



Norwegian University
of Life Sciences

Master's Thesis 2020 60 ECTS

Faculty of Biosciences (BIOVIT)

Whole genome sequencing with Oxford Nanopore and de novo genome assemblies of lipid- producing fungi in phylum Mucoromycota

Kai Fjær

Master of Biotechnology, Genetics

Summary

Phylum Mucoromycota consists of economically and ecologically important fungi, including industrial producers of lipids, enzymes and fermented foods and beverages, symbionts and decomposers of plants, as well as fungi causing post-harvest diseases and opportunistic infections in humans. The phylum includes great candidates for sustainable production lipids, but very few have so far been the subject of genomic research.

In this project, the genomes of eleven lipid-producing strains were sequenced and assembled and placed phylogenetically in the phylum. Genomic DNA was extracted using bead-beating and sequenced with Oxford Nanopore Technologies' PromethION platform. Three barcoded sequencing runs produced 62.7 gigabases and 18.9 million reads, of which 13 million reads were used to create *de novo* assemblies.

Extracting high-quality DNA from Mucoromycota fungi is challenging. When extracting DNA for long-read sequencing, care should be taken to avoid mechanical fragmentation of the DNA and to inactivate and inhibit DNA-degrading enzymes. Despite the comparatively short read lengths resulting from degradation of the extracted DNA, the nanopore reads resulted in highly contiguous assemblies for several strains.

Abbreviations

DNA	Deoxyribose nucleic acid
EDTA	Ethylenediaminetetraacetic acid
SDS	Sodium dodecyl sulfate
CTAB	Cetyltrimethylammonium bromide
MEA	Malt extract agar
MEB	Malt extract broth
PDA	Potato dextrose agar
dH ₂ O	distilled water
ATCC	The American Type Culture Collection
FRR	The Food Fungal Culture Collection
CCM	The Czech Collection of Microorganisms
VKM	The All-Russian Collection of Microorganisms
UBOCC	Université de Bretagne Occidentale Culture Collection
MTP	Microtiter plate
Tris-HCL	Tris(hydroxymethyl)aminomethane hydrochloride
rpm	Rotations per minute
Mb	Megabases (1 000 000 DNA bases)
MB	Megabytes
Gb	Gigabases (1 000 000 000 DNA bases)
GB	Gigabytes
NCBI	National Center for Biotechnology Information, U.S. National Library of Medicine.

Table of Contents

Summary	2
Abbreviations	3
1 Introduction	8
1.1 The Mucoromycota	8
1.1.1 The Mucoromycota fungi are nature’s recyclers and can help us towards a sustainable future.....	10
1.1.2 The mycelium.....	10
1.1.3 Reproduction and dispersal	11
1.1.4 Taxonomy and phylogenetics.....	12
1.2 Extracting DNA from filamentous fungi	14
1.3 Library prep - preparing the DNA for sequencing.....	15
1.4 Nanopore long-read sequencing.....	16
1.4.1 Genome assembly	17
1.4.2 The genomes of the Mucoromycota.....	20
1.5 Goals of the project	22
2 Materials and methods	22
2.1 Laboratory work.....	22
2.1.1 Cultivation of fungal strains.....	22
2.1.1.1 Origin of strain materials	22
2.1.1.2 Agar plate cultivation.....	22
2.1.1.3 Obtaining spore suspensions.....	25
2.1.1.4 Liquid cultures in microtiter plates	25
2.1.2 Biomass washing.....	26
2.1.3 Biomass handling and storage.....	26
2.1.3.1 Room temperature biomass handling protocol	26

2.1.3.2	Dry ice biomass handling protocol	26
2.1.4	DNA extraction	27
2.1.4.1	Mycelium disruption	27
2.1.4.2	Phenol-chloroform extraction	27
2.1.4.3	Isopropanol precipitation	28
2.1.4.4	Ethanol wash	28
2.1.4.5	Elution	29
2.1.5	DNA quality controls	29
2.1.5.1	NanoDrop – purity and concentration	29
2.1.5.2	Gel electrophoresis – DNA integrity	29
2.1.5.3	QuBit – concentration	30
2.1.5.4	Selection of samples for sequencing	30
2.1.6	DNA extraction – disruption method trials	30
2.1.6.1	Method one: bead-beading in TissueLyser	30
2.1.6.2	Method two: cold disruption - mortar and pestle cooled by dry ice	31
2.1.6.3	Method three: ultra-cold disruption - mortar and pestle cooled by liquid nitrogen	31
2.1.6.4	RNase A treatment	32
2.1.6.5	Quality control of DNA from the disruption method trials	32
2.1.7	Sequencing library preparation	32
2.1.8	Sequencing on ONT PromethION	32
2.1.8.1	Data acquisition	32
2.1.8.2	Basecalling and read quality filtering	32
2.2	Bioinformatics	33
2.2.1	Read statistics	33
2.2.2	Genome assembly	33
2.2.3	Assignment of taxonomic identity to assembly contigs using BLAST	34

2.2.4	Exploratory quality control of assemblies.....	34
2.2.5	Exploratory alignment of assemblies to reference genomes.....	34
2.2.6	Estimating expected assembly sizes.....	35
2.2.7	Finding G+C content of contigs.....	36
2.2.8	Exploratory assembly analysis.....	36
2.2.9	Contig filtering.....	36
2.2.10	Quality control of final assemblies.....	36
2.2.10.1	Assembly statistics.....	36
2.2.10.2	BUSCO completeness analysis.....	37
2.2.11	Visualisation of assembly contiguity.....	37
2.2.12	BUSCO phylogenomics.....	37
2.2.12.1	Outgroup selection.....	39
2.2.12.2	Multiple alignment of single-copy BUSCO genes.....	40
2.2.12.3	Gene concatenation supermatrix approach.....	40
2.2.12.3.1	Concatenation of alignments.....	40
2.2.12.3.2	Phylogenetic tree from gene concatenation supermatrix.....	40
2.2.12.4	Gene tree coalescence approach.....	41
2.2.12.4.1	Gene trees.....	41
2.2.12.4.2	Coalescence of gene trees to create species tree.....	42
2.2.12.5	Visualization of phylogenetic trees.....	42
3	Results.....	42
3.1	Obtaining biomass.....	42
3.2	DNA extraction.....	43
3.2.1	Disruption method trials.....	43
3.2.1.1	DNA integrity.....	43
3.2.1.2	Yield.....	44
3.2.1.3	DNA purity.....	44

3.2.2	Extraction of genomic DNA from fungal cultures.....	45
3.2.2.1	DNA integrity	45
3.2.2.2	Purity and yield – all samples	48
3.2.2.3	Purity and yield – sequenced samples	49
3.3	Sequencing	50
3.3.1	Sequencer performance and read statistics.....	50
3.3.2	Read statistics of barcoded samples	53
3.4	Genome assemblies	59
3.4.1	Genome assemblies of eleven fungal strains.....	59
3.4.2	Genome assembly of a bacterial contaminant.....	62
3.5	BUSCO phylogenomics	63
3.5.1	Gene concatenation approach.....	63
3.5.1.1	Phylogeny of phylum Mucoromycota	63
3.5.1.2	Placement of the sequenced strains.....	63
3.5.2	Gene tree coalescence approach.....	68
3.5.2.1	Comparison with the concatenation phylogeny	68
3.5.2.2	BUSCO completeness of Mucoromycota assemblies.....	68
4	Discussion	71
4.1	Extracting DNA from Mucoromycota fungi.....	71
4.2	Read length.....	73
4.3	Genome assemblies	73
4.3.1	<i>Mucor racemosus</i> UBOCC-A-111127 – a possible genome duplication	74
4.3.2	Bacterial sequences in Mucoromycota assemblies	75
4.4	BUSCO phylogenomics	76
5	Reference list.....	77
6	Appendix	85
6.1	Cultivation of fungal strains.....	85

1 Introduction

1.1 The Mucoromycota

The fungi of phylum Mucoromycota are near ubiquitous in nature. They grow as filamentous molds and are commonly found as decomposers in soil or on plants and plant materials, or as root symbionts (Spatafora et al., 2017). Mucoromycota is an economically important group with representatives used for industrial production of fatty acids, enzymes and metabolites, for production of fermented foods and beverages, and they are causative agents of post-harvest crop diseases and opportunistic infections in humans (Lennartsson et al., 2014; Spatafora et al., 2017; Walther et al., 2019; Wang et al., 2011). Some Mucoromycota species have a remarkable ability to accumulate lipids (Kosa et al., 2018; Wang et al., 2011) as well as the ability to grow on low-value substrates, making them promising candidates for sustainable production of lipids for fuels and for human and animal consumption (Papanikolaou et al., 2007; Tzimirotas et al., 2018).

Figure 1.1 shows an overview of the phyla in kingdom Fungi. Phylum Mucoromycota consists of three subphyla – Mucoromycotina, Mortierellomycotina and Glomeromycotina (Spatafora et al., 2016). The fungi in subphyla Mucoromycotina and Mortierellomycotina accumulate lipids and are generally easily culturable. Mucoromycotina fungi usually grow as decomposers on dead or dying plants. The fungi in subphylum Mortierellomycotina are also decomposers. They are found in soil and are often associated with plant roots, sometimes as root parasites (Spatafora et al., 2017; Spatafora et al., 2016). The fungi in subphylum Glomeromycotina have a lifestyle and a morphology unlike that of all other fungi. They can only live in symbiosis with plant roots and supply the plant with minerals in exchange for fixed carbon through tree-like structures (arbuscules) inside the plant cells. This mutualistic symbiosis is so successful that nearly all terrestrial plants live in symbiosis with a Glomeromycotina fungi (Spatafora et al., 2017; Tisserant et al., 2013). Fossil evidence from 400 million years ago suggests that living in symbiosis with Mucoromycota fungi enabled the early plants to colonize land (Bidartondo et al., 2011; Brundrett, 2002; Krings et al., 2013; Remy et al., 1994).

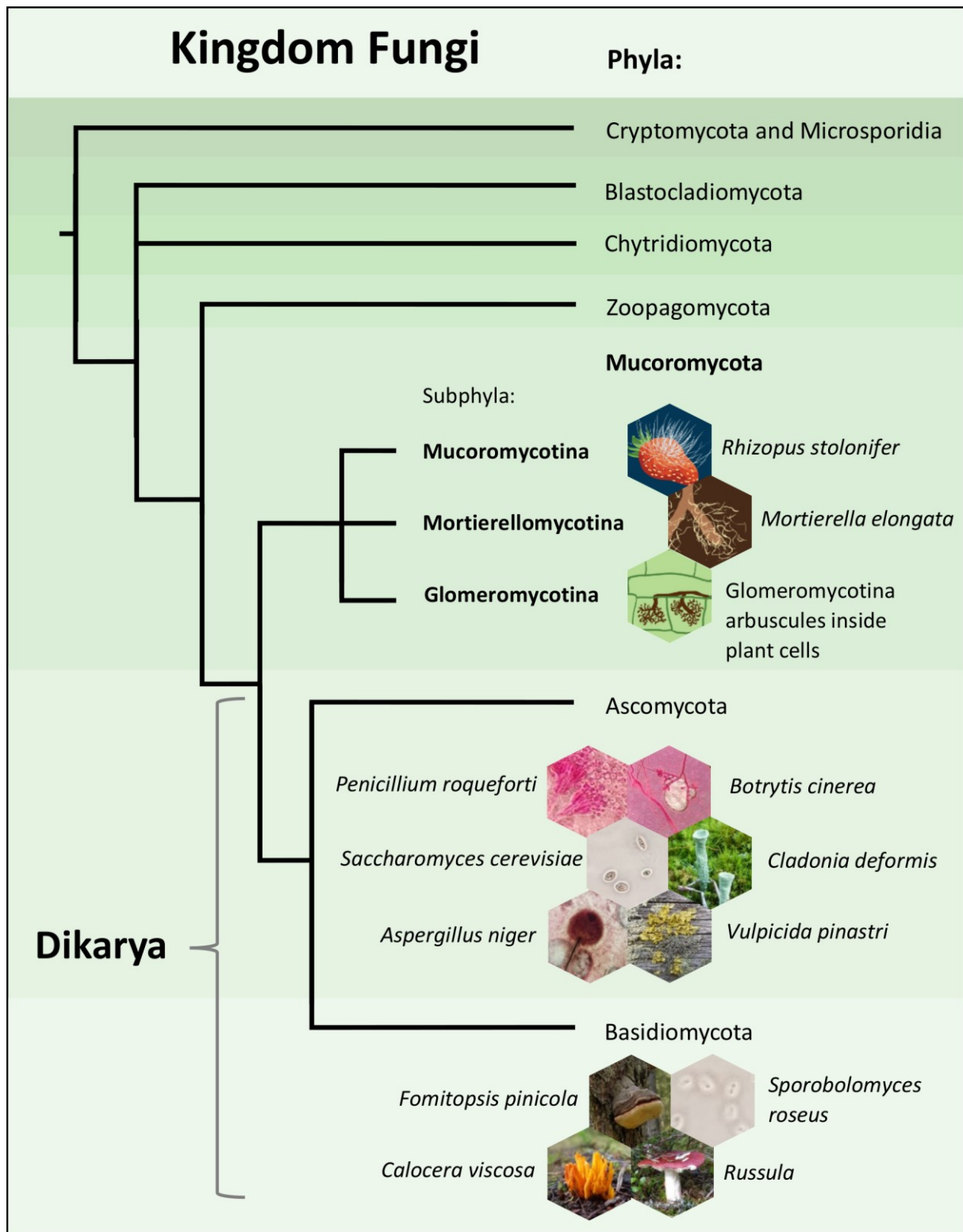


Figure 1.1. The fungal tree of life. The cladogram shows the phylogeny of the phyla in kingdom Fungi and is adapted from Spatafora et al. (2017). The pictures show a selection of the diversity within the phyla Mucoromycota, Ascomycota and Basidiomycota. *Rhizopus stolonifer* on strawberry: drawing adapted from photograph in Feliziani and Romanazzi (2016). *Mortierella elongata* interacting with roots of *Arabidopsis*: drawing adapted from

photograph in Weisenberger et al. (2013). Glomeromycotina arbuscules: drawing adapted from images by Brundrett (2008). The rest of the pictures are own photographs.

1.1.1 The Mucoromycota fungi are nature's recyclers and can help us towards a sustainable future

Many species in subphyla Mucoromycotina and Mortierellomycotina are able to rapidly accumulate lipids in culture. They can produce a wide array of lipid compounds, including low value lipids ideal for production of biodiesel and high value lipids such as essential fatty acids suitable for use in pharmaceuticals and food supplements (Kosa et al., 2018; Papanikolaou et al., 2007). The lipid composition can vary considerably between strains of lipid-producing species (Kosa et al., 2018). Due to their lifestyle as decomposers in nature, the fungi are able to grow on many kinds of organic materials, including low-value substrates such as waste from forestry and agriculture, food waste, waste glycerol and rest fat from animals (Blomqvist et al., 2019; Magdouli et al., 2014; Tzimirotas et al., 2018). These abilities make the fungi ideal candidates for sustainable lipid production. Studying the genomes of lipid-producing strains can give insight into the mechanisms of the fungi's lipid metabolism and help direct and optimize the development of large-scale, environmentally sustainable oil production (Grigoriev et al., 2011; Sharma, 2015).

1.1.2 The mycelium

Filamentous fungi grow as thick-walled tubes known as hyphae. The hyphae of a fungal colony can form a branched network called a mycelium. The mycelium expands by growth at the hyphal tips, where digestive enzymes are secreted, and nutrients absorbed. Nutrients that are absorbed in a hyphal tip can be transported to any place in the mycelial network, which enables the mycelium to span across substrates of both high and low nutritional content (Money, 2016b). Many Mucoromycota fungi lack cross-walls in their hyphae, which means that nuclei, vesicles and other cellular structures and organelles can be transported to any location in the mycelium (Naranjo-Ortiz & Gabaldon, 2020). The Mucoromycota fungi accumulate lipids in organelles called lipid bodies, see Figure 1.2, which are mainly composed of triacylglycerols with a wide diversity of possible fatty acids (Wang et al., 2011). The lipid bodies are energy storage organelles formed as a stress response to low nitrogen levels (Papanikolaou et al., 2007). Some species of Mucoromycota house symbiotic bacteria. The bacteria disperse by hitching a ride in the reproductive spores of the fungus (Bonfante & Desiro, 2017; Mondo et al., 2017b; Uehling et al., 2017).

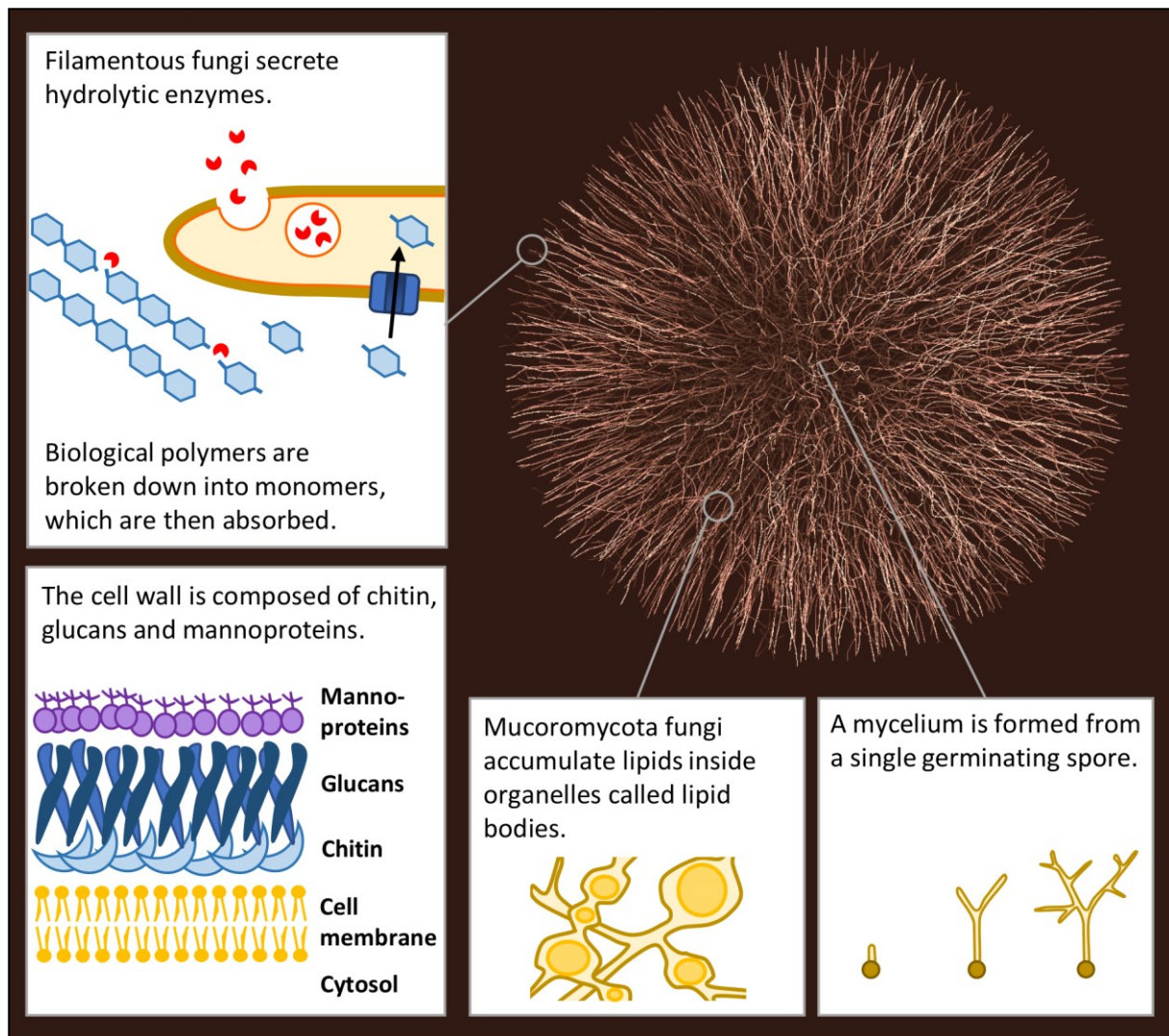


Figure 1.2. The mycelium. The depicted mycelium on dark background was simulated using the *Tricholoma* model in the software Neighbour-Sensing mathematical model of hyphal growth (Moore & Meškauskas, 2017). The illustrations on white background show different aspects of the mycelia of Mucoromycota fungi. The cell wall composition illustration is adapted from a figure in Vega and Kalkum (2012).

1.1.3 Reproduction and dispersal

In addition to spreading by expanding their mycelial networks, fungi in phylum Mucoromycota reproduce by producing spores that are dispersed by wind or by animals. Asexual spores are created through mitosis in sporangia at the tips of spore-bearing hyphae (see Figure 1.3.A). These are the spores that are normally used when inoculating cultures in growth media. In species where sexual reproduction is observed, it is carried out through the production of zygospores. Zygospores are formed when two hyphae of opposite mating types

meet and merge. Nuclei from the two haploid hyphae, meaning that they each have one set of chromosomes, are brought together in the zygospores, making it diploid. The hardy, thick-walled zygospore germinates and sprouts a sporangium that produces haploid spores through meiosis (Lee & Heitman, 2014). The steps in the sexual reproduction cycle is shown in Figure 1.3.B-F. During meiosis, mixing of genetic material occurs, giving sexually reproducing species the advantage of increased genetic diversity. The appearance of the spores and spore-bearing structures varies widely within the phylum, from complex and ornamented to simple like the ones shown in Figure 1.3.

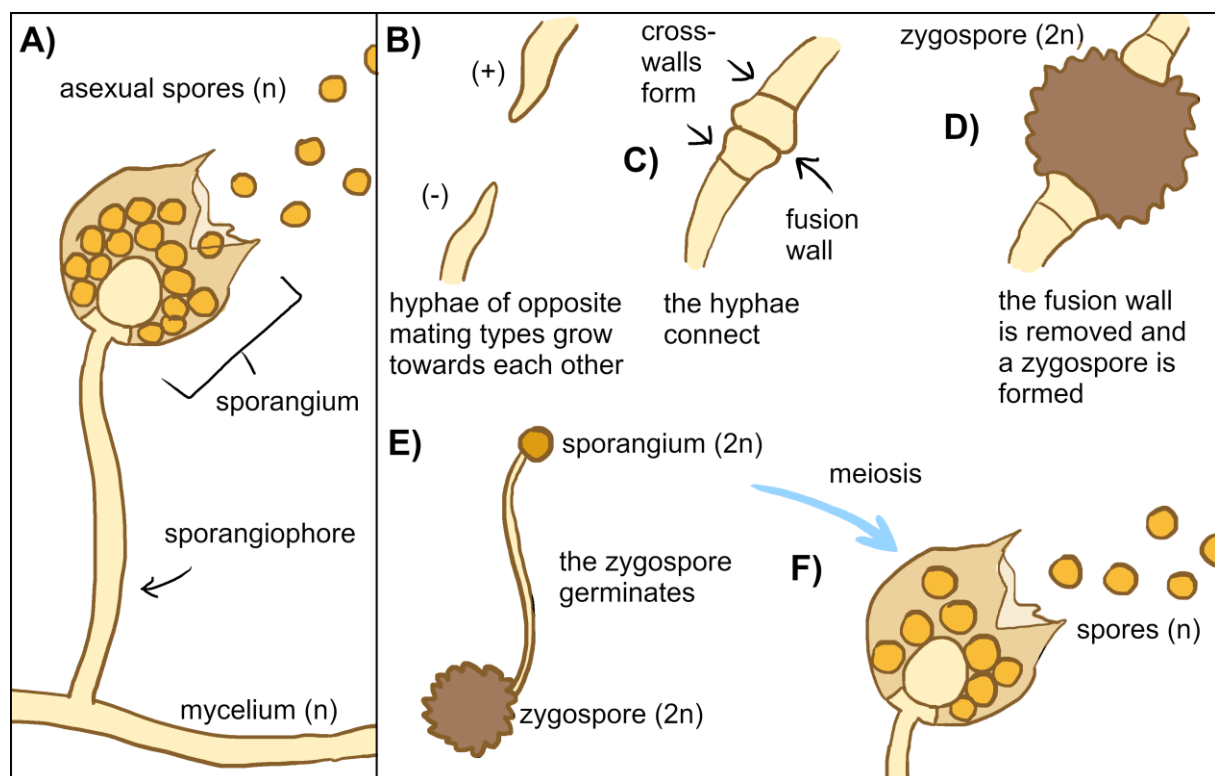


Figure 1.3. Reproduction of Mucoromycota fungi. A) Asexual reproduction by the production of sporangiospores (mitospores). B)-F) Sexual reproduction by the production of zygospores. A-D are adapted from Blakeslee (1904), Hocking (1967) and O'donnell et al. (1976). E is adapted from Gauger (1961). Ploidity is denoted by n (haploid) and $2n$ (diploid).

1.1.4 Taxonomy and phylogenetics

Traditionally, fungi were classified on basis of morphology and ecology. Following the introduction of DNA sequencing, it turned out that similarity in morphology, lifestyle and role in an ecosystem not necessarily corresponds to evolutionary relatedness in the fungal kingdom. For many years, the classification of Mucoromycota fungi was based on

morphological traits of the mycelium, spores and spore-bearing structures, but when researchers started using DNA sequences to infer relatedness they discovered that the traits had little to do with evolutionary relatedness (Hoffmann et al., 2013; Walther et al., 2019). In the last decades, phylogenetic studies using differences in DNA sequences to infer evolutionary relationships (phylogenies) have caused the taxonomy of kingdom Fungi to be rearranged over and over (Money, 2016a). This is also the case within phylum Mucoromycota. The phylogeny of the phylum is currently not clearly resolved, and there are many examples of fungal isolates being reclassified multiple times (Hoffmann et al., 2013). For several years, only a few short DNA sequences were used in phylogenetic studies of fungi. In recent time, however, DNA sequencing technologies has gone through rapid developments and the price of sequencing has dropped, making it possible to sequence more genomes and using sequences from all across the genomes to infer phylogenies. The use of genome-scale data to infer evolutionary relationships is called phylogenomics (Young & Gillung, 2020). In phylogenomics, there are two ways to build a phylogenetic tree from multiple genes. In the gene concatenation approach, gene sequence alignments are concatenated into one large alignment which is analyzed to make the species tree. In the coalescence approach, a phylogenetic tree is inferred for each of the genes separately before the gene trees are coalesced into a species tree (Gadagkar et al., 2005).

