated from BSCs collected from UC Reserve Boyd Deep Canyon (Reserve doi:10.21973/N3V66D), land of the traditional ancestral Cahuilla Tribe (White 2001; Hooper 1920). The word *tulheliweneti* in Cahuilla means, black-spread-be-noun, while *Taxaw* means body, and *tesnikish* means yellow. In Cahuilla traditions, Taxaw tesnikish is the deity that resided in Boyd Deep valley. To recognize the origins, we consulted with Tribal linguists and speakers to derive their names from Cahuilla and Serrano terms that describe either geographic region, oral tradition, or features that help distinguish them.

Results

Biological soil crust collection

Biological soil crusts were collected from the two regions described in **Table 1**. The method of sterile collection is through the use of sterile empty Petri dishes, stamped on top, with a sterile spatula used to grab the soil from underneath, flip the sample into the dish, place the lid on top and seal it with parafilm.

Culturing Results

Samples material was used to culture fungi using a collection of techniques in our laboratory at UC Riverside. More than 50+ fungal culture specimens were collected from biological soil crusts, 18 were MCF cultures. The pour plate protocol I used allows for fast-growing fungi to appear on the top, while MCF grows within the bottom layer of media in Petri dishes. Samples were excised from deep in the media and allowed to grow on fresh MEYE. Each sample had the DNA extracted and each strain ITS was Sanger sequenced to determine the identity. JES_112 was initially identified as a *Knufia* species. JES_115 and JES_119 were identified as Dothideales species, but there remained ambiguity on the exact species identification as a near perfect sequence match was not found for ITS1 to assign a name.

Table 1. Three species were collected from biological soil crusts collected from two desert locations in Southern California. Strain JES_112 was isolated from Sheephole Valley Wilderness in the Mojave desert, while JES_115 and JES_119 were both collected from biological soil crusts collected from the campground site in the UC Natural Reserve of Boyd Deep Canyon.

Species	Date	Location	Coordinates
JES_112	November 2017	Sheephole Valley Wilderness, Joshua Tree National Park, CA	34.2274064°N, -115.552759°W
JES_115	July 2019	Boyd Deep Canyon Reserve, CA	34.018282°N, -116.026291°W
JES_119	July 2019	Boyd Deep Canyon Reserve, CA	34.018282°N, -116.026291°W

Sequencing and assembly of MCF isolates

To gain information on the genome content for the recovered MCF isolates, we sequenced and assembled the genomes of the three isolates (**Table 2**). Each isolate was sequenced with Illumina MiSeq and Oxford Nanopore (ONT) technology and constructed into a hybrid assembly that combined both read types (**Table 2)**. The depth of coverage ranged from 25-43x Illumina coverage across the three specimens, while Nanopore ranged from 0.83-194.58x. Assessment of the genome completeness was performed with BUSCO (Benchmarking Universal Single-Copy Orthologs) and results indicated the genomes had ranges of 73-92.9% complete gene copies found, with 73-92.8% of the markers appearing as single copies in the assembly, a low fraction of duplication (0.1-0.2). Contig counts ranged from 38-480, while the average genome assembly size is about 30Mbps for all isolates. The L50 statistic, a measure of the count of the smallest number of contigs whose sum makes up half of the genome size, ranged from 9-88, and the N50 statistic, a measure

of the smallest contig length for covering half of the genome size, ranging from 117kbp - 1.232 Mbp.

To further assess completeness, I tested for the presence of telomeric repeat units "TTTAGGG/CCCTAA". These were identified as repeat arrays at both ends of contigs in JES_115 and JES_119, 61 total telomeres and 55 total telomeres respectively, though the results were inconclusive for JES_112. No telomeric repeat-containing contigs were detected in JES_112 which could be due to the fragmented nature of the genome assembly. The fragmented assembly also is reflected in the lower BUSCO score for this genome (Table 2).

The number of predicted tRNA genes ranged from 76-197, the total number of protein-coding genes predicted ranged from 9,029-10,920, and secondary metabolite cluster counts ranged from 15-27, 27 resulting from JES_112.

Table 2. Genome assembly and annotation summary statistics. All strains were sequenced with Illumina and Nanopore sequencing. With the added sequencing reads, we were able to glean telomere information along with genome completeness. It is clear that JES_112 initially was very contaminated as the BUSCO score is lower, while JES_115 and JES_119 are much closer to a more complete genome.

	JES_112	**JES_115**	**JES_119**
Scaffold Count	478	64	38
Total Length	32,593,754 bp	32,607,911 bp	29,383,553 bp
Minimum Length	561	4823	1783
Maximum Length	427,480	1,739,523	2,828,030
Mean Scaffold Length	68,187.77	509,498.61	773,251.39
Scaffold L50	87	14	9
Scaffold N50	117,971	966,560	1,232,081
Scaffold L90	282	31	24
Scaffold N90	32,261	507,175	558,687
BUSCO Complete % Illumina	84.7	90.2	97.4
BUSCO Complete % Nanopore	73.4	92.6	92.9
BUSCO Duplicate % Nanopore	0.2	0.1	0.1
BUSCO Single % Nanopore	73.2	92.5	92.8
GC%	48.25	52.55	51.98
Illumina coverage	34.58x	43.45x	25.38x
Nanopore coverage	0.83x	20.44x	194.58x
Telomere Count	0, 0F, 0R	61, 29 F, 32 R	55, 29 F, 26 R
Number of Genes	10,996	9,304	9,386

mRNA genes	10,924	9,226	9,189
tRNA genes	72	78	197
GO Terms	6,001	5,089	5,321
InterProScan	8,523	7,079	7,291
EGGNOG	9,934	8,487	8,511
PFAM	7,520	5,982	6,219
CAZYme	309	257	346
MEROPS	344	284	291
Secretion	674	532	631

Mating type loci determination

The mating types were determined using prior work on Chaetothyriales and Dothideomycetes fungi including *Exophiala dermatitidis*, and *Aureobasidium pullulans* (Metin et al. 2019; Teixeira et al. 2017). Using collected nucleotide and protein FASTA sequences, BLAST was used to identify the genes found in these loci. The genes in this locus are generally flanked by SLA2 and APN2 and may sometimes have alternative genes within. **Figure 2** was generated using JES_112, *Exophiala aquamarina* CBS 119918, *Exophiala dermatitidis* NIH/UT8656, *Exophiala dermatitidis* CBS 115663, and *Capronia coronata* CBS 617.96. The MAT locus is flanked by two genes, SLA2 (purple) and APN2 (orange). The MAT 1-1 gene, MAT 1-1-4 (green), and MAT 1-1-1 (teal) are observed in *Exophiala dermatitidis* CBS 115663 and *Capronia coronata*, while *Exophiala dermatitidis* NIH/UT8656 contains MAT 1-2 (pink) as does *Capronia coronata*. Our organism in question, *Neophaeococcomyces mojaviensis* contains the MAT 1-2 gene along with the

MAT 1-1-1 gene. This is similarly seen with *Neophaeococcomyces aloes* FJII-L3-CM-P2. This is confusing as the MAT 1-1-1 gene usually pairs with MAT 1-1-4. The gene arrow in blue is indicated as an L-type calcium channel domain, while the gene in yellow is an unidentified functional gene or domain.

Figure 3 was generated using JES_115 and JES_119 and two other species including, *Aureobasidium pullulans* EXF-150 and *Coniosporium apollinis* CBS 100218. All species described here only have the MAT 1-2 gene. The MAT locus is flanked by two genes, SLA2 (purple) and APN2 (orange). Dothideomycetes mating type loci look very different from the Chaetothyriales mating type loci. Other genes included in this locus concern only a single MAT locus, MAT 1-2 (pink), and rRNA adenine demethylase (yellow). Certain genes like the AP endonuclease (teal) and a hypothetical protein (blue) are not seen in all four genomes.

MCF morphology

Strains showed limited and slow growth with black colonies on all media tested. Strain JES_119 produced abundant unilateral, hyaline budding cells which soon inflated, developed cruciate septation, and became melanized. Strain JES_115 was hyphal with lateral chains of swollen cells which occasionally detached. Strain JES_112 lacked sporulation, the thallus consisting of melanized densely septate hyphae with a series of somewhat inflated cells

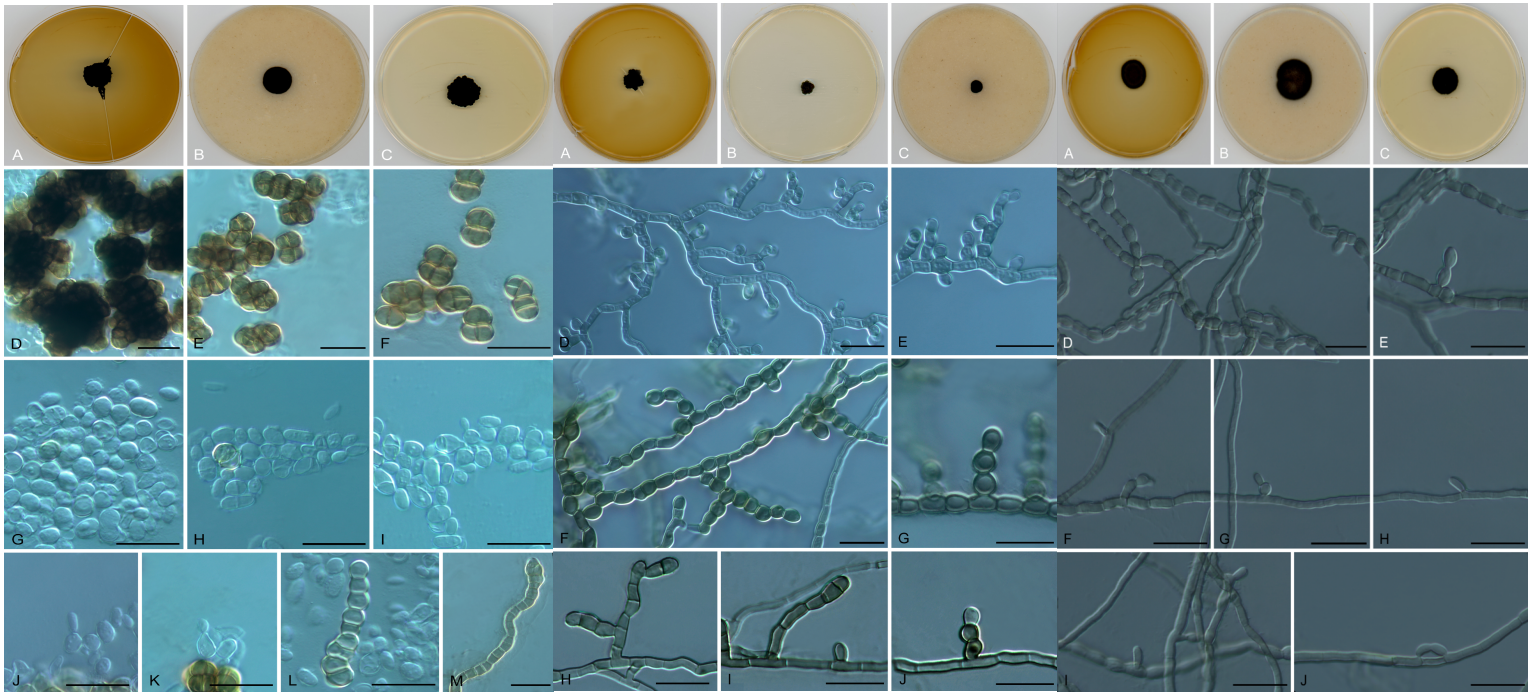


Figure 1. Microscopic morphology of isolated MCF. (A) *Taxawa tesnikishii* JES_119, colonies on MEA, OA, and PDA agars, meristematic cells and budding cells; (B) *Coniosporium tulheliwenetii* JES_115, colonies on MEA, OA, and PDA agars, hyphae with lateral branches of swollen cells reluctantly leading to arthroconidia; (C) *Neophaeococcomyces mojaviensis* JES_112, colonies on MEA, OA, and PDA agars, hyphae producing intercalary swollen cells without conidiation. Microscopic mounts were made in lactic acid and examined using a light microscope. The scale depicted here is 25µm.

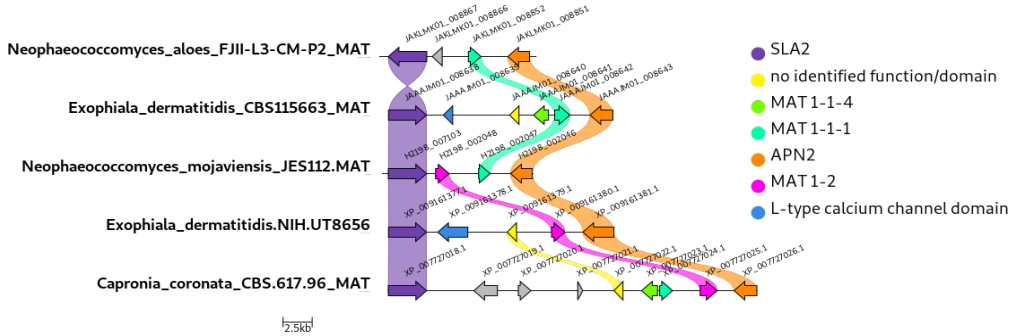


Figure 2. Mating type loci determined for Chaetothyriales JES_112 here called *Neophaeococcomyces mojaviensis.* Other MCFs added to this analysis include *Neophaeococcomyces aloes* FJII-L3-CM-P2*, Exophiala dermatitidis* NIH/UT8656, *Exophiala dermatitidis* CBS 115663, and *Capronia coronata* CBS 617.96. The MAT locus is flanked by two genes, SLA2 (purple) and APN2 (orange). The MAT 1-1 gene, MAT 1-1-4 (green), and MAT 1-1-1 (teal) are observed in *Exophiala dermatitidis* CBS 115663 and *Capronia coronata*, while *Neophaeococcomyces aloes* and *Exophiala dermatitidis* NIH/UT8656 contain MAT 1-2 (pink) as does *Capronia coronata*. Our organism in question, *Neophaeococcomyces mojaviensis* contains the MAT 1-2 and MAT 1-1-1 gene which is confusing as it is not common to lose the MAT 1-1-4. The gene arrow in blue is indicated as an L-type calcium channel domain, while the gene in yellow is an unidentified functional gene or domain.

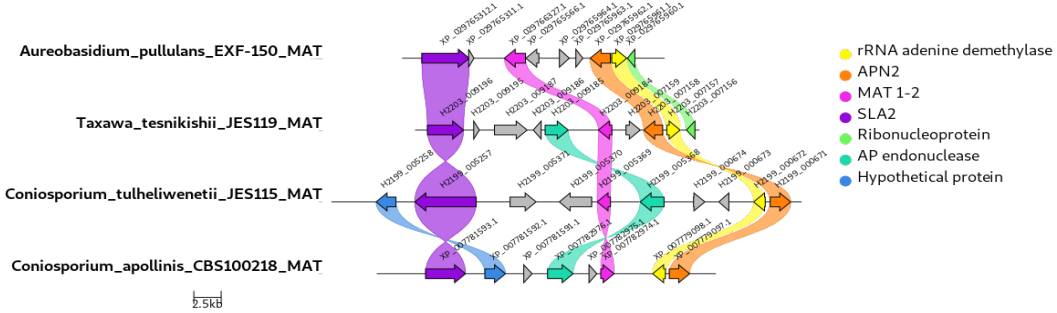


Figure 3. Mating type loci determined for Dothideomycetes JES_115 and JES_119, here called *Coniosporium tulheliwenetii* and *Taxawa tesnikishi.* Other MCFs in this analysis include *Aureobasidium pullulans* EXF-150, and *Coniosporium apollinis CBS* 100218. All species described here only have the MAT 1-2 gene. The MAT locus is flanked by two genes, SLA2 (purple) and APN2 (orange). Dothideomycetes mating type loci look very different from the Chaetothyriales mating type loci. Other genes included in this locus include only a single MAT locus, MAT 1-2 (pink), and rRNA adenine demethylase (yellow). Certain genes like the AP endonuclease (teal) and a hypothetical protein (blue) are not seen in all four genomes.

Secondary metabolite gene cluster predictions

During the annotation step, I processed the genomes through antiSMASH (**Table 3**), to detect Biosynthetic Gene Clusters (BGCs) of secondary metabolites from the annotated genome sequence. Once predicted metabolite of note, beta-lactone was found in all three MCF. A further literature search indicates the importance of beta-lactone in industry as there have been multiple products derived from these chemicals including, antifungals and anti-cancer agents (Robinson et al. 2019). A BLASTP search of the JES_112 beta-lactone gene identified homologs to Acetyl-CoA synthetase + 2-isopropylmalate synthases + bile acid dehydrogenase, zinc-binding alcohol dehydrogenase + acetoacetate-CoA + NAD-dependent alcohol dehydrogenase, and AMP-binding protein + 2-isopropylmalate synthases.

A second predicted metabolite of note was the siderophore gene cluster. These genes are involved in iron transport, is a more well-known metabolite that could provide the organism with virulence and protection by sequestering iron molecules from its surrounding environment (Johnson 2008; Sass et al. 2019) and was primarily detected in JES_112.

Additionally, the BSC search found a fungal-RiPPs (Ribosomal and post-transcriptionally modified proteins) in JES_119. The homolog was best classified as a gamma-glutamyltransferase + cytochrome protein. Fungal-RiPPs have been used as an alternative source of bioactive metabolites and could be of industrial interest (Vogt and Künzler 2019). Additional BGCs were found including Type-1 Polyketide synthases (T1PKS), Type-3

Polyketide synthases, non-ribosomal peptide synthases, non-ribosomal peptide synthase-like, and terpenes.

Table 3. Predicted secondary metabolite gene clusters. The BGCs predicted include multiple copies of PKS and NRPS genes, but siderophores were only recovered in JES_112. There were a higher number of BGCs predicted in JES_112 (27) our Chaetothyriales species, than JES_115 and JES_119 (15) our Dothideomycetes. This table lists the numbers of each. JES_112 has a much higher number (27) of secondary metabolites than the other two (15).

	T1PKS	**Terpene**	**T3PKS**	**NRPS**	**NRPS-like**	**Siderophore**	**Arylpolyene**	**Betalactone**	**Fungal RiPP**
JES_112	6	6	2	4	6	1	1	1	0
JES_115	5	4	0	2	3	0	0	1	0
JES_119	1	4	0	1	7	0	0	1	1

Genome Phylogenetics and Four-Locus Gene Trees

The BUSCO protein-coding marker based phylogenomic reconstruction of the strains **(Figure 4)** places JES_112, JES_115, and JES_119 among the 196 Chaetothyriales and Dothideomycetes representative species. This tree places JES_112 on a branch with *Neophaeococcomyces* although initial assignment was *Knufia* based on ITS similarity alone. The phylogenetic analysis places JES_119 near, but distinct, from the genera Aureobasidium and Delphinella. The species JES_115 is branched with the taxonomically confusing genus *Coniosporium.*

The multi-gene internal-transcribed space (ITS1), large ribosomal subunit (LSU) beta-tubulin, calmodulin tree (**Figure 5)** places the new isolates in context of named species in the known genera (**Table 3**).

Forty-three isolates were used, and the alignment of the sequences were trimmed, and concatenated to run the IQTree program and generate a concise multi-loci gene tree for *Neophaeococcomyces mojaviensis* JES112. Here this strain is placed in the genus *Neophaeococcomyces*, along with *Exophiala dermatitidis* NIH/UT8656, *Exophiala radicis* P2854, *Exophiala oligospoerma* CBS265.49, *Exophiala oligospoerma* CBS127587, and *Exophiala xenobiotica* CBS 11817, *Fonsecaea pedrosoi* CBS 271.37, *Capronia coronata* CBS 617.96, *Cladophialophora immunda* CBS834.96, *Cladophialophora carrionii* CBS260.83, *Cladophialophora yegresii* CBS114405, *Neophaeococcomyces aloes* FIJI-L3-CM-P2, *Neophaeococcomyces catenatus* CBS650.76, *Neophaeococcomyces oklahomaensis* EMSL3313, *Knufia petricola* MA5789, *Knufia separata* CGMCC3.17337, *Knufia mediterranea* CCFEE6211, *Knufia mediterranea* CCFEE6205, *Knufia sp.* MV-2018-MLT-8, *Knufia walvisbayicola* CBS146989, *Knufia tsunedae* FMR10621, *Bradymyces yunnanensis* CGmCC3.17350, *Bradymyces alpinus* CCFEE 5493, *Strelitziana africana* ICMP 21758 *and* ICMP 21760, and *Tuber melanosporum* Mel8 as the outgroup.

The third phylogenetic tree, **Figure 6**, uses the same method described above as in **Figure 5**. 25 strains of Dothideomycetes MCF were included to place JES_115 and JES_119 into their proper clades. MCF strains included a collection of *Coniosporium apollinis* CBS 100218, CBS 352.97, *Aureobasidium mangrovei* IBRCM30265, *Aureobasidium sp.* AWRI4619-CO-2020, *Zalaria obscura* FKI-L7-BK-DRAB2, *Polyplosphaeria fusca* CBS 125.425, *Leptosphaeria maculans* JN3, *Leptosphaeria biglobosa* CA1, *Phaeosphaeria sp.* MPI-PUGE-AT-0046c, *Alternaria solani* NL03003, *Cenococcum geophilum* 1.58, *Venturia pyrina* ICMP11032, *Coniosporium uncinatum* CBS 100219, and CCFEE 6149,

Scorias spongiosa HLSS1, *Polychaeton citri* CBS 116435, *Zymoseptoria tritici* CBS 100329, *Neohortaea acidophila* CBS 113389, *Salinomyces thailandica* CCFEE 6315, *Hortaea werneckii* EXF-2000, *Baudoinia panamericana* UAMH 10762, *Teratosphaeria guachensis* CMW 17545, *Friedmanniomyces simplex* CCFEE 5184, *Friedmanniomyces endolithicus* CCFEE 5311, and *Tuber melanosporum* Mel8 as the outgroup. JES_115 groups near *Coniosporium spp., w*hile JES_119 groups near *Aureobasidium spp.,* which is similar to **Figure 4**, but still seems apart from the other known species.

Taxonomy

Strain JES_112 clusters in the order *Chaetothyriales*, relatively close to the numerous rock-inhabiting species of the genus *Knufia*, and amidst the species of the small genus *Neophaeococcomyces*. Both genera are assigned to the family *Trichomeriaceae*, containing 'sooty molds' with oligotrophic lifestyles colonizing inert surfaces (Chomnunti et al. 2012). The cluster *Neophaeococcomyces,* based on *N. catenatus* as generic type species, as yet contains only four species, which mostly were isolated from low nutrient habitats, such as the bark of *Aloe*, stone, or air. Several species of the genus show budding with melanized cells, and frequently cells cohere in short chains. Often lateral branches of short inflated cells are inserted on hyphae; the cells may be released as arthroconidia which then swell or appear yeast-like, or remain coherent. Strain JES_112 showed inflated cells, but these remained intact, and often no clear distinction could be made between main and lateral hyphae. No conidia were observed.

Strain JES_115 was morphologically similar to *Knufia* species in the presence of lateral chains of swollen cells inserted on main hyphae, the cells eventually being released as conidia, but was phylogenetically remote. It clustered with some species of rock-inhabiting described in *Coniosporium,* in the order *Pleosporales*. Interestingly, the original descriptions, which were made before sequence data were available, classified them all into a single genus. The genus *Coniosporium* is one of the classical genera of meristematic fungi. However, most species were described in the 19th century, including the generic type species *C. olivaceum*, and voucher specimens are lost. Nevertheless, Haridas et al. (2020) introduced the family *Coniosporiaceae* and order *Coniosporales* based on the few recently described rock-inhabiting fungi, *C. apollinis* and *C. uncinatum* (Haridas et al. 2020). It should be noted that the family *Coniosporiaceae* was already used by Nannfeldt for the genus *Coniospora*.

Strain JES_119 is a meristematic black yeast near *Aureobasidium* in the order *Dothideales*. The fungus shows unilateral budding from hyaline cells. The yeast cells subsequently inflate and become cruciately septate and melanized. Similar morphology can be observed in cultural states of e.g. *Sydowia* and *Comminutispora*, and have been described in the asexual genus *Phaeotheca*. Thambugala et al. (2014) erected the family *Aureobasidiaceae* for fungi in this relationship, but also this name existed already, *Aureobasidiaceae* (Thambugala et al. 2014) Ciferri (1958), based on the same species, *Aureobasidium pullulans* (Ciferri 1958). Only a small number of extant species has been sequenced, and the genome trees show an even larger taxon sampling effect, but given our current inability to assign strain JES_119 to any genus, we introduce a new genus here, *Taxawa*.

Discussion

BSCs are an untapped resource of novel microorganisms, using untraditional methods of culturing may be a method to uncover more unavailable species for sequencing. Using a pour plate method for culturing allows the microorganism to avoid complete oxygen saturation (or oxygen toxicity), and instead can grow in low to micro-aerophilic concentrations of oxygen (Stevens 1995). Collecting MCF using this method has proved to be efficient, as fast-growing fungi overtake the surface of plates, while MCF can tolerate lower oxygen concentrations, thus localize to the bottom of plates. Preparing axenic cultures of MCF is then relatively easy to prepare and can also increase the rate of isolation as MCF tend to take two weeks to grow. A hypothesis about the facultative aerobic or microaerophilic nature of MCF could also be seen in the phenotype of these species. Morphological growth on surface of petri plates shows unusual furrowing and clumping growth (JES_119, JES_115), some strains (JES_112) grow underneath the surface of the plate which seem to be MCF adaptations to the high concentration of oxygen present on petri plates.

The high melanin content of MCF complicate efforts to extract high molecular weight (HMW) DNA for sequencing, however adjustments to standard fungal CTAB can dramatically improve the gDNA purity. In the case for these BSC MCF, an increase in phenol:chloroform and chloroform washes allow for melanin to be removed completely from the DNA samples. Another consideration to help concentrate more DNA is to incubate the DNA sample in -20C overnight during the DNA precipitation step. Using long read sequencing technology (Oxford ONT) it is best to have a very high concentration of

clean DNA to process and thus could be the limiting factor for these sequencing procedures. Using hybrid methods of sequencing (Illumina and Nanopore) provides scientists the proper backbone to develop *de-novo* assemblies (Stevens 1995; Khezri et al. 2021; Chen et al. 2020; Saud et al. 2021). It is vital to have as near complete assemblies of species as possible, this provides the scientific community with the best tools to do proper genomic assessments with. For future research endeavors, a second nanopore flowcell would be necessary to help build a better assembly for strain JES_112 (*Neophaeococcomyces mojaviensis*) as the telomere analysis failed on this first round of assembly. The other two strains had enough depth of coverage that allowed for different programs to optimize the assembly and have more closed telomere contigs.

Scientific literature describing novel MCF always note the clonal or asexual nature, describing "missing" sexual states and thus always being in a haploid state. The MAT locus is one of the pivotal regions in the fungal genome that provides more accurate ploidy and sexual states of fungi, and using this method to help describe the three new MCF helped point us in the direction of understanding the sexual states. JES_115 and JES_119 as the two Dothideomycetes representatives indicate heterothallic sexual states having only the MAT1-2 gene, vs homothallic species having representatives of both MAT1-1 and MAT1-2 genes. Better genome assembly would help better describe the JES_112 MAT locus as there is the presence of MAT1-2 but also a portion of the MAT1-1 gene (as it is composed of two genes, MAT1-1-1 and MAT1-1-4). The high melanin content of MCF could prove to be difficult when testing for ploidy in the scientific "gold standard" method (flow cytometry). But future research would most benefit from having this method optimized to

understand not only the new MCF described in this work but also for future MCF considerations. It could just be the sexual states have not yet been described or seen for MCFs and may require a particular environmental stimulus to produce the proper structures to reproduce. There has also been an increase in literature describing clonal MCF having high hybridization rates that create pseudo but stable diploids without the process of recombination (Gostinčar et al. 2022). This was not detected in these three species described in this paper.

MCF scientist Sybren de Hoog helped with the visualization and description of the fungal features for each MCF described here. MCF morphology indicates usual slow growth (about two weeks) seen in most MCF described previously (Selbmann et al. 2014; Chowdhary et al. 2014; Hölker et al. 2004; Zhao et al. 2010; Kejžar et al. 2013). JES_119 as it develops over time increases its melanin production and develops from a bright orange-yellow color to a more deeply pigmented black. The budding cells are unilateral and are uncommon in other described MCF. Dr. de Hoog describes JES_119 features as uncommon in other described MCF and believes the phenotype along with phylogenetic descriptions of this species is in fact a previously undescribed genus and species in Class Dothideomycetes as *Taxawa tesnikishii.* The species JES_115 seems very common to Dr. de Hoog, but due to its complicated taxonomy history, *Coniosporium* will be described definitely here in this thesis work. Future work on this publication will require confirmation with Mycobank along with other tests to place JES_115 definitively with *Coniosporium* and to indicate a new species. The species JES_112 is one of the few described *Neophaeococcomyces* available anywhere and thus an important contribution to the

scientific community. All species described here will provide the scientific community with identified species genomic data for future work.

Finding MCF from extreme environments such as BSCs may point to harbored secondary metabolites either absorbed from surrounding species through horizontal gene transfer (HGT) or adapted from time. MCF are known to use secondary metabolites to help survive such extremes and are thus an untapped resource for the identification of novel metabolites. JES_112 requires another attempt in long-read sequencing to produce a more complete assembly, but preliminary secondary metabolite analysis has indicated a propensity for bacterial secondary metabolites. The nature of this species is still being understood, as other work on *Neophaeococcomyces* has shown an obligate bacterial hitchhiker (Carr 2022). JES_115 and JES_119 both contain about 15 secondary metabolites detected by antiSMASH, which is a considerable number when compared to other fungi (Rateb and Ebel 2011). The roles of secondary metabolites in modern human industry applications are vast (Demain 2014) and thus would be important to test and compare across all MCF.

Genome phylogenetics have come a long way and have become more accurate using optimized models in using hundreds of genes (Jeffroy et al. 2006; Rokas and Chatzimanolis 2008; Li et al. 2020; James et al. 2020; Li et al. 2021). The opposite can also be said about using minimum but stable gene loci to do a similar assessment of fungal species (Gatesy et al. 2007; Feau et al. 2011; Walker et al. 2012; Shen et al. 2017). This is a hotly contested area in fungal species determination and thus may result in an adjustment for the final version of this paper.

Figure 4. Phylogenetic tree of Dothideomycetes and Chaetothyriales MCF and non-MCF fungi including 3 strains described in this paper. The Phylogenetic tree was generated using 196 fungal protein sequences using protein and nucleotide BUSCO markers, and a combination of ASTRAL and RaXML programs. This tree shows the Chaetothyriales and Dothideomycetes and places all of our three species within the clade. Our JES_112 or *Neophaeococcomyces mojaviensis* is placed near the Chaetothyriales clade, specifically near *Knufia* and *Arthrocladium spp.* While JES_115 or *Coniosporium tulheliwenetii* and JES_119 *Taxawa tesnikishii* are both found in the Dothideomyces clade but in different clades. JES_115 groups with another *Coniosporium spp.*, while JES_119 groups near but still apart from *Delphinella* and *Aureobasidium spp.*.

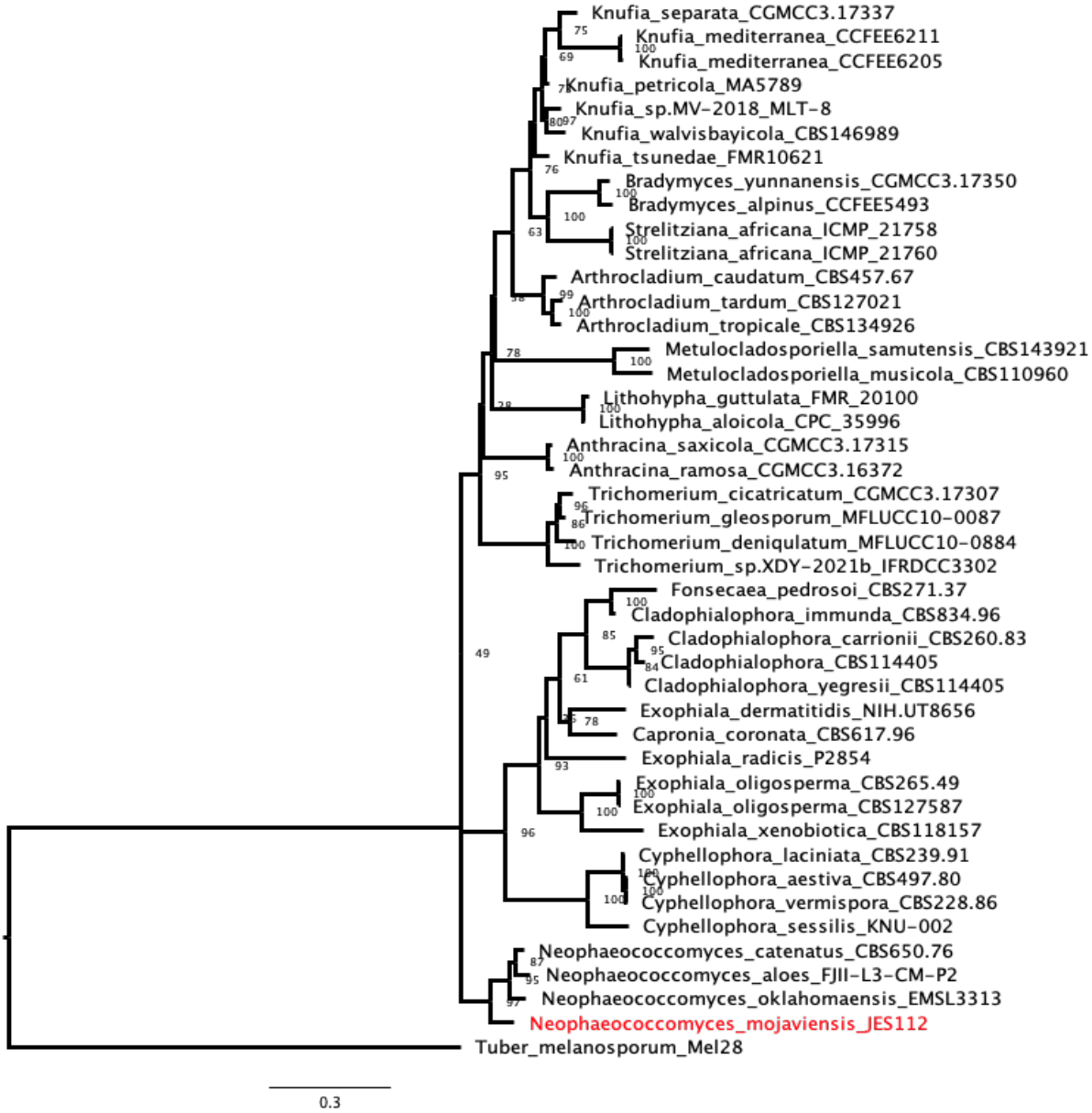


Figure 5. Chaetothyriales multi-locus gene tree. Multi-loci phylogenetic tree using 4 highly-conserved genes including, beta-tubulin, calmodulin, internal-transcribed sequence (ITS), and the large subunit of the Ribosome (LSU). Seven representative strains with loci were used, which were aligned, trimmed, and concatenated to run the IQTree program and generate a concise multi-loci gene tree for *Neophaeococcomyces mojaviensis* JES112. Here this strain is placed in the *Neophaeococcomyces* genera.

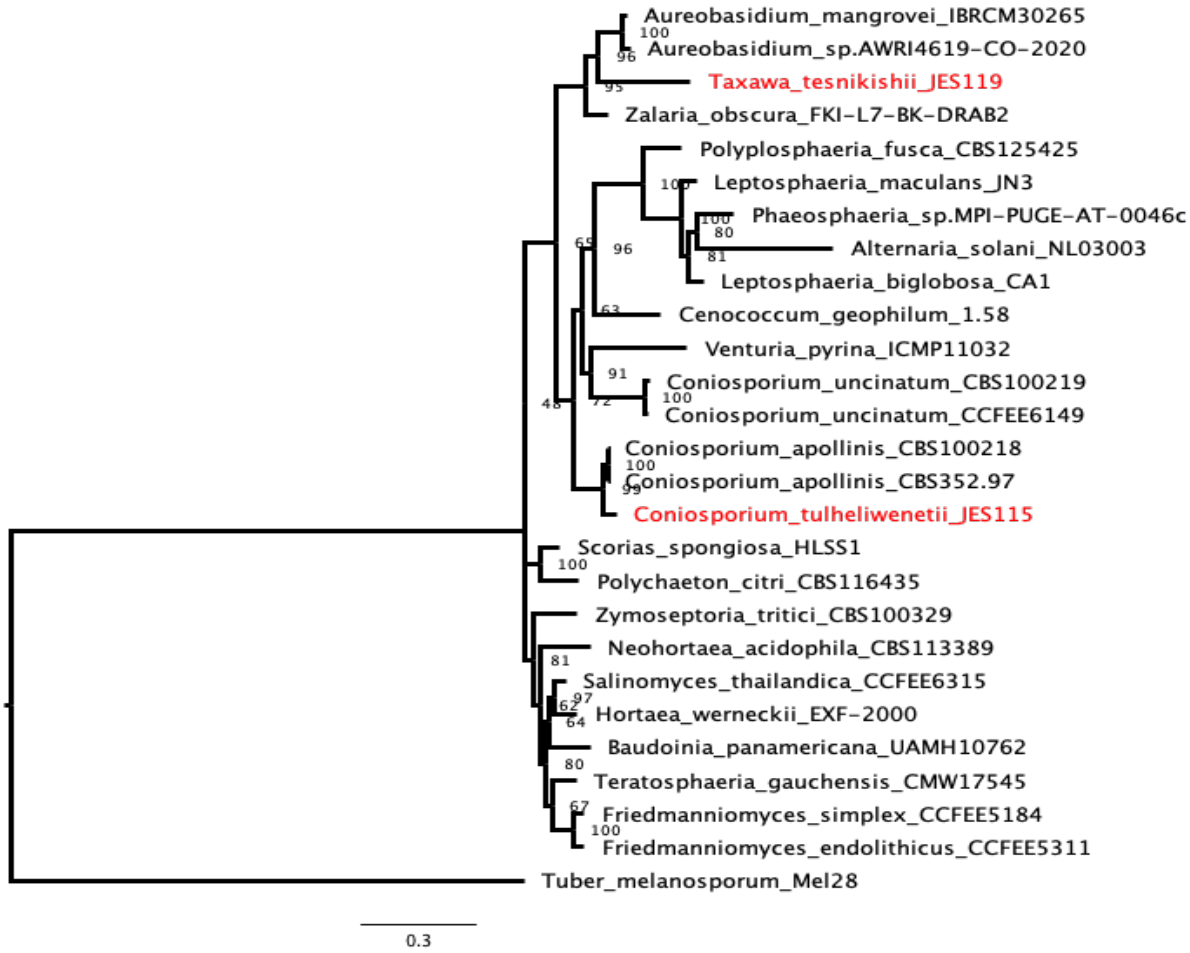


Figure 6. Dothideomycetes multi-locus gene tree. Multi-loci phylogenetic tree using 4 highly-conserved genes including, beta-tubulin, calmodulin, internal-transcribed sequence (ITS), and the large subunit of the Ribosome (LSU). Twenty-five representative strains with loci were used, which were aligned, trimmed, and concatenated to run the IQTree program and generate a concise multi-loci gene tree for JES_115 and JES_119 or *Coniosporium tulheliwenetii* and *Taxawa tesnikishii* respectively. Here strain JES_115 is grouping with other *Coniosporium spp.*, while JES_119 is grouping with but distinctly apart from *Aureobasidium spp.*.

Table 4. Gene Tree Nucleotide Accession numbers. Collected accession numbers of alternative strains to build Figures 5 and 6. The remaining empty spaces indicate the need to pull the sequences using BLAST. Once collected each gene populated a gene FASTA file, aligned, clipped and IQTREE2, a tree-building program, was run to determine the gene models for each gene. Once determined, the gene alignment lengths were calculated and partitioning scripts were described and used to build a larger tree, again using IQTREE2.

Species	Beta-Tubulin	Calmodulin	Internal-Transcribed Spacer	Large Subunit (Ribosomal)
Neophaeococcomyces mojaviensis JES112	-	-	-	-
Taxawa tesnikishii JES119	-	-	-	-
Alternaria solani NL03003	MK388240.1	MH243808.1	NR_172408.1	ON416546.1
Anthracina ramosa CGMCC3.16372	KP226556.1	-	NR_172152.1	KP174923.1
Anthracina saxicola CGMCC3.17315	KP226554.1	-	NR_172151.1	KP174921.1
Arthrocladium caudatum CBS457.67	LT558710.1	-	NR_145398.1	MH870732.1
Arthrocladium tardum CBS 127021	-	-	NR_154706.1	NG_057089.1
Arthrocladium tropicale CBS 134926	KX822406.1	-	NR_154724.1	NG_057119.1
Aureobasidium mangrovei IBRCM30265	XM_029904852.1	XM_041025809.1	NR_174637.1	NG_078639.1
*Aureobasidium sp.*AWRI4619-CO-2020	XM_029904852.1	XM_041025809.1	MK256457.1	MK968773.1
Baudoinia panamericana UAMH10762	XM_007673885.1	XM_007675264.1	NR_153615.1	KT186497.1

Bradymyces alpinus CCFEE 5493	LN589970.1	-	NR_132844.1	NG_058641.1
Bradymyces yunnanensis CGMCC 3.17350	KP226545.1	-	KP174865.1	KP174926.1
Capronia coronata CBS 617.96	XM_007728982.1	XM_007721556.1	NR_154745.1	NG_067831.1
Cenococcum geophilum 1.58	-	-	KC967410.1	MN817694.1
Cladophialophora carrionii CBS 260.83	KF928582.1	XM_008724000.1	MH861582.1	MH873312.1
Cladophialophora_yegresii CBS 114405	EU137209.1	-	NR_111284.1	KX822323.1
Cladophialophora immunda CBS 834.96	EU137203.1	-	NR_111283.1	MH874242.1
Coniosporium apollinis CBS 100218	-	XM_007785495.1	NR_159787.1	GU250898.1
Coniosporium uncinatum CBS 100219	-	-	NR_145343.1	NG_058806.1
Coniosporium uncinatum CCFEE 6149	-	-	MZ573424.1	MZ573424.1
Cyphellophora aestiva CBS 239.91	MW297547.1	-	MH861291.1	MH873056.1
Cyphellophora laciniata CBS 239.91	KF928604.1	-	NR_121335.1	MH873933.1
Cyphellophora sessilis KNU-002	KX431943.1	-	KX431941.1	KX431942.1
Cyphellophora vermispora CBS	KC455227.1	-	KC455244.1	KC455257.1

228.86				
Exophiala dermatitidis NIH.UT8656	-	XM_009160450.1	JH226131.1	ON543213.1
Exophiala oligosperma CBS 265.49	EF551507.1	XM_016404560.1	MH856519.1	MH868049.1
Exophiala oligosperma CBS 127587	KF928570.1	-	MH864631.1	MH876068.1
Exophiala radicis P2854	KT723463.1	-	NR_158397.1	NG_069319.1
Exophiala xenobiotica CBS 118157	NW_013562490.1	XM_013454656.1	OP132917.1	MH875251.1
Fonsecaea pedrosoi CBS 271.37	XM_013430809.1	XM_013426864.1	MT023597.1	KF155185.1
Friedmanniomyces endolithicus CCFEE 5311	-	-	NR_144917.1	GU250367.1
Friedmanniomyces simplex CCFEE 5184	-	-	NR_155079.1	GU250368.1
Hortaea werneckii EXF-2000	-	-	NR_145338.1	NG_057773.1
Knufia petricola MA5789	-	-	NR_111871.1	MT126669.1
Knufia mediterranea CCFEE 6211	-	-	KP791793.1	KR781080.1
Knufia mediterranea CCFEE 6205	-	-	KP791794.1	KR781081.1
Knufia separata CGMCC 3.17337	KP226564.1	-	NR_172233.1	KP174931.1

Knufia tsunedae FMR10621	-	-	NR_132842.1	HG003672.1
Knufia walvisbayicola CBS 146989	MZ078266.1	-	NR_173046.1	NG_076737.1
Knufia sp. MV-2018 MLT-8	-	-	MT636977.1	MT636937.1
Leptosphaeria biglobosa CA1	AY749005.1	-	NR_111620.1	JX681093.1
Leptosphaeria maculans JN3	AY749012.1	XM_003840719.1	M96384.1	JF740306.1
Lithohypha aloicola CPC35996	MN556837.1	-	NR_166313.1	NG_068316.1
Lithohypha guttulata FMR 20100	OX346320.1	-	OX346376.1	OX346377.1
Metulocladosporiella musicola CBS 110960	-	MG934537.1	MH862870.1	NG_067447.1
Metulocladosporiella samutensis CBS 143921	-	MG934540.1	NR_161056.1	MG934453.1
Neohortaea acidophila CBS 113389	XM_033733287.1	XM_033733930.1	NR_169648.1	OL739260.1
Neophaeococcomyces aloes FJII-L3-CM-P2	-	-	NR_132069.1	KF777234.1
Neophaeococcomyces catenatus CBS 650.76	-	-	AF050277.1	MH872793.1
Neophaeococcomyces oklahomaensis EMSL3313	-	-	MW798187.1	MW798182.1
*Phaeosphaeria sp.*MPI-PUGE-AT-0046c	-	-	NR_155675.1	NG_068582.1

Polychaeton citri CBS 116435	-	-	GU214649.1	OM238161.1
Polyplosphaeria fusca CBS 125425	-	-	NR_119397.1	MH875147.1
Salinomyces thailandica CCFEE 6315	-	-	NR_171710.1	MH875167.1
Scorias spongiosa HLSS1	-	-	NR_154422.1	MH866910.1
Strelitziana africana ICMP 21758	-	-	MK210538.1	MK210499.1
Strelitziana africana ICMP 21760	-	-	MK210540.1	MK210501.1
Teratosphaeria gauchensis CMW 17545	-	KF902726.1	NR_144976.1	EU019295.2
Trichomerium cicatricatum CGMCC3.17307	KP226540.1	-	NR_172232.1	KP174947.1
Trichomerium deniqulatum MFLUCC10-0884	-	-	NR_132965.1	NG_059479.1
Trichomerium gleosproum MFLUCC10-0087	-	-	JX313656.1	JX313662.1
Trichomerium sp. XDY-2021a-IFRDCC3302	-	-	MN901488.1	MZ822131.1
Tuber melanosporum	MZ458420.1	XM_002841184.1	GQ917052.1	JF807486.1
Venturia pyrina ICMP11032	EU853839.1	-	NR_169905.1	AY849967.1
Zalaria obscura FKI-L7-BK-	-	-	NR_153466.1	MT103097.1

DRAB2				
Zymoseptoria tritici CBS 100329	**XM_003856727.1**	**KF254299.1**	**KP852517.1**	**MK788267.1**

Conclusion

Biological soil crusts are an untapped resource for novel microorganisms as evidenced by our work. The MCF we collected are incredibly slow growing and could be easily missed during culturing attempts. The method described here (and detailed in **Appendix 1**) improved the chances for isolation of these species. MCFs have been isolated from different types of extreme environments including BSC's.

An interesting consideration for species JES_112 or *Neophaeococcomyces mojaviensis* is through personal communications with another MCF scientist, Erin Carr, who also isolated a species of *Neophaococcomyces*, named "Crusty" from a biological soil crust. Their species was associated with a pink *Methylobacterium* symbiont, which she calls "Light Pinky" and "Dark Pinky". Her assessment of this symbiosis suggests the role of the symbiont is involved in aerobic anoxygenic photosynthesis and auxin production (Carr 2022). Through our multiple culturing attempts, there was no indication of a pink symbiont, but the possible bacteria signal within our secondary metabolite assessment could be attributed to a contamination from this bacterium. JES_112 will be the second known genome of the genus *Neophaeococcomyces*.

Species JES_119 or *Taxawa tesnikishi* is the first isolated of this phenotype. Speaking to Sybren de Hoog, his assessment of this species was closer to *Aureobasidium*, yet phenotypically is very different from identified *Aureobasidium* species. Using fungal phenotype and phylogenomics, we were able to determine the microorganism is different from known *Aureobasidium* species and may be of a new genus, hence the naming *Taxawa*

as its genus name. *Taxawa tesnikishi* is the first known genus and species, now available in Genbank.

Species JES_115 or *Coniosporium tulheliwenetii* was also collected from the BSC's from Boyd Deep Canyon and was thus honored with a Cahuilla term used to describe BSC phenotypic features seen on the soil. The genus Coniosporium has historically been confusing, as the earliest strains were collected in the late 1800s. Initial collections of *Coniosporium* were placed in Chaetothyriales, while later similar MCF phenotypes were also grouped with *Coniosporium*, to be later considered as Dothideomycetes instead. A careful reorganization and assessment of the genus is desperately needed.

As scientists, we should make more effort in preserving indigenous cultural regions. One way to bring awareness is to highlight the beauty of language that is not normally used in colonization, like Latin. The MCF we isolated and named take particular Cahuilla terms, knowledge, and mythos, and apply them to the features we see with our fungi. The newly minted genus *Taxawa* and species *tesnikish* takes the knowledge and mythos of the Cahuilla demi-god which inhabited the land and bring it to a new appreciation. While the term tulheliwenetii describes BSC features found on the ground. As the Cahuilla language does not contain terms for microorganisms' phenotypic features and mythos is a more suitable application for our naming scheme.

The advancement of technology from when we initially sequenced these three strains, to now has increased exponentially and we were lucky enough to be able to use Nanopore sequencing to prepare a higher coverage genome. Amplicon sequencing has provided

scientists with a better understanding of the breadth of fungi flora that is still unknown, taking the time to culture, sequence, identify, and to even name these species can provide future scientists with names instead of generic fungal species identities.

Acknowledgments

This work was performed (in part) at the University of California Natural Reserve System (Boyd Deep Canyon Desert Research Center) Reserve DOI: (doi:10.21973/N3V66D) (UCNRS Information Manager 1965) and at Sheephole Valley Wilderness on Bureau Land Mangament (BLM) territory, near Joshua Tree National Park.

References:

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bell AA, Wheeler MH. 1986. Biosynthesis and Functions of Fungal Melanins. Annu Rev Phytopathol. 24(1):411–451.

Belnap J, Harper KT. 1995. Influence of cryptobiotic soil crusts on elemental content of tissue of two desert seed plants. Arid Soil Research and Rehabilitation. 9(2):107–115.

Beraldi-Campesi H, Retallack GJ. 2016b. Terrestrial Ecosystems in the Precambrian. In: Weber B, Büdel B, Belnap J, editors. Biological Soil Crusts: An Organizing Principle in Drylands. Cham: Springer International Publishing. p. 37–54.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49(W1):W29–W35.

Brown J, Boyd J. 1922. History of San Bernardino and Riverside Counties: With Selected Biography of Actors and Witnesses of the Period of Growth and Achievement. Western Historical Association.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP : eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics and Bioinformatics. 2(2). doi:10.1093/nargab/lqaa026. http://dx.doi.org/10.1093/nargab/lqaa026.

Bushnell B. 2014. BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States) Report No.: LBNL-7065E. https://www.osti.gov/biblio/1241166.

Carr EC. 2022. Discovering Novel Polyextremotolerant Fungi, and Determining Their Ecological Role Within the Biological Soil Crust Consortium.

Casadevall A, Rosas AL, Nosanchuk JD. 2000. Melanin and virulence in *Cryptococcus neoformans*. Current Opinion in Microbiology. 3(4):354–358. doi:10.1016/s1369-5274(00)00103-x.

Casadevall A, Steenbergen JN, Nosanchuk JD. 2003. “Ready made” virulence and “dual use” virulence factors in pathogenic environmental fungi — the *Cryptococcus neoformans* paradigm. Current Opinion in Microbiology. 6(4):332–337. doi:10.1016/s1369-5274(03)00082-1.

Chomnunti P, Bhat DJ, Jones EBG, Chukeatirote E, Bahkali AH, Hyde KD. 2012. Trichomeriaceae, a new sooty mould family of Chaetothyriales. Fungal Divers. 56(1):63–76.

Chowdhary A, Perfect J, de Hoog GS. 2014. Black Molds and Melanized Yeasts Pathogenic to Humans. Cold Spring Harb Perspect Med. 5(8):a019570.

Ciferri R. 1958. Affinity between fungi of the verrucous dermatitis and sooty-molds. Mycopathol Mycol Appl. 9(3):197–200.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 25(11):1422–1423.

Dadachova E, Bryan RA, Howell RC, Schweitzer AD, Aisen P, Nosanchuk JD, Casadevall A. 2008. The radioprotective properties of fungal melanin are a function of its chemical composition, stable radical presence and spatial arrangement. Pigment Cell Melanoma Res. 21(2):192–199.

Dadachova E, Casadevall A. 2008. Ionizing radiation: how fungi cope, adapt, and exploit with the help of melanin. Curr Opin Microbiol. 11(6):525–531.

Demain AL. 2014. Valuable Secondary Metabolites from Fungi. In: Martín J-F, García-Estrada C, Zeilinger S, editors. Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites. New York, NY: Springer New York. p. 1–15.

Evans RD, Ehleringer JR. 1993. A break in the nitrogen cycle in aridlands? Evidence from δp15N of soils. Oecologia. 94(3):314–317.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42(Database issue):D222–30.

Gilchrist CLM, Chooi Y-H. 2021 Jan 18. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics. doi:10.1093/bioinformatics/btab007.

Gómez BL, Nosanchuk JD. 2003. Melanin and fungi. Curr Opin Infect Dis. 16(2):91–96.

Gorbushina AA, Whitehead K, Dornieden T, Niesse A, Schulte A, Hedges JI. 2003. Black fungal colonies as units of survival: hyphal mycosporines synthesized by rock-dwelling microcolonial fungi. Can J Bot. 81(2):131–138.

Gostinčar C, Sun X, Černoša A, Fang C, Gunde-Cimerman N, Song Z. 2022. Clonality, inbreeding, and hybridization in two extremotolerant black yeasts. Gigascience. 11. doi:10.1093/gigascience/giac095.

