Genomic Analysis of *Exophiala SP. ETNP2018*, a Marine Fungal Strain Isolated From an Oxygen Minimum Zone

Margaret Kelly Bernish

Follow this and additional works at: https://scholarcommons.sc.edu/etd

Part of the Animal Sciences Commons

**Recommended Citation**


This Open Access Thesis is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.
GENOMIC ANALYSIS OF EXOPHIALA SP. ETNP2018, A MARINE FUNGAL STRAIN ISOLATED FROM AN OXYGEN MINIMUM ZONE

by

Margaret Kelly Bernish

Bachelor of Science

University of North Carolina at Chapel Hill, 2020

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Marine Science

College of Arts and Sciences

University of South Carolina

2023

Accepted by:

Xuefeng (Nick) Peng, Director of Thesis

Annie Bourbonnais, Reader

Joshua Stone, Reader

Cheryl L. Addy, Interim Vice Provost and Dean of the Graduate School
ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Xuefeng (Nick) Peng and the members of the Peng Lab. To my fellow graduate students and great friends: Maddie and Birch, as well as the undergraduate members of the Peng Lab: Sam, Julianna, Sydney, Lily, Liam, and Dylan—thank you all so much for your support. Many thanks to my committee members, Dr. Bourbonnais and Dr. Stone, as well, for your guidance and help along the way. Further thanks to Dr. Annie Bourbonnais for collecting seawater samples from the ETNP. Lastly, I would like to mention my dear friends Abby, Meredith, and Darcy. Thank you for being there for me.
ABSTRACT

Microbial metabolic activity is central to biogeochemical cycling in virtually all ecosystems. Fungi are important parts of microbial communities, serving as decomposers, parasites, and symbionts. Most of our understanding of fungal ecology is derived from the study of isolated, cultured terrestrial fungi. However, the marine environment includes thousands of divergent, uncultured fungi. A number of these species, which include saprobes, parasites, and endophytes, can survive and sporulate in a wide range of salinities. Other such species face the additional challenge of limited nutrient accessibility. The genus *Exophiala* includes agents of mycoses in both humans and in marine animals. This research seeks to characterize the genome of *Exophiala* sp. ETNP2018, a novel strain of marine fungi isolated from the Eastern Tropical North Pacific (ETNP) oxygen minimum zone, revealing the genetic adaptations of this strain to its unique environment, including low oxygen conditions and nutrient limitation. Structural and functional annotation, as well as a comparison to the representative species of its genus (*Exophiala aquamarina*), serves to elucidate the lifestyle of this novel strain. Furthermore, this research will, through the identification of protein-encoding sequences, determine the mechanisms used by this species to break down and metabolize biomass in the ocean.
TABLE OF CONTENTS

Acknowledgements ........................................................................................................ iii

Abstract ......................................................................................................................... iv

List of Tables ................................................................................................................ vii

List of Figures ................................................................................................................ ix

List of Abbreviations ..................................................................................................... xii

Chapter 1: Optimizing a DNA Extraction Protocol to Extract High Molecular Weight Genomic DNA from Marine Fungi Isolated from Salt Marsh Sediments ........................................................................ 1

1.1 Introduction ........................................................................................................... 1

1.2 Methodology ......................................................................................................... 2

1.3 Results .................................................................................................................. 6

1.4 Discussion ............................................................................................................. 11

Chapter 2: Identification of Marine Fungi using Marker Genes .................................. 13

2.1 Introduction ........................................................................................................... 13

2.2 Methodology ......................................................................................................... 15

2.3 Results .................................................................................................................. 17

2.4 Discussion ............................................................................................................. 21

Chapter 3: The genome of *Exophiala sp.* ETNP2018, isolated from the Eastern Tropical North Pacific Ocean ................................................................. 23

3.1 Introduction ........................................................................................................... 23

3.2 Methodology ......................................................................................................... 24
3.3 Genomic Analysis of *Exophiala sp.* ETNP2018.................................25

3.4 Discussion.....................................................................................38

References..........................................................................................44

Appendix A: Supplementary Figures..................................................51
LIST OF TABLES

Table 1.1 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA extracted using different lysis buffer volumes .................................................................7

Table 1.2 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA extracted using different bead beating times................................................................................................................8

Table 1.3 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA samples (with one replicate each) extracted using different kits................................................................................................................9

Table 1.4 Comparison of lysis temperature and bead beating time for DNA extracted using the DNeasy Plant Mini Kit .................................................................10

Table 1.5 Comparison of various vortexing times for DNA extracted using the DNeasy Plant Mini Kit .................................................................10

Table 2.1 Display of the various samples used in experiments, their dates and locations of sampling, and their denotation in this document...........................................................................................................16

Table 3.1 Genomic characteristics of the Exophiala sp. ETNP2018.................................26

Table 3.2 Classification of glycoside hydrolases, polysaccharide lyases, and carbohydrate esterases according to catalytic domains .................................................................................................................26

Table 3.3 Most commonly occurring protein families/domains, listed alongside their respective descriptions and number of occurrences within the Exophiala sp. ETNP2018 genome .................................................34

Table 3.4 Genes encoding biosynthetic gene clusters (BGCs) ........................................35

Table 3.5 Clusters of orthologous groups (COGs) detected in Exophiala sp. ETNP2018 ..................................................................................................................35
Table 3.6 Overview of gene statistics of *Exophiala sp.* ETNP2018 and *Exophiala aquamarina*...
LIST OF FIGURES

Figure 1.1 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted using various volumes of Buffer AP1 .................................................................7

Figure 1.2 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted with the various bead beating times........................................................................8

Figure 1.3 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted from the corresponding kits...........................................................................................9

Figure 1.4 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted using the various vortex times..........................................................................................11

Figure 2.1 Sampling locations in the ETNP oxygen minimum zone. Colormap shows oxygen concentration at the surface from World Ocean Atlas 2018 (yearly average from 2018). ........................................15

Figure 2.2 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): 6, N1, M2, M30, M36, and a negative control .........................................................................................17

Figure 2.3 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): M36, M36 (R2), M15, M2, M30, M2 (R2), M15 (R2), M30 (R2), M47, and a negative control .................................................................18

Figure 2.4 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): diluted M15, concentrated M15, M11, and a negative control .................................................................................19

Figure 2.5 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the corresponding primer sets (labeled in red): three M15 samples, M36, M3 (positive control), and a negative control .........................................................................................20
Figure 2.6 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the corresponding primer sets (labeled in red): M3 (positive control), M15 (R1), M15 (R2), and a negative control.

Figure 3.1 Sampling location of *Exophiala* sp. ETNP2018. Colormap shows oxygen concentration at 40 meters depth from World Ocean Atlas 2018 (yearly average from 2018).

Figure 3.2 Classification of CAZymes detected in *Exophiala* sp. ETNP2018, according to Table 3.2.

Figure 3.3 Classification of glycoside hydrolases (top) and auxiliary activity enzymes (bottom; AA) detected in the *Exophiala* sp. ETNP2018 genome within the ‘Other’ category described in Figure 3.2.

Figure 3.4 Reconstructed metabolic pathway for nitrogen in *Exophiala* sp. ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA.

Figure 3.5 Reconstructed metabolic pathway for starch/glucose in *Exophiala* sp. ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA.

Figure 3.6 Reconstructed mitogen-activated protein kinase (MAPK) high osmolarity glycerol (HOG) pathway in *Exophiala* sp. ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA.

Figure 3.7 Reconstructed xylulose monophosphate pathway (formaldehyde assimilation) in *Exophiala* sp. ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA.

Figure 3.8 Classification of transporter protein families within the *Exophiala* sp. ETNP2018 genome.

Figure A.1 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 19. Measurements were taken on 29 December 2021.