In 2016, on the basis of phylogenetic analyses using genome-scale data, Spatafora et al. abandoned the phylum Zygomycota and split it into two phyla: Mucoromycota and Zoopagomycota. Glomeromycotina, which was earlier considered its own phylum, turned out to share a most recent common ancestor with Mortierellomycotina and Mucoromycotina and was thus included in phylum Mucoromycota. Whether Glomeromycotina is more related to Mortierellomycotina or Mucoromycota is currently not known, and it is necessary to include more taxa in phylogenetic analyses to resolve the relationship. The relationship of Mucoromycota to the other phyla of kingdom Fungi is not clearly resolved, either. The subkingdom Dikarya, consisting of the diverse and well-studied phyla Ascomycota and Basidiomycota, diverged from Mucoromycota roughly 600-700 million years ago (Chang et al., 2019; Samarakoon et al., 2017; Tedersoo et al., 2018). Whether Dikarya diverged from Glomeromycotina, Mortierellomycotina or Mucoromycotina, or from a most recent common ancestor of all or two of them (Chang et al., 2019), is currently not certain, but Dikarya is placed closest to Glomeromycota in some analyses (Spatafora et al., 2016; Tedersoo et al.,

2018). Sequencing a wider selection of fungal genomes and coupling phylogenetic analyses with fossil evidence will give us a clearer picture of the evolutionary history of fungi.

1.2 Extracting DNA from filamentous fungi

In order to sequence the genome of a fungus, the DNA has to be collected from the cells and be purified. Extracting pure DNA of good quality from filamentous fungi is challenging (Umesha et al., 2016). Filamentous fungi have tough cell walls that withstand usual methods of cell lysis (Fredricks et al., 2005; van Burik et al., 1998). In many species, the cell wall also contains large amounts of polysaccharides and glycoproteins which poses additional challenges to extracting DNA (Inglis et al., 2018; Kües, 2007). To obtain DNA from filamentous fungi, organic extraction methods are often used. The mycelium is first broken open either by mechanical force or by hydrolytic enzymes that break chemical bonds in the cell wall. After breaking open the cell walls, a detergent is used to break apart membranes and cellular structures before an organic solvent is used to remove proteins and lipids.

Obtaining high molecular weight DNA suitable for long-read sequencing is difficult. Care must be taken at every step to not cause damage when handling and processing a sample containing DNA. Because genomic DNA molecules easily can be broken into smaller pieces when subjected to mechanical force, several changes are often made to DNA extraction protocols to ensure the integrity of the DNA. Instead of mixing a sample by vortexing or pipetting, the sample tube may be gently flicked or inverted. Pipetting is ideally done with pipette tips that have wide openings, and unnecessary pipetting is avoided. When it comes to extracting high molecular weight DNA from filamentous fungi, crushing the mycelium with a mortar and pestle cooled by liquid nitrogen is generally the preferred method (Inglis et al., 2018; Pacific Biosciences, 2020; Quick & Loman, 2018). The pestle breaks open the cell walls while the DNA remains largely undamaged. However, working with liquid nitrogen poses safety hazards, and because the mortar and pestle needs to be cleaned between each sample, the method is time-consuming when extracting DNA from many samples. In this project, bead-beating with steel beads was chosen instead as the method for mycelium disruption. Despite the disadvantage of often causing shearing of DNA (Quick & Loman,

2018), the method is convenient and allows many samples to be processed simultaneously (Inglis et al., 2018; Muller et al., 1998).

Following mycelium disruption, a lysis buffer is mixed with the crushed mycelium. The lysis buffer contains a detergent, for instance sodium dodecyl sulfate (SDS) that breaks apart membranes and cellular structures and dissolves and denatures proteins (Umesha et al., 2016). The lysis buffer also contains Tris-HCL, which keeps the pH of the cell lysate stable as organelles and cellular compartments with differing acidity are broken open. The lysis buffer contains salt, often NaCl, which helps in separating proteins from the DNA and in keeping proteins dissolved. EDTA, ethylenediaminetetraacetic acid, is added to the lysis buffer to protect the DNA. EDTA binds magnesium ions that DNases would otherwise utilize to break DNA into smaller pieces (Heikrujam et al., 2020).

In organic DNA extraction, a mixture of phenol and chloroform is often used as organic solvent. The phenol denatures proteins by bringing non-polar amino acid residues to the protein surface, and pulls proteins, lipids and hydrophobic compounds into the organic phase. DNA is water soluble due to the negative charges of the phosphate groups in its backbone and is kept inside the aqueous phase as unwanted compounds are removed with the organic phase. Chloroform gives the organic phase extra density, so that the organic phase always is below the aqueous phase after centrifugation and so that the phase separation is sharp. The DNA is separated from the lysis buffer and cell lysate solution by adding an alcohol, which allows Na^+ ions to bind to the DNA's phosphate groups, thus breaking the hydration shell around the DNA and encouraging the molecules to aggregate and fall out of solution. The precipitated DNA is washed with ethanol and eluted in water or a pH-stabilizing buffer (Heikrujam et al., 2020).

1.3 Library prep - preparing the DNA for sequencing

DNA is not ready for sequencing straight from the cell. The process of making a DNA sample ready for sequencing is called sequencing library preparation, or library prep for short. The sample, in ready-to-be-sequenced form, is called the sequencing library. In order to sequence DNA on a nanopore device, an adapter DNA molecule bound to a motor enzyme is attached to each end of the DNA molecules. The motor enzyme is needed to pull a DNA strand

through the nanopore during sequencing (Jain et al., 2016). Before attaching adapter molecules to the ends of the DNA molecules, barcode DNA molecules may be added so that several samples from different sources may be sequenced at the same time. The barcode molecules contain unique base pair sequences, which allows sorting of sequenced DNA according to sample of origin (Wick et al., 2018).

1.4 Nanopore long-read sequencing

To sequence DNA means to read the sequence of nucleotides and represent it digitally as a string of A's, T's, C's and G's, letters representing the nucleotide bases. A sequencer produces reads, which are text files containing the base sequence inferred from the sequenced DNA molecule. The files are usually in the format FASTQ, which contains the base sequence and a quality score for each of the bases that says how likely it is that the base was identified correctly.

The Oxford Nanopore sequencing technology reads the DNA sequence directly as the DNA molecule is pulled through a nanopore. The nanopore sits across a membrane that divides two solutions of differing ionic strength. An ionic current flows through the nanopore from the solution with highest ionic strength, which is the solution the DNA is added to. A DNA molecule connects with the nanopore and the motor enzyme attached to DNA molecule unzips the double helix and pushes one strand through the nanopore. The passing of the DNA molecule through the nanopore disrupts the ionic current. Because the DNA bases have different sizes, they block different proportions of the ionic flow through the pore, resulting in fluctuations in the current. The changes in ionic current across the membrane is measured over time, and the electric signal is translated to base sequence by a machine learning software (Jain et al., 2016; Pollard et al., 2018). The translation from raw signal into bases, known as base calling, is less accurate in Oxford Nanopore sequencing compared to other technologies. Nanopore reads have an error rate of around 10% (Jain et al., 2016) meaning that on average, every tenth base is called incorrectly. In contrast, Illumina sequencing, which is the most widely used technology and which produces short reads, has an error rate of around 0.1% (Glenn, 2011). Despite the disadvantage of a high error rate, nanopore reads are

very useful when reconstructing, i.e. assembling, genomes from reads (Jain et al., 2016; Pollard et al., 2018).

1.4.1 Genome assembly

To assemble a genome means to reconstruct an organism's genomic sequence from reads. In the assembly process, reads are pieced together based on sequence overlap – reads that span the same sequence in the genome can be stitched together at the shared bases. Many reads covering a location in the genome make it more likely that the location is accurately represented. The number of reads covering one specific base is that bases' read coverage. A high read coverage across the genome is necessary to obtain a high-quality assembly. A set of overlapping reads form a contig, which represents a part of the genomic sequence, for instance a piece of a chromosome. The genome assembly consists of contigs, whose base sequences are represented in text files in the FASTA format (see **Feil! Fant ikke referansekinden.**). If the length of the gap between two contigs is known, the contigs can be stitched together to a scaffold which consist of the two contig base sequences separated by a string of N's (N meaning any base in the FASTA format).

Because nanopore reads are long, they are very useful for assembling genomes. Just as a jigsaw puzzle is easier to assemble from large pieces than from small, assembling a genome accurately is easier when using long reads compared to short reads. Nanopore reads can be tens and even hundreds of thousands of bases long (Jain et al., 2016), which means that they can span repetitive and low-complexity genomic regions that short reads would not be able to resolve. Using long reads can reduce the frequency of breaks and of misassemblies, where the reconstructed genome is pieced together wrongly, thus resulting in continuous assemblies that accurately represents the true genomic sequence.

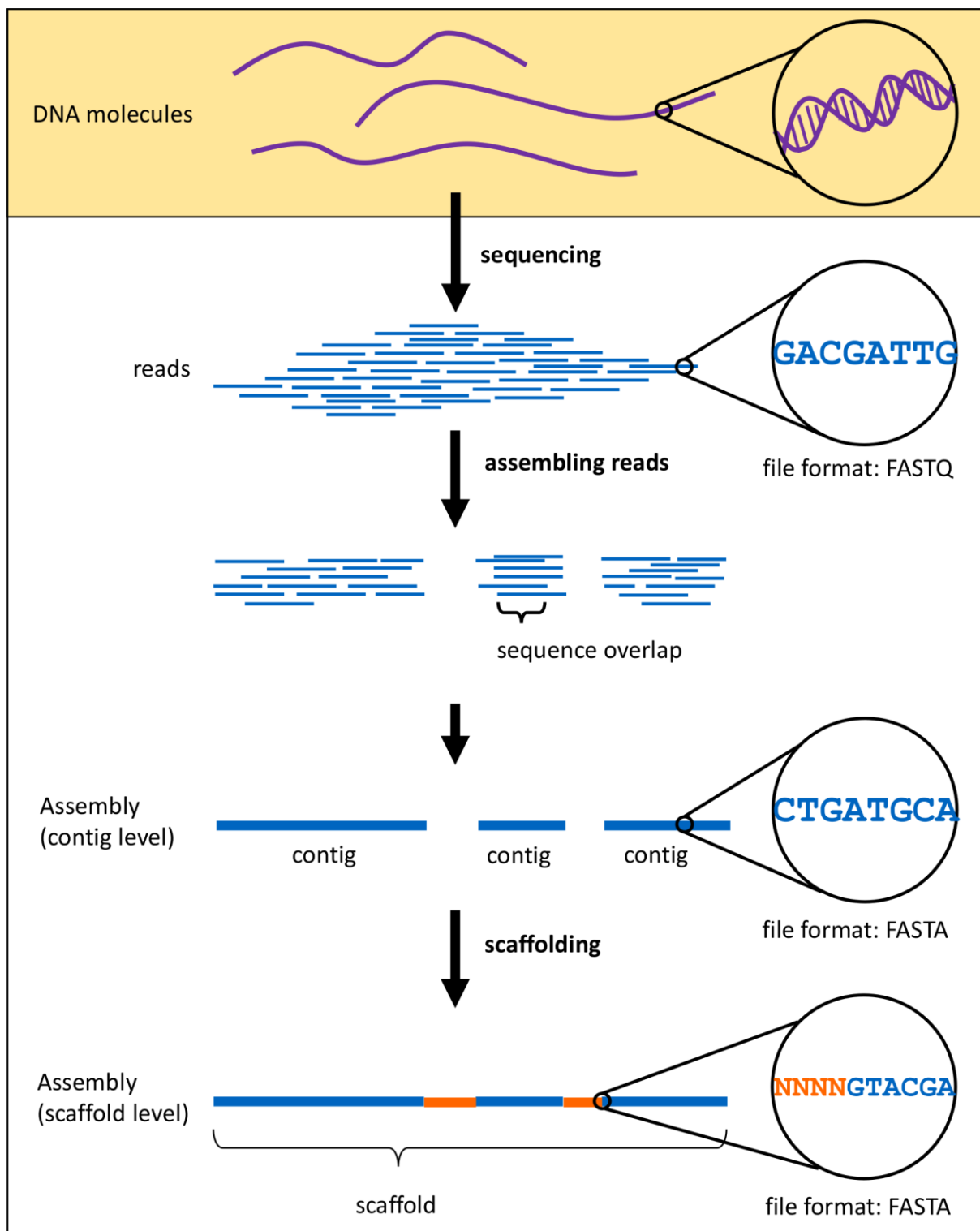


Figure 1.4 Genome assembly from sequencing reads.

Nanopore reads are suitable for *de novo* assembly, i.e. building an assembly from scratch rather than using a reference to build the assembly (see Figure 1.5). The *de novo* assembly method was used in this project. Using long reads for *de novo* assembly enables detection of genomic variations within a species, variations that can span large regions of DNA. These types of variations are often not detected when comparing short-read assemblies but can have large impacts on the biology of the organism.

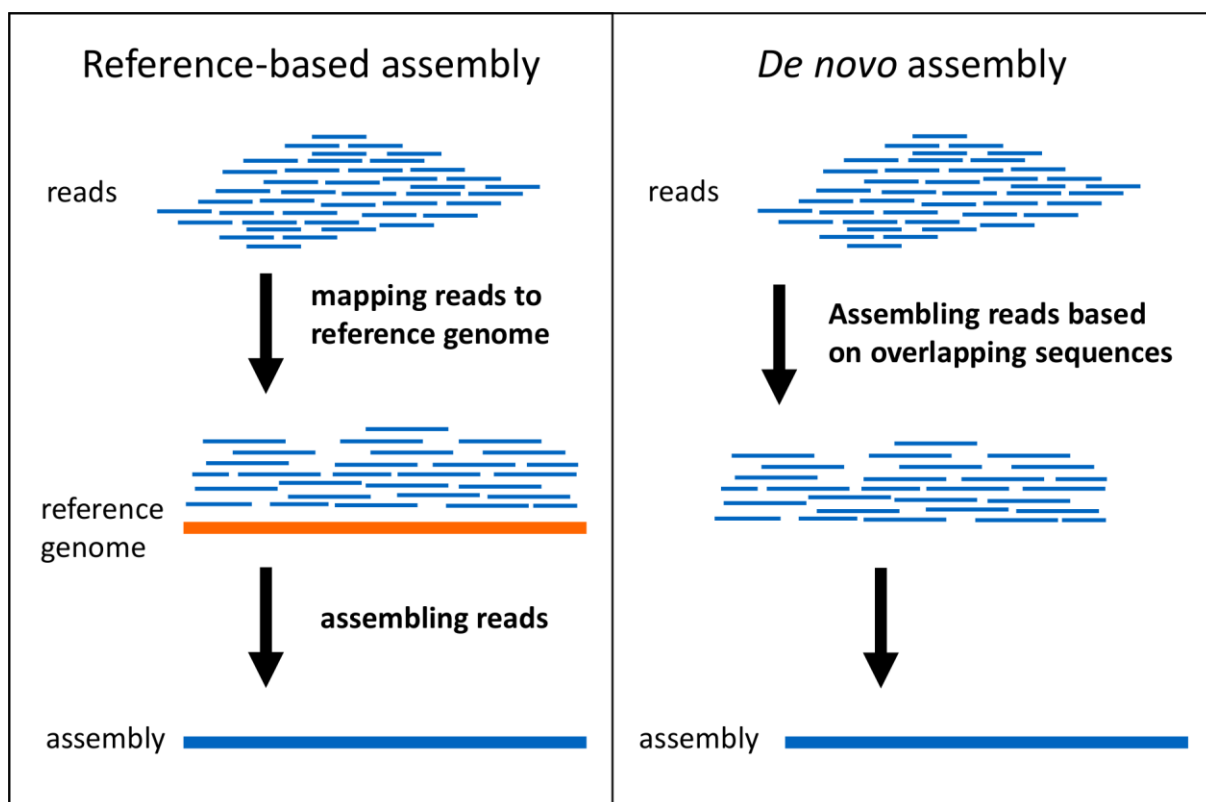


Figure 1.5 Reference-based genome assembly versus *de novo* genome assembly.

1.4.2 The genomes of the Mucoromycota

The genomes of Mucoromycota fungi are small, ranging from around 20 Mb (Mondo et al., 2017a) to 570 Mb (Morin et al., 2019), most genome assemblies being around 40 Mb long (JGI Mycocosm; Nordberg et al., 2014). The genomes are assumed to always be haploid (Gryganskyi & Muszewska, 2014). Research into Mucoromycota genomes only recently gained traction, see Figure 1.6. From five genome assemblies in 2013, 51 new assemblies were released in NCBI's GenBank database in 2014. Currently there are about 160 assemblies in GenBank.

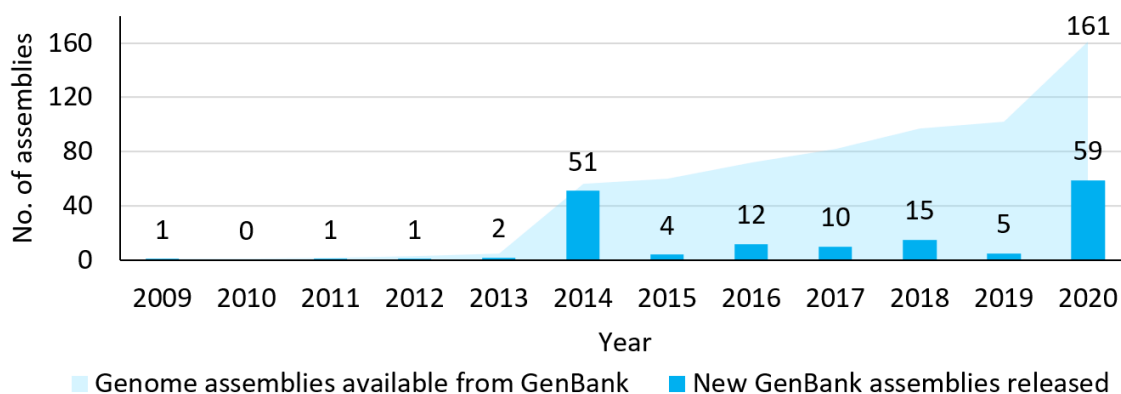


Figure 1.6. Number of Mucoromycota assembly releases in GenBank per year. The number of new assembly releases (blue, number in labels) and cumulative number of assemblies (light blue, number in 2020 shown in upper label) of Mucoromycota genomes in NCBI's GenBank database from 2014 to 2020.

The majority of the sequenced strains are medically relevant isolates belonging to order Mucorales, most of them species of *Rhizopus* (Chibucos et al., 2016). There are 27 species of Mucorales that can cause mucormycosis, an opportunistic, lethal infection in people with suppressed immune system (Soare et al., 2020; Walther et al., 2019). Genomic research into the pathogenic strains is done to uncover the mechanisms of infection and to develop methods

for prevention and treatment of mucormycosis. Some isolates of *Glomeromycotina* has also been sequenced in efforts to understand their evolution and their ecologically crucial symbioses with land plants (Morin et al., 2019; Tisserant et al., 2013; Venice et al., 2020).

Although many *Mucoromycota* fungi have remarkable abilities to accumulate lipids, some playing important roles in industrial production and others being candidates for future industrial applications, only a handful of strains has been sequenced in order to unveil the genomic underpinnings of lipid accumulation. The second *Mucoromycota* fungi to be sequenced was *Mortierella alpina* ATCC 32222 in 2011 (Wang et al.), a strain used in industrial production of arachidonic acid, a fatty acid used in infant formula and dietary supplements (Mamani et al., 2019). To investigate the mechanisms of lipid metabolism in this strain, a genome-scale metabolic model was constructed (Ye et al., 2015), and mutant strains has been created using genetic engineering to optimize production of nutritionally important fatty acids (Kikukawa et al., 2016; Kikukawa et al., 2018; Okuda et al., 2015; Sakamoto et al., 2017).

Other effective lipid producers that have been sequenced are *Umbelopsis isabellina* NBRC 7884 (Takeda et al., 2014) and *Mortierella alpina* CDC-B6842 (Etienne et al., 2014). *Mucor lusitanicus* CBS277.49 was sequenced by Corrochano et al. (2016) and had its genome and lipid metabolism investigated in Wei et al. (2013). Tang et al. (2015) sequenced the highly effective lipid producer *Mucor circinelloides* WJ11 and compared the genome to that of *Mucor lusitanicus* CBS277.49. Comparing the genomes of strains that have differing lipid-producing abilities can expand the understanding of the mechanisms behind lipid accumulation and allow optimization and direction of fatty acid production. Sequencing a broader selection of lipid-producing strains can help us identify the best strains for industrial applications, it can help us in optimizing cultivation processes and it enables genetic engineering. Sequencing and investigating the genomes of lipid-producing strains can accelerate and better direct the development of environmentally sustainable production of fuels (Grigoriev et al., 2011) and of lipids for human and animal consumption (Sharma, 2015).

1.5 Goals of the project

The goals of this project was to extract genomic DNA and sequence the genomes of lipid-producing strains in phylum Mucoromycota using long-read nanopore sequencing, generate genome assemblies and placing the sequenced strains inside the phylum using phylogenetic analyses with genome-scale data.

2 Materials and methods

2.1 Laboratory work

2.1.1 Cultivation of fungal strains

Figure 2.1 shows a visual overview of the cultivation steps.

2.1.1.1 Origin of strain materials

The strains cultured in this project were previously cultured in the study presented in Kosa et al. (2018). The strains were originally obtained in lyophilized form or in agar plates or agar slants from the Université de Bretagne Occidentale Culture Collection (UBOCC; Plouzané, France), the All-Russian Collection of Microorganisms (VKM; Moscow, Russia), the Czech Collection of Microorganisms (CCM; Brno Czech Republic), the Food Fungal Culture Collection (FRR; Commonwealth Scientific and Industrial Research Organisation, North Ryde, Australia) and the American Type Culture Collection (ATCC; VA, USA). From the original materials, there were produced stock cultures that were used for inoculating cultures in this project. The stock cultures were stored at -80°C and consisted of cryovials containing asexual fungal spores (sporangiospores) in glycerol-water solution.

2.1.1.2 Agar plate cultivation

The strains were cultured on either potato dextrose agar (39 g/L potato dextrose agar powder) or malt extract agar (20 g/L malt extract broth powder and 20 g/L agar powder). Cultures where the biomass was processed with the dry ice biomass handling protocol had 40 mg/L chloramphenicol added to the medium. Each strain was cultured on two plate replicates. Each plate was inoculated with three droplets of stock spore suspension deposited with single-use

plastic loops. The plates were incubated at 25°C for a minimum of 4 days in a VWR INCU-line 68R incubator (VWR International). Table 2.1 and Table 2.2 lists cultivation time and agar type used of the cultured strains. A full list of strains used in the project, including strains that did not have their DNA extracted, is available in the Appendix in Table 6.3.

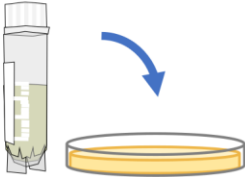

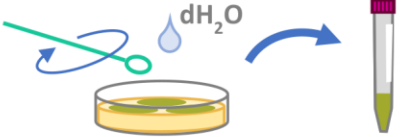
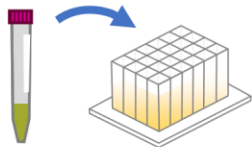


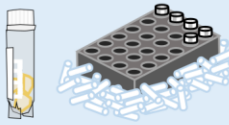

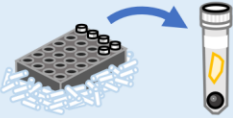


Cultivation and biomass handling steps	Biomass handling protocol	
 <p>Inoculation with stock spore suspensions.</p>  <p>3 droplets per plate; 2 plate replicates per strain.</p>	<p>Room temperature protocol</p> <p>No antibiotic added to the medium.</p>	<p>Dry ice protocol</p> <p>40 mg/L chloramphenicol added to the medium.</p>
<p>Incubation at 25°C for 4-11 days.</p>		
<p>Making fresh spore suspensions.</p> 	<p>No antibiotic added to the medium.</p>	
<p>Inoculation of liquid cultures.</p> 		
<p>Incubation at 25°C for 2-8 days in rotating shaker.</p>		
<p>Washing biomass with dH₂O by vacuum filtering.</p> <p>Transferring biomass to storage tubes.</p> 	<p>On lab bench for 5-60 min.</p>  <p>Room temperature</p>	<p>On dry ice for 20-80 min.</p>  <p>-78.5 °C</p>
<p>Freezer storage.</p>	<p>-21°C</p>	<p>-80°C</p>
	<p>Transfer to micro-tube and kept at room temperature for 60-90 min.</p> 	<p>On dry ice for 5 min before transfer to microtube.</p> 
<p>Preparation for DNA extraction.</p>	 <p>Addition of lysis buffer and steel ball.</p>	 <p>Addition of lysis buffer.</p>
<p>Crushing of mycelium and DNA extraction.</p>		

Figure 2.1. Steps in cultivation and biomass handling. The figure shows an overview of cultivation and handling of biomass samples and highlights the differences between the two protocols used.

2.1.1.3 Obtaining spore suspensions

Following incubation at 25°C, spore suspensions were created from agar plates with sporulating colonies. 2 to 6 mL of autoclaved, distilled water was added to the agar plates, the spores were mixed into the water by scraping the colonies with single-use plastic loops. The spore suspensions were transferred to 15 mL SuperClear Centrifuge tubes (VWR International) and stored at 4°C until inoculation of liquid cultures.