Green TGA, Allan Green TG, Proctor MCF. 2016. Physiology of Photosynthetic Organisms Within Biological Soil Crusts: Their Adaptation, Flexibility, and Plasticity. Biological Soil Crusts: An Organizing Principle in Drylands.:347–381. doi:10.1007/978-3-319-30214-0_18.

Green TGA, Broady PA. 2001. Biological soil crusts of Antarctica. In: Ecological Studies. Berlin, Heidelberg: Springer Berlin Heidelberg. (Ecological studies: analysis and synthesis. Berlin, Heidelberg, New York NY). p. 133–139.

Gümral R, Tümgör A, Saraçlı MA, Yıldıran ŞT, Ilkit M, de Hoog GS. 2014. Black yeast diversity on creosoted railway sleepers changes with ambient climatic conditions. Microb Ecol. 68(4):699–707.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9(1):R7.

Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, et al. 2020. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. Stud Mycol. 96:141–153.

Hölker U, Bend J, Pracht R, Tetsch L, Müller T, Höfer M, de Hoog GS. 2004. *Hortaea acidophila*, a new acid-tolerant black yeast from lignite. Antonie Van Leeuwenhoek. 86(4):287–294.

de Hoog GS, Hermanides-Nijhof EJ. 1977. The Black Yeasts and Allied Hyphomycetes, Studies in Mycology. Baarn, The Netherlands: Centraalbureau voor Schimmelcultures.

de Hoog GS, Matos T, Sudhadham M, Luijsterburg KF, Haase G. 2005. Intestinal prevalence of the neurotropic black yeast *Exophiala (Wangiella) dermatitidis* in healthy and impaired individuals. Mycoses. 48(2):142–145.

Hoog GS de, Queiroz-Telles F, Haase G, Fernandez-Zeppenfeldt G, Angelis DA, H. G. Gerrits van den Ende A, Matos T, Peltroche-Llacsahuanga H, Pizzirani-Kleiner AA, Rainer J, et al. 2000a. Black fungi: clinical and pathogenic approaches. Med Mycol. 38(sup1):243–250.

Hooper L. 1920. The Cahuilla Indians. University of California Press.

Housman DC, Yeager CM, Darby BJ, Sanford RL, Kuske CR, Neher DA, Belnap J. 2007. Heterogeneity of soil nutrients and subsurface biota in a dryland ecosystem. Soil Biol Biochem. 39(8):2138–2149.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

James TY, Stajich JE, Hittinger CT, Rokas A. 2020. Toward a Fully Resolved Fungal Tree of Life. Annu Rev Microbiol. 74:291–313.

Johnson L. 2008. Iron and siderophores in fungal–host interactions. Mycol Res. 112(2):170–183.

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

Kejžar A, Gobec S, Plemenitaš A, Lenassi M. 2013. Melanin is crucial for growth of the black yeast *Hortaea werneckii* in its natural hypersaline environment. Fungal Biology. 117(5):368–379. doi:10.1016/j.funbio.2013.03.006. http://dx.doi.org/10.1016/j.funbio.2013.03.006.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics. 5:59.

Kroeber AL. 1925. Handbook of the Indians of California. U.S. Government Printing Office.

Kurbessoian T. 2022. tania-k/Chapter2.3MCF: V1. https://zenodo.org/record/7259500.

Lal R. 2004b. Carbon sequestration in dryland ecosystems. Environ Manage. 33(4):528–544.

Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol. 14:82.

Little AEF, Currie CR. 2008. Black yeast symbionts compromise the efficiency of antibiotic defenses in fungus-growing ants. Ecology. 89(5):1216–1222.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44(W1):W54–7.

Maestre FT, Soliveres S, Gotelli NJ, Quero JL, Berdugo M. 2012. Response to Comment on “Plant Species Richness and Ecosystem Multifunctionality in Global Drylands.” Science. 337(6091):155–155. doi:10.1126/science.1220620. http://dx.doi.org/10.1126/science.1220620.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 20(16):2878–2879.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 38(10):4647–4654.

Mehdiabadi NJ, Schultz TR. Natural history and phylogeny of the fungus-farming ants (Hymenoptera: Formicidae: Myrmicinae: Attini). [accessed 2022 Mar 29].

Metin B, Döğen A, Yıldırım E, de Hoog GS, Heitman J, Ilkit M. 2019. Mating type (MAT) locus and possible sexuality of the opportunistic pathogen *Exophiala dermatitidis*. Fungal Genetics and Biology. 124:29–38. doi:10.1016/j.fgb.2018.12.011.

Nosanchuk JD, Stark RE, Casadevall A. 2015. Fungal melanin: What do we know about structure? Front Microbiol. 6:1463.

Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation. https://zenodo.org/record/4054262.

Pietrasiak N. 2005. Field guide to classify biological soil crusts for ecological site evaluation. NRCS.

Pietrasiak N, Regus JU, Johansen JR, Lam D, Sachs JL, Santiago LS. 2013. Biological soil crust community types differ in key ecological functions. Soil Biol Biochem. 65:168–171.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204.

Rateb ME, Ebel R. 2011. Secondary metabolites of fungi from marine habitats. Nat Prod Rep. 28(2):290–344.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46(D1):D624–D632.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3(2):217–223.

Robinson SL, Christenson JK, Wackett LP. 2019. Biosynthesis and chemical diversity of β-lactone natural products. Nat Prod Rep. 36(3):458–475.

Rozenstein O, Karnieli A. 2015. Identification and characterization of Biological Soil Crusts in a sand dune desert environment across Israel–Egypt border using LWIR emittance spectroscopy. J Arid Environ. 112:75–86.

Sass G, Ansari SR, Dietl A-M, Déziel E, Haas H, Stevens DA. 2019. Intermicrobial interaction: *Aspergillus fumigatus* siderophores protect against competition by *Pseudomonas aeruginosa*. PLoS One. 14(5):e0216085.

Saunte DM, Tarazooie B, Arendrup MC, de Hoog GS. 2012a. Black yeast-like fungi in skin and nail: it probably matters. Mycoses. 55(2):161–167.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Ecol Evol. 1(5):126.

Siletti CE, Zeiner CA, Bhatnagar JM. 2017. Distributions of fungal melanin across species and soils. Soil Biol Biochem. 113:285–293.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:31.

SMIT A. F. A. 2004. Repeat-Masker Open-3.0. http://www.repeatmasker.org.

Sofi MY, Shafi A, Masoodi KZ. 2022. NCBI BLAST. Bioinformatics for Everyone.:95–102. doi:10.1016/b978-0-323-91128-3.00021-5.

Staley JT, Palmer F, Adams JB. 1982. Microcolonial fungi: common inhabitants on desert rocks? Science. 215(4536):1093–1095.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34(Web Server issue):W435–9.

Steenbergen JN, Casadevall A. 2003. The origin and maintenance of virulence for the human pathogenic fungus Cryptococcus neoformans. Microbes Infect. 5(7):667–675.

Stevens TO. 1995. Optimization of media for enumeration and isolation of aerobic heterotrophic bacteria from the deep terrestrial subsurface. J Microbiol Methods. 21(3):293–303.

Taborda CP, da Silva MB, Nosanchuk JD, Travassos LR. 2008b. Melanin as a virulence factor of Paracoccidioides brasiliensis and other dimorphic pathogenic fungi: a minireview. Mycopathologia. 165(4-5):331–339.

Teixeira MM, Moreno LF, Stielow BJ, Muszewska A, Hainaut M, Gonzaga L, Abouelleil A, Patané JSL, Priest M, Souza R, et al. 2017. Exploring the genomic diversity of black yeasts and relatives (Chaetothyriales, Ascomycota). Stud Mycol. 86:1–28.

Thambugala KM, Ariyawansa HA, Li Y-M, Boonmee S, Hongsanan S, Tian Q, Singtripop C, Bhat DJ, Camporesi E, Jayawardena R, et al. 2014. Dothideales. Fungal Divers. 68(1):105–158.

Thanh VN, Hien DD. 2019. *Moniliella floricola* sp. nov., a species of black yeast isolated from flowers. International Journal of Systematic and Evolutionary Microbiology. 69(1):87–92. doi:10.1099/ijsem.0.003099. http://dx.doi.org/10.1099/ijsem.0.003099.

UCNRS Information Manager. 1965. Boyd deep canyon desert research center. doi:10.21973/N3V66D.

Vicente VA, Attili-Angelis D, Pie MR, Queiroz-Telles F, Cruz LM, Najafzadeh MJ, de Hoog GS, Zhao J, Pizzirani-Kleiner A. 2008. Environmental isolation of black yeast-like fungi involved in human infection. Stud Mycol. 61:137–144.

Vogt E, Künzler M. 2019. Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics. Appl Microbiol Biotechnol. 103(14):5567–5581.

Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, et al. 2020. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol Biol Evol. 37(2):599–603.

Warren SD, Clair LL St, Stark LR, Lewis LA, Pombubpa N, Kurbessoian T, Stajich JE, Aanderud ZT. 2019. Reproduction and dispersal of biological soil crust organisms. Front Ecol Evol. 7. doi:10.3389/fevo.2019.00344.

Weber B, Bowker M, Zhang Y, Belnap J. 2016. Natural Recovery of Biological Soil Crusts After Disturbance. In: Weber B, Büdel B, Belnap J, editors. Biological Soil Crusts: An Organizing Principle in Drylands. Cham: Springer International Publishing. p. 479–498.

Wheeler MH. 1983. Comparisons of fungal melanin biosynthesis in ascomycetous, imperfect and basidiomycetous fungi. Trans Br Mycol Soc. 81(1):29–36.

White PM. 2001. California Indians and Their Reservations: An Online Dictionary. San Diego State University Library.

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 20(1):129.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. Ggtree : An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 8(1):28–36.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46(W1):W95–W101.

Zhao J, Zeng J, de Hoog GS, Attili-Angelis D, Prenafeta-Boldú FX. 2010. Isolation and identification of black yeasts by enrichment on atmospheres of monoaromatic hydrocarbons. Microb Ecol. 60(1):149–156.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics. 29(21):2669–2677. doi:10.1093/bioinformatics/btt476. http://dx.doi.org/10.1093/bioinformatics/btt476.

Molecular Dating and Whole-Genome Duplication of the Micro-Colonial Fungi Genus *Friedmanniomyces*

Abstract

Micro-colonial fungi are colonizers of inhospitable ecological habitats such as Continental Antarctica, including the McMurdo Dry Valleys. The valleys are recognized as the coldest arid deserts on Earth. In the tundras of Antarctica, microorganisms find refuge inside rocks which harbor crypto-endolithic (rock-dwelling) communities. The genus *Friedmanniomyces* is exclusively associated with these ecosystems, two species include: *Friedmanniomyces endolithicus CCFEE5311* and *Friedmanniomyces simplex CCFEE5184*. In this work, we investigate genome evolution to understand important genetic factors of this extraordinarily specialized genus, showing that the genus diverged in the late Cretaceous (ca. 65 million years ago), when Antarctica reached its present geographic location. Since ploidy variations are a recurrent trait of stress-adapted microorganisms, we assessed allele frequencies, K-mer counts, CNV analysis, Gaussian statistics, and mating type locus determination of 25 strains of *Friedmanniomyces endolithicus* collected from Antarctica over a forty-year time span. Here, we report evidence on three different ploidy results, haploid, diploid and triploid with possible tetraploidy in a small population of *Friedmanniomyces endolithicus*. To clarify the ploidy analysis attempted here, a deeper and longer sequencing of these micro-colonial fungi are necessary to close genome gaps, indicate proper gene counts and assess other duplication markers like transposons. Studies on the genome evolution of this intriguing group of fungi may enable understanding of important genetic factors that govern their success in

the extremes, providing critical insights for further genomics analysis, including niche adaptation, fitness and evolution.

Introduction

History of Antarctica

Antarctic geological history has geospatially and biologically changed in the last 200 million years (Mya), as it was originally located at a much higher latitude than where it is found now. The continental Antarctic crust joined the South American, African, Indian, and Australian regions, generating the landmass of Gondwana, the southern part of the supercontinent Pangea (Crame 1992; Askin 2013; Maritati et al. 2020). Continental drift split Gondwana and moved Antarctica toward the South Pole (Fitzgerald 2002). The formation of the sea separating Antarctica and Australia and the splitting of South America from the Antarctic Peninsula was crucial for establishing the cold "Circumpolar Current" flowing clockwise around the continent and keeping warm waters away from the coast (Lyle et al. 2007). (**Figure 1**).

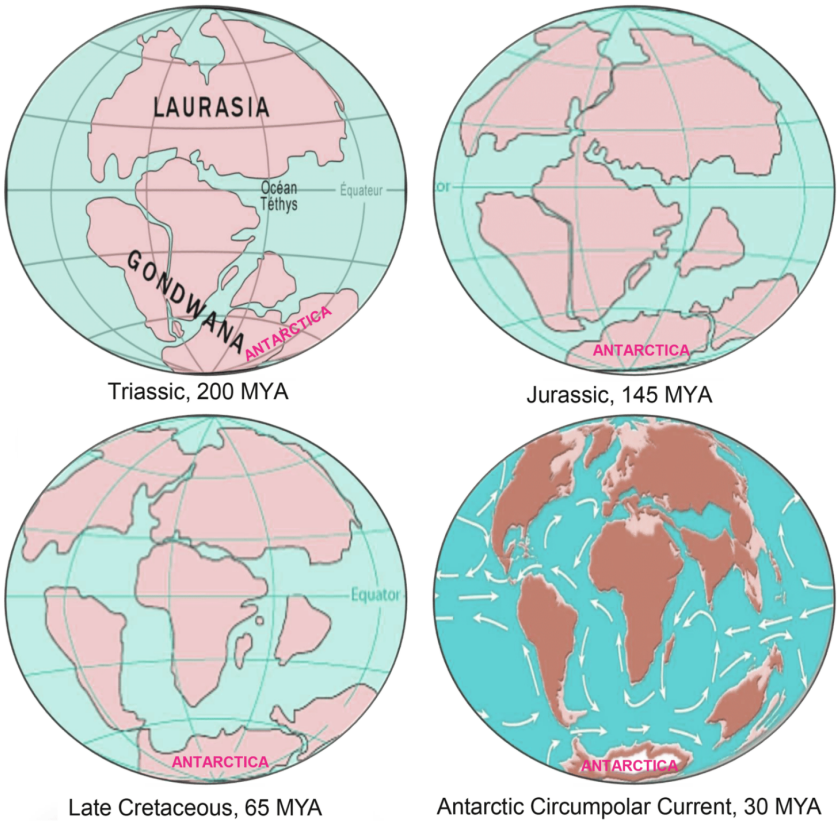


Figure 1. During the Triassic period, Gondwana was found closer to the remaining continents. As time went by, Antarctica slowly detached and migrated toward the south pole of the planet. The "Circumpolar currents" flow in the seas separating Antarctica, Australia, and South America. These currents were vital to producing the freezing Antarctic that we know today.

As temperatures dropped and ice on Antarctica spread, the freezing process reinforced itself due to increased sunlight reflection to space (higher albedo) and enhanced the cooling of the continent: the ice sheets continued to grow, the white surface became wider, and as a result, the continent became cooler (Perkins 2019). Combined, the effects pushed Antarctica into a deeper freeze and culminated in the extinction of higher life forms (Albanese et al. 2020). Glacier movements and strong winds eroded the region, causing Antarctica to regress to pristine-like conditions where life is represented almost exclusively by highly adapted microorganisms (Rogers 2007).

Mosses and epilithic lichens are the major primary producers along the coast of Continental Antarctica. As you move towards the inner continent, the predominant, if not sole, life-form are endolithic microbial communities (de la Torre et al. 2003; Cary et al. 2010; Archer et al. 2017; Coleine et al. 2021; Coleine et al. 2021) which are complex communities of striated microorganisms all adapted to the extremes of Antarctica. In fact, the endolithic niche supplies microorganisms with more buffered conditions due to the thermal inertia of rocks; less severe thermal fluctuations and UV protection from high solar radiation give microbes a chance of survival under conditions incompatible with active life otherwise (Friedmann 1982; Albanese et al. 2020; Coleine et al. 2021; Omelon 2008; Coleine et al. 2020; Fanelli et al. 2021).

Cryptoendolithic Communities in Antarctica

One of the most well-studied endolithic communities in Antarctica is the cryptoendolithic lichen-dominated community, which has adapted by advancing into the airspaces of quartzite sedimentary rock, recognized as Beacon sandstone (Friedmann et al. 1993; Nienow et al. 2003). These self-supporting microbial ecosystems host among the most extremotolerant and adapted microbes known to date. These microbes evolved in complete genetic and geographic isolation under immense pressure conditions over a timescale of evolutionary significance (Rogers 2007). Lichens forgo thallus organization and expand as filamentous hyphae in obligate symbiosis with algae which function as primary producers, while other fungi and bacteria function as consumers (Coleine et al. 2021). Among them are the melanized black meristematic fungi which are conspicuously present. These black

meristematic fungi play a fundamental protective aspect in the entire community by screening excessive harmful solar emissions and establishing a melanized barrier above the stratification of separate biological compartments (Selbmann et al. 2013).

Micro-colonial Fungi

Black meristematic fungi are morphologically similar but phylogenetically diverse, organisms that have convergently evolved common features enabling them to flourish in the extremes (Egidi et al. 2014). Black meristematic, micro-colonial (MCF), or rock-inhabiting fungi (RIF) have adapted to survive on and within the rock substrates (Selbmann et al. 2015; Gueidan et al. 2011; Onofri et al. 2014). The genus *Friedmanniomyces* contains two species, *F. endolithicus,* and *F. simplex,* and is endemic only to the Antarctic and is the most widespread genera (Onofri et al. 1999; Selbmann et al. 2005) among Antarctic RIF. In particular, *F. endolithicus* has been isolated from all endolithic colonized rock samples analyzed with high throughput sequencing (Coleine et al. 2018; Coleine et al. 2021) and is the most frequently isolated RIF species (Selbmann et al. 2015). The increased incidence of this taxon shows a remarkable ability to thrive in prohibitive conditions in one of Earth's harshest environments, on par with Mars (Onofri et al. 2019; Coleine et al. 2020).

A few genomic traits were found to be overrepresented in *Friedmanniomyces* strains, such as responses to X-ray radiation, DNA damage, and salt tolerance stress; while genetic traits associated with meristematic growth and cold adaptation were overrepresented in *F. endolithicus* (Coleine et al. 2021); this suggests that these characters may represent an evolutionary advantage, giving fungi more opportunity to adapt and survive in the hostile

conditions of the Antarctic desert. Aside from its psychrophilic profile (Selbmann et al. 2005), very little information is available on the physiology of these organisms. It was observed that responses to sub-optimal temperature (28° C; 13° C higher than the optimal growth temperature) lead to a downregulation of protein expression without consequent heat-shock protein production (Tesei et al. 2012), maybe as an adaptation to its almost frozen habitat as seen in other extreme thriving eukaryotic organisms (Hofmann et al. 2000). The peculiar morpho-physiological and genetic adaptations together with the restricted and exclusive distribution in the Antarctic suggest that the evolution of these taxa may have been related to the geological history of the continent.

Whole Genome Duplication and Diploidization

Polyploidy is a genetic process that shaped the evolution of eukaryotic organisms. We have a deeper knowledge of specific eukaryotes, including important agricultural plants like cotton, and wheat, or organisms like the gray tree frog or goldfish (Gerhardt 1994; Halacka and Lusková 2000). In the fungal realm, incidences of polyploidy are less common (Albertin and Marullo 2012). Most understanding of polyploidy has been derived from yeasts, which indicate ancient polyploidization (palaeo-polyploidization) events like whole genome duplication (WGD) which leads to diploidization (Smith 1987; Wolfe and Shields 1997; Pawlowska and Taylor 2004). Diploidization occurs when a polyploid organism returns to diploid, this can be due to many mechanisms, including partial or full chromosomal loss, rearrangement of genes in a genome, sequence divergence, and gene deletions (Albertin and Marullo 2012). There are instances where fungal pathogens can undergo whole genome duplication events during infection and while obtaining antifungal

drug resistance (Gerstein et al. 2017). Ploidy is a very important adaptation fungi have been using as different mutations can have different effects in different ploidy backgrounds (Selmecki et al. 2015; Gerstein 2013). The opposite has also been proposed where nutrient limitation or stress in fact makes it more difficult for the species to prepare the multiple ploidies, so the species will drop the extra ploidies (Zörgö et al. 2013; Gerstein et al. 2017). It is important to think about why there would be different ploidies found in the same species of fungi and if it's a means to better protect and adapt the fungus in its environment.

The majority of MCF ploidy is still unknown but discussed extensively (Teixeira et al. 2017; Gostinčar et al. 2022). As MCFs have been collectively described as haploid, *H. werneckii* was also considered so. Sinha and Gostincar (Sinha et al. 2017) concluded an endoreduplication, but later it was learned that the diploid variants were produced through two haploid hybridizations that could occur due to two divergent strains (Gostinčar et al. 2018; Gostinčar et al. 2021). Genome duplication and high heterozygosity (~4.6%) were observed when a comparative phylogenomic analysis was performed (Romeo et al. 2020). Gostincar observed two strains of MCFs, *Hortaea werneckii* and *Aureobasidium melanogenum*, and noticed haploid, diploid, and tetraploid ploidies, while still being clonal which was confirmed with linkage disequilibrium analysis (Gostinčar et al. 2022). The occasional diploid and polyploid strains form stable and highly heterozygous intraspecific hybrids. The mechanism of these hybridization events remains unknown and needs to be analyzed in other MCFs. It is a strange circumstance where presumed asexual MCF diploids are found but are not produced through fungal recombination.

To determine the ploidy of fungi there are five general ways. 1) To use flow cytometry (Todd et al. 2018), though using this method with filamentous fungi is difficult as seen with *Aspergillus* and *Ashbya* species (Anderson et al. 2015) but possible (Bleichrodt and Read 2019). 2) To use K-mer heterozygosity assessment (Liu et al. 2013) as an assembly-independent method for estimating genome characteristics (like genome size, repeat structure, and heterozygous rate). There are many tools to accomplish this and all have their benefits. 3) Calculating SNP distribution of allele frequencies and observing if there are alternative alleles at each locus/site (Yoshida et al. 2013; Ngoot-Chin et al. 2021). The mating type locus could also be another region to observe (Lengeler et al. 2001). 4) Using Gaussian mixture models along with the maximum likelihood to select the most plausible ploidy model (Weiß et al. 2018; Margarido and Heckerman 2015). 5) Copy number variation (CNV) analysis to observe large-scale read-depth differences to observe intra-genome hybridization (Hughes et al. 2013; Pillet et al. 2012; Muller et al. 2009; Yoshida et al. 2013; Li et al. 2017).

MAT Loci

Another method of assessing the ploidy of an organism is to observe the mating (MAT) genes found in the MAT loci. The MAT loci are great indicators of sexual reproduction, yet we are unfamiliar with MCF sexual states (Teixeira et al. 2017). The MAT locus of *Neurospora crassa* is required for mating during sexual development but it also mediates heterokaryon incompatibility following fungal fusion during vegetative growth (Xiang and Glass 2004; Beadle and Coonradt 1944; Garnjobst and Wilson 1956). Sexual reproduction

in Ascomycetes relies on the presence of a compatible mating partner determined by the presence of one or two idiomorphic (non-allelic versions) HMG-box domain genes (Idnurm 2011; Wilken et al. 2017; Turgeon and Yoder 2000). The cassette model was verified by experiments indicating a haploid cell can switch genotypes at the MAT loci through a gene-conversion process (Haber 1998). The usual loci genes that fall into the cassette are APN2 (Apurinic-apyimidinic endonuclease 2) and SLA2 (Src-like-adapter 2) which generally flank the MAT loci at opposite ends (Butler et al. 2004). Obligately outcrossing, or heterothallic, species have two idiomorphs of the MAT genes, specifically MAT1-1 and MAT1-2, where each individual species may have one or the other gene in order to cross (Coppin et al. 1997; Wilken et al. 2017). While obligate selfing, or homothallic, species harbor both the MAT1-1 and MAT1-2 genes in the same genome (Lin and Heitman 2014; Heitman et al. 2007).

Very recently two MAT ORFs were detected, specifically in the Teratosphaeriaceae family (Ascomycota, Pezizomycota, Dothideomycetes, Dothideomycetidae, Mycosphaerellales, Teratosphaeriaceae), originally described as MATORF1 and MATORF2 (Arzanlou et al. 2010), and found on each idiomorph. Certain Teratosphaeriaceae species contain MATORF2 with MAT1-1 idiomorph, while MATORF1 was found with the MAT1-2 idiomorph (Aylward et al. 2020; Havenga et al. 2020). These genes were later called MAT1-1-10 and MAT1-2-12 (Aylward et al. 2020; Havenga et al. 2020). In the Teratosphaeriaceae family, the MAT locus is not necessarily flanked by SLA2 and APN2, and thus has been a challenge to confirm the entire MAT locus (Aylward et al. 2022).

Functionally, these two genes are still in the process of being understood, but there are some good hypotheses. Aylward et al (Aylward et al. 2022) observed these genes in the context of MAT loci within the Teratosphaeriaceae and noticed the alternative MAT genes resemble that of a spliced mRNA transcript. Their thoughts about that pattern are possibly through Long Interspersed Nuclear Elements (LINEs) which are a class of retrotransposons known to reverse transcribe mature mRNA to be integrated into a genome forming a pseudogene (Dewannieux and Heidmann 2005; Esnault et al. 2000). It is important to think and confirm the MAT locus for MCF to further understand the asexual nature of these species and their main ways of reproducing.

The aim of this work was to 1) Use a molecular clock approach to assess the evolutionary timescale of the endemic Antarctic fungal genus *Friedmanniomyces* evolution in regards to Antarctic geographic history and collected fungal fossil data. 2) Collect public *Friedmanniomyces* data, along with twenty-four new *F. endolithicus* genomes to be assembled and annotated for public release 3) Investigated ploidy of twenty-five strains of *F. endolithicus* isolated from the last forty years using K-mer heterozygosity determination, CNV analysis, Gaussian statistics, MAT locus determination, and SNP distributions of allele frequencies.

Here we investigated the divergence of *Friedmanniomyces* from other Dothideomycetes micro-colonial fungi using multiple population genomics approaches. Based on *Friedmanniomyces* distribution and abundance in only Antarctica, 1) We hypothesize that the genus *Friedmanniomyces* likely arose as Antarctica diverged from Pangea 65 million years ago. 2) As *Friedmanniomyces* is an endemic MCF isolated from cryptolithic samples in Antarctica, *F. endolithicus* collected from near coastal, lower elevation regions of Antarctica are more similar with haploid or diploid ploidy, while strains isolated from extreme isolation and elevations like mountains would have diploid to polyploidy. 3) *Friedmanniomyces* will follow the unusual compilation with LINE mating type locus fragments as seen in other Dothideomycetes MCF.

Materials and Methods

Public samples available in Genbank

The MCF *Friedmanniomyces endolithicus*, strain CCFEE 5311 (NAJP00000000.1) and *Friedmanniomyces simplex*, strain CCFEE 5184 (NAJQ00000000.1), were isolated from Antarctic cryptoendolithic communities collected in the Victoria Land (Continental Antarctica) at Ford Peak 75.430S 160.270E in 1996 and Battleship Promontory 76.550S 160.550E (Southern Victoria Land, McMurdo Dry Valleys) in 1997, respectively (Selbmann et al. 2005). Other publicly available fungal species used in this study include *Hortaea werneckii* EXF-2000 (MUNK00000000.1), *Aspergillus fumigatus Af293* (AAHF00000000.1), *Aspergillus ruber* CBS 135680 (AWRT00000000.1), *Cladophialophora immunda* CBS 83496 (JYBZ00000000.1), *Cladosporium*

cladosporioides TYU (NOXB00000000.1), *Exophiala dermatitidis NIH/UT8656* (AFPA00000000.1), *Exophiala mesophila* CBS 40295 (JYBW00000000.1), *Fusarium graminearum* PH-1 (AACM00000000.2), *Magnaporthe oryzae 70-15* (AACU00000000.3), *Neurospora crassa OR74A* (AABX00000000.3), *Rachicladosporium antarcticum CCFEE 5527* (NAJO00000000.1), *Saccharaomyces cerevisae* S288C (GCA_002057685.1), *Yarrowia lipolytica* CLIB122 (GCA_000002525.1).

Endolithic black fungi sequencing

Twenty-five strains of *F. endolithicus* were collected from endolithic communities in Antarctica. **Table 1** helps depict the Antarctic geographic locations (latitude and longitude), along with the altitude and distance from the ocean. Whole genome sequences of *F. endolithicus* CCFEE 5311 and *F. simplex* CCFEE 5184 were recently reported by Coleine et al. (Coleine et al. 2020). The remaining 24 species of *F. endolithicus* will be released with this publication under Bioproject PRJNA666634.

Genomic DNA was sheared with Covaris S220 ultrasonic homogenizer and the sequencing library was constructed using a NeoPrep TruSeq Nano DNA sample prep kit (Illumina) at the University of California, Riverside Genomics Core. Low-coverage whole genome sequencing (2x300 bp paired-end) was run on the Illumina Miseq platform.

Table 1. Twenty-five strains of *Friedmanniomyces endolithicus* were collected from Antarctica.

Strain	Metadata	Latitude	Longitude	Year Collected	Altitude	Sample Type
CCFEE_ 5001	Timber Peak, Northern Victoria Land, Antarctica	-74° 10' 0.0474"	162° 22' 59.8794"	1997	2754	sandstone
CCFEE_ 5193	Timber Peak, Northern Victoria Land, Antarctica	-74° 10' 0.0474"	162° 22' 59.8794"	1997	2754	sandstone
CCFEE_ 5195	Boomerang Glacier, Northern Victoria Land, Antarctica	-74° 32' 59.928"	163° 54' 0"	1997	484	granite
CCFEE_ 5199	Trio Nunataks, Northern Victoria Land, Antarctica	-75° 30' 0"	159° 41' 59.9994"	1997	1191	sandstone
CCFEE_ 5200	Trio Nunataks, Northern Victoria Land, Antarctica	-75° 30' 0"	159° 41' 59.9994"	1997	1191	sandstone

CCFEE_5208	Promontory between Widowmaker Pass and Olson Nunatak, Northern Victoria Land, Antarctica	-74° 55' 13.0794"	167° 22' 44.3994"	1997	200	sandstone
CCFEE_524	Linnaeus Terrace, McMurdo Dry Valleys, Antarctica	-76° 10' 5.0844"	167° 44' 12.12"	1981	1238	soil
CCFEE_5273	Arenaria Convoy Range, Antarctica	-76° 46' 59.879"	160° 45' 8.9064"	2004	1794	sandstone
CCFEE_5275	Arenaria Linnaeus Terrace, Antarctica	-76° 10' 5.0844"	167° 44' 12.12"	2004	1238	sandstone
CCFEE_5277	Arenaria Battleship Promontory, Antarctica	-76° 54' 41.856"	160° 54' 13.3986"	2004	1126	sandstone

CCFEE_5281	Arenaria Battleship Promontory, Antarctica	-76° 54' 41.856"	160° 54' 13.3986"	2004	1126	sandstone
CCFEE_5283	Arenaria Battleship Promontory, Antarctica	-76° 54' 41.856"	160° 54' 13.3986"	2004	1126	sandstone
CCFEE_5307	Battleship, Antarctica	-76° 54' 41.856"	160° 54' 13.3986"	2004	1126	sandstone
CCFEE_5311	Ford Peak, Antarctica	-75° 43' 0.0114"	160° 26' 59.9994"	2004	1239	sandstone
CCFEE_5311_v1	Ford Peak, Antarctica	-75° 43' 0.0114"	160° 26' 59.9994"	2004	1239	sandstone
CCFEE_5486	C/O IL CAMPO sample 4, Antarctica	76°54’39.7”	160°54’35.0”	2004	1105	sandstone
CCFEE_6074	Stewart Heights, Antarctica	73°29’26”	163°54'44”	2011	2670	sandstone

CCFEE_ 6081	Mt Bowen, Antarctica	75°45’24’’	161°03’46’’	2011	1874	sandstone
CCFEE_ 6082	Pudding Butte, Antarctica	75°45’24’’	161°03’46’’	2011	1874	granite
CCFEE_ 6096	Anderson Ridge, Antarctica	74°42’51”	162°37’04’’	2010	500	sandstone
CCFEE_ 6249	Timber Peak, site 2 Antarctica	74°10’11’’	162°25’34"	2010	2800	sandstone
CCFEE_ 6250	Trio Nunatak site 3, Antarctica	75°30’02’’	159°40’28’’	2010	1000	sandstone
CCFEE_ 6464	Mn New Zealand, Antarctica	75°30’02’’	159°40’28’’	2016	2965	sandstone
CCFEE_ 670	McMurdo Dry Valleys, Antarctica	-76° 10' 5.0844"	162° 44' 12.12"	1981	45	sandstone
CCFEE_ 690	McMurdo Dry Valleys, Antarctica	-76° 10' 5.0844"	162° 44' 12.12"	1981	45	sandstone

Genome assembly and annotation

Genome assemblies were constructed for the three MCF isolates using MiSeq Illumina sequencing. All genomes were *de novo* assembled with the AAFTF pipeline (v.0.2.3) (Palmer and Stajich 2022) which performs read QC and filtering with BBTools bbduk (*v.38.86)* (Bushnell 2014) followed by SPAdes (*v.3.15.2*) (Bankevich *et al.* 2012) assembly using default parameters, followed by screening to remove short contigs < 200 bp and contamination using NCBI's VecScreen. The BUSCO ascomycota_odb10 database (Manni *et al.* 2021) was used to determine how complete the assembly 3 isolates were.

We predicted genes in each near-complete genome assembly with Funannotate (*v1.8.1*) (Palmer and Stajich 2020). A masked genome was created by generating a library of sequence repeats with the RepeatModeler pipeline (Smit and Hubley 2008). These species-specific predicted repeats were combined with fungal repeats in the RepBase (Bao *et al.* 2015) to identify and mask repetitive regions in the genome assembly with RepeatMasker (*v.4-1-1*) (SMIT 2004). To predict genes, *ab initio* gene predictors SNAP (*v.2013_11_29*) (Korf 2004) and AUGUSTUS (*v.3.3.3*) (Stanke *et al.* 2006) were used along with additional gene models by GeneMark.HMM-ES (*v.4.62_lic*) (Brůna *et al.* 2020), and GlimmerHMM (*v.3.0.4*) (Majoros *et al.* 2004) utilize a self-training procedure to optimize *ab initio* predictions. Additional exon evidence to provide hints to gene predictors was generated by DIAMOND BLASTX alignment of SwissprotDB proteins and polished by Exonerate (*v.2.4.0*) (Slater and Birney 2005). Finally, EvidenceModeler (*v.1.1.1*) (Haas *et al.* 2008) generated consensus gene models in Funannotate that were constructed using

default evidence weights. Non-protein-coding tRNA genes were predicted by tRNAscan-SE (*v.2.0.9*) (Lowe and Chan 2016).

The annotated genomes were processed with antiSMASH (*v.5.1.1*) (Blin *et al.* 2021) to predict secondary metabolite biosynthesis gene clusters. These annotations were also incorporated into the functional annotation by Funannotate. Putative protein functions were assigned to genes based on sequence similarity to InterProScan5 (*v.5.51-85.0*) (Jones *et al.* 2014), PFAM (*v.35.0*) (Finn *et al.* 2014), Eggnog (*v.2.1.6-d35afda*) (Huerta-Cepas *et al.* 2019), dbCAN2 (*v.9.0*) (Zhang *et al.* 2018) and MEROPS (*v.12.0*) (Rawlings *et al.* 2018) databases relying on NCBI BLAST (*v.2.9.0+)* (Sofi *et al.* 2022) and HMMer (*v.3.3.2*) (Potter *et al.* 2018). Gene Ontology terms were assigned to protein products based on the inferred homology based on these sequence similarity analyses. The final annotation produced by Funannotate was deposited in NCBI as a genome assembly with gene model annotation.

Telomeric repeat determination

Identification of telomeric repeat sequences was performed using the FindTelomeres.py script (https://github.com/JanaSperschneider/FindTelomeres). Briefly, this searches for chromosomal assembly with a regular expression pattern for telomeric sequences at each scaffold's 5' and 3' end. Telomere repeat sequences were also predicted using A Telomere Identification toolkit (tidk) (*v.0.1.5*) "explore" option (https://github.com/tolkit/telomeric-identifier).

Mating type determination

The Mating Type (MAT) locus was identified by searching for homologous MAT genes. Using ncbi-blast+ (v.2.9.0) BLASTX, and BLASTP searches were performed on all 25 species of *Friedmanniomyces endolithicus*. The identified homologous regions were examined for their conserved synteny of the MAT locus using clinker (Gilchrist and Chooi 2021) and a custom Biopython script (Cock *et al.* 2009) (ZENODO) to extract the annotated region of the genome which contained the locus.

Phylogenomics and divergence time estimation

Phylogenetic distance was estimated using the PHYling toolkit (https://github.com/stajichlab/PHYling_unified). Profile-Hidden-Markov models of phylogenetic markers were searched in the predicted protein sequences using HMMER3 (*v3.3.2*) (Potter et al. 2018). A total of 555 conserved orthologous markers were identified with hmmsearch (cutoff = $1E^{-10}$) in the 5 Eurotiomycetes, 5 Dothideomycetes, 5 root strains which fall under the Sordariomycetes and Ascomycetes clades (**Table 2**). Protein sequence homologs for each phylogenetic marker were aligned with hmmalign to marker profile-HMM. The protein and coding sequences of the markers were concatenated into an alignment denied by each gene marker. Tree generation was completed using the PHYling alignment output and the maximum likelihood statistical method used in IQ-Tree (*v2.1.1*) (Minh et al. 2020).

Divergence times were estimated using the penalized likelihood algorithm implemented in r8s (*v1.81*) (Sanderson 2003) utilizing fungal most common ancestor estimates generated

by Lutzoni and team (Lutzoni et al. 2018). The three most common ancestor estimates were used in this analysis: Eurotiomycetes 376.27 Mya, Dothideomycetes 350.14 Mya, and Saccharomycotina 402.29 Mya (**Table 3**). The HMMER tool esl-alistat was used to compute average identity in the multiple sequence alignment generated by PHYling. This number was used as the variable for nlines in the r8s configuration script. The R packages Treeio (*v1.10.0*) (Wang et al. 2020) and Deeptime (*v0.0.52*) (Gearty et al. 2022) were used to plot molecular clock divergence time results. Data files and R code are available in the GitHub repository DOI: 10.5281/zenodo.7339211

Table 2. Strains and fossil time points used in the molecular clock analysis.

Strains	Isolation source	Genbank	SRA	Reference
Aspergillus fumigatus Af293	Human died from Aspergillosis.	AAHF00000000.1	SRR068952	(Nierman et al. 2005)
Aspergillus ruber CBS 135680	Halophilic Dead Sea	AWRT00000000.1	SRR427135	(Kis-Papo et al. 2014)
Cladophialophora immunda CBS 83496	Human, sub-cutaneous ulcer	JYBZ00000000.1	SRS1103389	(Sterflinger et al. 2015)
Cladosporium cladosporioides TYU	*Taxus cuspidata*	NOXB00000000.1	SAMN07420618	(Chen et al. 2014)
Exophiala dermatitidis NIH.UT8656	Human strain	AFPA00000000.1	SRS281040	(Chen et al. 2014)
Exophiala mesophila CCFEE 6314	Antarctic cryptoendolithic communities	NAJM00000000.1	SRR5223779	(Coleine et al. 2019)
Friedmanniomyces endolithicus CCFEE 5311	Antarctic cryptoendolithic communities	NAJP00000000.1	SRR5223777	(Coleine et al. 2020)

Friedmanniomyces simplex CCFEE 5184	Antarctic cryptoendolithic communities	NAJQ00000000.1	SRR5223776	(Coleine et al. 2020)
Fusarium graminearum PH-1	Wheat and barley	AACM00000000.2	ERS670574	(Cuomo et al. 2007)
Hortaea werneckii EXF2000	Saltern	MUNK00000000.1	SRR866616	(Sinha et al. 2017)
Neurospora crassa OR74A	Bread mold	AABX00000000.3	SRS691181	(Galagan et al. 2003)
Magnaporthe oryzae 70-15	Rice	AACU00000000.3	SAMN02953596	(Xue et al. 2012)
Rachicladosporium antarcticum CCFEE 5527	Antarctic cryptoendolithic communities	NAJO00000000.1	SRR5223785	(Coleine et al. 2017)
Saccharomyces cerevisiae S288C	Baker's yeast	GCA_000146045.2	ERS1360911	(Fisk et al. 2006)
Yarrowia lipolytica CLIB122	Cross isolate from a Paris sewer and American corn processing plant	GCA_000002525.1	ERS610106	(Magnan et al. 2016)

Table 3. Time tree orientation dimensions. Estimated median age and 95% highest posterior density (HPD) of major fungal clades (crown nodes) considered in this study. Data have been retrieved from Lutzoni et al. 2018 (Lutzoni et al. 2018).

Number	Clade	Median age (Ma)	Age 95% HPD (Ma)
1	Eurotiomycetes	376.27	342.92-414.92
2	Dothideomycetes	350.14	311.98-393.00
3	Sordariomycetes	219.14	201.33-239.61
4	Saccharomycotina	402.29	346.03-463.77

Heterozygosity testing using K-mers

To estimate the extent of genome duplication, K-mer frequency analysis with GenomeScope (*v2.0*) (Ranallo-Benavidez et al. 2020) was employed. First, K-mer frequencies within raw sequencing reads were analyzed and calculated using Jellyfish (*v2.3.0*) (Marcais and Kingsford 2012) to estimate major genome characteristics such as size, heterozygosity, and repetitiveness. Analysis was performed with k=21. "Jellyfish count" and "jellyfish histo" tags were used to generate two-tab delimited files of frequency counts. These tabulated files were then added to the GenomeScope v.2 website (http://qb.cshl.edu/genomescope/genomescope2.0/) to generate the data visuals and genome statistics. Once the analysis was complete, K-mer 21 was picked as the value to continue with the remaining K-mer counting analysis. Other K-mer assessing tools were

also used to confirm GenomeScope results, including K-mer Analysis Tool (KAT) (Mapleson et al. 2017) and findGSE (Sun et al. 2018) (**Appendix 3**)

Calculating SNP distribution of allele frequencies

Sequence variation among isolates was assessed using the best practices of the Genome Analysis ToolKit GATK (*v. 4.0.4.0*) (McKenna *et al.* 2010; Franke and Crowgey 2020) to identify SNPs and Insertion/Deletions (INDEL). Illumina paired-end reads were aligned to CCFEE_5001 assembly with BWA (*v.0.7.17*) (Li and Durbin 2010) and processed with Samtools (*v.1.8*) (Li *et al.* 2009) and Picard Toolkit (Institute; Toolkit) AddOrReplaceReadGroups and MarkDuplicates (*v.2.18.3*). The alignments were further improved by realigning reads near inferred INDELs using GATK tools RealignerTargetCreator, IndelRealigner, and PrintReads. Genotypes were inferred with the GATK Haplotype and GenotypeGVCF methods to produce a single VCF file of the identified variants. Low-quality SNPs were further filtered using GATK VariantFiltration and finally, SelectVariants was used with the parameters: mapping quality (score < 40), quality by depth (<2 reads), Strand Odds Ratio (SQR > 4.0), Fisher Strand Bias (> 200), and Read Position Rank Sum Test (< -20) to retain only high-quality polymorphisms. The ploidy for each strain was visualized with R using the vcfR (Knaus and Grünwald 2017) package.

Copy number variation

Copy number variation (CNV) was examined by plotting 10kbp window-based read coverage of the short-read alignments of each isolate. The depth of coverage was calculated using Mosdepth (Pedersen and Quinlan 2018) and visualized with R using the ggplot2 package (Wickham 2016).

Using Gaussian statistics to confirm ploidy

Gaussian statistics were used to calculate and confirm the ploidy of each *F. endolithicus* species using the program nQuire, (Weiß et al. 2018). Illumina read files were re-mapped onto the reference assembled genome (CCFEE_5001) to a BAM format using the program BWA (*v.0.7.17*). BAM files were then used to run the nQuire "create" program to create frequency bins. To tackle the problem of the base frequency having a high number of statistical noise (due to mis-mappings or could be elevated due to high repeat regions/genomes), nQuire provides a "denoise" step to clear the incongruencies. To assess the ploidy level, nQuire uses the Gaussian Mixture Model (GMM) combined with the Expectation-Maximization (EM) algorithm in the form of the command "lrdmodel". The program also provides two other options "modeltest", which provides each of the three fixed models used in the lrdmodel option along with "estmodel" which provides the free model option. In short, the maximized log-likelihood of the three fixed models can be subtracted from the maximized log-likelihood of the free model to obtain the three-delta log-likelihoods. The lowest delta log-likelihood of a particular fixed model indicates the optimum and is in the highest support of the particular ploidy level. The final assessment is the command "histotest" which reports statistics for each ploidy considered, including

the sum of squared residuals (SSR) of ideal vs empirical histograms, the slope (y-y), the standard error of the slope, and the R2 of the regression of y-values. Another way to detect the ploidy of each sample is to observe the histotest results - optimum results include low SSR values, a positive slope with low standard error, and a high R2.

Results

Twenty-five strains of Friedmanniomyces were collected from Antarctica

Over a span of 30 years, multiple strains of *F. endolithicus* were collected from Antarctica, including the original two species, *Friedmanniomyces endolithicus* CCFEE 5311 and *Friedmanniomyces simplex* CCFEE 5184. A map of Antarctica's McMurdo Valley region (**Figure 2**) highlights where all 26 species of Friedmanniomyces were collected. Metadata collected on each of twenty-five strains of F. endolithicus is collected in **Table 1.** Google Earth was used to visualize all the coordinates of each species, along with elevation data to be used for further ploidy analysis. Within the 25 strains, two are *Friedmanniomyces endolithicus* 5311, and the name 5311_v1 was added to the originally submitted *F. endolithicus* 5311 as there seemed to be inherent differences between the two sets of sequencing files.

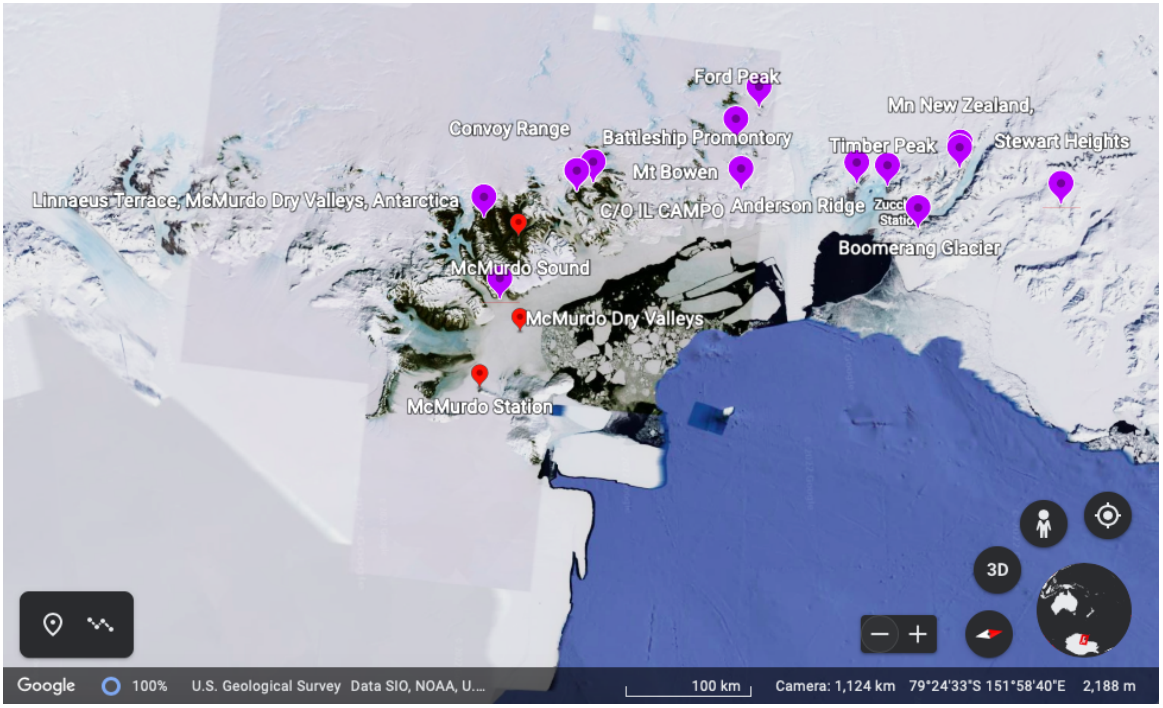


Figure 2. Map of Antarctica's McMurdo Valley region. The purple indicators show the approximate regions which our 25 strains of F. endolithicus were isolated from. Coordinates were mapped out and labeled as the regions from which samples were collected along with elevation data.

Genome Assembly and Annotation

Genome assembly statistics and characteristics (**Table 4**) were generated from our assembly pipeline along with BUSCO results and telomeres mapped out for each strain. The average coverage for all 25 *F. endolithicus* Illumina sequencing was around 124x, while the average genome length was around 48.3Mbp. The average L50 was very large, around 150 where the smallest L50 (CCFEE_5001) was 10, while the largest L50 belonged to CCEE_6464 with 1337 contigs. Certain species like CCFEEE_5001, CCFEE_5486, and CCFEE_6074 have a markedly lower genome length (around 20 Mbp), while other species like CCFEE_5199, CCFEE_5200, CCFEE_5273, and CCFEE_5277 have a higher genome length (around 66Mbp).

MAT Loci Assessment

The mating types were determined using prior work on Teratosphiaceae fungi observed in the publication by Aylward et al (Aylward et al. 2022), the protein sequences were saved and used to BLAST into each *F. endolithicus* genome. The genes in this locus are generally adjacent to AIP2, APN2, and HIS7, and may sometimes have alternative genes as seen with LINE retrotransposons. BLAST results indicated certain species had multiple copies of these regions, some had more than 2 copies of the MAT locus. Species that contained only one copy of the MAT locus include CCFEE 5001, CCFEE 5486, and CCFEE 6074 (**Figure 3**). Species that contained three copies of the MAT locus include CCFEE 5199, CCFEE 5200, and CCFEE 5273 (**Figure 4**). The remaining 19 species had two copies of the MAT locus including CCFEE 5193, CCFEE 5195, CCFEE 5208, CCFEE 524, CCFEE 5275, CCFEE 5277, CCFEE 5281, CCFEE 5281, CCFEE 5283, CCFEE 5307, CCFEE 5311, CCFEE 5311_v1, CCFEE 6081, CCFEE 6082, CCFEE 6096, CCFEE 6249, CCFEE 6464, CCFEE 670, and CCFEE 690 (**Figure 5**). An observation with CCFEE 5277 was that it contained 3 flanking regions, but had the third set of MAT genes missing. This could be due to its high contig count which could be disrupting the proper protein prediction. Unlike other Dothideomycetes species or even Chaetothyriales MCF, *F. endolithicus* did not have SLA2 to flank the end of the MAT locus, and instead, SLA2 was found 10kbp-50kbp away from the MAT genes, this was also seen in Aylward et al (2022) for other Teratosphiaceae.

Table 4. Genome Assembly Statistics on all 25 strains of *F. endolithicus*.

STRAIN	COVERAGE	CONTIG COUNT	TOTAL LENGTH	L50	N50	GC%	BUSCO Complete	Telomeres Found
CCFEE_5001	289.33	889	23,497,767	10	714,947	56.65	96.9	9, 7F, 2R
CCFEE_5193	119.50	4,715	49,049,759	112	134,755	56.17	97.4	35, 17F, 18R
CCFEE_5195	77.93	3,448	47,100,715	37	408,586	56.55	97.6	35, 21F, 14R
CCFEE_5199	489.18	1,711	67,311,006	125	152,028	56.59	97.8	36, 21F, 15R
CCFEE_5200	126.14	4,477	67,647,515	144	128,909	56.47	97.5	58, 24F, 34R
CCFEE_5208	169.68	2,542	46,417,718	31	428,208	56.6	97.6	33, 19F, 14R
CCFEE_524	82.05	39,528	76,162,173	285	63,248	57.55	96.6	50, 22F, 28R
CCFEE_5273	92.29	6,315	66,197,846	129	152,618	56.47	97.8	57, 29F, 28R
CCFEE_5275	27.36	9,685	48,327,421	105	139,649	56.1	97.2	63, 32F, 31R
CCFEE_5277	296.48	3,169	64,614,953	104	173,150	56.56	97.7	34, 13F, 21R
CCFEE_5281	39.58	5,633	46,973,427	143	95,340	56.33	96.9	37, 20F, 17R
CCFEE_5283	107.60	2,438	45,325,156	102	127,184	56.48	97	7, 6F, 1R

CCFEE_5307	44.15	4,936	46,710,760	144	95,722	56.33	96.8	35, 18F, 17R
CCFEE_5311	81.75	3,279	47,023,468	126	119,801	56.33	97.3	42, 23F, 19R
CCFEE_5311_v1	70.99	427	46,752,167	39	383,223	56.5	97	52, 24F, 28R
CCFEE_5486	51.40	10,382	25,750,566	298	24,252	55.07	80.9	33, 22F, 11R
CCFEE_6074	28.82	2,941	24,317,134	13	641,321	56.44	97.1	27, 10F, 17R
CCFEE_6081	54.18	3,493	46,275,398	114	123,582	56.33	96.7	30, 18F, 12R
CCFEE_6082	175.33	1,594	46,114,487	36	425,051	56.62	97.3	24, 12F, 12R
CCFEE_6096	173.45	3,199	46,547,415	40	358,973	56.59	97.6	36, 12F, 24R
CCFEE_6249	65.42	2,951	46,441,624	118	110,133	56.35	96.7	34, 16F, 18R
CCFEE_6250	207.70	2,906	46,543,679	32	501,394	56.59	97.5	29, 16F, 13R
CCFEE_6464	16.67	7,778	41,880,500	1337	9,320	56.43	87.9	22, 14F, 8R
CCFEE_670	162.74	2,427	46,280,107	111	130,608	56.39	96.3	31, 14F, 17R
CCFEE_690	66.68	8,350	49,767,767	127	108,649	55.52	97.1	35, 18F, 17R

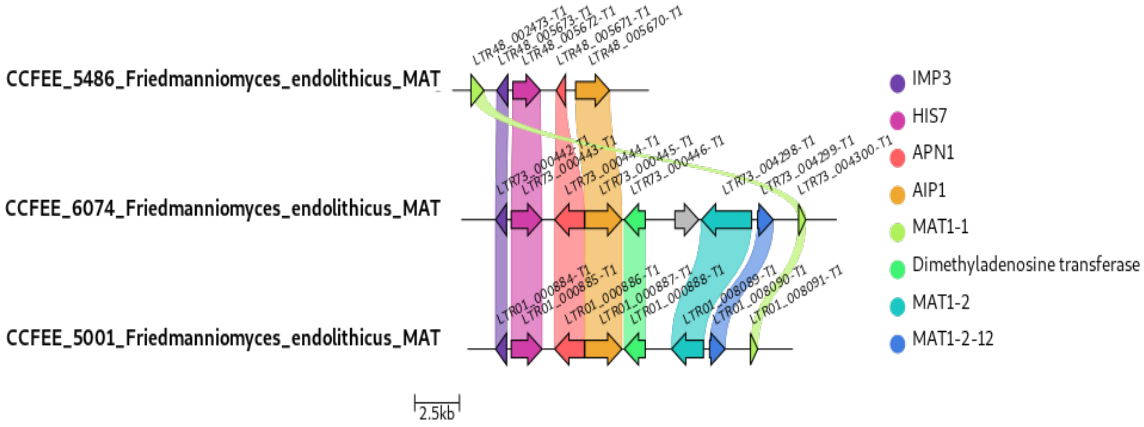


Figure 3. MAT loci comparing three haploid species of *F. endolithicus.* Single copies of MAT locus for species CCFEE 5486, CCFEE 6074 and CCFEE 5001 are indicated here. Purple indicates a flanking IMP3, HIS7 is colored in pink, APN1 is colored in orange, AIP1 is in yellow, Dimethyladenosine transferase in turquoise, MAT1-2 in teal, MAT1-1 in lime-green and MAT1-2-12 (accessory MAT gene) in blue). MAT1-1 and MAT1-2 seem almost always present in all loci, while the accessory genes alternate. Genome CCFEE_4856 may be a very shattered genome and all genes are not predicted properly in its MAT locus.