Figure A.2 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 35. Measurements were taken on 7 January 2022.
Figure A.3 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 41. Measurements were taken on 11 January 2022.
LIST OF ABBREVIATIONS

CE .......................................................................................... Carbohydrate esterase
ETNP................................................................................. Eastern Tropical North Pacific Ocean
GH......................................................................................... Glycoside hydrolase
OD.......................................................................................... Optical density
OMZ...................................................................................... Oxygen minimum zone
PFAM.................................................................................. Protein family
PL....................................................................................... Polysaccharide lyase
YEP ..................................................................................... Yeast extract and peptone
CHAPTER 1

OPTIMIZING A DNA EXTRACTION PROTOCOL TO EXTRACT HIGH MOLECULAR WEIGHT GENOMIC DNA FROM MARINE FUNGI ISOLATED FROM SALT MARSH SEDIMENTS

1.1 INTRODUCTION

Long-read sequencing has revolutionized genetic research through the retrieval of highly contiguous assemblies, which allows for better resolution of repetitive genomes and for the detection of structural variants. However, successful reads rely on the extraction of high-quality genomic DNA (gDNA), which has proven to be a challenge when studying filamentous fungal species (Lee et al., 1988). Filamentous fungi often contain multi-layered cell walls and high concentrations of polysaccharides, which bind to gDNA. The disruption of the cell wall often leads to fragmented gDNA, which is less suitable for long-read sequencing (Yu et al., 2004). As marine fungi attract attention for their role in biogeochemical cycling, the need for reproducible high-molecular weight gDNA extraction methods becomes important. We begin the process of optimizing a DNA extraction method for filamentous fungi by comparing various commercial kits. After selecting one kit that obtains high quality DNA without the use of the dangerous chemicals that are commonly used in manual DNA extraction methods, we adjust some of the steps in an effort to optimize the protocol for extraction of high-molecular weight
DNA. We present an optimized method, adapted from the Qiagen DNeasy Plant Mini Kit, that yields high quality, high-molecular weight DNA from an isolated strain of *Rhodotorula sp.*, a pigmented basidiomycetous yeast isolated from salt marsh sediments.

1.2 METHODOLOGY

Yeast extract and peptone (YEP) are common supplements to complex medium used to grow microbial cultures. 1 L of YEP medium was prepared using the following protocol:

1. 1 L of seawater base was added to a 2 L flask.

2. The flask was placed on a magnetic stirring plate with a stir bar.

3. The following components were added to the flask:

   10 mL of MOPs buffer (1M, pH = 7.2)

   10 mL of NH$_4$Cl (100x)

   10 mL of Na$_2$SO$_4$ (100x)

   10 mL phosphate (100x)

   1 mL of trace metal solution

   10 mL of glucose stock solution

   2.5 g yeast extract

   2.5 g Ultrapure bacteriological peptone
4. The solution was placed into multiple 200 mL bottles and autoclaved on a liquid cycle (20 minutes at 121°C).

Samples were collected at Clambank Creek, a research site in the North Inlet-Winyah Bay National Estuarine Research Reserve System. *Rhodotoroula sp.* was isolated, identified, and inoculated into 10 mL culture tubes supplemented with 5 mL of YEP media, 0.2 mL of 10 mM penicillin-streptomycin solution and 0.2 mL of 10 mM sodium nitrite (NaNO₂) solution.

Cultures were incubated in dark at room temperature (20 °C) for 24 hours and harvested through centrifugation at 12,000g for 2 minutes and 30 seconds. The optical density (OD) at 600 nm of each sample was measured as a proxy for cell growth using a NanoDrop ND-2000 spectrophotometer. Samples were transferred to MN Type A Bead Beating Tubes (cat. No. 740786.50, supplied with 0.6-0.8 mm ceramic beads) using a pipette, then were lysed via the Biospec Mini-Beadbeater-16 bead beating machine. DNA extractions were carried out using the Qiagen DNeasy Plant Mini Kit protocol (unless otherwise stated) with bead beating times, lysis buffer volume, and lysis temperature and duration being changed according to each experiment. The purity and concentration of the DNA was determined using a NanoDrop ND-2000 spectrophotometer via the calculation of A260/A280 and A260/A230 ratios. DNA fragment size and integrity were analyzed using the Agilent TapeStation 4150 system, supplied with Agilent Genomic DNA ScreenTape.

Buffer solutions are commonly used in DNA extraction protocols. Tris-hydroxymethyl aminomethane, or tris, is a buffer used in combination with
ethylenediaminetetraacetic acid (EDTA) to maintain a stable pH during the cell lysis process. 100 mL of 1M Tris-HCl buffer, adjusted to have a pH of 8, was prepared using the following protocol:

1. 121.1 g of Tris base was added to a 1 L beaker.
2. 800 mL of distilled water was added.
3. The beaker was placed on a magnetic stirring plate with a stir bar.
4. A pH meter was placed into the solution to observe the pH.
5. Concentrated (6M) hydrochloric acid (HCl) solution was slowly added using a pipette to reduce the pH to 8.0.
6. Distilled water was added to the solution until the total volume reached 1 L.
7. The solution was put into a 1 L bottle and autoclaved on a liquid cycle (20 minutes at 15 psi).

EDTA is a commonly used chelating agent and is necessary during cell lysis to protect DNA from degradation. A solution of 0.5M EDTA, adjusted to have a pH of 8.0, was prepared using the following protocol:

1. 18.61 g of EDTA disodium salt, dihydrate was added to a 100 mL flask.
2. 80 mL of distilled water was added to the flask.
3. The flask was placed on a magnetic stirring plate with a stir bar.
4. A pH meter was placed into the solution to observe the pH.
5. Sodium hydroxide (NaOH) pellets were added to the solution, 1-2 pellets at a time, to adjust the pH to 8.0.

6. 100 mL of distilled water was added to the solution once the EDTA was fully dissolved.

7. The solution was autoclaved on a liquid cycle (20 minutes at 15 psi).

We hypothesized that a higher volume of Buffer AP1 would protect the fungal biomass from shearing during the bead beating process, resulting in longer DNA fragments. The first experiment consisted of lysing re-suspended fungal matter (distributed into 4 tubes, each approximately 37.5 mg in mass) with 1.2, 1, 0.55, and 0.45 mL of Buffer AP1. After bead beating, only 450 µL of buffer/fungus mixture was used in the lysis step for each sample.

The next experiment consisted of varying bead beating times between samples. 1 mL of Buffer AP1 was added to re-suspended fungal matter (distributed into 3 tubes, each approximately 37.5 mg in mass) before distributing 450 µL of the mixture into bead beating tubes. Samples were placed into the bead beater for 30, 45, and 60 seconds respectively.

We hypothesized that the Qiagen DNeasy Plant Mini Kit would yield the highest quality DNA due to its inclusion of multiple washing steps, which remove unwanted polysaccharides and other compounds. Water was added to a centrifuged fungal pellet for resuspension. 200 µL of the mixture was distributed into bead beating tubes. Three kits were used in the following experiment: the Qiagen DNeasy Plant Mini Kit, the Qiagen DNeasy PowerSoil Pro Kit, and the Zygem Quick-DNA Fungal Bacteria MiniPrep Kit.
Each replicate was vortexed using a Vortex Adapter for 1.5-2 mL tubes for 5 minutes before being extracted following each respective protocol.

The next two experiments dealt with the lysis step of the DNA extraction protocol. We hypothesized that carrying out the lysis step at room temperature, as opposed to 65°C, as recommended by the Qiagen DNeasy Plant Mini Kit protocol, would reduce DNA shearing and result in longer DNA fragments. With the remaining water/fungal mixture from the kit experiment, lysis temperature was tested using the Qiagen DNeasy Plant Mini Kit. The lysis step was initiated for both samples by vortexing them using a Vortex Adapter for 1.5-2 mL tubes for 5 minutes. One sample was incubated at 65°C for 10 minutes, while the other was incubated at room temperature for 10 minutes.

Finally, we hypothesized that replacing the bead beating step in the Qiagen DNeasy Plant Mini Kit with vortexing would reduce DNA shearing and yield longer and higher quality DNA fragments. The fungal pellet was diluted with 1800 µL of water. 200 µL of the fungi/water mixture was added to bead beating tubes. 800 µL of buffer AP1 was added to each tube. The samples were then vortexed for the following durations: 5 seconds, 10 seconds, 30 seconds, and 40 seconds. The rest of the Qiagen DNeasy Plant Mini Kit protocol was followed as normal.

1.3 RESULTS

Increased lysis buffer volume (exceeding 1 mL) led to significantly lower concentrations of DNA being recovered (Table 1.1). However, increased lysis buffer
volumes also led to longer DNA fragments being recovered. Fragment lengths ranged from approximately 8,000 to 26,000 base pairs (Table 1.1, Figure 1.1).