2.1.1.4 Liquid cultures in microtiter plates

Duetz system deep well 24-square microtiter plates (EnzyScreen, Heemstede, Netherlands) were used for the liquid cultures. The strains were cultivated in liquid medium containing 20 g/L malt extract broth powder. 40 mg/L chloramphenicol was added to the medium of cultures where the dry ice biomass handling protocol was used. In each well, 7 mL medium was inoculated with 20 µL freshly made spore suspension. A minimum of two well replicates were inoculated per strain. The cultures were incubated at 25°C for a minimum of 2 days in Kuhner Shaker X Climo-Shaker ISF1-X shaking incubator (Kuhner AG, Birsfelden, Switzerland) with shaking rate of 245 rpm or in a Thermo Scientific™ MaxQ™ 4000 Benchtop Orbital Shaker (Thermo Fisher Scientific, Waltham, Massachusetts, the United States) with shaking rate of 400 rpm. The incubation times of the sequenced strains are listed in Table 2.1 and Table 2.2, and full lists of all fungi that were cultured and had their DNA extracted are found in Table 6.1 and Table 6.2 in the appendix.

Table 2.1. Sequenced strains cultured without antibiotic and where biomass was processed with the room temperature biomass handling protocol.

Sequen-cing sample number	Strain	Agar medium	Agar plate incubation time (days)	Liquid culture incubation time (days)
1	<i>Mucor lanceolatus</i> UBOCC-A-109193	MEA	5	2
2	<i>Mortierella hyalina</i> UBOCC-A-101349	PDA	5	2
3	<i>Mucor hiemalis</i> UBOCC-A-109197	MEA	5	2
4	<i>Umbelopsis ramanniana</i> CCM F-622	PDA	5	2
5	<i>Lichtheimia corymbifera</i> CCM 8077	MEA	4	2

MEA = malt extract agar.

PDA = potato dextrose agar.

Table 2.2. Sequenced strains cultured with 40 mg/L chloramphenicol added to the medium and where biomass was processed with dry ice biomass handling protocol.

Sequencing sample number	Strain	Agar medium	Agar plate incubation time (days)	Liquid culture incubation time (days)
6	<i>Mucor plumbeus</i> UBOCC-A-109204	MEA	4	3
7	<i>Absidia glauca</i> CCM 450	MEA	4	3
8	<i>Mucor racemosus</i> UBOCC-A-111127	MEA	4	3
9	<i>Mucor plumbeus</i> UBOCC-A-111132	MEA	5	2
10	<i>Rhizopus stolonifer</i> CCM F-445	MEA	6	2
11	<i>Amylomyces rouxii</i> CCM F-220	MEA	5	2
12	<i>Lichtheimia corymbifera</i> VKM F-513	MEA	7	2

MEA = malt extract agar.

PDA = potato dextrose agar.

2.1.2 Biomass washing

Biomass in the form of mycelium pellets was transferred from the liquid cultures using single-use plastic loops and/or glass pipettes to a Whatman No. 1 filter paper (GE Whatman, Maidstone, UK) on a vacuum flask setup. The biomass was vacuum filtered and washed with distilled water to remove growth medium.

2.1.3 Biomass handling and storage

Two protocols for biomass handling and storage were used, here named the room temperature protocol and the dry ice protocol. A visual overview of the differences between the protocols is shown in Figure 2.1.

2.1.3.1 Room temperature biomass handling protocol

After washing, the biomass was scraped off the filter paper and transferred to a 15 mL SuperClear Centrifuge tubes (VWR International). The biomass tubes were kept at room temperature for up to an hour before storage at -21°C. Prior to DNA extraction, approximately 50 mg biomass was transferred to a 2 mL screw cap centrifuge tube (2.0 SC Micro Tube PCR-PT, REF 72.693.465, SARSTEDT AG & Co. KG, Nümbrecht, Germany) and kept at room temperature for 60 to 90 minutes.

2.1.3.2 Dry ice biomass handling protocol

After washing, the biomass was transferred from the filter paper to 1.8 mL cryovials, plastic tubes that withstand low temperatures (CryoPure Tube white, REF 72.379, SARSTEDT,

Nümbrecht, Germany,). Each cryovial was filled with approximately 50 mg biomass and immediately placed in a metal block cooled with dry ice (-78.5°C). The biomass was kept on dry ice until storage at -80°C and between storage and initiation of DNA extraction.

2.1.4 DNA extraction

2.1.4.1 Mycelium disruption

Approximately 50 mg biomass was transferred to a 2 mL screw cap centrifuge tube together with a 5 mm stainless steel ball (Stainless Steel Beads, 5 mm (200), Cat No./ID: 69989, QIAGEN, Venlo, Netherlands) and 600 µL lysis buffer without SDS. The final concentrations of the lysis buffer (after later addition of SDS) was 0.2 M Tris-HCl pH 8.0 (Trizma® hydrochloride solution 1 M, Sigma-Aldrich, St. Louis, Missouri, the United States), 0.25 M NaCl (Sodium chloride solution BioUltra, for molecular biology, ~5 M in H₂O, 71386, Sigma-Aldrich, St. Louis, Missouri, the United States), 0.025 M EDTA (Ethylenediaminetetraacetic acid disodium salt solution 0.5 M, E7889, Sigma-Aldrich, St. Louis, Missouri, the United States) and 0.5% SDS. To break down RNA and thus avoid co-extracting RNA with the DNA, 1 µL RNase A (1.25 mL, 100 mg/mL, Mat. No. 1007885, Lot No. 154017058, QIAGEN, Venlo, Netherlands) was added to each sample tube, which was then placed on ice. To crush the mycelium, each screw cap tube was placed in a TissueLyser bead mill (QIAGEN, Venlo, Netherlands) which was run at 30 Hz for 1 minute, two to three times. Between every 1-minute run, the tube placements were changed to the opposite side of the adapter (closest or furthers apart from the TissueLyser) to ensure that all samples were disrupted with approximately equal force. After mycelium disruption, 65 µL lysis buffer without SDS and 35 µL 10% SDS was added to each tube which was placed on a rotating rack running at 50 rpm. The SDS, which lyses membranes and cellular structures and dissolves lipids and proteins, was added after using the TissueLyser to prevent foaming during mycelium disruption.

2.1.4.2 Phenol-chloroform extraction

The approximately 700 µL lysis buffer and cell lysate solution was transferred to a 2 mL PhaseLock tube (5PRIME Phase Lock Gel Light 2 mL, Cat# 2302820, Quantabio, Beverly, Massachusetts, the United States) before 700 µL of a 7:3 buffer saturated phenol and chloroform mix was added and thoroughly mixed with the sample by placing the PhaseLock tube on a rotating rack for 30 minutes at 50 rpm at room temperature. The phenol in the

organic phase denatures proteins and dissolves proteins and lipids, separating them from the water-soluble DNA which remains in the aqueous phase. The sample was then centrifuged at 6000 g for 15 minutes at 4 °C. The chloroform gives extra density to the organic phase, ensuring a clear phase separation where the organic phase is at the bottom even if the aqueous phase has high salt concentrations. The PhaseLock gel in the PhaseLock tubes has a density between that of the aqueous and organic phase and creates a barrier between the phases during centrifugation. This enables one to more easily collect the aqueous phase without also collecting organic phase or cell fragments and proteins that collect at the interphase.

550 µL of the aqueous phase was transferred to a fresh PhaseLock tube and 550 µL chloroform was added. Chloroform dissolves nonpolar proteins, lipids and phenol that still are in the aqueous phase. The sample was mixed on a rotating rack for 10 minutes at 50 rpm before centrifugation at 6000 g for 15 minutes at 4 °C. Aqueous phase, containing DNA, was transferred in volumes of 50 µL from the PhaseLock tube to a fresh 1.5 mL DNA LoBind Safe-Lock tube (cat no 022431021, Eppendorf, Hamburg, Germany).

2.1.4.3 Isopropanol precipitation

To precipitate the DNA, 0.6 volumes of isopropanol cooled on ice was added per volume of aqueous solution. The isopropanol has a lower dielectric constant than water and promotes Na⁺ ions from the lysis buffer to bind to the negatively charged phosphate groups on the DNA's backbone. This breaks the hydration shell that water molecules created around the DNA molecules and encourages the DNA molecules to aggregate and precipitate out of the solution. The isopropanol also dissolves remaining phenol and chloroform. The precipitation step was carried out on ice or at -20 °C for 30-60 minutes, or overnight at -20 °C. Following the precipitation step, the sample was centrifuged at 8000 g for a minimum of 15 minutes at 4 °C to collect the DNA precipitate in a pellet.

2.1.4.4 Ethanol wash

The supernatant containing isopropanol and lysis buffer was discarded. The sample tube was spun on a mini centrifuge to collect droplets of supernatant which was then discarded. 1 mL of fresh 70% ethanol was added to remove excess salts from the DNA pellet. The tube was rotated slowly to wash the walls and lid without dislodging the pellet. The tube was centrifuged at 8000 g and 4 °C for a minimum of 5 minutes before the supernatant was

discarded and the wash step repeated. The tube was then placed upside down on a tissue paper until the ethanol had evaporated.

2.1.4.5 Elution

The washed, dried DNA was eluted in 50 μL elution buffer containing 10 mM of Tris-HCl pH 8.0. The tube was gently rotated before it was placed at 4 $^{\circ}\text{C}$ overnight or for up to several days to elute the DNA in the elution buffer. After quality checks of the extracted DNA, the DNA samples were stored at -21 $^{\circ}\text{C}$.

2.1.5 DNA quality controls

2.1.5.1 NanoDrop – purity and concentration

1 μL of each sample was measured on a NanoDrop 8000 microvolume spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, the United States) to detect presence of contaminating compounds and to estimate nucleic acid concentrations. NanoDrop measures the UV absorption at wavelengths where nucleic acids and common contaminants absorb light.

2.1.5.2 Gel electrophoresis – DNA integrity

To get an overview of how much of the extracted genomic DNA was intact and how much that had been broken into smaller fragments, extracted DNA was quality checked using gel electrophoresis. Small 0.5% agarose gels were made from 0.2 g agarose powder and 40 mL TAE (Tris-acetate-EDTA) buffer. The gels were made with 0.7 μL Sybr Safe DNA-binding dye. To each well, there was added either 1 μL of Thermo Scientific GeneRuler 1 kb DNA Ladder (#SM0311, Thermo Fisher Scientific, Waltham, Massachusetts, the United States), 10 μL ladder mix containing 0.25 μL Thermo Scientific GeneRuler High Range DNA Ladder (0.5 $\mu\text{g}/\mu\text{L}$, 50 μg , #SM1351, Lot 00418526, Thermo Fisher Scientific, Waltham, Massachusetts, the United States), 8 μL nuclease free dH_2O (Axiom™ Water 96 rxn, REF 901522, Thermo Fisher Scientific, Waltham, Massachusetts, the United States) and 2 μL loading dye, or 11 μL sample mix containing 2 μL loading dye, 1 μL eluted DNA and 8 μL nuclease free dH_2O . When using normalized input concentration, 50-100 ng DNA was used and nuclease free dH_2O was added to give a total volume of 11 μL . The gels were run at 50 V for 2 hours and photographed using a gel doc imaging system.

The loading dyes used in gel electrophoresis assays were NEB Gel Loading Dye Purple (6X) (#B7024S, New England BioLabs Inc., Massachusetts, the United States) and Thermo Scientific TriTrack DNA Loading Dye (6X) (R1161, Thermo Fisher Scientific, Waltham, Massachusetts, the United States).

2.1.5.3 QuBit – concentration

Samples that were judged as potential candidates for sequencing on basis of DNA integrity seen on agarose gels and NanoDrop measurements were measured on a QuBit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, the United States) to gain more accurate estimates of DNA concentrations. The samples were measured with Qubit™ dsDNA BR Assay Kit per manufacturer's instructions. The QuBit assay works by adding a fluorescent dye that binds selectively to DNA and calculating the concentration from the measured signal with the help of a standard curve made with samples of known DNA concentrations.

2.1.5.4 Selection of samples for sequencing

Each sample of extracted DNA was given a score from 0 to 10 based on DNA integrity observed on agarose gels, DNA concentration measured with QuBit and purity of the DNA measured on NanoDrop. The 12 samples with the highest scores were selected to be sequenced. The selected samples, numbered from 1 to 12, and the strains that originally were used to inoculate plate cultures were: 1 *Mucor lanceolatus* UBOCC-A-109193, 2 *Mortierella hyalina* UBOCC-A-101349, 3 *Mucor hiemalis* UBOCC-A-109197, 4 *Umbelopsis ramanniana* CCM F-622, 5 *Lichtheimia corymbifera* CCM 8077, 6 *Mucor plumbeus* UBOCC-A-109204, 7 *Absidia glauca* CCM 450, 8 *Mucor racemosus* UBOCC-A-111127, 9 *Mucor plumbeus* UBOCC-A-111132, 10 *Rhizopus stolonifer* CCM F-445, 11 *Amylomyces rouxii* CCM F-220 and 12 *Lichtheimia corymbifera* VKM F-513.

2.1.6 DNA extraction – disruption method trials

Prior to choosing TissueLyser disruption as the mycelium crushing method used in this project, two other methods were tested and compared. The sections below describe the steps where the protocols deviated from the TissueLyser protocol.

2.1.6.1 Method one: bead-beading in TissueLyser

DNA was extracted from 58 mg *Rhizopus oryzae* CCM 8075 mycelium. RNase A was not added to the lysis buffer.

2.1.6.2 Method two: cold disruption - mortar and pestle cooled by dry ice

DNA was extracted from 67 mg *Rhizopus oryzae* CCM 8075 mycelium. The mycelium was crushed using a mortar and pestle cooled by dry ice in ethanol (99%, denatured) bath and transferred to a microcentrifuge tube using a spatula. 700 µL lysis buffer was added to the mortar to collect remaining mycelium powder and transferred to the microcentrifuge tube. The lysis buffer contained SDS from the start and no RNase A was added. The tube contents were thawed, mixed gently, and transferred to a PhaseLock tube to which 700 µL of 25:24:1 Phenol:Chloroform:Isoamyl (saturated with 10 mM Tris, pH 8.0 and mM EDTA, Sigma-Aldrich, St. Louis, Missouri, the United States) was subsequently added. The tube was placed on a rotating rack for 60 minutes at 50 rpm at room temperature, before DNA extraction was completed as described above.

2.1.6.3 Method three: ultra-cold disruption - mortar and pestle cooled by liquid nitrogen

DNA was extracted from 1.463 g *Umbelopsis ramanniana* VKM F-502 mycelium, 1.466 g *Cunninghamella echinulata* VKM F-531 mycelium and 1.819 g *Cunninghamella blakesleeana* CCM F-705 mycelium. The mycelium was ground to a fine powder with a cooled pestle in a mortar cooled and filled 1/3 with liquid nitrogen. The mycelium powder was transferred to a 50 mL screw cap centrifuge tube and 10 mL lysis buffer was added. The lysis buffer contained SDS from the start and no RNase A was added. 7 mL phenol and 3 mL chloroform was added to the 50 mL tube which was left at room temperature for 1 hour and mixed by inversion every 30 minutes. The tube was centrifuged at 6000 g for 15 minutes before the supernatant was transferred to a fresh 50 mL tube and an equal volume of chloroform was added. The tube was mixed gently by inversion and centrifuged at 6000 g for 5 minutes. The supernatant was transferred to a fresh 50 mL tube and 0.6 volumes of cooled isopropanol was added per volume of supernatant. The tube was placed on ice for 30 minutes to allow the DNA to precipitate before a centrifugation step for 15 minutes at 6000 g. The isopropanol supernatant was carefully decanted off and the DNA was washed with 70% ethanol and air dried for 5 minutes. The washed and dried DNA pellet was resuspended in 500 µL elution buffer and transferred to a fresh 1.5 mL tube. The DNA extraction of the three samples was carried out in the laboratory facilities of Oslo Mycology Group at the Department of Biosciences at the University of Oslo. The eluted DNA was placed on ice for 4 hours until storage at -21 °C.

2.1.6.4 RNase A treatment

Following DNA extraction, aliquots of 10 µL of extracted DNA was moved to fresh 1.5 mL centrifuge tubes and treated with RNase. 1 µL of RNase A was added to each sample aliquot, which were then incubated at 37 °C for 70 minutes.

2.1.6.5 Quality control of DNA from the disruption method trials

NanoDrop measurements, and a gel electrophoresis assay with input nucleic acid concentrations normalized to 50-100 ng, were performed on samples that had received RNase A treatment and on samples that had not.

2.1.7 Sequencing library preparation

To allow for the twelve selected DNA samples to be pooled and sequenced simultaneously, a unique barcode sequence was attached to the DNA molecules of each sample. This allows the sequencing platform software to sort the reads according to sample of origin. The sequencing library was prepared according to the protocol “Native barcoding genomic DNA (with EXP-NBD104, with EXP-NBD114, and SQK-LSK109)” (Oxford Nanopore Technologies, 2019). The kits used for library preparation were Ligation Sequencing Kit (SQK-LSK109) and Native Barcoding Expansion 1-12 (EXP-NBD104) (Oxford Nanopore Technologies, Oxford, UK).

2.1.8 Sequencing on ONT PromethION

The DNA was sequenced on a PromethION sequencer, on two FLO-PRO002 flow cells (Oxford Nanopore Technologies, Oxford, UK). The first flow cell was used for one sequencing run lasting 25 hours. The second flow cell was used for two sequencing runs lasting 24 and 72 hours and was washed with a nuclease flush step between the runs so that it could be reused and more sequencing library could be loaded.

2.1.8.1 Data acquisition

Sequencing data was acquired with MinKNOW Core version 3.6.1. The MinKNOW software produced data in the form of FAST5 files containing raw signal data.

2.1.8.2 Basecalling and read quality filtering

The raw signal data in FAST5 files was basecalled with Guppy version 3.2.8 (Oxford Nanopore Technologies, Oxford, UK), which produced FASTQ files containing sequencing

reads and one summary file for each sequencing run. Reads were sorted according to barcode sequence by Guppy (demultiplexing). Guppy was used to filter the reads based on the mean read basecall quality of each read. Reads that passed the quality threshold was saved to disk for use in genome assemblies. The quality threshold corresponds to a Phred score of 7, meaning a basecall accuracy of 0.8005.

2.2 Bioinformatics

2.2.1 Read statistics

Read statistics for the sequencing runs and total read statistics were calculated by PycoQC (Leger & Leonardi, 2019) using the command `pycoQC --summary_file summary_file_produced_by_guppy.txt --html_outfile report_to_be_created.html`, and read statistics for each barcoded sample was calculated by NanoPlot (De Coster et al., 2018) using the commands `NanoPlot --readtype 1D -f pdf --fastq reads_file.fastq` and `NanoPlot --readtype 1D -f pdf --summary_file summary_file_produced_by_guppy.txt`. Plots showing cumulative yield and number of active pores over time were created by NanoPlot and the plot showing read quality over time was created by PycoQC. To make read length and quality distribution plots of the twelve barcoded samples. BBmap version 37.48 (Bushnell) was used to sort reads according to read length in bins of 50 bp and according to read quality in bins of 1 Phred score unit. The command `bbmap/37.48/readlength.sh in=reads_file.fastq out=readlength_histogram.txt bin=50 -Xmx8g max=180000` was used for read length and `bbmap/37.48/reformat.sh in=reads_file.fastq aqhist=average_read_quality_histogram.txt` was used for read quality.

2.2.2 Genome assembly

For each sample, an assembly was built from reads that passed the quality threshold. The *de novo* long read assembler Flye, version 2.7-b1585 (Kolmogorov et al., 2019), was used to make the assemblies. Flye was called with the command `flye --nano-raw reads_file.fastq --genome-size 40m --meta`. The expected genome size was set to 40 Mb and the metagenomic option was activated to compensate for possible

contamination. Flye produces FASTA files containing contigs and in some cases scaffolds comprising contig sequences connected with 100 N's.

2.2.3 Assignment of taxonomic identity to assembly contigs using BLAST

To get an overview of what type of organism each contig originated from and thus detecting contamination, and to see which genes or genomic regions the contigs correspond to in the genomes of close relatives, each contig sequence was queried against NCBI's Nucleotide database using BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990). The Nucleotide database contains genome, gene and transcript sequences that are collected from several sources, including GenBank, RefSeq and Protein Data Bank (PDB). The algorithm BLASTn megablast was used to find highly similar database sequences. The expectation value cutoff was set to $1e^{-25}$ to ensure only accurate hits. The top BLAST hit was considered the taxonomic identity of the contig.

The contigs of sample number 1 to 6 were queried against the May 22nd of 2020 Nucleotide database using local BLAST. Because of lack of disk space after downloading the Nucleotide database locally, the contigs of sample 7 to 12 were instead queried against the April 17th of 2014 Nucleotide database on a remote server. The remote server was accessed through an installation of the Galaxy bioinformatics platform (Afgan et al., 2018) at NMBU, through the NeLS portal (Norwegian e-Infrastructure for Life Sciences).

2.2.4 Exploratory quality control of assemblies

The assembly quality control software QUASt (Gurevich et al., 2013) was used to obtain assembly statistics (N50, overall read coverage and G+C content). QUASt was called with the command `quast.py --nanopore reads_file.fastq --no-snps --no-sv assembly_file.fasta`.

2.2.5 Exploratory alignment of assemblies to reference genomes

If the genome assembly of a known close relative of the sequenced strain was available, it was used as a reference genome in QUASt for comparison (size, G+C content and sequence alignment to reference genome). QUASt was run with the command `quast.py -r reference_genome.fasta --labels "Sample vs Reference" --nanopore reads_file.fastq --no-snps --no-sv`

assembly_file.fasta. If taxonomic assignment by BLAST or initial reference alignment in QUASt suggested that the assembly contained different organisms than intended, alignment softwares were used to identify possible sources of the sequenced DNA. In addition to QUASt, the software MashMap for fast approximate alignments (Jain et al., 2017; Jain et al., 2018) and the software IGV, Integrative Genomics viewer (Robinson et al., 2011; Thorvaldsdóttir et al., 2012) were used. The assemblies used as reference genomes are listed in Table 2.3.

Table 2.3. List of genome assemblies used as reference genomes in alignment software.

Sample	Assembly used as reference genome	GenBank accession version/JGI assembly version and short name (reference)	Software used for alignment
1	<i>Ralstonia pickettii</i> 12J	GCA_000020205.1	QUASt
3	<i>Mucor irregularis</i> B50 <i>Mucor irregularis</i> B7584	GCA_000587855.1 GCA_000697435.1	MashMap MashMap
5	<i>Lichtheimia corymbifera</i> 008-049	GCA_000697175.1 (Chibucos et al., 2015)	QUASt
6	Sample 9 assembly	na	QUASt
7	<i>Absidia glauca</i> CBS 101.48 substr. RVII-324 met-	GCA_900079185.1	QUASt
8	<i>Mucor racemosus</i> B9645 <i>Mucor racemosus</i> f. <i>racemosus</i> UBOCC-A-109155 Sample 11 assembly	GCA_000697255.1 (Chibucos et al., 2016) Mucrac1 (Lebreton et al., 2020) na	QUASt, MashMap MashMap MashMap
9	Sample 6 assembly	na	QUASt
10	<i>Rhizopus stolonifer</i> NRRL 66455	v1.0, Rhisto1	QUASt
11	<i>Amylomyces rouxii</i> NRRL 5866 <i>Mucor lusitanicus</i> CBS277.49 <i>Mucor circinelloides</i> 1006PhL <i>Mucor circinelloides</i> B8987 <i>Mucor circinelloides</i> WJ11 Sample 8 assembly	Amyrou1 draft genome (James, 2018, used with permission) v2.0 Mucci2 (Corrochano et al., 2016) GCA_000401635.1 GCA_000696935.1 GCA_001276145.1 na	QUASt QUASt MashMap MashMap, IGV MashMap MashMap

2.2.6 Estimating expected assembly sizes

For previously sequenced species, the expected assembly size was defined as the median size of available genome assemblies. For species with no publicly available genomic sequence and

where phylogenetic placement was known from literature, the expected assembly size was defined as the mean assembly size of the species in the same phylogenetic clade. For sample 6 *Mucor plumbeus* UBOCC-A-109204 and sample 9 *Mucor plumbeus* UBOCC-A-111132, the expected assembly size was defined as the mean of the two assemblies.

2.2.7 Finding G+C content of contigs

The software SeqKit (Shen et al., 2016) was used to find the G+C (%) content of each contig, by using the command `seqkit fx2tab assemblyfile.fasta --gc --gc-skew --header-line --length -name`. Deviating G+C content can signify that a contig is from a contaminant organism, that it contains a mitochondrial sequence or that it contains repetitive sequences.

2.2.8 Exploratory assembly analysis

The contigs of each assembly was explored by building data frames in R (R Core Team, 2020) of contig characteristics such as length, read coverage, G+C content and taxonomic assignment by BLAST. Contig length was plotted against read coverage and G+C content to get an overview of the contig composition of each assembly.

2.2.9 Contig filtering

Contigs with zero BLAST hits to the Nucleotide database were excluded from the assemblies, except for the assembly of sample 7 *Absidia glauca* CCM 450. Contigs with BLAST hits to bacteria were removed. Contigs with low read coverage compared to the rest of the contigs within the assembly were also excluded. No contig filtering was performed on the assemblies of sample 8 and 11. The final assemblies were written to new files using SeqKits `grep` and `sort` commands:

```
seqkit grep --pattern-file  
list_of_contigs_to_keep.txt assembly_file.fasta >  
final_assembly_file_unsorted.fasta followed by seqkit sort --by-  
length -reverse final_assembly_file_unsorted.fasta >  
final_assembly_file.fasta.
```

2.2.10 Quality control of final assemblies

2.2.10.1 Assembly statistics

QUAST was used to acquire assembly statistics of the final assemblies.

2.2.10.2 BUSCO completeness analysis

The BUSCO software (Seppey et al., 2019) is used to estimate how completely the genome assembly represents the actual genomic sequence. The software searches the genome assembly for a collection of near-universal single-copy orthologs (single-copy genes in current species that descended from genes in the species' last common ancestor). The full collection is expected to be present in the genome. If a BUSCO gene is missing or fragmented, it is assumed to be due to low read coverage or misassemblies, meaning that the assembly is incomplete. If a BUSCO gene is duplicated, it is assumed that the sequence is erroneously represented more than once in the assembly, which can happen when reads are misassembled. BUSCO version 4.0.6 was run with genome mode on the twelve assemblies using the command `busco -m genome -i assembly_file.fasta --lineage_dataset mucoromycota_odb10`. The BUSCO gene set used was `mucoromycota_odb10`, which consists 1614 genes from the OrthoDB catalogue (Kriventseva et al., 2018).