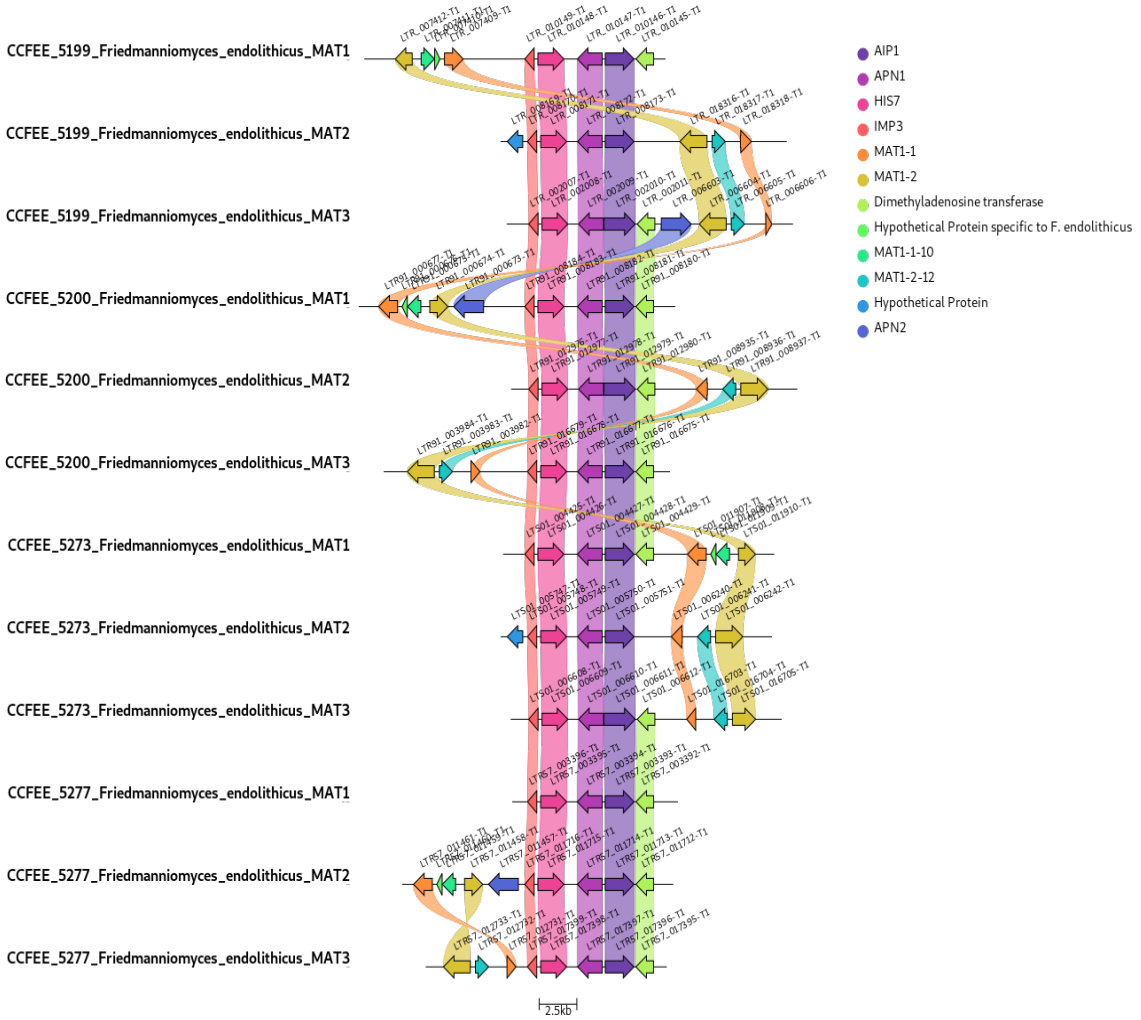


Figure 4. MAT loci comparing four triploid ploidy species of *F. endolithicus*. The MAT locus is split per species to show the clear copies and differences at each intra-locus as well as between species inter-locus. Each locus is flanked with IMP3, HIS7, AIP1, and APN genes with a new predicted dimethyl-adenosine transferase. MAT1-1 and MAT1-2 seem almost always present in all loci, while the accessory genes alternate.

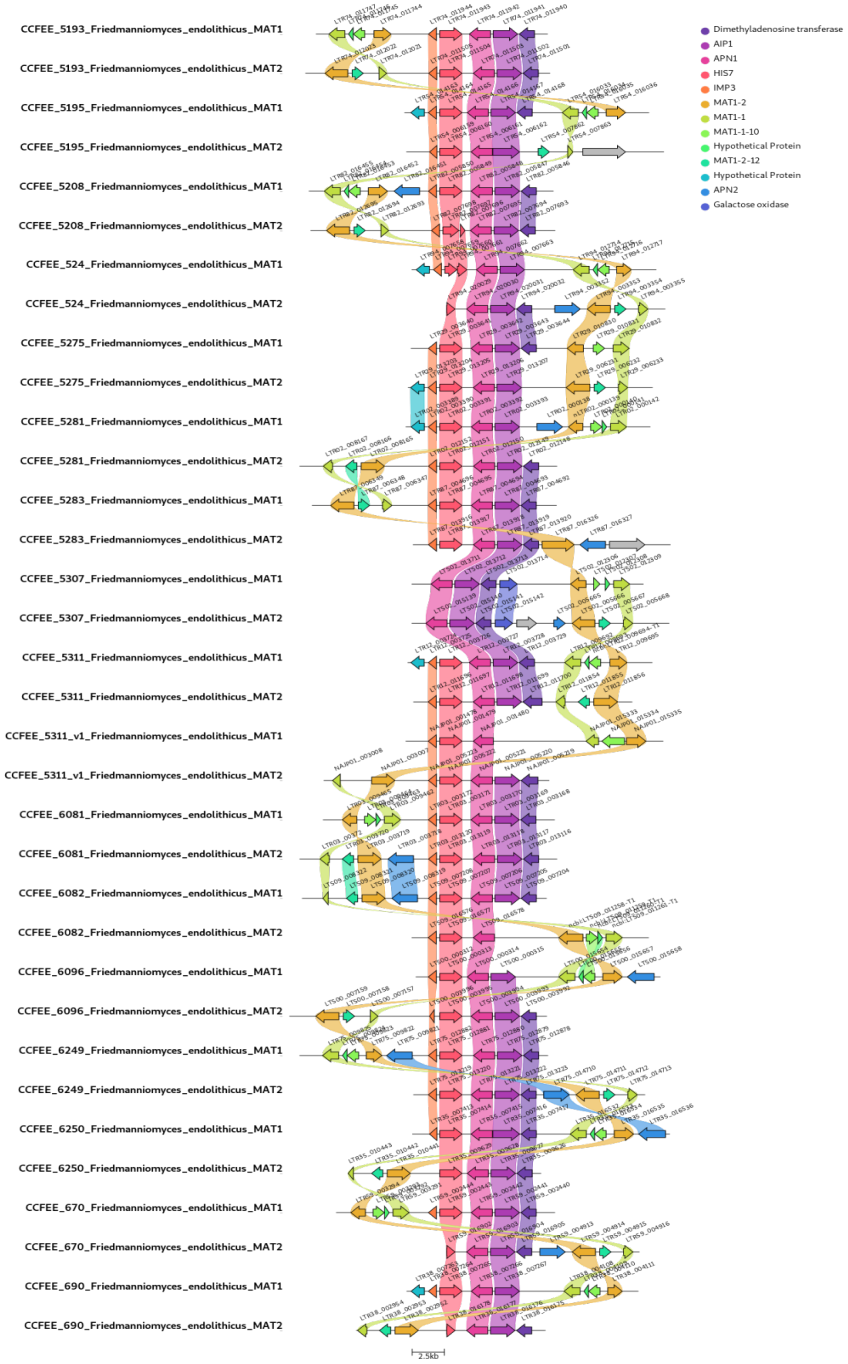


Figure 5. MAT loci comparing eighteen diploid species of *F. endolithicus*. MAT locus is flanked with IMP3, AIP1, HIS7, and APN genes, and now seeing dimethyl-adenosine transferase joins the group. It is noticeable that the MAT1-1 and MAT1-2 genes are accompanied by the accessory MAT genes, MAT1-1-10 and MAT1-2-12. MAT1-1 and MAT1-2 seem almost always present in all loci, while the accessory genes alternate.

The genes flanking *F. endolithicus* genomes include IMP3, HIS7, AIP1, APN2, and dimethyl-adenosine transferase. Single-copy MAT locus genomes, CCFEE_5486, CCFEE_5001, and CCFEE_6074 all contain the MAT1-1 genes, while CCFEE_5001 and CCFEE_6074 contain both copies of the MAT gene including the accessory protein MAT1-2-12 (**Figure 3**). Double-copy MAT locus genomes, CCFEE_5193, CCFEE_5195, CFEE_5208, CCFEE_524, CCFEE_5275, CCFEE_5307, CCFEE_5311, CCFEE_5311_v1, CCFEE_6081, CCFEE_6082, CCFEE_6096, CCFEE_6249, CCFEE_6250, CCFEE_6464, CCFEE_670, and CCFEE_690 all contain MAT1-1 and MAT1-2 genes (**Figure 5**). It seems the accessory MAT genes (MAT1–1-10 and MAT1-2-12) alternate in each MAT locus. The naming of MAT1 and MAT2 are arbitrary as chromosome arrangement is still too random to confirm, which shows the random result we see in the figure. Triple copy MAT locus genomes, CCFEE_5199, CCFEE_5200, CCFEE_5273, and CCFEE_5277 almost all have the MAT1-1 and MAT1-2 genes (except for CCFEE_5277 first MAT locus). There also seems to be an overall pattern where one locus contains only both MAT1-1 and MAT1-2 genes, while the other two loci would have the two genes along with the accessory proteins. It could be where the locus one only the representative MAT1-1 and MAT1-2 genes are the ancestor versions of the MAT loci, while the other two are remnants from the LINE transposon or some type of duplication event and LINE transposon event.

Friedmanniomyces spp. diverged after the Antarctic land mass reached the South Pole around the Paleogene or Cenozoic era

Using Lutzoni's calculated plant fossil dates for fungal lineages including fungal Phyla, Orders, and Classes information (Lutzoni et al. 2018) (**Table 3**), we calibrated the nodes on a phylogenetic tree generated from previously described phyla and lineage (**Figure 6**). The analysis indicates the genus *Friedmanniomyces* diverged from other Dothideomycetes like *Hortaea werneckii* on an average of 48.42 (±12.7) million years ago, indicating a drift in evolution around the time Antarctica settled in the latitude seen today around the Late Cretaceous and now Cenozoic (Fitzgerald 2002). The fifteen fungi observed in this phylogenetic tree include 5 Dothideomycetes strains, 5 Eurotiomycetes strains, 3 Sordariomycetes strains, and 2 Saccharomycotina strains, showcasing representatives from 4 different groups, including micro-colonial fungi and non-micro-colonial fungi from the same Class. The phylogenetic tree includes time ranges to orient the divergence of the chosen lineages as best as possible.

The species belonging to the class Dothideomycetes (*Rachicladosporium antarcticum*, *Cladosporium cladosporioides, Friedmanniomyces* spp., and *Hortaea werneckii*) diverged around 311.98 Mya, the range given by Lutzoni was around 311.98-393 Mya, on average about 350.14 Mya (Lutzoni et al. 2018), which is estimated near the middle of the Carboniferous. Specifically, *F. endolithicus* and *F. simplex,* two strains isolated from the Antarctic cryptoendolithic communities (Coleine et al. 2020), the analysis identified the divergence between both fungi around 48.42 Mya around the early Paleogene or Cenozoic. *H. werneckii,* isolated from a saltern in Slovenia (Sinha et al. 2017), clustered with the two

Friedmanniomyces spp., confirming that these three strains fall in the Dothideomycetes and are sister to the *Friedmanniomyces* species as expected. Our findings indicated that the divergence between both genera happened around 186.84 Mya during the Jurassic. The species *R. antarcticum* was isolated from the Antarctic crypto-endolithic communities, while *C. cladosporioides* from the seeds of *Taxus cuspidata* in South Korea (Coleine et al. 2017). These two species diverged apart from each other more than 206.47 Mya during the beginning of the Triassic. Five species belonging to Eurotiomycetes (*Cladophialophora immunda*, *Exophiala mesophila, E. dermatitidis, Aspergillus fumigatus, and Aspergillus ruber*) have been added to the evolutionary analysis. We found that *E. dermatitidis* and C. *immunda*, were both described as human opportunistic pathogens (Sterflinger et al. 2015; Chen et al. 2014). We identified the divergence between both fungi around 85.08 Mya from each other during the middle Cretaceous. *E. mesophila*, isolated from the Antarctic cryptoendolithic communities (Coleine et al. 2019), diverged from the other two human pathogenic strains around 112.53 Mya in the late Cretaceous. Instead, *A. fumigatus* and *A. ruber*, isolated from the human body and halophilic environment (Nierman et al. 2005; Kis-Papo et al. 2014), respectively, the analysis identified the divergence between both fungi around 103.73 Mya which also falls in the Cretaceous. The two *Aspergillus* spp. fungi diverged from the other three micro-colonial fungi Eurotiomycetes around 342.92 Mya, the range determined by Lutzoni was around 342.92-414.92 Mya, on average of about 376.27 Mya (Lutzoni et al. 2018), indicating a deep divergence in the Devonian.

Three Sordariomycetes species were also included in this analysis to help create a better rooting for this phylogenetic tree. The analysis identifies *Magnaporthe oryzae,* a pathogen of rice (Xue et al. 2012), diverged from *Neurospora crassa*, traditionally isolated from bread (Galagan et al. 2003), around 170.55 Mya during the Jurassic. The analysis indicates species *Fusarium graminearum* (Cuomo et al. 2007) diverged from the other two Sordariomycetes around 206.95 Mya, the range determined by Lutzoni was around 201.33-239.61 Mya, on average about 219.14 Mya (Lutzoni et al. 2018), late in the Triassic. The five Dothideomycetes seem to have diverged from the Eurotiomycetes around 368.24 Mya in the early Devonian, as the analysis identifies Sordariomycetes diverged from the Dothideomycetes and Eurotiomycetes around 375.78 Mya.

Finally, two Ascomycota spp. strains were also added to our study. *Saccharomyces cerevisiae* and *Yarrowia lipolytica* (Fisk et al. 2006; Magnan et al. 2016) diverged around 272.50 Mya during the Permian. When comparing these two strains to the remaining tree, the analysis identifies the divergence early on, around 414.92 Mya, the range determined by Lutzoni was around 346,03-463.77 Mya (Lutzoni et al. 2018), on average about 402.29 Mya during the early Devonian. All strains used for this analysis with NCBI, and location details are depicted in **Table 2**.

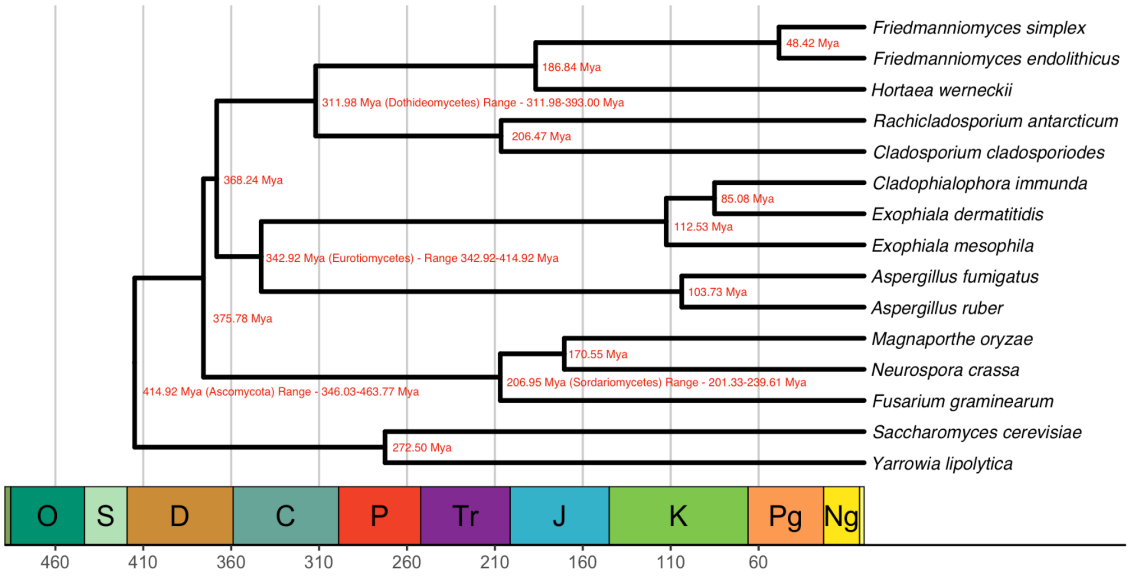


Figure 6. Micro-Colonial Fungi time tree using point estimated fossil data. A dated phylogeny based on Maximum Likelihood reconstruction from concatenated protein alignment of 555 genes observed in 15 genomes. The time tree was inferred with the r8s tool using fixed node divergence times using data available by (Lutzoni et al. 2018). The bottom scale indicates geological epochs. Eras (million years ago, Mya): Ng - *Neogene* (23 - 2.6 Mya), Pg - *Paleogene* (66 - 23 Mya), K - Cretaceous (145 - 66 Mya), J - Jurassic (201.3 - 145 Mya), T - Triassic (251.9 - 201.3 Mya), P - Permian (298.9 - 251.9 Mya), C - Carboniferous (358.9 - 298.9 Mya), D - Devonian (419.2 - 358.9 Mya), S - Silurian (443.8 - 419.2 Mya), and O - Ordovician (485.4 - 445.8 Mya). Era information was collected from Cohen et al. (2013). Estimated node divergence times and taxonomic classification are indicated in red. The Ascomycota node was used to fix the tree indicated by (Ascomycota), while Eurotiomycetes (Eurotiomycetes), Sordariomycetes (Sordariomycetes), and Dothideomycetes (Dothideomycetes) are also labeled in two other fixed nodes. Labels are in red.

Using K-mer counting programs to determine the ploidy of Friedmanniomyces endolithicus

GenomeScope2 results indicate the 25 *F. endolithicus* species have a mixed ploidy state. **Figure 7** highlights 3 out of the 25, while the remaining species are available to view in (Appendix 3). **Figure 7A** describes species CCFEE_5001, our haploid representative. We see a singular shoulder around the 500-coverage range and about 95.6% unique, meaning

4.4% of its genome contains repetitive elements. **Figure 7B** describes species CCFEE_5195, our diploid representative. We see a double shoulder around the 50 and 110 coverage range and about 76.6% unique, meaning 33.4% of its genome contains repetitive elements. While **Figure 7C** describes species CCFEE_5199, our triploid representative. We see a similar double shoulder as in Figure 7B, but the coverage range is further ahead at 210 and 450, with about 67.2% unique, meaning 42.7% of its genome contains repetitive elements. GenomeScope2 provided us with a summary of homozygosity and heterozygosity statistics visualized in **Table 5** and summarized for ploidy averages in **Table 6**. For haploid strains, the homozygosity rate is closer to 1, with some discrepancies including CCFEE_6250 and CCFEE_6082, CCFEE_5311. Certain diploid strains have a higher heterozygosity rate than the described polyploidy strains including strains CCFEE_5195, CCFEE_5208, CCFEE_5311_v1, and CCFEE_6096. These differences are also reflected in the average table (**Table 6**), indicating a lower heterozygosity rate in polyploidy strains than in the diploid strains. The genome repeat length is 3 times greater in polyploids than seen in the average haploid and diploid genomes. The average diploid genomes also seem to have a higher error rate (0.3288%) than the haploid (0.2313%) or polyploid (0.2454%) genomes.

Table 5. Description of average heterozygosity and homozygosity calculated using K-mer counting programs. The program Genomescope2 was used to assess each species of *F. endolithicus*. Statistics are described below including the homozygous mean and heterozygous mean which helps depict the ploidy of each species. Other details including the genome haploid length, genome repeat length, and genome unique length are shared. The model fit percentage and error rate are other statistics used to confirm the validity of the described data. The final column is the ploidy as described by the results seen in this table.

Strain	Homozygous (aa) mean	Heterozygous (Aa) mean	Genome Haploid Length	Genome Repeat Length	Genome Unique Length	Model Fit	Error Rate	Ploidy
CCFEE_5001	100.00%	0.00%	11,494,919bp	389,109.5bp	11,105,809bp	95.59%	0.25%	Haploid
CCFEE_5193	92.19%	7.80%	25,276,383bp	4,786,639.5bp	20,489,744bp	92.70%	0.23%	Diploid
CCFEE_5195	90.05%	9.94%	25,715,551bp	6,012,377.5bp	19,703,174.5bp	92.49%	0.24%	Diploid
CCFEE_5199	91.20%	8.79%	34,403,202.5bp	11,299,608.5bp	23,103,594bp	91.61%	0.23%	Triploid
CCFEE_5200	90.41%	9.58%	33,633,708.5bp	10,129,601bp	23,504,107.5bp	91.97%	0.29%	Triploid
CCFEE_5208	89.01%	10.98%	24,072,681bp	3,900,016.5bp	20,172,664.5bp	93.93%	0.24%	Diploid
CCFEE_524	92.60%	7.39%	27,332,245bp	5,800,752.5bp	21,531,492.5bp	95.00%	0.57%	Diploid

CCFEE_5273	90.61%	9.38%	33,468,283.5bp	10,137,417.5bp	23,330,865.5bp	93.61%	0.23%	Triploid
CCFEE_5275	90.01%	9.98%	23,133,697bp	2,276,230bp	20,852,466.5bp	97.74%	0.32%	Diploid
CCFEE_5277	90.08%	9.91%	31,594,459.5bp	8,349,318.5bp	23,245,140bp	89.12%	0.24%	Triploid
CCFEE_5281	92.80%	7.19%	23,850,389.5bp	2,435,958bp	21,414,431bp	96.28%	0.25%	Diploid
CCFEE_5283	90.53%	9.46%	23,392,650bp	6,850,972.5bp	16,541,677bp	94.52%	0.19%	Diploid
CCFEE_5307	93.47%	6.52%	24,705,260.5bp	4,498,978.5bp	20,206,282bp	96.56%	0.23%	Diploid
CCFEE_5311	99.98%	0.01%	56,246,193.5bp	21,262,407.5bp	34,983,759bp	83.45%	0.21%	Diploid
CCFEE_5311_v1	88.50%	11.49%	24,165,181.5bp	4,559,100.5bp	19,606,081bp	93.90%	0.77%	Diploid
CCFEE_5486	99.73%	0.26%	23,165,786bp	1,331,875.5bp	21,833,910bp	92.45%	0.23%	Haploid
CCFEE_6074	99.95%	0.04%	23,901,042bp	1,342,057.5bp	22,558,984.5bp	97.31%	0.28%	Haploid

CCFEE_6081	92.79%	7.20%	24,593,996bp	3,225,513.5bp	21,368,483bp	94.40%	0.24%	Diploid
CCFEE_6082	99.95%	0.04%	46,459,808bp	7,606,733bp	38,853,075bp	94.59%	0.19%	Diploid
CCFEE_6096	88.58%	11.41%	25,056,428.5bp	5,742,638bp	19,313,790bp	91.38%	0.22%	Diploid
CCFEE_6249	94.30%	5.69%	26,270,163.5bp	5,671,311.5bp	20,598,852bp	92.44%	0.22%	Diploid
CCFEE_6250	99.94%	0.05%	49,234,769.5bp	10,613,545bp	38,621,224.5bp	91.98%	0.19%	Diploid
CCFEE_6464	93.24%	6.75%	24,066,013.5bp	2,876,296.5bp	21,189,717bp	96.97%	0.58%	Diploid
CCFEE_670	92.09%	7.90%	23,85,367bp	3,891,224bp	19,961,143bp	94.64%	0.24%	Diploid
CCFEE_690	93.47%	6.55%	24,456,670.5bp	3,607,871bp	20,848,798.5bp	95.93%	0.32%	Diploid

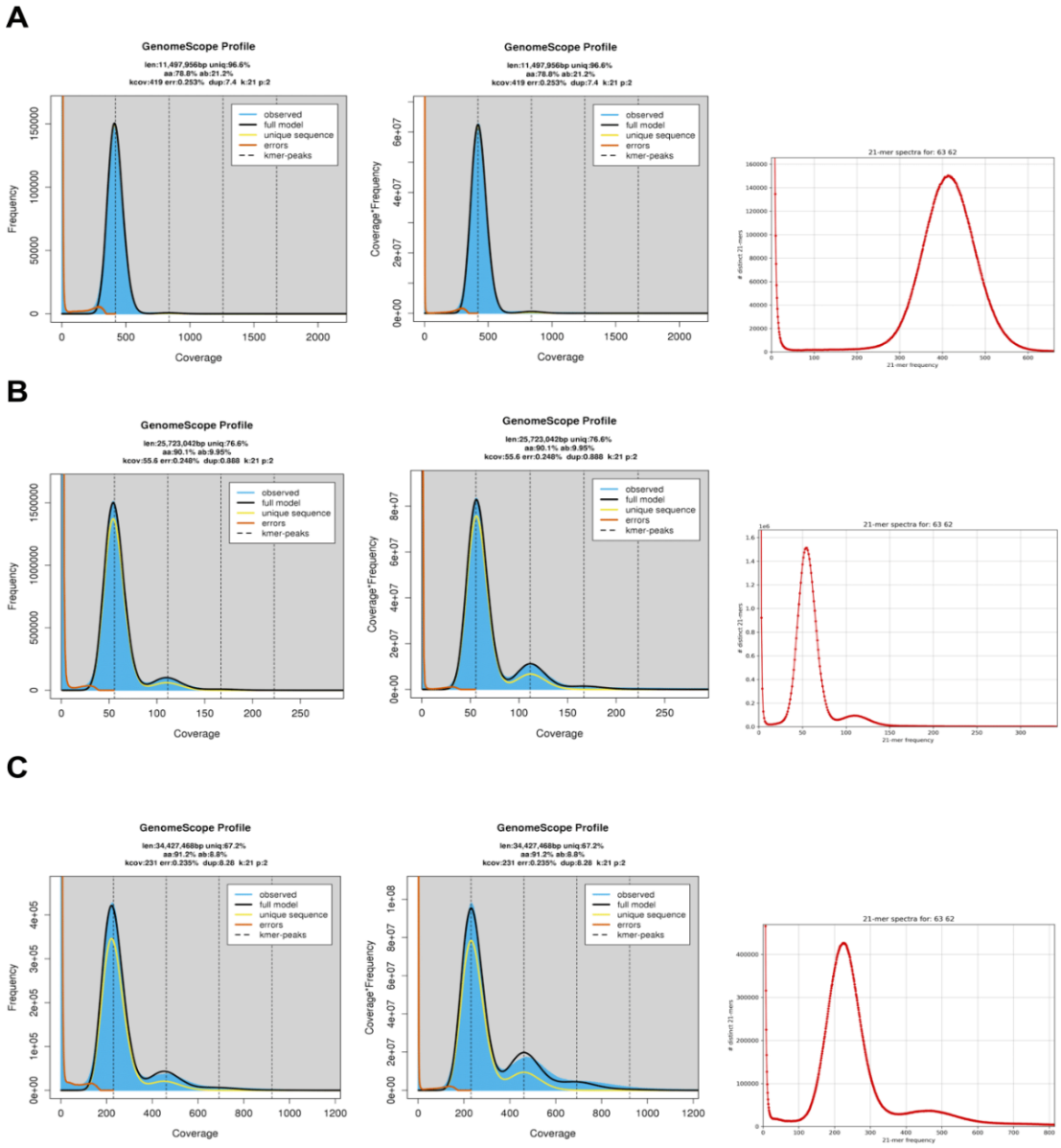


Figure 7. K-mer heterozygosity histograms were calculated first with jellyfish and then visualized using Genomescope2 and KAT. Described here are the three different ploidy we see in our 25 strains of *F. endolithicus*. **(A)** describes species CCFEE_5001, our results indicated a haploid ploidy as seen here with a singular shoulder. **(B)** describes species CCFEE_5195, our results indicated a diploid ploidy, here we see one long shoulder, and the second shoulder is a much smaller one. While (C) describes species CCFEE_5199, our results indicated a triploid ploidy. Here we see two shoulders like seen in the diploid species, but the second shoulder seems much larger and is seen around the 450-coverage range, unlike the diploid species.

Table 6. Average 21-k-mer average frequency details on Haploid, Diploid, and Triploid *F. endolithicus* details. Observing 21-mer average frequency details generated from GenomeScope v.2 for all 25 species of *Friedmanniomyces endolithicus,* averaged details on haploid, diploid, and triploid strains.

	Average Haploid	**Average Diploid**	**Average Polyploid**
Homozygous (aa) mean	99.91%	91.57%	92.46%
Heterozygous (Aa) mean	0.0822%	8.42%	7.53%
Genome Haploid Length	30,851,264.9bp	24,720,522bp	37,869,169.50bp
Genome Repeat Length	4,256,664.10bp	4,409,058.67bp	12,235,670.60bp
Genome Unique Length	26,594,601bp	20,253,253bp	25,633,493bp
Model Fit	94.38%	94.59%	89.95%
Error Rate	0.23%	0.32%	0.24%

Using Gaussian statistics to confirm ploidy

Generally, flow cytometry is the best tool to test and confirm the ploidy of microorganisms, but due to the highly melanized nature along with their slow growth, MCF flow cytometry isn't a common procedure. Instead, there are new tools that allow for estimating ploidy using WGS data including Gaussian statistics. Using the nQuire tool generated statistical results all described in **Table 7**. Statistics calculated include the "free model" seen in the first column, then "fixed models" of diploid, triploid and tetraploid in the next three columns. The final three columns are the subtracted product (or delta log-likelihoods) of the "free model" - "fixed model" for each ploidy. This is the "optimum model" for the

empirical data under the assumptions of the GMM model. The higher the delta log-likelihood the further it is from the optimum and thus not the proper ploidy to consider. Observing columns 5-7, the bolded values are the lowest optimum number and could be predicted as the ploidy of that genome. The only issue with this program is the haploid species are not represented here well. Prior informed haploid species were omitted from the analysis (CCFEE_5001, CCFEE_5486, and CCFEE_6074).

Based on **Table 7**, species CCFEE_5193 and CCFEE_5195 had the lowest delta log-likelihood as tetraploid. CCFEE_5199, CCFEE_5200, CCFEE_5273, and CCFEE_5277 had the lowest delta log-likelihood as a triploid. While CCFEE_524, CCFEE_5208, CCFEE_5275, CCFEE_5281, CCFEE_5283, CCFEE_5307, CCFEE_5311, CCFEE_5311_v1, CCFEE_6081, CCFEE_6082, CCFEE_6096, CCFEE_6249, CCFEE_6250, CCFEE_670, and CCFEE_690 had the lowest delta log-likelihood as a diploid.

Further statistical analysis was performed using the "histotest" program offered by nQuire. A positive correlation to detecting ploidy per genome is to observe a low SSR, a positive slope with low standard error, and a high r2. Results that fit within the empirical-optimized range are highlighted in yellow.

Results were identical to those seen with the other tool used and visualized in Table 7. Table 8 breaks down the calculations for each ploidy including the SRRs, y-slope with a standard error value, and the r^2 value. Species CCFEE_5193 and CCFEE_5195 had optimized results in the tetraploid analysis. CCFEE_5199, CCFEE_5200, CCFEE_5273,

and CCFEE_5277 had optimized results in the triploid analysis. While the remaining strains, CCFEE_524, CCFEE_5208, CCFEE_5275, CCFEE_5281, CCFEE_5283, CCFEE_5307, CCFEE_5311, CCFEE_5311_v1, CCFEE_6081, CCFEE_6082, CCFEE_6096, CCFEE_6249, CCFEE_6250, CCFEE_670, and CCFEE_690 had optimized results in the diploid analysis.

Table 7. Results obtained from Gaussian statistical analysis on *F. endolithicus* aligned genome bam files which helped deduce their ploidy. The free model testing was performed along with three fixed model calculations for each ploidy (diploid, triploid, and tetraploid), while the last three columns take the second column with the free model test, and subtract it against each fixed test to receive the delta log-likelihood results for each ploidy. The optimum ploidy results are highlighted in bold, as these results are the lowest in comparison to the other two delta values.

Species	Free model test	Diploid fixed test	Triploid fixed test	Tetraploid fixed test	Delta Diploid	Delta Triploid	Delta Tetraploid
CCFEE_5193	18763.92	9127.67	14710.70	16857.67	9636.24	4053.21	**1906.25**
CCFEE_5195	4838.08	6028.72	11635.99	13893.65	8809.35	3202.09	**944.42**
CCFEE_5199	1963500.55	1293285.58	1766312.73	1046265.68	670214.96	**197187.81**	917234.87
CCFEE_5200	2119418.38	1421463.2	1892344.43	1047831.54	697955.18	**227073.94**	1071586.84
CCFEE_5208	2486530.45	2301577.04	952385.26	1507767.96	**184953.41**	1534145.19	978762.49
CCFEE_524	1622090.14	1557052.71	702126.19	940283.922	**65037.42**	919963.95	681806.22
CCFEE_5273	1893652.37	1378359.48	1663442.97	1002263.59	515292.89	**230209.40**	891388.78
CCFEE_5275	1698275.87	1660764.34	948897.79	940351.37	**37511.53**	749378.07	757924.50
CCFEE_5277	2008509.46	1398361.75	1718271.7	978884.16	610147.70	**290237.76**	1029625.3
CCFEE_5281	1565030.52	1520189.9	704597.65	887399.97	**44840.61**	860432.86	677630.55
CCFEE_5283	1673589.34	1588303.1	689110.75	986580.94	**85286.24**	984478.59	687008.39
CCFEE_5307	1468457.3	1431789.78	697936.65	829125.96	**36667.51**	770520.65	639331.33

CCFEE_5311	1656796.13	1607025.88	721911.46	940491.59	**49770.24**	934884.66	716304.53
CCFEE_5311_v1	2592187.83	2490752.09	1047201.51	1489638.1	**101435.74**	1544986.31	1102549.73
CCFEE_6081	1629417.15	1569150.17	701615.74	937564.57	**60266.98**	927801.40	691852.57
CCFEE_6082	2193667.37	2079094.52	926193.78	1280789.98	**114572.85**	1267473.59	912877.39
CCFEE_6096	2207086.94	2080623.33	928371.73	1297533.56	**126463.60**	1278715.2	909553.37
CCFEE_6249	1435869.14	1405705.93	708485.47	797925.84	**30163.20**	727383.662	637943.29
CCFEE_6250	2564429.84	2351896.06	951921.32	1573994.84	**212533.78**	1612508.52	990434.99
CCFEE_6464	1040105.54	1022833.46	664460.93	586128.46	**17272.08**	375644.61	453977.07
CCFEE_670	1646623.92	1541414.09	687558.42	985476.46	**105209.82**	959065.49	661147.45
CCFEE_690	1634524.73	1572422.52	704714.18	945468.31	**62102.21**	929810.54	689056.41

Table 8. The second part of the Gaussian statistical analysis on *F. endolithicus* aligned genome bam files, thus indicating each of their ploidy. This analysis provides for each ploidy the sum of squared residuals (SRRs) of the empirical vs ideal histograms calculated before, the y-slope value along with its standard error value, and an r^2 of the regression of the y-values. A positive correlation to detecting ploidy per genome is to observe a low SSR, a positive slope with low standard error, and a high r^2. Results that fit within the empirical-optimized range are highlighted in yellow.

5193	Diploid:	5195	Diploid:	5199	Diploid:	5200	Diploid:	5208	Diploid:
	Norm SSR: 0.0452975		Norm SSR: 0.0494307		Norm SSR: 0.0493356		Norm SSR: 0.0501808		Norm SSR: 0.00615847
	y-y slope: -0.0534355, with std.Err: 0.0193057		y-y slope: -0.100182, with std.Err: 0.0204109		y-y slope: -0.0883858, with std.Err: 0.0285288		y-y slope: -0.0900431, with std.Err: 0.0333261		y-y slope: 0.612842, with std.Err: 0.0082057
	r^2: 0.114925		r^2: 0.289937		r^2: 0.139921		r^2: 0.110108		r^2: 0.989533
	Triploid:		**Triploid:**		**Triploid:**		**Triploid:**		**Triploid:**
	Norm SSR: 0.0210683		Norm SSR: 0.0212036		Norm SSR: 0.00853802		Norm SSR: 0.00760899		Norm SSR: 0.0502551
	y-y slope: -0.0312339, with std.Err: 0.02959		y-y slope: -0.0244664, with std.Err: 0.0351111		y-y slope: 0.333189, with std.Err: 0.0110952		y-y slope: 0.376632, with std.Err: 0.0154817		y-y slope: -0.42792, with std.Err: 0.102589
	r^2: 0.0185347		r^2: 0.00816283		r^2: 0.938593		r^2: 0.909346		r^2: 0.227738

	Tetraploid:		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**
	Norm SSR: 0.00504333		Norm SSR: 0.00583889		Norm SSR: 0.0147604		Norm SSR: 0.0171148		Norm SSR: 0.0204536
	y-y slope: 0.240392, with std.Err: 0.0343797		y-y slope: 0.214458, with std.Err: 0.0472431		y-y slope: -0.303125, with std.Err: 0.0574435		y-y slope: -0.408394, with std.Err: 0.059825		y-y slope: 0.162612, with std.Err: 0.180474
	r^2: 0.453157		r^2: 0.258857		r^2: 0.320636		r^2: 0.441292		$r^2$2: 0.0135734

524	**Diploid:**	**5273**	**Diploid:**	**5275**	**Diploid:**	**5277**	**Diploid:**	**5281**	**Diploid:**
	Norm SSR: 0.010519		Norm SSR: 0.0456392		Norm SSR: 0.0198022		Norm SSR: 0.0473856		Norm SSR: 0.0129466
	y-y slope: 0.505977, with std.Err: 0.0178265		y-y slope: -0.0483964, with std.Err: 0.0263989		y-y slope: 0.323068, with std.Err: 0.0248734		y-y slope: -0.0553148, with std.Err: 0.0344927		y-y slope: 0.463141, with std.Err: 0.02444
	r^2: 0.931762		r^2: 0.0538942		r^2: 0.740887		r^2: 0.0417683		r^2: 0.858888
	Triploid:		**Triploid:**		**Triploid:**		**Triploid:**		**Triploid:**

	Norm SSR: 0.0433037		Norm SSR: 0.0103459		Norm SSR: 0.0315071		Norm SSR: 0.00803601		Norm SSR: 0.0410926
	y-y slope: -0.354951, with std.Err: 0.0879217		y-y slope: 0.27224, with std.Err: 0.0174439		y-y slope: -0.184583, with std.Err: 0.066939		y-y slope: 0.364929, with std.Err: 0.0193166		y-y slope: -0.32294, with std.Err: 0.0848506
	r^2: 0.216451		r^2: 0.805001		r^2: 0.114163		r^2: 0.858142		r^2: 0.19712
	Tetraploid:		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**
	Norm SSR: 0.0190579		Norm SSR: 0.0144595		Norm SSR: 0.0164515		Norm SSR: 0.0168001		Norm SSR: 0.0196275
	y-y slope: -0.016819, with std.Err: 0.15459		y-y slope: -0.315586, with std.Err: 0.0457464		y-y slope: -0.193303, with std.Err: 0.107806		y-y slope: -0.389193, with std.Err: 0.0616888		y-y slope: -0.117555, with std.Err: 0.146602
	r^2: 0.000200583		r^2: 0.446481		r^2: 0.0516765		r^2: 0.402853		r^2: 0.0107807

5283	**Diploid:**	**5307**	**Diploid:**	**5311**	**Diploid:**	**5311_v1**	**Diploid:**	**6081**	**Diploid:**
	Norm SSR: 0.00867958		Norm SSR: 0.0141964		Norm SSR: 0.0115146		Norm SSR: 0.0090874		Norm SSR: 0.0111126

	y-y slope: 0.548027, with std.Err: 0.0145972		y-y slope: 0.431394, with std.Err: 0.0230583		y-y slope: 0.488447, with std.Err: 0.0209905		y-y slope: 0.543141, with std.Err: 0.0176215		y-y slope: 0.500808, with std.Err: 0.0219571
	r^2: 0.959823		r^2: 0.855753		r^2: 0.901747		r^2: 0.941529		r^2: 0.89814
	Triploid:		**Triploid:**		**Triploid:**		**Triploid:**		**Triploid:**
	Norm SSR: 0.0462096		Norm SSR: 0.0383997		Norm SSR: 0.0423334		Norm SSR: 0.0461578		Norm SSR: 0.0436974
	y-y slope: -0.391449, with std.Err: 0.0929414		y-y slope: -0.285857, with std.Err: 0.080147		y-y slope: -0.340063, with std.Err: 0.0868324		y-y slope: -0.389636, with std.Err: 0.0931509		y-y slope: -0.360612, with std.Err: 0.0884469
	r^2: 0.231162		r^2: 0.177368		r^2: 0.206323		r^2: 0.22872		r^2: 0.219817
	Tetraploid:		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**
	Norm SSR: 0.0196791		Norm SSR: 0.0183305		Norm SSR: 0.0194473		Norm SSR: 0.0202381		Norm SSR: 0.019875
	y-y slope: 0.0412379, with std.Err: 0.164902		y-y slope: -0.117211, with std.Err: 0.136697		y-y slope: -0.0679304, with std.Err: 0.151455		y-y slope: 0.006468, with std.Err: 0.165097		y-y slope: -0.0576807, with std.Err: 0.155684

	r^2: 0.00105884		r^2: 0.012308		r^2: 0.00339805		r^2: 2.60135e-05		r^2: 0.00232121

6082	**Diploid:**	**6096**	**Diploid:**	**6249**	**Diploid:**	**6250**	**Diploid:**	**6464**	**Diploid:**
	Norm SSR: 0.00939822		Norm SSR: 0.00909184		Norm SSR: 0.0150068		Norm SSR: 0.00529174		Norm SSR: 0.0268176
	y-y slope: 0.525537, with std.Err: 0.0128159		y-y slope: 0.531704, with std.Err: 0.0115346		y-y slope: 0.407576, with std.Err: 0.0201538		y-y slope: 0.642381, with std.Err: 0.00855448		y-y slope: 0.243326, with std.Err: 0.0406425
	r^2: 0.966103		r^2: 0.972984		r^2: 0.873927		r^2: 0.989645		r^2: 0.377925
	Triploid:		**Triploid:**		**Triploid:**		**Triploid:**		**Triploid:**
	Norm SSR: 0.0443371		Norm SSR: 0.0446731		Norm SSR: 0.0357219		Norm SSR: 0.0523635		Norm SSR: 0.0289169
	y-y slope: -0.370522, with std.Err: 0.0890953		y-y slope: -0.37446, with std.Err: 0.0897567		y-y slope: -0.244001, with std.Err: 0.0762631		y-y slope: -0.444076, with std.Err: 0.10784		y-y slope: -0.0993131, with std.Err: 0.0738787
	r^2: 0.226685		r^2: 0.2278		r^2: 0.147849		r^2: 0.223248		r^2: 0.0297181

	Tetraploid:		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**		**Tetraploid:**
	Norm SSR: 0.0184127		Norm SSR: 0.018301		Norm SSR: 0.0172089		Norm SSR: 0.021331		Norm SSR: 0.0179708
	y-y slope: 0.0530793, with std.Err: 0.157552		y-y slope: 0.0722511, with std.Err: 0.15871		y-y slope: -0.115581, with std.Err: 0.12771		y-y slope: 0.202395, with std.Err: 0.188627		y-y slope: -0.2502, with std.Err: 0.112107
	$r^{2:}$ 0.00192007		r^2: 0.00350031		r^2: 0.0136924		r^2: 0.0191402		r^2: 0.0778498

670	**Diploid:**	**690**	**Diploid:**
	Norm SSR: 0.00836395		Norm SSR: 0.0105434
	y-y slope: 0.549465, with std.Err: 0.0100704		y-y slope: 0.507145, with std.Err: 0.0186455
	r^2: 0.980567		r^2: 0.926139
	Triploid:		**Triploid:**

	Norm SSR: 0.0453777		Norm SSR: 0.0434456
	y-y slope: -0.374743, with std.Err: 0.0931405		y-y slope: -0.355588, with std.Err: 0.0884779
	r^2: 0.215299		r^2: 0.214924
	Tetraploid:		**Tetraploid:**
	Norm SSR: 0.0186312		Norm SSR: 0.0193417
	y-y slope: 0.0955685, with std.Err: 0.163189		y-y slope: -0.0274596, with std.Err: 0.155391
	r^2: 0.00577934		r^2: 0.000528997

SNP biallelic analysis confirms ploidy for each of 25 strains of F. endolithicus

Bialleleic frequencies were calculated using the program GATK then visualized using two visual methods. VCF data along with metadata provided by our collaborators is visualized into correlated data (**Figure 8).** We visualized 25 strains of *F. endolithicus* using their biallelic SNP data ploidy, year collected, and elevation. The coordination biplots indicate principal components (PCs), PC1 and PC2 to both be at 50%. We see some inferences of that with the diploid strains (**Figure 8A**). There are two haploid outliers CCFEE_5486 and CCFEE_5001, while CCFEE_6074 is grouped nearly identically to a cluster of diploids. The four triploid species CCFEE_5199, CCFEE_5200, CCFEE_5273, and CCFEE_5277 are also grouped tightly with other diploids. Two outliers for diploids include CCFEE_5311_v1 and CCFEE_5193, where CCFEE_5311_v1 clusters closer to CCFEE_5486. Adding a layer of metadata complexity allows for other patterns to be visualized. When adding the years metadata to our ploidy analysis (**Figure 8B)** we can visualize ploidy and correlate it to the years *F. endolithicus* was collected from Antarctic endolithic communities. We see a tight clustering of species collected from 1997, 2004, 2010, and 2011. The initial collections from 1981, specifically CCFEE_670 and CCFEE_690 both cluster among three different time collections, while CCFEE_524 clusters are apart but closer to 1997 and 2004. An attempt to use elevation data to visualize ploidy data was attempted (**Figure 8C**). The best clustering groups are those collected from 1000m, 2000m, and 3000m elevations. There still seems to be variability in these results and outliers. Elevation has some hold, but the years collected hold a higher effect on the 25 strains than other metadata used in this analysis.

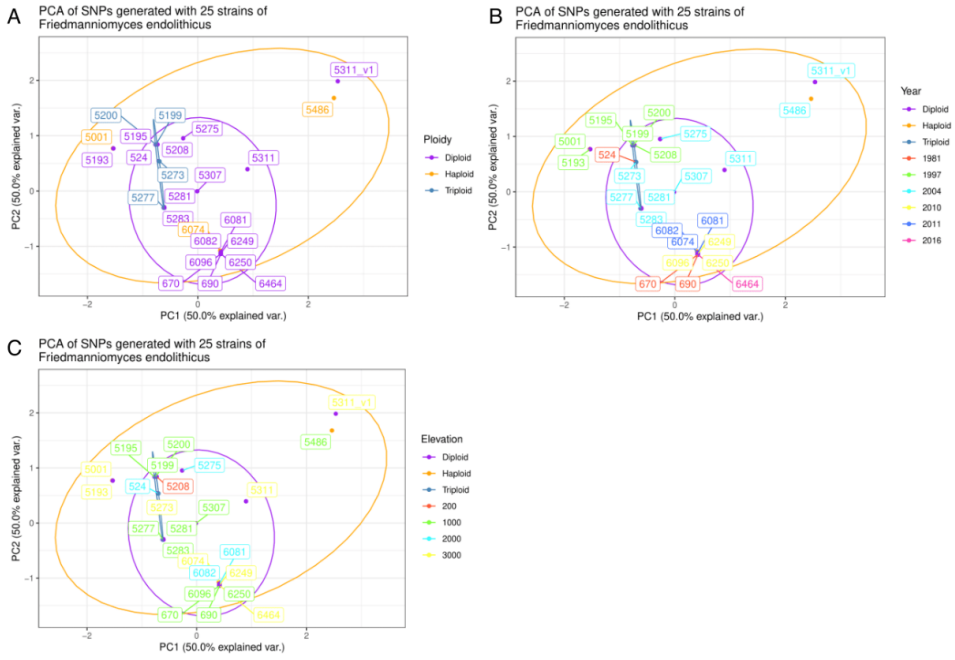


Figure 8. PCA plots were drawn to characterize metadata of collected *F. endolithicus* along with ploidy information. (A) Draws out the different ploidy seen in the biallelic SNP data. Purple indicates diploids, haploids are visualized in orange, while triploids are seen in blue. (B) Taking the same data of diploids, haploids, and triploids an added layer of the years collected is combined into this PCA plot The colors indicate years, pink is the latest collection from 2016, navy blue is from 2011, yellow is from 2010, teal blue from 2004, lime green from 1997, and red are from the earliest collection of 1981. (C) Takes the same diploid, haploid, and triploid data to add the elevation of the different strains collected in this workThe colors indicate elevation, red is 200m from the ocean level, lime green is 1000m from the ocean level, teal blue is 2000m from the ocean level, and yellow is 3000m from the ocean level.

The same biallelic SNP data was used to generate histograms of frequency counts, (**Figure 9).** Three main visuals are seen in these groups of figures, specifically of haploid (CCFEE_5001, CCFEE_5486, CCFEE_6074), triploid (CCFEE_5199, CCFEE_5200, CCFEE_5273, CCFEE_5277), with the remainder resembling diploid frequencies. A second iteration of the histograms was generated, but this time omitting the homozygous frequencies to better show the middle shoulders. This was added on top of each figure post R analysis.

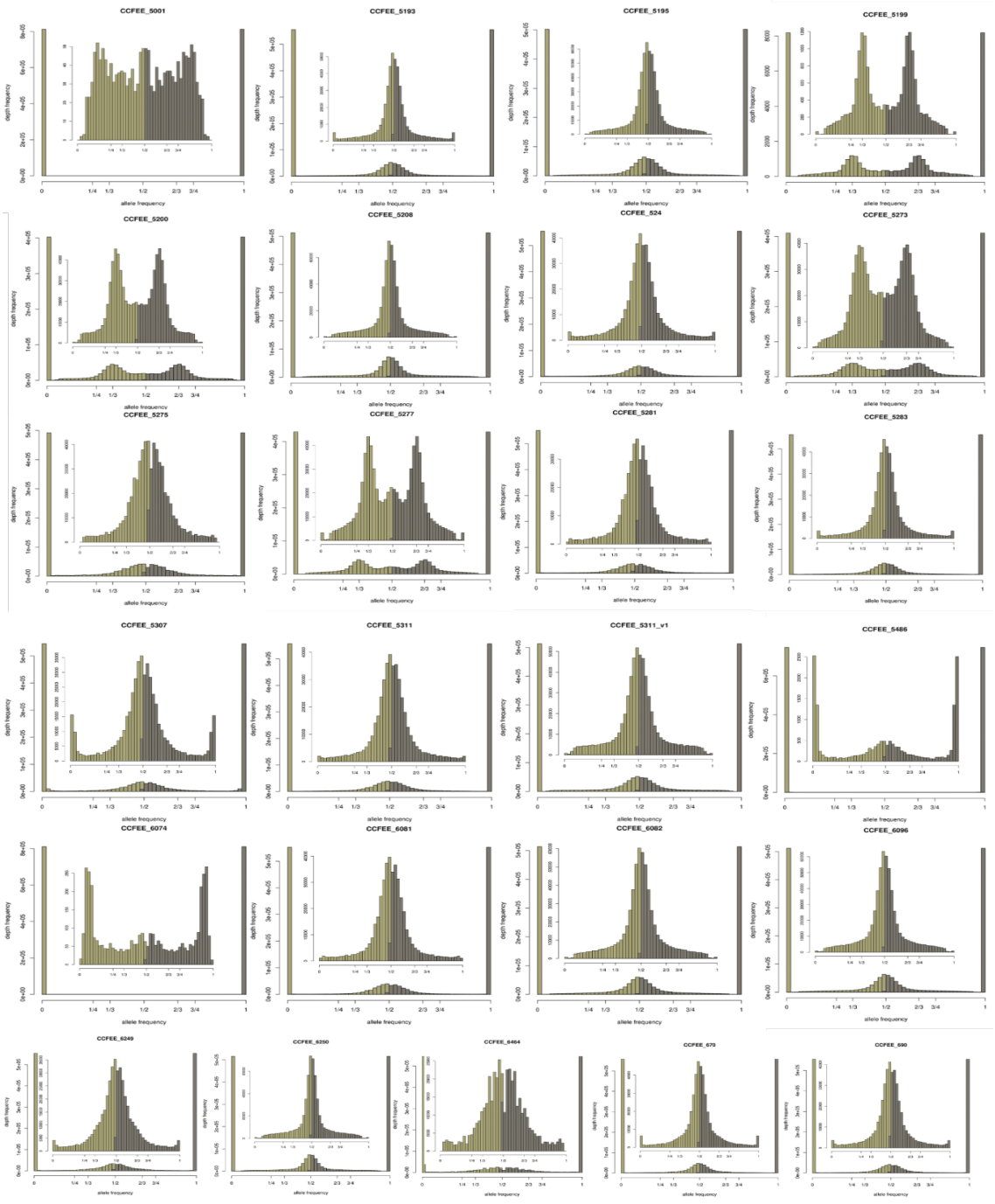


Figure 9. Ploidy analysis using allele frequency determined in the VCF output from GATK indicates the variety of ploidy seen in all 25 strains of *F. endolithicus*. Allele frequencies for every 25 strains of *F. endolithicus* are mapped along with the homozygous allele frequencies. A second visual of the frequencies is zoomed in and mapped without the homozygous allele and superimposed onto the same figure. Species frequencies that fall hard in the ½ mark indicate diploid species, while species with ⅓ and ⅔ marks indicate a triploid species.