Table 1.1 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA extracted using different lysis buffer volumes

<table>
<thead>
<tr>
<th>Lysis Buffer (Buffer AP1) Volume (mL)</th>
<th>DNA concentration (ng/μL)</th>
<th>260/280</th>
<th>260/230</th>
<th>Size (bp) of highest % integrated area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>15.8</td>
<td>1.87</td>
<td>2.73</td>
<td>10,939</td>
</tr>
<tr>
<td>1</td>
<td>21.5</td>
<td>1.92</td>
<td>6.93</td>
<td>9,615</td>
</tr>
<tr>
<td>0.550</td>
<td>48.9</td>
<td>1.85</td>
<td>2.86</td>
<td>8,610</td>
</tr>
<tr>
<td>0.450</td>
<td>58.4</td>
<td>1.86</td>
<td>3.39</td>
<td>8,476</td>
</tr>
</tbody>
</table>

Figure 1.1 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted using various volumes of Buffer AP1

When using the Biospec Mini-Beadbeater-16 bead beating machine, there was not a clear association between bead beating time and the length of extracted DNA
fragments. However, the longest DNA fragments were recovered when the samples were beaten for 45 seconds (Table 1.2, Figure 1.2).

Table 1.2 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA extracted using different bead beating times

<table>
<thead>
<tr>
<th>Bead beating time (s)</th>
<th>DNA concentration (ng/μL)</th>
<th>260/280</th>
<th>260/230</th>
<th>Size (bp) of highest % integrated area</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>6.8</td>
<td>2.20</td>
<td>3.39</td>
<td>11,735</td>
</tr>
<tr>
<td>45</td>
<td>6.8</td>
<td>2.30</td>
<td>-1.46</td>
<td>12,046</td>
</tr>
<tr>
<td>60</td>
<td>11</td>
<td>1.88</td>
<td>-1.36</td>
<td>10,345</td>
</tr>
</tbody>
</table>

Figure 1.2 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted with the various bead beating times

The QuickDNA Fungal/Bacteria Mini Kit yielded the shortest extracted DNA fragments, while the DNeasy Plant Mini Kit and the DNeasy PowerSoil Kit yielded fragment sizes more comparable to one another (Table 1.3).
Table 1.3 DNA concentration, 260/280, and 260/230 values, as well as the fragment length of the highest % integrated area for genomic DNA samples (with one replicate each) extracted using different kits

<table>
<thead>
<tr>
<th>Sample ID (kit used)</th>
<th>DNA concentration (ng/μL)</th>
<th>260/280</th>
<th>260/230</th>
<th>Size (bp) of highest % integrated area</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (Dneasy Plant Mini Kit)</td>
<td>263.9</td>
<td>1.88</td>
<td>2.45</td>
<td>7,983</td>
</tr>
<tr>
<td>P2 (Dneasy Plant Mini Kit)</td>
<td>247.7</td>
<td>1.87</td>
<td>2.51</td>
<td>8,862</td>
</tr>
<tr>
<td>Z1 (QuickDNA Fungal/Bacteria Mini Kit)</td>
<td>138.1</td>
<td>1.91</td>
<td>0.18</td>
<td>5,267</td>
</tr>
<tr>
<td>Z2 (QuickDNA Fungal/Bacteria Mini Kit)</td>
<td>91.0</td>
<td>1.93</td>
<td>1.30</td>
<td>5,427</td>
</tr>
<tr>
<td>S1 (Dneasy PowerSoil Kit)</td>
<td>308.6</td>
<td>1.87</td>
<td>1.60</td>
<td>10,343</td>
</tr>
<tr>
<td>S2 (Dneasy PowerSoil Kit)</td>
<td>258.7</td>
<td>1.86</td>
<td>0.86</td>
<td>7,919</td>
</tr>
</tbody>
</table>

Figure 1.3 TapeStation gel image showing the ladder (leftmost), alongside the DNA fragments extracted from the corresponding kits
Table 1.4 Comparison of lysis temperature and bead beating time for DNA extracted using the DNeasy Plant Mini Kit

<table>
<thead>
<tr>
<th>Lysis temperature</th>
<th>Bead beating time (min)</th>
<th>DNA concentration (ng/μL)</th>
<th>260/280</th>
<th>260/230</th>
<th>Size (bp) of highest % integrated area</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>5</td>
<td>263.9</td>
<td>1.88</td>
<td>2.45</td>
<td>7,983</td>
</tr>
<tr>
<td>65</td>
<td>5</td>
<td>247.7</td>
<td>1.87</td>
<td>2.51</td>
<td>8,862</td>
</tr>
<tr>
<td>65</td>
<td>1</td>
<td>272.3</td>
<td>1.87</td>
<td>2.47</td>
<td>8,194</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>207.0</td>
<td>1.88</td>
<td>2.55</td>
<td>7,771</td>
</tr>
</tbody>
</table>

For cultures grown for 24 hours before being harvested via centrifugation, vortexing rather than bead beating produced longer DNA fragments.

Table 1.5 Comparison of various vortexing times for DNA extracted using the DNeasy Plant Mini Kit

<table>
<thead>
<tr>
<th>Vortexing time (s)</th>
<th>DNA concentration (ng/μL)</th>
<th>260/280</th>
<th>260/230</th>
<th>Size (bp) of highest % integrated area</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>138.4</td>
<td>1.88</td>
<td>2.79</td>
<td>54,423</td>
</tr>
<tr>
<td>10</td>
<td>123.8</td>
<td>1.87</td>
<td>2.84</td>
<td>50,223</td>
</tr>
<tr>
<td>30</td>
<td>240.6</td>
<td>1.86</td>
<td>2.61</td>
<td>58,650</td>
</tr>
<tr>
<td>45</td>
<td>167.9</td>
<td>1.86</td>
<td>2.90</td>
<td>56,945</td>
</tr>
</tbody>
</table>
1.4 DISCUSSION

The success of high-molecular weight DNA extraction relies on the efficient isolation of genomic DNA from the cell. The breakdown of the fungal cell wall by lysing is a crucial step during extraction (Rodrigues et al., 2018; Scharf et al., 2020). The fungal cell wall often consists of chitin, lipids, peptides, and other compounds that are resistant to enzymatic digestion and chemical breakdown, which can contaminate genomic DNA products (Karakousis et al., 2006). Fungal cell walls are often compact, making good quality fungal DNA hard to obtain. Various cell disruption processes have been tested for efficiency in fungal DNA extraction, such as heating, bead milling using grinding homogenizers, and grinding in liquid nitrogen (Turan et al., 2015; Karakousis et al., 2006). However, these methods vary in success and have their own setbacks and
difficulties; for example, grinding with mortar and pestle in liquid nitrogen is generally accepted as the most efficient method for fungal cell wall disruption, but often leads to sample loss and contamination (Karakousis et al., 2006).

Through testing lysis temperature, lysis time, amount of lysis buffer, as well as using various DNA extraction kits, we optimized a protocol for extracting high molecular weight genomic DNA from three marine fungi species, adapted from the Qiagen DNeasy Plant Mini Kit. The incorporation of vortexing instead of bead-beating at the end of the lysis step allows the buffer to break down the thick cell wall without shearing the DNA, producing longer fragments of DNA that are suitable for long-read sequencing technologies. We recommend using this optimized gDNA extraction protocol, which involves vortexing samples for 30 seconds rather than bead beating in order to reduce DNA shearing, as well as diluting samples with 1.2 mL of Buffer AP1 to further protect DNA from shearing, in order to obtain high-molecular weight DNA from marine fungi. This protocol provides a relatively safe and easy alternative to other high-molecular weight DNA extraction protocols described in literature, which often include the use of cetyltrimethylammonium bromide (CTAB), a known carcinogen used for the removal of polysaccharides from genomic DNA, as well as phenol and chloroform, both of which are considered hazardous materials (Zhang et al., 2010). More replicates are necessary to determine whether this extraction protocol is suitable for extracting high-molecular weight DNA from *Rhodotorula sp.* In addition, while this method may work for this particular strain of *Rhodotorula*, additional experiments with other filamentous fungal strains would be necessary to determine whether this protocol works with strains that have different cell wall structures or polysaccharide compositions.
CHAPTER 2
IDENTIFICATION OF MARINE FUNGI USING MARKER GENES

2.1 INTRODUCTION

Oxygen minimum zones (OMZs) are oxygen gas (O$_2$) deficient layers in the water column. Thermal stratification, low circulation, upwelling, and heterotrophic consumption of primary producers are all potential drivers of oxygen deficiencies in the water column (Vik et al., 2020). These regions are notable for fulfilling a unique role in the oceanic nitrogen cycle. Microbial communities are responsible for the denitrification of nitrate (NO$_3^-$), a limiting nutrient in the open ocean, into nitrous gas (N$_2$), resulting in a significant loss of nitrogen to the atmosphere – some estimates exceeding 30% of nitrogen loss from the ocean, and 16% of nitrogen loss from terrestrial and oceanic ecosystems combined (Lam & Kuypers, 2011). OMZs, both in sediments and within the water column, have also been associated with the anaerobic oxidation of ammonium (NH$_4^+$) via anammox (NO$_2$) (Paulmier & Ruiz-Pino, 2009). OMZs, like salt marshes, are associated with the production of N$_2$O, a byproduct of anaerobic denitrification (Peng & Valentine, 2021).