2.2.11 Visualisation of assembly contiguity

The length and read coverage of the contigs of each assembly was visualized using a modified version of the script “contiguity.R” by Nicholls (2019), as seen in (Nicholls et al., 2019). The script uses the package `ggplot2` (Wickham, 2016).

2.2.12 BUSCO phylogenomics

The phylogeny of phylum Mucoromycota was inferred by using genomic data in the form of BUSCO gene sequences. By virtue of being single-copy orthologs, the genes used in the BUSCO completeness analysis may also be used in phylogenetic analyses. All 1614 genes in the `mucoromycota_odb10` gene set were used to infer the phylogeny of 97 genome assemblies including 11 assemblies from this project and 86 assemblies downloaded from NCBI's GenBank database and JGI's Mycocosm portal. The downloaded assemblies are listed in Table 2.1. Only complete gene sequences present in one copy were used in the analyses.

Table 2.4. List of genome assemblies of 86 taxa used in phylogenomic analyses.

Taxon	GenBank accession version/JGI assembly version and short name (reference)
Absidia glauca CBS 101.48 substr. RVII-324 met-	GCA_900079185.1
Absidia repens NRRL 1336	v1.0 Absrep1 (Mondo et al., 2017a)
Actinomucor elegans JCM 22485	GCA_001599635.1

<i>Apophysomyces elegans</i> B7760	GCA_000696995.1
<i>Apophysomyces trapeziformis</i> B9324	GCA_000696975.1
<i>Apophysomyces variabilis</i> NCCPF 102052	GCA_002749535.1 (Prakash et al., 2017)
<i>Aspergillus nidulans</i> FGSC A4 (Outgroup, phylum Ascomycota)	GCA_000011425.1 (Wortman et al., 2009)
<i>Bifiguratus adelaidae</i> AZ0501	GCA_002261195.1
<i>Cokeromyces recurvatus</i> B5483	GCA_000697235.1
<i>Cunninghamella bertholletiae</i> 175	GCA_000697215.1
<i>Cunninghamella bertholletiae</i> B7461	GCA_000697315.1
<i>Cunninghamella elegans</i> B9769	GCA_000697015.1
<i>Diversispora epigaea</i> IT104	GCA_003547095.1
<i>Endogone</i> sp. FLAS 59071	Endsp1 (Chang et al., 2019)
<i>Gigaspora margarita</i> BEG34	GCA_009809945.1 (Venice et al., 2020)
<i>Gigaspora rosea</i> DAOM 194757	GCA_003550325.1
<i>Glomus cerebriforme</i> JS1	GCA_003833025.1
<i>Gongronella</i> sp. w5	GCA_001650995.1
<i>Hesseltinella vesiculosa</i> NRRL3301	v2.0 Hesve2finisherSC (Mondo et al., 2017a)
<i>Jimgerdemannia flammicorona</i> AD002	Jimfl AD 1 (Chang et al., 2019)
<i>Jimgerdemannia flammicorona</i> GMNB39	Jimfl_GMNB39_1 (Chang et al., 2019)
<i>Jimgerdemannia lactiflua</i> OSC166217	Jimlac1 (Chang et al., 2019)
<i>Lichtheimia corymbifera</i> 008-049	GCA_000697175.1 (Chibucos et al., 2015)
<i>Lichtheimia corymbifera</i> B2541	GCA_000697475.1 (Chibucos et al., 2015)
<i>Lichtheimia ramosa</i> B5399	GCA_000738555.1 (Chibucos et al., 2015)
<i>Lichtheimia ramosa</i> B5792	GCA_000697395.1 (Chibucos et al., 2015)
<i>Lichtheimia ramosa</i> JMRC FSU:6197	GCA_000945115.1 (Linde et al., 2014)
<i>Lobosporangium transversale</i> NRRL 3116	v1.0 Lobtra1 (Mondo et al., 2017a)
<i>Mortierella alpina</i> ATCC 32222	GCA_000240685.2 (Wang et al., 2011)
<i>Mortierella alpina</i> B6842	GCA_000507065.1 (Etienne et al., 2014)
<i>Mortierella alpina</i> CCTCC M207067	GCA_001021685.1
<i>Mortierella elongata</i> AG-77	v2.0 Morel2 (Uehling et al., 2017)
<i>Mortierella verticillata</i> NRRL 6337	GCA_000739165.1 (Seif et al., 2005)
<i>Mucor ambiguus</i> NBRC 6742	GCA_000950595.1
<i>Mucor circinelloides</i> 1006PhL	GCA_000401635.1
<i>Mucor circinelloides</i> B8987	GCA_000696935.1
<i>Mucor circinelloides</i> f. lusitanicus MU402	v1.0 Muccir1_3
<i>Mucor circinelloides</i> WJ11	GCA_001276145.1
<i>Mucor endophyticus</i> UBOCC-A-113049	Mucend1 (Lebreton et al., 2020)
<i>Mucor fuscus</i> UBOCC-A-109160	Mucfus1 (Lebreton et al., 2020)
<i>Mucor irregularis</i> B50	GCA_000587855.1
<i>Mucor irregularis</i> B7584	GCA_000697435.1
<i>Mucor lanceolatus</i> UBOCC-A-109153	Muclan1 (Lebreton et al., 2020)
<i>Mucor lusitanicus</i> CBS277.49	v2.0 Mucci2 (Corrochano et al., 2016)
<i>Mucor racemosus</i> B9645	GCA_000697255.1 (Chibucos et al., 2016)
<i>Mucor racemosus</i> f. racemosus UBOCC-A-109155	Mucrac1 (Lebreton et al., 2020)
<i>Mucor velutinous</i> B5328	GCA_000696895.1
<i>Parasitella parasitica</i> CBS 412.66 substr. NGI-315	GCA_000938895.1
ade-	

Phycomyces blakesleeanus NRRL1555	v2.0 Phylb2 (Corrochano et al., 2016)
Rhizomucor miehei CAU432	GCA_000611695.1 (Zhou et al., 2014)
Rhizomucor pusillus	GCA_900175165.2
Rhizophagus cerebriforme DAOM 227022	v1.0 Rhice1_1 (Morin et al., 2019)
Rhizophagus clarus HR1	GCA_003203555.1
Rhizophagus diaphanus MUCL43196	v1.0 Rhidi1
Rhizophagus irregularis A1	v1.0 RhiirA1_1 (Chen et al., 2018)
Rhizophagus irregularis A4	v1.0 RhiirA4 (Chen et al., 2018)
Rhizophagus irregularis A5	v1.0 RhiirA5 (Chen et al., 2018)
Rhizophagus irregularis B3	v1.0 RhiirB3 (Chen et al., 2018)
Rhizophagus irregularis C2	v1.0 RhiirC2 (Chen et al., 2018)
Rhizophagus irregularis DAOM 181602	v1.0 Gloin1 (Tisserant et al., 2013)
Rhizophagus irregularis DAOM 197198	v2.0 Rhiir2_1 (Tisserant et al., 2013)
Rhizopus azygosporus CBS 357.93	GCA_003325435.1 (Gryganskyi et al., 2018)
Rhizopus delemar RA 99-880	GCA_000149305.1 (Ma et al., 2009)
Rhizopus delemar Type I NRRL 21789	GCA_000697155.1
Rhizopus delemar Type II NRRL 21446	GCA_000738605.1
Rhizopus delemar Type II NRRL 21477	GCA_000738585.1
Rhizopus microsporus ATCC11559	v1.0 Rhimi_ATCC11559_1 (Lastovetsky et al., 2016)
Rhizopus microsporus var. chinensis CCTCC M201021	Rhich1 (Wang et al., 2013)
Rhizopus microsporus var. microsporus ATCC52813	v1.0 Rhimi1_1 (Mondo et al., 2017b)
Rhizopus oryzae 97-1192	GCA_000697195.1
Rhizopus oryzae 99-133	GCA_000697135.1
Rhizopus oryzae 99-892	GCA_000697725.1
Rhizopus oryzae B7407	GCA_000696915.1
Rhizopus oryzae HUMC 02	GCA_000697605.1
Rhizopus oryzae Type I NRRL 13440	GCA_000697075.1
Rhizopus oryzae Type I NRRL 18148	GCA_000697095.1
Rhizopus oryzae Type I NRRL 21396	GCA_000697115.1
Rhizopus stolonifer B9770	GCA_000697035.1
Rhizopus stolonifer LSU 92-RS-03	GCA_003325415.1 (Gryganskyi et al., 2018)
Saksenaea oblongisporus B3353	GCA_000697495.1
Saksenaea vasiformis B4078	GCA_000697055.1
Syncephalastrum monosporum B8922	GCA_000697355.1
Syncephalastrum racemosum NRRL 2496	v1.0 Synrac1
Thermomucor indicae-seudaticae HACC 243	GCA_000787465.1
Umbelopsis isabellina B7317	GCA_000697415.1
Umbelopsis isabellina NBRC 7884	GCA_000534915.1

2.2.12.1 Outgroup selection

The strain *Aspergillus nidulans* FGSC A4 of phylum Ascomycota was chosen as outgroup due to being distantly related to the Mucoromycota, being a model organism with a well

characterized genome sequence and due to having a large fraction (823 of 1614 genes) of the mucoromycota_odb10 gene set present in single-copy, complete form. The outgroup was not used to root the phylogenetic trees or constrain the tree topology when building or drawing the trees, allowing the outgroup to be placed according to the data used.

2.2.12.2 Multiple alignment of single-copy BUSCO genes

The amino acid sequences of single-copy, complete BUSCO genes were collected from the BUSCO output folders of the 97 genome assemblies. The sequences of each individual gene were aligned with MAFFT v7.310 (Nakamura et al., 2018). MAFFT was used with the FFT-NS-i (Accurate but slow) strategy and with an iterative refinement method with a maximum of 16 iterations.

2.2.12.3 Gene concatenation supermatrix approach

A gene concatenation phylogenomics approach was used to build a maximum likelihood phylogram of the 97 taxa. In this approach, the multiple alignments of the BUSCO genes were concatenated and a phylogenetic tree was built from the supermatrix containing the 1614 gene alignments.

2.2.12.3.1 Concatenation of alignments

The alignment files produced by MAFFT were concatenated with AMAS version 1.0 (Borowiec, 2016). AMAS was called with the command `python3 /path/bin/AMAS.py concat --in-files alignment_file.aln --in-format fasta --data-type aa --concat-part partition_file_created_by_AMAS.txt --out-format fasta`. AMAS produced a supermatrix alignment file containing the aligned amino acid sequences, and a partition file containing information on each of the 1614 partitions corresponding to each of the 1614 genes. The partition file was modified to include the amino acid substitution model WAG (Whelan & Goldman, 2001) in each partition, so that the partition file was in the format “WAG, gene_1 = start-end; WAG, gene_2 = start-end; [...]”.

2.2.12.3.2 Phylogenetic tree from gene concatenation supermatrix

The phylogeny of phylum Mucoromycota was inferred by The software IQ-tree 2 (Minh et al., 2020), using the maximum likelihood method. The supermatrix alignment file and the

corresponding partition file were used as input. IQ-tree 2 was called with the command `/path/iqtree-2.0.6-Linux/bin/iqtree2 -s supermatrix.aln -p partition_file_created_by_AMAS.txt -B 1000 --sampling GENESITE -m WAG`. IQ-tree 2 was used with Ultrafast Bootstrap Approximation (Hoang et al., 2017) with 1 000 bootstrap replicates. IQ-tree 2 was run with partitioned analysis (Chernomor et al., 2016) with the `-p` option meaning the branch length were linked across partitions and that the rates of evolution were assumed to be constant across partitions. To reduce the chance of wrong branches receiving full support, the resampling strategy GENESITE (Gadagkar et al., 2005), which first resamples partitions and next resamples sites within resampled partitions. No tree topology constraint was applied to the phylogenetic tree. The consensus tree made from the 1 000 bootstrap replicates, with ultrafast bootstrap approximation support values, was chosen as the final tree.

2.2.12.4 Gene tree coalescence approach

As an alternative to the concatenation approach, the coalescence approach to phylogenomics was also used. Phylogenetic trees were made for each individual gene, and the best-scoring maximum likelihood trees of each of the genes were merged to a species tree.

2.2.12.4.1 Gene trees

Phylogenetic trees of each of the 1614 BUSCO genes were inferred using the maximum likelihood method with the software RAxML (Stamatakis, 2014) with 100 bootstrap replicates. RAxML was called with the command `/path/standard-RAxML-master/raxmlHPC-PTHREADS-AVX -T 2 -p 12345 -x 12345 -f a -m PROTGAMMAAUTO --print-identical-sequences -N 100 -s gene_alignment_file.aln`. RAxML was run with the model PROTGAMMAAUTO, meaning that the best amino acid substitution model is automatically selected by the software and that the heterogeneity of substitution rates among sites follows a gamma distribution. When RAxML encountered alignment files with columns containing solely ambiguous amino acid characters (B, Z, J, X) and missing data, the software created files with those column removed. For these genes, the reduced alignment files were used to make gene trees instead of using the original files.

2.2.12.4.2 Coalescence of gene trees to create species tree

The best-scoring maximum likelihood trees of each of the genes were concatenated into one file, which was used as input to ASTRAL version 5.7.3 (Rabiee et al., 2019; Zhang et al., 2018) to infer a species tree. ASTRAL was run with the command `java -jar /path/Astral/astral.5.7.3.jar --input all_best_RAxML_gene_trees_for_astral_ML.tree --output astral_coalesced_species_tree_mucoromycota.tree -t 1`. The branch support values were calculated as the quartet score (Sayyari & Mirarab, 2016), i.e. the percent of quartet trees (unrooted tree with four leaves) found in the concatenated gene tree file that are present in the species tree.

2.2.12.5 Visualization of phylogenetic trees

The phylogenetic trees were visualized in Interactive Tree Of Life (iTOL) (Letunic & Bork, 2016) and additional information was added in Inkscape (Inkscape Project, 2020). The trees are unrooted and the uppermost taxon in the rectangular trees correspond to the first taxon in the input file. The Plots of BUSCO results were created with the BUSCO companion script “generate_plot.py” (Zdobnov, 2020). The script utilizes R and ggplot2.

3 Results

3.1 Obtaining biomass

Most of the strains that were cultured produced sporulating cultures on agar plates and mycelium pellets in liquid medium. Each liquid culture microtiter well replicate produced up to 640 mg mycelial biomass, usually about 150 mg. Breaking the biomass into smaller pieces was most convenient when frozen, but also possible at room temperature by using tweezers. Figure 3.1.A-I shows a small selection of the diversity of colony morphology in subphyla Mucoromycotina and Mortierellomycotina, while Figure 3.1.M shows six strains in liquid culture, illustrating the various morphologies the mycelia can take in liquid culture.

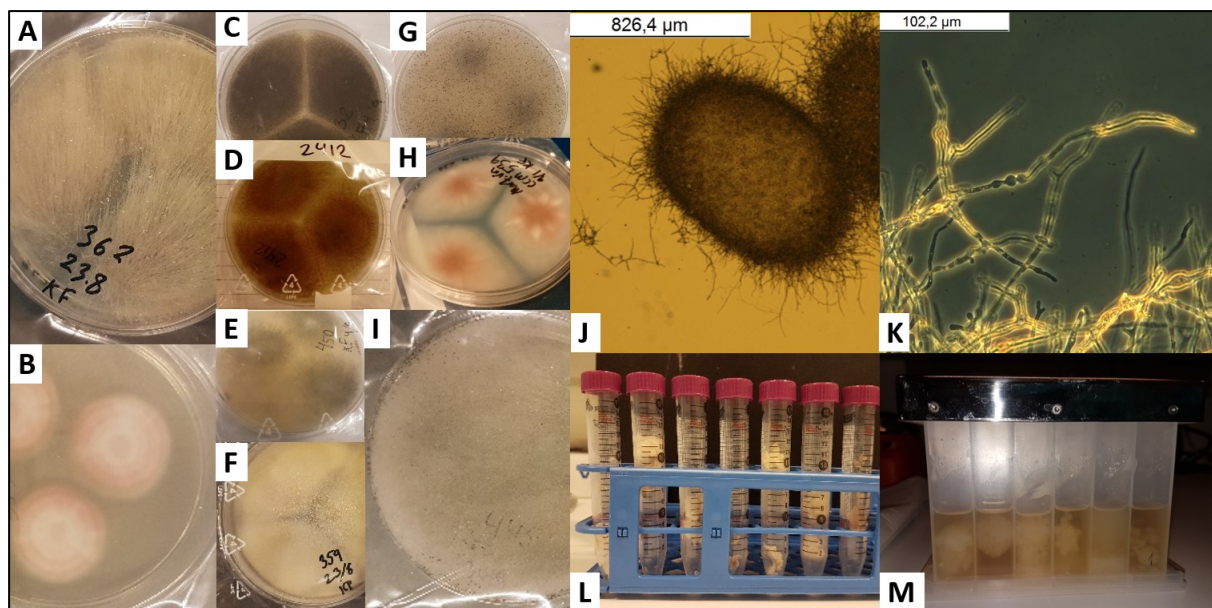


Figure 3.1. Cultures and biomass pellets. A) *Mucor mucedo* UBOCC-A-101362. B) *Umbelopsis ramanniana* CCM F-622. C) *Mucor plumbeus* UBOCC-A-111132. D) *Mucor plumbeus* FRR 2412. E) *Absidia glauca* CCM 450. F) *Mucor hiemalis* UBOCC-A-101359. G) *Rhizopus oryzae* CCM 8116. H) *Umbelopsis vinacea* CCM F-539. I) *Rhizopus stolonifer* CCM F-445. J) *Lichtheimia corymbifera* VKM F-513 biomass pellet from liquid culture, scale bar showing 826.4 µm. K) The edge of the biomass pellet, scale bar showing 102.2 µm. L) Washed biomass pellets before freezer storage. M) Microtiter plate with liquid cultures, six strains visible.

3.2 DNA extraction

3.2.1 Disruption method trials

3.2.1.1 DNA integrity

Fragmentation of genomic DNA into smaller pieces was observed in the extracted samples from all of the three methods tested. Both the TissueLyser method and the cold disruption method gave degraded DNA, indicated by the smears of DNA fragments visible on the gel in Figure 3.2 (samples B and A). Of the two, the TissueLyser method resulted in the highest fraction of degraded to undegraded DNA. However, a considerable fraction of the extracted DNA had high integrity, indicated by the clear bands of high molecular weight DNA. The ultra-cold disruption method gave variable results among the three strains tested. In sample C, most of the genomic DNA was intact, as seen by the clear band of high molecular weight DNA and smear only being faintly visible. In sample E, the lack of a clear band suggest degradation of all the extracted DNA, and no DNA from sample D is visible. The smears near

the bottom of the gel shows that all methods caused RNA to be extracted together with the DNA. The RNase treatment was successful in all samples, shown by the lack of RNA smears in treated samples.

3.2.1.2 Yield

NanoDrop measurements of undiluted, extracted DNA indicated yields of 750-2500 ng/ μ L nucleic acids. The cold disruption method gave about twice as much DNA as the TissueLyser method, and the yields from the ultra-cold disruption method varied between the three strains tested. The measured concentrations are shown in Figure 3.2 C.

3.2.1.3 DNA purity

The purity of the DNA, as indicated by NanoDrop A_{260}/A_{230} and A_{260}/A_{280} absorbance ratios, were equivalent in the samples from the three methods. Measurements are shown in Figure 3.2 A and B. The DNA appeared to be of high purity, with measured ratios falling within or close to the ideal values. The A_{260}/A_{230} ratios were lower in RNase treated samples compared to the samples before treatment, indicating higher concentrations of contaminating protein and phenol and likely caused by evaporation of elution buffer during treatment.

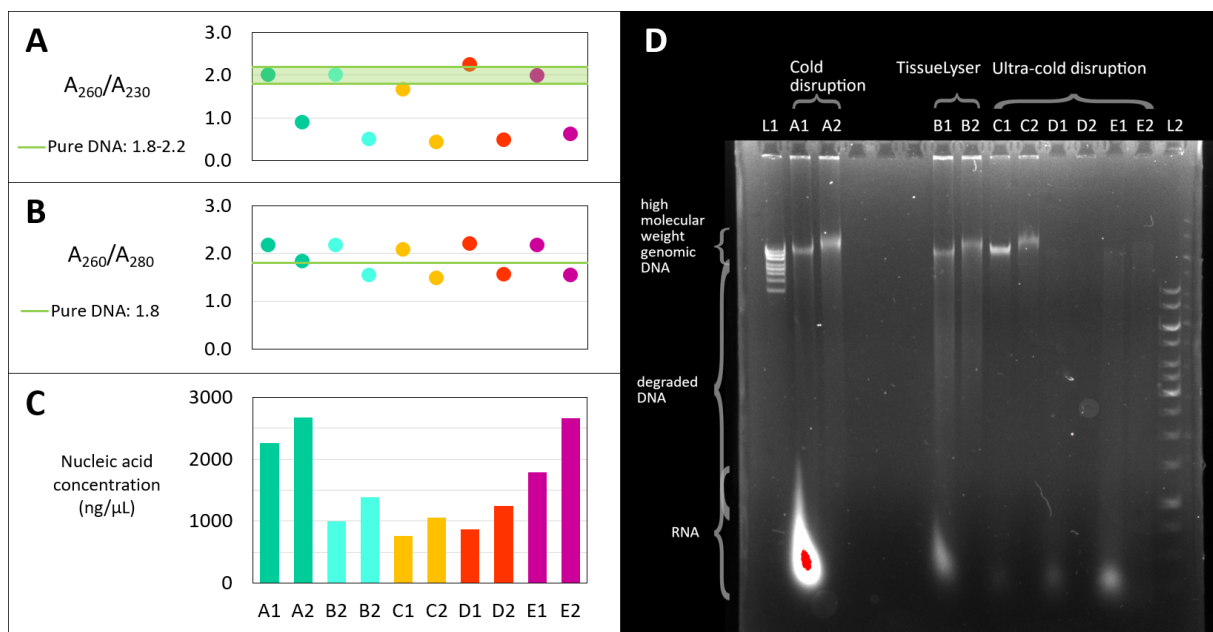


Figure 3.2. Purity, concentration and integrity of DNA extracted in the disruption method trials. A) NanoDrop measurements of A_{260}/A_{230} absorbance ratio. B) NanoDrop measurements of A_{260}/A_{280} absorbance ratio. C) NanoDrop measurements of nucleic acid concentrations. D) Agarose gel showing extracted DNA. Explanation of sample labels: the number 1 denotes the sample before RNase treatment, and the number 2 denotes the sample after RNase treatment. Letters symbolizes sample and strain used: A and B -

Rhizopus oryzae CCM 8075, C - *Umbelopsis ramanniana* VKM F-502, D - *Cunninghamella echinulata* VKM F-531, E - *Cunninghamella blakesleeana* CCM F-705. Ladders used: L1 – GeneRuler HR ladder, L2 - GeneRuler 1kb ladder.

3.2.2 Extraction of genomic DNA from fungal cultures

3.2.2.1 DNA integrity

Gel electrophoresis pictures of samples processed with the room temperature and the dry ice biomass handling protocols are shown in Figure 3.3 and Figure 3.4, respectively. No difference in DNA integrity was observed between the protocols. The extracted DNA showed a high degree of fragmentation in most samples. Only a few of the samples showed clear high-molecular weight band on the gels.

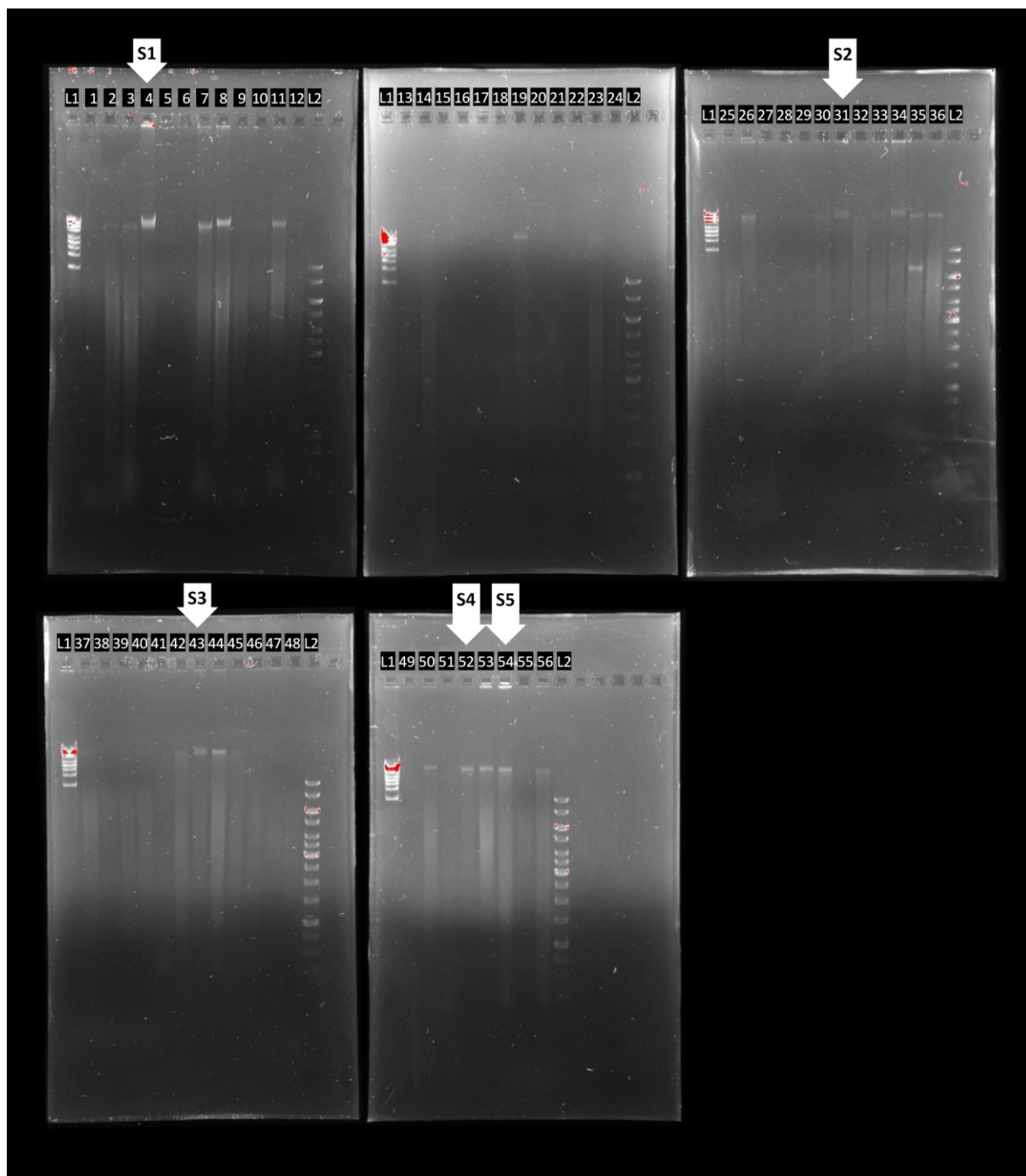


Figure 3.3. Gel pictures showing DNA from samples processed with the room temperature biomass handling protocol. The wells are numbered 1 to 56 according to the order the samples were processed during DNA extraction. The strain of each sample is listed in Table 6.1. Five of the samples were selected for sequencing, here labeled S1 to S5 in white arrows.