Copy number variation shows no significant aneuploidies present

A final assessment was done to observe genome-based anomalies through a 10kb window sliding view to observe for chromosomal abnormalities. In order to have a visible understanding of each genome's patterns, the 25 species were split into three groups of eight (**Figure 11**). Visually we do not detect any patterns that could help differentiate these species based on their ploidy. The final species, CCFEE_5311_v1 was compared to the re-sequenced CCFEE_5311 and was observed for any major differences (**Figure 10**).

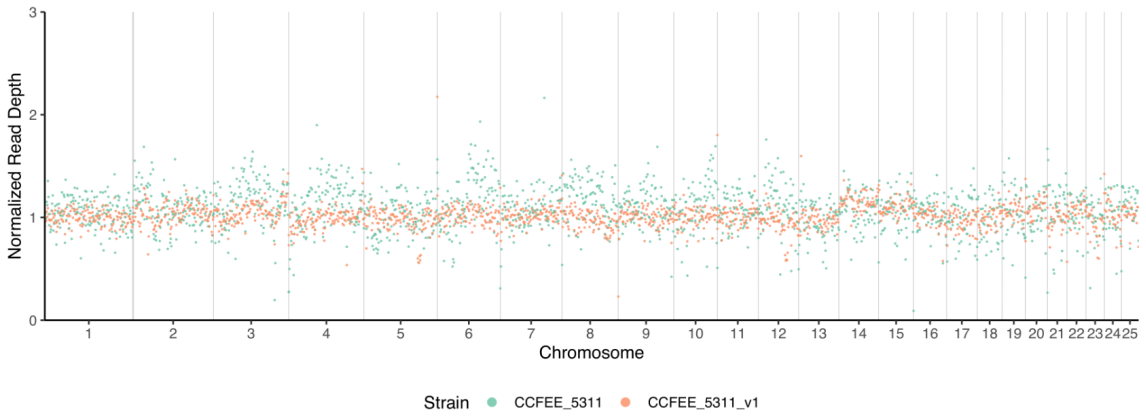


Figure 10. Using a 10kbp sliding window view to observe two nearly identical species of *F. endolithicus* and showing minor differences across all 25 genomes. Both sequenced versions of *F. endolithicus 5311* were plotted and visualized to observe for large chromosomal differences that could be spacing the two apart from each other in the PCA plots.

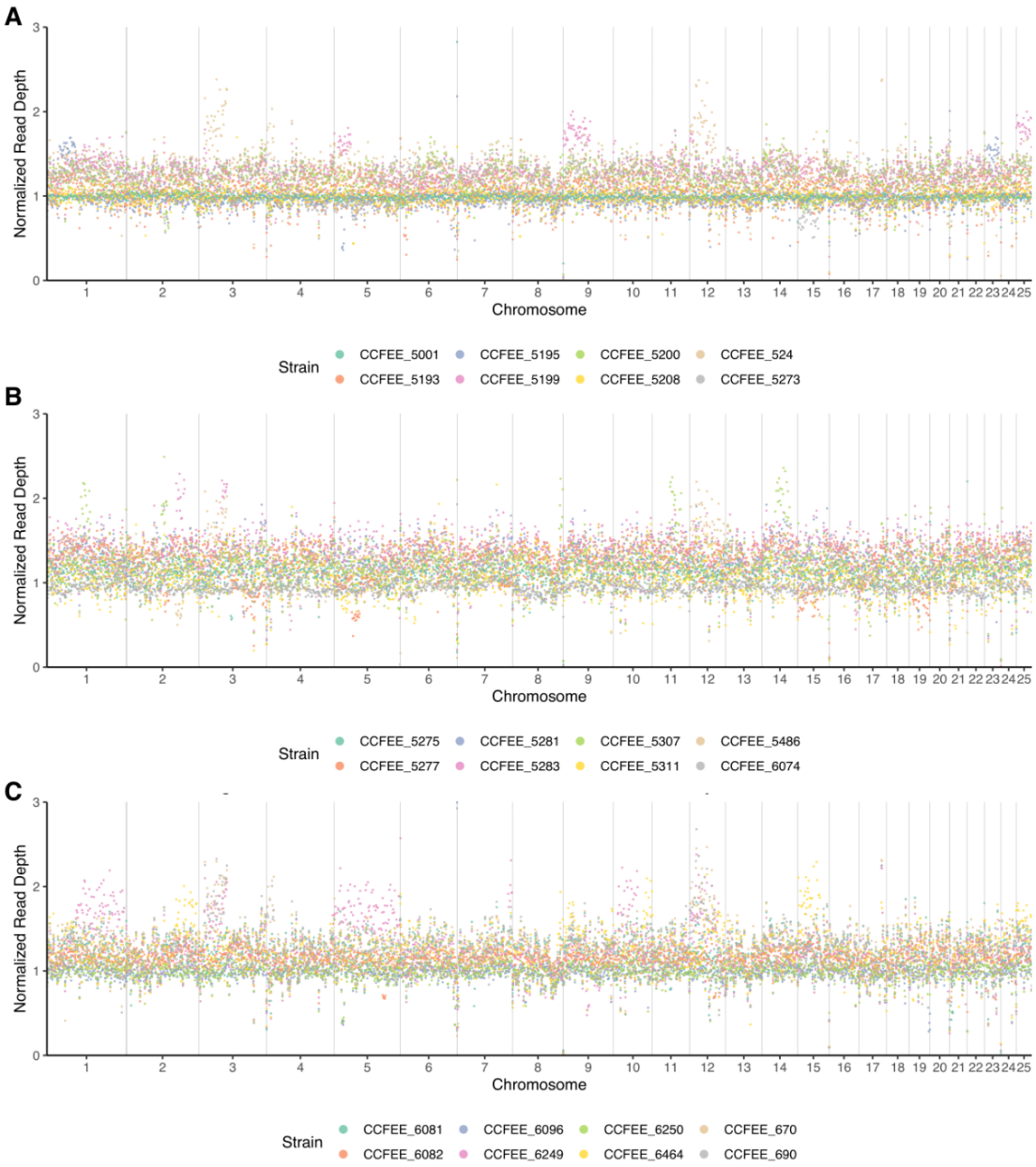


Figure 11. Copy number variation (CNV) plot depicting all 25 species of *F. endolithicus* separated into three figures. 10kbp sliding window view of all 25 species of *F. endolithicus* split into three graphs. These figures' main focus was to show there are no aneuploidies or other chromosomal issues that could account for the ploidy count we are experiencing with our other series of analyses. (**A**) Depicts species CCFEE_5001, CCFEE_5193, CCFEE_5195, CCFEE_5199, CCFEE_5200, CCFEE_524, CCFEE_5208, and CCFEE_5273. (**B**) Depicts species CCFEE_5275, CCFEE_5277, CCFEE_5281, CCFEE_5283, CCFEE_5307, CCFEE_5311, CCFEE_5486, and CCFEE_6074. The mapping does not show any anomalies, indicating no large chromosomal change. (C) Depicts the last 8 CCFEE_6081, CCFEE_6082, CCFEE_6096, CCFEE_6249, CCFEE_6250, CCFEE_6464, CCFEE_670, and CCFEE_690.

Discussion

We present an analysis of the twenty-six *Friedmanniomyces* genomes and surrounding micro-colonial fungi to determine ploidy structure, heterozygosity, and evolution through fossil time predictions, population genomics, and comparative genomics.

Endemic Friedmanniomyces spp. collected over 40-year period

Collaborative efforts with scientists allowed us to have a large sample collection to process and analyze endemic species from Antarctica. As a result of my work, there is a collection of twenty-five species of *Friedmanniomyces endolithicus* (**Table 1**) and one species of *Friedmanniomyces simplex* species available for further study. The limitations of Illumina become obvious once species with highly repetitive elements are sequenced and become difficult to assemble to the chromosomal level (**Table 4**). With more species to be sequenced, along with transcriptomics of this species will help unlock more unknown adaptations *Friedmanniomyces spp.* may have when it comes to surviving the extremes of the South Pole.

Friedmanniomyces spp. diverged when the Antarctic reached the South Pole

We used a molecular clock approach to place the evolutionary timescale of the endemic Antarctic fungi *F. endolithicus* and *F. simplex* to diverge 48 million years ago. The chronogram also compared the two strains to 13 other fungal species that fall within the Dothideomycetes, Eurotiomycetes, Sordariomycetes, and Saccharomycotina orders. Our tree-building efforts combined with fossil records have placed the divergence of *Friedmanniomyces* species to be in the scope of when the landmass of Antarctica settled

in the region it is found in today. Our chronogram (**Figure 6**) indicates the two *Friedmanniomyces* species diverged about 48.42 million years ago. The remaining Dothideomycetes, Eurotiomycetes, Sordariomycetes, and Saccharomyces strains were used to help orient both *Friedmanniomyces* species genealogically. Cryptoendolithic communities are protected from the harshness of the extreme cold and desiccation of the Antarctic tundra. As the Antarctic split away from Pangea and settled to the bottom pole of the planet, the temperature shifts disallowed for fungal survival and instead forced them to adapt and hide within rocks (Friedmann and Weed 1987). The cold deserts of the McMurdo Dry Valleys are extremely stressful to microorganismal growth and could be the reason as to only having endemic to Antarctica MCF species, like *Friedmanniomyces*.

Ploidy estimation indicates different stages of ploidy found in 25 species of Friedmanniomyces

Whole genome duplication events are correlated in other fungi with stress adaptations. *Hortaea werneckii* along with *Aureobasidium melanogenum*, were recently studied for their ploidy differences within a large population (Gostinčar et al. 2022). The first indication of a difference in ploidy was after the assembly of each genome and observing each genome length. The genome lengths of haploids fall around 20Mbp, diploids around 40Mbp, and triploids around 60Mbp. Observing the Kmer heterozygosities for each species also indicated a higher number of repetitive elements present in triploid species (about 40% of its genome), compared to diploids (30%) and haploids (5%). Long-read sequencing, along with a hybrid assembly of long-read and short-read sequencing could easily help answer these repetitive element questions and may bring clarity to all ploidy statuses. The

nQuire Gaussian statistics helped confirm some of the multiple ploidy species but brought some uncertainty to our haploid species. The results for two species, CCFEE_5001 and CCFEE_5193 indicate a tetraploidy. Further research in understanding tetraploidy indicates much difficulty to obtain confirmation unless flow cytometry is performed on these species. We fear the high melanin nature may provide increased difficulties with this process and might require preparing melanin knock-out strains to properly visualize these two species. Using the SNP biallelic frequency generated data was another measure to help determine the ploidy of our strains (**Figure 9**). I wanted to see correlations to the years and elevation the strains were isolated from (**Figure 8**). It is interesting to see clusters forming around particular years, it would be important to assess the temperature ranges of the years each sample was collected.

MAT Loci

All 25 species of *Friedmanniomyces* contained both MAT1-1 and MAT1-2 genes, indicating their homothallic nature. The presence of multiple mating type locus in diploid and triploid species with copies of alternative MAT genes (MAT1-1-10 and MAT1-2-12) has me thinking about the role of multiple loci and the benefits of having the LINE retrotransposon-induced genes included. This self-mating nature could allow for a rise of multi-ploidy organisms (Ene and Bennett 2014) and could be what I am observing. Whole genome duplications and polyploidization events are rare in fungi but have previously been observed in micro-colonial fungi, specifically *H. werneckii*, where it was described to undergo interspecific hybridization (Sinha et al. 2017; Gostinčar et al. 2018; Zalar et al. 2019). It is possible that a similar event is occurring in these *Friedmanniomyces* strains.

The trouble with suggesting large genome duplications for ploidy comes with the stress and energy of having to maintain such a large genome. Usually, after duplication events, large-scale gene deletions, reductions, and genome shuffling occurs and can be a method to detect if a whole genome duplication has occurred (Campbell et al. 2016) . Though their slow-growing nature is something to consider and may in fact provide the micro-colonial fungi with all the resources it needs to grow once the environment permits (Gostinčar et al. 2022) without the need for replicating often.

Conclusion

Cryptoendolithic communities are incredibly important resources for extremeotolerant species. Extreme-adapted microorganisms develop duplicate gene burdens as a means of survival. Tracking fungal speciation using fossil data helped direct a time when the speciation event could have occurred, which coincided with the time of large continental shifts to the areas we know today. Our population of *Friedmanniomyces endolithicus* has been collected from a variety of years and locations in Antarctica. This population has a variety of ploidies and it may be a way to protect itself from the extremes Antarctica offers. Further studies on the genome evolution and duplication of micro-colonial fungi may enable a deeper understanding of specific genetic factors that govern their success in extreme environments; specifically looking into niche adaptation, fitness, and evolution.

Supplementary Materials

Data and Analysis can be found here:
DOI: 10.5281/zenodo.733921

References

Albanese D, Coleine C, Rota-Stabelli O, Onofri S, Tringe SG, Stajich JE, Selbmann L, Donati C. 2020. Antarctic cryptoendolithic bacterial lineages of pre-Cambrian origin as proxy for Mars colonization. bioRxiv. doi:10.1101/2020.02.27.967604.

Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. Proc Biol Sci. 279(1738):2497–2509.

Anderson CA, Roberts S, Zhang H, Kelly CM, Kendall A, Lee C, Gerstenberger J, Koenig AB, Kabeche R, Gladfelter AS. 2015. Ploidy variation in multinucleate cells changes under stress. Mol Biol Cell. 26(6):1129–1140.

Archer SDJ, de los Ríos A, Lee KC, Niederberger TS, Cary SC, Coyne KJ, Douglas S, Lacap-Bugler DC, Pointing SB. 2017. Endolithic microbial diversity in sandstone and granite from the McMurdo Dry Valleys, Antarctica. Polar Biol. 40(5):997–1006.

Arzanlou M, Crous PW, Zwiers L-H. 2010. Evolutionary dynamics of mating-type loci of *Mycosphaerella spp.* occurring on banana. Eukaryot Cell. 9(1):164–172.

Askin RA. 2013. Late cretaceous-early tertiary antarctic outcrop evidence for past vegetation and climates. In: The Antarctic Paleoenvironment: A Perspective on Global Change: Part One. Washington, D. C.: American Geophysical Union. (Antarctic research series). p. 61–74.

Aylward J, Havenga M, Dreyer LL, Roets F, Wingfield BD, Wingfield MJ. 2020. Genomic characterization of mating type loci and mating type distribution in two apparently asexual plantation tree pathogens. Plant Pathol. 69(1):28–37.

Aylward J, Havenga M, Wingfield BD, Wingfield MJ, Dreyer LL, Roets F, Steenkamp ET. 2022. Novel mating-type-associated genes and gene fragments in the genomes of Mycosphaerellaceae and Teratosphaeriaceae fungi. Mol Phylogenet Evol. 171:107456.

Beadle GW, Coonradt VL. 1944. Heterocaryosis in *Neurospora Crassa*. Genetics. 29(3):291–308.

Blakeslee AF. 1976. Sexual Reproduction in the Mucorineae. J. Cramer.

Bleichrodt R-J, Read ND. 2019. Flow cytometry and FACS applied to filamentous fungi. Fungal Biol Rev. 33(1):1–15.

Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH. 2004. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci U S A. 101(6):1632–1637.

Campbell MA, Ganley ARD, Gabaldón T, Cox MP. 2016. The Case of the Missing Ancient Fungal Polyploids. Am Nat. 188(6):602–614.

Carter-House D, E Stajich J, Unruh S, Kurbessoian T. 2020. Fungal CTAB DNA Extraction v1. doi:10.17504/protocols.io.bhx8j7rw.

Cary SC, McDonald IR, Barrett JE, Cowan DA. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol. 8(2):129–138.

Chen Z, Martinez DA, Gujja S, Sykes SM, Zeng Q, Szaniszlo PJ, Wang Z, Cuomo CA. 2014. Comparative genomic and transcriptomic analysis of *Wangiella dermatitidis*, a major cause of phaeohyphomycosis and a model black yeast human pathogen. G3. 4(4):561–578.

Coleine C, Biagioli F, de Vera JP, Onofri S, Selbmann L. 2021. Endolithic microbial composition in Helliwell Hills, a newly investigated Mars-like area in Antarctica. Environ Microbiol. 23(7):4002–4016.

Coleine C, Gevi F, Fanelli G, Onofri S, Timperio AM, Selbmann L. 2020. Specific adaptations are selected in opposite sun exposed Antarctic cryptoendolithic communities as revealed by untargeted metabolomics. PLoS One. 15(5):e0233805.

Coleine C, Masonjones S, Selbmann L, Zucconi L, Onofri S, Pacelli C, Stajich JE. 2017. Draft Genome Sequences of the Antarctic Endolithic Fungi *Rachicladosporium antarcticum* CCFEE 5527 and *Rachicladosporium sp*. CCFEE 5018. Genome Announc. 5(27). doi:10.1128/genomeA.00397-17.

Coleine C, Masonjones S, Sterflinger K, Onofri S, Selbmann L, Stajich JE. 2020. Peculiar genomic traits in the stress-adapted cryptoendolithic Antarctic fungus *Friedmanniomyces endolithicus.* Fungal Biol. 124(5):458–467.

Coleine C, Selbmann L, Masonjones S, Onofri S, Zucconi L, Stajich JE. 2019. Draft Genome Sequence of an Antarctic Isolate of the Black Yeast Fungus *Exophiala mesophila*. Microbiol Resour Announc. 8(19). doi:10.1128/MRA.00142-19.

Coleine C, Stajich JE, de Los Ríos A, Selbmann L. 2021. Beyond the extremes: Rocks as ultimate refuge for fungi in drylands. Mycologia. 113(1):108–133.

Coleine C, Stajich JE, Zucconi L, Onofri S, Pombubpa N, Egidi E, Franks A, Buzzini P, Selbmann L. 2018. Antarctic Cryptoendolithic Fungal Communities Are Highly Adapted and Dominated by Lecanoromycetes and Dothideomycetes. Front Microbiol. 9:1392.

Coppin E, Debuchy R, Arnaise S, Picard M. 1997. Mating types and sexual development in filamentous ascomycetes. Microbiol Mol Biol Rev. 61(4):411–428.

Crame JA. 1992. Evolutionary history of the polar regions. Hist Biol. 6(1):37–60.

Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, et al. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science. 317(5843):1400–1402.

Dewannieux M, Heidmann T. 2005. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. Cytogenet Genome Res. 110(1-4):35–48.

Egidi E, de Hoog GS, Isola D, Onofri S, Quaedvlieg W, de Vries M, Verkley GJM, Stielow JB, Zucconi L, Selbmann L. 2014. Phylogeny and taxonomy of meristematic rock-inhabiting black fungi in the Dothideomycetes based on multi-locus phylogenies. Fungal Divers. 65(1):127–165.

Ene IV, Bennett RJ. 2014. The cryptic sexual strategies of human fungal pathogens. Nat Rev Microbiol. 12(4):239–251.

Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 24(4):363–367.

Fanelli G, Coleine C, Gevi F, Onofri S, Selbmann L, Timperio AM. 2021. Metabolomics of Dry Versus Reanimated Antarctic Lichen-Dominated Endolithic Communities. Life. 11(2). doi:10.3390/life11020096. http://dx.doi.org/10.3390/life11020096.

Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, Issel-Tarver L, Schwartz K, Sethuraman A, Botstein D, Cherry JM, et al. 2006. Saccharomyces cerevisiae S288C genome annotation: a working hypothesis. Yeast. 23(12):857–865.

Fitzgerald P. 2002. Tectonics and landscape evolution of the Antarctic plate since the breakup of Gondwana, with an emphasis on the West Antarctic Rift System. Royal Society of New Zealand Bulletin.

Friedmann EI. 1982. Endolithic microorganisms in the antarctic cold desert. Science. 215(4536):1045–1053.

Friedmann EI, Kappen L, Meyer MA, Nienow JA. 1993. Long-term productivity in the cryptoendolithic microbial community of the Ross Desert, Antarctica. Microb Ecol. 25(1):51–69.

Friedmann EI, Weed R. 1987. Microbial trace-fossil formation, biogenous, and abiotic weathering in the Antarctic cold desert. Science. 236(4802):703–705.

Fulton TM, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep. 13(3):207–209.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature. 422(6934):859–868.

Garnjobst L, Wilson JF. 1956. Heterocaryosis and Protoplasmic Incompatibility in *Neurospora crassa.* Proceedings of the National Academy of Sciences. 42(9):613–618.

Gearty W, AEgit, Yu G, Guidoni K. 2022. willgearty/deeptime:*v0.3.0.* https://zenodo.org/record/7308261.

Gerhardt HC. 1994. Reproductive character displacement of female mate choice in the grey treefrog, *Hyla chrysoscelis*. Anim Behav. 47(4):959–969.

Gerstein AC. 2013. Mutational effects depend on ploidy level: all else is not equal. Biol Lett. 9(1):20120614.

Gerstein AC, Lim H, Berman J, Hickman MA. 2017. Ploidy tug-of-war: Evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. Evolution. 71(4):1025–1038.

Glass NL, Grotelueschen J, Metzenberg RL. 1990. *Neurospora crassa* A mating-type region. Proc Natl Acad Sci U S A. 87(13):4912–4916.

Glass NL, Lee L. 1992. Isolation of *Neurospora crassa A* mating type mutants by repeat induced point (RIP) mutation. Genetics. 132(1):125–133.

Gostinčar C, Stajich JE, Zupančič J, Zalar P, Gunde-Cimerman N. 2018. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. BMC Genomics. 19(1):364.

Gostinčar C, Sun X, Černoša A, Fang C, Gunde-Cimerman N, Song Z. 2022. Clonality, inbreeding, and hybridization in two extremotolerant black yeasts. Gigascience. 11. doi:10.1093/gigascience/giac095.

Gostinčar C, Zalar P, Gunde-Cimerman N. 2022. No need for speed: slow development of fungi in extreme environments. Fungal Biol Rev. 39:1–14.

Gueidan C, Ruibal C, de Hoog GS, Schneider H. 2011. Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. Fungal Biol. 115(10):987–996.

Haber JE. 1998. Mating-type gene switching in *Saccharomyces cerevisiae*. Annu Rev Genet. 32:561–599.

Halacka K, Lusková V. 2000. Polyploidy in *Carassius auratus* in the lower reaches of the river Dyje-determination using the size of erythrocyte nuclei. In: [4. Czech Ichthyological Conference], Vodnany (Czech Republic), 10-12 May 2000.

Havenga M, Wingfield BD, Wingfield MJ, Roets F, Dreyer LL, Tatham CT, Duong TA, Wilken PM, Chen S, Aylward J. 2020. Mating strategy and mating type distribution in six global populations of the Eucalyptus foliar pathogen *Teratosphaeria destructans*. Fungal Genet Biol. 137:103350.

Heitman J, Kronstad JW, Casselton LA, Taylor JW. 2007. Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press.

Hofmann GE, Buckley BA, Airaksinen S, Keen JE, Somero GN. 2000. Heat-shock protein expression is absent in the antarctic fish *Trematomus bernacchii* (family Nototheniidae). J Exp Biol. 203(Pt 15):2331–2339.

Hughes KW, Petersen RH, Lodge DJ, Bergemann SE, Baumgartner K, Tulloss RE, Lickey E, Cifuentes J. 2013. Evolutionary consequences of putative intra-and interspecific hybridization in agaric fungi. Mycologia. 105(6):1577–1594.

Ibrahim AS, Magee BB, Sheppard DC, Yang M, Kauffman S, Becker J, Edwards JE Jr, Magee PT. 2005. Effects of ploidy and mating type on virulence of *Candida albicans*. Infect Immun. 73(11):7366–7374.

Idnurm A. 2011. Sex and speciation: the paradox that non-recombining DNA promotes recombination. Fungal Biol Rev. 25(3):121–127.

Kis-Papo T, Weig AR, Riley R, Peršoh D, Salamov A, Sun H, Lipzen A, Wasser SP, Rambold G, Grigoriev IV, et al. 2014. Genomic adaptations of the halophilic Dead Sea filamentous fungus *Eurotium rubrum*. Nat Commun. 5:3745.

Knaus BJ, Grünwald NJ. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour. 17(1):44–53.

Lengeler KB, Cox GM, Heitman J. 2001. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect Immun. 69(1):115–122.

Li Y, Shen H, Zhou Q, Qian K, van der Lee T, Huang S. 2017. Changing Ploidy as a Strategy: The Irish Potato Famine Pathogen Shifts Ploidy in Relation to Its Sexuality. Mol Plant Microbe Interact. 30(1):45–52.

Lin X, Heitman J. 2014. Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism. In: Sex in Fungi. Washington, DC, USA: ASM Press. p. 35–57.

Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA, Swofford DL, Hibbett D, et al. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. Nat Commun. 9(1):5451.

Lyle M, Gibbs S, Moore TC, Rea DK. 2007. Late Oligocene initiation of the Antarctic Circumpolar Current: Evidence from the South Pacific. Geology. 35(8):691–694.

Magnan C, Yu J, Chang I, Jahn E, Kanomata Y, Wu J, Zeller M, Oakes M, Baldi P, Sandmeyer S. 2016. Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity. PLoS One. 11(9):e0162363.

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 33(4):574–576.

Marcais G, Kingsford C. 2012. Jellyfish: A fast k-mer counter. [accessed 2022 Nov 16].

Margarido GRA, Heckerman D. 2015. ConPADE: genome assembly ploidy estimation from next-generation sequencing data. PLoS Comput Biol. 11(4):e1004229.

Maritati A, Danišík M, Halpin JA, Whittaker JM, Aitken ARA. 2020. Pangea Rifting Shaped the East Antarctic Landscape. Tectonics. 39(8):e2020TC006180.

Metzenberg RL, Glass NL. 1990. Mating type and mating strategies in Neurospora. Bioessays. 12(2):53–59.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 37(5):1530–1534.

Muller H, Thierry A, Coppée J-Y, Gouyette C, Hennequin C, Sismeiro O, Talla E, Dujon B, Fairhead C. 2009. Genomic polymorphism in the population of *Candida glabrata*: gene copy-number variation and chromosomal translocations. Fungal Genet Biol. 46(3):264–276.

Ngoot-Chin T, Zulkifli MA, van de Weg E, Zaki NM, Serdari NM, Mustaffa S, Zainol Abidin MI, Sanusi NSNM, Smulders MJM, Low ETL, et al. 2021. Detection of ploidy and chromosomal aberrations in commercial oil palm using high-throughput SNP markers. Planta. 253(2):63.

Nienow JA, Friedmann EI, Ocampo-Friedmann R. 2003. Endolithic Microorganisms in Arid Regions. In: Encyclopedia of Environmental Microbiology. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, et al. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature. 438(7071):1151–1156.

Omelon CR. 2008. Endolithic Microbial Communities in Polar Desert Habitats. Geomicrobiol J. 25(7-8):404–414.

Onofri, Pagano, Zucconi, Tosi. 1999. *Friedmanniomyces endolithicus* (Fungi, Hyphomycetes), anam-gen and sp nov, from continental Antarctica. Nova Hedwigia.

Onofri S, Selbmann L, Pacelli C, Zucconi L, Rabbow E, de Vera J-P. 2019. Survival, DNA, and Ultrastructural Integrity of a Cryptoendolithic Antarctic Fungus in Mars and Lunar Rock Analogs Exposed Outside the International Space Station. Astrobiology. 19(2):170–182.

Onofri S, Selbmann L, Zucconi L, Pagano S. 2004. Antarctic microfungi as models for exobiology. Planet Space Sci. 52(1):229–237.

Onofri S, Zucconi L, Isola D, Selbmann L. 2014. Rock-inhabiting fungi and their role in deterioration of stone monuments in the Mediterranean area. Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology. 148(2):384–391.

Pawlowska TE, Taylor JW. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. Nature. 427(6976):733–737.

Perkins S. 2019. Albedo is a simple concept that plays complicated roles in climate and astronomy. Proceedings of the National Academy of Sciences. 116(51):25369–25371.

Pillet L, Fontaine D, Pawlowski J. 2012. Intra-genomic ribosomal RNA polymorphism and morphological variation in *Elphidium macellum* suggests inter-specific hybridization in foraminifera. PLoS One. 7(2):e32373.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432.

Rogers AD. 2007. Evolution and biodiversity of Antarctic organisms: a molecular perspective. Philos Trans R Soc Lond B Biol Sci. 362(1488):2191–2214.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19(2):301–302.

Selbmann, De Hoog, Mazzaglia. 2005. Fungi at the edge of life: cryptoendolithic black fungi from Antarctic desert. Stud Mycol.

Selbmann L, Grube M, Onofri S, Isola D, Zucconi L. 2013. Antarctic epilithic lichens as niches for black meristematic fungi. Biology. 2(2):784–797.

Selbmann L, de Hoog GS, Zucconi L, Isola D, Ruisi S, van den Ende AHGG, Ruibal C, De Leo F, Urzì C, Onofri S. 2008. Drought meets acid: three new genera in a dothidealean clade of extremotolerant fungi. Stud Mycol. 61:1–20.

Selbmann L, Zucconi L, Isola D, Onofri S. 2015. Rock black fungi: excellence in the extremes, from the Antarctic to space. Curr Genet. 61(3):335–345.

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, et al. 2015. Polyploidy can drive rapid adaptation in yeast. Nature. 519(7543):349–352.

Sinha S, Flibotte S, Neira M, Formby S, Plemenitaš A, Cimerman NG, Lenassi M, Gostinčar C, Stajich JE, Nislow C. 2017. Insight into the recent genome duplication of the halophilic yeast *Hortaea werneckii:* combining an improved genome with gene expression and chromatin structure. G3: Genes, Genomes, Genetics. 7(7):2015–2022.

Smith MM. 1987. Molecular evolution of the *Saccharomyces cerevisiae* histone gene loci. J Mol Evol. 24(3):252–259.

Staben C, Yanofsky C. 1990. *Neurospora crassa* a mating-type region. Proc Natl Acad Sci U S A. 87(13):4917–4921.

Sterflinger K, Lopandic K, Blasi B, Poynter C, de Hoog S, Tafer H. 2015. Draft Genome of *Cladophialophora immunda*, a Black Yeast and Efficient Degrader of Polyaromatic Hydrocarbons. Genome Announc. 3(1). doi:10.1128/genomeA.01283-14. http://dx.doi.org/10.1128/genomeA.01283-14.

Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. Bioinformatics. 34(4):550–557.

Teixeira MM, Moreno LF, Stielow BJ, Muszewska A, Hainaut M, Gonzaga L, Abouelleil A, Patané JSL, Priest M, Souza R, et al. 2017. Exploring the genomic diversity of black yeasts and relatives (Chaetothyriales, Ascomycota). Stud Mycol. 86:1–28.

Tesei D, Marzban G, Zakharova K, Isola D, Selbmann L, Sterflinger K. 2012. Alteration of protein patterns in black rock inhabiting fungi as a response to different temperatures. Fungal Biol. 116(8):932–940.

Todd RT, Braverman AL, Selmecki A. 2018. Flow Cytometry Analysis of Fungal Ploidy. Curr Protoc Microbiol. 50(1):e58.

de la Torre JR, Goebel BM, Friedmann EI, Pace NR. 2003. Microbial diversity of cryptoendolithic communities from the McMurdo Dry Valleys, Antarctica. Appl Environ Microbiol. 69(7):3858–3867.

Turgeon BG, Yoder OC. 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal Genet Biol. 31(1):1–5.

Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, et al. 2020. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol Biol Evol. 37(2):599–603.

Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018. nQuire: a statistical framework for ploidy estimation using next generation sequencing. BMC Bioinformatics. 19(1):122.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD. 2017. Which MAT gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. Fungal Biol Rev. 31(4):199–211.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 387(6634):708–713.

Xiang Q, Glass NL. 2004. The control of mating type heterokaryon incompatibility by vib-1, a locus involved in het-c heterokaryon incompatibility in *Neurospora crassa*. Fungal Genet Biol. 41(12):1063–1076.

Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, Zhao W, Shen M, Zhang H, Li C, et al. 2012. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. PLoS Genet. 8(8):e1002869.

Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, et al. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. Elife. 2:e00731.

Zalar P, Zupančič J, Gostinčar C, Zajc J, de Hoog GS, De Leo F, Azua-Bustos A, Gunde-Cimerman N. 2019. The extremely halotolerant black yeast *Hortaea werneckii* - a model for intraspecific hybridization in clonal fungi. IMA Fungus. 10:10.

Zörgö E, Chwialkowska K, Gjuvsland AB, Garré E, Sunnerhagen P, Liti G, Blomberg A, Omholt SW, Warringer J. 2013. Ancient Evolutionary Trade-Offs between Yeast Ploidy States. PLoS Genetics. 9(3):e1003388. doi:10.1371/journal.pgen.1003388.

In host evolution of *Exophiala dermatitidis* in cystic fibrosis lung micro-environment

Abstract

Individuals with cystic fibrosis (CF) are susceptible to chronic lung infections that lead to inflammation and irreversible lung damage. While most respiratory infections that occur in CF are caused by bacteria, some are dominated by fungi such as the slow-growing black yeast *Exophiala dermatitidis*. Here, we analyze isolates of *E. dermatitidis* cultured from two samples, collected from a single subject two years apart. One isolate genome was sequenced using long-read Nanopore technology as an in-population reference to use in comparative single nucleotide polymorphism (SNP) and insertion-deletion (INDEL) variant analyses of twenty-three isolates. We then used population genomics and phylo-genomics to compare the isolates to each other as well as the type strain *E. dermatitidis NIH/UT8656.* Within the CF lung population, three *E. dermatitidis* clades were detected, each with varying mutation rates. Overall, the isolates were highly similar suggesting that they were recently diverged. All isolates were MAT 1-1, which was consistent with their high relatedness and the absence of evidence for mating or recombination between isolates. Phylogenetic analysis grouped sets of isolates into clades that contained isolates from both early and late time points indicating there are multiple persistent lineages. Functional assessment of variants unique to each clade identified alleles in genes that encode transporters, cytochrome P450 oxidoreductases, iron acquisition and DNA repair processes. Consistent with the genomic heterogeneity, isolates showed some stable

phenotype heterogeneity in melanin production, subtle differences in antifungal minimum inhibitory concentrations, growth on different substrates. The persistent population heterogeneity identified in lung-derived isolates is an important factor to consider in the study of chronic fungal infections, and the analysis of changes in fungal pathogens over time may provide important insights into the physiology of black yeasts and other slow-growing fungi *in vivo*.

Introduction

Cystic fibrosis (CF) is an autosomal recessive disorder caused by mutations in the cystic fibrosis transmembrane regulator (CFTR) gene that impair the balance of salts and water across epithelia. In the lungs, these ion transport defects cause viscous mucus which contributes to respiratory infections that cause most of the morbidity and mortality in CF populations (Riordan *et al.* 1989; Davis 2006; Ferec and Cutting 2012). Microbial colonization of mucosal plugs results in recurring infection and inflammation that cause irreversible lung damage and declining lung function (Turcios 2020). Bacteria, particularly *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Burkholderia cepacia,* and *Stenotrophomonas maltophilia* are pathogens that frequently dominate CF respiratory infections (Burns *et al.* 1998; Mariani-Kurkdjian and Bingen 2003; Horré *et al.* 2004; Steinkamp *et al.* 2005; Tunney *et al.* 2008; Pihet *et al.* 2009; Zhao *et al.* 2012) and are sometimes isolated with various species of fungi. Clinically significant fungi in CF lung infections include *Exophiala dermatitidis*, *Scedosporium apiospermum*, *Aspergillus fumigatus*, *Candida albicans,* and *Clavispora (Candida) lusitaniae* (Kusenbach *et al.* 1992;

Cimon *et al.* 2000; Defontaine *et al.* 2002; Horré *et al.* 2004, 2009; Parize *et al.* 2014; Chen *et al.* 2018; Demers *et al.* 2018; Jong *et al.* 2020). The consequences of fungal infection on CF outcomes are not well understood but are influenced by the genotypes of both the host and microbes (Burns *et al.* 1998; Nagano *et al.* 2007; Pihet *et al.* 2009; Packeu *et al.* 2012).

Exophiala dermatitidis, previously named *Hormiscium dermatitidis* and *Wangiella dermatitidis*, are taxonomically classified in Phylum Ascomycota, Order Chaetothyriales, Family Herpotrichiellaceae. To date, about 40 species in the *Exophiala* genus have been identified, seventeen of which are known to cause disease in mammals. Among these, *E. dermatitidis* is the most clinically prevalent with reported mortality rates of 25–80% in systemic and invasive cases, even though fatal systemic cases are relatively rare (Kirchhoff *et al.* 2019). Clinical presentations of this fungus include phaeohyphomycosis, keratitis, chromoblastomycosis, and even several neural diseases and meningitis (Revankar *et al.* 2002; Matos *et al.* 2002; Uijthof *et al.* 2009; Revankar and Sutton 2010; Seyedmousavi *et al.* 2014; Song *et al.* 2017; Kirchhoff *et al.* 2019; Lavrin *et al.* 2020). The first instance of *E. dermatitidis* to be isolated from a sputum culture procured from a cystic fibrosis patient was in 1990 (Haase *et al.* 1990; Kusenbach *et al.* 1992). Many studies have since isolated *E. dermatitidis* from CF sputum cultures (Rath *et al.* 1997; Diemert *et al.* 2001; Horré *et al.* 2004; Griffard *et al.* 2010; Packeu *et al.* 2012), some of which have led to the development of later state mycosis disease (Sudfeld *et al.* 2010; Kondori *et al.* 2011; Song *et al.* 2017; Grenouillet *et al.* 2018), treatment (Packeu *et al.* 2012; Mukai *et al.* 2014) or even death (Klasinc *et al.* 2019).

Exophiala species are black yeasts that are classified by three defining features. They produce melanin through the 1-8 dihydroxynapthalene (DHN) biosynthesis pathway, exhibit morphological plasticity or meristematic growth (yeast cells, hyphae, or even pseudohyphae), and have membrane-associated carotenoids and intracellular mycosporine-like amino acids (Hoog *et al.* 2000; Nosanchuk and Casadevall 2003, 2006; Saunte *et al.* 2012; Smith and Casadevall 2019). All of these properties likely contribute to the extreme resistance to environmental stresses including desiccation, UV or solar exposure, and ionizing radiation. These resistance traits may also contribute to success in growth in mammalian hosts and their ability to cause disease in susceptible hosts. The *E. dermatitidis* strain NIH/UT8656 genome was sequenced and assembled into 11 complete and contiguous chromosomes (Robertson *et al.* 2012; Chen *et al.* 2014; Schultzhaus *et al.* 2020; Malo *et al.* 2021) which has enabled comparative genomics and identification of the genes which may underlie its resilience (Robertson *et al.* 2012; Schultzhaus *et al.* 2020; Malo *et al.* 2021) and success in human host colonization (Kondori *et al.* 2011; Kirchhoff *et al.* 2019).

Reservoirs of *E. dermatitidis* have long been associated with hot and humid tropical origins (Sudhadham *et al.* 2008). *E. dermatitidis* are isolated from many man-made substrates found in humidifiers (Nishimura and Miyaji 1982), saunas (Matos *et al.* 2002), and dishwashers (Zupančič *et al.* 2016; Babič *et al.* 2018). Babic et al. (Babič *et al.* 2018) concluded that *Exophiala* tends to be found in locations with oligotrophic conditions or where rubber seals and humidity act as an enrichment or trapping mechanism which supports *Exophiala* persistence. *E. dermatitidis* and related species are also found in broad

environmental niches including wasp nests (Conti-Díaz *et al.* 1977), healthy bats (Reiss and Mok 1979), lesions of toads (Frank *et al.* 1970), and rotten wood (Dixon *et al.* 1980) which suggests human patients acquire infections from environmental exposure (Sudhadham *et al.* 2011).

In light of several studies that chronic CF-related fungal infections diversify over time (Kim et al. 2015; Demers et al. 2018; Ross et al. 2021; Jones et al. 2021), here we report both phenotypic and genomic diversity among *E. dermatitidis* isolates from a single individual. Population genomic analyses identified multiple lineages that persisted over two years. This data set allows us to test the hypothesis that, as seen in previous studies, the CF lung environment supports stably diverged populations of a clonally derived yeast.

Methods

Sputum-derived isolate cultures

Frozen sputum was obtained from a specimen bank in which samples were obtained in accordance with protocols approved by the Dartmouth-Hitchcock Institutional Review Board. Aliquots of sputum were plated onto Sabouraud Dextrose Agar (SAB) medium as described previously (Grahl *et al.* 2018). Individual isolates were obtained and banked; isolates are listed in **Table 1.**

Table 1. Collection and MIC values for CF patient-derived *E. dermatitidis* isolates. Information on the 23 CF isolates cultured from one patient's sputum across three years. The table summarizes phylogenetic clade designation, itraconazole MIC, date of collection, and classification as Early or Late.

Strain ID	Clade Node	MIC ng/ml	Date Isolated	Early or Late
Ex2	I	500	October 2, 2014	Early
Ex1	I	250	October 2, 2014	Early
Ex12	I	250	August 4, 2016	Late
Ex14	I	62.5	August 4, 2016	Late
Ex6	II	250	October 2, 2014	Early
Ex10	II	62.5	October 2, 2014	Early
Ex22	II	62.5	August 17, 2016	Late
Ex17	II	250	August 4, 2016	Late
Ex16	II	500	August 4, 2016	Late
Ex7	II	500	October 2, 2014	Early
Ex19	II	250	August 4, 2016	Late
Ex23	II	250	August 17, 2016	Late
Ex3	II	125	October 2, 2014	Early
Ex8	II	125	October 2, 2014	Early
Ex4/DCF04	II	250	October 2, 2014	Early
Ex15	III	250	August 4, 2016	Late

Ex18	III	62.5	August 4, 2016	Late
Ex20	III	125	August 17, 2016	Late
Ex21	III	125	August 17, 2016	Late
Ex5	III	250	October 2, 2014	Early
Ex9	III	500	October 2, 2014	Early
Ex13	III	250	August 4, 2016	Late
Ex11	III	250	October 2, 2014	Early

DNA extraction and sequencing

Exophiala isolates were grown in Yeast Peptone Dextrose media (YPD) for approximately 24 hours in 5 ml roller-drum cultures at 37℃. Cells were spun down for 5 minutes at 5000 RCF and washed thrice with deionized water. Genomic DNA was extracted from cell pellets using the MasterPure yeast DNA purification kit (Epicentre). Melanin was removed from genomic DNA using the *OneStep*™ PCR Inhibitor Removal kit (Zymo Research). Genomic DNA was measured by Nanodrop and diluted to ~20ng/µl. DNA extractions were sent to Novogene, (Novogene Corporation Inc., Cambridge, United Kingdom) for 2x150bp sequencing on an Illumina NovoSeq 6000. DNA from isolate DCF04 was also extracted and sequenced on Oxford Nanopore (ONT) platform with library preparation and sequencing following the manufacturer's directions (Oxford Nanopore, Oxford United Kingdom). Flow cell versions FAK67997 and FAK73296 were used along with base-calling using guppy (*v. 3.4.4+a296cb*) (Wick *et al.* 2019).

Genome assembly and annotation

Genome assemblies were constructed for the twenty-three *E. dermatitidis* isolates from Illumina sequencing. One isolate, Ex4, was also re-sequenced using Oxford Nanopore technology. All genomes were *de novo* assembled with AAFTF pipeline (*v.0.2.3*) (Palmer and Stajich 2022) which performs read QC and filtering with BBTools bbduk (*v.38.86)* (Bushnell 2014) followed by SPAdes (*v.3.15.2*) (Bankevich *et al.* 2012) assembly using default parameters, followed by screening to remove short contigs < 200 bp and contamination using NCBI's VecScreen. The BUSCO ascomycota_odb9 database (Manni *et al.* 2021) was used to determine how complete the assembly was for all 23 isolates of *E. dermatitidis.* A hybrid assembly of isolate DCF04/Ex4 was generated using MaSUrCa (*v.3.3.4*) (Zimin *et al.* 2013) as the assembler using both Nanopore and Illumina sequencing reads. General default parameters were used except: CA_PARAMETERS=cgwErrorRate=0.15, NUM_THREADS=16, and JF_SIZE=200000000. The updated genome was then scaffolded to strain NIH/UT8656 accession GCF_000230625 using Ragtag (*v.1.0.2*) (Alonge *et al.* 2019) which uses minimap2 (*v. 2.17-r941)* (Li 2018) to further link scaffolds based on shared co-linearity of these isolates' genomes.

We predicted genes in this near-complete genome assembly with Funannotate (*v1.8.1*) (Palmer and Stajich 2020). A masked genome was created by first generating a library of sequence repeats with the RepeatModeler pipeline (Smit and Hubley 2008). These species-specific predicted repeats were combined with fungal repeats in the RepBase (Bao *et al.* 2015) to identify and mask repetitive regions in the genome assembly with RepeatMasker

(*v.4-1-1*) (SMIT A. F. A 2004). To predict genes, *ab initio* gene predictors SNAP (*v.2013_11_29*) (Korf 2004) and AUGUSTUS (*v.3.3.3*) (Stanke *et al.* 2006) were trained using the Funannotate 'train' command based on the full-length transcripts constructed by Genome-Guided run of Trinity (*v.2.11.0*) (Grabherr *et al.* 2011) using RNA-Seq from published E. *dermatitis* SRA accession SRS282040. The assembled transcripts were aligned with PASA (*v.2.4.1*) (Haas *et al.* 2008) to produce full-length spliced alignments and predicted open reading frames for training the *ab initio* predictors and as informant data for gene predictions. Additional gene models were predicted by GeneMark.HMM-ES (*v.4.62_lic*) (Brůna *et al.* 2020), and GlimmerHMM (*v.3.0.4*) (Majoros *et al.* 2004) utilize a self-training procedure to optimize *ab initio* predictions. Additional exon evidence to provide hints to gene predictors was generated by DIAMOND BLASTX alignment of SwissprotDB proteins and polished by Exonerate (*v.2.4.0*) (Slater and Birney 2005). Finally, EvidenceModeler (*v.1.1.1*) (Haas *et al.* 2008) generated consensus gene models in Funannotate that were constructed using default evidence weights. Non-protein-coding tRNA genes were predicted by tRNAscan-SE (*v.2.0.9*) (Lowe and Chan 2016).

The annotated genome was processed with antiSMASH (*v.5.1.1*) (Blin *et al.* 2021) to predict secondary metabolite biosynthesis gene clusters. These annotations were also incorporated into the functional annotation by Funannotate. Putative protein functions were assigned to genes based on sequence similarity to InterProScan5 (*v.5.51-85.0*) (Jones *et al.* 2014), Pfam (*v.35.0*) (Finn *et al.* 2014), Eggnog (*v.2.1.6-d35afda*) (Huerta-Cepas *et al.* 2019), dbCAN2 (*v.9.0*) (Zhang *et al.* 2018) and MEROPS (*v.12.0*) (Rawlings *et al.* 2018) databases relying on NCBI BLAST (*v.2.9.0+)* (Sofi *et al.* 2022) and HMMer (*v.3.3.2*)

(Potter *et al.* 2018). Gene Ontology terms were assigned to protein products based on the inferred homology based on these sequence similarity analyses. The final annotation produced by Funannotate was deposited in NCBI as a genome assembly with gene model annotation (JAJGCF000000000).

Copy number variation (CNV) was examined by plotting 10kbp window-based read coverage of the short-read alignments of each isolate. The depth of coverage was calculated using Mosdepth (Pedersen and Quinlan 2018) and visualized with R using the ggplot2 package (Wickham 2016).

The Mating Type (MAT) locus was identified by searching for homologous MAT genes (HMPREF08862) and (HMPREF05727) in this study's 23 *E dermatitidis* isolate genomes with cblaster (Gilchrist *et al.* 2021). A homothallic black yeast, *Capronia coronata* CBS 617.96 (AMWN00000000.1) (Teixeira *et al.* 2017), which has both MAT genes, was also incorporated into the analyses and visualization. The identified homologous regions were examined for their conserved synteny of the MAT locus using clinker (Gilchrist and Chooi 2021) and a custom Biopython script (Cock *et al.* 2009) (Kurbessoian 2022) to extract the annotated region of the genome which contained the locus.

Identification of telomeric repeat sequences was performed using FindTelomeres.py script (https://github.com/JanaSperschneider/FindTelomeres). Briefly, this searches for chromosomal assembly with a regular expression pattern for telomeric sequences at the 5' and 3' end of each scaffold. Telomere repeat sequences were also predicted using A

Telomere Identification toolkit (tidk) (*v.0.1.5*) "explore" option (https://github.com/tolkit/telomeric-identifier).

Identification of sequence variation

Sequence variation among isolates was assessed using the best practices of the Genome Analysis ToolKit GATK (*v. 4.0.4.0*) (McKenna *et al.* 2010; Franke and Crowgey 2020) to identify SNPs and Insertion/Deletions (INDEL). Illumina paired-end reads were aligned to isolate DCF04 assembly with BWA (*v.0.7.17*) (Li and Durbin 2010) and processed with Samtools (*v.1.8*) (Li *et al.* 2009) and Picard Toolkit (Institute; Toolkit) AddOrReplaceReadGroups and MarkDuplicates (*v.2.18.3*). The alignments were further improved by realigning reads near inferred INDELs using GATK tools RealignerTargetCreator, IndelRealigner, and PrintReads. Genotypes were inferred with the GATK Haplotype and GenotypeGVCF methods to produce a single VCF file of the identified variants. Low-quality SNPs were further filtered using GATK VariantFiltration and finally, SelectVariants was used with the parameters: mapping quality (score < 40), quality by depth (<2 reads), Strand Odds Ratio (SQR > 4.0), Fisher Strand Bias (> 200), and Read Position Rank Sum Test (< -20) to retain only high-quality polymorphisms. Finally, an additional stringent series of three filtering steps implemented in bcftools (*v. 1.12*)(Li et al. 2019) was used on the VCF file to remove calls that were below the 1000 quality score threshold, where any individual isolate had a "no call", and where the standard deviation in read depth (DP) was above or below a standard deviation value of 1 for an individual SNP. SnpEff (*v.4.3r*) (Cingolani *et al.* 2012) was used to score the impact of the identified variants using the Funannotate annotated DCF04 genome GFF3 file.

Variant calling was performed on two sets of individuals, one limited to the twenty-three CF patient population isolates with reads aligned to the DCF04 isolate and one using the *E. dermatitidis* NIH/UT8656 type strain. Pairwise isolate comparisons of SNP and INDEL were counted to generate isolate correlation heatmaps for both variant types using UPGMA clustering. A custom script make_diagonal.sh uses plink (*v.2.00a3LM_AVX2*) (Chang *et al.* 2015) to count all pairwise differences between individuals in the VCF files stratified by SNPs or INDELs. A custom Perl script transformed pairwise counts into a matrix of isolated differences observed for both SNP and INDEL variants. Counts were summarized as heatmaps with an R script. To summarize the matrix plots, a distance plot using regression statistics was applied to both SNP counts and INDEL counts. A regression plot and statistics for the slope of the progression line, Pearson's R, R-squared, and the *p*-value were computed with an R script. All scripts developed for this manuscript are available at the GitHub (Kurbessoian 2022) project linked in this paper.

A calculation of the population mutation rate was performed on each isolate based on the number of SNPs shared among a pair of isolates. The formula to calculate the mutation rate per year for each isolate is as follows: (SNP Pairwise Value) / (Adjusted Genome Length) / Pair / Year. The time between isolated collections was 22 months. The value used for the "Adjusted Genome Length" was collected from running the assembler AAFTF pipeline (*v.0.2.3*) (Palmer and Stajich 2022). A one-way ANOVA was run on the grouped calculated mutation rates for each isolate to determine significance.

Genome comparison with dot-plot was constructed with D-GENIES (Cabanettes and Klopp 2018) using minimap2 and default parameters through the website for the tool. Longer isoform proteins were extracted for each strain genome annotation in order to call a more accurate gene count. Using Orthofinder (Emms and Kelly 2019) and DIAMOND ultra-sensitive parameters (Buchfink *et al.* 2015), assessment of overlap in the predicted protein-coding gene sequences from the genomes of *E. dermatitidis DCF04* and *E. dermatitidis NIH/UT8656* protein genomes was generated.

Phylogenetics relationships of the isolates

SNPs from polymorphic sites were extracted from the VCF files as multi-fasta files using BCFTools (Li *et al.* 2019) and a custom script make_strain_tree.sh. A Maximum Likelihood phylogenetic tree was constructed from the multi-fasta file using IQTree (*v. 2.0.4*) (Minh *et al.* 2020) and the model parameters [-m GTR+ASC]. The chosen nucleotide substation model was GTR+F+ASC selected based on Bayesian information criteria (BIC). Statistical support for the tree nodes was evaluated from 1000 bootstrap replicates using UFBoot ultra-fast bootstrapping approximation (Hoang *et al.* 2018). The tree was visualized using iTOL (Letunic and Bork 2016).