The Eastern Tropical North Pacific, or the ETNP, accounts for approximately half of the total area of OMZs (Faull et al., 2020). Various publications on the microbial ecology in the ETNP, as well as other OMZs, have focused on bacteria and archaea rather than microbial eukaryotes (Duret et al., 2015). Despite this, marine fungi are thought to
play a role in organic matter recycling in the water column (Amend et al., 2019). Furthermore, incubation experiments with ETNP fungal strains revealed that fungal production of N₂O gas accounted for approximately 20% of total N₂O production, suggesting that fungal denitrification contributes to N₂O production in OMZs (Peng & Valentine, 2021).

The identification and classification of members of the fungal kingdom has proven to be a bottleneck for fungal research (Banos et al., 2018). Taxon identification of fungi through the sequencing of marker genes relies on the availability of reference sequences, suitable primer sets, and resolution power that spans across different fungal groups (Reich & Labes, 2017). Identification through the targeting of the internal transcribed spacer (ITS) region has a broader range of species resolution of fungi compared to other marker genes; however, a lack of reference sequences often limits identification to higher taxonomic levels (Nilsson et al., 2016). Other fungal phylogenetic markers include the large and small ribosomal RNA gene sequences, the latter of which having more publicly available reference sequences from fungal taxa (Panzer et al., 2015).

In December 2021, the R/V Sally Ride departed from Puerto Caldera, Costa Rica, travelling to various stations in the ETNP oxygen minimum zone. Three stations were visited to study fungal diversity, resulting in the compilation of isolated strains in a fungal culture inventory. Genomic DNA isolated from various marine fungal strains from the culture inventory was amplified using different primer sets to aid in their identification, and to investigate the size of fragments produced by primer sets designed to amplify fungi-specific 18S rRNA regions.
2.2 METHODOLOGY

Aboard the R/V Sally Ride, at three stations located at 13.227°N, 90.730°W, 12.09°N, 90.130°W, and 14.24°N, 104.94°W (stations S19, S35, and S41, respectively) (Figure 2.1), seawater was collected at multiple depths using 30 L Niskin bottles.

Figure 2.1 Sampling locations in the ETNP oxygen minimum zone. Colormap shows oxygen concentration at 100 meters depth from World Ocean Atlas 2018 (yearly average from 2018)

Fungal strains isolated and cultured from the cruise were entered into an ETNP fungal culture inventory. Samples from said inventory (Table 2.1) were extracted using the Qiagen DNeasy Plant Mini Kit DNA. We expected that the genomic DNA extracted from the ETNP fungal culture inventory would produce DNA fragments of about 700 base pairs in length when amplified using the ITS5 and ITS4 primer set (5’-GCAAGTAAAAGTCGTAACAAGG, 5’-TCCTCCGCTTATTGATATGC, respectively) (White et al., 1990). We also expected that genomic DNA extracted from fungal cultures
would produce DNA fragments about 500 base pairs in length when amplified using the Fun18S1 and LR5 primer set (5’-CCATGCATGTCTAAGTWAA, 5’-TCCTGAGGGAAACTTCG, respectively) which targets the fungi-specific 18S ribosomal RNA region (Banos et al., 2018). Various samples from the ETNP culture inventory were extracted using the Qiagen DNeasy Plant Mini Kit (Table 2.1) and subsequently amplified using the following components: 4 µL Thermo Scientific 5X Phusion Buffer, 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 0.4 µL 10 mM dNTPs, and 0.2 µL Thermo Scientific Phusion Plus Polymerase.

Table 2.1 Display of the various samples used in experiments, their dates and locations of sampling, and their denotation in this document

<table>
<thead>
<tr>
<th>Sample Full Name</th>
<th>Identification/Culture Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-R1-G1 An S19-E1 250m From Rd3 2022.6.7</td>
<td>M15</td>
</tr>
<tr>
<td>O1-R2 S35-E2 125m From Rd2 White 2022.6.6</td>
<td>M3</td>
</tr>
<tr>
<td>O3-R2-G0 An S35-E2 350m From Rd4 2022.6.13</td>
<td>M36</td>
</tr>
<tr>
<td>O2-R1 S41 150m From Rd4 Pink 2022.6.7</td>
<td>M11</td>
</tr>
<tr>
<td>O3-R2 S35-E2 350m From Rd3 Pink 2022.7.7</td>
<td>M47</td>
</tr>
<tr>
<td>O1-R3 S35-E2 125m From Rd3 2022.6.3</td>
<td>M2</td>
</tr>
<tr>
<td>O3-R1 S41 200m From Rd3 2022.5.31</td>
<td>M30</td>
</tr>
<tr>
<td>O1-R2 S35-E2 125m From Rd3 White 2022.6.7</td>
<td>N1</td>
</tr>
<tr>
<td>O2-R3-G0 An S35-E2 140m From Rd2 2022.6.7</td>
<td>6</td>
</tr>
</tbody>
</table>

Samples were loaded into strip tubes with 6.6 µL of master mix (components listed above), 12.4 µL of autoclaved Millipore Milli-Q water, and 1 µL of DNA (depending on the amount of DNA present in the sample, sometimes more DNA was used with less water). Each PCR run included a negative control of just water. Samples were amplified in a thermocycler, with annealing temperatures depending on the specific primers used. To analyze the size of targeted DNA fragments, the PCR products were then run on 2% agarose gels alongside a 1 kb ladder. Gels were imaged using the Gel Doc EZ System’s Image Lab Software. Specific bands were excised from UV-
illuminated gels, processed with the Zymo Research Gel DNA Recovery Kit, and sent to Eton Bioscience for Sanger sequencing. Sequences were analyzed using NCBI’s Basic Local Alignment Search Tool (BLAST).

2.3 RESULTS

Samples 6, N1, M2, M30, and M36 were amplified using the universal primers 515F-Y and 926-R, which target the V4-V5 region of 16S SSU rRNA. The samples all produced bands that signify smaller fragments than were expected for the fungal V4-V5 regions. Samples 6, N1, and M36 produced fragments of about 500 bp in length. Samples M2 and M30 produced fragments of about 600 bp in length. Samples M2, M30, and M47 produced fragments of about 600 bp in length.

Figure 2.2 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): 6, N1, M2, M30, M36, and a negative control

Two replicates of sample M36, M15, two replicates of sample M2, two replicates of sample M30, and M47 were amplified using the universal primers 515F-Y and 926.
Sample M36 produced fragments of about 400 bp in length. Samples M2, M30, and M47 produced fragments of about 600 bp in length.

Figure 2.3 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): M36, M36 (R2), M15, M2, M30, M2 (R2), M15 (R2), M30 (R2), M47, and a negative control.

Sample M15 (a white, filamentous colony collected from the ETNP on June 7th, 2022), was extracted again and diluted by a factor of 1:10. One sample of the diluted M15, one of the concentrated M15, and one of M11 were amplified using the universal primers 515F-Y and 926-R and imaged using gel electrophoresis. Both M15 samples produced fragments of about 500 bp in length, as well as faint fragments of about 400 bp in length. M11 produced fragments of about 600 bp in length as well as fragments of about 800 bp in length.
Figure 2.4 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the universal primer set (515F-Y/926-R): diluted M15, concentrated M15, M11, and a negative control.

Three samples of M15, extracted on different dates using the Qiagen DNeasy Plant Mini Kit, as well as M36, were diluted by a factor of 1:10 and amplified. M3, a known strain of *Emericellopsis maritima* was included as a positive control. These samples were amplified using primers designed for the ITS3, ITS4, and ITS5 regions. They were also amplified using the LR0R/LR5 primer set (5'-ACCCGCTGAACCTTAAGC, 5'-TCCTGAGGGAAACTTCG, respectively), which is designed to amplify the 28S LSU rRNA, and two primer sets designed to amplify fungi-specific rRNA 18S regions, FF390/FR1 and Fun18S1/FR1 (Tedersoo et al., 2015). M15, M36, and M3 produced bands of about 600 base pairs in length when amplified with the ITS5/ITS4 primer set. The ITS3/ITS4 primer set produced no bands amongst any of the samples. M3 and one sample of M15 produced bands of about 750 base pairs in length when amplified with the LR0R/LR5 primer set. Samples M3 and the same sample of M15 produced larger bands (approximately 900-1000 base pairs in length) when amplified using the Fun18S1/FR1 primer set. M15 was sent to Eton Biosciences Inc. for
Sanger sequencing. Identification through NCBI’s BLAST revealed that M15 was likely contaminated by the positive control.

Figure 2.5 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the corresponding primer sets (labeled in red): three M15 samples, M36, M3 (positive control), and a negative control.

Two replicates of M15 were amplified alongside positive and negative controls using the universal primer set, the LR0R/LR5 primer set, and the ITS5/ITS4 primer set. The second replicate of M15 produced fragments of about 750 bp in length with both the LR0R/LR5 and the ITS5/ITS4 primer sets, without contamination from the positive control. These two PCR products were sent to Eton Biosciences Inc. for Sanger Sequencing. The strain amplified using the LR0R/LR5 primer set was identified as *Aureobasidium sp.*, and the strain amplified using the ITS primer set was identified as *Pseudomonas putida*, a gram-negative bacterium.
2.4 DISCUSSION

A majority of the samples (N1, M47, 6, M2, M30), thought to be novel strains of marine fungi, produced PCR products for 16S instead of the expected 18S fragments when amplified with universal primers. Sample M36, previously thought to be an isolated strain of marine fungi, did not amplify when using the LR0R/LR5 and the Fun18S1/FR1 primer sets, indicating the DNA product is not from a fungus. One replicate of sample M15 was successfully amplified using both the LR0R/LR5 and the ITS5/ITS4 primer sets, producing bands of about 750 base pairs in length and without visible contamination in the positive or negative controls (Figure 2.5). When sequenced, however, both replicates were identified as strains of bacteria. Another replicate of sample M15 was successfully amplified using the Fun18S1/FR1 primer set (Figure 2.4), signifying that the DNA contained sequences unique to fungi. However, when amplified using the Fun18S1/FR1 primer set alongside a positive control (*Emericellopsis maritima*; M3) and a negative control, the negative control also produced a band, signifying contamination of

Figure 2.6 Gel electrophoresis image showing the 1 kilo bp ladder (leftmost), alongside samples amplified using the corresponding primer sets (labeled in red): M3 (positive control), M15 (R1), M15 (R2), and a negative control.
the DNA product. This sample appeared to be contaminated by the positive control after being sequenced and identified using NCBI’s BLAST (Altschul et al., 1990; Sayers et al., 2022). Thus, it is evident that Sample M15 is not a pure fungal isolate and is instead a mixed culture of bacteria and fungi. Before Sample M5 can be sent for sequencing and identification, the culture must be further purified to rid it of any bacterial growth. Due to time constrains, we decided to analyze the genome of *Exophiala sp.*, which had already been isolated and sequenced.
CHAPTER 3

THE GENOME OF *EXOPHIALA SP*. ETNP2018, ISOLATED FROM THE EASTERN TROPICAL NORTH PACIFIC OCEAN

3.1 INTRODUCTION

*Exophiala*, an anamorph genus characterized by the production of heads of conidia, comprises of over forty identified species, with various members being associated with pathogenicity in both human and cold-blooded animal hosts (Shoff & Perfect, 2021). Animals that have a marine stage in their life cycle, or otherwise have smooth, moist skin, are especially susceptible to infections by *Exophiala* species (de Hoog et al., 2011). Most studies on members of the *Exophiala* genus concern species that are either pathogenic to humans or that have been isolated from various fish hatcheries (Saraiva et al., 2019). *Exophiala sp*. ETNP2018, however, is a novel strain isolated from water samples from the eastern tropical North Pacific (ETNP) in 2018. This strain was isolated, cultured, and sequenced in order to investigate its adaptations to life in the water column.

*Exophiala sp*. ETNP2018 is a novel strain isolated from water samples from the eastern tropical North Pacific (ETNP) in 2018. This strain was isolated, cultured, and sequenced in order to investigate its adaptations to life in the water column. Preliminary phylogenetic analysis reveals that this strain is more closely related to *Exophiala*
aquamarina, a pathogenic strain known to cause systemic infections in fish, than to strains associated with human pathogenicity, such as *Exophiala dermatitidis* and *Exophiala spinefera* (Harris et al., 2009; Teixeira et al., 2017).

3.2 METHODOLOGY

*Exophiala sp.* ETNP2018 was isolated from a minimal seawater-based media supplemented with laminarin. The inoculum originated from the ETNP at 40 meters depth (Figure 3.1), collected on March 18th, 2018, at 23:50:23 UTC.

![Figure 3.1 Sampling location of Exophiala sp. ETNP2018. Colormap shows oxygen concentration at 40 meters depth from World Ocean Atlas 2018 (yearly average from 2018)](image)

The genome assembly of *Exophiala sp.* ETNP2018 was analyzed and annotated using Funannotate 1.8.14 (Palmer & Stajich, 2019). Repetitive contiguous sequences, or contigs, were removed. The assembly was then sorted by fragment lengths and masked. Proteins were predicted using Augustus, a program designed to predict genes in eukaryotic genome sequences (Keller et al., 2011). The protein sequences annotated (with
a minimum contig length of 1,500 base pairs) via Funannotate’s annotation function were submitted to BlastKOALA, a server which performs ortholog assignments according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Sequences were also submitted to EggNOG-mapper, a tool that uses orthologous groups and phylogenies from the EggNOG v5.0 database (Huerta-Cepas et al., 2019). Secondary metabolites biosynthesis gene clusters were mapped using antiSMASH v6.0 (Blin et al., 2021). Transporter proteins were detected using a BLAST analysis alongside the Transporter Classification (TC) Database (Saier et al., 2016). Carbohydrate-active enzymes (CAZymes) were annotated using dbCAN2, a web server for automated CAZyme annotation (Zhang et al., 2018). The dbCAN2 server filter was set to use the three available tools: HMMR, DIAMOND, and dbCAN_sub. Results were imported into Excel and filtered so that only the results found by at least two of the three tools remained.

Genome characteristics of *Exophiala sp.* ETNP2018 were compared to those of *Exophiala aquarmarina*, a representative genome of the *Exophiala* genus. The information available in the Jane Goodall Institute (JGI) Genome Portal database, including genome size, GC content, and metabolic diversity (such as CAZymes and biosynthetic gene clusters), was used to directly compare to *Exophiala sp.* ETNP2018.

### 3.3 GENOMIC ANALYSIS OF *EXOPHIALA SP.* ETNP2018

Our assembly of the genome of *Exophiala sp.* ETNP2018 is 40,300,402 base pairs in length, with 168 contiguous sequences (contigs) and a GC-content of 49.48% (Table 3.1). There are approximately 14,277 genes within the genome, with an average length of 1,602 base pairs (Table 3.1).
Table 3.1 Genomic characteristics of the *Exophiala* sp. ETNP2018

<table>
<thead>
<tr>
<th>Number of contigs</th>
<th>N50</th>
<th>N90</th>
<th>Number of genes</th>
<th>Average gene length (bp)</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>168</td>
<td>2,074,523</td>
<td>535,603</td>
<td>14,277</td>
<td>1,602</td>
<td>49.48</td>
</tr>
</tbody>
</table>

Carbohydrate-active enzymes, or CAZymes, control the diversity of the complex carbohydrates used by cells to mediate various biological functions, including carbon storage and intracellular recognition (Lombard et al., 2014). This suite of enzymes is responsible for the synthesis, modification, and degradation of these compounds. The publicly available CAZyme database (CAZy) describes over 200 experimentally characterized CAZyme families, including members of the glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE), and carbohydrate-binding molecule (CBM) families (Drula et al., 2022). The *Exophiala* sp. ETNP2018 genome displays a number of genes associated with the production of CAZymes from various families described across literature and often characterized as cellulases, hemicellulases, pectinases, and esterases, depending on the structures they act upon (Table 3.2, Figure 3.2).

Table 3.2 Classification of glycoside hydrolases, polysaccharide lyases, and carbohydrate esterases according to catalytic domains

<table>
<thead>
<tr>
<th>Glycoside Hydrolase (GH)</th>
<th>Polysaccharide Lyase (PL)</th>
<th>Carbohydrate Esterase (CE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cellulase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 6, 7, 8, 9, 12, 44, 48, 51, 74, 124</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hemicellulase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5, 8, 10, 11, 16, 26, 30, 43, 44, 51, 62, 98, 141</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pectinase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1, 2, 3, 9, 10</td>
<td></td>
</tr>
<tr>
<td><strong>Esterase</strong></td>
<td></td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 12, 13, 15</td>
</tr>
</tbody>
</table>
Figure 3.2 Classification of CAZymes detected in *Exophiala sp.* ETNP2018, according to Table 3.2
Glycoside hydrolase 16 (GH16) is a large, diverse family involved in the degradation and production of cell wall polysaccharides in both marine and terrestrial organisms. GH16 is separated into various subfamilies based on a large-scale sequence-specific nuclease (SSN) analysis of the GH16 sequence space (Viborg et al., 2019). The genome of *Exophiala sp.* ETNP2018 contains members of various GH16 subfamilies associated with fungal genomes, such as GH16_1, GH16_2, GH16_18, and GH16_19.
The largest GH16 subfamily, GH6_1, consists of almost exclusively fungal enzymes. GH16_2, similarly, includes enzymes reported in fungi, as well as some members associated with plant-damaging oomycetes such as water molds. The subfamilies GH16_18 and GH16_19 consist of fungal enzymes associated with chitin β(1,3)/β(1,6)-glucosyltransferases (Viborg et al., 2019). The genome of *Exophiala* sp. ETNP2018 also includes occurrences of the GH16_3, a sequence-diverse subfamily associated with Endo-β(1,3)-glucanase and/or endo-β(1,3)/β(1,4)-glucanase activity (Viborg et al., 2019).

Glycoside hydrolase 72 (GH72) is a part of a larger suite of GH families associated with the hyphal elongation of filamentous fungi (Ao & Free, 2017). These enzyme families, including GH16 (β-1,3-glucanases), GH17 (β-1,3-glucanases), GH72(β-1,3-glucanases), GH18 (chitinases), and GH76 (α-1,6-mannanases) are all present in the genome of *Exophiala* sp. ETNP2018 (Supplementary Figure S1). GH72 consists of transglycosylases that are able to change the length of β-1,3-glucanases, assisting in the insertion of β-1,3-glucan into the cell wall (Mazáň et al., 2011). GH17 and GH18 are believed to be associated with the cross-linking of glucan and chitin polymers (Gastebois et al., 2010). GH76 is associated with the cross-linking of cell wall glycoprotein-associated galactomannans and outer chain mannans with the cell wall matrix (Solanki et al., 2022). The five glycoside hydrolase families represent the cross-linking necessary to create cell wall matrices in fungi (Ao & Free, 2017).

Glycoside hydrolase family 5 (GH5), known as the ‘cellulase family A,’ is another large GH family with various described subfamilies detected in the *Exophiala* sp. ETNP2018 genome. GH5_5, one such subfamily, is composed of secreted fungal and bacterial enzymes that are associated with endo-β-1,4-glucanase activity, though the
Exophiala sp. ETNP2018 genome lacks the modular protein family CBM1 that often accompanies GH5_5 endo-β-1,4-glucanases in cellulose degradation (Aspeborg et al., 2012). GH5_15, a subfamily associated with β-1,6-glucanase activity, mycoparasitic activity, and cell wall recycling, is also present in the Exophiala sp. ETNP2018 genome (Aspeborg et al., 2012).

The genome of Exophiala sp. ETNP2018 contains enzyme-encoding genes used in the metabolism of nitrogen—specifically, in assimilatory nitrate reduction, the transport of extracellular nitrate and nitrite across the cell membrane, and in the hydrolysis of aromatic nitrile compounds into ammonia (Figure 3.4).

Figure 3.4 Reconstructed metabolic pathways for nitrogen in Exophiala sp. ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA

At first glance, it appears that the dissimilatory nitrate reduction pathways, which includes a suite of enzymes such as NarGHI, NapAB, NirK, NirS, NorBC, and NosZ, is not fully conserved in the Exophiala sp. ETNP2018 genome (Ferguson, 1998). The
process of denitrification was previously thought to be conducted only by prokaryotes through four reducing steps; however, fungi are able to utilize a separate pathway, which has been characterized as the fungal denitrification system. This system consists of a copper-containing nitrite reductase (NirK; E.C. 1.7.2.1), as well as a cytochrome P450 nitric oxide reductase (P450nor; E.C. 1.7.1.14). P450nor has been demonstrated to receive electrons from nicotinamide adenine dinucleotide, reducing nitric oxide (NO) to nitrous oxide (N₂O). The fungal denitrification system, unlike its prokaryotic counterpart, is demonstrated to lack the final step of nitrous oxide reduction, in which nitrous oxide is reduced into dinitrogen (N₂). The Exophiala sp. ETNP2018 genome contains the gene encoding P450nor (Figure 3.4), demonstrating that this gene is capable of reducing nitric oxide to nitrous oxide.

The central carbohydrate metabolic pathways are fully conserved in the Exophiala sp. ETNP2018 genome, including those necessary for conducting glycolysis, pyruvate oxidation, and the citrate cycle. Other carbohydrate metabolic pathways, including the Leloir pathway associated with galactose degradation, glycogen biosynthesis and degradation, and the glyoxylate cycle are fully conserved. Within the starch and glucose metabolic pathway network, only the pathways for glycogen biosynthesis and degradation are fully conserved (Figure 3.5). Other starch and glucose metabolism pathways are partially conserved in the Exophiala sp. ETNP2018 genome, indicating that this strain may metabolize simpler carbohydrates or has another means of obtaining energy (Figure 3.5).
One of the main storage polysaccharides found in brown algae is laminarin, a structure formed by 1,3-β-glucans with β-1,6 linkages (Balabanova et al., 2018). The genome of Exophiala sp. ETNP2018 includes genes that encode enzymes associated with the breakdown of laminarins, including endo-beta-1,3-1,4 glucanase, beta-1,3-glucanase, and beta-1,6-glucanase. Other polymers found in macroalgae that is generally uncommon in terrestrial organisms include fucans/fuciodans and ulvans (Balabanova et al., 2018). The genome of Exophiala sp. ETNP2018 includes a gene associated with the breakdown of fucan, α-L-Fucosidase. However, it lacks any genes associated with the production of enzymes used for the breakdown of substrates such as alginates and oligo alginates, which are found in marine plants.

Marine fungi can produce enzymes needed for salt tolerance, thermostability, and other novel physiological characteristics (Velmurugan & Lee, 2012). However, these enzymes are also produced by their terrestrial relatives, with few studies demonstrating
that they are different from their terrestrial counterparts (Verbist et al., 2000).

Halotolerance in fungi is associated with the mitogen-activated protein kinase (MAPK) high osmolarity cascade. The genome of *Exophiala sp.* ETNP2018 contains the osmosensors involved in activating the MAPK high osmolarity cascade (*Sho1* and *Sln1*), along with their respective cascades (*Sln1-Ypd1-Ssk1-Ssk2-Pbs2-Hog1, Sho1-Cdc42-Ste20* (or *Cla4*-Ste11-Pbs2-Hog1)) (Figure 3.6) that are activated in response to changes in osmolarity. Fungi utilize the HOG-MAPK pathway to reduce osmotic pressure and to prevent water losses through the accumulation of glycerol in the cytoplasm.

Figure 3.6 Reconstructed mitogen-activated protein kinase (MAPK) high osmolarity glycerol (HOG) pathway in *Exophiala sp.* ETNP2018, according to the functional KEGG Orthology annotation data extracted from BlastKOALA

A complete pathway for formaldehyde assimilation via the xylulose monophosphate pathway was conserved in the *Exophiala sp.* ETNP2018 genome (Figure 3.7).
3,942 distinct protein families were identified within the genome of *Exophiala sp.* ETNP2018. PF07690, a major facilitator family, was the most occurring protein family, followed by PF00172, a fungal binuclear cluster domain, and the short chain dehydrogenase family, PF00106, a group of enzymes used in the mediation of biochemical reactions in the respiratory chain pathway (Table 3.3).

Table 3.3 Most commonly occurring protein families/domains, listed alongside their respective descriptions and number of occurrences within the *Exophiala sp.* ETNP2018 genome

<table>
<thead>
<tr>
<th>PFAM ID</th>
<th>Description</th>
<th># of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF07690</td>
<td>Major facilitator superfamily</td>
<td>680</td>
</tr>
<tr>
<td>PF00172</td>
<td>Fungal binuclear cluster domain</td>
<td>557</td>
</tr>
<tr>
<td>PF00106</td>
<td>Short chain dehydrogenase family</td>
<td>359</td>
</tr>
<tr>
<td>PF04082</td>
<td>Fungal-specific transcription factor domain</td>
<td>353</td>
</tr>
<tr>
<td>PF13561</td>
<td>Enoyl-Acyl carrier protein reductase domain</td>
<td>345</td>
</tr>
<tr>
<td>PF08659</td>
<td>KR domain</td>
<td>254</td>
</tr>
<tr>
<td>PF00083</td>
<td>Glucose transporter family</td>
<td>222</td>
</tr>
<tr>
<td>PF00067</td>
<td>Cytochrome P450 domain</td>
<td>173</td>
</tr>
</tbody>
</table>

AntiSMASH identified 5 biosynthetic gene clusters related to secondary metabolite production within the *Exophiala sp.* ETNP 2018 genome assembly. These clusters included non-ribosomal synthetase-like (NRPS-like) cluster 20.1, terpene cluster
1.2, type 1 polyketide synthase (T1PKS) clusters 3.1 and 4.1, and non-ribosomal synthetase (NRPS) cluster 5.1 (Table 3.4).

Table 3.4 Genes encoding biosynthetic gene clusters (BGCs)

<table>
<thead>
<tr>
<th>Region</th>
<th>Type</th>
<th>Most similar known cluster</th>
<th>Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>Terpene</td>
<td>Squalestatin S1</td>
<td>40</td>
</tr>
<tr>
<td>3.1</td>
<td>1,3,6,8-tetrahydroxynaphthalene</td>
<td>T1PKS</td>
<td>100</td>
</tr>
<tr>
<td>4.1</td>
<td>Isoterrein</td>
<td>T1PKS</td>
<td>18</td>
</tr>
<tr>
<td>5.1</td>
<td>Acetylaranotin</td>
<td>NRPS</td>
<td>20</td>
</tr>
<tr>
<td>20.1</td>
<td>Choline</td>
<td>NRPS-like</td>
<td>100</td>
</tr>
</tbody>
</table>

A total of 6,558 genes (~ 46% of total genes) were recognized as orthologous genes based on the KOG database (Table 3.5). Highly represented categories of gene orthologs of known function include amino acid metabolism and transport (713 orthologs), intracellular trafficking and secretion (537), energy production and conversion (465), and lipid metabolism (439).

Table 3.5 Clusters of orthologous groups (COGs) detected in *Exophiala sp. ETNP2018*

<table>
<thead>
<tr>
<th>COG functional classification</th>
<th>Number of found clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoskeleton</td>
<td>106</td>
</tr>
<tr>
<td>Intracellular trafficking and secretion</td>
<td>537</td>
</tr>
<tr>
<td>Cell cycle control and mitosis</td>
<td>92</td>
</tr>
<tr>
<td>Cell wall/membrane biogenesis</td>
<td>91</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>267</td>
</tr>
<tr>
<td>Energy production and conversion</td>
<td>465</td>
</tr>
<tr>
<td>Amino acid metabolism and transport</td>
<td>713</td>
</tr>
<tr>
<td>Inorganic ion transport and metabolism</td>
<td>399</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>439</td>
</tr>
<tr>
<td>Translation</td>
<td>351</td>
</tr>
<tr>
<td>Unknown function</td>
<td>3098</td>
</tr>
</tbody>
</table>

The *Exophiala sp. ETNP2018* genome encodes over 600 transporter proteins (608 genes) identified by the transporter classification database system (TCDB), accounting
for approximately 4% of the total predicted genes (Saier et al., 2021). A vast majority of these transporter proteins are classified into the major facilitator superfamily (MFS; Figure 3.8), which are responsible for membrane transport of small solutes across cell membranes (Pao et al., 1998). This class of transporters is also associated with the growth of fungi under stressful conditions (Xu et al., 2014).

Figure 3.8 Classification of transporter protein families within the *Exophiala sp.* ETNP2018 genome

Transporters, according to the transport protein classification system (TC), are separated into five classes: channels/pores (TC1), electrochemical potential-driven transporters (TC2), primary active transporters (TC3), group translocators (TC4), and transmembrane electron carriers (TC5), with two additional classes to characterize accessory factors involved in transport (TC8) and transporter systems that are not fully defined or characterized (TC9) (Milton et al., 2016). The genome of *Exophiala sp.* ETNP2018 encodes 302 transporters (~50% of all transporters identified in the genome) from TC2, which encompasses the Major Facilitator Superfamily. TC3 and TC4 make up
the next two most abundant transporter classes in *Exophiala sp.* ETNP2018 25% and 19%, respectively. 25 transporter genes were classified into TC9.

*Exophiala aquamarina* CBS 119918 (hereby referred to as *Exophiala aquamarina*) is one of over forty described *Exophiala* species. This waterborne strain, as well as other members of the *Exophiala* genus, are known to cause infections in cold-blooded animals (Saraiva et al., 2019). The genome statistics for both *Exophiala aquamarina* and *Exophiala sp.* ETNP2018 are described in Table 3.6 (Teixeira et al., 2017).

Table 3.6 Overview of gene statistics of *Exophiala sp.* ETNP2018 and *Exophiala aquamarina*

<table>
<thead>
<tr>
<th></th>
<th>Exophiala sp. ETNP2018</th>
<th>Exophiala aquamarina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembled genome size</td>
<td>40,300,402</td>
<td>40,982,715</td>
</tr>
<tr>
<td>N50</td>
<td>2,074,523</td>
<td>2,236,061</td>
</tr>
<tr>
<td>Total number of genes</td>
<td>14,222</td>
<td>13,118</td>
</tr>
<tr>
<td>Average gene length</td>
<td>1,602</td>
<td>1,485</td>
</tr>
<tr>
<td>Biosynthetic gene clusters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terpene</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NRPS-like</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>T1PKS</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CAZymes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auxiliary activities (AA)</td>
<td>82</td>
<td>52</td>
</tr>
<tr>
<td>Glycoside hydrolases (GH)</td>
<td>186</td>
<td>184</td>
</tr>
<tr>
<td>Carbohydrate-binding molecules (CBM)</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Carbohydrate esterases (CE)</td>
<td>13</td>
<td>8</td>
</tr>
</tbody>
</table>

The genome of *Exophiala sp.* ETNP2018 is 1.7% shorter than that of *Exophiala aquamarina*. The number of predicted genes within the *Exophiala sp.* ETNP2018 genome, however, outnumbers that of *Exophiala aquamarina* by 7.7%. Both genomes include biosynthetic gene clusters such as choline and 1,3,6,8-tetrahydroxynaphthalene. One gene cluster, encoding for clavaric acid, was specific to *Exophiala aquamarina*. 

37
Neither of the genomes contain the enzymes necessary for dissimilatory nitrate reduction or for denitrification; however, they both include genes encoding nitrilase, a hydrolase that acts on a wide range of aromatic nitriles, as well as the genes necessary for assimilatory nitrate reduction.

3.4 DISCUSSION

The genome of *Exophiala sp.* ETNP2018 was sequenced and analyzed, revealing its potential to survive and metabolize as an obligate marine fungus. The *Exophiala sp.* ETNP2018 genome assembly was 40,300,402 base pairs in length, approximately 1.7% shorter than that of the sequenced reference genome of *Exophiala aquamarina*. A functional annotation shows that this strain includes genes involved in various metabolic, molecular, and cellular processes, including, but not limited to, transporter activities, intracellular trafficking, signal transduction, cell wall and membrane biogenesis, the metabolism of carbohydrates such as sucrose, galactose, and mannose, and the biosynthesis of various secondary metabolites. Of the secondary metabolites able to be produced by *Exophiala sp.* ETNP2018, one is associated with NRPS-like peptides. *Exophiala sp.* ETNP2018 also has a wide range of CAZymes, suggesting that it is able to utilize a wide range of substrates and is capable of a generalist lifestyle. Of these CAZymes, *Exophiala sp.* ETNP2018 encodes for members of the carbohydrate-binding module family, including CBM20, CBM32, CBM67, and CBM87. CBMs aid in the binding of otherwise insoluble substrates, including cellulose and galactose. Interestingly, the genome of *Exophiala sp.* ETNP2018 lacks genes within the CBM1 classification. CBM1 is associated with cellulose-binding in fungi. Several members of AA1, a CAZyme family of laccases associated with lignin degradation, were conserved in the
Exophiala sp. ETNP2018 genome, indicating that this strain may be capable of degrading lignocellulose. The presence of various CBMs, AA1 laccases, and CEs demonstrates the ability to degrade woody substrates.

Annotation via GO, KEGG, and EggNOG databases revealed that Exophiala sp. ETNP2018 contains the genes involved in central carbon metabolic pathways, as well as those involved in other carbohydrate metabolic pathways such as galactose degradation (via the Leloir pathway) as well as glycogen biosynthesis and degradation. Genes encoding for various glycosylases, hydrolases, and cellulases were also detected.

The assimilatory reduction of nitrate to ammonia consists of two steps, the first of which utilizes nitrate reductase (NR; E.C. 1.7.1.1) to reduce nitrate to nitrite, and the second of which utilizes an assimilatory nitrite reductase (NIT-6; E.C. 1.7.1.4) to transform the nitrite into ammonia (Garrett & Amy, 1979). Most fungal species are capable of assimilating nitrate and/or nitrite, though the energetic cost associated with this process causes it to be repressed by the presence of ammonia (Pfannmüller et al., 2017). Therefore, it is not surprising that Exophiala sp. ETNP2018 contains genes encoding for the enzymes necessary for assimilatory nitrate reduction. This genome also contains the gene encoding nitrilase, an enzyme that acts on a wide range of aromatic nitriles to produce ammonia (Pace & Brenner, 2001). The ability to convert nitriles has been demonstrated in an isolated deep-sea bacterial strain, as well as in nitrile-contaminated soils, polluted river mud, and shallow marine sediments (Heald et al., 2001). Without a metabolomic analysis, it is uncertain as to under which conditions Exophiala sp. ETNP2018 utilizes the nitrile conversion pathway, if at all.
The prokaryotic denitrification pathway, utilized by bacteria in anaerobic conditions, produces N₂O as a byproduct, and has been associated with a significant loss of nitrogen from both terrestrial and oceanic ecosystems. The process of eukaryotic denitrification, however, has been demonstrated to terminate at the production of N₂O from the reduction of NO. The release of N₂O, a potent greenhouse gas, by fungi has been reported in various environments, including coastal marine sediments and oxygen minimum zones (Cathrine & Raghukumar, 2009; Peng & Valentine, 2021). While *Exophiala* sp. ETNP2018 appears to be capable of the reduction of NO to N₂O through the fungal nitric oxide reductase-encoding gene *P450nor*, experimental data is required to determine whether this strain is able to conduct fungal denitrification and produce N₂O.

There is an absence of polyphenol oxidases within the genome of *Exophiala* sp. ETNP2018, demonstrating that this strain is unable to degrade polyphenols as well as gallotannins and ellagitannins from terrestrial sources (Cammann et al., 1989). However, there are a number of genes in *Exophiala* sp. ETNP2018 that produce enzymes capable of utilizing various polysaccharides that are common in marine ecosystems, including laminarins and fucoidans, both of which are synthesized by brown algae, and which serve as storage glucans and cell wall glycans, respectively (Cuskin & Lowe, 2020). While fucoidans are typically too large to be utilized as drugs, low-molecular weight fucoidans have been extensively studied for their anti-inflammatory properties (Vischuk et al., 2011). Fucoidanases, or enzymes that are able to cleave fucoidans into oligomers, have been found in marine bacteria, invertebrates, and some fungi (Rodriguez-Jasso et al., 2010; Silchenko et al., 2005). The complete bacterial degradation pathway of glycan structures, a metabolically-expensive process first discovered in *Lentimonas* sp. CC4 by
Sichert et al., consists of over 46 enzymes, including those found in glycoside hydrolase families GH29, GH95, and GH141 (Cuskin & Lowe, 2020; Sichert et al., 2020). The fungal degradation of algal polysaccharides, however, is poorly characterized and often associated with the same fungal enzymes used for plant-polysaccharide-degrading enzymes (Balabanova et al., 2018). The β-1,3-linkage, an abundant structure in marine substrates, is cleaved and degraded by enzymes belonging to various GH families, including GH3, GH5, GH16, GH17, GH26, and GH64, all of which have members that are present in the *Exophiala* sp. ETNP2018 genome. In addition, members of the carbohydrate-binding molecule family 6 (CBM6), which serves as an auxiliary domain for the reception of laminarin, are found within the *Exophiala* sp. ETNP2018 genome. To utilize sugars from marine polysaccharides, the attached sulfate groups must be removed. The genome of *Exophiala* sp. ETNP2018 includes modular sulfatases (E.C. 3.1.6.1) with glycoside hydrolase catalytic modules, often associated with glycoside hydrolase family GH78, or the rhamnosidases, demonstrating that this species is able to utilize energy from the degradation of marine polysaccharides (Helbert, 2017).

The genome of *Exophiala* sp. ETNP2018 contains the gene that encodes deoxyribodipyrimidine photo-lyase (hereby referred to as photo-lyase; E.C. 4.1.99.3). Photo-lyases are DNA repair enzymes that activate when a cell is exposed to ultraviolet light (Ries et al., 2000). The presence of a DNA photo-lyase-encoding gene could indicate that *Exophiala* sp. ETNP2018 is not adapted to deep-sea environments (Hagestad et al., 2021). Other fungal strains have partially lost the ability to produce photo-lyases, including white-nose fungi isolated from cave-dwelling bats (Palmer et al., 2018).
Methylo-trophic fungi, which harvest carbon and energy from methanol, have been extensively studied due to their practical applications. Unlike methylo-trophic bacteria, which span diverse genera and subclasses, these yeasts belong to a select number of genera such as *Candida* and *Kuraishia* (Yurimoto & Sakai, 2019). The successful utilization of methanol relies on the ability of the microbe to convert three C1 molecules into one C3 compound (van der Klei et al., 2006). In bacteria, this can be achieved using three different pathways: the ribulose bisphosphate cycle, the ribulose monophosphate cycle, and the serine pathway. In fungi, however, the only described C1 assimilation method is associated with the xylulose monophosphate cycle (van der Klei et al., 2006).

The genome of *Exophiala* sp. ETNP2018 contains a gene used to produce alcohol oxidase, an enzyme that is used in the initial oxidation of methanol, the first step in the fungal methanol metabolism pathway. The hydrogen peroxide formed by this reaction is decomposed into water and oxygen by catalase, which is also present in *Exophiala* sp. ETNP2018. The genome also contains a gene associated with the production of dihydroxyacetone synthase (DHAS), an enzyme that catalyzes the first step of formaldehyde assimilation by fixing formaldehyde to xylulose 5-phosphate (van der Klei et al., 2006). All of the genes needed for formaldehyde assimilation (via the xylulose monophosphate cycle) are conserved in the *Exophiala* sp. ETNP2018. This demonstrates that this strain is capable of metabolizing methanol, which may be indicative of the anoxic conditions of the water in which it resides.

The genus *Exophiala*, first described by Carmichael in 1966, is commonly associated with black yeast-like fungi that include annellidic conidiogenesis and yeast-like states (de Hoog et al., 2011). *Exophiala aquamarina*, originally sequenced by the
Broad Institute as part of a black yeast comparative genome project, was compared to the genome of *Exophiala sp.* ETNP2018 (Teixeira et al., 2017). The genome of *Exophiala sp.* ETNP2018 is 1.7% shorter than that of *Exophiala aquamarina*, yet it contains 1,104 more genes. The genomes of *Exophiala sp.* ETNP2018 and *Exophiala aquamarina* encode various members of the glycoside hydrolase family (186 and 184 members, respectively), members of the carbohydrate-binding molecule family (11 and 18 members, respectively), and members of the carbohydrate esterase family (13 and 8 members, respectively). *Exophiala aquamarina* contains less genes that encode glycoside hydrolase and carbohydrate-binding molecule family members associated with the degradation of the β-1,3-linkage found in marine substrates; it is missing genes encoding members of GH5, GH26, and CBM6, all of which are conserved in the *Exophiala sp.* ETNP2018 genome. Both genomes encode the enzyme nitrite reductase (E.C. 1.7.1.4), used in assimilatory nitrate reduction. They also both contain the gene encoding photo-lyase (E.C. 4.1.99.3), indicating that neither strain has lost the ability to survive under UV radiation. The genome of *Exophiala aquamarina* also contains the gene encoding a modular sulfatase (E.C. 3.1.6.1), alongside glycoside hydrolase catalytic modules, demonstrating that both species are able to at least partially degrade complex marine polysaccharides. However, it is worth noting that the genome of *Exophiala aquamarina* does not encode all of the genes necessary for the xylulose monophosphate cycle associated with methanol oxidation and metabolism.
REFERENCES


APPENDIX A

SUPPLEMENTARY FIGURES

Figure A.1 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 19. Measurements were taken on 29 December 2021.
Figure A.2 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 35. Measurements were taken on 7 January 2022.
Figure A.3 CTD measurements of temperature (a), salinity (b), dissolved oxygen (c), and photosynthetically available radiation (PAR) (d) at Station 41. Measurements were taken on 11 January 2022