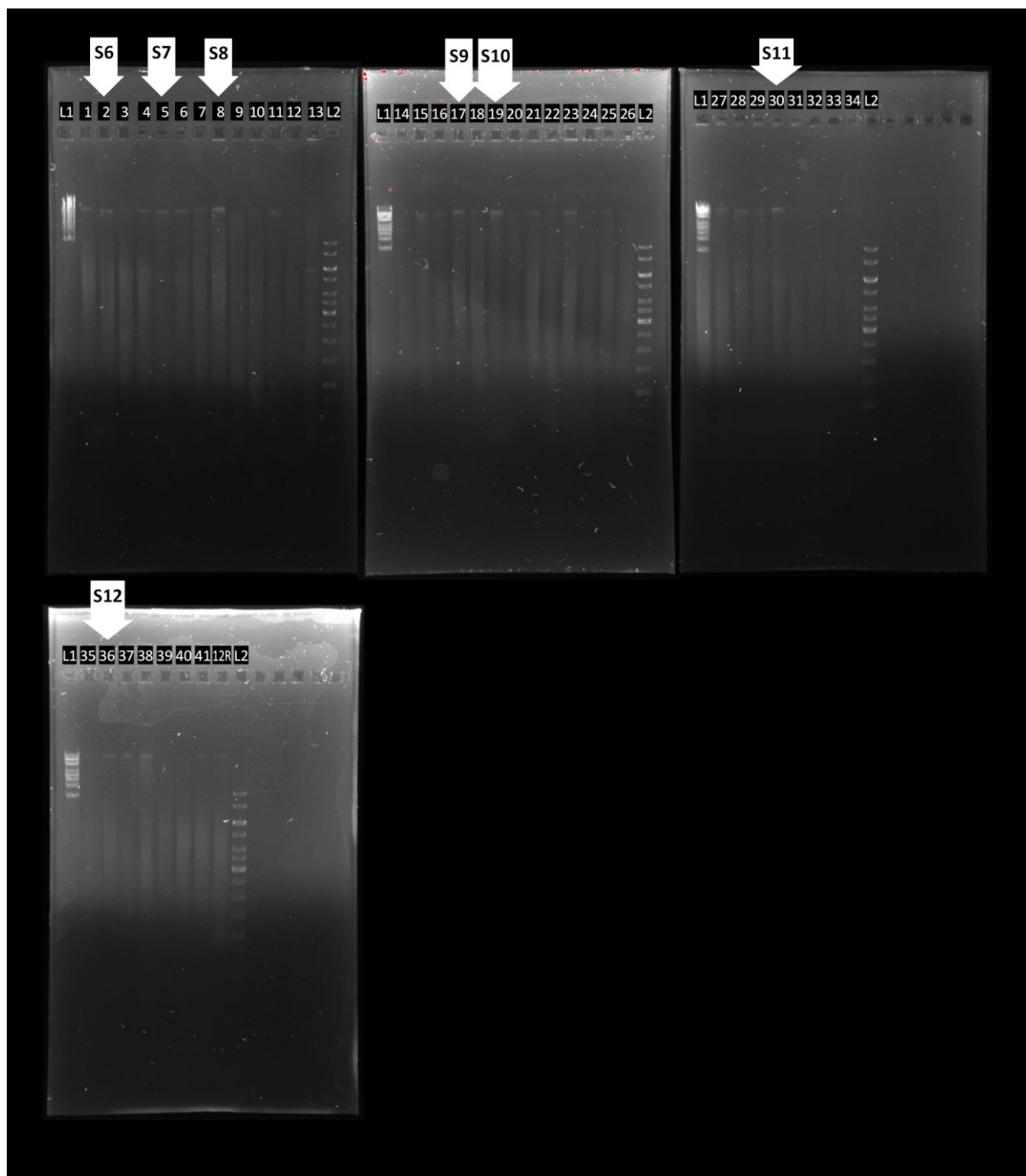


Figure 3.4. Gel pictures showing DNA from samples processed with the dry ice biomass handling protocol. The wells are numbered 1 to 41 according to the order the samples were processed during DNA extraction. The sample 12R is an extraction replicate of the same biomass sample as sample number 12. The strain of each sample is listed in Table 6.2. Seven of the samples were selected for sequencing, here labeled S6 to S12 in white arrows.

3.2.2.2 Purity and yield – all samples

NanoDrop measurements estimated that on average, a biomass sample of around 50 mg generated a yield of 11 µg nucleic acids. The average and median of the yield and the concentration in 50 µL elution buffer is given in Table 3.1. The extracted DNA was measured on NanoDrop to assess the purity. Most samples had a A_{260}/A_{230} absorbance ratio below the ideal range, suggesting presence of contaminating proteins and/or phenolic compounds. Other than slightly lower absorbance ratios on average, there was no difference in DNA purity between biomass handling protocols.

Table 3.1. Yield and nucleic acid concentration of extracted DNA. Yield is calculated from the nucleic acid concentration reported by NanoDrop*. SD = standard deviation.

Biomass handling protocol	Number of samples measured	Nucleic acid concentration (ng/µL)			Yield (ng)		
		Median	Mean	SD	Median	Mean	SD
Room temperature	55	140	216	242	7020	10815	12076
Dry ice	42	201	225	137	10045	11259	6856
Total	97	169	220	203	8430	11007	10153

* NanoDrop is known to overestimate nucleic acid concentrations, so the actual concentrations are most likely lower.

Table 3.2. Purity of extracted DNA. The purity of the extracted DNA is indicated by absorbance ratios reported by NanoDrop. SD = standard deviation.

Biomass handling protocol	Number of samples measured	A_{260}/A_{230}		A_{260}/A_{280}	
		Mean	SD	Mean	SD
Room temperature	55	1.3	0.5	1.9	0.2
Dry ice	42	1.0	0.2	1.7	0.1
Total	97	1.2	0.4	1.8	0.2

3.2.2.3 Purity and yield – sequenced samples

Twelve DNA samples were chosen from the samples with highest DNA integrity as observed on gel and with highest purity as indicated by NanoDrop measurements. Table 3.3 shows the yield and DNA concentration measured with Qubit for the twelve sequenced samples. The Qubit concentrations are also shown in Figure 3.5 C, where the red line marks the recommended concentration in a sample used for sequencing library preparation. Sample 1 and 4 had lower concentrations than the recommendation. Figure 3.5 A and B shows A_{260}/A_{230} and A_{260}/A_{280} absorbance ratios, indicating presence of phenolic compounds, carbohydrates and/or EDTA and presence of protein and/or phenolic compounds in samples with low ratios.

Table 3.3. Yield and concentration of the DNA extracted from samples selected for sequencing. DNA concentration was measured with QuBit and used to calculate yield.

Sample number	DNA concentration (ng/ μ L)	Yield (ng)
1	17	850
2	55.8	2790
3	54.4	2720
4	14.6	730
5	79.4	3970
6	57	2850
7	37.2	1860
8	54	2700
9	42	2100
10	40.2	2010
11	24.6	1230
12	66.4	3320

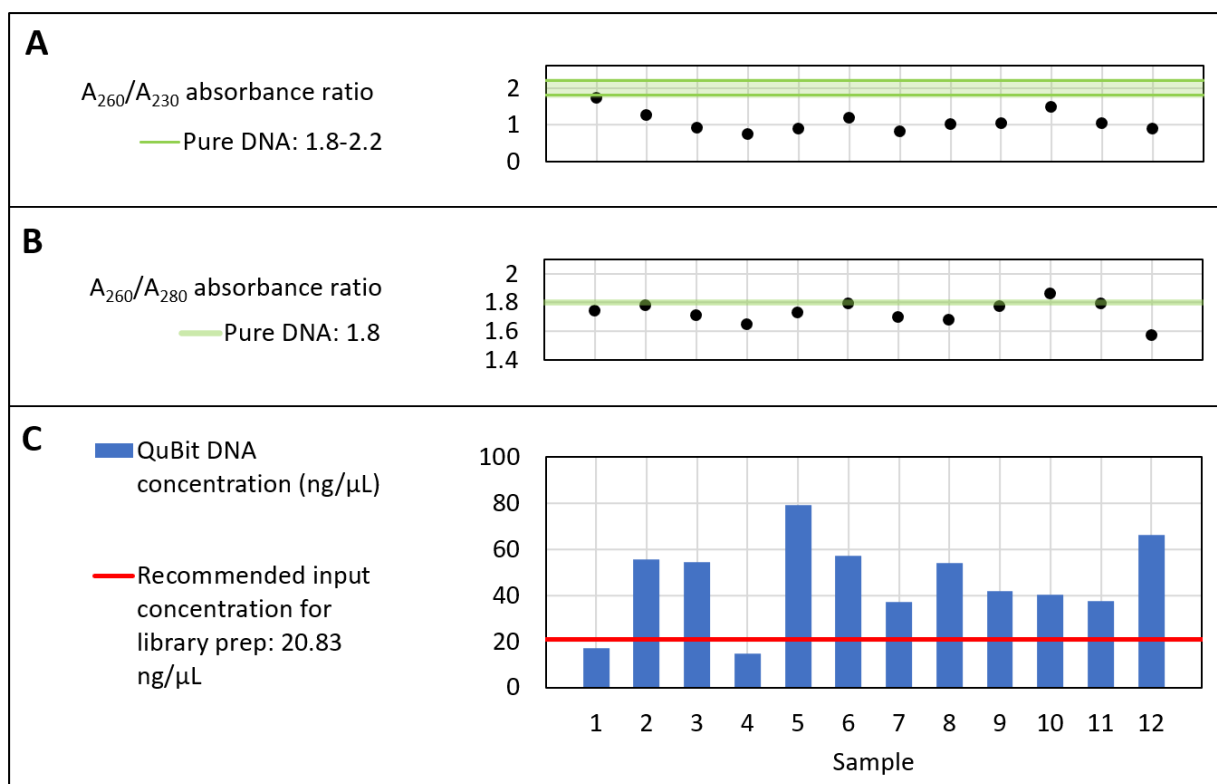


Figure 3.5. Purity and DNA concentration of samples selected for sequencing. A) A_{260}/A_{230} absorbance ratio measured with NanoDrop. The ideal range of absorbance ratio values is marked in green. B) A_{260}/A_{280} absorbance ratio measured with NanoDrop. The ideal absorbance ratio for pure DNA is marked in green. C) DNA concentration in ng/μL measured with QuBit. The recommended DNA concentration for samples used in sequencing library preparation, 20.83 ng/μL, is marked in red.

3.3 Sequencing

3.3.1 Sequencer performance and read statistics

The twelve barcoded samples were sequenced in three multiplexed runs, on two flow cells. The runs produced a total of 62.7 billion bases (Gb) and 18.9 million reads, of which 13 million reads passed the quality threshold and were used downstream for genome assembly. Read statistics and run information for each of the sequencing runs are listed in Table 3.4. Sequencing run number 1 and 2 were stopped after 25 and 24 hours, respectively, at which point 9.3 and 7.8 million reads had been produced. Figure 3.6 shows cumulative yield in reads and change in number of active nanopores over time for the three sequencing runs. The number of active pores dropped from around 1700 at the beginning of the run to around 700 and to below 600 pores. The flow cell used in run 2 was washed and used for a third sequencing run with a new load of sequencing library. The flow cell wash was not successful

in recovering many pores from an inactive state (see Figure 3.6 B and C), but the flow cell produced an additional 1.7 million reads during the first 30 hours of the run, after which the number of active pores had dropped close to zero. Figure 3.7 shows how the quality of the produced reads changed over experiment time. The mean quality of total reads dropped over the course of each sequencing run, to below 7 in run 3 when the number of active pores dropped towards zero.

Table 3.4. Sequencing run information and read statistics per run. Read statistics are shown for total produced reads and for reads that passed the quality threshold (an average Phred quality score of 7 or above across the read).

Run information				Read statistics				
Sequencing run	Flow cell	Active channels*	Run length (h)	Total reads (reads that passed the quality threshold)				
				Bases (Gb)	Reads (M)	Median read length (bp)	Read length N50 (bp)	Median read quality
Run 1	First	2529	25	29.3 (22.2)	9.3 (6.5)	2480 (2790)	4650 (4760)	9.1 (10.2)
Run 2	Second	2391	24	26.8 (21.5)	7.8 (5.8)	2770 (3040)	4830 (4930)	9.8 (10.7)
Run 3	Second	1924	72	6.6 (4.6)	1.8 (1.2)	2860 (3130)	5280 (4930)	8.3 (9.6)
All runs	na	na	121	62.7 (48.3)	18.9 (13.4)	2640 (2930)	4790 (4890)	9.3 (10.3)

* A PromethION flow cell has up to 3000 channels, each containing a sensor and four nanopores, of which one pore can be active at any given time.

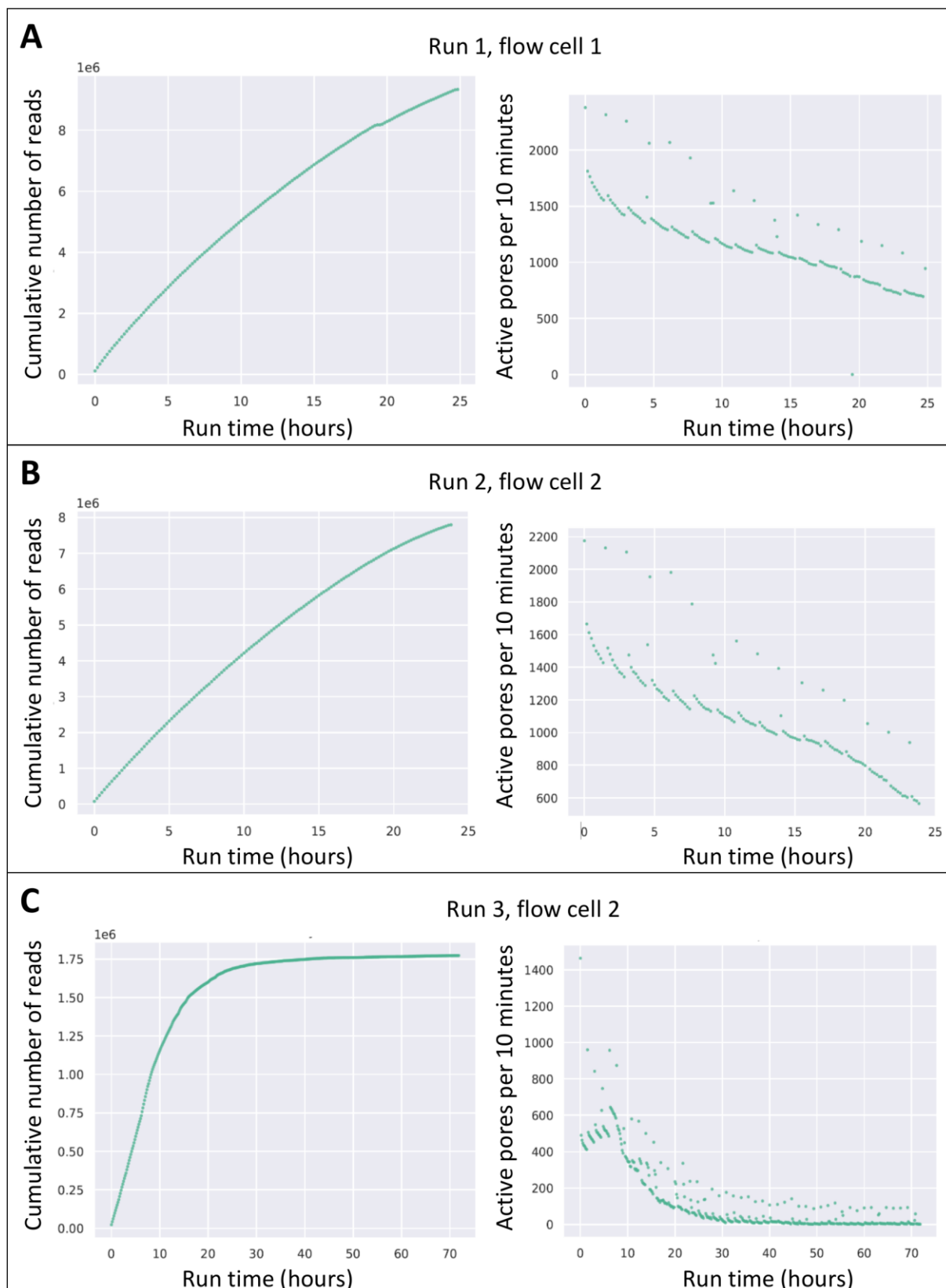


Figure 3.6. Cumulative number of reads produced and number of active nanopores over time. Note that the plots for each of the sequencing runs have axes with different scales. A)

Sequencing run 1, 25 hours. B) Sequencing run 2, 24h hours. C) Sequencing run 3, 72 hours.

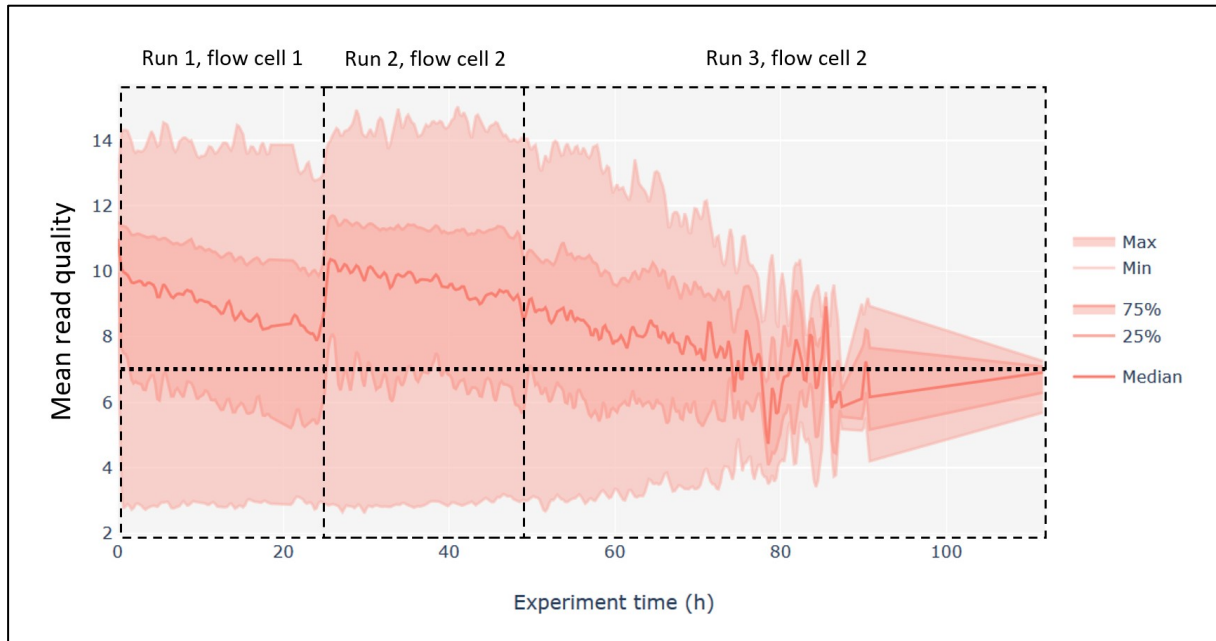


Figure 3.7. Read quality over experiment time. The figure shows the read quality median and the 25th and 75th percentile. Read quality is given as the average Phred quality score across the read. The dotted horizontal line shows the quality threshold (Phred score at seven or above). The dashed line rectangles mark the sequencing runs. The first 25 hours on the x axis represent run 1, the next 24 hours represent run 2 and the last 72 hours on the axis represents run 3.

3.3.2 Read statistics of barcoded samples

Twelve barcoded samples were sequenced in the three PromethION runs. On average, a total of 4 billion bases and 1.1 million reads were produced per sample, of which a total of 1 million reads (3.8 billion bases) were used to make genome assemblies. The median and longest read length varied between the samples, from the shortest in sample 5 with a median of 1208 bp and a maximum read length of 98 kb to the longest in sample 1 with a median of 3849 bp and a maximum read length of 166 kb. Read statistics of each of the samples are shown in Table 3.5 and the distribution of read length and read quality score for each sample is shown in Figure 3.8 and Figure 3.9. The read quality distribution was similar for all samples. 5.4 million reads failed to be sorted to any sample.

Table 3.5. Read statistics for each sequenced sample. Read statistics for total produced reads and reads that passed the quality threshold (average quality score of 7 or above) are given for each of the twelve samples in the barcoded run. Barcode number corresponds to sample number of the sequenced sample.

Barcode (sample number)	Total produced reads (reads that passed the quality threshold)					Longest read to pass the quality threshold (kb)
	Bases (Gb)	Reads (M)	Median read length (bp)	Read length N50 (bp)	Median read quality	
1	3.5 (3.3)	0.7 (0.7)	3849 (3864)	5581 (5579)	10.0 (10.6)	166
2	4.3 (4.0)	1.2 (1.1)	2472 (2572)	6265 (6280)	10.0 (10.6)	145
3	3.4 (3.1)	1.0 (0.9)	2998 (3033)	4726 (4729)	10.0 (10.5)	146
4	4.1 (3.9)	1.1 (1.1)	2504 (2537)	5431 (5433)	10.1 (10.7)	139
5	3.7 (3.4)	1.7 (1.6)	1208 (1236)	3840 (3849)	10.0 (10.5)	98
6	5.3 (5.0)	1.6 (1.5)	2971 (2989)	4203 (4206)	10.6 (11.3)	115
7	3.3 (3.1)	1.2 (1.1)	1915 (1938)	4410 (4411)	10.1 (10.7)	114
8	4.3 (4.1)	1.0 (1.0)	3682 (3693)	5000 (5000)	10.8 (11.4)	167
9	4.2 (4.0)	1.0 (0.9)	3629 (3641)	5000 (5003)	10.8 (11.4)	119
10	3.4 (3.2)	0.8 (0.7)	3446 (3465)	5674 (5676)	10.5 (11.1)	140
11	4.9 (4.5)	1.2 (1.1)	3429 (3449)	4955 (4957)	10.4 (11.0)	117
12	3.7 (3.5)	1.1 (1.0)	3013 (3028)	4188 (4189)	10.2 (10.8)	131
	14.5	5.4	1824	4471	5.0	

Reads not classified to barcode	(3.1)	(0.8)	(3001)	(4982)	(8.8)	147
---------------------------------------	-------	-------	--------	--------	-------	-----

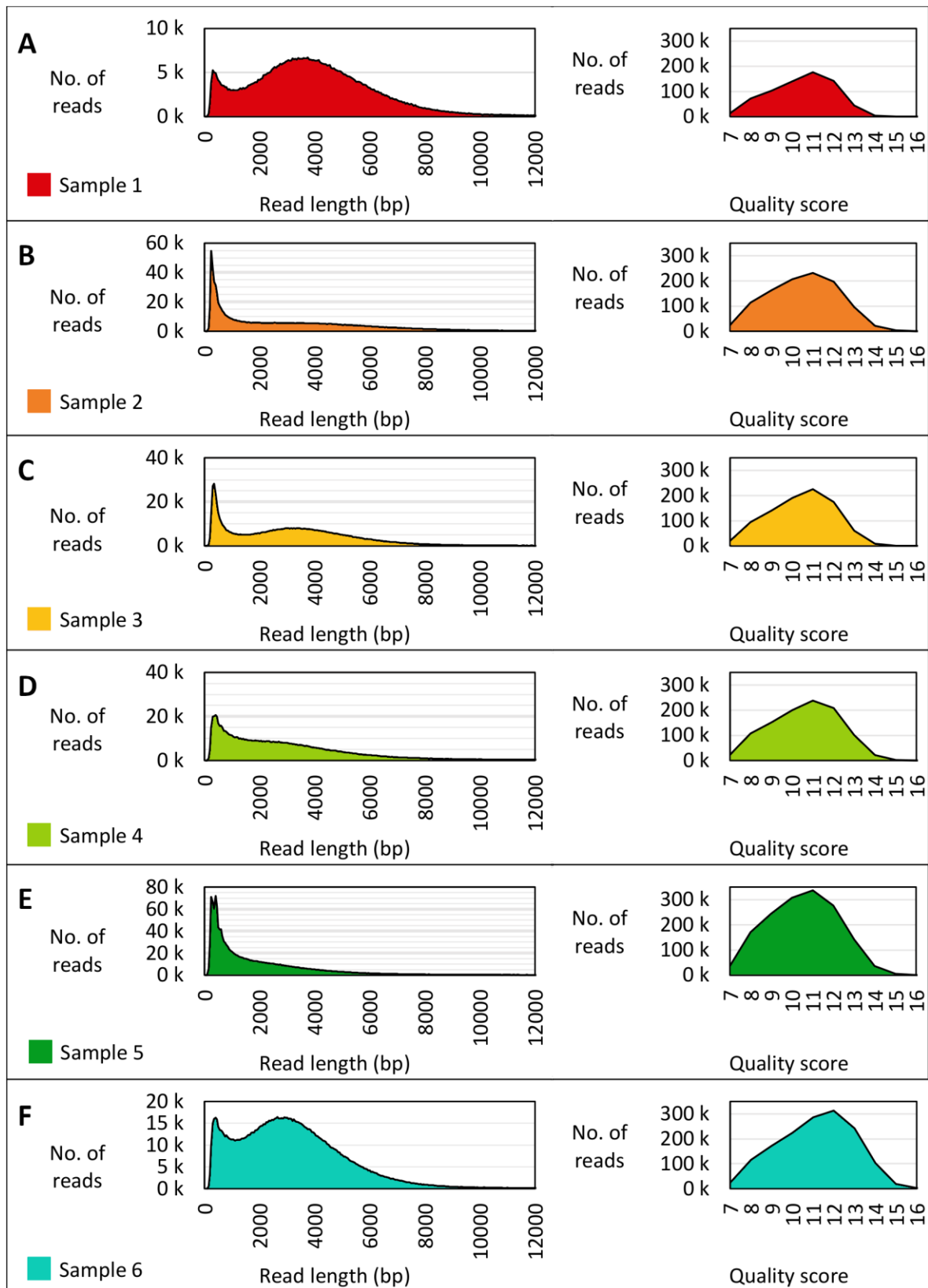


Figure 3.8. Read length and read quality distributions of samples 1 to 6. The read length distribution histograms show number of reads of each read length up to 12 kb, bin size is 50 bp. Note that the y axes of the read length distribution plots vary between samples. Read

quality distribution histograms show the number of reads with average quality score between 7 and 16, bin size is 1. A) Sample 1. B) Sample 2. C) Sample 3. D) Sample 4. E) Sample 5. F) Sample 6.