Phenotype assays

Exophiala isolates were streaked from -80˚C onto yeast extract peptone dextrose (YPD) plates (2% glucose, 2% yeast extract, 1% peptone) and allowed to grow for 48 hours at 37℃. Overnight cultures were started from YPD patches inoculated into 5 ml of liquid YPD and grown for approximately 24 hours in 5 ml rolling barrel cultures at 30℃. For

minimum inhibatory concentration (MIC) assays, cultures were spun down for 5 minutes at 5000 RCF and washed thrice in deionized water. Cells were counted on a hemocytometer and added to a final concentration of 1000 colony-forming units per well in a 96-well flat-bottom dish, then grown at 37˚C for 72 hours before measuring the final MIC.

Results

A molecular and culture-based analysis of a series of sputum samples identified an individual with CF with a chronic lung infection caused by *E. dermatitidis* (Grahl *et al.* 2018). *E. dermatitidis* isolates were recovered from banked sputum samples, collected two years apart. *Staphylococcus aureus* and *Candida albicans* were also identified in clinical cultures from the patient in the intervening years between the two time points (**Figure 1**). The subject's antimicrobial use history included Aztreonam, Azithromycin, Tobramycin, Ciprofloxacin, and Doxycycline, and the patient's lung function, measured by percent predicted forced expiratory volume (%FEV1), ranged between 80 and 49% during this time period. While *E. dermatitidis* was not detected in the first clinical microbiological analysis, perhaps due to its extremely slow growth out of clinical samples (Grahl *et al.* 2018) or suppression by bacteria, it was detected in the second clinical analysis. Twenty-three isolates (eleven from the early time point and twelve from the late time point) were selected for further population genomic study.

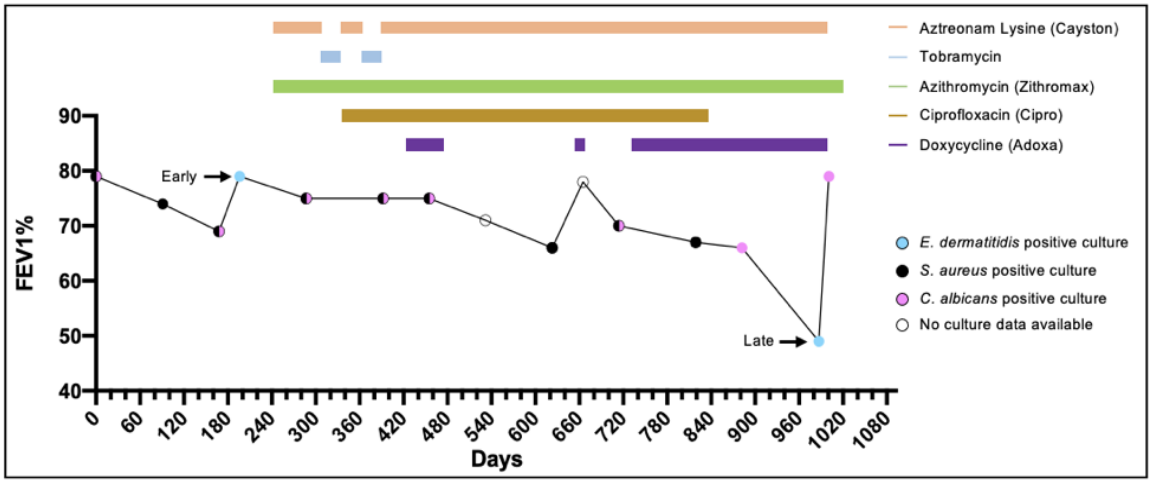


Figure 1. FEV1 pulmonary function data. FEV1 pulmonary function data were tested over the course of three years, and sputum cultures were acquired at each indicated time point and assessed for the presence of pathogens other than the mixed bacteria that compose normal upper respiratory flora. Colored bars indicate the duration of treatment for each listed antimicrobial given during infection.

Sequencing and assembly of E. dermatitidis isolates

To gain information on the genetic variation and potential population structure for the recovered *E. dermatitidis* isolates, we sequenced and assembled the genomes of the twenty-three isolates (**Table 3**). The depth of coverage ranged from 11-47x coverage across all 23 Illumina sequenced samples. The BUSCO (Benchmarking Universal Single-Copy Orthologs) is another tool used to determine genome assembly completeness; within our dataset, it ranged from 99-99.3 % complete, 98.3-99.3 single copies present, 0-1 ranged in duplicates, with a range of 0-0.3 in missing BUSCOs. Contig counts ranged from 44-1422, while the average genome assembly size is 26,706,323 Mbp, L50 about 1,078,277, and N50 of 9.

To further examine the fine-scale variation within the CF isolates we sought to generate a within-population high-quality reference genome. The isolate DCF04 was sequenced using Oxford Nanopore (ONT) long reads to a genome depth of coverage at 11x and a hybrid genome assembly constructed from both ONT and Illumina reads (**Table 1; Supplemental Table 2).** The 26.6 Mb assembly contained only 3.19% of identified repetitive elements. The candidate telomeric repeat units "TTTAGGG/CCCTAA" were identified as repeat arrays at both ends of five scaffolds, but also found as single pairs in the remaining 4 scaffolds, as would be expected for 9 complete chromosomes (**Table 4**). Sixty-four tRNA models were predicted from the genome. While the total number of genes predicted was 10030, 9599 of which are protein-coding. 15 secondary metabolite clusters, 37 biosynthetic enzymes, and 49 small COGs were predicted with antiSMASH.

Table 2. Genome assembly summary statistics for reference isolate *E. dermatitidis* DCF04. The table indicates assembly summary statistics of the hybrid assembly using Illumina and Nanopore sequencing of isolate DCF04. Summary statistics of contig and scaffold lengths are presented with genome completeness data calculated with BUSCO using the ascomycota_odb9 database.

Genome Assembly Statistics	***E. dermatitidis* DCF04**
Scaffold Count	44
Total length	26,633,774 bp
Minimum length	501
Maximum length	4,314,646
Mean contig length	59,1861.64
Scaffold L50	4
Scaffold N50	3,672,342
Scaffold L90	7
Scaffold N90	2,864,055
Contig Count	165
Contig L50	21
Contig N50	451,966
BUSCO Complete %	99%
BUSCO Duplicate %	0%
BUSCO Single %	99%
GC%	51.40

Comparing genome assembly and annotation of DCF04 to NIH/UT8656

Our DCF04 isolate is genetically close to the public strain *Exophiala dermatitidis NIH/UT8656* sequenced with Sanger sequencing technology (BioProject: PRJNA225511, Assembly: GCF_000230625.1). To assess the differences between the two genome assemblies we compiled summary statistics (**Table 2** and **Table 3**) and interrogated the predicted gene content of both genomes. The DCF04 assembly had 165 contigs linked into 44 scaffolds, while NIH/UT8656 comprised 238 contigs linked into 10 scaffolds. Note that DCF04 scaffolds were derived by a comparative assembly against the NIH/UT8656 assembly to achieve the best assembly after checking for rearrangements. The total length of the genome assembly is nearly the same for both DCF04 at 26.6Mb and NIH/UT8656 at 26.4Mb. The summary statistics for scaffold L50 and N50 are also nearly identical for DCF04 at 4 and 3.7Mb and NIH/UT8656 were 4 and 3.6Mb. A dot-plot comparing the two genome assemblies revealed minimal rearrangements or discontinuity suggesting high similarity of the two isolated genomes (**Figure 2**).

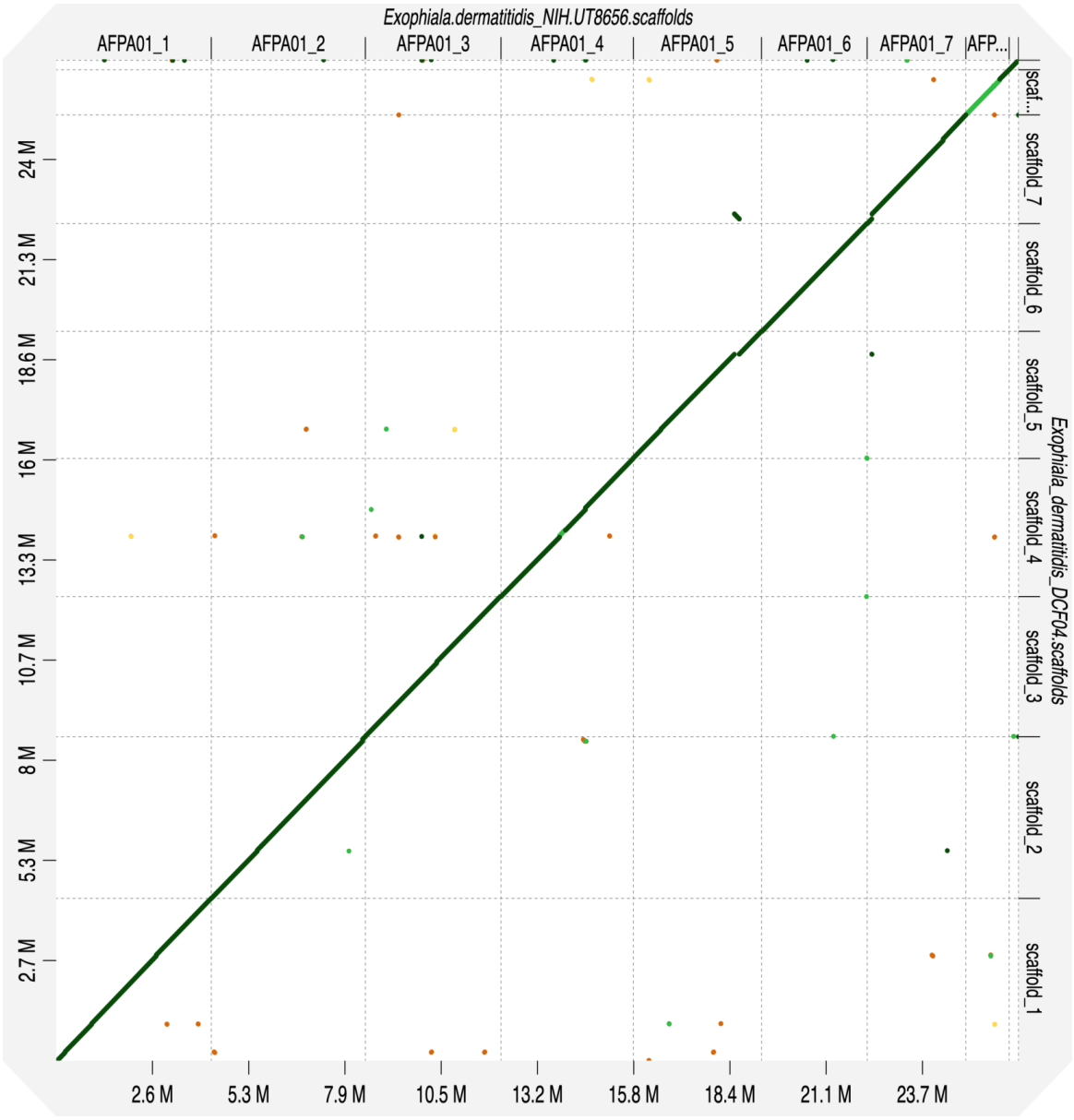


Figure 2. Genome dot plot of *E. dermatitidis* DCF04 and *E. dermatitidis* NIH/UT8656. D-GENIES web application was used to generate a dot plot representation to compare genome assembly content and test for structural changes or rearrangements. Both iterations of the plot indicate a majority of the two genomes are nearly identical to each other.

Table 3. Genome assembly statistics for 23 CF isolates. Assembly statistics for scaffolded assemblies of isolates and NIH/UT8656 previously published genome. The table indicates assembly summary statistics of the assembly using Illumina sequencing of all 23 CF isolates. Summary statistics of scaffold lengths are presented with genome completeness data calculated with BUSCO using the ascomycota_odb9 database.

Strain	# Reads	# Bases (Mb)	Depth of Coverage	Contig Count	Assembly Size (bp)	N50	L50	BUSCO Complete	Acc #
Ex1	6138963	1841.69	61.39	104	26,631,992	951,691	11	99.0	JAJGCI000000000
Ex2	6517791	1955.34	65.18	90	26,627,755	951,565	9	99.0	JAJGCH000000000
Ex3	5804279	1741.28	58.04	87	26,628,775	1,002,554	9	99.0	JAJGCG000000000
Ex4	5875173	1762.55	58.75	175	26,652,590	966,306	10	99.0	JAJGCF000000000
Ex5	7066295	2119.89	70.66	102	26,631,945	951,690	10	99.0	JAJGCE000000000
Ex6	5423418	1627.03	54.23	411	26,713,974	1,053,417	8	99.3	JAJGCD000000000
Ex7	5628088	1688.43	56.28	197	27,280,555	951,673	9	99.3	JAJGCC000000000
Ex8	6026426	1807.93	60.26	171	26,647,902	1,000,046	8	99.0	JAJGCB000000000
Ex9	6534312	1960.29	65.34	93	26,632,463	1,108,162	8	99.3	JAJGCA000000000
Ex10	5717134	1715.14	57.17	103	26,621,718	885,042	9	99.0	JAJGBZ000000000
Ex11	5842789	1752.84	58.43	160	26,649,686	1,106,267	8	99.0	JAJGBY000000000
Ex12	6952094	2085.63	69.52	58	26,624,850	1,256,779	8	99.0	JAJJCG000000000
Ex13	4814549	1444.36	48.15	503	26,784,658	946,443	9	99.0	JAJGBX000000000
Ex14	6120366	1836.11	61.20	136	26,657,265	965,535	10	99.0	JAJGBW000000000

Ex15	5670365	1701.11	56.70	431	26,771,446	821,229	10	99.0	JAJJCF000000000
Ex16	5359740	1607.92	53.60	502	26,733,512	976,501	8	99.0	JAJJCE000000000
Ex17	5216653	1565.00	52.17	229	26,658,537	965,533	9	99.0	JAJJCD000000000
Ex18	5358369	1607.51	53.58	601	26,787,297	569,204	17	99.0	JAJJCC000000000
Ex19	5742268	1722.68	57.42	275	26,670,959	951,665	9	99.0	JAJJCB000000000
Ex20	5955090	1786.53	59.55	1422	26,980,824	853,597	9	99.0	JAJJCA000000000
Ex21	6395829	1918.75	63.96	61	26,621,742	999,909	10	99.0	JAJGBV000000000
Ex22	5452693	1635.81	54.53	212	26,657,178	951,688	10	99.3	JAJJBZ000000000
Ex23	5445813	1633.74	54.46	151	26,650,368	1,019,798	9	99.0	JAJGBU000000000
Ex4/DCF04	5991351	307.88	11.56	44	26,633,774	3,672,342	4	99.0	JAJGCF000000000
NIH/8656				10	26,376,767	3,623,950	4	98.6	AFPA00000000.1

Table 4. Telomere Recovery for DCF04. Table depicting telomere recovery results for DCF04 CF *E. dermatitidis*. Candidate telomeric repeat units "TTTAGGG/CCCTAA" were identified as repeat arrays at both ends of five scaffolds, but also found as single pairs in the remaining 4 scaffolds, as would be expected for 9 complete chromosomes.

find teloemeres.py script		Telomeres found: 14 (6 forward, 7 reverse)						
scaffold_1	Forward (start)	CCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCC	**tidk explore**					
scaffold_1	Reverse (end)	TTAGGGTTTAGGGTTTAGGGTTAGGGTTTAGGGTTTAGGGTTTAGGGTT	**id**	**Start position**	**End position**	**Repeat number**	**Repeat sequence**	**Sequence length**
scaffold_2	Forward (start)	CCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACC	scaffold_3	57044	57444	6	CCCC	4
scaffold_2	Reverse (end)	GGTTTAGGGTTTAGGGTTTAGGGTTTAGGGGTTTAGGGTTTAGGGTTTAGG	scaffold_8	1092092	1092212	6	TTGA	4
scaffold_3	Reverse (end)	GTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGG	scaffold_1	7	63	8	CCTA AAC	7
scaffold_4	Forward (start)	AAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTA	scaffold_2	7	49	6	CCTA AAC	7
scaffold_4	Reverse (end)	GTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGG	scaffold_4	7	63	8	AAA CCCT	7
scaffold_5	Forward (start)	CCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACCCTAAACC	scaffold_5	7	49	6	CCT AAAC	7
scaffold_5	Reverse (end)	TAGGGTTTAGGGTTTAGGGTTTAGGGGTTTAGGGTTTAGGGTTTAGGGTTT	scaffold_6	7	63	8	CCT AAAC	7

scaffold_6	Forward (start)	CCTAAACCCTAAACCCTAAACCCTA AACCCTAAACCCTAAACCCTAAACC	scaffold_9	7	63	8	AAC CCTA	7
scaffold_6	Reverse (end)	AGGGTTTAGGGTTTAGGGTTTAGG GTTTAGGGTTTAGGGTTTAGGGTTTA	scaffold_2	21480	21640	12	TCCT GTCC	8
scaffold_8	Reverse (end)	TTAGGGTTTAGGGTTTAGGGTTTAG GGTTTAGGGTTTAGGGTTTAGGGTT	scaffold_4	42	220136	7	GTA GGCG TGC AGTG	14
scaffold_9	Forward (start)	AACCCTAAACCCTAAACCCTAAACC CTAAACCCTAAACCCTAAACCCTAA						
scaffold_1	Forward (start)	CCTAAACCCTAAACCCTAAACCCTA AACCCTAAACCCTAAACCCTAAACC						

Table 5. OrthoFinder summary comparing DCF04 and NIH/UT8656. Comparison of the shared and unique orthogroups found in the annotated proteomes of *E. dermatitidis* strains NIH/UT8656 and DCF04. A majority of orthogroups (8,256; 99%) had members from both strains, of these 15 were single-copy orthogroups containing a single protein-coding gene from each strain. There were 10 orthogroups unique to DCF04 encompassing 34 protein-coding genes and 5 orthogroups unique to NIH/UT8656 made up of 24 protein-coding genes. DCF04 contained 705 unassigned genes, while NIH/UT8656 had 475.

OrthoFinder Results	***E. dermatitidis* DCF04**	***E. dermatitidis* NIH/UT8656**
Number of genes	9535	9285
Number of genes in orthogroups	8830	8810
Number of unassigned genes (species-specific)	705	475
Number of genes in species-specific orthogroups	34	24
Total number of genes specific to each species	739	499

Summary Statistics	
Number of genes	18,820
Number of genes in orthogroups	17,640
Number of unassigned genes	1,180
Number of orthogroups	8,271
Number of orthogroups with both taxa present	8,256
Number of single-copy orthogroups	7,818

Lineage-specific paralogs

Orthogroup	**DCF04**	**NIH/UT8656**	**Total**	**Function**
OG0000035	0	8	8	ABC multidrug transporter [Exophiala dermatitidis]
OG0000038	0	7	7	AAT family amino acid transporter [Exophiala dermatitidis NIH/UT8656]

OG0000108	0	5	5	5-oxoprolinase (ATP-hydrolysing) [Exophiala dermatitidis NIH/UT8656]
OG0008269	0	2	2	hypothetical protein HMPREF1120_05993 [Exophiala dermatitidis NIH/UT8656] DUF300-domain-containing protein
OG0008270	0	2	2	hypothetical protein HMPREF1120_08910 [Exophiala dermatitidis NIH/UT8656] P-loop containing nucleoside triphosphate hydrolase protein
OG0000020	8	0	8	ABC multidrug transporter [Exophiala dermatitidis]
OG0000043	6	0	6	AAT family amino acid transporter [Exophiala dermatitidis]
OG0000101	5	0	5	5-oxoprolinase (ATP-hydrolysing) [Exophiala dermatitidis]
OG0000020	8	0	8	ABC multidrug transporter [Exophiala dermatitidis]
OG0000043	6	0	6	AAT family amino acid transporter [Exophiala dermatitidis]
OG0000101	5	0	5	5-oxoprolinase (ATP-hydrolysing) [Exophiala dermatitidis]
OG0000377	3	0	3	hypothetical protein A1O5_09736 [Cladophialophora psammophila]
OG0001397	2	0	2	hypothetical protein KCU88_g6189 [Aureobasidium melanogenum]

Lineage-specific unassigned paralogs

Full list can be found in Zenodo as OrthoFinder/Orthogroups_UnassignedGenes.tsv
doi: 10.5281/zenodo.7106110

The genome content was further compared using OrthoFinder (Emms and Kelly 2019) (**Table 5).** OrthoFinder identified 8,256 orthologous groups or 17,640 orthologous protein-coding genes between the two genomes. Both strains had a number of unassigned genes that belong to orthogroups, 705 for DCF04 and 475 for NIH/UT8656, along with 58 genes that were assigned, 34 DCF04 genes and 24 NIH/UT8656 genes. Of the 34 isolate-specific assigned genes in DCF04, 21 had identifiable fungal homologs with NIH/UT8656 including ABC multidrug transporters, AAT family amino acid transporter, 5-oxoprolinase and hypothetical protein/P-loop containing nucleoside triphosphate hydrolase protein. Two of the 13 assigned protein-coding genes resulted in a ribonuclease HI protein, while the remaining assigned 11 resulted in hypothetical protein matches. 4 out of the 5 orthogroups specific to NIH/UT8656 (22 protein-coding genes) were identical to the 4 seen in DCF04, while 1 orthogroup matched to a hypothetical protein/DUF300-domain-containing protein. When analyzing the unassigned genes, these functionally had no known paralog on NCBI. We believe these results reflect differences in gene prediction pipelines as much as it could be due to gene content differences (Weisman *et al.* 2022**.**

All CF isolates are MAT1-1 mating type

The small-scale genome synteny evaluation tool clinker was used to visualize slices of the genome adjacent to the identified MAT loci. All the lung isolates including DCF04/Ex4 *E. dermatitidis* isolate encoded a MAT1-1 locus (**Figure 3, Figure 4, Table 6**). For all other *E. dermatitidis* isolates and *Capronia* species analyzed, *SLA2*, an SRC-like adapter protein, and *APN2*, apurinic-apyrimidinic endonuclease 2, flanked the MAT loci in all isolates, and a hypothetical protein between SLA2 and MAT1-1-4 was also present across isolates. The

MAT locus found in strain NIH/UT8656 has the MAT1-2 mating type, while our clinical isolates have the MAT 1-1 mating type. Consistent with MAT loci described in *E. dermatitidis* (Metin *et al.* 2019) DCF04 had both MAT1-1-4 and MAT1-1-1 genes (**Figure 3**). The homothallic outgroup *Capronia coronata CBS617.96* genome (Teixeira *et al.* 2017) contains both MAT1-1 and MAT1-2 genes.

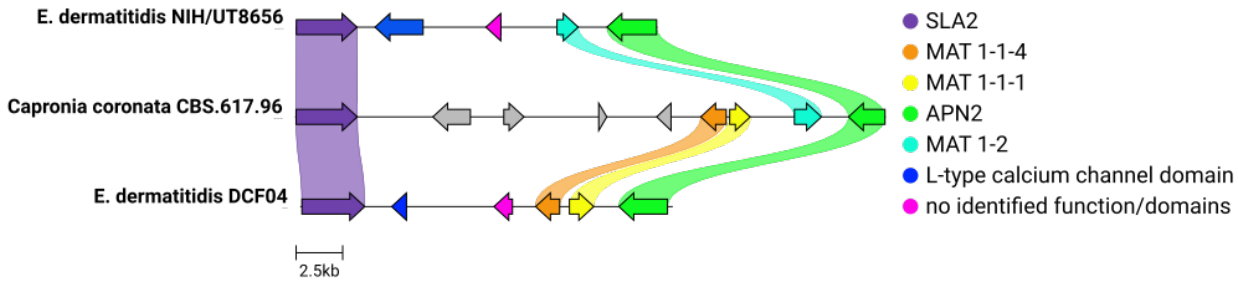


Figure 3. Mating-type determination of 4 clinical isolates of *E. dermatitidis*. Gene content, order, and orientation of the MAT locus from *E. dermatitidis* DCF04 and NIH/UT8656 strains and the Chaetothyriales black yeast *Capronia coronata CBS 617.98*. The locus is flanked by two genes SLA2 (purple) and APN2 (green). The MAT 1-1 gene, MAT 1-1-4 (orange), and MAT 1-1-1 (yellow) are observed in DCF04 strains while the reference strain NIH/UT8656 has a MAT 1-2 (teal) gene. In addition, two genes are predicted in the interval between SLA2 and the MAT genes. One is an L-type calcium channel domain (blue), which is only found in the *E. dermatitidis* strains, and a gene with no identified function or domains (pink) which is syntentically adjacent to the MAT genes in all strains and species.

Figure 4. Mating-type determination of 23 clinical isolates of *E. dermatitidis*. Gene content, order, and orientation of the MAT locus from 23 CF isolated *E. dermatitidis*, NIH/UT8656 strain, and the Chaetothyriales black yeast *Capronia coronata* CBS 617.98. The locus is flanked by two genes SLA2 (purple) and APN2 (green). The MAT 1-1 gene, MAT 1-1-4 (orange), and MAT 1-1-1 (yellow) are observed in all 23 CF strains while the reference strain NIH/UT8656 has a MAT 1-2 (teal) gene. In addition, two genes are predicted in the interval between SLA2 and the MAT genes. One is an L-type calcium channel domain (blue), which is only found in the *E. dermatitidis* strains, and a gene with no identified function or domains (pink) which is syntentically adjacent to the MAT genes in some strain.

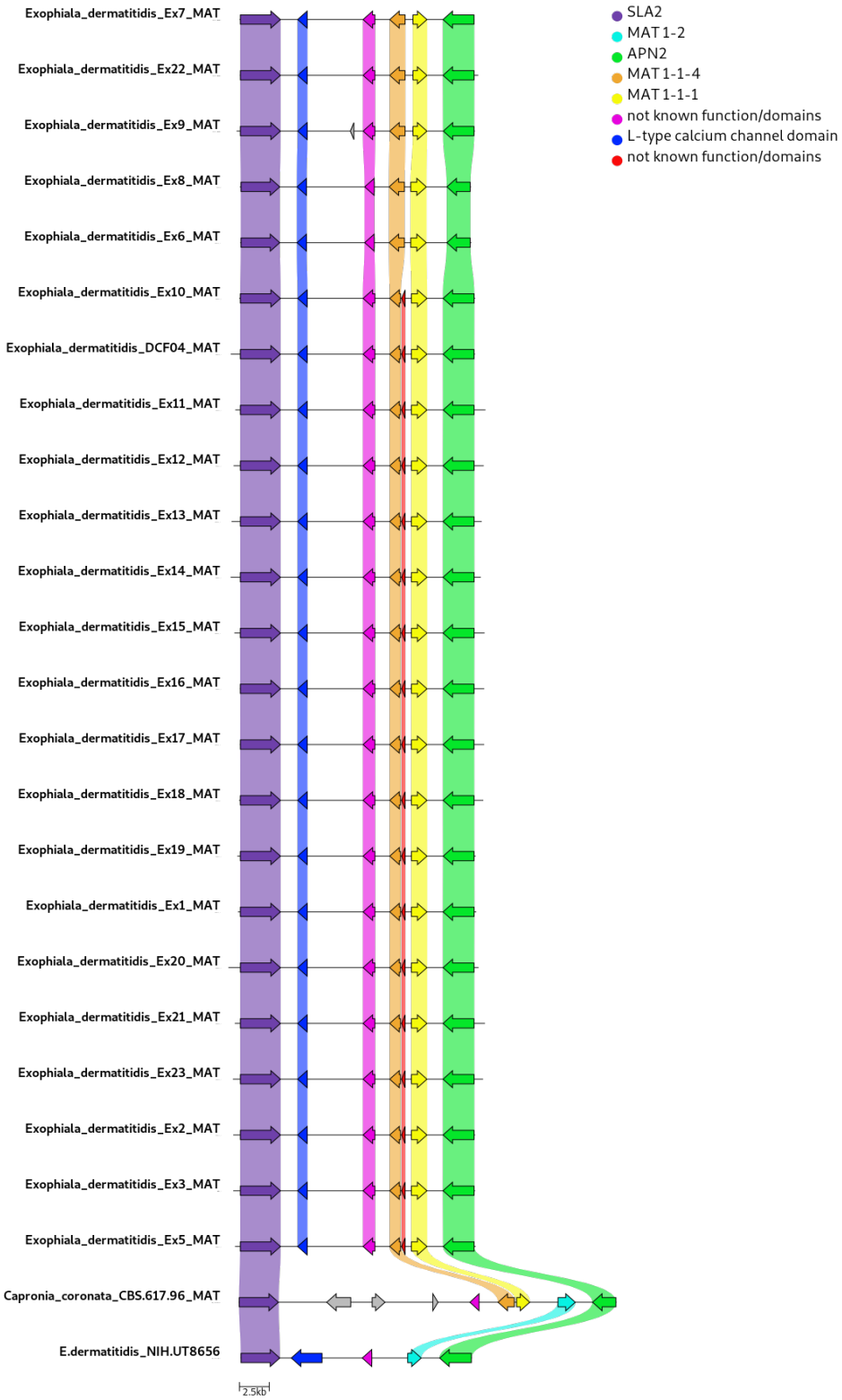

Exophiala_dermatitidis_Ex7_MAT
Exophiala_dermatitidis_Ex22_MAT
Exophiala_dermatitidis_Ex9_MAT
Exophiala_dermatitidis_Ex8_MAT
Exophiala_dermatitidis_Ex6_MAT
Exophiala_dermatitidis_Ex10_MAT
Exophiala_dermatitidis_DCF04_MAT
Exophiala_dermatitidis_Ex11_MAT
Exophiala_dermatitidis_Ex12_MAT
Exophiala_dermatitidis_Ex13_MAT
Exophiala_dermatitidis_Ex14_MAT
Exophiala_dermatitidis_Ex15_MAT
Exophiala_dermatitidis_Ex16_MAT
Exophiala_dermatitidis_Ex17_MAT
Exophiala_dermatitidis_Ex18_MAT
Exophiala_dermatitidis_Ex19_MAT
Exophiala_dermatitidis_Ex1_MAT
Exophiala_dermatitidis_Ex20_MAT
Exophiala_dermatitidis_Ex21_MAT
Exophiala_dermatitidis_Ex23_MAT
Exophiala_dermatitidis_Ex2_MAT
Exophiala_dermatitidis_Ex3_MAT
Exophiala_dermatitidis_Ex5_MAT
Capronia_coronata_CBS.617.96_MAT
E.dermatitidis_NIH.UT8656
2.5kb
SLA2
MAT1-2
APN2
MAT1-1-4
MAT1-1-1
not known function/domains
L-type calcium channel domain
not known function/domains

Table 6. Mating-type determination locus name descriptions of *E. dermatitidis*. This table describes all the proteins depicted in Figure 2 of the gene content, order, and orientation of MAT locus from *E. dermatitidis* isolates in this study, strain NIH/UT8656 along with the Chaetothyriales black yeast *Capronia coronata CBS 617.98*.

Species	SLA2 (purple)	L-domain (blue)	No known function/domain (pink)	MAT 1-2 (teal)	MAT 1-1-4 (orange)	MAT 1-1-1 (yellow)	APN2 (green)
NIH/UT8656	HMPREF1120_08859	HMPREF1120_08860	HMPREF1120_08861	HMPREF1120_08862	x	x	HMPREF1120_08863
DCF04	HRR96_009253	HRR96_009234	HRR96_009235	x	HRR96_009236	HRR96_009237	HRR96_009238
CBS 617.98	A1O1_07962	x	A1O1_07968	A1O1_07971	A1O1_07969	A1O1_07970	A1O1_07972

Chromosome copy number variation across E. dermatitidis isolates

Copy number variation of full or partial chromosomes was evaluated by calculating the depth of coverage using 10kb sliding windows (**Figure 5).** The read depths of windows across *E. dermatitidis* isolate genomes from this study were compared across all 9 chromosomes. Visual scanning of the plots identified an anomaly of at least 1.5x higher coverage on chromosome 5 in Ex3 (**Figure 5A**). A similar but much smaller region of chromosome 5 appears to have 2-2.25x coverage and may be duplicated as an aneuploid in Ex13. Additional partial 1.25-1.5x coverage for part of the left arm of chromosome 2 in isolate Ex18 is also observed.

The plots of isolates Ex15, Ex18, and Ex20 indicate 0.5x less normalized coverage in chromosomes 1 and 3 (**Figure 5B and Figure 5C**), a possible sign of a segmental chromosomal aneuploidy event. The continuation of the coverage pattern between chromosomes 1 & 3 seems to complement each other for these isolates. Haploid organisms, like *Exophiala dermatitidis*, could benefit from genomic plasticity through expansion or contraction to enable adaptation in a new environment (Selmecki *et al.* 2010; Legrand *et al.* 2019). Noting that Ex3 was isolated in the early time point, this genome copy may contribute to an adaptive mechanism to support colonization and persistence in the human host. Other isolates sharing potential aneuploidies (Ex15, Ex18, Ex20) are from the late time point and also have the highest mutation rate among this population as observed from the phylogenetic tree and mutation rate calculation (**Figure 6 and Table 11**).

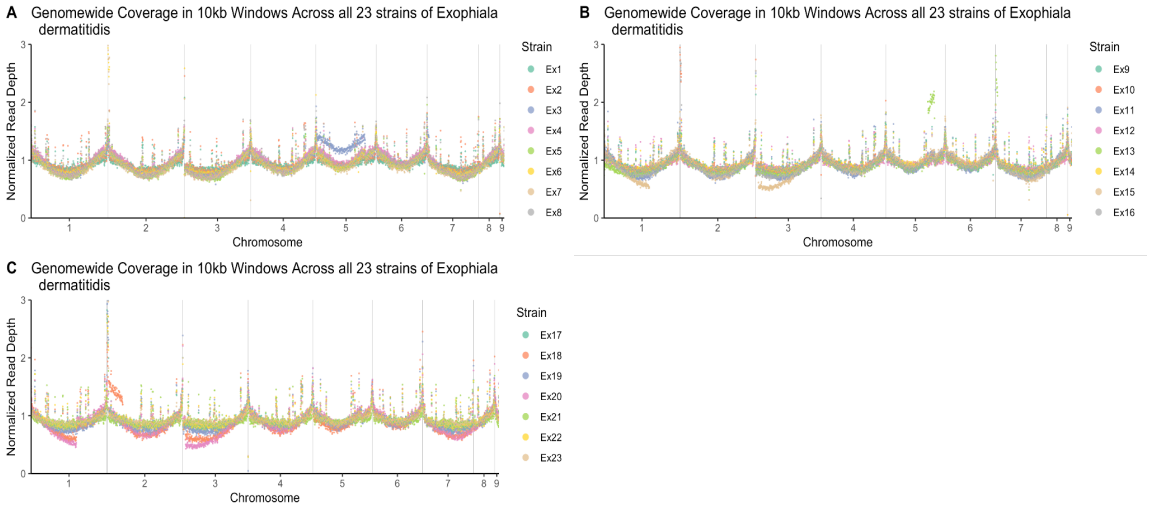


Figure 5. Genome sequencing depth coverage visualization to test for copy number variation across isolates. Visualization of the depth of coverage was generated by plotting a normalized read depth across chromosomes for isolates **A)** Ex 1-8, **B)** Ex9-16, and **C)** Ex17-23.

SNP genotyping and SNP-based phylogenetic analysis of 23 E. dermatitidis isolates

We used the DCF04/Ex4 isolate as a reference for variant identification within the 23 CF isolate collection. The DCF04/Ex4 VCF file generated 441 variants between twenty-two isolates and the DCF04 reference isolate. A phylogenetic tree was generated using the filtered SNP data results and rooted with the earliest diverging group containing Ex1, Ex2, Ex12, and Ex14 (**Figure 6**). When repeated for out-population reference strain NIH/UT8656 the analysis resulted in 11,000 variants between the NIH/8656 strain and the CF isolates (Kurbessoian 2022). Another phylogenetic tree was generated using a secondary dataset created with the NIH strain to be used as a root (**Figure 7**).

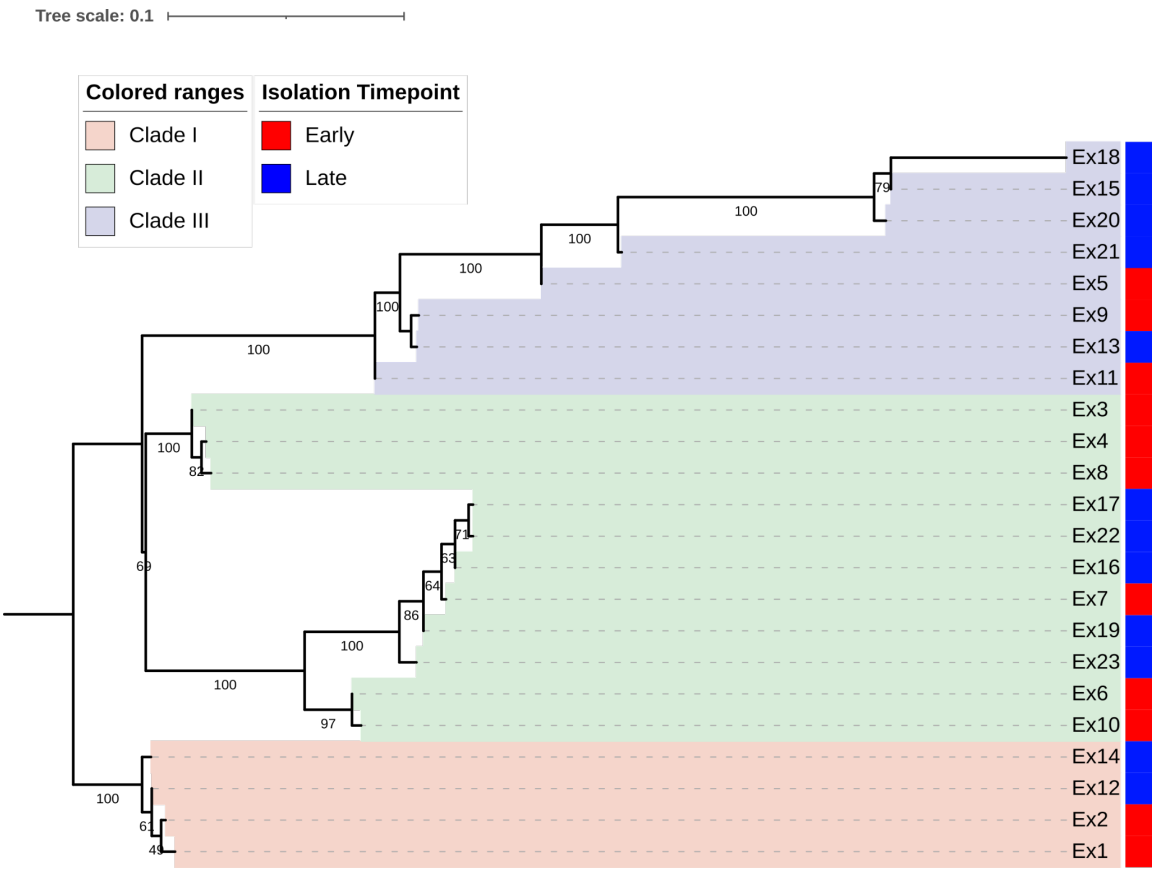


Figure 6: Phylogenetic tree of 23 isolates. A Maximum-Likelihood phylogenetic tree constructed from the Single Nucleotide Variants by IQTREE2 was identified from the isolate resequencing data. The tree is rooted with the Clade I branch based on additional analyses that included NIH/UT8656 as an outgroup.

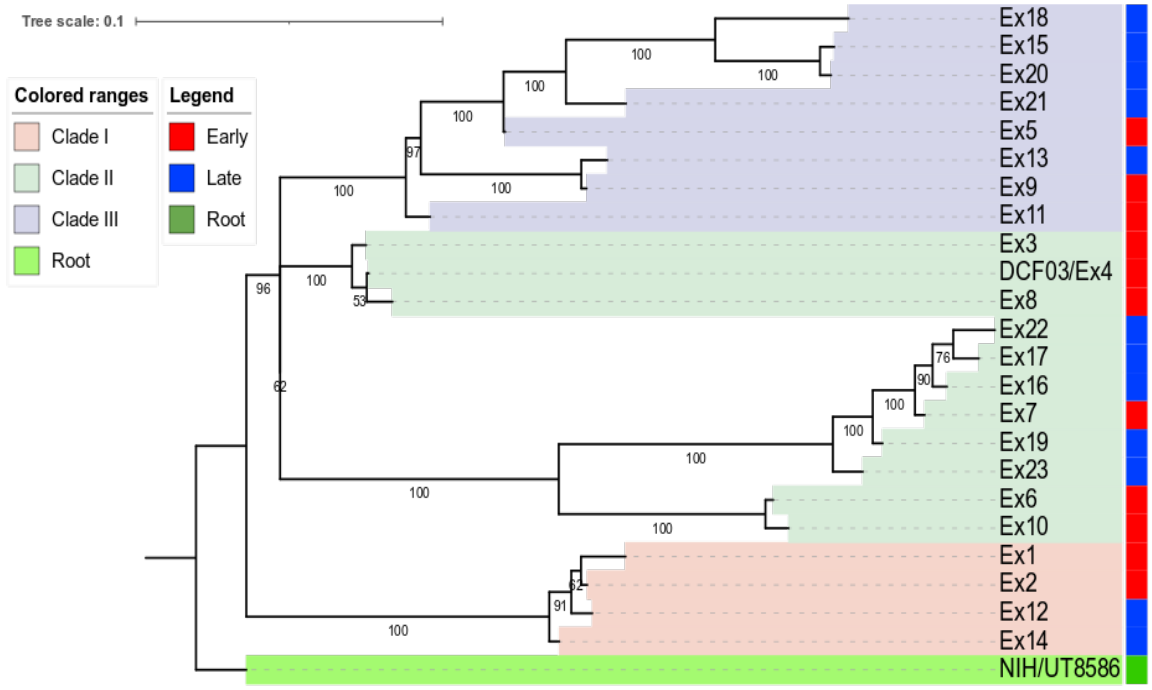


Figure 7. Phylogenetic tree of 24 isolates. A Maximum-Likelihood phylogenetic tree constructed from the Single Nucleotide Variants by IQTREE2 was identified from the isolate resequencing data. The tree is rooted with NIH/UT8656 as an outgroup. Isolates are labeled as one of three clades based on the phylogenetic relationships and the isolation time point is indicated with a red (early) or blue (late) colored box.

Three clades (Clade I, Clade II, and Clade III) were identified based on the tree. Clade I is composed of two early collected and two late collected isolates of *E. dermatitidis* Ex1, Ex2, Ex12, and Ex14. Clade II is composed of three subgroups, one of which contains only early isolates and a second group with both early isolates and late isolates. The third group in Clade II contains two early collected isolates Ex6 and Ex10. Clade III has two main groups composed of both early and late isolates. The subgroup of Clade III had a long branch length suggesting more divergence. Interestingly, the CNV plot (**Figure 5**) showed these three isolates, Ex15, Ex20, and Ex18 contained similar CNV differences in chromosomes 1 and 3 when compared to the other isolates.

Non-synonymous and synonymous SNP and INDEL pairwise differences

A dissimilarity matrix was constructed comparing the overlaps of SNPs and INDELs collated for all pairs of isolates. As expected, isolates that are closest to each other in the SNP-based phylogenetic tree had fewer differences in their SNP composition (**Figure 8A**). Our analysis was further supported by the observation that the number of SNP differences correlated with the number of INDELs detected in pairwise analyses (**Figure 8B**). Within Clade I and Clade II, isolates had few SNP and INDEL differences, indicating the evolutionary distance between them was small. The SNP and INDEL counts within Clade III were much higher indicating a higher divergence within this group of isolates as compared to other groups.

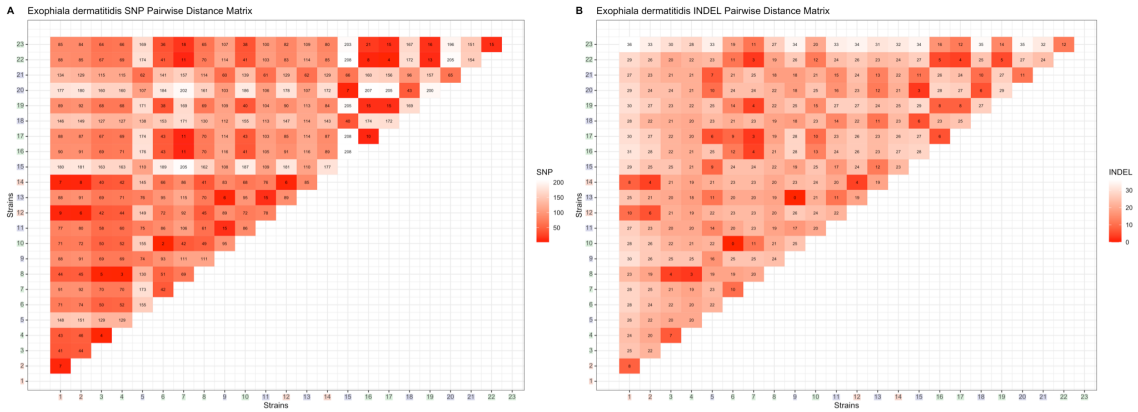


Figure 8. Matrix of *E. dermatitidis* CF isolates SNP and INDEL pairwise dissimilarities. The number of SNPs **(A)** and **(B)** INDELs that differ among pairs of isolates. The more similar isolate pairs have lower (darker red) numbers and the color approaches white indicating more dissimilar isolates. Clade I and Clade II isolates generally differ by very few SNP and INDELs (noting a few exceptions) consistent with their inferred near phylogenetic relationships. Within Clade III isolate pairs differ by more SNPs which may indicate a higher mutation rate within these isolates.

Table 7. Genes with SNP variants were found stratified by clade. Isolates collected in the early (red) and late (blue) time points are labeled in the comparisons. The functions of genes of interest found to have nonsynonymous mutations among the entire clade of comparisons are reported with the total count of nonsynonymous and synonymous (NonSyn/Syn) variants found in each clade to identify the relative frequency of these changes and general functional differences in the types of genes with variants across the population clades.

Clade	Isolates	Nonsynonymous mutations of interest in clade	Number of NonSynonymous/Synonymous SNPs unique to clade
I	**Ex1, Ex2, Ex12, Ex14**	MADS box Transcription factor, G4 quadruplex binding protein, MFS transporter SP family sugar:H+ symporter, Bud site selection Bud4, DNA polymerase alpha subunit A, GTPase-activating protein	59/7
II	a) **Ex3, Ex4, Ex8** b) **Ex22, Ex17, Ex16, Ex7, Ex19, Ex23** c) **Ex6, Ex10**	a. All hypothetical. b. Ex7 - Ex22 GO terms indicate Signal Transduction Mechanism and Transcription Factor, Ex7-Ex17 Zinc finger binding protein TFIIIA, Ex16 queuine tRNA-ribosyltransferase, MFS transporter SP family sugar:H+ symporter, NAD-dependent histone deacetylase SIR2 and cytochrome P450, family 7, subfamily B (oxysterol 7-alpha-hydroxylase) c. None.	56/8
III	a) **Ex9, Ex13, Ex11** b) **Ex18, Ex15, Ex20, Ex21, Ex5**	a. Ex13 & Ex9, Ex11 Ap-3 complex subunit delta, DNA repair protein RAD50, and Regulator of nonsense transcripts 1-like protein. b. Ex18 & Ex15, Ex20 an MFS transporter DHA2 family methylenomycin A resistance protein, sulfite reductase (ferredoxin), glycerol ethanol-ferric requiring protein, polyketide synthase, MFS transporter SP family solute carrier family 2 and a DNA repair protein RAD20	125/20

Non-synonymous SNP Differences seen between clades

The human-readable snpEff table which best described the variant analysis along with functional gene annotation can be found in **Table 8 (Github),** but is summarized in **Table 7**. Clade I is composed of four isolates: two (Ex1, Ex2) and two isolates from the late time point (Ex12, Ex14). Two genes with mutations are of note, a MADS-box transcription factor (HMPREF1120_06786) and a G4 quadruplex nucleic acid binding protein (HMPREF1120_02174). The MADS-box transcription factor is part of the MADS-box proteins with a highly conserved 56 amino acid DNA-binding domain, some containing a weakly conserved K-box domain that is involved in the dimerization of transcription factors (Shore and Sharrocks 1995). In fungi, isolates with knocked-out MADS-box genes have reduced virulence (Damveld *et al.* 2005; Qu *et al.* 2014; Xiong *et al.* 2016). Mutations in *E. dermatitidis* MADS-box regions may increase their pervasiveness and tolerance of the lung environment. Previous studies have found G4 quadruplex nucleic acid binding proteins. The helical complex is formed through guanine-rich nucleic acid sequences and is found at the telomeric regions of chromosomes.

Table 8. Functional impact of identified variants. (See GitHub) Filtered human readable snpEff tabular results for all CF 23 *E. dermatitidis*. Annotations have been added to the final list to better help assess the function of each protein while also observing variants detected from GATK.

Clade II contains 11 isolates, six from the early time point and five from the later time point, that fall into three different subgroups. Ex3, Ex4/DCF04, and Ex8 fell into one group, Ex22, Ex17, Ex16, Ex7, Ex19 and Ex23 fell into the second group, and Ex6 and

Ex10 formed a third group. Group one and group three both contain only early isolates, while the second group contains the majority of the later isolates.

Clade III, which contains eight isolates, is also divided into two groups: one composed of Ex18, Ex15, Ex20, Ex21, and the other group of Ex9, Ex13, and Ex11. Five are later isolated while the other three are from the early isolations. The most significant group is the second consisting of Ex9, Ex13, and Ex11. Four genes were found to have distinct mutational differences when comparing Ex13 (a late isolate) to the early isolates Ex9 and Ex11. These four genes are the Ap-3 complex subunit delta (HMPREF1120_00143), DNA repair protein RAD50 (HMPREF1120_05599), a regulator of nonsense transcripts 1-like protein (HMPREF1120_06837) which is similar to helicase-RNA complex, and a gene with no identified function (HMPREF1120_02556). The second group had three isolates (Ex18, Ex15, and Ex20) with a higher mutation rate than *E. dermatitidis* in Clade III. There are about 13 instances of hypothetical proteins, while the other 12 instances are predicted genes. Only six of the twelve are genes of note: an MFS transporter DHA2 family methylenomycin A resistance protein (HMPREF1120_00012), sulfite reductase (ferredoxin) (HMPREF1120_00943), glycerol ethanol-ferric requiring protein (HMPREF1120_01007), polyketide synthase (HMPREF1120_03173), MFS transporter SP family solute carrier family 2 (HMPREF1120_06771) and a DNA repair protein RAD50 (HMPREF1120_05599). Research on polyketide synthases in micro-colonial fungi and *E. dermatitidis* has been found to impact phenotypes and adjust the melanin synthesis pathway, resistance or susceptibility to antifungals, and extreme environment adaptability (Paolo *et al.* 2006).

Analysis of SNP differences between clades

Finally, a single non-synonymous mutation in a gene orthologous to *S. cerevisiae MRS4* (HMPREF1120_06597), a mitochondrial iron transporter. One allele of the *MRS4* allele encoded a protein that was identical in sequence to the allele in the reference NIH strain UT8656, present in the isolates in Clade I (Ex1, Ex2, Ex12, Ex14), and a subclade of clade II (Ex3, Ex4/DCF04, Ex8). The remainder of the isolates had a second allele with a non-synonymous mutation in the 40th amino acid position, converting a glutamic acid residue to glycine. The functional consequences and differences of these Mrs4 alleles will be described in a separate manuscript.

Testing for the enrichment of evolutionary patterns within clades

A pairwise comparison of the synonymous and nonsynonymous SNP differences was performed on fifteen pairs of isolates identified as early and late members of the same lineage. **Table 3** summarizes the significant results with a focus on candidate genes that may relate to lung pathogenicity. A comprehensive list of gene differences among all pairwise comparisons is in **Table 9 (Github)** and includes hypothetical proteins with no identified function. This analysis tested differences between early and late isolates found in the same clade to focus on changes that may have occurred within the host.

Table 9. All functional SNP and INDEL results for early and late pair (See GitHub). Results indicated in this table include all hypothetical or undescribed results along with results described in Tables 2 and 3. Proteins listed include Protein ID numbers to better facilitate identification.

Table 10. Genes with nonsynonymous SNP variants of interest were observed in pairwise comparisons of *E. dermatitidis* CF lung isolates. Genes with notable non-synonymous differences were identified in pairwise comparisons of early and late time point isolates. Full results of all pairwise differences are in **Supplemental Table 7**. For continuity, gene locus names are in parentheses and refer to the locus names in the reference strain *E. dermatitidis* NIH/UT8656.

Early vs. Late	Function
Ex5 & Ex18	MFS transporter, DHA2 family, methylenomycin A resistance protein (HMPREF1120_00012) DNA repair protein RAD50 (HMPREF1120_05599) cytochrome P450 oxidoreductase (HMPREF1120_01361) sulfite reductase (ferredoxin) (HMPREF1120_00943) sulfite oxidase (HMPREF1120_01306) MFS transporter, DHA1 family, multidrug resistance protein polyketide synthase (HMPREF1120_6570) MFS transporter, SP family, sugar:H+ symporter (HMPREF1120_04157) tyrosinase (HMPREF1120_04514)
Ex9 & Ex13	G2/mitotic-specific cyclin 3/4 (HMPREF1120_00797) cytochrome P450 oxidoreductase (HMPREF1120_01361) DEAD-box RNA helicase HelA (HMPREF1120_02010) MFS transporter, DHA1 family, multidrug resistance protein (HMPREF1120_01715) MFS transporter, SP family, sugar:H+ symporter (HMPREF1120_04157) DNA repair protein RAD50 (HMPREF1120_05599) ISU1 - iron-binding protein (HMPREF1120_06751) MFS transporter, SIT family, siderophore-iron:H+ symporter (HMPREF1120_07838)

The analysis considered all pairwise combinations between the early Ex1 and Ex2 and the late Ex12 and Ex14 isolates from Clade I. These isolates show very little genetic differentiation (about 6-9 variant SNPs and 6-10 variant INDELs). The contrast of early Ex1 vs Ex2 isolates to the late isolates (Ex12, Ex14) found variants in three genes: AFG2- an ATPase (HMPREF1120_06104), a DNA polymerase (HMPREF1120_07994), GYP1- a GTPase activating protein (HMPREF1120_08601), and MFS transporter (HMPREF1120_04157) homologs. These variants are an interesting observation and warrant extra analysis in the future.