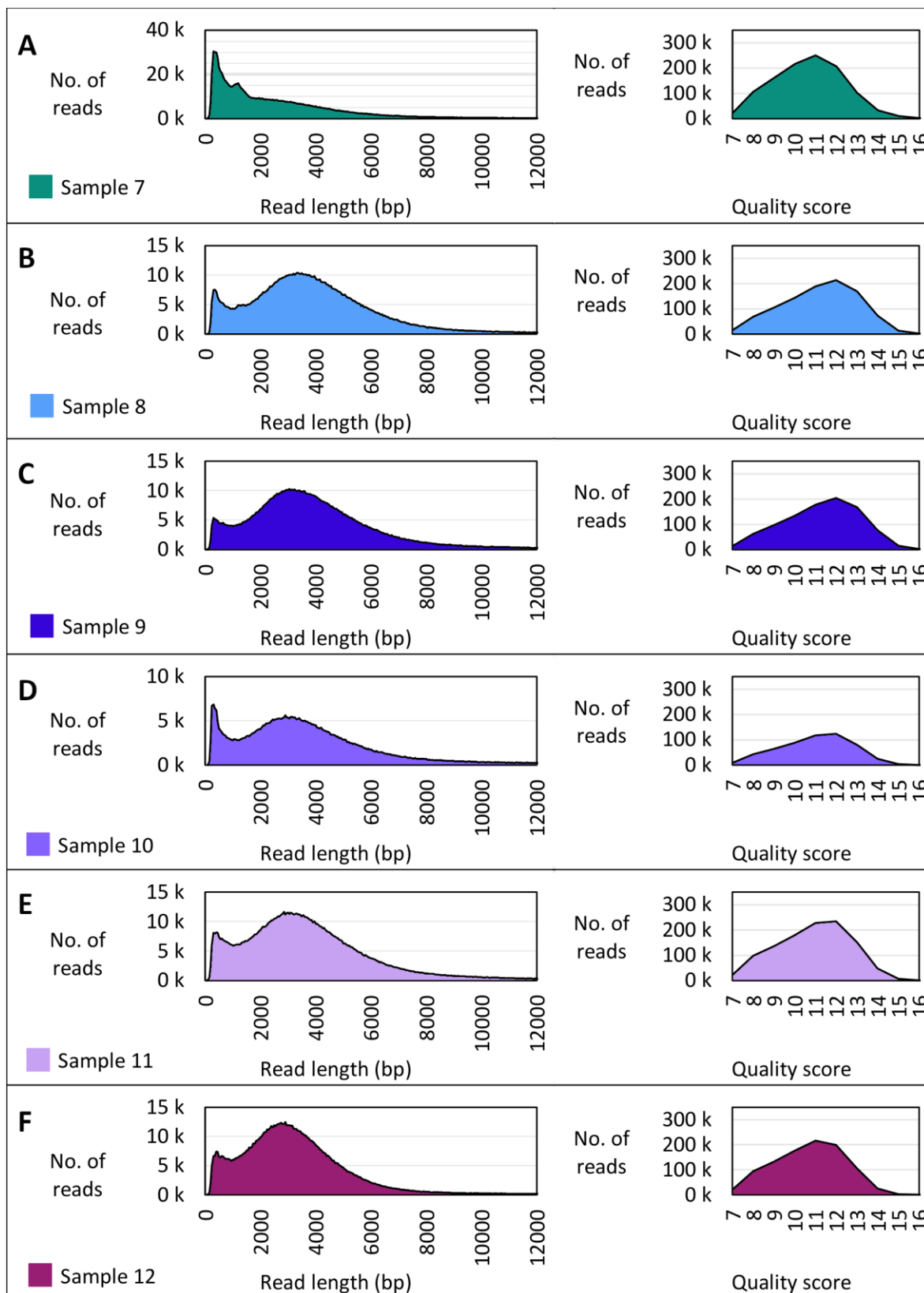


Figure 3.9. Read length and read quality distributions of samples 7 to 12. The read length distribution histograms show number of reads of each read length up to 12 kb, bin size is 50 bp. Note that the y axes of the read length distribution plots vary between samples. Read

quality distribution histograms show the number of reads with average quality score between 7 and 16, bin size is 1. A) Sample 7. B) Sample 8. C) Sample 9. D) Sample 10. E) Sample 11. F) Sample 12.

3.4 Genome assemblies

3.4.1 Genome assemblies of eleven fungal strains

In this project, the genomes of eleven lipid-producing Mucoromycota strains were assembled, including three species that had not been sequenced before. The first assembly of *Umbelopsis ramanniana*, here represented by *Umbelopsis ramanniana* CCM F-622, is 24.34 Mb in size and is highly contiguous with 18 contigs and an N50 value of 1646 kb. The first two assemblies of *Mucor plumbeus*, of the strains UBOCC-A-109204 and UBOCC-A-111132, are 50.95 and 48.93 Mb in length. *Mortierella hyalina* UBOCC-A-101349 is the first of its species to have a genome assembly, which is 47.18 Mb long. The assemblies had BUSCO completeness scores between 82.9% and 97.2%, the lower scores indicating too low read coverage to represent the BUSCO genes and/or genomic regions that are difficult to assemble. Assembly statistics and results of BUSCO completeness analyses are shown in Table 3.6 and Table 3.7. Stacked bar charts showing contig lengths and coverage for each of the assemblies can be found in Figure 3.10.

The assembly of sample 11 was expected to be of *Amylomyces rouxii* CCM F-220, however, the assembly showed little similarity to the only existing assembly of the species (draft assembly of *Amylomyces rouxii* NRRL 5866, used with permission from James (2018)) when aligned in QUAST, as 1232 of the 1417 contigs did not align at all. Alignments to assemblies of *Mucor* species showed that the assembly is likely of a *Mucor circinelloides*.

The assemblies of *Mucor racemosus* UBOCC-A-111127 (sample 8) and of the *Mucor circinelloides* (sample 11) were larger than expected and consisted of shorter and more contigs than the other assemblies. The BUSCO analysis reported a high number of duplicated single-copy genes. Alignment to assemblies of *Mucor circinelloides* and *Mucor racemosus* in QUAST showed that some blocks of genomic sequence aligned twice or more to the corresponding block in the references.

Table 3.6. Assembly statistics of sample 1-6. The table lists assembly statistics and results from the BUSCO completeness analysis for the assemblies of *Mucor lanceolatus* UBOCC-A-109193 (Sample 1), *Mortierella hyalina* UBOCC-A-101349 (sample 2), *Mucor hiemalis* UBOCC-A-109197 (sample 3), *Umbelopsis ramanniana* CCM F-622 (sample 4), *Lichtheimia corymbifera* CCM 8077 (sample 5) and *Mucor plumbeus* UBOCC-A-109204 (sample 6). Expected read coverage is calculated from total number of bases in reads used in the assembly, divided by the expected assembly size.

Sample number	1	2	3	4	5	6
Species	<i>Mucor lanceolatus</i>	<i>Mortierella hyalina</i>	<i>Mucor hiemalis</i>	<i>Umbelopsis ramanniana</i>	<i>Lichtheimia corymbifera</i>	<i>Mucor plumbeus</i>
Strain	UBOCC-A-109193	UBOCC-A-101349	UBOCC-A-109197	CCM F-622	CCM 8077	UBOCC-A-109204
Assembly size (Mb)	0.05	47.18	44.81	24.34	37.86	50.95
Expected size (Mb)	43	45	34	22	37	50
Proportion of expected size	0.0	1.0	1.3	1.1	1.0	1.0
Average read coverage (fold)	39	87	25	175	98	114
Expected read coverage (fold)	77	89	92	175	93	100
Number of contigs	5	44	117	18	54	53
Number of scaffolds	0	1	1	0	1	0
Contig N50 (kb)	12	2565	629	1646	2042	1589
GC content (%)	39.4	48.4	48.5	42.9	43.5	31.5
BUSCO completeness score (%)	0.0	93.0	89.3	90.8	96.3	93.4
Complete and single-copy BUSCO genes (%)	0.0	91.8	88.4	90.3	95.4	92.9
Complete and duplicated BUSCO genes (%)	0.0	1.2	0.9	0.5	0.9	0.5

Fragmented BUSCO genes (%)	0.0	0.7	0.7	1.2	0.6	1.3
Missing BUSCO genes (%)	100.0	6.3	10.0	8.0	3.1	5.3

Table 3.7. Assembly statistics of sample 7-12. The table lists assembly statistics and results from the BUSCO completeness analysis for the assemblies of *Absidia glauca* CCM 450, (sample 7), *Mucor racemosus* UBOCC-A-111127 (sample 8), *Mucor plumbeus* UBOCC-A-111132 (sample 9), *Rhizopus stolonifer* CCM F-445 (sample 10), a *Mucor circinelloides* (sample 11) and *Lichtheimia corymbifera* VKM F-513 (sample 12). Expected read coverage is calculated from total number of bases in reads used in the assembly, divided by the expected assembly size.

Sample number	7	8	9	10	11	12
Species	<i>Absidia glauca</i>	<i>Mucor racemosus</i>	<i>Mucor plumbeus</i>	<i>Rhizopus stolonifer</i>	<i>Mucor circinelloides</i>	<i>Lichtheimia corymbifera</i>
Strain	CCM 450	UBOCC-A-111127	UBOCC-A-111132	CCM F-445		VKM F-513
Assembly size (Mb)	42.10	85.90	48.93	33.08	53.63	32.64
Expected size (Mb)	49	66	50	40	46	37.00
Proportion of expected size	0.9	1.3	1.0	0.8	1.2	0.9
Average read coverage (fold)	75	49	93	112	119	116
Expected read coverage (fold)	64	63	80	80	98	94
Number of contigs	88	1392	22	50	1417	28
Number of scaffolds	2	0	1	2	0	0
Contig N50 (kb)	1645	130	2891	1423	97.531	2446
GC content (%)	44.5	32.8	31.6	36.0	39.2	43.5
BUSCO completeness score (%)	94.4	96.5	94.0	82.9	97.2	95.5
Complete and single-copy BUSCO genes (%)	89.4	39.2	93.6	80.0	79.9	94.4

Complete and duplicated BUSCO genes (%)	5.0	57.3	0.4	2.9	17.3	1.1
Fragmented BUSCO genes (%)	0.8	0.4	1.4	3.0	0.3	0.8
Missing BUSCO genes (%)	4.8	3.1	4.6	14.1	2.5	3.7

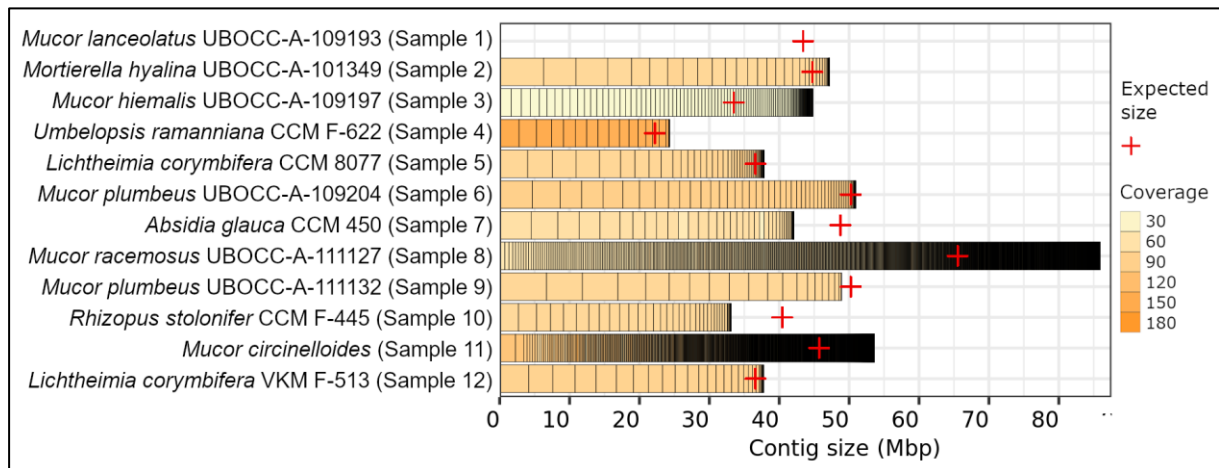


Figure 3.10. Assembly contiguity plot. The stacked bar charts show the lengths of the contigs (bars with black outline) in each of the twelve genome assemblies. Average read coverage of each contig is indicated by the color gradient. The expected assembly sizes are marked with red crosses.

3.4.2 Genome assembly of a bacterial contaminant

The assembly of sample 1, *Mucor lanceolatus* UBOCC-A-109193, contained five contigs with BLAST hits to relatives of the strain, and five high-coverage contigs with BLAST hits to *Ralstonia pickettii* 12J and 12D. Alignment to *Ralstonia pickettii* 12J in QUAST confirmed it as a close relative to 12J, and alignment of bacterial contigs from sample 3 to the *Ralstonia pickettii* of sample 1 strongly indicated that the same bacterium was present in both assemblies. The assembled *Ralstonia pickettii* genome is 5 Mb long and consists of two chromosomes and three plasmids. The assembly has a G+C content of 63.7%, N50 of 3.5 Mbp and average read coverage of 379x.

3.5 BUSCO phylogenomics

3.5.1 Gene concatenation approach

3.5.1.1 Phylogeny of phylum Mucoromycota

Figure 3.11 shows a phylogram of taxa in phylum Mucoromycota sampled widely from the available genome assemblies in NCBI's GenBank database and JGI's Mycocosm resource, as well as the strains sequenced in this project. The branch lengths are proportional to amino acid substitution rate. Branch support values are shown in Figure 3.12, a rectangular representation of the same phylogram.

For subphyla Mortierellomycotina and Glomeromycotina the phylogeny is in agreement with the current assignments to family in NCBI's Taxonomy database (Federhen, 2012), and for subphylum Mucoromycotina the phylogeny is in agreement with the family concepts provided by Hoffmann et al. (2013). The genus *Apophycomyces* clustered with the genus *Saksenaea*, thus supporting Hoffmann et al. (2013) in including the genus in family Saksenaeaceae. The strain *Mucor racemosus* B9645 was clustered with the Mucoraceae clade, supporting its initial assignment to *Mucor racemosus* (Chibucos et al., 2016) and not the later reassignment to *Rhizopus microsporus* (Gryganskyi et al., 2018). The outgroup species *Aspergillus nidulans* of phylum Ascomycota was placed closest to subphylum Mortierellomycotina.

3.5.1.2 Placement of the sequenced strains

Lichtheimia corymbifera CCM 8077 (sample 5) and *Lichtheimia corymbifera* VKM F-513 (sample 12) clustered together and was placed in a clade of *Lichtheimia corymbifera* and *L. ramosa*. *Umbelopsis ramanniana* CCM F-622 (sample 4) was placed as sister to *Umbelopsis isabellina*. *Mucor hiemalis* UBOCC-A-109197 (sample 3) clustered unexpectedly with *Mortierella hyalina* UBOCC-A-101349 (sample 2), the two of the forming a sister clade to *Mortierella elongata* AG-77, both branches with 100% ultrafast bootstrap support. *Absidia glauca* CCM 450 (sample 7) was grouped with *Absidia glauca* CBS 101.48 substr. RVII-324 met-, the two of them forming a sister clade to *Absidia repens* NRRL 1336. *Rhizopus stolonifer* CCM F-445 (sample 10) was placed with two other of its species, and *Mucor plumbeus* UBOCC-A-109204 (sample 6) and UBOCC-A-111132 (sample 9) clustered with *Mucor racemosus*. *Mucor racemosus* UBOCC-A-111127 (sample 8) was placed with *Mucor*

racemosus with low (72% ultrafast bootstrap) branch support. The *Mucor circinelloides* (sample 11) was grouped with *Mucor circinelloides* with low (74%) branch support.

Figure 3.11. Phylogram of the phylum Mucoromycota based on the concatenated alignment of 1614 single-copy orthologous genes. The depicted tree is the maximum likelihood consensus tree which is made from 1000 bootstrap replicates, and which has branch lengths optimized for the original alignment. The branch lengths are proportional to the average number of amino acid substitutions per site, with the scale bar showing one substitution per site. Branch support values are shown in Figure 3.12. Clades are labeled with family assignments according to Hoffmann et al. (2013) for Mucoromycotina and NCBI'S Taxonomy database for Mortierellomycotina and Glomeromycotina (Federhen, 2012).



Figure 3.12. Rectangular phylogram of the phylum Mucoromycota based on the concatenated alignment of 1614 single-copy orthologous genes. The maximum likelihood consensus tree is made from 1000 bootstrap replicates. The branch support values is given

as ultrafast bootstrap (UFBoot) supports. The clade should be relied on if the support is $\geq 95\%$ (Hoang et al., 2017). The branch lengths are proportional to the average number of amino acid substitutions per site, with the scale bar showing one substitution per site. Clades are labeled with family assignments according to Hoffmann et al. (2013) for Mucoromycotina and NCBI'S Taxonomy database for Mortierellomycotina and Glomeromycotina (Federhen, 2012).

3.5.2 Gene tree coalescence approach

Phylogenetic trees were inferred for each of the 1614 single-copy orthologs and then coalesced to form a species tree with branch support values indicating level of agreement between gene trees. Figure 3.11 shows the resulting phylogenetic tree presented as a cladogram, which also shows the BUSCO completeness analysis results for all of the 97 assemblies.

3.5.2.1 Comparison with the concatenation phylogeny

The two phylogenies are for the most part in agreement. One exception is in the species *Rhizopus stolonifer*, where the strain *Rhizopus stolonifer* LSU 92-RS-03 diverged at the base of the *Rhizopus* clade, rather than clustering with the other two *Rhizopus stolonifer* strains. Another exception is in the clade of *Mucor plumbeus* and *Mucor racemosus*. In the coalesced species tree, *Mucor plumbeus* UBOCC-A-109204 (sample 6) clustered with *Mucor racemosus* B9645, the two of them forming a sister clade to *Mucor plumbeus* UBOCC-A-111132 (sample 9), instead of the two *Mucor plumbeus* strains forming a clade like in the concatenation phylogeny.

3.5.2.2 BUSCO completeness of Mucoromycota assemblies

Most of the Mucoromycota assemblies have near full completeness score. The genomes assembled in this project has a larger fraction of missing genes than the other assemblies in their respective clades. The assemblies of strains in subphylum Glomeromycotina and in family Endogonaceae of subphylum Mucoromycotina show more missing genes than other taxa, indicating that the BUSCO mucoromycota_odb10 gene set is poorly adapted to these lineages. The number of duplicated BUSCO genes are consistent within the species *Rhizopus oryzae* and *Rhizopus delemar*. A handful of assemblies show substantial numbers of duplicated genes, among them the large and fragmented assemblies of *Mucor racemosus* UBOCC-A-111127 (sample 8) and *Mucor circinelloides* (sample 11).

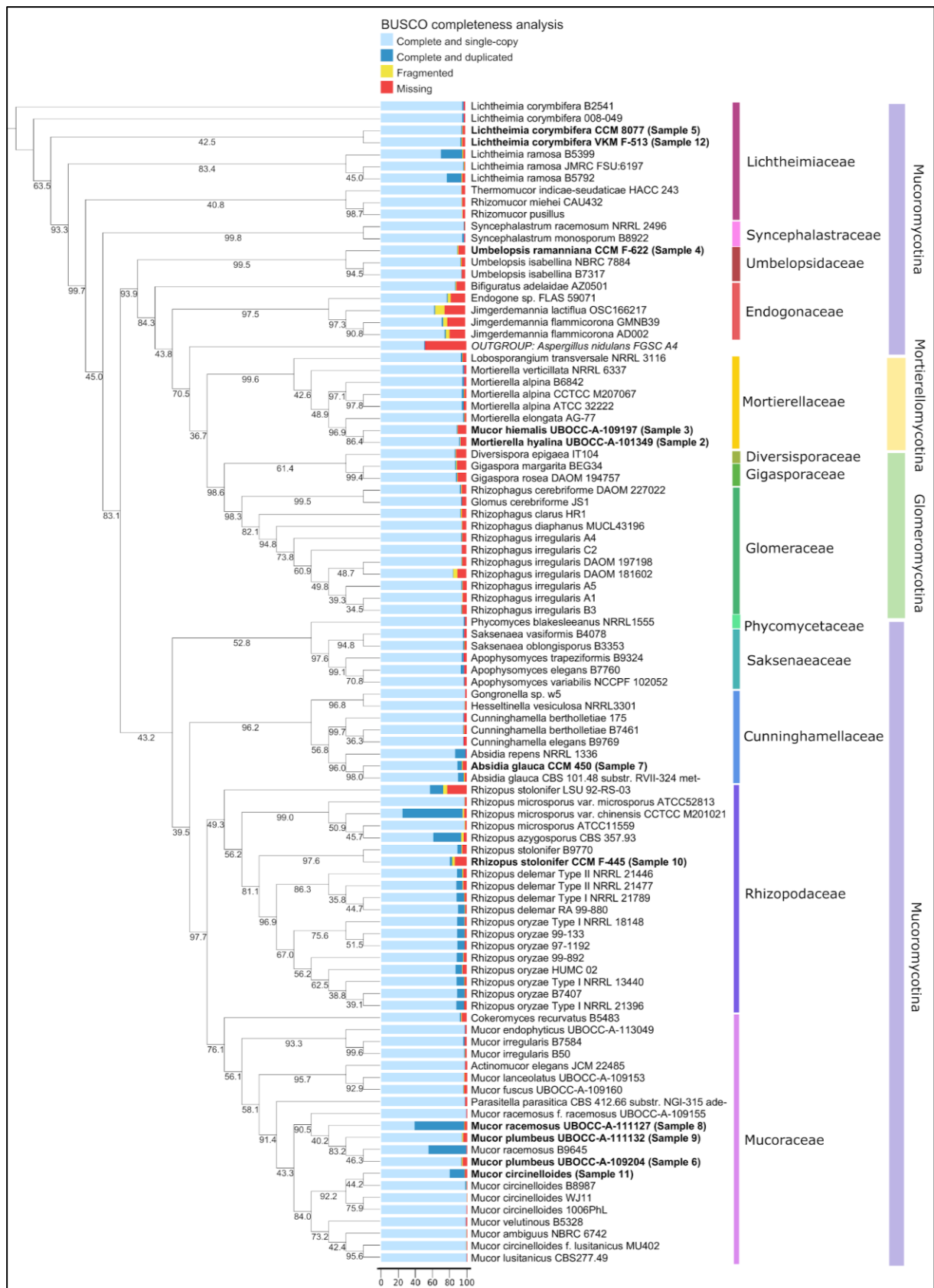


Figure 3.13. Cladogram of 97 taxa in phylum Mucoromycota based on gene trees of 1614 single-copy orthologous genes. The cladogram is a consensus species tree made by

coalescing the best maximum likelihood gene trees for each gene. The gene trees were generated from 100 bootstrap replicates in RAXML and were coalesced with ASTRAL. The branch support values are given as percent of quartet trees in the gene trees that are present in the species tree. A lower support value means that more gene trees are in disagreement. Clades are labeled with family assignments according to Hoffmann et al. (2013) for Mucoromycotina and NCBI'S Taxonomy database for Mortierellomycotina and Glomeromycotina (Federhen, 2012). The bar plot between the tree and the leaf labels show the BUSCO completeness analysis results for all the 97 genome assemblies.

4 Discussion

Phylum Mucoromycotina includes many strains that are great candidates for sustainable lipid production on an industrial scale and for research into lipid accumulation in fungi. Very few strains have however been the subject of genome sequencing and genomic research.

Sequencing more genomes of lipid-producing strains enables further understanding of their lipid metabolism, and can help direct and optimize the development of large-scale, sustainable production of lipids. Sequencing Mucoromycota genomes also enables comparative approaches to learn about the biological mechanisms of the Mucoromycota fungi's roles as industrial producers, decomposers, opportunistic pathogens, plant parasites and plant symbionts.

In this project, fungal DNA was extracted from mycelium using a bead-beating method and sequenced on an Oxford Nanopore PromethION sequencer. The comparatively short nanopore reads generated contiguous assemblies of eleven Mucoromycota strains.