Pairwise comparisons of the Clade II isolates contrasted the early Ex7 with each of the four late isolates Ex16, Ex17, Ex19, and Ex23. There was more variation among these isolates (about 11-169 variant SNPs and 3-16 variant INDELs) than observed in Clade I isolates. Isolate Ex23 also appeared to have substantially more differences from all others indicating it may be more distantly related or where its corresponding ancestral early strain was not sampled. A variant of note, NAD-dependent histone deacetylase SIR2, is involved with chromosomal remodeling specifically with phenotype transcription modification (Freire-Benéitez *et al.* 2016).

Comparison of Clade III group 1 member Ex5 (early) to Ex15,18,20,21 identified the highest number of variants as compared to the other clades (from 62-138 variant SNPs and 7-10 variant INDELs). Clade III group 2 members Ex9 (early), Ex11 (early), and Ex13 had a similar number of variants as Clades I & II (6-15 variant SNPs and 0-17 variant INDELs). The five pairwise comparisons for group 1 revealed a non-synonymous mutation in the RAD50-DNA repair protein (HMPREF1120_05599). Group 1 members in the phylogenetic tree also appear to have a long branch length indicating a potentially higher diversification rate than the other clades. It may be that the RAD50 changes could have an impact on DNA repair and contribute to the higher mutation rate observed. In addition, non-synonymous mutations were identified in a cytochrome P450 oxidoreductase (HMPREF1120_01361), MFS transporter sugar to H+ symporter (HMPREF1120_04157), and GTPase activating proteins (HMPREF1120_08601). SNP variants with predicted functional impact on interactions with the host and CF lungs were found in the sulfite

oxidase (HMPREF1120_01306), sulfite reductase (HMPREF1120_00943), polyketide synthase (HMPREF1120_03173), and tyrosinase (HMPREF1120_04514) genes.

Interestingly, the group 2 (Clade III) isolates Ex9, Ex11 and Ex13 had more variants in iron-binding (HMPREF1120_06751) and siderophore-related (HMPREF1120_07838) genes than group 1. We took a candidate gene approach and tested if the siderophore NPRS SidC (HMPREF1120_07636) had accumulated any specific mutations in these lineages, but we did not identify any non-synonymous variants in this gene across the CF lung isolates. Though, the presence of SNP variants in iron-binding and siderophore transporters indicates there may be selective pressure for *E. dermatitidis* to obtain iron in the CF lung environment, in which iron is strictly regulated. Further analysis of these variants will point to supporting phenotypic differences allowing for effective colonization of CF lung environments.

Mutation rate calculation to test for different rates of evolution between clades

We calculated a mutation rate for each of the three clades (**Table 11**) by taking an average of all the pairwise comparisons with a clade. The Clade I average of the four pairwise calculations for all combinations of early and late isolates was 6.78E-08 variants/pair/2 years. The average rate for the five comparisons of early to late isolates in Clade II was 2.20E-07 variants/pair/2 years. The average rate for the six comparisons of early to late isolates in Clade III the first group values were averaged and determined to be 7.39E-07 variants/pair/2 years, while the second group values were averaged and determined to be 1.13E-07 variants/pair/2 years. The Clade III group 1 isolates appear to be a faster evolving

group and may have acquired variants allowing improved adaptation to the lung environment. One-way ANOVA test revealed a statistically significant difference between all Clades (F-value = 7.0711, p-value = 0.00104) and a Tukey posthoc test across the clades indicated Clade III (Group 1) has a significantly different mutation rate. Calculations underlying the mutation rate values are detailed in **Supplemental Table 8**.

Table 11. Average mutation rates for each clade/group. Mutation rates were calculated for each clade of isolates and two sub-groups of Clade III. Calculations were summarized as the median of all pairs of isolates within a group or clade.

Groups or Clades	**Pairs**	**Mutation Rates**
Clade I	Ex1+Ex12, Ex1+Ex14, Ex2+Ex12, Ex2+Ex12	6.76E-08 variants/pair/months
Clade II	Ex7+Ex16, Ex7+Ex17, Ex7+Ex19, Ex7+Ex22, Ex7+Ex23	2.20E-07 variants/pair/months
Clade III (Group I)	Ex5+Ex15, Ex5+Ex18, Ex5+Ex20, Ex5+Ex21	7.39E-07 variants/pair/months
Clade III (Group II)	Ex9+Ex11, Ex9+Ex13, Ex11+Ex13	1.13E-07 variants/pair/months

Table 12. All 23 mutation rates were calculated (See Github). Mutation rates for each 23 CF *E. dermatitidis* isolates were calculated using the formula described in the methods. One-way ANOVA was run to detect significance (p-value= 0.00104, F-value = 7.0711) along with Tukey multiple comparisons of means indicating Clade III had the highest significance among the six pairwise comparisons.

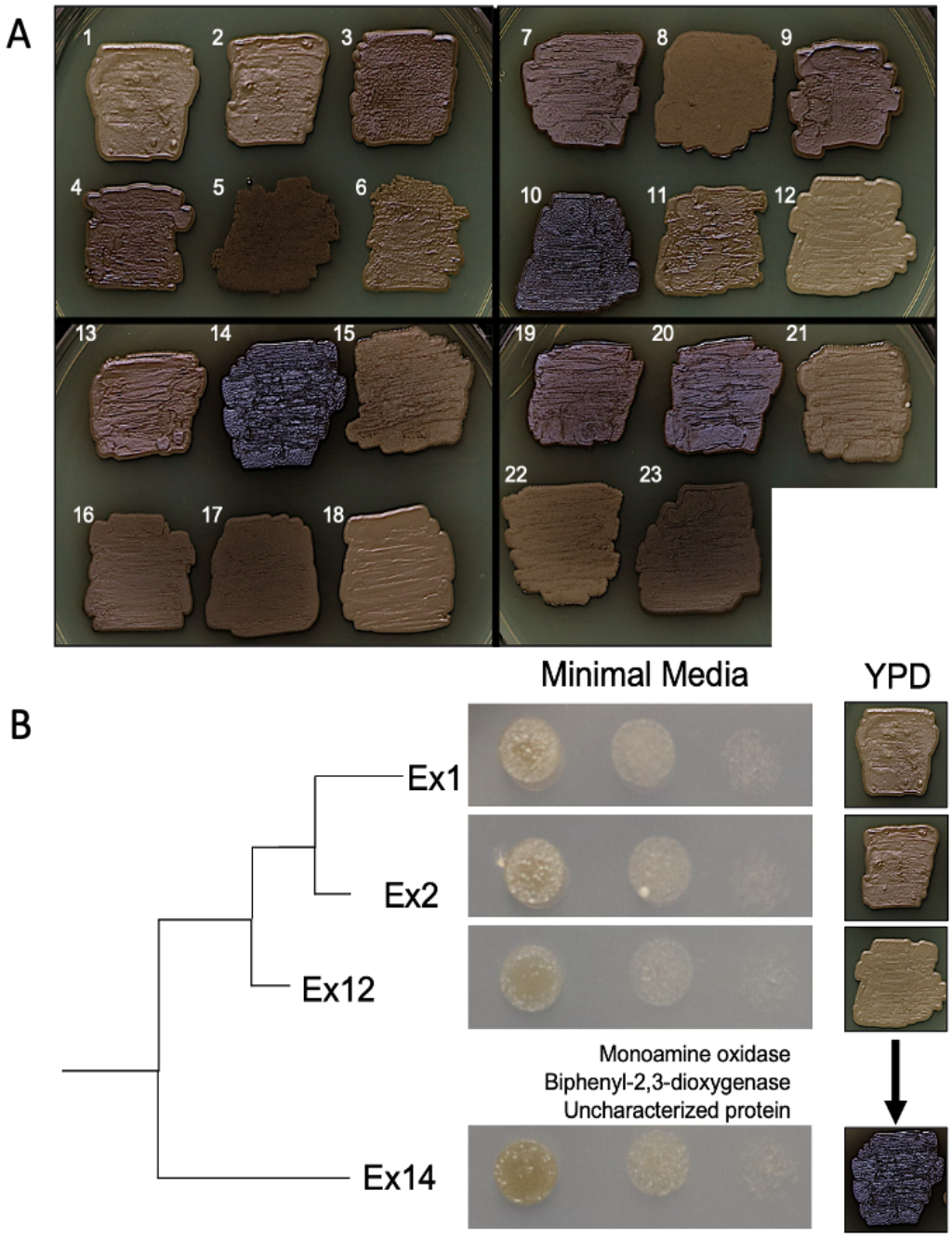


Figure 9. ***E. dermatitidis*** **clades and their comparisons to melanin phenotype and MIC. (A)** Isolates were struck from freezer stock onto YPD plates and imaged at after 120-hour incubation h at 37˚C to compare melanin accumulation. **(B)** Strong differences in melanin production between related isolates may be due to intergenic mutations in the 5' UTR of a biphenyl-2,3-dioxygenase encoding gene were observed, with as few as 3 SNPs differentiating low melanin producers from the high melanin producer.

Heterogeneous phenotypes observed in the E. dermatitidis isolates

The twenty-three *E. dermatitidis* isolates (Ex1-11 from the early time point and isolates Ex12-23 from the late time point) were heterogeneous for traits in a number of ways including pigmentation, antifungal sensitivity, and auxotrophy. We noted differences in melanin pigmentation across isolates (**Figure 9A**) that did not correlate with either time point or phylogenetic clade. Within Clade 1, Ex1, Ex2, and Ex12 had light melanization, but isolate Ex14 was one of the strongest melanin producers (**Figure 9B**). In a search for genetic determinants responsible for this phenotypic range, we identified three candidate intergenic mutations which differentiated Ex14 from the other members of the clade.

The 23 isolates also varied in sensitivity to itraconazole, a recommended treatment for *E. dermatitidis* infections (Fothergill *et al.* 2009; Mukai *et al.* 2014), over a ~10-fold range (minimum inhibitory concentrations (MIC) from 0.0625-0.5 µg/ml) (**Figure 10**). Heterogeneity in amino acid auxotrophy scored as no growth on a minimal medium that was rescued by supplementation of amino acids (**Figure 10**), was also observed. Lastly, there were stable differences in filamentation across isolates in two of the three clades; closely related strains did not have similar morphologies (**Figure 11**).

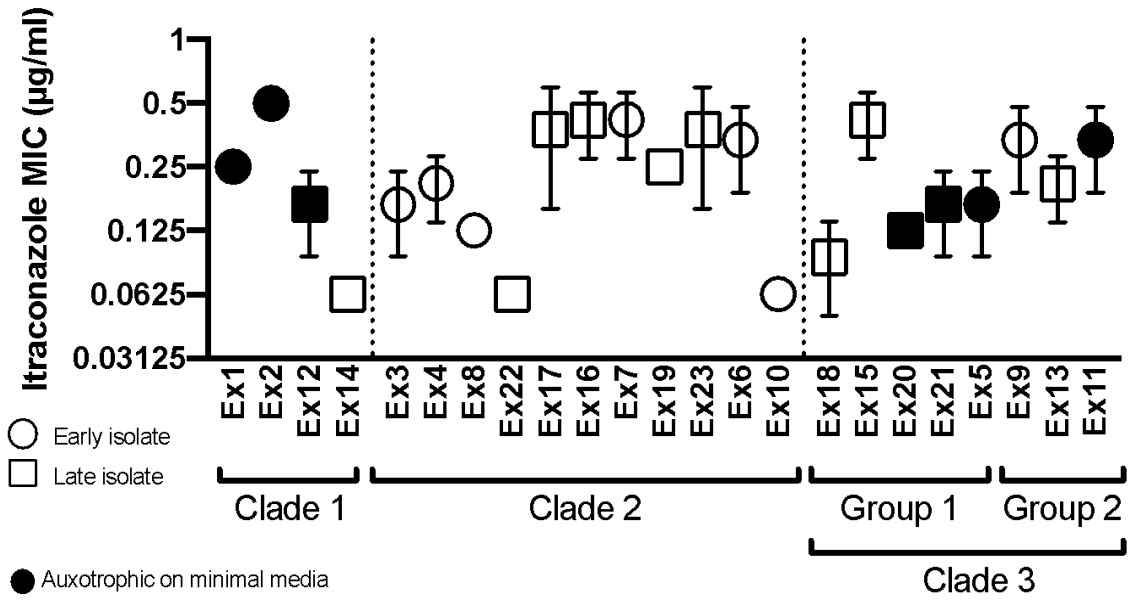


Figure 10. Itraconazole MIC testing on 23 isolates of *E. dermatitidis,* while comparing clades. Isolates were grown for ~24 hours in liquid YPD at 30˚C in a rolling barrel culture, then inoculated into 96-well flat bottom plates at a concentration of 1000 CFU/well and allowed to grow for 72 hours at 37˚ before determining minimum inhibitory concentration (MIC), which is represented on the figure as the mean of three biological replicates. MIC was determined through visual inspection, as the concentration of azole which was sufficient to restrict visible growth. For determination of auxotrophy, cells were grown as described above, CFU equilibrated, and spotted onto rich media and minimal media (YNB) with and without the addition of casamino acids. Isolates with decreased growth on YNB relative to YPD and could be rescued with the addition of amino acids are noted as auxotrophic.

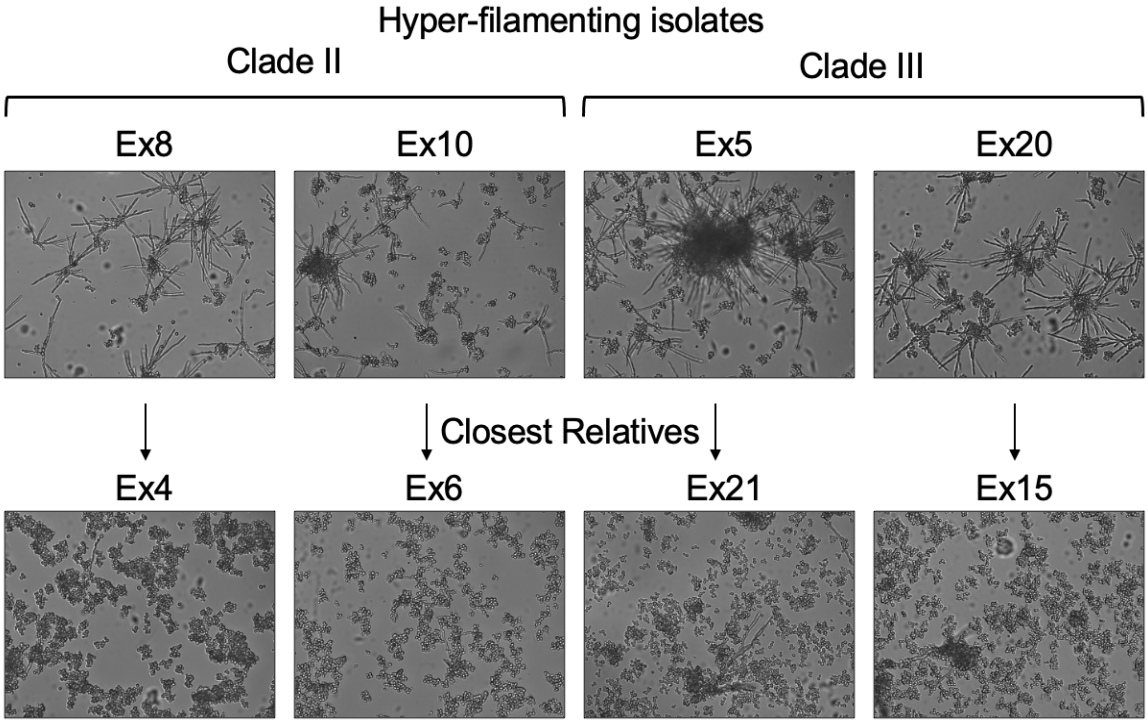


Figure 11. Cell phenotype microscopy. Isolates were inoculated at 1000 cfu/ml into RPMI-1640 in glass-bottom 24-well culture dishes, and incubated atwere grown for 18 h in RPMI, at 37˚C in 5% CO, and for 18 hours. Cultures were then imaged at 63X magnification, DIC with a Zeiss Axiovert 200M inverted microscope using DIC microscopy[2]. The first row of images depicts isolates displaying hyper-filamentous attributes and are distributed throughout their respective clades. The second row depicts the closest related isolate to each respective isolate in the first row, which lacks similar phenotypes.

Discussion

Fungus-dominant population

While opportunistic fungal pathogens are often culturable from the sputum of patients with cystic fibrosis, they do not commonly present as the dominant microbe or as a risk for infection. In this paper, we have presented a clinical case where a population of the black yeast *E. dermatitidis* was the predominant microbe concomitant with a lung exacerbation event. Other reported clinical cases have shown that *E. dermatitidis* has an underappreciated role as a CF pathogen (Haase *et al.* 1990; Kusenbach *et al.* 1992; Kondori *et al.* 2014). Previous sputum isolates revealed the presence of this uncommon fungi two years prior, indicating that it persisted in low concentrations before having the opportunity to dominate the lung microbiome. Because *E. dermatitidis* is relatively slow-growing (Rath *et al.* 1997; Sudhadham *et al.* 2011; Malo *et al.* 2021), it may be easily missed during routine clinical microbiological identification. To better understand the CF disease and the fungal infections associated with a long-term disease such as this one, it is important to consider testing for slow-growing microorganisms such as black yeasts.

Phenotypic heterogeneity; Melanin production, amino acid metabolism, filamentation, and drug resistance

Identification of key evolutionary strategies are crucial to the understanding of microbial pathogenesis in clinically relevant settings. Developing methods for evaluating population structure and heterogeneity for these human-associated microbes will aid in future studies. Initial observations of these isolates indicated a striking amount of heterogeneity in melanin production, which propelled an investigation into the genotypic diversity of the

population. Population heterogeneity is an important factor to consider in treatment, and determining adaptive responses through genomics can help identify selective stimuli in chronic infections (Demers *et al.* 2018).

Melanin production in fungi is associated with increased survival in stress conditions, virulence, dissemination, and host immune response (Eisenman and Casadevall 2012). *E. dermatiditis* produce DHN-melanin from 1,8 dihydroxynaphthalene polymerization, the more common and better characterized pathway in fungi; compared to L-DOPA produced DOPA-melanin or pyomelanin produced from degradation of tyrosine (Jia et al. 2021). DHN synthesis initiates from endogenous acetyl CoA or melanyl CoA, which is converted to 1,3,6,8 tetrahydroxynaphthalene by polyketide synthase, and converted to DHN melanin through a series of reductions and polymerizations. We observed striking levels of variation in melanin production, even in closely related isolates with few mutations differentiating them. Future experiments will determine whether these SNPs play a role in melanin synthesis or cell wall localization.

We expected that differing levels of melanin, which comprises part of the cell wall in black yeasts, would have an effect on antifungal resistance. While drug resistance to clinically relevant antifungals Itraconazole and Voriconazole varied slightly between isolates, there was no consistent pattern of evolution that corresponded either with a time of isolation or melanin production (**Figure 7**). We found amino acid auxotrophs in the population, and the isolation of auxotrophs in CF infections has been previously described in bacterial isolates, such as *P. aeruginosa* (Barth and Pitt 1995; Thomas *et al.* 2000). This repeated occurrence of filamenting phenotypes in phylogenetically diverse isolates may indicate

evolution towards hyper-filamentation or a non-mutation-based difference between isolates such as "switching" or phase variation, or epigenetic regulation.

We found amino acid auxotrophs in the population, and the isolation of auxotrophs in CF infections has been previously described in bacterial isolates, such as *P. aeruginosa* (Barth and Pitt 1995; Thomas *et al.* 2000). We also identified isolates which filamented in overnight cultures, and could be further stimulated to form hyphae in classic filamenting conditions of RPMI with 5% CO2. These isolates were not clustered in the population, indicating repeated evolution towards this morphology. This repeated occurrence of filamenting phenotypes in phylogenetically diverse isolates may indicate evolution towards hyper-filamentation or a non-mutation-based difference between isolates such as "switching" or phase variation, or epigenetic regulation.

Siderophore gene mutations and iron acquisition

Successful microbial invasions require iron, a critical growth-limiting factor, which must be sequestered from the surrounding environment through the release of iron chelating proteins called siderophores (Neilands 1995; Mossialos and Amoutzias 2009). For example, *Pseudomonas aeruginosa* produces the fluorescent siderophores, pyoverdine and pyochelin, which are used to sequester iron from the lung environment and as cofactors for respiratory proteins needed for surface motility and biofilm maturation (Haase *et al.* 1990; Banin *et al.* 2005; Matilla *et al.* 2007). Another example, the fungus *Aspergillus fumigatus* produces two hydroxamate-type siderophores: triacetylfusarinine (SidD) and ferricrocin (SidC), while also producing fusarinine and hydroxyferricrocin type siderophores (Schrettl

et al. 2007; Haas 2014). In a study by Zajc et al., (Zajc *et al.* 2019), indicated 10 predicted siderophores in four *Exophiala dermatitidis* genomes, two of which include triacetylfusarinine and ferricrocin. Ferricrocin is an internal siderophore used to store iron and is essential for sexual development and contributes to oxidative stress resistance (Schrettl *et al.* 2007; Tyrrell and Callaghan 2016). Triacetylfusarinine is used to facilitate hyphal growth under iron-depleted conditions (Schrettl *et al.* 2007; Tyrrell and Callaghan 2016).

The *E. dermatitis* locus (HMPREF1120_07636) depicts the non-ribosomal peptide synthase SidC necessary for siderophore synthesis (Zajc *et al.* 2019; Malo *et al.* 2021). Polymicrobial infections persist in the CF lung environment through production and scavenging of extracellular siderophores which aid microbes in competition for resources (Tyrrell and Callaghan 2016; Sass *et al.* 2019; Yan *et al.* 2022). Microbes can use, obtain, and sequester iron in siderophores from hemoglobin found in red blood cells and lactoferrin contained in mucosal secretions. The competition and use of iron is an important dynamic in polymicrobial infections including fungal-bacteria competition within plant and animal hosts, and may sometimes assist in promoting the growth of their host (Crowley *et al.* 1991; Johnson 2008; Aznar *et al.* 2014; Kim 2018; Mochochoko *et al.* 2021; Pohl Carolina H. and Noverr Mairi C. 2005).

Short and long read sequencing, in-population vs outside reference

While short-read sequencing provided us with the basic and high-quality read depth for each CF isolate, using the Oxford Nanopore long-read technology sequencing to build on

top of a short-read assembly provided us with a well-assembled genome. This genome was then used as an in-population reference allowing for a better recovery of variants due to the read mapping to a closer isolate. When using our in-population reference, compared to the NIH/UT8656 strain (Robertson *et al.* 2012; Chen *et al.* 2014; Schultzhaus *et al.* 2020; Malo *et al.* 2021), our quantification of population-specific number of variants is higher (due to higher sensitivity) while still maintaining relevance in our study system.

The temporal resolution of variation accumulation in a fungal CF lung population

Contrasting mutations accumulating in isolates collected from two time points allowed comparison in mutation rate differences among three different genetic lineages of *E. dermatitidis* and evaluated if any changes could suggest adaptations that enabled lineages to survive the lung environment. Average mutation rates within a Clade tended to be similar indicating common diversification events. Statistical comparisons of mutation rates between Clades indicate significance seen between Clade I and Clade III group 1, Clade II and Clade III group 1 and Clade III group 1 and group 2. The main difference here is the greatly increased diversification in Clade III group 1 which is shown with the phylogenetic tree in **Figure 4**, chromosomal aneuploidies seen in **Figure 3**, the very high SNP and INDEL counts seen in **Figure 5**, the increased mutation rates seen in Table 4 and hyper-filamentous phenotypes see in **Figure 8**. We propose the mutations found in *RAD50* may have contributed to the increased diversification seen in this group and future studies will test this hypothesis (Goldman *et al.* 2002; Krogh and Symington 2004).

These clades persisted over a two-year span. Are there physical separations disallowing SNPs in these sub-populations to mix? When observing the functional assessment of variants unique to each clade, we identified alleles in genes that encode for transporters, cytochrome P450 oxidoreductases, iron acquisition, and DNA repair processes. The iron acquisition can be a possible virulence method *E. dermatitidis* could be used to persist in the lung environment as seen in *Aspergillus fumigatus* and *Pseudomonas aeruginosa* (Neilands 1995; Matilla *et al.* 2007; Schrettl *et al.* 2007). Mutations in transporters could be an evolutionary move to adapt to antibiotics or antifungal treatments. The finding of a non-synonymous mutation, *MRS3/4* a mitochondrial iron transporter, may also help point to the pathogenicity and virulence (Murante, 2022, In preparation). Though, the presence of certain SNP variants in iron-binding and siderophore transporters suggests that there could be other means by which *E. dermatitidis* is obtaining the elusive iron molecules from its environment. It would be important to observe the type of variants (e.g. stop codons or a loss of function) produced in these iron-related genes to begin searching for other virulence genes. Further analysis of these variants will point to supporting phenotypic differences allowing for effective colonization of CF lung environments. It's clear that this mutation is not found in the rapidly diverging Clade III and only in the Clade I cluster with very few variants. The isolates we have sequenced only contain one mating type, indicating the low likelihood that even if the isolates have close proximity to each other they are unable to mate. As these isolates were collected from sputum plates, it could be certain isolates have co-localized within different lobes of the lung and could have nutritional or another ecological partitioning that could be driving these clades to diverge and remain

diverged. Our results suggest that there may have been a diversification event that occurred early perhaps during the initial colonization. The clades then stabilized once co-localizing into their respective niches, or there may have been a second 'inoculation' event from a common, stably heterogeneous environmental source at some point in the two-year timeline (Warren *et al.* 2011).

Conclusions

As it has become increasingly clear, collecting a single strain and using it as a metric to assess a single environment and moment of time is inaccurate and a more population approach must be used to best assess microbial infection (Demers *et al.* 2018). Having a closer reference strain to assess variants is also necessary to observe true variants and not a result of years of strain formation. Our results indicate the CF lung environment supports stably diverged populations of clonally derived yeasts.

Data availability

Sequence data generated for isolates with Illumina and Oxford Nanopore technology are deposited in NCBI Sequence Read Archive linked under BioProject PRJNA628510. The assembled genomes of each CF isolate (Ex1-23) are available under accessions listed in **Supplemental Table 2**. The assembled and annotated genome of the in-population reference isolate (DCF04) is available at accession JAJGCF000000000. All analysis pipelines, custom scripts used for data analysis, and raw variant data in the variant call format are available in GitHub repository https://github.com/tania-k/CF_Exophiala_dermatitidis and archived in Zenodo (Kurbessoian 2022).

References

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20(1):224. doi:10.1186/s13059-019-1829-6.

Aznar A, Chen NWG, Rigault M, Riache N, Joseph D, Desmaële D, Mouille G, Boutet S, Soubigou-Taconnat L, Renou J-P, et al. 2014. Scavenging iron: a novel mechanism of plant immunity activation by microbial siderophores. Plant Physiol. 164(4):2167–2183. doi:10.1104/pp.113.233585.

Babič MN, Zupančič J, Gunde-Cimerman N, de Hoog S, Zalar P. 2018. Ecology of the Human Opportunistic Black Yeast *Exophiala dermatitidis* Indicates Preference for Human-Made Habitats. Mycopathologia. 183(1):201–212. doi:10.1007/s11046-017-0134-8.

Banin E, Vasil ML, Peter Greenberg E. 2005. Iron and *Pseudomonas aeruginosa* biofilm formation. Proceedings of the National Academy of Sciences. 102(31):11076–11081. doi:10.1073/pnas.0504266102.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477. doi:10.1089/cmb.2012.0021.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11. doi:10.1186/s13100-015-0041-9.

Barth AL, Pitt TL. 1995. Auxotrophic variants of *Pseudomonas aeruginosa* are selected from prototrophic wild-type strains in respiratory infections in patients with cystic fibrosis. Journal of Clinical Microbiology. 33(1):37–40. doi:10.1128/jcm.33.1.37-40.1995.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49(W1):W29–W35. doi:10.1093/nar/gkab335.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP : eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics and Bioinformatics. 2(2). doi:10.1093/nargab/lqaa026.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60. doi:10.1038/nmeth.3176.

Burns JL, Emerson J, Stapp JR, Yim DL, Krzewinski J, Louden L, Ramsey BW, Clausen CR. 1998. Microbiology of sputum from patients at cystic fibrosis centers in the United States. Clin Infect Dis. 27(1):158–163. doi:10.1086/514631.

Bushnell B. 2014. BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States) Report No.: LBNL-7065E.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 6:e4958. doi:10.7717/peerj.4958.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4:7. doi:10.1186/s13742-015-0047-8.

Chen M, Kondori N, Deng S, van den Ende AHGG, Lackner M, Liao W, de Hoog GS. 2018. Direct detection of *Exophiala* and *Scedosporium* species in sputa of patients with cystic fibrosis. Medical Mycology. 56(6):695–702. doi:10.1093/mmy/myx108.

Chen Z, Martinez DA, Gujja S, Sykes SM, Zeng Q, Szaniszlo PJ, Wang Z, Cuomo CA. 2014. Comparative genomic and transcriptomic analysis of *Wangiella dermatitidis*, a major cause of phaeohyphomycosis and a model black yeast human pathogen. G3. 4(4):561–578. doi:10.1534/g3.113.009241.

Cimon B, Carrère J, Vinatier JF, Chazalette JP, Chabasse D, Bouchara JP. 2000. Clinical Significance of *Scedosporium apiospermum* in Patients with Cystic Fibrosis. European Journal of Clinical Microbiology & Infectious Diseases. 19(1):53–56. doi:10.1007/s100960050011.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6: 80--92. doi:10.4161/fly.19695.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 25(11):1422–1423. doi:10.1093/bioinformatics/btp163.

Conti-Díaz IA, Mackinnon JE, Civila E. 1977. Isolation and identification of black yeasts from the external environment in Uruguay. Pan Amer Health Org Sci Publ. 356:109–114.

Crowley DE, Wang YC, Reid CPP, Szaniszlo PJ. 1991. Mechanisms of iron acquisition from siderophores by microorganisms and plants. In: Chen Y, Hadar Y, editors. Iron Nutrition and Interactions in Plants, 11–17 June 1989, Jerusalem, Israel, 1989. Dordrecht: Springer Netherlands. p. 213–232.

Damveld RA, Arentshorst M, Franken A, vanKuyk PA, Klis FM, van den Hondel CAMJJ, Ram AFJ. 2005. The *Aspergillus niger* MADS-box transcription factor RlmA is required for cell wall reinforcement in response to cell wall stress. Mol Microbiol. 58(1):305–319. doi:10.1111/j.1365-2958.2005.04827.x.

Davis PB. 2006. Cystic Fibrosis Since 1938. American Journal of Respiratory and Critical Care Medicine. 173(5):475–482. doi:10.1164/rccm.200505-840oe.

Defontaine A, Zouhair R, Cimon B, Carrère J, Bailly E, Symoens F, Diouri M, Hallet J-N, Bouchara J-P. 2002. Genotyping Study of *Scedosporium apiospermum* Isolates from Patients with Cystic Fibrosis. Journal of Clinical Microbiology. 40(6):2108–2114. doi:10.1128/jcm.40.6.2108-2114.2002.

Demers EG, Biermann AR, Masonjones S, Crocker AW, Ashare A, Stajich JE, Hogan DA. 2018. Evolution of drug resistance in an antifungal-naive chronic *Candida lusitaniae* infection. Proceedings of the National Academy of Sciences. 115(47):12040–12045. doi:10.1073/pnas.1807698115.

Diemert D, Kunimoto D, Sand C, Rennie R. 2001. Sputum isolation of *Wangiella dermatitidis* in patients with cystic fibrosis. Scand J Infect Dis. 33(10):777–779. doi:10.1080/003655401317074644.

Dixon DM, Shadomy HJ, Shadomy S. 1980. Dematiaceous fungal pathogens isolated from nature. Mycopathologia. 70(3):153–161. doi:10.1007/BF00443026.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Ferec C, Cutting GR. 2012. Assessing the Disease-Liability of Mutations in CFTR. Cold Spring Harb Perspect Med. 2(12):a009480. doi:10.1101/cshperspect.a009480.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42(Database issue):D222–30. doi:10.1093/nar/gkt1223.

Fothergill AW, Rinaldi MG, Sutton DA. 2009. Antifungal susceptibility testing of *Exophiala spp*.: a head-to-head comparison of amphotericin B, itraconazole, posaconazole and voriconazole. Med Mycol. 47(1):41–43. doi:10.1080/13693780802512451.

Frank W, Roester U, Others. 1970. Amphibia as carriers of *Hormiscium (Hormodendrum) dermatitidis* a causative agent of chromo-blastomycosis (chromomycosis) in man. Z Tropenmed Parasitol. 21(1):93–108.

Franke KR, Crowgey EL. 2020. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. Genomics Inform. 18(1):e10. doi:10.5808/GI.2020.18.1.e10.

Freire-Benéitez V, Gourlay S, Berman J, Buscaino A. 2016. Sir2 regulates stability of repetitive domains differentially in the human fungal pathogen *Candida albicans*. Nucleic Acids Res. 44(19):9166–9179. doi:10.1093/nar/gkw594.

Gilchrist CLM, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. 2021. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. Bioinformatics Advances. 1(1):vbab016. doi:10.1093/bioadv/vbab016.

Gilchrist CLM, Chooi Y-H. 2021 Jan 18. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics. doi:10.1093/bioinformatics/btab007.

Goldman GH, McGuire SL, Harris SD. 2002. The DNA damage response in filamentous fungi. Fungal Genet Biol. 35(3):183–195. doi:10.1006/fgbi.2002.1344.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652. doi:10.1038/nbt.1883.

Grahl N, Dolben EL, Filkins LM, Crocker AW, Willger SD, Morrison HG, Sogin ML, Ashare A, Gifford AH, Jacobs NJ, et al. 2018. Profiling of Bacterial and Fungal Microbial Communities in Cystic Fibrosis Sputum Using RNA. mSphere. 3(4). doi:10.1128/mSphere.00292-18.

Grenouillet F, Cimon B, Pana-Katatali H, Person C, Gainet-Brun M, Malinge M-C, Le Govic Y, Richaud-Thiriez B, Bouchara J-P. 2018. *Exophiala dermatitidis* Revealing Cystic Fibrosis in Adult Patients with Chronic Pulmonary Disease. Mycopathologia. 183(1):71–79. doi:10.1007/s11046-017-0218-5.

Griffard EA, Guajardo JR, Cooperstock MS, Scoville CL. 2010. Isolation of *Exophiala dermatitidis* from pigmented sputum in a cystic fibrosis patient. Pediatr Pulmonol. 45(5):508–510. doi:10.1002/ppul.21187.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9(1):R7. doi:10.1186/gb-2008-9-1-r7.

Haas H. 2014. Fungal siderophore metabolism with a focus on *Aspergillus fumigatus*. Nat Prod Rep. 31(10):1266–1276. doi:10.1039/c4np00071d.

Haase G, Skopnik H, Kusenbach G. 1990. *Exophiala dermatitidis* infection in cystic fibrosis. Lancet. 336(8708):188–189. doi:10.1016/0140-6736(90)91721-l.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 35(2):518–522. doi:10.1093/molbev/msx281.

Hoog GS de, Queiroz-Telles F, Haase G, Fernandez-Zeppenfeldt G, Angelis DA, H. G. Gerrits van den Ende A, Matos T, Peltroche-Llacsahuanga H, Pizzirani-Kleiner AA, Rainer J, et al. 2000. Black fungi: clinical and pathogenic approaches. Med Mycol. 38(sup1):243–250. doi:10.1080/mmy.38.s1.243.250.

Horré R, Marklein G, Siekmeier R, Nidermajer S, Reiffert SM. 2009. Selective isolation of *Pseudallescheria* and *Scedosporium* species from respiratory tract specimens of cystic fibrosis patients. Respiration. 77(3):320–324. doi:10.1159/000167419.

Horré R, Schaal KP, Siekmeier R, Sterzik B, de Hoog GS, Schnitzler N. 2004. Isolation of Fungi, Especially *Exophiala dermatitidis*, in Patients Suffering from Cystic Fibrosis. Respiration. 71(4):360–366. doi:10.1159/000079640.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314. doi:10.1093/nar/gky1085.

Institute. "Picard Toolkit." Broad Institute, GitHub Repository. Picard Toolkit.

Johnson L. 2008. Iron and siderophores in fungal–host interactions. Mycol Res. 112(2):170–183. doi:10.1016/j.mycres.2007.11.012.

Jones JT, Liu K-W, Wang X, Kowalski CH, Ross BS, Mills KAM, Kerkaert JD, Hohl TM, Lofgren LA, Stajich JE, et al. 2021. Aspergillus fumigatus Strain-Specific Conidia Lung Persistence Causes an Allergic Broncho-Pulmonary Aspergillosis-Like Disease Phenotype. mSphere. 6(1). doi:10.1128/msphere.01250-20.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Jong CCM de, de Jong CCM, Slabbers L, Engel TGP, Yntema JB, van Westreenen M, Croughs PD, Roeleveld N, Brimicombe R, Verweij PE, et al. 2020. Clinical relevance of *Scedosporium spp.* and *Exophiala dermatitidis* in patients with cystic fibrosis: A nationwide study. Medical Mycology. 58(7):859–866. doi:10.1093/mmy/myaa003.

Kim DH. 2018. Bacterial Siderophores Promote Animal Host Iron Acquisition and Growth. Cell. 175(2):311–312. doi:10.1016/j.cell.2018.09.020.

Kim SH, Clark ST, Surendra A, Copeland JK, Wang PW, Ammar R, Collins C, Tullis DE, Nislow C, Hwang DM, et al. 2015. Global Analysis of the Fungal Microbiome in Cystic Fibrosis Patients Reveals Loss of Function of the Transcriptional Repressor Nrg1 as a Mechanism of Pathogen Adaptation. PLoS Pathog. 11(11):e1005308. doi:10.1371/journal.ppat.1005308.

Kirchhoff L, Olsowski M, Rath P-M, Steinmann J. 2019. *Exophiala dermatitidis*: Key issues of an opportunistic fungal pathogen. Virulence. 10(1):984–998. doi:10.1080/21505594.2019.1596504.

Klasinc R, Riesenhuber M, Bacher A, Willinger B. 2019. Invasive Fungal Infection Caused by *Exophiala dermatitidis* in a Patient After Lung Transplantation: Case Report and Literature Review. Mycopathologia. 184(1):107–113. doi:10.1007/s11046-018-0275-4.

Kondori N, Gilljam M, Lindblad A, Jönsson B, Moore ERB, Wennerås C. 2011. High rate of *Exophiala dermatitidis* recovery in the airways of patients with cystic fibrosis is associated with pancreatic insufficiency. J Clin Microbiol. 49(3):1004–1009. doi:10.1128/JCM.01899-10.

Kondori N, Lindblad A, Welinder-Olsson C, Wennerås C, Gilljam M. 2014. Development of IgG antibodies to *Exophiala dermatitidis* is associated with inflammatory responses in patients with cystic fibrosis. J Cyst Fibros. 13(4):391–399. doi:10.1016/j.jcf.2013.12.007.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics. 5:59. doi:10.1186/1471-2105-5-59.

Krogh BO, Symington LS. 2004. Recombination proteins in yeast. Annu Rev Genet. 38:233–271. doi:10.1146/annurev.genet.38.072902.091500.

Kurbessoian T. 2022. tania-k/CF_Exophiala_dermatitidis: Version 6 update.

Kusenbach G, Skopnik H, Haase G, Friedrichs F, Döhmen H. 1992. *Exophiala dermatitidis* pneumonia in cystic fibrosis. Eur J Pediatr. 151(5):344–346. doi:10.1007/BF02113255.

Lavrin T, Konte T, Kostanjšek R, Sitar S, Sepčič K, Prpar Mihevc S, Žagar E, Župunski V, Lenassi M, Rogelj B, et al. 2020. The Neurotropic Black Yeast *Exophiala dermatitidis* Induces Neurocytotoxicity in Neuroblastoma Cells and Progressive Cell Death. Cells. 9(4). doi:10.3390/cells9040963.

Legrand M, Jaitly P, Feri A, d'Enfert C, Sanyal K. 2019. *Candida albicans*: An Emerging Yeast Model to Study Eukaryotic Genome Plasticity. Trends Genet. 35(4):292–307. doi:10.1016/j.tig.2019.01.005.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(W1):W242–5. doi:10.1093/nar/gkw290.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 26(5):589–595. doi:10.1093/bioinformatics/btp698.

Li H, Handsaker B, Danecek P, McCarthy S, Marshall J. 2019. BCFtools.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44(W1):W54–7. doi:10.1093/nar/gkw413.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 20(16):2878–2879. doi:10.1093/bioinformatics/bth315.

Malo ME, Schultzhaus Z, Frank C, Romsdahl J, Wang Z, Dadachova E. 2021. Transcriptomic and genomic changes associated with radio-adaptation in *Exophiala dermatitidis*. Comput Struct Biotechnol J. 19:196–205. doi:10.1016/j.csbj.2020.12.013.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 38(10):4647–4654. doi:10.1093/molbev/msab199.

Mariani-Kurkdjian P, Bingen E. 2003. Bactéries pathogènes dans la mucoviscidose. Archives de Pédiatrie. 10:S342–S346. doi:10.1016/S0929-693X(03)90050-9.

Matilla MA, Ramos JL, Duque E, de Dios Alché J, Espinosa-Urgel M, Ramos-González MI. 2007. Temperature and pyoverdine-mediated iron acquisition control surface motility of *Pseudomonas putida*. Environmental Microbiology. 9(7):1842–1850. doi:10.1111/j.1462-2920.2007.01286.x.

Matos T, de Hoog GS, de Boer AG, de Crom I, Haase G. 2002. High prevalence of the neurotrope *Exophiala dermatitidis* and related oligotrophic black yeasts in sauna facilities. Mycoses. 45(9-10):373–377. doi:10.1046/j.1439-0507.2002.00779.x.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303. doi:10.1101/gr.107524.110.

Metin B, Döğen A, Yıldırım E, de Hoog GS, Heitman J, Ilkit M. 2019. Mating type (MAT) locus and possible sexuality of the opportunistic pathogen *Exophiala dermatitidis*. Fungal Genetics and Biology. 124:29–38. doi:10.1016/j.fgb.2018.12.011.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 37(5):1530–1534. doi:10.1093/molbev/msaa015.

Mochochoko BM, Ezeokoli OT, Sebolai O, Albertyn J, Pohl CH. 2021. Role of the high-affinity reductive iron acquisition pathway of *Candida albicans* in prostaglandin E2 production, virulence, and interaction with *Pseudomonas aeruginosa*. Medical Mycology. 59(9):869–881. doi:10.1093/mmy/myab015.

Mossialos D, Amoutzias GD. 2009. Role of siderophores in cystic fibrosis pathogenesis: foes or friends? Int J Med Microbiol. 299(2):87–98. doi:10.1016/j.ijmm.2008.06.008.

Mukai Y, Nureki S-I, Hata M, Shigenaga T, Tokimatsu I, Miyazaki E, Kadota J-I, Yarita K, Kamei K. 2014. *Exophiala dermatitidis* pneumonia successfully treated with long-term itraconazole therapy. J Infect Chemother. 20(7):446–449. doi:10.1016/j.jiac.2014.02.006.

Nagano Y, Cherie Millar B, Johnson E, Goldsmith CE, Stuart Elborn J, Rendall J, Moore JE. 2007. Fungal infections in patients with cystic fibrosis. Reviews in Medical Microbiology. 18(1):11–16. doi:10.1097/mrm.0b013e3282e1c70a.

Neilands JB. 1995. Siderophores: structure and function of microbial iron transport compounds. J Biol Chem. 270(45):26723–26726. doi:10.1074/jbc.270.45.26723.

Nishimura K, Miyaji M. 1982. Studies on a saprophyte of *Exophiala dermatitidis* isolated from a humidifier. Mycopathologia. 77(3):173–181. doi:10.1007/bf00518803.

Nosanchuk JD, Casadevall A. 2003. The contribution of melanin to microbial pathogenesis. Cell Microbiol. 5(4):203–223. doi:10.1046/j.1462-5814.2003.00268.x.

Nosanchuk JD, Casadevall A. 2006. Impact of Melanin on Microbial Virulence and Clinical Resistance to Antimicrobial Compounds. Antimicrobial Agents and Chemotherapy. 50(11):3519–3528. doi:10.1128/aac.00545-06.

Packeu A, Lebecque P, Rodriguez-Villalobos H, Boeras A, Hendrickx M, Bouchara J-P, Symoens F. 2012. Molecular typing and antifungal susceptibility of *Exophiala* isolates from patients with cystic fibrosis. J Med Microbiol. 61(Pt 9):1226–1233. doi:10.1099/jmm.0.042317-0.

Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation.

Palmer JM, Stajich JE. 2022. Automatic assembly for the fungi (AAFTF): genome assembly pipeline.

Paolo WF Jr, Dadachova E, Mandal P, Casadevall A, Szaniszlo PJ, Nosanchuk JD. 2006. Effects of disrupting the polyketide synthase gene WdPKS1 in *Wangiella [Exophiala] dermatitidis* on melanin production and resistance to killing by antifungal compounds, enzymatic degradation, and extremes in temperature. BMC Microbiol. 6:55. doi:10.1186/1471-2180-6-55.

Parize P, Billaud S, Bienvenu AL, Bourdy S, le Pogam MA, Reix P, Picot S, Robert R, Lortholary O, Bouchara J-P, et al. 2014. Impact of *Scedosporium apiospermum* complex seroprevalence in patients with cystic fibrosis. Journal of Cystic Fibrosis. 13(6):667–673. doi:10.1016/j.jcf.2014.01.011.

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 34(5):867–868. doi:10.1093/bioinformatics/btx699.

Pihet M, Carrere J, Cimon B, Chabasse D, Delhaes L, Symoens F, Bouchara J-P, Pihet M, Carrere J, Cimon B, et al. 2009. Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis – a review. Medical Mycology. 47(4):387–397. doi:10.1080/13693780802609604.

Pohl Carolina H., Noverr Mairi C. Competition for Iron during Polymicrobial Infections May Increase Antifungal Drug Susceptibility—How Will It Impact Treatment Options? Infect Immun. 0(0):e00057–22. doi:10.1128/iai.00057-22.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204. doi:10.1093/nar/gky448.

Qu X, Yu B, Liu J, Zhang X, Li G, Zhang D, Li L, Wang X, Wang L, Chen J, et al. 2014. MADS-box transcription factor SsMADS is involved in regulating growth and virulence in *Sclerotinia sclerotiorum*. Int J Mol Sci. 15(5):8049–8062. doi:10.3390/ijms15058049.

Rath P-M, Muller K-D, Dermoumi H, Ansorg R. 1997. A comparison of methods of phenotypic and genotypic fingerprinting of *Exophiala dermatitidis* isolated from sputum samples of patients

with cystic fibrosis. Journal of Medical Microbiology. 46(9):757–762. doi:10.1099/00222615-46-9-757.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46(D1):D624–D632. doi:10.1093/nar/gkx1134.

Reiss NR, Mok WY. 1979. *Wangiella dermatitidis* isolated from bats in Manaus Brazil. Sabouraudia. 17(3):213–218.

Revankar SG, Patterson JE, Sutton DA, Pullen R, Rinaldi MG. 2002. Disseminated phaeohyphomycosis: review of an emerging mycosis. Clin Infect Dis. 34(4):467–476. doi:10.1086/338636.

Revankar SG, Sutton DA. 2010. Melanized fungi in human disease. Clin Microbiol Rev. 23(4):884–928. doi:10.1128/CMR.00019-10.

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 245(4922):1066–1073. doi:10.1126/science.2475911.

Robertson KL, Mostaghim A, Cuomo CA, Soto CM, Lebedev N, Bailey RF, Wang Z. 2012. Adaptation of the black yeast *Wangiella dermatitidis* to ionizing radiation: molecular and cellular mechanisms. PLoS One. 7(11):e48674. doi:10.1371/journal.pone.0048674.

Ross BS, Lofgren LA, Ashare A, Stajich JE, Cramer RA. 2021. *Aspergillus fumigatus* In-Host HOG Pathway Mutation for Cystic Fibrosis Lung Microenvironment Persistence. MBio. 12(4):e0215321. doi:10.1128/mBio.02153-21.

Sass G, Ansari SR, Dietl A-M, Déziel E, Haas H, Stevens DA. 2019. Intermicrobial interaction: *Aspergillus fumigatus* siderophores protect against competition by *Pseudomonas aeruginosa*. PLoS One. 14(5):e0216085. doi:10.1371/journal.pone.0216085.

Saunte DM, Tarazooie B, Arendrup MC, de Hoog GS. 2012. Black yeast-like fungi in skin and nail: it probably matters. Mycoses. 55(2):161–167. doi:10.1111/j.1439-0507.2011.02055.x.

Schultzhaus Z, Romsdahl J, Chen A, Tschirhart T, Kim S, Leary D, Wang Z. 2020. The response of the melanized yeast *Exophiala dermatitidis* to gamma radiation exposure. Environmental Microbiology. 22(4):1310–1326. doi:10.1111/1462-2920.14936.

Schrettl M, Bignell E, Kragl C, Sabiha Y, Loss O, Eisendle M, Wallner A, Arst HN, Haynes K, Haas H. 2007. Distinct Roles for Intra- and Extracellular Siderophores during *Aspergillus fumigatus* Infection. PLoS Pathogens. 3(9):e128. doi:10.1371/journal.ppat.0030128.

Selmecki A, Forche A, Berman J. 2010. Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot Cell. 9(7):991–1008. doi:10.1128/EC.00060-10.

Seyedmousavi S, Netea MG, Mouton JW, Melchers WJG, Verweij PE, de Hoog GS. 2014. Black yeasts and their filamentous relatives: principles of pathogenesis and host defense. Clin Microbiol Rev. 27(3):527–542. doi:10.1128/CMR.00093-13.

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. Eur J Biochem. 229(1):1–13. doi:10.1111/j.1432-1033.1995.0001l.x.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:31. doi:10.1186/1471-2105-6-31.

Smit A. F. A. 2004. Repeat-Masker Open-3.0.

Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0.

Smith DFQ, Casadevall A. 2019. The Role of Melanin in Fungal Pathogenesis for Animal Hosts. Curr Top Microbiol Immunol. 422:1–30. doi:10.1007/82_2019_173.

Sofi MY, Shafi A, Masoodi KZ. 2022. NCBI BLAST. Bioinformatics for Everyone.:95–102. doi:10.1016/b978-0-323-91128-3.00021-5.

Song Y, de Sande WWJL, Moreno LF, van den Ende BG, Li R, de Hoog S. 2017. Comparative Ecology of Capsular Exophiala Species Causing Disseminated Infection in Humans. Frontiers in Microbiology. 8. doi:10.3389/fmicb.2017.02514.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34(Web Server issue):W435–9. doi:10.1093/nar/gkl200.

Steinkamp G, Wiedemann B, Rietschel E, Krahl A, Gielen J, Bärmeier H, Ratjen F. 2005. Prospective evaluation of emerging bacteria in cystic fibrosis. Journal of Cystic Fibrosis. 4(1):41–48. doi:10.1016/j.jcf.2004.10.002.

Sudfeld CR, Dasenbrook EC, Merz WG, Carroll KC, Boyle MP. 2010. Prevalence and risk factors for recovery of filamentous fungi in individuals with cystic fibrosis. Journal of Cystic Fibrosis. 9(2):110–116. doi:10.1016/j.jcf.2009.11.010.

Sudhadham M, Gerrits van den Ende AHG, Sihanonth P, Sivichai S, Chaiyarat R, Menken SBJ, van Belkum A, de Hoog GS. 2011. Elucidation of distribution patterns and possible infection routes of the neurotropic black yeast *Exophiala dermatitidis* using AFLP. Fungal Biol. 115(10):1051–1065. doi:10.1016/j.funbio.2010.07.004.

Sudhadham M, Prakitsin S, Sivichai S, Chaiyarat R, Dorrestein GM, Menken SBJ, de Hoog GS. 2008. The neurotropic black yeast *Exophiala dermatitidi*s has a possible origin in the tropical rainforest. Studies in Mycology. 61:145–155. doi:10.3114/sim.2008.61.15.

Teixeira MM, Moreno LF, Stielow BJ, Muszewska A, Hainaut M, Gonzaga L, Abouelleil A, Patané JSL, Priest M, Souza R, et al. 2017. Exploring the genomic diversity of black yeasts and relatives (Chaetothyriales, Ascomycota). Stud Mycol. 86:1–28. doi:10.1016/j.simyco.2017.01.001.

Thomas SR, Ray A, Hodson ME, Pitt TL. 2000. Increased sputum amino acid concentrations and auxotrophy of *Pseudomonas aeruginosa* in severe cystic fibrosis lung disease. Thorax. 55(9):795–797. doi:10.1136/thorax.55.9.795.

Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, et al. 2008. Detection of Anaerobic Bacteria in High Numbers in Sputum from Patients with Cystic Fibrosis. American Journal of Respiratory and Critical Care Medicine. 177(9):995–1001. doi:10.1164/rccm.200708-1151oc.

Turcios NL. 2020. Cystic Fibrosis Lung Disease: An Overview. Respiratory Care. 65(2):233–251. doi:10.4187/respcare.06697.

Tyrrell J, Callaghan M. 2016. Iron acquisition in the cystic fibrosis lung and potential for novel therapeutic strategies. Microbiology. 162(2):191–205. doi:10.1099/mic.0.000220.

Uijthof JMJ, Hoog GS, Cock AWA, Takeo K, Nishimura K. 2009. Pathogenicity of strains of the black yeast *Exophiala (Wangiella) dermatitidis*: an evaluation based on polymerase chain reaction. Mycoses. 37(7-8):235–242. doi:10.1111/j.1439-0507.1994.tb00419.x.

Warren AE, Boulianne-Larsen CM, Chandler CB, Chiotti K, Kroll E, Miller SR, Taddei F, Sermet-Gaudelus I, Ferroni A, McInnerney K, et al. 2011. Genotypic and phenotypic variation in *Pseudomonas aeruginosa* reveals signatures of secondary infection and mutator activity in certain cystic fibrosis patients with chronic lung infections. Infect Immun. 79(12):4802–4818. doi:10.1128/IAI.05282-11.

Weisman CM, Murray AW, Eddy SR. 2022. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. Curr Biol. 32(12):2632–2639.e2. doi:10.1016/j.cub.2022.04.085.

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 20(1):129. doi:10.1186/s13059-019-1727-y.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

Xiong D, Wang Y, Tian L, Tian C. 2016. MADS-Box Transcription Factor VdMcm1 Regulates Conidiation, Microsclerotia Formation, Pathogenicity, and Secondary Metabolism of Verticillium dahliae. Front Microbiol. 7:1192. doi:10.3389/fmicb.2016.01192.

Yan K, Yin H, Wang J, Cai Y. 2022. Subtle relationships between *Pseudomonas aeruginosa* and fungi in patients with cystic fibrosis. Acta Clinica Belgica. 77(2):425–435. doi:10.1080/17843286.2020.1852850.

Zajc J, Gostinčar C, Černoša A, Gunde-Cimerman N. 2019. Stress-Tolerant Yeasts: Opportunistic Pathogenicity Versus Biocontrol Potential. Genes. 10(1):42. doi:10.3390/genes10010042.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46(W1):W95–W101. doi:10.1093/nar/gky418.

Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, et al. 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proceedings of the National Academy of Sciences. 109(15):5809–5814. doi:10.1073/pnas.1120577109.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics. 29(21):2669–2677. doi:10.1093/bioinformatics/btt476.

Zupančič J, Babič MN, Zalar P, Gunde-Cimerman N. 2016. The Black Yeast *Exophiala dermatitidis* and Other Selected Opportunistic Human Fungal Pathogens Spread from Dishwashers to Kitchens. PLOS ONE. 11(2):e0148166. doi:10.1371/journal.pone.0148166.

Conclusion

The great advancements in science are what allow us to take a finer approach to understand the biological world around us. Mycology is still a budding field and has so much more potential to discover, which makes this a really exciting moment to be in. As sequencing technology becomes more affordable (Preston et al. 2020) and with optimization for long-read technology (De Coster et al. 2021), it has become easier to achieve complete genome assemblies for further study. As there is an increase in consideration for fungal associates in clinical and environmental circumstances, black yeasts are a vital member to be brought into the conversation.

Choosing the term micro-colonial fungi (MCF) for the entire dissertation while introducing all the other nomenclature it is known for was only done to bring the focus to their timely length of enigmatic visibility but to look at them as a group. MCFs have been studied for a very long time, each person taking the time to describe them in such a way, that when the next person identifies them they too have to add their own nomenclature. In the end, we see these microorganisms that appear to be everywhere on this planet down to almost every extreme we can think of. And yet we can also find them in the most common spaces, living with us, and sometimes living within us.

Chapter 1 described MCFs from the deserts of southern California, and into these interesting complex microbial features called biological soil crusts (BSC). Using previously described culture-dependent methods (Kurbessoian 2019) we were able to isolate multiple different types of MCF. Three identified species, when sequenced and

compared to other known MCF, were different enough and required a naming schema. Observing the native American traditional lands and culture, we reached out to the reigning tribes to consult on nomenclature. Thankfully one of the two groups responded and with a linguist, we were able to devise a new species name, and a new genus and species name that followed Cahuilla verbiage. The word tulheliweneti, meaning black-spread-be-noun, would be the term Cahuilla tribe would use to describe how a BSC would appear to them. Thus, the name *Coniosporium tulhelowenetii* was devised to describe our MCF. As Cahuilla do not have words for microorganisms, we looked into mythology to help describe our second species which needed a new genus and species name. In Cahuilla traditions, Taxaw tesnikish is the deity that resided in Boyd Deep valley. Thus, the name of our second organism is *Taxawa* (body) *tesnikishii* (yellow). Our third MCF was collected near Joshua tree in Sheephole Valley Wilderness on BLM land, unfortunately, we did not receive a similar positive outcome from the Serrano tribe folk, and decided to name our third *Neophaeococcomyces mojaviensis*, after the Mojave Desert it was isolated from. Sequencing availability allowed us to produce high-quality genome assemblies due to Nanopore technology. The pipelines developed in the Stajich lab gave me the confidence in assembly and annotation, and to finally submit them to the scientific community.

In chapter 2 I assessed *Friedmanniomyces*, an endemic MCF isolated from the Antarctica's McMurdo Dry Valley region. A hypothesis was to consider the endemic nature of *Friedmanniomyces* species and see if we can place on a molecular clock the divergence of this species from other Dothideomycetes. Using fungal fossil records and molecular clock calculations the time of divergence was detected to be around 50 million years ago. When

observing the time Antarctica, as a continent, drifted away from Pangea and into the South Pole to be around 60 million years ago (Fitzgerald 2002) has us considering this as a viable hypothesis. Another question was to observe the ploidy of these species. Generally, MCF have been described to be asexual and haploid. An example of an MCF that voided the haploid description was *Hortaea werneckii* with its two haploid hybridization events creating a diploid-like ploidy (Gostinčar et al. 2018; Gostinčar et al. 2021). Twenty-five strains of *Friedmanniomyces endolithicus* were recently sequenced through the help of our Italian collaborators Dr. Laura Selbmann and Dr. Claudia Coleine and were the main species to consider in our work. Traditionally, the gold standard to observe ploidy is to use flow cytometry, but with the advent of whole genome sequencing, there are new tools we can use to determine ploidy computationally. We had confirmation through use of multiple tools that strains expressed diploidy and triploidy, but there is some more work to be done to confirm if two of the four strains have tetraploidy instead of haploidy. We also used this method to generate biallelic SNPs along with the metadata collected of the 25 strains to see if there could be a correlation between the ploidy and another measure. The year each sample was collected seemed to have a higher correlation with the ploidy states, unlike the elevation, which we initially believed was more of a differentiating factor. The fact that there are microorganisms surviving extreme tundra conditions, like the *Friedmanniomyces*, is a marvel and it is interesting to consider the conditions where these species would increase or decrease their ploidy to increase their reproductive chances.

In Chapter 3 I observed persistent MCFs, *Exophiala dermatitidis*, isolated from a cystic fibrosis patient over the span of three years. The slow-growing nature of MCF could be

the reason why initial identification or consideration of this species was not met. As each species was collected and observed, there was a drastic heterogeneity in melanin production and this propelled us to take a genomic and population approach to observe for any variant changes that could also be aiding in persistence. Melanin production is a virulence factor many microorganisms take advantage of to persist and further infect their hosts (Eisenman and Casadevall 2012). By building our population structure tree, we were not able to pinpoint specific lineages with more melanization and correlate it to the variants we were detecting, in the mutation rates we calculated, or in the aberrant chromosomal structures we observed. Instead, certain variants for siderophores, iron-related genes, and increased filamentation appeared, along with mutation rates of a particular hyper-variant set of isolates, and had us consider the functional means MCF had to endure to persist in another extreme environment, the human body. The mating type was also an important consideration (Metin *et al.* 2019). Prior consideration of *E. dermatitidis* pointed to the species being haploid and clonal. Our work confirmed the ploidy and clonal nature of these strains, suggesting there was no recombination occurring within the lungs of this patient. This work should have clinicians take pause and consider leaving microbial colony plates longer than the expected 48 hours to observe for slow-growing clinical case fungi such as *Exophiala dermatitidis*.

References:

De Coster W, Weissensteiner MH, Sedlazeck FJ. 2021. Towards population-scale long-read sequencing. Nat Rev Genet. 22(9):572–587.

Eisenman HC, Casadevall A. 2012. Synthesis and assembly of fungal melanin. Appl Microbiol Biotechnol. 93(3):931–940.

Fitzgerald P. 2002. Tectonics and landscape evolution of the Antarctic plate since the breakup of Gondwana, with an emphasis on the West Antarctic Rift System and the …. Royal Society of New Zealand Bulletin.

Gostinčar C, Stajich JE, Kejžar A, Sinha S, Nislow C, Lenassi M, Gunde-Cimerman N. 2021. Seven Years at High Salinity—Experimental Evolution of the Extremely Halotolerant Black Yeast Hortaea werneckii. Journal of Fungi. 7(9):723.

Gostinčar C, Stajich JE, Zupančič J, Zalar P, Gunde-Cimerman N. 2018. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. BMC Genomics. 19(1):364.

Metin B, Döğen A, Yıldırım E, de Hoog GS, Heitman J, Ilkit M. 2019. Mating type (MAT) locus and possible sexuality of the opportunistic pathogen Exophiala dermatitidis. Fungal Genetics and Biology. 124:29–38. doi:10.1016/j.fgb.2018.12.011.

Kurbessoian T. 2019. Cultivating melanized fungi from biological soil crust and rock surfaces v1. doi:10.17504/protocols.io.3fxgjpn.

Preston, VanZeeland, Peiffer. Innovation at Illumina: The road to the $600 human genome. Nat Portf.

Appendix A. Cultivating Melanized Fungi from Biological Soil Crust and Rock Surfaces V.1

dx.doi.org/10.17504/protocols.io.3fxgjpn

Tania Kurbessoian; Renata Haro; Jason E. Stajich

ABSTRACT

As the interest to understand melanized fungi becomes more of a focus due to pathological diseases, there needs to be a clearer method to isolate and identify the fungi from their surroundings. Culture independent analysis has helped bloom scientist understandings of the vast quantities of microorganisms around us, but culturing and identifying has always been a struggle. We must look into utilizing different microbiological techniques to help better understand, isolate, and to apply Koch's postulates to prove they are the cause of said disease. Growing and isolating fungi has always been an issue especially if the fungi in question is a slow-growing fungus, where fast growing fungi or other microorganisms can grow and surpass the field of view and compete with the slow growing fungi. A series of dilutions, antibiotics and oligotrophic media can all counteract the issues to provide you with a clear window to help isolate your fungi in question.

There is still a need to verify that living strains of organisms are actually present in the environment. One approach is to use "culture dependent" method to obtain strains of organisms present in the biocrust. This is achieved by plating biocrust soils onto microbiological media in order to further isolate life strains. Serial dilution is used to isolate the fungi from the soil. This method involves making a soil slurry by diluting the

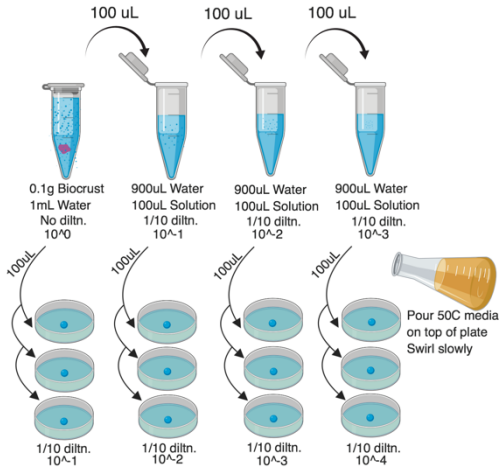
soil sample with increasing higher proportion of water or media in order to get a low starting concentration of starting spores or material growing on the Petri dishes.

Materials

Prepare Glucose Asparagine Agar (GAA)/1L
0.500g Dibasic Potassium Phosphate
0.500g Asparagine
10.00g Glucose
15.00g Agar

1. Autoclave

2. Prepare filtered antibiotics for 1-liter agar solution, either Gentamycin, Tetracycline, Kanamycin, Streptomycin or Chloramphenicol will work.



3. Now place 0.1 g of sample (either biocrust, soil or rock surface can be used) placed in microcentrifuge tube.
    a. Add 1000 µL Sterile H2O into the same tube.
    b. Label this 10^0.

4. Vortex (on highest setting) sample for 10 minutes until solution becomes a slurry.

Preparing the Serial Dilution

5. Prepare 3 other microcentrifuge tubes with 900 µL sterile H2O.
    a. Label them 10^-1, 10^-2, 10^-3.
    b. Take100 µL slurry solution from 10^0 and add it to 10^-1. Vortex 10^-1.

c. Then take 100 µL 10^-1 and add it to 10^-2. Vortex 10^-2.
d. Then take 100 µL 10^-2 and add it to 10^-3. Vortex 10^-3.

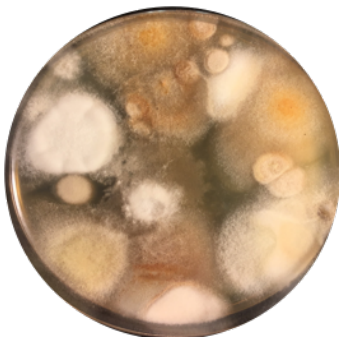
6. Set up 12 empty Petri Plates and label them with appropriate labels. I.e. Date, Media (list antibiotic used), Crust type, Dilution, Iteration (Plate 1, Plate 2 or Plate 3).
    a. Pipette out 100 µL 10^0 into first three plates that were properly labeled for 10^0.
    b. Pipette out 100 µL 10^-1 into next three plates labeled for 10^-1.
    c. Pipette out 100 µL 10^-2 into next three plates labeled for 10^-2.
    d. Pipette out 100 µL 10^-3 into next three plates labeled for 10^-3.

7. Make sure to have media cooled (to 50 °C *) and ready to go at this point.
*which means you can touch it with your hands without burning yourself.
    a. Add the media slowly to avoid bubbles. Fill until agar closes all gaps on plate. Fill up to three plates at a time.
    b. When three plates have been filled start to slowly mix the plate and slurry solution. Do a very careful "Figure 8" motion. Make sure to mix slowly and evenly without touching the lid of the plate.
    c. If it does touch, just replace with a new lid from an unused empty plate.
    d. Continue repeating 5.1 and 5.2 until all plates have been filled and are cooling.

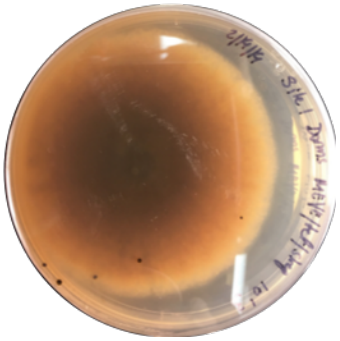
8. Let plates air dry for 24 hours, then parafilm to allow for long-term storage.

9. New fungi should start popping up on undiluted and next dilution plates in the next couple of days. Most dilute plates will also have sparser filamentous fungi growing.
Melanized fungi will take about 1-2 weeks to show up on plates.

10. Make sure to check on the underside of plates for black dots. Those are the fungi you are looking for!



This is an example of the least diluted plate fungal growth.



This is the back of a plate. Notice the little black dots. These are the melanized fungi/yeasts to collect.

Appendix B. Low Biomass, high contamination Illumina DNA prep using DNeasy PowerSoil (Pro) Kit

dx.doi.org/10.17504/protocols.io.6qpvr4ey3gmk/v1

Tania Kurbessoian[1], Jason E Stajich[1], Sonia L. Ghose[2]
[1]UC Riverside
[2]UC Davis

ABSTRACT
This is an addendum to the already optimized DNeasy PowerSoil Kit. This can also be applied to the PowerSoil Pro Kit. Use this protocol for Low Biomass or high contamination samples.

This protocol was used to clear out chelators in a hydrocarbon contaminated soil sample. Basal levels of DNA were collected and used successfully in Illumina sequencing.

BEFORE STARTING

1. Add sample into the PowerBead tube.

2. Add 60µL of Solution C1.
    a. Make sure the sample has not precipitated. If so, heat sample to 60ºC until precipitate is dissolved into solution.

3. Vortex to mix and incubate at 65ºC for 10 minutes.

4. Bead beat at "homogenize" setting: 90 seconds bead beating, 60 seconds rest, then another 90 seconds (3 minutes total of actual beating).

5. Centrifuge at 10,000 x g for 30 s.

6. Remove the supernatant or upper aqueous layer if you have one, to the new tube.

7. Add 100 uL Solution C2 and 100µL of Solution C3. Vortex to mix. Incubate at 4ºC or on ice for 5 minutes.
    a. *This is for low humic soils (100µL of each). If high humic soil, add 150µL C2 followed by 150µL C3.

8. Centrifuge to pellet (1 min at 10,000 x g.) and transfer the supernatant to a new tube. Avoid pellet.

9. Ideally you will have 650µL of lysate, and can add 650µL of Solution C4 and 650µL of 100% ethanol.
    a. Make sure you mix C4 solution well. If new kit add ethanol.
    b. If you have 700µL, add 700µL of Solution C4 and 600µL of 100% ethanol. (Keep it as close to 1:1 as possible)

10. Load the lysate to the spin filter 650µL at a time and bind in three steps, alternating with centrifugation (1 min at 10,000 x g). Discard flow-through each time.

11. If the membrane is not stained brown, wash with 650µL of 100% ethanol (centrifuge 1 min at 10,000 x g and discard flow-through).
    a. If the membrane is stained brown, per sample, prepare a mix of 300µL of Solution C4 and 370µL of 100% ethanol. Wash the column with this mixture first. Follow this wash with the 100% ethanol wash.

12. Wash with 500µL of solution C5 (centrifuge 1 min at 10,000 x g and discard flow-through)

13. Dry the spin column for 2 minutes 10,000 x g. Transfer to a clean tube.

14. Elute in 60µL of buffer C6 (Heat this buffer to 60ºC before eluting). Let the buffer sit on the membrane 5 minutes before elution.

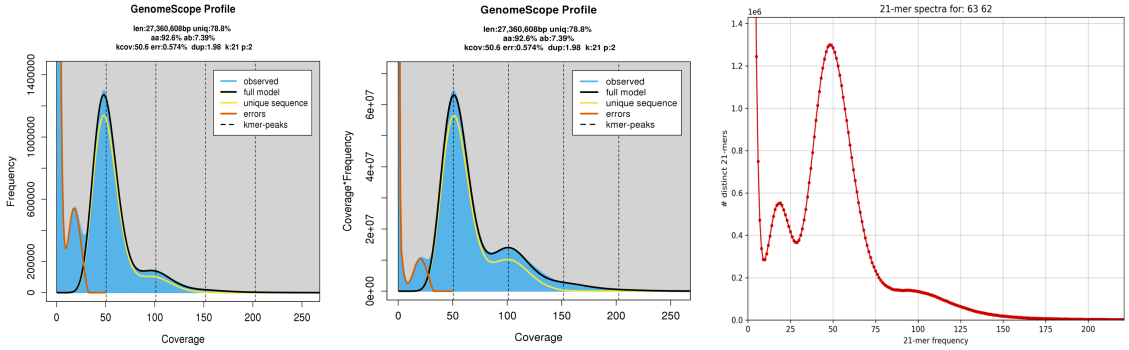
15. Then centrifuge for 1 min 10,000 x g into your storage tube.

Appendix C. K-mer frequency data on *Friedmanniomyces endolithicus*.

CCFEE_524

GenomeScope version 2.0

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

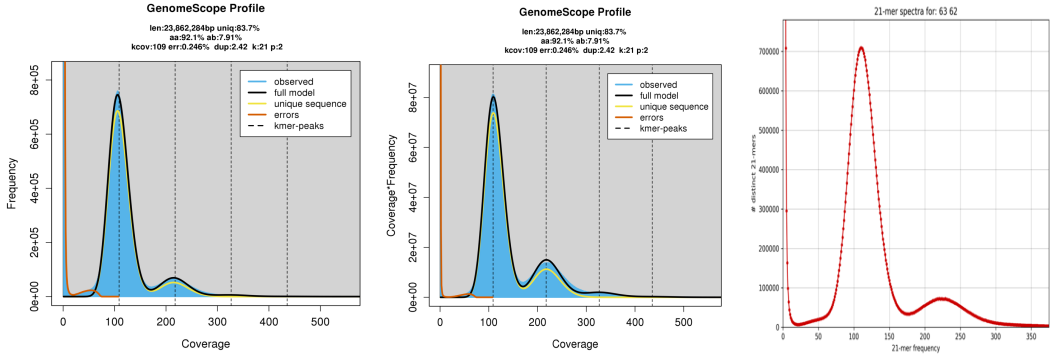


p = 2

property	min	max
Homozygous (aa)	92.4862%	92.7248%
Heterozygous (ab)	7.27525%	7.51379%
Genome Haploid Length	27,303,882 bp	27,360,608 bp
Genome Repeat Length	5,794,733 bp	5,806,772 bp
Genome Unique Length	21,509,149 bp	21,553,836 bp
Model Fit	93.5158%	96.4899%
Read Error Rate	0.573587%	0.573587%

CCFEE_670

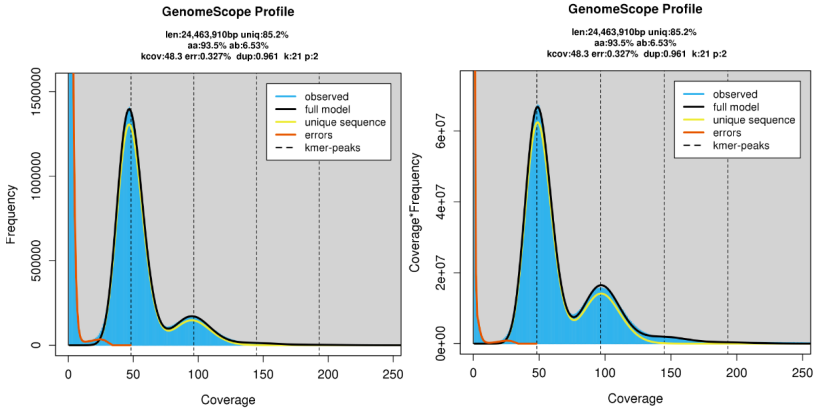
Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram



p = 2

property	min	max
Homozygous (aa)	92.0089%	92.1727%
Heterozygous (ab)	7.82734%	7.9911%
Genome Haploid Length	23,842,450 bp	23,862,284 bp
Genome Repeat Length	3,889,606 bp	3,892,842 bp
Genome Unique Length	19,952,844 bp	19,969,442 bp
Model Fit	94.3233%	94.966%
Read Error Rate	0.246369%	0.246369%

CCFEE_690

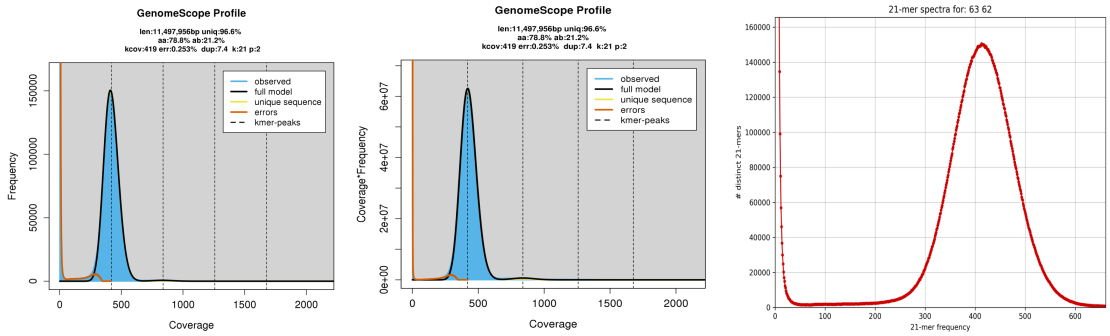


p = 2

property	min	max
Homozygous (aa)	93.4501%	93.4909%
Heterozygous (ab)	6.50912%	6.54989%
Genome Haploid Length	24,449,431 bp	24,463,910 bp
Genome Repeat Length	3,606,803 bp	3,608,939 bp
Genome Unique Length	20,842,627 bp	20,854,970 bp
Model Fit	94.7323%	97.1429%
Read Error Rate	0.326953%	0.326953%

CCFEE_5001

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

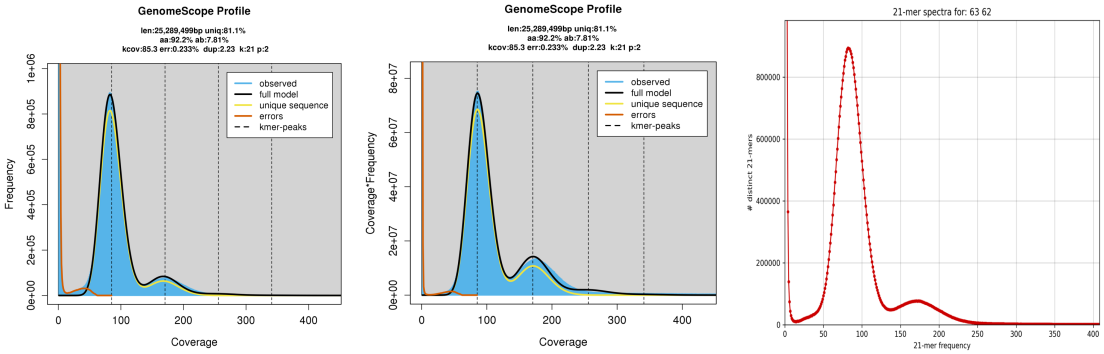


p = 2

property	min	max
Homozygous (aa)	0%	100%
Heterozygous (ab)	0%	100%
Genome Haploid Length	11,491,882 bp	11,497,956 bp
Genome Repeat Length	389,007 bp	389,212 bp
Genome Unique Length	11,102,875 bp	11,108,743 bp
Model Fit	94.9056%	96.2807%
Read Error Rate	0.252778%	0.252778%

CCFEE_5193

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

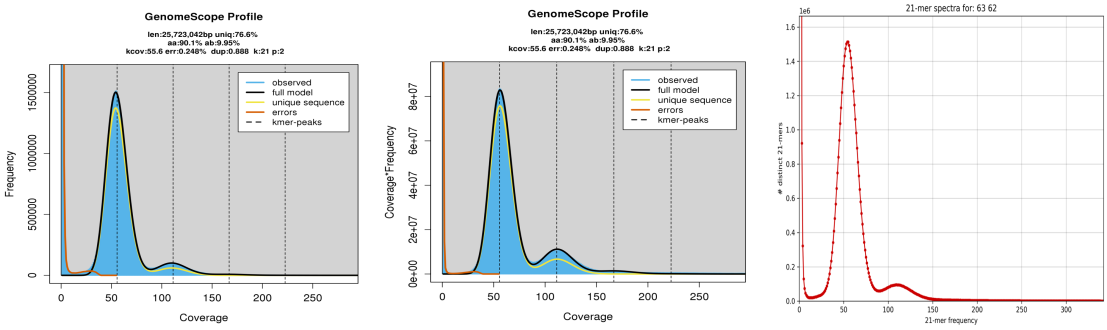


p = 2

property	min	max
Homozygous (aa)	92.1082%	92.2812%
Heterozygous (ab)	7.71876%	7.89183%
Genome Haploid Length	25,263,267 bp	25,289,499 bp
Genome Repeat Length	4,784,156 bp	4,789,123 bp
Genome Unique Length	20,479,112 bp	20,500,376 bp
Model Fit	90.9235%	94.4957%
Read Error Rate	0.23337%	0.23337%

CCFEE_5195

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

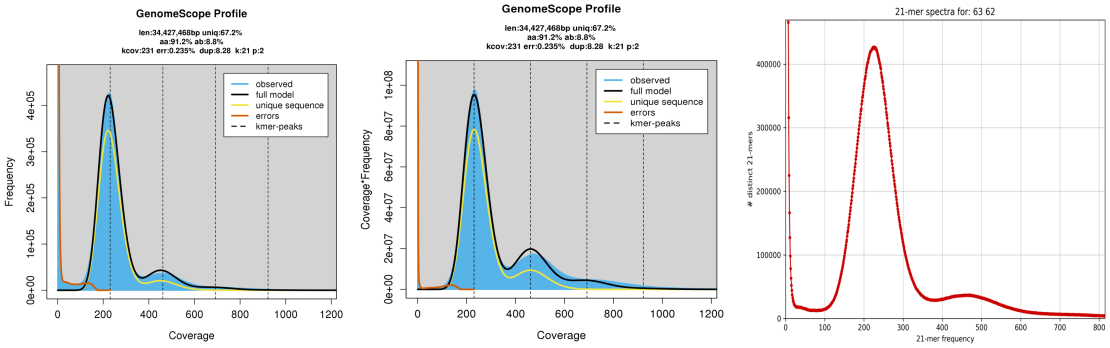


p = 2

property	min	max
Homozygous (aa)	89.7517%	90.3579%
Heterozygous (ab)	9.64212%	10.2483%
Genome Haploid Length	25,708,060 bp	25,723,042 bp
Genome Repeat Length	6,010,626 bp	6,014,129 bp
Genome Unique Length	19,697,435 bp	19,708,914 bp
Model Fit	88.8422%	96.1484%
Read Error Rate	0.248133%	0.248133%

CCFEE_5199

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

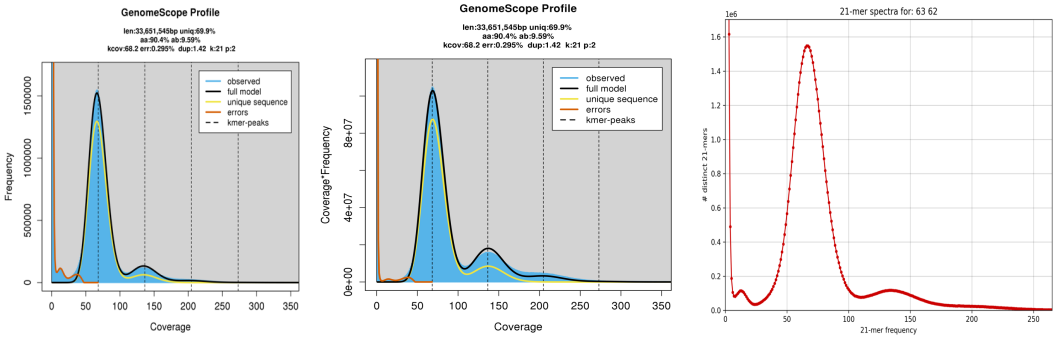


p = 2

property	min	max
Homozygous (aa)	89.8495%	92.5579%
Heterozygous (ab)	7.44214%	10.1505%
Genome Haploid Length	34,378,937 bp	34,427,468 bp
Genome Repeat Length	11,291,639 bp	11,307,578 bp
Genome Unique Length	23,087,298 bp	23,119,890 bp
Model Fit	91.1939%	92.0274%
Read Error Rate	0.235391%	0.235391%

CCFEE_5200

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

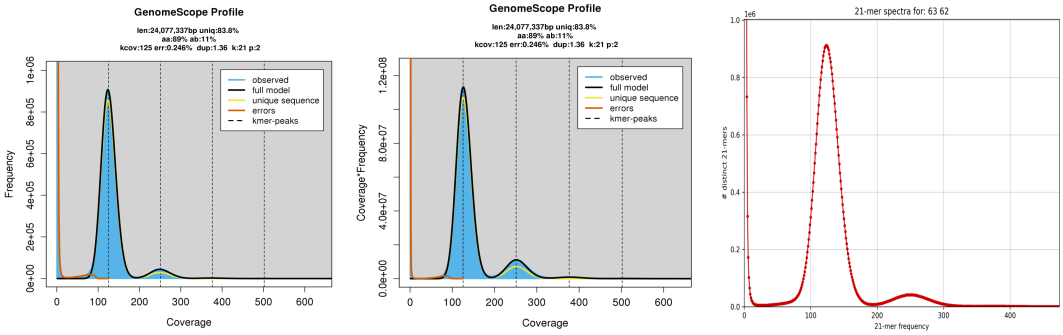


p = 2

property	min	max
Homozygous (aa)	88.186%	92.6384%
Heterozygous (ab)	7.36156%	11.814%
Genome Haploid Length	33,615,872 bp	33,651,545 bp
Genome Repeat Length	10,124,229 bp	10,134,973 bp
Genome Unique Length	23,491,643 bp	23,516,572 bp
Model Fit	90.5017%	93.4451%
Read Error Rate	0.295091%	0.295091%

CCFEE_5208

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

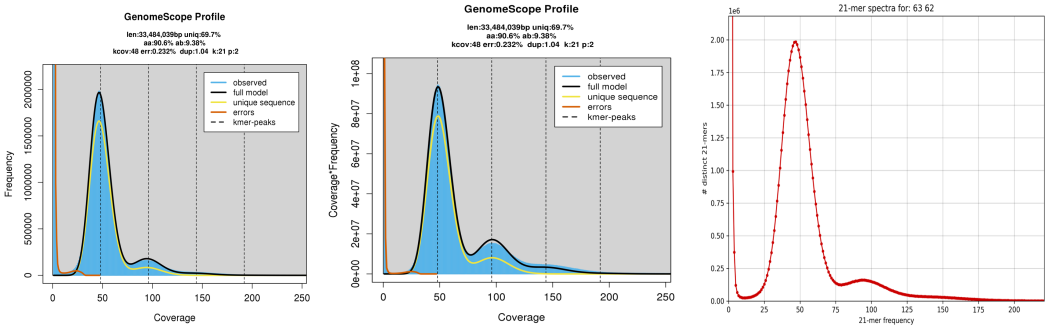


p = 2

property	min	max
Homozygous (aa)	88.6404%	89.3801%
Heterozygous (ab)	10.6199%	11.3596%
Genome Haploid Length	24,068,025 bp	24,077,337 bp
Genome Repeat Length	3,899,262 bp	3,900,771 bp
Genome Unique Length	20,168,763 bp	20,176,566 bp
Model Fit	91.5495%	96.3241%
Read Error Rate	0.245519%	0.245519%

CCFEE_5273

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

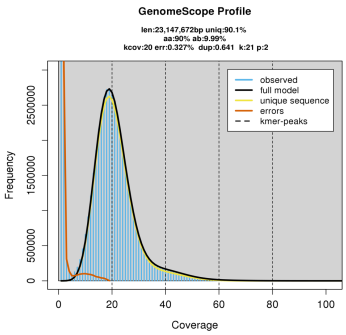


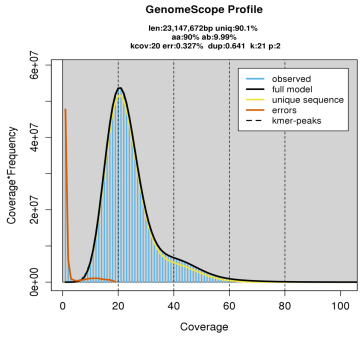
p = 2

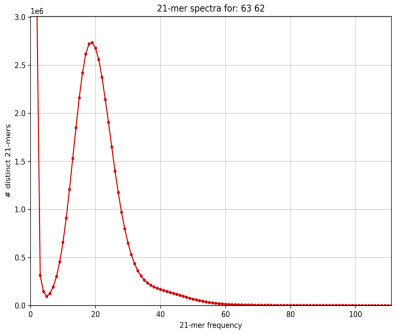
property	min	max
Homozygous (aa)	89.0959%	92.1349%
Heterozygous (ab)	7.86508%	10.9041%
Genome Haploid Length	33,452,528 bp	33,484,039 bp
Genome Repeat Length	10,132,645 bp	10,142,190 bp
Genome Unique Length	23,319,882 bp	23,341,849 bp
Model Fit	92.5195%	94.7015%
Read Error Rate	0.231745%	0.231745%

CCFEE_5275

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





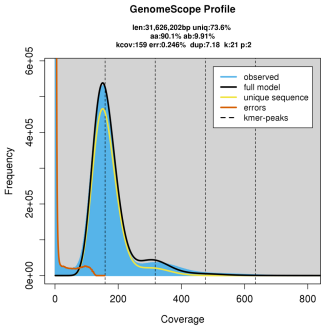


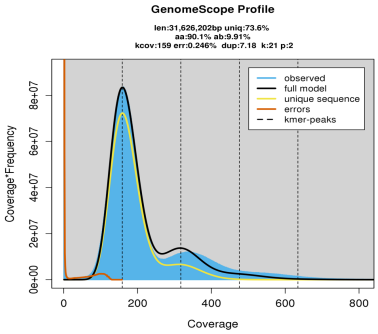
p = 2

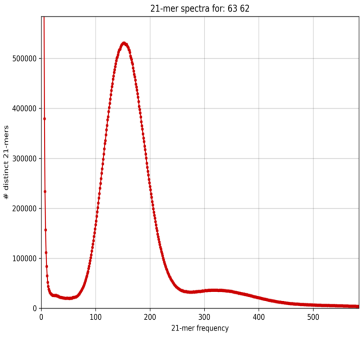
property	min	max
Homozygous (aa)	89.9469%	90.0768%
Heterozygous (ab)	9.92316%	10.0531%
Genome Haploid Length	23,119,722 bp	23,147,672 bp
Genome Repeat Length	2,279,852 bp	2,282,608 bp
Genome Unique Length	20,839,870 bp	20,865,063 bp
Model Fit	96.141%	99.3445%
Read Error Rate	0.327357%	0.327357%

CCFEE_5277

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





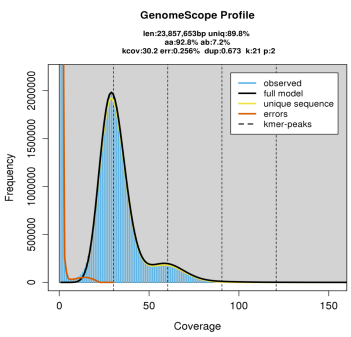


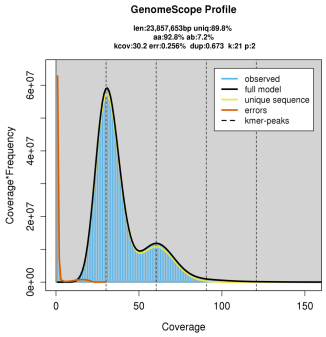
p = 2

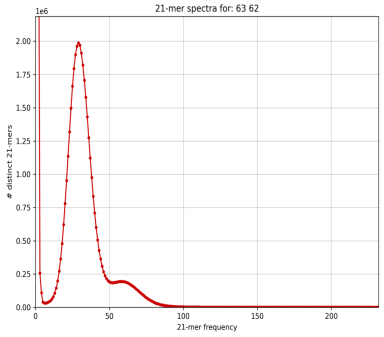
property	min	max
Homozygous (aa)	85.6634%	94.51%
Heterozygous (ab)	5.48996%	14.3366%
Genome Haploid Length	31,562,717 bp	31,626,202 bp
Genome Repeat Length	8,340,930 bp	8,357,707 bp
Genome Unique Length	23,221,786 bp	23,268,494 bp
Model Fit	88.7521%	89.4936%
Read Error Rate	0.245633%	0.245633%

CCFEE_5281

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





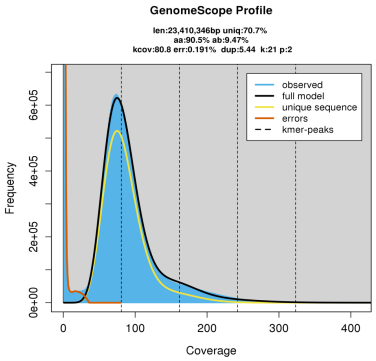


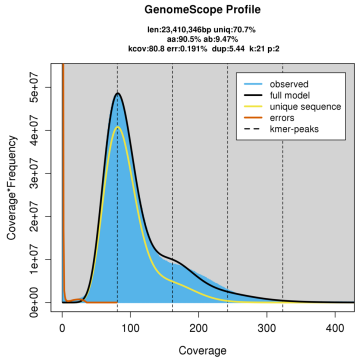
p = 2

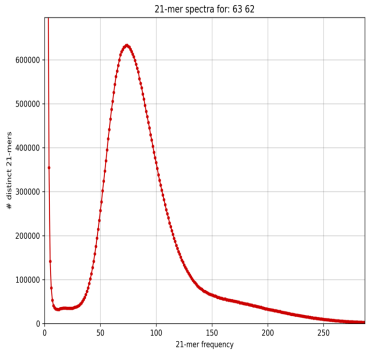
property	min	max
Homozygous (aa)	92.7827%	92.8268%
Heterozygous (ab)	7.17324%	7.21726%
Genome Haploid Length	23,843,126 bp	23,857,653 bp
Genome Repeat Length	2,435,216 bp	2,436,700 bp
Genome Unique Length	21,407,909 bp	21,420,953 bp
Model Fit	93.9693%	98.6016%
Read Error Rate	0.25607%	0.25607%

CCFEE_5283

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





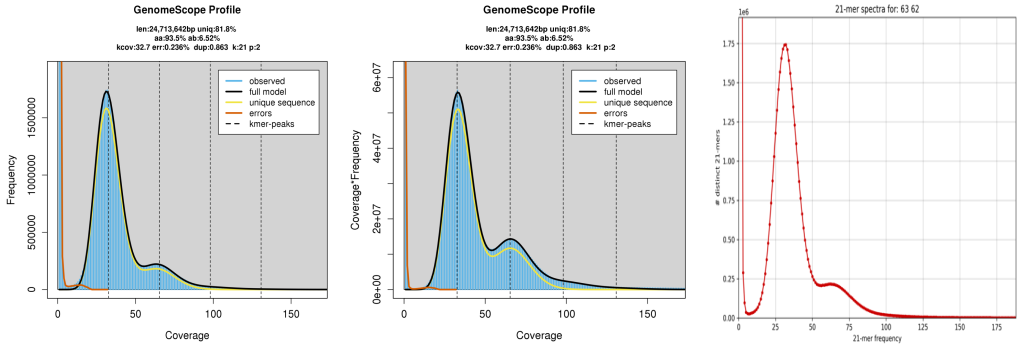


p = 2

property	min	max
Homozygous (aa)	89.0529%	92.0108%
Heterozygous (ab)	7.98924%	10.9471%
Genome Haploid Length	23,374,954 bp	23,410,346 bp
Genome Repeat Length	6,845,790 bp	6,856,155 bp
Genome Unique Length	16,529,164 bp	16,554,190 bp
Model Fit	93.5033%	95.5555%
Read Error Rate	0.190519%	0.190519%

CCFEE_5307

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

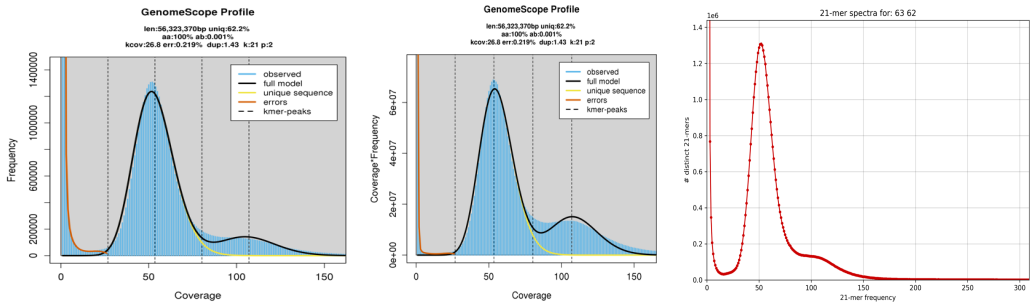


p = 2

property	min	max
Homozygous (aa)	93.4564%	93.4967%
Heterozygous (ab)	6.50328%	6.54363%
Genome Haploid Length	24,696,879 bp	24,713,642 bp
Genome Repeat Length	4,497,452 bp	4,500,505 bp
Genome Unique Length	20,199,427 bp	20,213,137 bp
Model Fit	95.2833%	97.8428%
Read Error Rate	0.235722%	0.235722%

CCFEE_5311

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

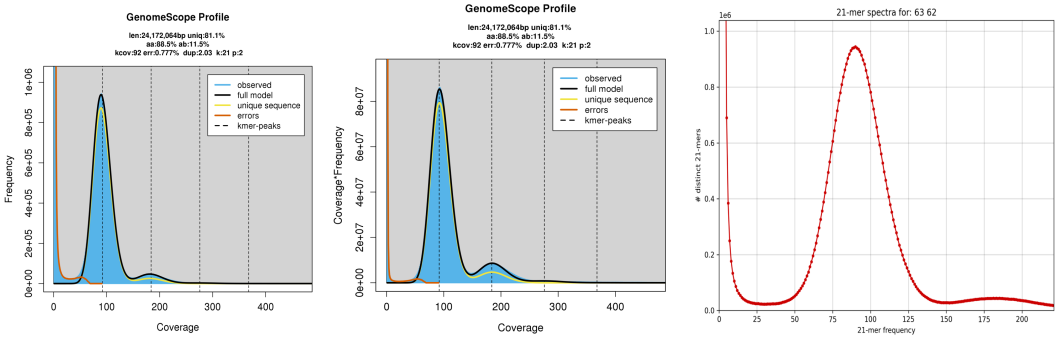


p = 2

property	min	max
Homozygous (aa)	99.972%	100%
Heterozygous (ab)	0%	0.0280474%
Genome Haploid Length	56,169,017 bp	56,323,370 bp
Genome Repeat Length	21,233,260 bp	21,291,609 bp
Genome Unique Length	34,935,757 bp	35,031,761 bp
Model Fit	75.5%	91.4153%
Read Error Rate	0.218963%	0.218963%

CCFEE_5311_v1

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

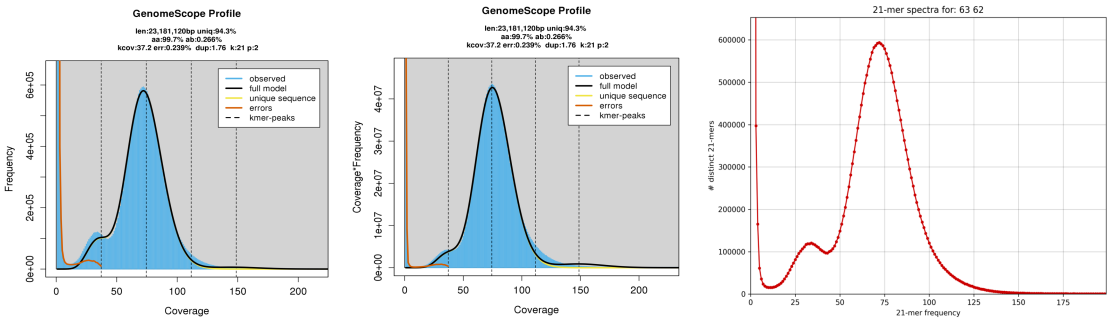


p = 2

property	min	max
Homozygous (aa)	87.2918%	89.7164%
Heterozygous (ab)	10.2836%	12.7082%
Genome Haploid Length	24,158,299 bp	24,172,064 bp
Genome Repeat Length	4,557,802 bp	4,560,399 bp
Genome Unique Length	19,600,497 bp	19,611,665 bp
Model Fit	91.078%	96.7316%
Read Error Rate	0.777225%	0.777225%

CCFEE_5486

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

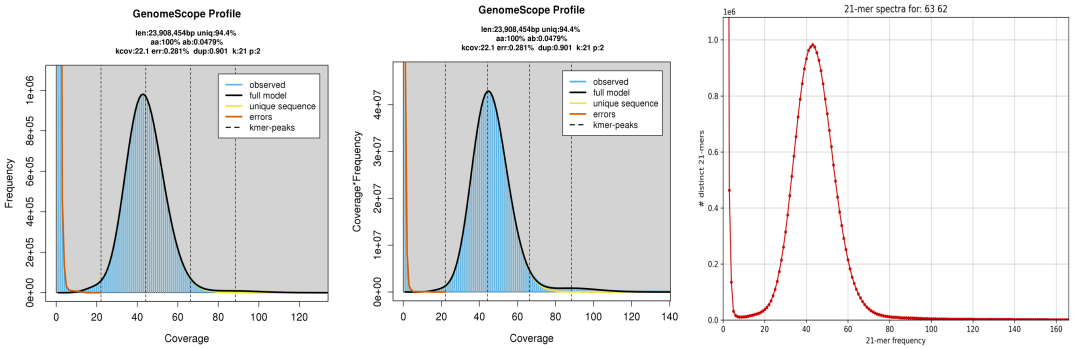


p = 2

property	min	max
Homozygous (aa)	99.7249%	99.7435%
Heterozygous (ab)	0.256494%	0.275132%
Genome Haploid Length	23,150,452 bp	23,181,120 bp
Genome Repeat Length	1,330,994 bp	1,332,757 bp
Genome Unique Length	21,819,458 bp	21,848,362 bp
Model Fit	91.2746%	93.6369%
Read Error Rate	0.239033%	0.239033%

CCFEE_6074

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

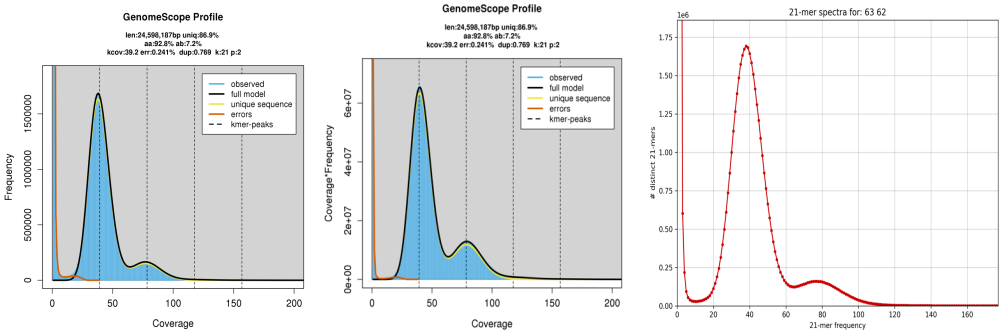


p = 2

property	min	max
Homozygous (aa)	99.9464%	99.9577%
Heterozygous (ab)	0.042251%	0.0535591%
Genome Haploid Length	23,893,630 bp	23,908,454 bp
Genome Repeat Length	1,341,641 bp	1,342,474 bp
Genome Unique Length	22,551,989 bp	22,565,980 bp
Model Fit	95.8086%	98.8234%
Read Error Rate	0.280607%	0.280607%

CCFEE_6081

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

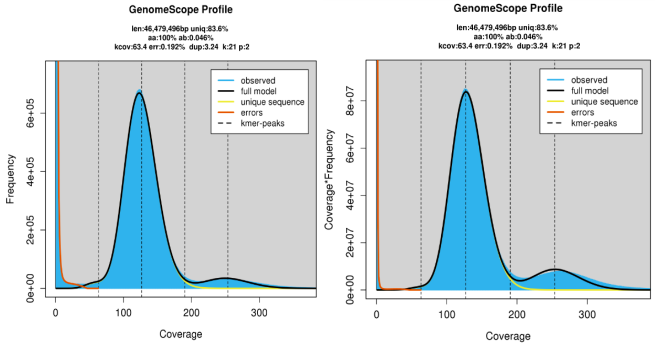


p = 2

property	min	max
Homozygous (aa)	92.7847%	92.813%
Heterozygous (ab)	7.18703%	7.21527%
Genome Haploid Length	24,589,805 bp	24,598,187 bp
Genome Repeat Length	3,224,964 bp	3,226,063 bp
Genome Unique Length	21,364,842 bp	21,372,124 bp
Model Fit	90.4782%	98.3287%
Read Error Rate	0.241174%	0.241174%

CCFEE_6082

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

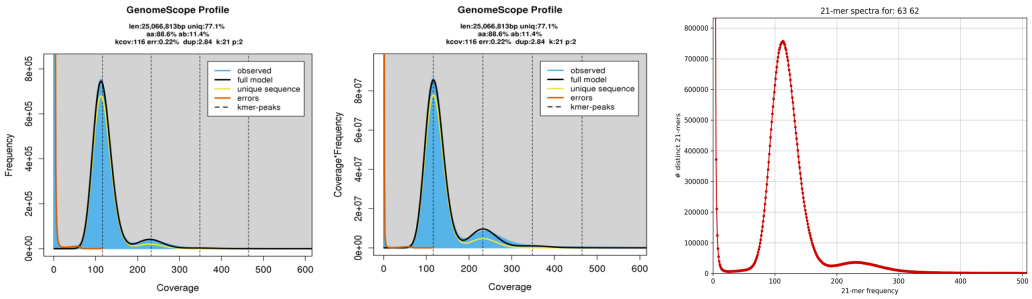


p = 2

property	min	max
Homozygous (aa)	99.9476%	99.9603%
Heterozygous (ab)	0.0397245%	0.0523672%
Genome Haploid Length	46,440,120 bp	46,479,496 bp
Genome Repeat Length	7,603,510 bp	7,609,956 bp
Genome Unique Length	38,836,611 bp	38,869,539 bp
Model Fit	92.9995%	96.1829%
Read Error Rate	0.19205%	0.19205%

CCFEE_6096

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram

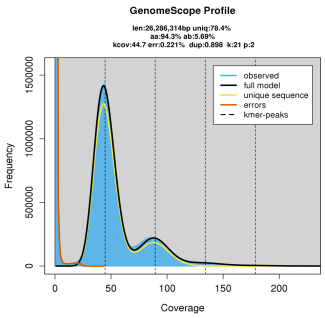


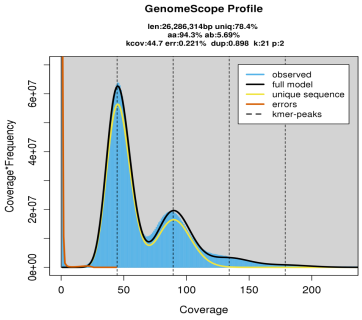
p = 2

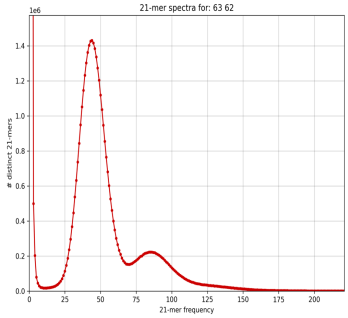
property	min	max
Homozygous (aa)	82.3891%	94.7873%
Heterozygous (ab)	5.21271%	17.6109%
Genome Haploid Length	25,046,044 bp	25,066,813 bp
Genome Repeat Length	5,740,258 bp	5,745,018 bp
Genome Unique Length	19,305,785 bp	19,321,795 bp
Model Fit	88.8361%	93.9397%
Read Error Rate	0.220213%	0.220213%

CCFEE_6249

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





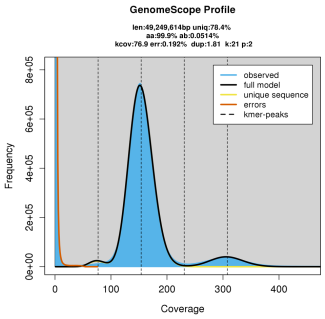


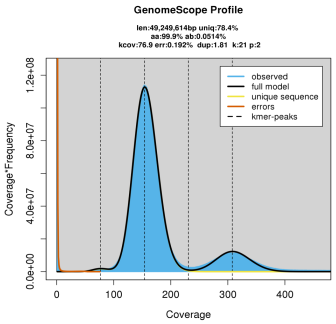
p = 2

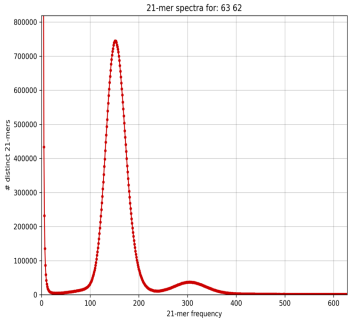
property	min	max
Homozygous (aa)	94.2759%	94.3345%
Heterozygous (ab)	5.66549%	5.72414%
Genome Haploid Length	26,254,013 bp	26,286,314 bp
Genome Repeat Length	5,667,825 bp	5,674,798 bp
Genome Unique Length	20,586,188 bp	20,611,516 bp
Model Fit	90.3973%	94.5023%
Read Error Rate	0.221096%	0.221096%

CCFEE_6250

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram





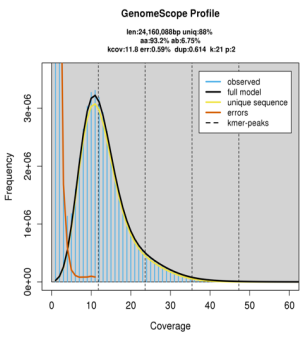


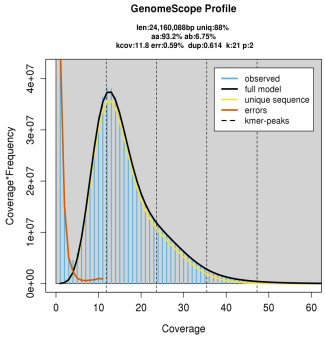
p = 2

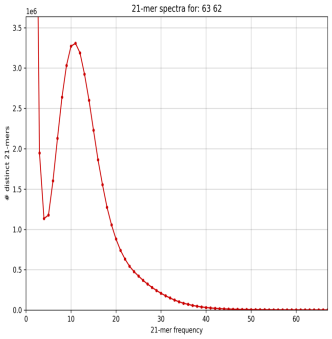
property	min	max
Homozygous (aa)	99.9428%	99.9544%
Heterozygous (ab)	0.0455983%	0.0572166%
Genome Haploid Length	49,219,925 bp	49,249,614 bp
Genome Repeat Length	10,610,345 bp	10,616,745 bp
Genome Unique Length	38,609,580 bp	38,632,869 bp
Model Fit	88.2774%	95.6884%
Read Error Rate	0.192162%	0.192162%

CCFEE_6464

Frequency/Coverage & Coverage*Frequency/Coverage & KAT Histogram







p = 2

property	min	max
Homozygous (aa)	93.1978%	93.2923%
Heterozygous (ab)	6.70767%	6.80217%
Genome Haploid Length	23,971,939 bp	24,160,088 bp
Genome Repeat Length	2,865,053 bp	2,887,540 bp
Genome Unique Length	21,106,886 bp	21,272,548 bp
Model Fit	94.9804%	98.973%
Read Error Rate	0.589645%	0.589645

Appendix D. The lost chapter-A sticky situation: Obtaining Clean DNA from High Hydrocarbon Contaminated Soils Reveals Micro-Colonial Fungi

Introduction

Popular references to tar provide us with vivid images of large mastodons and saber-tooth tigers being ensnared in thick, sticky molasses-like substance that oozes from the ground. Though, tar is not exactly the accurate term to be used here, as what erupts from the surface of the Earth is more formally known as naturally occurring asphalt. Asphalt, bitumen, creosote, or even pitch are different varieties of a similar hydrocarbon-containing substrate that is released and collected in large pools (Murali Krishnan and Rajagopal 2003). This material is created from bones, wood, and other organic substances that undergo a process called destructive distillation, which involves heating the organic materials in the absence of air (Murali Krishnan and Rajagopal 2003). Carbon-14 dating of the fossil material extracted from the Brea Pits indicates a range from 12,000-up to 40,0000 years (Howard 1960). Rancho La Brea or "La Brea Tarpits" is an example of a naturally occurring asphalt-accumulating region in the heart of Los Angeles, California, and is now recognized as one of the world's first 100 IUGS Geological Heritage sites (Hilario and International Union of Ge...). This geological marvel also contains the K-Pg boundary indicating the visible line of iridium deposit left by the asteroid that destroyed millions of dinosaurs and plant life eons ago which also helped place it on IUGS's list.

Rancho La Brea

A fascinating thought about these natural deposits is the millions of well-preserved fossils and bones of the unfortunate creatures caught in the viscous material. Geologists and many interns have been working on extracting bones, pollen, and seeds to compile a good assessment of the Quaternary period (2.6 million years ago). Microbial life, on the other hand, hasn't had much understanding in this environment, understandably so as the material is so thick and viscous causing issues with DNA extraction. Many researchers at La Brea have suggested that there is no DNA available in the asphalt, or the asphalt-soaked soils surrounding the seeps (personal communication). But there have been many successful attempts to culture and work with organisms indicating there is life and even DNA in these highly contaminated sites.

Microbes in Rancho La Brea

Kim and Crowley were fortunate enough to be the first to survey the microbial diversity using 16S rRNA and DNA of dioxygenase genes (Kim and Crowley 2007; Belcher et al. 2012). Their work was the first to take samples from Lot 91 (the main attraction in La Brea) and to culture out archaea and bacteria. Fluorescent staining was also attempted to visualize the complex microstructure of the microbial communities present in these extreme environments. But other than these two bodies of work, there has been no further push into Rancho La Brea on microbiome analysis.

Micro-colonial fungi are also found in hydrocarbon-contaminated soils

Hydrocarbon soil contamination is a common problem, and probably an issue we need to consider more of in the coming future. There has been prior work looking into creosote contamination sites, specifically concerning black yeasts (Gümral et al. 2014). Multiple species of *Exophiala* are commonly isolated from these sites, as they seem to be the most tolerant of these naturally occurring soil contaminants and some seem to even contain the proper mechanism to break down the chemical (Lladó et al. 2013; Seyedmousavi et al. 2011; Döğen et al. 2013; Zhao et al. 2010). But the hydrocarbon degradation is not limited to just this species, other genera of micro-colonial fungi (MCF) have also shown hydrocarbon degradation potential (Atagana et al. 2006).

Boldened by prior culturing results an attempt was made to collect, culture, and isolate DNA out of crude samples collected from Rancho La Brea in 2019. Below will be a general depiction of our early attempts and what possible future work will look like regarding the Rancho and fungal hydrocarbon degradation exploration.

Methods

Sampling

Once granted access to the site, we were able to collect soil samples from "boxes" found all over the rancho. The story behind the samples has to do with clearing soil for expanding LACMA's parking garage underneath the museum. Lots of displaced soil opened up a clearing where multiple sticky features with animal bones were found. These were "boxed up" and brought next door to the rancho to be further studied. We chose 6 boxes and one main lot (Lot 91) to collect our soil samples from. The scientists there work on each box a

little at a time slowly unearthing the pieces of bones from ancient wildlife. The samples are named: “Box 13”, “Box 7-P23-7A”, “Box P23-10A”, “Box 1860C-K10”, “Box 5B”, “Box 9”, and “Lot 91”.

Culturing and Isolation

Media for isolation, cultivation, and collection used were Malt Extract Yeast Extract (MEYE) media for nutrient-rich complex media and nutrient-poor, oligotrophic media Glucose Asparagine Agar (GAA). The recipe for both media can be found in our protocols.io and the step-by-step protocol for culturing MCFs (Kurbessoian 2019), also found in **Appendix 1**). The mixing of the sample with the media provides an environment with different layers of dissolved oxygen. Fast-growing fungi require full oxygenation and grow on the surface of the media, while slow-growing MCFs propagate on the bottom of the culture plate, requiring low levels of oxygen. This also allowed a much easier way to culture and select the strains. Once grown, axenic MCFs are isolated by using a flamed loop to pick and place onto clean media.

DNA extraction

Each isolate was grown on MEYE for approximately 1.5 weeks at 22℃. Genomic DNA was extracted from hyphae collected from MEYE plates using the CTAB protocol described in our protocols.io document (Carter-House et al. 2020). Melanin was removed using multiple iterations of phenol:chloroform and chloroform washes of our sample. Genomic DNA for one species, *Exophiala xenobiotica*, was measured by Nanodrop and diluted to ~28ng/ul. DNA extractions were sent to the UCR Genomics core (Riverside,

CA) for 2x150bp sequencing on an Illumina NovoSeq 6000. Amplicon sequencing was also performed on all seven samples collected. The following updated protocol for soil DNA extraction with high humic acid, using Qiagen's DNeasy Powersoil kit (Qiagen product: 47016) found here (Kurbessoian 2022). Steps to prepare for amplification with EMP primers are described here (P. Smith 2021). SequalPrep Normalization Plate Kit was used for DNA for a clean-up and normalization according to the manufacturer's protocols (Thermofisher product: A1051001) with 25 ul of PCR product added.

Genome assembly and annotation

Initial genome assembly was constructed for one MCF isolated from Lot 91 culturing attempts using Illumina sequencing. The genome was *de novo* assembled with the AAFTF pipeline (v.0.2.3) (Palmer and Stajich 2022) which performs read QC and filtering with BBTools bbduk (*v.38.86)* (Bushnell 2014) followed by SPAdes (*v.3.15.2*) (Bankevich *et al.* 2012) assembly using default parameters, followed by screening to remove short contigs < 200 bp and contamination using NCBI's VecScreen. The BUSCO ascomycota_odb10 database (Manni *et al.* 2021) was used to determine how complete the assembly was.

We predicted genes in *E. xenobiotica* genome assembly with Funannotate (*v1.8.1*) (Palmer and Stajich 2020). A masked genome was created by generating a library of sequence repeats with the RepeatModeler pipeline (Smit and Hubley 2008). These species-specific predicted repeats were combined with fungal repeats in the RepBase (Bao *et al.* 2015) to identify and mask repetitive regions in the genome assembly with RepeatMasker (*v.4-1-1*) (SMIT 2004). To predict genes, *ab initio* gene predictors SNAP (*v.2013_11_29*) (Korf

2004) and AUGUSTUS (*v.3.3.3*) (Stanke *et al.* 2006) were used along with additional gene models by GeneMark.HMM-ES (*v.4.62_lic*) (Brůna *et al.* 2020), and GlimmerHMM (*v.3.0.4*) (Majoros *et al.* 2004) utilize a self-training procedure to optimize *ab initio* predictions. Additional exon evidence to provide hints to gene predictors was generated by DIAMOND BLASTX alignment of SwissprotDB proteins and polished by Exonerate (*v.2.4.0*) (Slater and Birney 2005). Finally, EvidenceModeler (*v.1.1.1*) (Haas *et al.* 2008) generated consensus gene models in Funannotate that were constructed using default evidence weights. Non-protein-coding tRNA genes were predicted by tRNAscan-SE (*v.2.0.9*) (Lowe and Chan 2016).

The annotated genomes were processed with antiSMASH (*v.5.1.1*) (Blin *et al.* 2021) to predict secondary metabolite biosynthesis gene clusters. These annotations were also incorporated into the functional annotation by Funannotate. Putative protein functions were assigned to genes based on sequence similarity to InterProScan5 (*v.5.51-85.0*) (Jones *et al.* 2014), Pfam (*v.35.0*) (Finn *et al.* 2014), Eggnog (*v.2.1.6-d35afda*) (Huerta-Cepas *et al.* 2019), dbCAN2 (*v.9.0*) (Zhang *et al.* 2018) and MEROPS (*v.12.0*) (Rawlings *et al.* 2018) databases relying on NCBI BLAST (*v.2.9.0+)* (Sofi *et al.* 2022) and HMMer (*v.3.3.2*) (Potter *et al.* 2018). Gene Ontology terms were assigned to protein products based on the inferred homology based on these sequence similarity analyses. The final annotation produced by Funannotate was deposited in NCBI as a genome assembly with gene model annotation.

Sequence Read Archive (SRA) files and genome assembly and annotation files can be found under BioProject PRJNA631111 to be released soon.

ITS amplicon assessment

Amplicon DNA was demultiplexed by the UCR Genomics core (Riverside, CA) and returned to our lab ready to be analyzed. Using R programming and DADA2 R packaging (Ombrello 2020) samples including our positive and negative controls were analyzed and visualized. RScripts are available on our Zenodo link (https://zenodo.org/badge/latestdoi/568646119).

Results

Introducing Exophiala xenobiotica isolated from Rancho La Brea

On February 18, 2009, it was announced that in 2006 16 fossil deposits had been removed from the ground during the construction of an underground parking garage for LACMA (Los Angeles County Museum of Art) which is located right next to the Tar Pits. These fossils were packaged and moved to the museum property. There are a total of 23 large accumulations of tar and specimens which are named "Project 23". *E. xenobiotica TK-68* was isolated from "Project 5B" or Box 5B which was one of the 23 accumulations that are stored on-site. Approximately 45 MCFs were isolated from Box 5B, one of which was later renamed TK-68. Sequencing the ITS region allowed us to identify the species to be an *Exophiala xenobiotica* and was thus named as such. The entire genome is around 30.2Mbp which is roughly similar to genome lengths of other available Exophiala xenobiotica found online. The other known species were sequenced and submitted by the Broad institute in

2015 and is about 31.4Mbp (GCA_000835505.1). **Figure 1** shows both colony morphology phenotypes and the microscopy cellular morphology phenotype *E. xenobiotica* TK-68 displays. The colony morphology had an interesting bi-phasic feature which was confirmed to not be contaminated. The microscopy image was visualized under 40X and shows many budding yeast cells as some are attached to long extended hyphae. **Table 1** lists the genome content information after genome assembly and annotation.

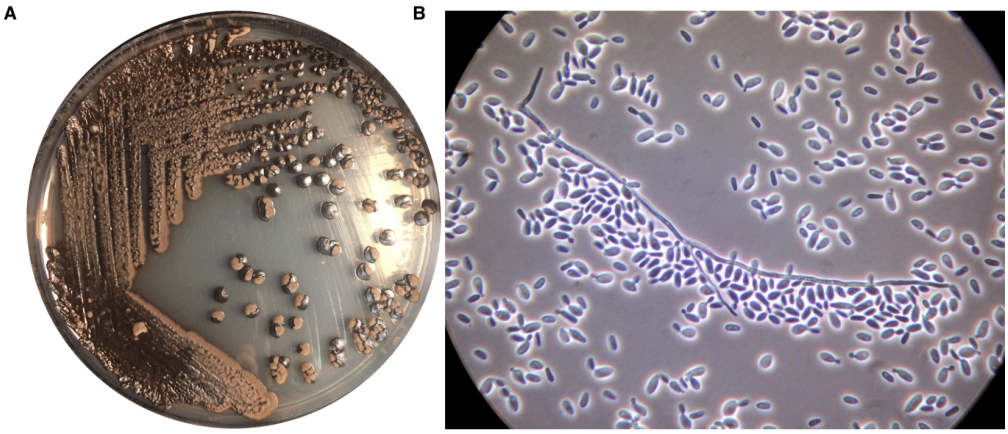


Figure 1. Plate colony morphology and cellular morphology of Exophiala xenobiotica isolated from Rancho La Brea. (A) Shows an aerial view of the species isolated from the first culturing efforts from Rancho La Brea. Individual colony morphology indicates an interesting biphasic morphology which was not due to contamination. (B) The same species were visualized under 40X microscopy indicating individual budding yeast features.

Table 1. Genome summary statistics generated on genome assembly of Exophiala xenobiotica TK68.

Species	*Exophiala xenobiotica TK68*
Contig Count	646
Total Length	30,280,774 bp
L50	15
N50	676,939
BUSCO Complete	97.9
Coverage	60x
Genes	11,317
mRNA	11,272
tRNA	45
GOterms	6,604
Interproscan	8,969
Eggnog	10,431
PFAM	7,886
CAZYME	321
MEROPS	356
Secretion	634

ITS Illumina sequencing

Preliminary analysis was performed on the ITS data using the DADA2 script available on our Zenodo link. **Figure 2** describes early data analysis on the ITS amplicon samples we received back from the UCR core. Data was relatively clean, though not of great quality. A rarefaction curve was derived from the ASV data associated with each dataset. Sample

Box 5B has the highest species richness and has been sequenced to reach full identification. Samples Box13 and Box P23-10A have the lowest number in diversity but also a low number of reads were recovered from their dataset. The mean Shannon Diversity index was calculated for each dataset indicating the diversity and evenness of each sample. Using this metric, sample Box 9, Box 5B and Box 1860C-K10 have higher diversity rates than the remaining four samples. Sample Box 91 seems to have the lowest Shannon Diversity when compared to the other 6 datasets. The Bray-Curtis metric was applied to our seven datasets to quantify dissimilarity and indicate which group of samples is more similar or different from the other. Box 9 and Box P23-10A are grouped together in the top right corner, while Box 7-P23-7A and Box 5B are grouped closer to each other, while the other two samples Lot 91 and Box 1860C-K10 are more dissimilar from all other samples. The final metric used to assess our data was the Relative abundance analysis where we observed all ASVs across the seven datasets to observe for the most abundant species. Multiple MCFs are present in this data set including *Exophiala* (the most abundant genera), *Cladosporium*, *Elasticomyces*, *Scytalidium,* and *Vermiconia*. Species from different phyla were also abundant including, *Basidiobolus*, *Bifiguratus*, and *Arrhenia* a mushroom-producing fungus.

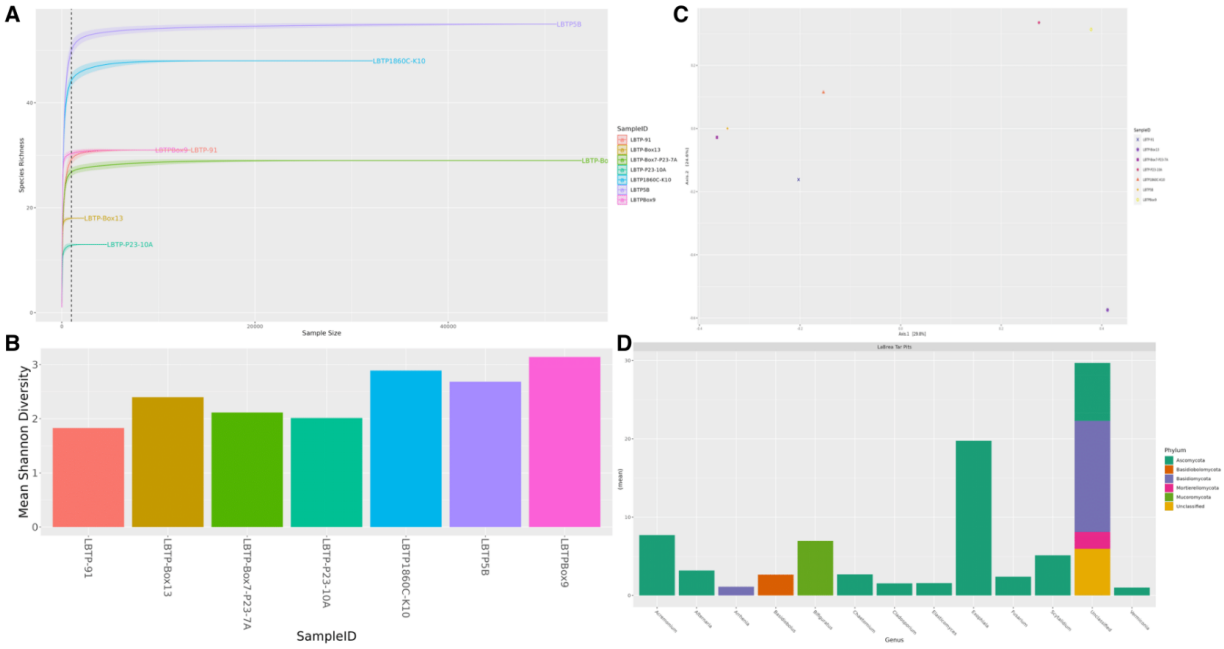


Figure 2. A combination of different statistical methods was used to tease apart ITS data collected from Rancho La Brea. (A) Rarefaction curves using 10,000 reads mark as an indicator of which samples need to be tested more, and which have reached the plateau. Sample Box 5B has the highest species richness and has been sequenced to reach full identification. Samples Box13 and Box P23-10A have the lowest number in diversity but also a low number of reads were recovered from their dataset. (B) Mean Shannon diversity plot or Alpha diversity plot comparing the diversity found in each sample. Using this metric, sample Box 9, Box 5B and Box 1860C-K10 have higher diversity rates than the remaining four samples. (C) Data from each sample set was plotted using the Bray-Curtis metric used to quantify dissimilarity and indicate which group of samples is more similar or different from each other. Box 9 and Box P23-10A are grouped together in the top right corner, while Box 7-P23-7A and Box 5B are grouped closer to each other, while the other two samples Lot 91 and Box 1860C-K10 are more dissimilar from all other samples. (D) Relative abundance was determined across all seven samples and compiled into one data to indicate the most abundant fungal species present. Multiple MCFs are present in this data set including *Exophiala* (the most abundant genera), *Cladosporium*, *Elasticomyces*, *Scytalidium,* and *Vermiconia*. Species from different phyla were also abundant including, *Basidiobolus*, *Bifiguratus*, and *Arrhenia* a mushroom-producing fungus.

Discussion

Here we demonstrated using culture-dependent methods of isolating and collecting large numbers of MCF which are alive in high hydrocarbon-contaminated soils. Along with culture-independent methods, where for the first time, was able to detect DNA from heavily

contaminated soils and indicated a large variety of fungal species DNA. Using the combined efforts of both culture-dependent and culture-independent we were able to connect amplicon ASV species of *Exophiala* to the cultured sample *Exophiala*.

As this was a simple preliminary study we believe there is huge potential in the work and the results present in our analysis. Future research would be to use more ecological sampling techniques for better metadata collection but also assessment of our data with low statistical noise. The 16S was also sequenced on this dataset using the EMP primers and could provide us with a deeper understanding of the total microbial population available in these highly contaminated and complex habitats.

The isolation of multiple MCF from this sample set tells us the fungal population is alive and thriving in this environment and may be impacting the hydrocarbon degradation. A deeper culturing attempt should be made to collect MCF but also secondary fungal species which could be providing an important role and function in this habitat. It is also an obvious resource for understanding hydrocarbon degradation pathways which could be vital for addressing the world's plastic problem and using microbial means of remediation.

Conclusion

As the plastic problem continues to grow on our planet, looking for alternative solutions is vital to ensure future generations' chance of survival. MCF have been found in a variety of extreme environments, and interestingly have been found in the hydrocarbon-soaked soils found in Rancho La Brea in the heart of Los Angeles, California. Using culture-dependent methods of isolating species will be important to amass a large variety of MCF which could be grown in larger quantities for remediation efforts. Amplicon sequencing was also used for the first time on heavily contaminated soils with viable data results linking amplicon ASV to cultured species.

References

Atagana HI, Haynes RJ, Wallis FM. 2006. Fungal Bioremediation of Creosote-Contaminated Soil: A Laboratory Scale Bioremediation Study Using Indigenous Soil Fungi. Water Air Soil Pollut. 172(1):201–219.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11.

Belcher RW, Huynh KV, Hoang TV, Crowley DE. 2012. Isolation of biosurfactant-producing bacteria from the Rancho La Brea Tar Pits. World J Microbiol Biotechnol. 28(12):3261–3267.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49(W1):W29–W35.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP : eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics and Bioinformatics. 2(2). doi:10.1093/nargab/lqaa026. http://dx.doi.org/10.1093/nargab/lqaa026.

Bushnell B. 2014. BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States) Report No.: LBNL-7065E. https://www.osti.gov/biblio/1241166.

Carter-House D, Stajich JE, Unruh S, Kurbessoian T. 2020. Fungal CTAB DNA Extraction. Protocols io. doi:10.17504/protocols.io.bhx8j7rw.

Döğen A, Ilkit M, de Hoog GS. 2013. Black yeast habitat choices and species spectrum on high altitude creosote-treated railway ties. Fungal Biol. 117(10):692–696.

Döğen A, Kaplan E, Ilkit M, de Hoog GS. 2013. Massive contamination of *Exophiala dermatitidis* and *E. phaeomuriformis* in railway stations in subtropical Turkey. Mycopathologia. 175(5-6):381–386.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42(Database issue):D222–30.

Gümral R, Tümgör A, Saraçlı MA, Yıldıran ŞT, Ilkit M, de Hoog GS. 2014. Black yeast diversity on creosoted railway sleepers changes with ambient climatic conditions. Microb Ecol. 68(4):699–707.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9(1):R7.

Howard H. 1960. Significance of carbon-14 dates for Rancho la Brea. Science. 131(3402):712–714.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

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

Kim J-S, Crowley DE. 2007. Microbial diversity in natural asphalts of the Rancho La Brea Tar Pits. Appl Environ Microbiol. 73(14):4579–4591.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics. 5:59.

Kurbessoian T. 2019. Cultivating melanized fungi from biological soil crust and rock surfaces v1. doi:10.17504/protocols.io.3fxgjpn.

Kurbessoian T. 2022. Low Biomass, high contamination Illumina DNA prep using DNeasy PowerSoil (Pro) Kit v1. doi:10.17504/protocols.io.6qpvr4ey3gmk/v1.

Lladó S, Gràcia E, Solanas AM, Viñas M. 2013. Fungal and bacterial microbial community assessment during bioremediation assays in an aged creosote-polluted soil. Soil Biol Biochem. 67:114–123.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44(W1):W54–7.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 20(16):2878–2879.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 38(10):4647–4654.

Murali Krishnan and J, Rajagopal KR. 2003. Review of the uses and modeling of bitumen from ancient to modern times. Appl Mech Rev. 56(2):149–214.

Ombrello AK. 2020. DADA2. Encyclopedia of Medical Immunology.:251–257. doi:10.1007/978-1-4614-8678-7_118.

Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation.

Palmer JM, Stajich JE. 2022. Automatic assembly for the fungi (AAFTF): genome assembly pipeline.

P. Smith D. 2021. Modified EMP ITS Illumina Amplicon Protocol v1. doi:10.17504/protocols.io.6qpvrdmeogmk/v1.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204.

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46(D1):D624–D632.

Seyedmousavi S, Badali H, Chlebicki A, Zhao J, Prenafeta-Boldú FX, De Hoog GS. 2011. Exophiala sideris, a novel black yeast isolated from environments polluted with toxic alkyl benzenes and arsenic. Fungal Biol. 115(10):1030–1037.

Sofi MY, Shafi A, Masoodi KZ. 2022. NCBI BLAST. Bioinformatics for Everyone.:95–102. doi:10.1016/b978-0-323-91128-3.00021-5.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:31.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34(Web Server issue):W435–9.

SMIT A. F. A. 2004. Repeat-Masker Open-3.0. http://www.repeatmasker.org.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46(W1):W95–W101.

Zhao J, Zeng J, de Hoog GS, Attili-Angelis D, Prenafeta-Boldú FX. 2010. Isolation and identification of black yeasts by enrichment on atmospheres of monoaromatic hydrocarbons. Microb Ecol. 60(1):149–156.

Appendix E. Black yeast tree with 304 species, using collected available data from GenBank along with collaborator data involving Antarctic cryptoendolithic samples

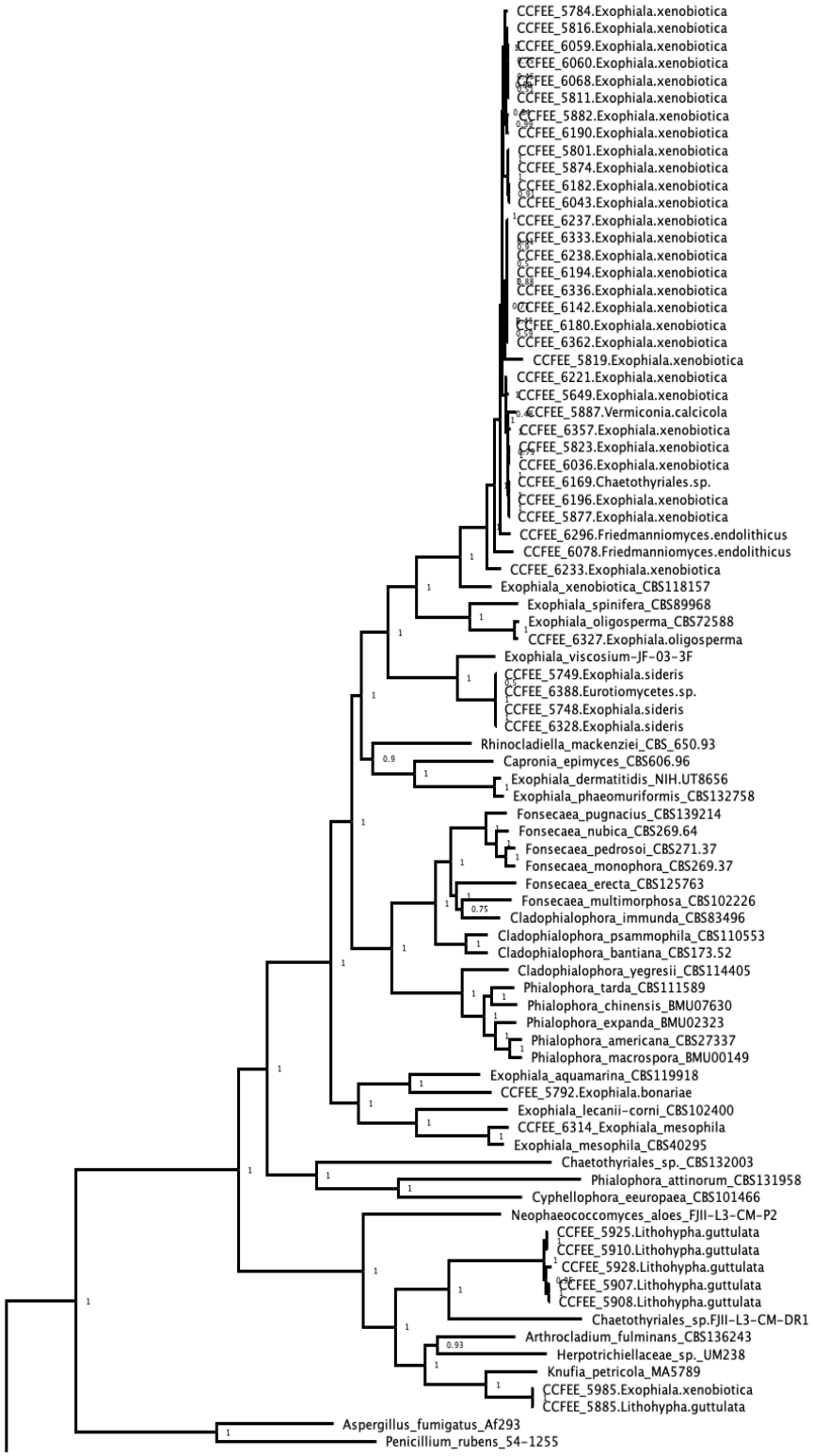
Our collaborative efforts with Dr. Laura Selbmann and Dr. Claudia Coleine for the past 8 years have given fruit to many publications on micro-colonial fungi isolated from Antarctica. In 2020 I received sequencing reads of 180+ species of micro-colonial fungi collected by the Italian team ready to be assembled, annotated, and added to a large phylogenetic tree. With the help of Dr. Stajich and I have been able to collect other species that could help bring clarity to the phylogenetic tree.

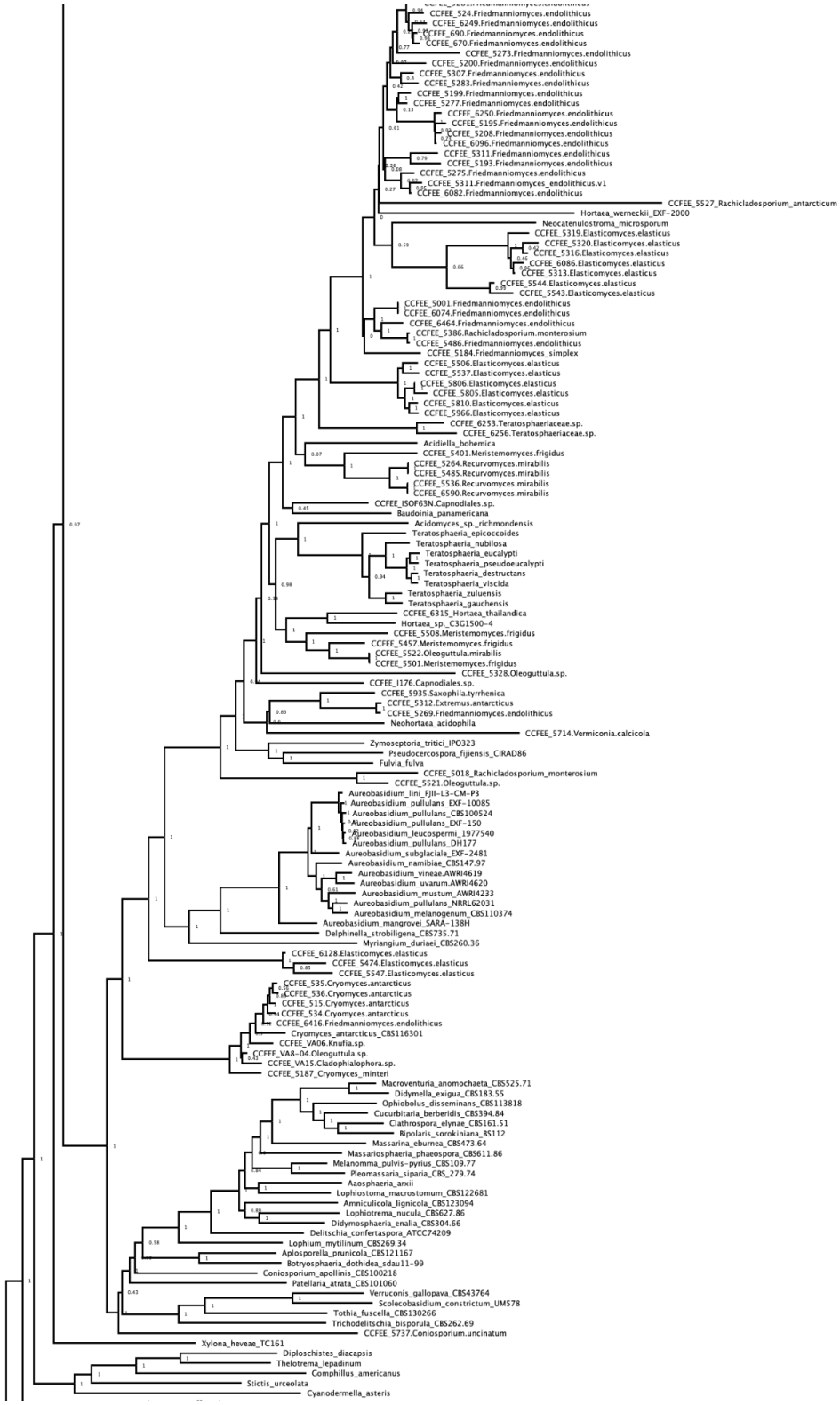
Methods

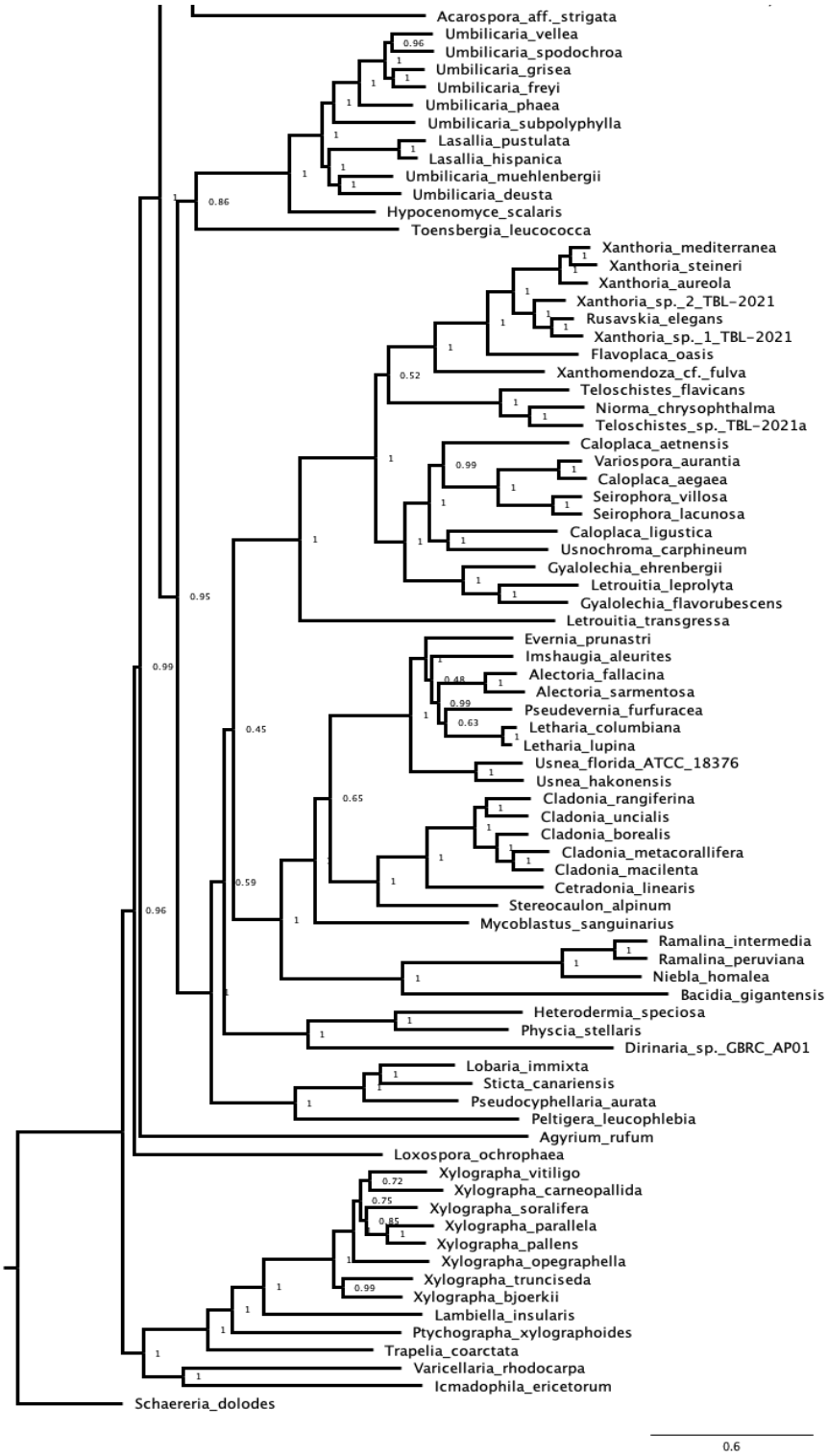
Genome Phylogenetic Tree Generation

Public genomic data were collected from NCBI's Genbank also seen in Table 1. Using BUSCO (v.5.3.2) ascomycota_odb10 (eukaryote 2020-09-10) dataset 1706 total BUSCO groups were searched in each genome. Another custom script found in the tool PHYling_unified (https://github.com/stajichlab/PHYling_unified) (busco_to_phyling.py) takes individual BUSCO family generated FASTA nucleotide sequences and fasta amino acid sequences and combines these results per species creating a combined BUSCO nucleotide FASTA sequence and a BUSCO amino acid FASTA sequence. The PHYling_unified Python script (PHYling.py) is then run on both datasets to generate final alignment files for CDS and PEP datasets per species. The Maketree script contained in the PHYling_unified toolset then takes each BUSCO family generated nucleotide and amino acid sequence across all species, aligned and builds FastTree

(*v.2.1.11*) (Price et al. 2010) files for each BUSCO family. This MakeTree script is repeated for both CDS and PEP datasets generated from BUSCO. The program ASTRAL (*v.5.15.5*) was used to combine all known species CDS and PEP-generated tree files into one tree file. RaXML-NG (*v.1.1.0*) (Stamatakis 2014) was used to add in the branch lengths using the alignment file generated from the PHYling.py step. A final (rename_tree_nodes.pl) script was used to take prefix names to rename the final tree. The tree was visualized using the R packages ggtree (*v.3.4.0*) (Yu et al. 2017), treeio (*v.1.20.0*) (Wang et al. 2020), and phytools (*v.1.0-3*) (Revell 2012).


CCFEE_5784.Exophiala.xenobiotica
CCFEE_5816.Exophiala.xenobiotica
CCFEE_6059.Exophiala.xenobiotica
CCFEE_6060.Exophiala.xenobiotica
CCFEE_6068.Exophiala.xenobiotica
CCFEE_5811.Exophiala.xenobiotica
CCFEE_5882.Exophiala.xenobiotica
CCFEE_6190.Exophiala.xenobiotica
CCFEE_5801.Exophiala.xenobiotica
CCFEE_5874.Exophiala.xenobiotica
CCFEE_6182.Exophiala.xenobiotica
CCFEE_6043.Exophiala.xenobiotica
CCFEE_6237.Exophiala.xenobiotica
CCFEE_6333.Exophiala.xenobiotica
CCFEE_6238.Exophiala.xenobiotica
CCFEE_6194.Exophiala.xenobiotica
CCFEE_6336.Exophiala.xenobiotica
CCFEE_6142.Exophiala.xenobiotica
CCFEE_6180.Exophiala.xenobiotica
CCFEE_6362.Exophiala.xenobiotica
CCFEE_5819.Exophiala.xenobiotica
CCFEE_6221.Exophiala.xenobiotica
CCFEE_5649.Exophiala.xenobiotica
CCFEE_5887.Vermiconia.calcicola
CCFEE_6357.Exophiala.xenobiotica
CCFEE_5823.Exophiala.xenobiotica
CCFEE_6036.Exophiala.xenobiotica
CCFEE_6169.Chaetothyriales.sp.
CCFEE_6196.Exophiala.xenobiotica
CCFEE_5877.Exophiala.xenobiotica
CCFEE_6296.Friedmanniomyces.endolithicus
CCFEE_6078.Friedmanniomyces.endolithicus
CCFEE_6233.Exophiala.xenobiotica
Exophiala_xenobiotica_CBS118157
Exophiala_spinifera_CBS89968
Exophiala_oligosperma_CBS72588
CCFEE_6327.Exophiala.oligosperma
Exophiala_viscosium-JF-03-3F
CCFEE_5749.Exophiala.sideris
CCFEE_6388.Eurotiomycetes.sp.
CCFEE_5748.Exophiala.sideris
CCFEE_6328.Exophiala.sideris
Rhinocladiella_mackenziei_CBS_650.93
Capronia_epimyces_CBS606.96
Exophiala_dermatitidis_NIH.UT8656
Exophiala_phaeomuriformis_CBS132758
Fonsecaea_pugnacius_CBS139214
Fonsecaea_nubica_CBS269.64
Fonsecaea_pedrosoi_CBS271.37
Fonsecaea_monophora_CBS269.37
Fonsecaea_erecta_CBS125763
Fonsecaea_multimorphosa_CBS102226
Cladophialophora_immunda_CBS83496
Cladophialophora_psammophila_CBS110553
Cladophialophora_bantiana_CBS173.52
Cladophialophora_yegresii_CBS114405
Phialophora_tarda_CBS111589
Phialophora_chinensis_BMU07630
Phialophora_expanda_BMU02323
Phialophora_americana_CBS27337
Phialophora_macrospora_BMU00149
Exophiala_aquamarina_CBS119918
CCFEE_5792.Exophiala.bonariae
Exophiala_lecanii-corni_CBS102400
CCFEE_6314_Exophiala_mesophila
Exophiala_mesophila_CBS40295
Chaetothyriales_sp._CBS132003
Phialophora_attinorum_CBS131958
Cyphellophora_eeuropaea_CBS101466
Neophaeococcomyces_aloes_FJII-L3-CM-P2
CCFEE_5925.Lithohypha.guttulata
CCFEE_5910.Lithohypha.guttulata
CCFEE_5928.Lithohypha.guttulata
CCFEE_5907.Lithohypha.guttulata
CCFEE_5908.Lithohypha.guttulata
Chaetothyriales_sp.FJII-L3-CM-DR1
Arthrocladium_fulminans_CBS136243
Herpotrichiellaceae_sp._UM238
Knufia_petricola_MA5789
CCFEE_5985.Exophiala.xenobiotica
CCFEE_5885.Lithohypha.guttulata
Aspergillus_fumigatus_Af293
Penicillium_rubens_54-1255


CCFEE_524.Friedmanniomyces.endolithicus
CCFEE_6249.Friedmanniomyces.endolithicus
CCFEE_690.Friedmanniomyces.endolithicus
CCFEE_670.Friedmanniomyces.endolithicus
CCFEE_5273.Friedmanniomyces.endolithicus
CCFEE_5200.Friedmanniomyces.endolithicus
CCFEE_5307.Friedmanniomyces.endolithicus
CCFEE_5283.Friedmanniomyces.endolithicus
CCFEE_5199.Friedmanniomyces.endolithicus
CCFEE_5277.Friedmanniomyces.endolithicus
CCFEE_6250.Friedmanniomyces.endolithicus
CCFEE_5195.Friedmanniomyces.endolithicus
CCFEE_5208.Friedmanniomyces.endolithicus
CCFEE_6096.Friedmanniomyces.endolithicus
CCFEE_5311.Friedmanniomyces.endolithicus
CCFEE_5193.Friedmanniomyces.endolithicus
CCFEE_5275.Friedmanniomyces.endolithicus
CCFEE_5311.Friedmanniomyces_endolithicus.v1
CCFEE_6082.Friedmanniomyces.endolithicus
CCFEE_5527_Rachicladosporium_antarcticum
Hortaea_werneckii_EXF-2000
Neocatenulostroma_microsporum
CCFEE_5319.Elasticomyces.elasticus
CCFEE_5320.Elasticomyces.elasticus
CCFEE_5316.Elasticomyces.elasticus
CCFEE_6086.Elasticomyces.elasticus
CCFEE_5313.Elasticomyces.elasticus
CCFEE_5544.Elasticomyces.elasticus
CCFEE_5543.Elasticomyces.elasticus
CCFEE_5001.Friedmanniomyces.endolithicus
CCFEE_6074.Friedmanniomyces.endolithicus
CCFEE_6464.Friedmanniomyces.endolithicus
CCFEE_5386.Rachicladosporium.monterosium
CCFEE_5486.Friedmanniomyces.endolithicus
CCFEE_5184.Friedmanniomyces_simplex
CCFEE_5506.Elasticomyces.elasticus
CCFEE_5537.Elasticomyces.elasticus
CCFEE_5806.Elasticomyces.elasticus
CCFEE_5805.Elasticomyces.elasticus
CCFEE_5810.Elasticomyces.elasticus
CCFEE_5966.Elasticomyces.elasticus
CCFEE_6253.Teratosphaeriaceae.sp.
CCFEE_6256.Teratosphaeriaceae.sp.
Acidiella_bohemica
CCFEE_5401.Meristemomyces.frigidus
CCFEE_5264.Recurvomyces.mirabilis
CCFEE_5485.Recurvomyces.mirabilis
CCFEE_5536.Recurvomyces.mirabilis
CCFEE_6590.Recurvomyces.mirabilis
CCFEE_ISOF63N.Capnodiales.sp.
Baudoinia_panamericana
Acidomyces_sp._richmondensis
Teratosphaeria_epicoccoides
Teratosphaeria_nubilosa
Teratosphaeria_eucalypti
Teratosphaeria_pseudoeucalypti
Teratosphaeria_destructans
Teratosphaeria_viscida
Teratosphaeria_zuluensis
Teratosphaeria_gauchensis
CCFEE_6315_Hortaea_thailandica
Hortaea_sp._C3G1500-4
CCFEE_5508.Meristemomyces.frigidus
CCFEE_5457.Meristemomyces.frigidus
CCFEE_5522.Oleoguttula.mirabilis
CCFEE_5501.Meristemomyces.frigidus
CCFEE_5328.Oleoguttula.sp.
CCFEE_I176.Capnodiales.sp.
CCFEE_5935.Saxophila.tyrrhenica
CCFEE_5312.Extremus.antarcticus
CCFEE_5269.Friedmanniomyces.endolithicus
Neohortaea_acidophila
CCFEE_5714.Vermiconia.calcicola
Zymoseptoria_tritici_IPO323
Pseudocercospora_fijiensis_CIRAD86
Fulvia_fulva
CCFEE_5018_Rachicladosporium_monterosium
CCFEE_5521.Oleoguttula.sp.
Aureobasidium_lini_FJII-L3-CM-P3
Aureobasidium_pullulans_EXF-10085
Aureobasidium_pullulans_CBS100524
Aureobasidium_pullulans_EXF-150
Aureobasidium_leucospermi_1977540
Aureobasidium_pullulans_DH177
Aureobasidium_subglaciale_EXF-2481
Aureobasidium_namibiae_CBS147.97
Aureobasidium_vineae.AWRI4619
Aureobasidium_uvarum.AWRI4620
Aureobasidium_mustum_AWRI4233
Aureobasidium_pullulans_NRRL62031
Aureobasidium_melanogenum_CBS110374
Aureobasidium_mangrovei_SARA-138H
Delphinella_strobiligena_CBS735.71
Myriangium_duriaei_CBS260.36
CCFEE_6128.Elasticomyces.elasticus
CCFEE_5474.Elasticomyces.elasticus
CCFEE_5547.Elasticomyces.elasticus
CCFEE_535.Cryomyces.antarcticus
CCFEE_536.Cryomyces.antarcticus
CCFEE_515.Cryomyces.antarcticus
CCFEE_534.Cryomyces.antarcticus
CCFEE_6416.Friedmanniomyces.endolithicus
Cryomyces_antarcticus_CBS116301
CCFEE_VA06.Knufia.sp.
CCFEE_VA8-04.Oleoguttula.sp.
CCFEE_VA15.Cladophialophora.sp.
CCFEE_5187_Cryomyces_minteri
Macroventuria_anomochaeta_CBS525.71
Didymella_exigua_CBS183.55
Ophiobolus_disseminans_CBS113818
Cucurbitaria_berberidis_CBS394.84
Clathrospora_elynae_CBS161.51
Bipolaris_sorokiniana_BS112
Massarina_eburnea_CBS473.64
Massariosphaeria_phaeospora_CBS611.86
Melanomma_pulvis-pyrius_CBS109.77
Pleomassaria_siparia_CBS_279.74
Aaosphaeria_arxii
Lophiostoma_macrostomum_CBS122681
Amniculicola_lignicola_CBS123094
Lophiotrema_nucula_CBS627.86
Didymosphaeria_enalia_CBS304.66
Delitschia_confertaspora_ATCC74209
Lophium_mytilinum_CBS269.34
Aplosporella_prunicola_CBS121167
Botryosphaeria_dothidea_sdau11-99
Coniosporium_apollinis_CBS100218
Patellaria_atrata_CBS101060
Verruconis_gallopava_CBS43764
Scolecobasidium_constrictum_UM578
Tothia_fuscella_CBS130266
Trichodelitschia_bisporula_CBS262.69
CCFEE_5737.Coniosporium.uncinatum
Xylona_heveae_TC161
Diploschistes_diacapsis
Thelotrema_lepadinum
Gomphillus_americanus
Stictis_urceolata
Cyanodermella_asteris


Acarospora_aff._strigata
Umbilicaria_vellea
Umbilicaria_spodochroa
Umbilicaria_grisea
Umbilicaria_freyi
Umbilicaria_phaea
Umbilicaria_subpolyphylla
Lasallia_pustulata
Lasallia_hispanica
Umbilicaria_muehlenbergii
Umbilicaria_deusta
Hypocenomyce_scalaris
Toensbergia_leucococca
Xanthoria_mediterranea
Xanthoria_steineri
Xanthoria_aureola
Xanthoria_sp._2_TBL-2021
Rusavskia_elegans
Xanthoria_sp._1_TBL-2021
Flavoplaca_oasis
Xanthomendoza_cf._fulva
Teloschistes_flavicans
Niorma_chrysophthalma
Teloschistes_sp._TBL-2021a
Caloplaca_aetnensis
Variospora_aurantia
Caloplaca_aegaea
Seirophora_villosa
Seirophora_lacunosa
Caloplaca_ligustica
Usnochroma_carphineum
Gyalolechia_ehrenbergii
Letrouitia_leprolyta
Gyalolechia_flavorubescens
Letrouitia_transgressa
Evernia_prunastri
Imshaugia_aleurites
Alectoria_fallacina
Alectoria_sarmentosa
Pseudevernia_furfuracea
Letharia_columbiana
Letharia_lupina
Usnea_florida_ATCC_18376
Usnea_hakonensis
Cladonia_rangiferina
Cladonia_uncialis
Cladonia_borealis
Cladonia_metacorallifera
Cladonia_macilenta
Cetradonia_linearis
Stereocaulon_alpinum
Mycoblastus_sanguinarius
Ramalina_intermedia
Ramalina_peruviana
Niebla_homalea
Bacidia_gigantensis
Heterodermia_speciosa
Physcia_stellaris
Dirinaria_sp._GBRC_AP01
Lobaria_immixta
Sticta_canariensis
Pseudocyphellaria_aurata
Peltigera_leucophlebia
Agyrium_rufum
Loxospora_ochrophaea
Xylographa_vitiligo
Xylographa_carneopallida
Xylographa_soralifera
Xylographa_parallela
Xylographa_pallens
Xylographa_opegraphella
Xylographa_trunciseda
Xylographa_bjoerkii
Lambiella_insularis
Ptychographa_xylographoides
Trapelia_coarctata
Varicellaria_rhodocarpa
Icmadophila_ericetorum
Schaereria_dolodes
0.6