4.1 Extracting DNA from Mucoromycota fungi

Most of the extracted DNA samples in this project showed high degrees of degradation and showed indication of contaminating proteins or phenolic compounds. It is well known that it is very challenging to extract high-quality, pure DNA from filamentous fungi (Fredricks et al., 2005; Muller et al., 1998; van Burik et al., 1998), and the Mucoromycota fungi are no

different. Gryganskyi and Muszewska (2014) report DNA and RNA degradation by the fungi's own nucleases as a hindrance for Mucoromycota sequencing projects. Due to their role as decomposers, the fungi produce high amounts of hydrolytic enzymes including DNases. The dry ice protocol for handling biomass was developed to slow any enzymatic DNA degradation occurring in the mycelium samples. The lack of improvement in DNA integrity when switching from sample handling at room temperature to dry ice suggest that degradation by DNases, if occurring, is for the most part happening after crushing the mycelium.

Two results hints to presence of DNase in the extracted DNA samples. First, two of the three samples extracted with mortar and pestle in with liquid nitrogen showed especially high degrees of degradation (disruption method trials, Figure 3.2). Some strains containing more DNases than others could explain why degradation varied between the samples. Another indication of excess DNase is the high integrity of sample 1 seen in the gel assay (Figure 3.3) which sequencing revealed to contain mostly bacterial DNA and therefore would contain very low amounts fungal DNases. To reduce degradation, Gryganskyi and Muszewska (2014) recommends adding higher amounts of DNase inhibitors to the lysis buffer during DNA extraction and performing additional cleaning steps. Proteinase K can also be added to inactivate DNase, as recommended by Quick and Loman (2018) for nanopore sequencing. Adding the inhibitor EDTA to the elution buffer can protect the DNA during storage, though it is not suitable for applications that use polymerases (Lopata et al., 2019) or transposases (Ason & Reznikoff, 2004).

Bead-beating causes shearing of DNA (Quick & Loman, 2018) and may have caused a substantial amount of the DNA degradation observed in this project. Bead-beating has been used successfully by others to obtain undegraded DNA from fungi: Muller et al. (1998) obtained high-molecular weight DNA from filamentous fungi and yeasts and Inglis et al. (2018) reported undegraded DNA from fungal mycelium and plant tissues. However, in spite of obtaining undegraded DNA, Inglis et al. recommend to instead grind mycelium in a mortar cooled with liquid nitrogen when the DNA will be used for long-read sequencing. The mortar and pestle method is recommended for both PacBio (Pacific Biosciences, 2020) and Nanopore long-read technologies (Quick & Loman, 2018). Some of the DNA samples showed clear bands of genomic DNA in the gel electrophoresis assay, meaning that it could

be possible to use bead-beating and obtain a high-quality genome assembly. This could be achievable by making more replicates of fungal cultures and pooling the extracted DNA before performing size-selection steps to remove short DNA molecules.

4.2 Read length

Due to degradation of the extracted DNA, the nanopore reads acquired in this project were shorter than other reads produced when sequencing fungi. Panthee et al. (2018) used an enzyme-based method to lyse yeast cell walls and obtained nanopore reads with a mean length of 8.3 kb. Giordano et al. (2017) used gTUBE to shear yeast DNA to 18 kb, which resulted in reads with mean length between 8.3 and 9.0 kb and N50 values between 9.8 kb and 11.7 kb for the four sequenced strains. Shearing DNA with gTUBE is recommended when working with low amounts of input DNA, as it will create more DNA molecules, meaning that more nanopores will be active with sequencing and more data is generated (Oxford Nanopore Technologies, 2019). Dutreux et al. (2018) extracted DNA from spores with an enzyme-based lysis method and sheared the DNA to 8 kb and 20 kb with gTUBE. The reads of the three plant pathogen isolates had mean read length of 3.5 kb, 5.5 kb and 3.6 kb, N50 values of 6.4 kb, 7.2 kb and 7.1 kb and a maximum read size of 1.6 Mb, 1.8 Mb and 7.1 Mb.

4.3 Genome assemblies

Although they were generated from mostly short nanopore reads, most of the assemblies produced in this project were highly contiguous and comparable with long-read assemblies of related taxa found in NCBI's taxonomy database. The assembly of the *Mortierella* strain BCC40632 a hybrid assembly of nanopore and Illumina reads spanning 49.96 Mb has the same N50 value, 2565 kb, as the 47.18 Mb large assembly of *Mortierella hyalina* UBOCC-A-101349 (sample 2). *Mortierella elongata* AG-77 was also assembled with a hybrid approach, from long PacBio and short Illumina reads. The assembly is 49.85 Mb and the scaffold-N50 is 517 bp. The related *Lobosporangium transversale* NRRL 3116 has an N50 of 673 kb, a size of 42.77 Mb and a higher BUSCO completeness score than sample 2.

Absidia glauca CCM 450 (sample 7) gave an assembly of 42.10 Mb and an N50 value of 1645, which is equivalent to *Absidia repens* NRRL 1336 which was assembled with PacBio long reads, resulting in a 47.42 Mb assembly and an N50 value of 1295 kb. The BUSCO

completeness score is not equivalent, however, the assembly created in this project has more genes missing, indicative of lower base accuracy in the assembly. Increasing read coverage to make up for sequencing errors, or to use a hybrid approach of assembling both error-prone nanopore reads and high-accuracy Illumina short reads could significantly improve the assemblies.

The long and fragmented assemblies of *Mucor racemosus* UBOCC-A-111127 (sample 8) and of the *Mucor circinelloides* (sample 11) are however much less contiguous than their relatives *Mucor lusitanicus* MU402 and *Mucor lusitanicus* CBS 277.49, which are 36.82 Mb and 36.57 Mb long and have N50 values of 4574 kb and 4318 kb and which were sequenced with PacBio and Sanger respectively. The two *M. lusitanicus* strains have nearly 100% of the BUSCO genes complete and in single-copy, while the assemblies of sample 8 and 11 contained large fractions of duplicated genes.

4.3.1 *Mucor racemosus* UBOCC-A-111127 – a possible genome duplication

The assemblies of *Mucor racemosus* UBOCC-A-111127 (sample 8) and of *Mucor circinelloides* (sample 11) were characterized by short and numerous contigs, a larger than expected size and a high number of duplicated BUSCO genes. When aligned to related genomes in QUAST, blocks of the reference genome were sometimes represented twice or more in the assembly of sample 8 or 11. These are all signs of an assembly made from reads of related genomes (Naranjo-Ortiz & Gabaldon, 2020). As Mucoromycota fungi are assumed to be haploid (Gryganskyi & Muszewska, 2014), possible explanations were cross-contamination of cultures or DNA samples, or a genome duplication in the form of autopolyploidization or a hybridization event between two strains or species. Because the assembly of sample 11 did not appear to contain the expected species, cross-contamination was considered the most likely explanation for both samples.

However, in a recent paper by Lebreton et al. (2020), the authors note that the isolate *Mucor racemosus* B9645 (called *Rhizopus microsporus* CDC-B9645 in the paper) show a high degree of gene duplications and that 52% of the BUSCO genes were duplicated (see *Mucor racemosus* B9645 in Figure 3.13). The authors suggest that a whole or partial genome

duplication event has happened, and that a hybridization event between two species could explain why the isolate has been placed differently in different phylogenetic analyses. The isolate was originally identified as a *Mucor racemosus*, later reassigned to *Rhizopus microsporus* by Gryganskyi et al. (2018) based on 192 orthologous genes, and in this project placed in a clade of *Mucor racemosus* and *Mucor plumbeus* strains (Figure 3.12 and Figure 3.13).

Genome duplications event are considered very in fungi. This assumption is false, however, according to Albertin and Marullo (2012), Naranjo-Ortiz and Gabaldon (2020) and Campbell et al. (2016). Albertin and Marillo explain that while there has been research into polyploid *Saccharomyces* yeasts, they were overshadowed by the sheer volume of publications on haploid *Saccharomyces cerevisiae* model strains. According Albertin and Marullo, ancient duplication events are not actually extremely rare in fungi, they have just not been discovered. Even recent duplication events may go unnoticed as fragmented assemblies and bloated assembly sizes are easily overlooked (Naranjo-Ortiz & Gabaldon, 2020). Genome duplications are described in phylum Mucoromycota, specifically in the ancestor of genera *Mucor* and *Phycomyces* (Corrochano et al., 2016) and a recent whole-genome duplication in the genus *Rhizopus* (Ma et al., 2009; Naranjo-Ortiz & Gabaldon, 2020). And since species of *Mucor* can reproduce sexually (Lee & Heitman, 2014), hybridization events are certainly possible. Assembly features demonstrated by sample 8 and 11, however, can also be a consequence of sequencing and assembling DNA from two separate but related organisms. Particular care should be taken when culturing or isolating related strains and when processing DNA samples of related strains, to avoid creating erroneous genome assemblies.

4.3.2 Bacterial sequences in Mucoromycota assemblies

When sequencing a Mucoromycota genome one should take an extra look at bacterial sequences before writing them off as contamination, as many Mucoromycota fungi house endosymbiotic bacteria. While the bacterium sequenced in this project, *Ralstonia pickettii*, is a relative of the endosymbionts, it has not been observed as such and is instead a contaminant. If in doubt, it is useful to look at the gene content of the bacterium, which is often reduced in endosymbiont strains (Bonfante & Desiro, 2017).

4.4 BUSCO phylogenomics

There have been inferred a wide number of phylogenies of Mucoromycota taxa, some of them using one or a few sequences and taxa that were the subject of the specific study, for instance in Mondo et al. (2017b); Wang et al. (2011). These phylogenies often disagree, and while useful for studying the particular genes, are often not useful to study the evolution of the phylum as a whole. Other studies have used marker genes to try to establish a phylogeny of larger parts of the phylum (Tedersoo et al., 2018; Voigt & Wostemeyer, 2001; White et al., 2006) or even in the context of kingdom fungi (Ren et al., 2016) or the whole tree of life (Voigt & Wostemeyer, 2001).

A phylum-wide phylogeny of Mucoromycota genomes was inferred by using a gene concatenation approach with 1614 single-copy orthologous genes. Interestingly, the phylogeny agreed remarkably well with the phylogeny by Hoffmann et al. (2013) which was constructed with Bayesian analysis of only four genes: translation elongation factor 1-alpha, actin, 18S rDNA and 28S rDNA. The placement of species is in high accordance between the trees, as well as the branch lengths indicating number of amino acid substitutions. This demonstrates perhaps that more is not necessarily better in the field of phylogenetics. Indeed, not all of the BUSCO genes were equally informative. For some genes, up to 20 taxa shared an identical amino acid sequence. Performing tests to determine the contribution of each gene and then filtering out uninformative sequences would simplify any future analyses with the gene set and reduce the computational resources needed.

5 Reference list

- Albertin, W. & Marullo, P. (2012). Polyploidy in fungi: evolution after whole-genome duplication. *Proceedings of the Royal Society B: Biological Sciences*, 279 (1738): 2497-2509. doi: 10.1098/rspb.2012.0434.
- Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. (1990). Basic local alignment search tool. *J Mol Biol*, 215. doi: 10.1016/s0022-2836(05)80360-2.
- Bidartondo, M. I., Read, D. J., Trappe, J. M., Merckx, V., Ligrone, R. & Duckett, J. G. (2011). The dawn of symbiosis between plants and fungi. *Biology letters*, 7 (4): 574-577. doi: 10.1098/rsbl.2010.1203.
- Blakeslee, A. F. (1904). Sexual Reproduction in the Mucorineae. *Proceedings of the American Academy of Arts and Sciences*, 40 (4): 205-319. doi: 10.2307/20021962.
- Blomqvist, J., Langseter, A. M., Markina, D., Zimmermann, B., Kohler, A. & Shapaval, V. (2019). Production of polyunsaturated fatty acids from animal fat emulsion by oleaginous filamentous fungi. *Journal of Biotechnology*, 305: S15. doi: <https://doi.org/10.1016/j.jbiotec.2019.05.065>.
- Bonfante, P. & Desiro, A. (2017). Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in Mucoromycota. *ISME J*, 11 (8): 1727-1735. doi: 10.1038/ismej.2017.21.
- Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4: e1660-e1660. doi: 10.7717/peerj.1660.
- Brundrett, M. (2008). *MYCORRHIZAL ASSOCIATIONS: The Web Resource*. Available at: <https://mycorrhizas.info/vam.html>.
- Brundrett, M. C. (2002). Coevolution of roots and mycorrhizas of land plants. *New Phytologist*, 154 (2): 275-304. doi: 10.1046/j.1469-8137.2002.00397.x.
- Bushnell, B. *BBMap short read aligner, and other bioinformatic tools*. Available at: sourceforge.net/projects/bbmap/.
- Campbell, M. A., Ganley, A. R. D., Gabaldón, T. & Cox, M. P. (2016). The Case of the Missing Ancient Fungal Polyploids. *The American Naturalist*, 188 (6): 602-614. doi: 10.1086/688763.
- Chang, Y., Desirò, A., Na, H., Sandor, L., Lipzen, A., Clum, A., Barry, K., Grigoriev, I. V., Martin, F. M., Stajich, J. E., et al. (2019). Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytologist*, 222 (1): 511-525. doi: 10.1111/nph.15613.
- Chernomor, O., von Haeseler, A. & Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, 65 (6): 997-1008. doi: 10.1093/sysbio/syw037.
- Chibucos, M. C., Soliman, S., Gebremariam, T., Lee, H., Daugherty, S., Orvis, J., Shetty, A. C., Crabtree, J., Hazen, T. H., Etienne, K. A., et al. (2016). An integrated genomic and transcriptomic survey of mucormycosis-causing fungi. *Nature communications*, 7: 12218-12218. doi: 10.1038/ncomms12218.
- Corrochano, L. M., Kuo, A., Marcet-Houben, M., Polaino, S., Salamov, A., Villalobos-Escobedo, J. M., Grimwood, J., Alvarez, M. I., Avalos, J., Bauer, D., et al. (2016). Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Curr Biol*, 26 (12): 1577-1584. doi: 10.1016/j.cub.2016.04.038.
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34 (15): 2666-2669. doi: 10.1093/bioinformatics/bty149.

- Dutreux, F., Da Silva, C., d'Agata, L., Couloux, A., Gay, E. J., Istace, B., Lapalu, N., Lemainque, A., Linglin, J., Noel, B., et al. (2018). De novo assembly and annotation of three *Leptosphaeria* genomes using Oxford Nanopore MinION sequencing. *Sci Data*, 5: 180235. doi: 10.1038/sdata.2018.235.
- Etienne, K. A., Chibucos, M. C., Su, Q., Orvis, J., Daugherty, S., Ott, S., Sengamalay, N. A., Fraser, C. M., Lockhart, S. R. & Bruno, V. M. (2014). Draft Genome Sequence of *Mortierella alpina* Isolate CDC-B6842. *Genome announcements*, 2 (1): e01180-13. doi: 10.1128/genomeA.01180-13.
- Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic acids research*, 40 (Database issue): D136-D143. doi: 10.1093/nar/gkr1178.
- Feliziani, E. & Romanazzi, G. (2016). Postharvest decay of strawberry fruit: Etiology, epidemiology, and disease management. *Journal of Berry Research*, 6: 47-63.
- Fredricks, D. N., Smith, C. & Meier, A. (2005). Comparison of six DNA extraction methods for recovery of fungal DNA as assessed by quantitative PCR. *J Clin Microbiol*, 43 (10): 5122-8. doi: 10.1128/JCM.43.10.5122-5128.2005.
- Gadagkar, S. R., Rosenberg, M. S. & Kumar, S. (2005). Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 304B (1): 64-74. doi: 10.1002/jez.b.21026.
- Gauger, W. L. (1961). THE GERMINATION OF ZYGOSPORES OF RHIZOPUS STOLONIFER. *American Journal of Botany*, 48 (5): 427-429. doi: 10.1002/j.1537-2197.1961.tb11662.x.
- Giordano, F., Aigrain, L., Quail, M. A., Coupland, P., Bonfield, J. K., Davies, R. M., Tischler, G., Jackson, D. K., Keane, T. M., Li, J., et al. (2017). De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep*, 7 (1): 3935. doi: 10.1038/s41598-017-03996-z.
- Glenn, T. (2011). Field guide to next-generation DNA sequencers. *Molecular ecology resources*, 11: 759-69. doi: 10.1111/j.1755-0998.2011.03024.x.
- Grigoriev, I. V., Cullen, D., Goodwin, S. B., Hibbett, D., Jeffries, T. W., Kubicek, C. P., Kuske, C., Magnuson, J. K., Martin, F., Spatafora, J. W., et al. (2011). Fueling the future with fungal genomics. *Mycology*, 2 (3): 192-209. doi: 10.1080/21501203.2011.584577.
- Gryganskyi, A. & Muszewska, A. (2014). Whole Genome Sequencing and the Zygomycota. *Fungal Genomics & Biology*, 04. doi: 10.4172/2165-8056.1000e116.
- Gryganskyi, A. P., Golan, J., Dolatabadi, S., Mondo, S., Robb, S., Idnurm, A., Muszewska, A., Steczkiewicz, K., Masonjones, S., Liao, H.-L., et al. (2018). Phylogenetic and Phylogenomic Definition of *Rhizopus* Species. *G3 (Bethesda, Md.)*, 8 (6): 2007-2018. doi: 10.1534/g3.118.200235.
- Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29 (8): 1072-1075. doi: 10.1093/bioinformatics/btt086.
- Heikrujam, J., Kishor, R. & Behari Mazumder, P. (2020). The Chemistry Behind Plant DNA Isolation Protocols. In Oana-Maria Boldura, C. B. a. N. S. A. (ed.) *Biochemical Analysis Tools - Methods for Bio-Molecules Studies*: IntechOpen.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. (2017). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35 (2): 518-522. doi: 10.1093/molbev/msx281.

- Hocking, D. (1967). Zygosporangium initiation, development and germination in *Phycomyces blakesleeana*. *Transactions of the British Mycological Society*, 50 (2): 207-213. doi: [https://doi.org/10.1016/S0007-1536\(67\)80031-7](https://doi.org/10.1016/S0007-1536(67)80031-7).
- Hoffmann, K., Pawłowska, J., Walther, G., Wrzosek, M., de Hoog, G. S., Benny, G. L., Kirk, P. M. & Voigt, K. (2013). The family structure of the Mucorales: a synoptic revision based on comprehensive multigene-genealogies. *Persoonia*, 30: 57-76. doi: 10.3767/003158513X666259.
- Inglis, P. W., Pappas, M. C. R., Resende, L. V. & Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*, 13 (10): e0206085. doi: 10.1371/journal.pone.0206085.
- Inkscape Project. (2020). *Inkscape*. Available at: <https://inkscape.org>.
- Jain, C., Dilthey, A., Koren, S., Aluru, S. & Phillippy, A. M. (2017, 2017//). *A Fast Approximate Algorithm for Mapping Long Reads to Large Reference Databases*. Research in Computational Molecular Biology, Cham: Springer International Publishing.
- Jain, C., Koren, S., Dilthey, A., Phillippy, A. M. & Aluru, S. (2018). A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics*, 34 (17): i748-i756. doi: 10.1093/bioinformatics/bty597.
- Jain, M., Olsen, H. E., Paten, B. & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, 17 (1): 239. doi: 10.1186/s13059-016-1103-0.
- James, T. Y. (2018). *Amylomyces rouxii* NRRL 5866 Draft Metagenome-Assembled Genome DOE Joint Genome Institute Mycocosm. Available at: <https://mycocosm.jgi.doe.gov/Amyrou1/Amyrou1.home.html>.
- Kikukawa, H., Sakuradani, E., Ando, A., Okuda, T., Shimizu, S. & Ogawa, J. (2016). Microbial production of dihomogamma-linolenic acid by $\Delta 5$ -desaturase gene-disruptants of *Mortierella alpina* 1S-4. *Journal of Bioscience and Bioengineering*, 122 (1): 22-26. doi: <https://doi.org/10.1016/j.jbiosc.2015.12.007>.
- Kikukawa, H., Sakuradani, E., Ando, A., Shimizu, S. & Ogawa, J. (2018). Arachidonic acid production by the oleaginous fungus *Mortierella alpina* 1S-4: A review. *Journal of Advanced Research*, 11: 15-22. doi: <https://doi.org/10.1016/j.jare.2018.02.003>.
- Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37 (5): 540-546. doi: 10.1038/s41587-019-0072-8.
- Kosa, G., Zimmermann, B., Kohler, A., Ekeberg, D., Afseth, N. K., Mounier, J. & Shapaval, V. (2018). High-throughput screening of Mucoromycota fungi for production of low- and high-value lipids. *Biotechnol Biofuels*, 11: 66. doi: 10.1186/s13068-018-1070-7.
- Krings, M., Taylor, T. N. & Dotzler, N. (2013). Fossil evidence of the zygomycetous fungi. *Persoonia*, 30: 1-10. doi: 10.3767/003158513X664819.
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A. & Zdobnov, E. M. (2018). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47 (D1): D807-D811. doi: 10.1093/nar/gky1053.
- Kües, U. (ed.) (2007). *Wood production, wood technology, and biotechnological impacts*. Göttingen: Universitätsverlag Göttingen.

- Lebreton, A., Corre, E., Jany, J. L., Brillet-Gueguen, L., Perez-Arques, C., Garre, V., Monsoor, M., Debuchy, R., Le Meur, C., Coton, E., et al. (2020). Comparative genomics applied to Mucor species with different lifestyles. *BMC Genomics*, 21 (1): 135. doi: 10.1186/s12864-019-6256-2.
- Lee, S. C. & Heitman, J. (2014). Sex in the Mucoralean fungi. *Mycoses*, 57 Suppl 3 (0 3): 18-24. doi: 10.1111/myc.12244.
- Leger, A. & Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software*, 4: 1236. doi: 10.21105/joss.01236.
- Lennartsson, P. R., Taherzadeh, M. J. & Edebo, L. (2014). Rhizopus. In Batt, C. A. & Tortorello, M. L. (eds) *Encyclopedia of Food Microbiology (Second Edition)*, pp. 284-290. Oxford: Academic Press.
- 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 research*, 44 (W1): W242-W245. doi: 10.1093/nar/gkw290.
- Ma, L. J., Ibrahim, A. S., Skory, C., Grabherr, M. G., Burger, G., Butler, M., Elias, M., Idnurm, A., Lang, B. F., Sone, T., et al. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet*, 5 (7): e1000549. doi: 10.1371/journal.pgen.1000549.
- Magdouli, S., Yan, S., D. Tyagi, R. & Y. Surampalli, R. (2014). *Heterotrophic Microorganisms: A Promising Source for Biodiesel Production*, vol. 44.
- Mamani, L. D. G., Magalhães, A. I., Ruan, Z., Carvalho, J. C. d. & Soccol, C. R. (2019). Industrial production, patent landscape, and market trends of arachidonic acid-rich oil of *Mortierella alpina*. *Biotechnology Research and Innovation*, 3 (1): 103-119. doi: <https://doi.org/10.1016/j.biori.2019.02.002>.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37 (5): 1530-1534. doi: 10.1093/molbev/msaa015.
- Mondo, S. J., Dannebaum, R. O., Kuo, R. C., Louie, K. B., Bewick, A. J., LaButti, K., Haridas, S., Kuo, A., Salamov, A., Ahrendt, S. R., et al. (2017a). Widespread adenine N6-methylation of active genes in fungi. *Nat Genet*, 49 (6): 964-968. doi: 10.1038/ng.3859.
- Mondo, S. J., Lastovetsky, O. A., Gaspar, M. L., Schwardt, N. H., Barber, C. C., Riley, R., Sun, H., Grigoriev, I. V. & Pawlowska, T. E. (2017b). Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nat Commun*, 8 (1): 1843. doi: 10.1038/s41467-017-02052-8.
- Money, N. P. (2016a). Chapter 1 - Fungal Diversity. In Watkinson, S. C., Boddy, L. & Money, N. P. (eds) *The Fungi (Third Edition)*, pp. 1-36. Boston: Academic Press.
- Money, N. P. (2016b). Chapter 2 - Fungal Cell Biology and Development. In Watkinson, S. C., Boddy, L. & Money, N. P. (eds) *The Fungi (Third Edition)*, pp. 37-66. Boston: Academic Press.
- Moore, D. & Meškauskas, A. (2017). *Neighbour-Sensing mathematical model of hyphal growth*. Available at: <http://www.davidmoore.org.uk/CyberWEB/Downloads-page.htm>.
- Morin, E., Miyauchi, S., San Clemente, H., Chen, E. C. H., Pelin, A., de la Providencia, I., Ndikumana, S., Beaudet, D., Hainaut, M., Drula, E., et al. (2019). Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora*

- rosea highlights specific genetic features in Glomeromycotina. *New Phytologist*, 222 (3): 1584-1598. doi: 10.1111/nph.15687.
- Muller, F. M., Werner, K. E., Kasai, M., Francesconi, A., Chanock, S. J. & Walsh, T. J. (1998). Rapid extraction of genomic DNA from medically important yeasts and filamentous fungi by high-speed cell disruption. *J Clin Microbiol*, 36 (6): 1625-9.
- Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, 34 (14): 2490-2492. doi: 10.1093/bioinformatics/bty121.
- Naranjo-Ortiz, M. A. & Gabaldon, T. (2020). Fungal evolution: cellular, genomic and metabolic complexity. *Biol Rev Camb Philos Soc*. doi: 10.1111/brv.12605.
- Nicholls, S. M. (2019). *contiguity.R*. Available at: <https://gist.github.com/SamStudio8/6b16674bae8afec706bed80e6ac2da24>.
- Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. (2019). Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience*, 8 (5). doi: 10.1093/gigascience/giz043.
- Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I. V. & Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic acids research*, 42 (Database issue): D26-D31. doi: 10.1093/nar/gkt1069.
- O'donnell, K. L., Hooper, G. & Fields, W. G. (1976). *Zygosporogenesis in Phycomyces blakesleeanus*.
- Okuda, T., Ando, A., Negoro, H., Kikukawa, H., Sakamoto, T., Sakuradani, E., Shimizu, S. & Ogawa, J. (2015). Omega-3 eicosatetraenoic acid production by molecular breeding of the mutant strain S14 derived from *Mortierella alpina* 1S-4. *Journal of Bioscience and Bioengineering*, 120 (3): 299-304. doi: <https://doi.org/10.1016/j.jbiosc.2015.01.014>.
- Oxford Nanopore Technologies. (2019). *Native barcoding genomic DNA (with EXP-NBD104, with EXP-NBD114, and SQK-LSK109, version NBE_9065_v109_revU_14Aug2019*. Oxford, UK
- Oxford Nanopore Technologies.
- Pacific Biosciences. (2020). *Technical Note DNA Prep*: Pacific Biosciences of California, Inc. Available at: <https://www.pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf>.
- Panthee, S., Hamamoto, H., Ishijima, S. A., Paudel, A. & Sekimizu, K. (2018). Utilization of Hybrid Assembly Approach to Determine the Genome of an Opportunistic Pathogenic Fungus, *Candida albicans* TIMM 1768. *Genome Biol Evol*, 10 (8): 2017-2022. doi: 10.1093/gbe/evy166.
- Papanikolaou, S., Galiotou-Panayotou, M., Fakas, S., Komaitis, M. & Aggelis, G. (2007). Lipid production by oleaginous Mucorales cultivated on renewable carbon sources. *European Journal of Lipid Science and Technology*, 109 (11): 1060-1070. doi: 10.1002/ejlt.200700169.
- Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T. & Sandhu, M. S. (2018). Long reads: their purpose and place. *Hum Mol Genet*, 27 (R2): R234-R241. doi: 10.1093/hmg/ddy177.
- Quick, J. & Loman, N. J. (2018). DNA Extraction Strategies for Nanopore Sequencing. In *Nanopore Sequencing*, pp. 91-105: WORLD SCIENTIFIC.

- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Available at: <http://www.R-project.org/>.
- Rabiee, M., Sayyari, E. & Mirarab, S. (2019). Multi-allele species reconstruction using ASTRAL. *Molecular Phylogenetics and Evolution*, 130: 286-296. doi: <https://doi.org/10.1016/j.ympev.2018.10.033>.
- Remy, W., Taylor, T. N., Hass, H. & Kerp, H. (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences*, 91 (25): 11841. doi: 10.1073/pnas.91.25.11841.
- Ren, R., Sun, Y., Zhao, Y., Geiser, D., Ma, H. & Zhou, X. (2016). Phylogenetic Resolution of Deep Eukaryotic and Fungal Relationships Using Highly Conserved Low-Copy Nuclear Genes. *Genome biology and evolution*, 8 (9): 2683-2701. doi: 10.1093/gbe/evw196.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29 (1): 24-26. doi: 10.1038/nbt.1754.
- Sakamoto, T., Sakuradani, E., Okuda, T., Kikukawa, H., Ando, A., Kishino, S., Izumi, Y., Bamba, T., Shima, J. & Ogawa, J. (2017). Metabolic engineering of oleaginous fungus *Mortierella alpina* for high production of oleic and linoleic acids. *Bioresource Technology*, 245: 1610-1615. doi: <https://doi.org/10.1016/j.biortech.2017.06.089>.
- Samarakoon, M. C., Kd, H. & Ariyawansa, H. (2017). Divergence and ranking of taxa across the kingdoms Animalia, Fungi and Plantae. *Mycosphere*, 7: 1678-1689. doi: 10.5943/mycosphere/7/11/5.
- Sayyari, E. & Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution*, 33 (7): 1654-1668. doi: 10.1093/molbev/msw079.
- Seppy, M., Manni, M. & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In Kollmar, M. (ed.) *Gene Prediction: Methods and Protocols*, pp. 227-245. New York, NY: Springer New York.
- Sharma, K. (2015). Fungal genome sequencing: basic biology to biotechnology. *Critical Reviews in Biotechnology*, 36. doi: 10.3109/07388551.2015.1015959.
- Shen, W., Le, S., Li, Y. & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, 11 (10): e0163962. doi: 10.1371/journal.pone.0163962.
- Soare, A. Y., Watkins, T. N. & Bruno, V. M. (2020). Understanding Mucormycoses in the Age of “omics”. *Frontiers in Genetics*, 11 (699). doi: 10.3389/fgene.2020.00699.
- Spatafora, J., Aime, M., Grigoriev, I., Martin, F., Stajich, J. & Blackwell, M. (2017). The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies. *Microbiology Spectrum*, 5. doi: 10.1128/microbiolspec.FUNK-0053-2016.
- Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*, 108 (5): 1028-1046. doi: 10.3852/16-042.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30 (9): 1312-1313. doi: 10.1093/bioinformatics/btu033.
- Takeda, I., Tamano, K., Yamane, N., Ishii, T., Miura, A., Umemura, M., Terai, G., Baker, S. E., Koike, H. & Machida, M. (2014). Genome Sequence of the Mucoromycotina

- Fungus *Umbelopsis isabellina*, an Effective Producer of Lipids. *Genome announcements*, 2 (1): e00071-14. doi: 10.1128/genomeA.00071-14.
- Tang, X., Zhao, L., Chen, H., Chen, Y. Q., Chen, W., Song, Y. & Ratledge, C. (2015). Complete Genome Sequence of a High Lipid-Producing Strain of *Mucor circinelloides* WJ11 and Comparative Genome Analysis with a Low Lipid-Producing Strain CBS 277.49. *PLOS ONE*, 10 (9): e0137543. doi: 10.1371/journal.pone.0137543.
- Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D., May, T., Ryberg, M. & Abarenkov, K. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity*, 90 (1): 135-159. doi: 10.1007/s13225-018-0401-0.
- Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14 (2): 178-192. doi: 10.1093/bib/bbs017.
- Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., et al. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (50): 20117-20122. doi: 10.1073/pnas.1313452110.
- Tzimirotas, D., Afseth, N. K., Lindberg, D., Kjørlaug, O., Axelsson, L. & Shapaval, V. (2018). Pretreatment of different food rest materials for bioconversion into fungal lipid-rich biomass. *Bioprocess and Biosystems Engineering*, 41 (7): 1039-1049. doi: 10.1007/s00449-018-1933-0.
- Uehling, J., Gryganskyi, A., Hameed, K., Tschaplinski, T., Misztal, P. K., Wu, S., Desiro, A., Vande Pol, N., Du, Z., Zienkiewicz, A., et al. (2017). Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycoavidus cysteinexigens*. *Environ Microbiol*, 19 (8): 2964-2983. doi: 10.1111/1462-2920.13669.
- Umesha, S., Manukumar, H. M. & Raghava, S. (2016). A rapid method for isolation of genomic DNA from food-borne fungal pathogens. *3 Biotech*, 6 (2): 123. doi: 10.1007/s13205-016-0436-4.
- van Burik, J. A., Schreckhise, R. W., White, T. C., Bowden, R. A. & Myerson, D. (1998). Comparison of six extraction techniques for isolation of DNA from filamentous fungi. *Med Mycol*, 36 (5): 299-303.
- Vega, K. & Kalkum, M. (2012). Chitin, Chitinase Responses, and Invasive Fungal Infections. *International journal of microbiology*, 2012: 920459. doi: 10.1155/2012/920459.
- Venice, F., Ghignone, S., Salvioli di Fossalunga, A., Amselem, J., Novero, M., Xianan, X., Sędziewska Toro, K., Morin, E., Lipzen, A., Grigoriev, I. V., et al. (2020). At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita* provides insights into plant, endobacterial and fungal interactions. *Environmental Microbiology*, 22 (1): 122-141. doi: 10.1111/1462-2920.14827.
- Voigt, K. & Wostemeyer, J. (2001). Phylogeny and origin of 82 zygomycetes from 54 genera of the Mucorales and Mortierellales based on combined analysis of actin and translation elongation factor EF-1 alpha genes. *Gene*, 270: 113-120. doi: 10.1016/S0378-1119(01)00464-4.
- Walther, G., Wagner, L. & Kurzai, O. (2019). Updates on the Taxonomy of Mucorales with an Emphasis on Clinically Important Taxa. *J Fungi (Basel)*, 5 (4). doi: 10.3390/jof5040106.

- Wang, L., Chen, W., Feng, Y., Ren, Y., Gu, Z., Chen, H., Wang, H., Thomas, M. J., Zhang, B., Berquin, I. M., et al. (2011). Genome characterization of the oleaginous fungus *Mortierella alpina*. *PLoS One*, 6 (12): e28319. doi: 10.1371/journal.pone.0028319.
- Wei, H., Wang, W., Yarbrough, J. M., Baker, J. O., Laurens, L., Van Wychen, S., Chen, X., Taylor, L. E., II, Xu, Q., Himmel, M. E., et al. (2013). Genomic, Proteomic, and Biochemical Analyses of Oleaginous *Mucor circinelloides*: Evaluating Its Capability in Utilizing Cellulolytic Substrates for Lipid Production. *PLOS ONE*, 8 (9): e71068. doi: 10.1371/journal.pone.0071068.
- Weisenberger, A., Bonito, G., Lee, S. J., McKisson, J., Gryganskyi, A., Reid, C. D., Smith, M. F., Vaidyanathan, G. & Welch, B. (2013). A radioisotope based methodology for plant-fungal interactions in the rhizosphere. *IEEE Nuclear Science Symposium Conference Record*. doi: 10.1109/NSSMIC.2013.6829457.
- Whelan, S. & Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, 18 (5): 691-699. doi: 10.1093/oxfordjournals.molbev.a003851.
- White, M. M., James, T. Y., O'Donnell, K., Cafaro, M. J., Tanabe, Y. & Sugiyama, J. (2006). Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. *Mycologia*, 98 (6): 872-84.
- Wick, R. R., Judd, L. M. & Holt, K. E. (2018). Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol*, 14 (11): e1006583. doi: 10.1371/journal.pcbi.1006583.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York. Available at: <https://ggplot2.tidyverse.org>.
- Ye, C., Xu, N., Chen, H., Chen, Y. Q., Chen, W. & Liu, L. (2015). Reconstruction and analysis of a genome-scale metabolic model of the oleaginous fungus *Mortierella alpina*. *BMC Syst Biol*, 9: 1. doi: 10.1186/s12918-014-0137-8.
- Young, A. D. & Gillung, J. P. (2020). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45 (2): 225-247. doi: 10.1111/syen.12406.
- Zdobnov, E. (2020). *generate_plot.py*. Available at: https://gitlab.com/ezlab/busco/-/blob/master/scripts/generate_plot.py.
- Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19 (6): 153. doi: 10.1186/s12859-018-2129-y.

6 Appendix

6.1 Cultivation of fungal strains

Table 6.1. Strains cultured without antibiotic and where biomass was processed with the room temperature biomass handling protocol.

DNA extraction sample number	Strain	Agar medium	Agar plate incubation time (days)	Liquid culture incubation time (days)	Sequencing sample number
1	<i>Mucor racemosus</i> UBOCC-A-102007*	MEA	5	2	na
2 and 37	<i>Mucor circinelloides</i> UBOCC-A-102010*	MEA	5	2	na
3	<i>Mucor racemosus</i> UBOCC-A-111127*	MEA	5	2	na
4	<i>Mucor lanceolatus</i> UBOCC-A-109193	MEA	5	2	1
5	<i>Mucor fragilis</i> UBOCC-A-109196*	MEA	5	2	na
6	<i>Mucor plumbeus</i> UBOCC-A-109204*	MEA	5	2	na
7	<i>Umbelopsis vinacea</i> UBOCC-A-101347	PDA	5	2	na
8	<i>Mortierella zonata</i> UBOCC-A-101348	PDA	5	2	na
9	<i>Umbelopsis isabellina</i> UBOCC-A-101350*	PDA	5	2	na
10	<i>Umbelopsis isabellina</i> UBOCC-A-101351*	PDA	5	2	na
11	<i>Mortierella alpina</i> UBOCC-A-112046	PDA	5	2	na
12	<i>Mucor circinelloides</i> CCM 8328*	MEA	5	2	na
13	<i>Absidia cylindrospora</i> CCM F-52T*	MEA	5	2	na
14	<i>Mucor plumbeus</i> UBOCC-A-111132*	MEA	5	2	na
15	<i>Absidia glauca</i> CCM 450*	MEA	5	2	na
16	<i>Absidia glauca</i> CCM 451*	MEA	5	2	na
17	<i>Cunninghamella echinulata</i> VKM F-470*	PDA	7	5	na
18	<i>Umbelopsis ramanniana</i> VKM F-502*	PDA	7	5	na
19	<i>Umbelopsis vinacea</i> CCM F-513	MEA	7	5	na
20	<i>Cunninghamella echinulata</i> VKM F-531*	PDA	7	5	na
21	<i>Umbelopsis vinacea</i> CCM F-539*	PDA	8	5	na
22	<i>Cunninghamella blakesleeana</i> CCM F-705*	PDA	8	5	na
23	<i>Rhizopus oryzae</i> CCM 8076*	MEA	5	2	na
24	<i>Rhizopus oryzae</i> CCM 8116*	MEA	5	2	na
25	<i>Mucor mucedo</i> UBOCC-A-101361*	MEA	5	2	na
26	<i>Absidia glauca</i> CCM F-444	MEA	5	2	na
27	<i>Absidia coerulea</i> VKM F-627*	MEA	5	2	na
28	<i>Rhizopus microsporus</i> CCM F-792*	MEA	5	2	na

29	<i>Mucor circinelloides</i> FRR 5020*	MEA	5	2	na
30	<i>Rhizopus oryzae</i> CCM 8075*	MEA	5	2	na
31	<i>Mortierella hyalina</i> UBOCC-A-101349	PDA	5	2	2
32	<i>Cunninghamella echinulata</i> VKM F-439*	PDA	5	2	na
33	<i>Umbelopsis isabellina</i> VKM F-525*	PDA	5	2	na
34	<i>Mortierella alpina</i> ATCC® 32222_TT™	PDA	5	2	na
35	<i>Mucor plumbeus</i> UBOCC-A-111125	MEA	5	2	na
36	<i>Mucor plumbeus</i> UBOCC-A-111128	MEA	5	2	na
38	<i>Mucor circinelloides</i> UBOCC-A-105017*	MEA	5	2	na
39	<i>Mucor hiemalis</i> UBOCC-A-111119*	MEA	5	2	na
40	<i>Mucor racemosus</i> UBOCC-A-111130	MEA	5	2	na
41	<i>Mucor lanceolatus</i> UBOCC-A-110148*	MEA	5	2	na
42	<i>Mucor hiemalis</i> UBOCC-A-112185*	MEA	5	2	na
43	<i>Mucor hiemalis</i> UBOCC-A-109197	MEA	5	2	3
44	<i>Mucor plumbeus</i> UBOCC-A-109208	MEA	4	2	na
45	<i>Amylomyces rouxii</i> CCM F-220*	MEA	4	2	na
46	<i>Mucor fragilis</i> CCM F-236*	MEA	4	2	na
47	<i>Lichtheimia corymbifera</i> VKM F-513*	MEA	7	8	na
48	<i>Mucor flavus</i> VKM F-1003*	MEA	7	8	na
49	<i>Mucor lanceolatus</i> UBOCC-A-101355*	MEA	5	2	na
50	<i>Mucor plumbeus</i> CCM F-443	MEA	5	2	na
51	<i>Rhizopus stolonifer</i> CCM F-445*	MEA	4	2	na
52	<i>Umbelopsis ramanniana</i> CCM F-622	PDA	5	2	4
53	<i>Rhizopus microsporus</i> CCM F-718	MEA	5	2	na
54	<i>Lichtheimia corymbifera</i> CCM 8077	MEA	4	2	5
55	<i>Mucor racemosus</i> CCM 8190*	MEA	4	2	na
56	<i>Absidia coerulea</i> CCM 8230*	MEA	4	2	na

* Strains that also were cultured later with 40 mg/L chloramphenicol and where biomass was processed with the dry ice biomass handling protocol.

MEA = malt extract agar.

PDA = potato dextrose agar.

Table 6.2. Strains cultured with 40 mg/L chloramphenicol added to the medium and where biomass was processed with dry ice biomass handling protocol.

DNA extraction sample number	Strain	Agar medium	Agar plate incubation time (days)	Liquid culture incubation time (days)	Sequencing sample number
1	<i>Mucor racemosus</i> FRR 3336	MEA	4	3	na
2	<i>Mucor plumbeus</i> UBOCC-A-109204*	MEA	4	3	6

3	<i>Mucor racemosus</i> UBOCC-A-111130	MEA	4	3	na
4	<i>Mucor circinelloides</i> UBOCC-A-102010*	MEA	4	3	na
5	<i>Absidia glauca</i> CCM 450*	MEA	4	3	7
6	<i>Mucor circinelloides</i> CCM 8328*	MEA	4	3	na
7	<i>Mucor racemosus</i> UBOCC-A-102007*	MEA	4	3	na
8	<i>Mucor racemosus</i> UBOCC-A-111127*	MEA	4	3	8
9	<i>Absidia coerulea</i> CCM 8230*	MEA	9	3	na
10	<i>Mucor hiemalis</i> UBOCC-A-112185*	MEA	8	3	na
11	<i>Cunninghamella blakesleeana</i> CCM F-705*	PDA	8	3	na
12	<i>Umbelopsis isabellina</i> UBOCC-A-101350*	PDA	8	3	na
13	<i>Cunninghamella echinulata</i> VKM F-470*	PDA	7	3	na
14	<i>Cunninghamella echinulata</i> VKM F-531*	PDA	8	3	na
15	<i>Absidia glauca</i> CCM 451*	MEA	4	3	na
16	<i>Absidia cylindrospora</i> CCM F-52T*	MEA	4	3	na
17	<i>Mucor plumbeus</i> UBOCC-A-111132*	MEA	5	2	9
18	<i>Mucor racemosus</i> CCM 8190*	MEA	6	2	na
19	<i>Rhizopus stolonifer</i> CCM F-445*	MEA	6	2	10
20	<i>Mucor fragilis</i> CCM F-236*	MEA	5	2	na
21	<i>Rhizopus oryzae</i> CCM 8076*	MEA	5	2	na
22	<i>Mucor lanceolatus</i> UBOCC-A-110148*	MEA	5	2	na
23	<i>Rhizopus oryzae</i> CCM 8116*	MEA	5	2	na
24	<i>Mucor lanceolatus</i> UBOCC-A-101355*	MEA	6	2	na
25	<i>Mucor circinelloides</i> UBOCC-A-105017*	MEA	6	2	na
26	<i>Mucor circinelloides</i> FRR 5020*	MEA	6	2	na
27	<i>Rhizopus microsporus</i> CCM F-792*	MEA	6	2	na
28	<i>Absidia coerulea</i> VKM F-627*	MEA	6	2	na
29	<i>Rhizopus oryzae</i> CCM 8075*	MEA	6	2	na
30	<i>Amylomyces rouxii</i> CCM F-220*	MEA	5	2	11
31	<i>Umbelopsis ramanniana</i> VKM F-502*	PDA	11	2	na
32	<i>Umbelopsis vinacea</i> CCM F-539*	PDA	11	2	na
33	<i>Umbelopsis isabellina</i> VKM F-525*	PDA	10	2	na
34	<i>Umbelopsis isabellina</i> UBOCC-A-101351*	PDA	11	2	na
35	<i>Mucor mucedo</i> UBOCC-A-101361*	MEA	6	2	na
36	<i>Lichtheimia corymbifera</i> VKM F-513*	MEA	7	2	12
37	<i>Cunninghamella echinulata</i> VKM F-439*	PDA	6	2	na
38	<i>Mucor flavus</i> VKM F-1097	MEA	7	2	na
39	<i>Mucor flavus</i> VKM F-1003*	MEA	7	2	na
40	<i>Mucor fragilis</i> UBOCC-A-109196*	MEA	6	2	na

41	<i>Mucor hiemalis</i> UBOCC-A-111119*	MEA	6	2	na
----	---------------------------------------	-----	---	---	----

* Strains that also were cultured previously without antibiotic in the medium and where biomass was processed with the room temperature biomass handling protocol.

MEA = malt extract agar.

PDA = potato dextrose agar.

Table 6.3. List of Mucoromycota strains used in the project, including strains where DNA was not extracted.

Agar medium	Strain	Agar medium	Strain
MEA	<i>Absidia coerulea</i> CCM 8230	MEA	<i>Mucor flavus</i> VKM F-1110
MEA	<i>Absidia coerulea</i> VKM F-627	MEA	<i>Mucor fragilis</i> CCM F-236
MEA	<i>Absidia coerulea</i> VKM F-833	MEA	<i>Mucor fragilis</i> UBOCC-A-109196
MEA	<i>Absidia cylindrospora</i> CCM F-52T	MEA	<i>Mucor fragilis</i> UBOCC-A-113030
MEA	<i>Absidia cylindrospora</i> VKM F-1632	MEA	<i>Mucor hiemalis</i> FRR 5101
MEA	<i>Absidia cylindrospora</i> VKM F-2428	MEA	<i>Mucor hiemalis</i> UBOCC-A-101359
MEA	<i>Absidia glauca</i> CCM 450	MEA	<i>Mucor hiemalis</i> UBOCC-A-101360
MEA	<i>Absidia glauca</i> CCM 451	MEA	<i>Mucor hiemalis</i> UBOCC-A-109197
MEA	<i>Absidia glauca</i> CCM F-444	MEA	<i>Mucor hiemalis</i> UBOCC-A-111119
MEA	<i>Absidia glauca</i> UBOCC-A-101330	MEA	<i>Mucor hiemalis</i> UBOCC-A-112185
MEA	<i>Amylomyces rouxii</i> CCM F-220	MEA	<i>Mucor lanceolatus</i> UBOCC-A-101355
PDA	<i>Cunninghamella blakesleeana</i> CCM F-705	MEA	<i>Mucor lanceolatus</i> UBOCC-A-109193
PDA	<i>Cunninghamella blakesleeana</i> VKM F-993	MEA	<i>Mucor lanceolatus</i> UBOCC-A-110148
PDA	<i>Cunninghamella echinulata</i> VKM F-439	MEA	<i>Mucor mucedo</i> UBOCC-A-101353
PDA	<i>Cunninghamella echinulata</i> VKM F-470	MEA	<i>Mucor mucedo</i> UBOCC-A-101361
PDA	<i>Cunninghamella echinulata</i> VKM F-531	MEA	<i>Mucor mucedo</i> UBOCC-A-101362
MEA	<i>Lichtheimia corymbifera</i> CCM 8077	MEA	<i>Mucor plumbeus</i> CCM F-443
MEA	<i>Lichtheimia corymbifera</i> VKM F-507	MEA	<i>Mucor plumbeus</i> FRR 2412
MEA	<i>Lichtheimia corymbifera</i> VKM F-513	MEA	<i>Mucor plumbeus</i> FRR 4804
PDA	<i>Mortierella alpina</i> ATCC® 32222_TT™	MEA	<i>Mucor plumbeus</i> UBOCC-A-109204
PDA	<i>Mortierella alpina</i> UBOCC-A-112046	MEA	<i>Mucor plumbeus</i> UBOCC-A-109208

PDA	Mortierella alpina UBOCC-A-112047	MEA	Mucor plumbeus UBOCC-A-109210
PDA	Mortierella elongata VKM F-1614	MEA	Mucor plumbeus UBOCC-A-111125
PDA	Mortierella elongata VKM F-524	MEA	Mucor plumbeus UBOCC-A-111128
PDA	Mortierella gamsii VKM F-1402	MEA	Mucor plumbeus UBOCC-A-111132
PDA	Mortierella gamsii VKM F-1529	MEA	Mucor racemosus CCM 8190
PDA	Mortierella gamsii VKM F-1641	MEA	Mucor racemosus FRR 3336
PDA	Mortierella gemmifera VKM F-1252	MEA	Mucor racemosus FRR 3337
PDA	Mortierella gemmifera VKM F-1631	MEA	Mucor racemosus UBOCC-A-102007
PDA	Mortierella gemmifera VKM F-1651	MEA	Mucor racemosus UBOCC-A-109211
PDA	Mortierella globulifera VKM F-1408	MEA	Mucor racemosus UBOCC-A-111127
PDA	Mortierella globulifera VKM F-1448	MEA	Mucor racemosus UBOCC-A-111130
PDA	Mortierella globulifera VKM F-1495	MEA	Rhizopus microsporus CCM F-718
PDA	Mortierella humilis VKM F-1494	MEA	Rhizopus microsporus CCM F-792
PDA	Mortierella humilis VKM F-1528	MEA	Rhizopus microsporus VKM F-1091
PDA	Mortierella humilis VKM F-1611	MEA	Rhizopus oryzae CCM 8075
PDA	Mortierella hyalina UBOCC-A-101349	MEA	Rhizopus oryzae CCM 8076
PDA	Mortierella hyalina VKM F-1629	MEA	Rhizopus oryzae CCM 8116
PDA	Mortierella hyalina VKM F-1854	MEA	Rhizopus stolonifer CCM F-445
PDA	Mortierella zonata UBOCC-A-101348	MEA	Rhizopus stolonifer VKM F-399
PDA	Mortierella zonata VKM F-1409	MEA	Rhizopus stolonifer VKM F-400
MEA	Mucor circinelloides VI 04473	PDA	Umbelopsis isabellina UBOCC-A-101350
MEA	Mucor circinelloides CCM 8328	PDA	Umbelopsis isabellina UBOCC-A-101351
MEA	Mucor circinelloides FRR 4846	PDA	Umbelopsis isabellina VKM F-525
MEA	Mucor circinelloides FRR 5020	PDA	Umbelopsis ramanniana CCM F-622
MEA	Mucor circinelloides FRR 5021	PDA	Umbelopsis ramanniana VKM F-502
MEA	Mucor circinelloides UBOCC-A-102010	PDA	Umbelopsis vinacea CCM 8333
MEA	Mucor circinelloides UBOCC-A-105017	MEA	Umbelopsis vinacea CCM F-513
PDA	Mucor flavus CCM 8086	PDA	Umbelopsis vinacea CCM F-539
MEA	Mucor flavus VKM F-1003	PDA	Umbelopsis vinacea UBOCC-A-101347
MEA	Mucor flavus VKM F-1097		

MEA = malt extract agar.

PDA = potato dextrose agar.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway