FACULTY OF SCIENCES

FACULTY OF BIOSCIENCE ENGINEERING

FACULTY OF ENGINEERING AND ARCHITECTURE

METAGENOMIC STUDY OF LETTUCE POTTING SOIL MICROBIOME UNDER CHITIN TREATMENT AND FUNCTIONAL GENOMIC ANALYSIS OF *MORTIERELLA HYALINA*

A PLANT GROWTH-PROMOTING, CHITIN DEGRADING FUNGUS

Word count: 28125

Floris Voorthuijzen

Student number: 01503491

Promoters: Prof. Dr. Peter Dawyndt, Dr. Caroline De Tender

Supervisor: Dr. Annelies Haegeman

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science in Bioinformatics: Bioscience Engineering

Academic Year: 2019- 2020

GHENT UNIVERSITY

06/06/2020

Floris Voorthuijzen

Acknowledgements

I wish to express my deepest gratitude to all the friendly folks at the ILVO in Merelbeke for their help and hospitality. In particular, I would like to thank everyone in the monthly Microbiodem meetings for their input. Special thanks go to Caroline De Tender, whose vision for this thesis was of invaluable importance, Annelies Haegeman, who flawlessly took over the co-promotership in the second semester and guided me through the genome assembly, and Steve Baeyen, who was always eager to help.

Table of Contents

Abbreviations

16S rRNA	Ribosomal RNA subunit of size 16 Svedberg units
AA	Amino Acid
ABA	Abscisic acid
ANOVA	Analysis of Variance
ASV	Amplicon Sequence Variant
BAC	Bacterial Artificial Chromosomes
BAM	Binary Alignment Map
BLAST	Basic Local Alignment Search Tool
BLASTp	Protein-protein BLAST
BS	Bulk potting soil sampling
CH	Chitin treatment
CLI	Command Line Interface
CPM	Count Per Million
CT	Control treatment
Df	Degrees of Freedom
DNA	Deoxyribonucleic acid
EC	Enzyme Commission (number)
FWD	Forward (reads)
GC content	Guanine-Cytosine content
GFF	Generic Feature Format
GH	Glycoside Hydrolase
GlcNAc	N-acetylglucosamine
GLM	Generalized Linear Model
IAA	Indole-3-acetic acid
ILVO	Instituut voor Landbouw en Visserij Onderzoek (Institute for Agriculture and Fishery Research)
Indel	Insertion & Deletion
ITS	Internally Transcribed Spacer
Kb	Kilo bases (=10^3 bases)
KB	Kilobyte
LRT	Likelihood Ratio Test
MAMP	Microbe Associated Molecular Pattern
Mb	Mega bases (=10^6 bases)
MB	Megabyte
MDS	Multidimensional Scaling
Mean Sq	Mean square
ML	Maximum Likelihood
N-cycle	Nitrogen cycle
NGS	Next Generation Sequencing
NirA	Gene for a nitrogen assimilation transcription factor
Nit-4	Gene for a nitrogen assimilation transcription factor

NJ	Neighbour Joining
NPK	Nitrogen Phosphorus Potassium fertilizer
ORF	Open Reading Frame
OTU	Operational Taxonomic Unit
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PERMANOVA	Permutational Multivariate Analysis of Variance
PGPF	Plant Growth Promoting Fungi
PGPR	Plant Growth Promoting Rhizobacteria
PHRED	Quality standard for sequencing data
QC	Quality Control
REV	Reverse (reads)
Rhizo	Rhizosphere sampling
RNA	Ribonucleic acid
SGS	Second Generation Sequencing
SNP	Single Nucleotide Polymorphism
Sum Sq	Sum square
T4	Sampling after 4 weeks
T8	Sampling after 8 weeks
tBLASTn	Protein query BLAST in nucleotide database translated to amino acids in 6 ORF's
TGS	Third Generation Sequencing
V3-V4	Region in the bacterial 16S rRNA
WGS	Whole Genome Shotgun sequencing

Abstract

In this thesis we show that *Mortierella hyalina* is a plant growth-promoting fungus that thrives in chitin-rich soil, and we take a deep dive into its genome to find out why this is the case.

We began by analyzing the effect of chitin on the bacterial and fungal microbiome of lettuce grown in potting soil. The bacterial microbiome was analyzed with 16S V3-V4 metabarcoding, and ITS2 metabarcoding was used for the fungal community. Samples were taken from the rhizosphere and bulk potting soil of chitin-treated and control plants, 4 and 8 weeks after treatment. We found that the treatment and sampling location had a significant effect on both the fungal and bacterial microbiome, and that the bacterial community also changed significantly over time.

In the fungal community, one particular fungus stood out. The relative abundance of *Mortierella* increased to over 50% in all chitin-treated samples. Preliminary research has shown that *Mortierella spp.* can degrade chitin, and have certain plant growth-promoting properties. Therefore, we sequenced the whole genome of a *Mortierella* strain in our lab with as primary goal to confirm that this strain indeed has the ability to degrade chitin and to identify genes related in plant growth promotion. Phylogenetic analysis of this *Mortierella* strain classified it as *Mortierella hyalina*.

Since no genome assembly of *Mortierella hyalina* is publicly available, we performed *de novo* genome assembly with the state-of-the-art Shovill software. On this assembly we did a functional genomic analysis, starting off with the gene prediction tool AUGUSTUS. We first queried the genome for genes involved in the chitin degradation pathways to explain the immense success of this fungus in chitin-rich soil. Multiple genes that take part in every relevant chitin degrading pathway were identified. We then searched for genes involved in plant growth-promotion, plant resilience, and nitrogen-cycling. We found that the *Mortierella hyalina* genome is loaded with useful genes for biocontrol and nutrient cycling, and conclude that this fungus holds much potential to be a major asset for sustainable crop protection and cultivation.

Dutch Abstract

In deze thesis wordt aangetoond dat *Mortierella hyalina* een plantengroeibevorderende fungus is die gedijt in chitinerijke bodem en we nemen het genoom onder de loep om te zien waarom dat zo is.

We begonnen met het effect van chitine op het bacteriële en fungale microbioom van sla die gekweekt is in potgrond te analyseren. Het bacteriële microbioom werd bestudeerd met 16S V3-V4 metabarcoding en voor de fungale gemeenschap werd ITS2 metabarcoding gebruikt. Stalen werden genomen uit de rhizosfeer of uit de potgrond van chitinebehandelde en controle plantjes, 4 en 8 weken na de behandeling. We vonden dat de behandeling en de locatie van de staalname een significant effect hadden op zowel de bacteriële als de fungale gemeenschap en dat de bacteriële gemeenschap ook signficant veranderde doorheen de tijd.

In de fungale gemeenschap viel één bepaalde fungus op. Er was een toename van meer dan 50% in de relatieve overvloed van *Mortierella* in alle chitinebehandelde stalen. Voorafgaand onderzoek heeft aangetoond dat *Mortierella spp.* chitine kunnen afbreken en bepaalde plantengroei-bevorderende eigenschappen kunnen vertonen. Daarom hebben we het hele genoom van een *Mortierella* stam in ons labo gesequeneerd, waarmee we beogen te bevestigen dat deze stam inderdaad over de mogelijkheid beschikt om chitine af te breken. Tevens zullen we trachten in het *Mortierella* genoom een aantal genen te identificeren die een plantengroeibevorderende rol spelen. Een fylogenetische analyse classificeerde de betreffende *Mortierella* stam als *Mortierella hyalina*.

Omdat er geen publieke data over het *Mortierella hyalina* genoom ter beschikking is, hebben we een *de novo* genoomconstructie gemaakt met de moderne Shovill software. Op deze genoomconstructie hebben we een functionele genomische analyse gedaan, die aanving met het genpredictietool AUGUSTUS. Eerst doorzochten we het genoom naar genen die actief zijn in chitine afbrekende moleculaire systemen om het immense succes van deze fungus in chitinerijke bodem te kunnen verklaren. Verschillende genen die een rol spelen in elk relevant chitine afbrekend moleculair systeem werden geïdentificeerd. Daarna zochten we naar genen die een potentieel plantengroei-bevorderend effect hebben en genen die een rol spelen in het plantaardig afweersysteem en de stikstof cyclus. We vonden dat het *Mortierella hyalina* genoom boordevol nuttige genen zit voor biocontrole en nutriënten cycli en we concluderen dat deze fungus veel potentieel heeft om een belangrijke speler te worden in hernieuwbare gewasbescherming en -bebouwing.

Literature Research

1. Introduction

1.1 Lettuce cultivation in Flanders

The cultivation of lettuce is of great economic importance to Flanders. With a yearly production of 39,500 tonnes (Maertens et al., 2014), and an annual revenue of 37 million euros (Bergen, 2013), lettuce is the second most cultured greenhouse horticultural product in Flanders. Tomatoes (47.7%) are cultivated the most in greenhouses, and bell pepper completes the top 3 with a production area of 8.8% of the greenhouses (Platteau et al., 2019, Figure 1). In total, 790 hectares of land area is used for lettuce cultivation, which is approximately 14.8% of the total area for greenhouse horticulture. Most of these horticultural products are meant for export, which makes greenhouse horticulture strongly embedded in the region's economy (Maertens et al., 2014).

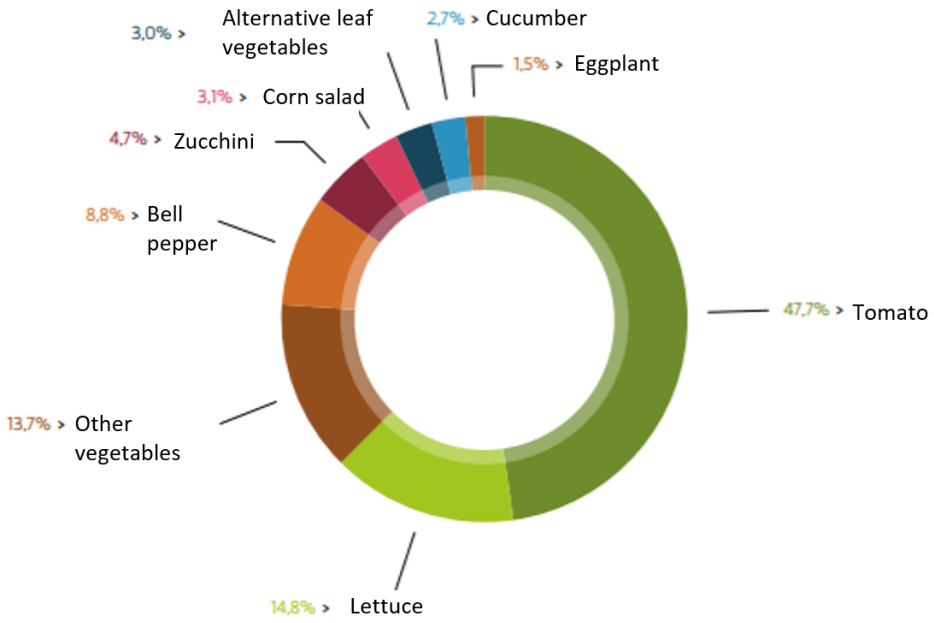


Figure 1: Area used for greenhouse horticulture in Flanders in 2017, adapted from Department of Agriculture and Fisheries

As consumer expectations rise, importers pressurize farmers to assure that their crops meet a certain quality standard. Cosmetic expectations are often the reason for food waste. Fruits or vegetables that are even slightly dented or discolored are discarded or downgraded, leaving the farmer with lower profit. In Flemish horticulture, approximately 283.000 tons of harvested biomass is wasted. This food waste is downgraded to non-human consumption such as animal feed, or is simply plowed back into the soil (Gellynck et al., 2017). As for lettuce, the average sales loss (product that cannot be sold due to unsatisfactory quality) in 2017 was 9.1% (Gellynck et al., 2017).

Next to food waste, two other factors are important in influencing horticultural profit: lower crop yield due to (1) nutritional deficiencies, often solved by incorporating mineral fertilizers, and (2) plant and human diseases (plant pathogens and zoonotic pathogens, Debode et al., 2016), from which plants are protected by chemical crop protection products.

1.2 Lettuce nutritional needs

Lettuce grows best at moderate temperatures of 25 °C during day and 8 °C during night. Since climatic conditions for lettuce horticulture in Flanders can be optimized in a greenhouse, lettuce is actively produced all year around. The soil should be well-drained, and pH should be between 6.0 and 7.0. Adequate nutrition should contain mainly nitrogen, phosphorus, and potassium, and secondly calcium and magnesium (Farm Services, 2017) to promote a good plant development. Farmers opt for fertilizer use to provide lettuce with a sufficient amount of nutrients. Mass producers now operate a monoculture-based lettuce cultivation system, but this horticultural intensification causes soil to be depleted much quicker, which has led to nutrient recycling being replaced by fertilizer usage (Barriere et al., 2014). Next to soil, lettuce is also grown in hydroponic systems where nutrient addition is paramount (Sambo et al., 2019).

1.3 Fertilizer

The most used fertilizer is NPK (Figure 2). NPK can be tailored to include other nutrients such as Ca and Mg, and separately produced components can be blended to obtain a custom fertilizer. Regardless of how flexible this way of nurturing the crop might appear, the use of fertilizer is hardly ever optimal for large scale cultivation as there are certain challenges that need to be faced.

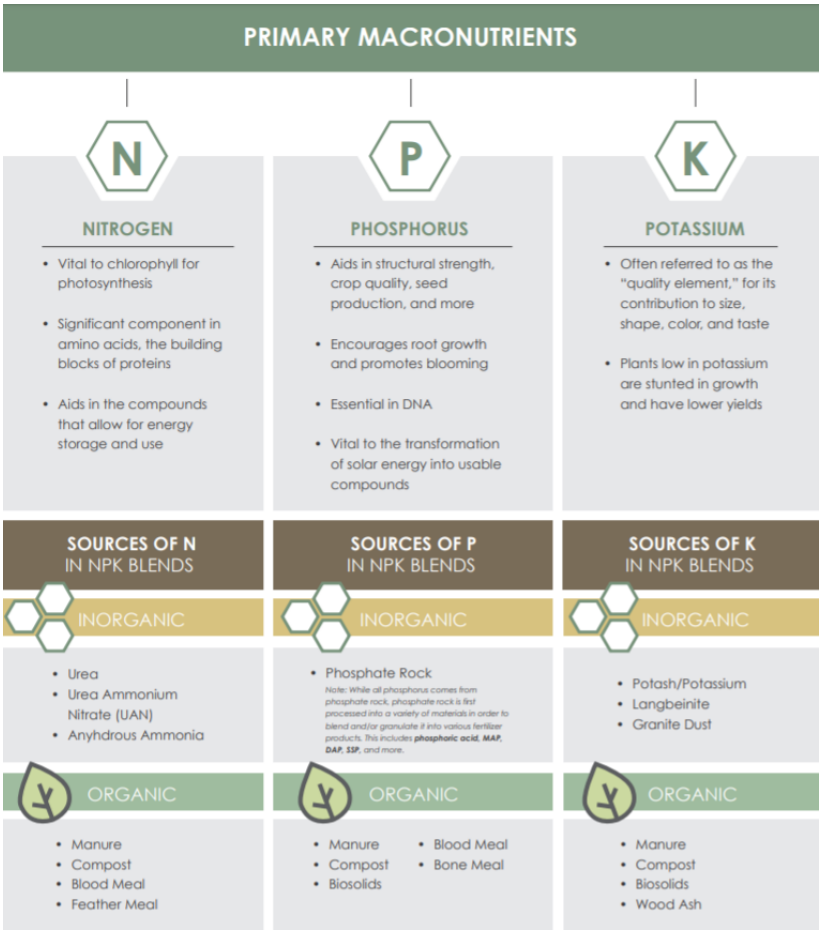


Figure 2: Primary macronutrients of NPK, adapted from FEECO.com/npk-fertilizer

Fertilizers release high amounts of nutrients at the time of application. This sudden surplus in nutrition can have adverse effects on not only the crop of interest, e.g. excessive nitrogen can lead to soft rot and tipburn (Farm Services, 2017), but also affect the environment. Nitrogen in soil exists in two forms: insoluble organic nitrogen, fixed in organic compounds, and water-soluble mineral nitrogen: ammonium and nitrate. Plants can only absorb the water-soluble mineral nutrients, which are often the main ingredients of a chemical fertilizer (Yara Vlaardingen, 2019).

When large amounts of nutrients are applied to the soil it can take a long time before they are fixed by plant roots or soil microbes, depending on soil type and microbial composition. Leaching occurs when heavy rainfall or irrigation washes away water-soluble minerals in the soil, depriving the crops of their nutrients and contaminating the ground water (Yara Vlaardingen, 2019). This is especially the case in lettuce horticulture, where constant irrigation is required to achieve a marketable yield (Farm Services, 2017). Nitrogen and phosphorus leaching are known to cause eutrophication of the ground water, which is detrimental to the biodiversity and sustainability of ecosystems and may cause oxygen depletion of the water body (Schindler et al., 2004). Sandy soils are considerably more affected by nutrient leaching than clay soils (Yara Vlaardingen, 2019). We therefore need smarter alternatives to prevent nutrient leaching while still supplying adequate nutrition to the plant, preferably tailored to the plant and the soil type.

1.4 Chemical crop protection

Lettuce is sensitive to various diseases caused by bacterial, viral, or fungal infection such as anthracnose, *Fusarium wilt* disease, downy mildew, soft rot and lettuce mosaic (Farm Services, 2017). In addition, other factors such as weeds competing for water and sunlight, and insects and nematodes feeding on the crop, can dramatically reduce crop productivity. Chemical crop protection provides a relatively cheap and effective way to combat these pests and maximize yield. The main classes of crop protection chemicals are herbicides, insecticides, and fungicides (Armstrong & Clough, 2009). Research by Oerke et al. (2006) has shown that without chemical crop protection the farmer's productivity could be halved.

Synthetic crop protection chemicals, being designed to kill living things, are often dangerous to the farmer who uses them and can have detrimental effects on the environment. Chemical crop protection products have been linked to cancer and neural damage (Alavanja et al., 2004) and are known to cause birth defects (Larsen et al., 2017). Farmers are being exposed to these chemicals at a far higher rate than the consumer (Damalas & Koutroubas, 2016). Especially in third-world countries, where few regulations on surrounding crop protection are in place, poor farmers who need to maximize their efforts to protect their crops in order to provide for their family are the biggest victims of this irresponsible use of dangerous synthetic chemicals (Ecobichon, 2001). In addition, these products often persist in the environment for a long time after initial application to the crop, which can have devastating effects on local wildlife due to their high biological activity. In the worst case, chemical crop protection products can, by the same principle of fertilizer leeching, enter the groundwater or water streams and kill plants, animals, and microorganisms alike, disturbing entire ecosystems (Hayes T.B., 2010).

The European commission's Good Agricultural Practice instructs farmers to use chemical crop protection products as little as possible and only when necessary (ECPA, 2014). Nevertheless, in this day and age chemical crop protection is used systematically as yield insurance, rather than sparsely in case of disease outbreak (Lamine et al., 2010). This overuse of chemical crop protection is to blame for trace amounts of chemicals on harvested produce, called residues. However small the amount of residue on consumer food may be; it is in fact still compliant with the European norm in 98.5% of cases (EFSA, 2019), a chemical crop protection-free or chemical crop protection-poor food industry seems a noble cause, particularly for the health of the farmer and for the environment. Just like the well-known problem of antibiotic resistance in human and animal pathogens, chemical crop protection resistance is a growing problem in lettuce pathogens, and it might have a devastating impact on the future of greenhouse horticulture (Davet et al., 1993, Brown et al., 2004). Alternative methods are to ensure the health and quality of horticultural produce, without the need for dangerous synthetic chemicals that can cause evolution of superbugs, destruction of ecosystems and birth defects in developing countries.

1.5 Research Perspectives

There might exist a way to suit the nutritional needs of lettuce in potting soil and protect it from disease without the need to spray chemicals on the crops. Addition of chitin to soil is linked to higher yield and a stimulated plant immune system (Sharp, 2013; Debode et al, 2016). The great advantage of this method is that chitin is a biodegradable and very abundant natural polymer, and its use is expected to have zero implications on the health of the farmer or consumer and very little effect on large scale ecosystems. We aim to study whether chitin has a similar effect on lettuce yield and health when added to potting soil.

2. Chitin

2.1 Structure, Presence & Properties

Chitin is the world's second most abundant biopolymer, only being surpassed by cellulose (Gooday, 1990). It is a structural polysaccharide, present in the cell wall of fungi, the exoskeletons of insects, arachnids, crustaceans, and the eggs of nematodes (Ramirez et al., 2010). Chitin is polymeric β-1,4-N-acetylglucosamine. The structure of chitin very much resembles that of cellulose, save for the acetamide group attached to C2 where cellulose has a hydroxyl group (Figure 3). Similarly to cellulose, chitin strands are linked together with hydrogen bonds and are organized into microfibrils, which provide support and structure to cellular components like the cell wall (Ohno, 2007).

Even though chitin is insoluble in water, it has potential to be impactful to a biome when chitinolytic microorganisms are present. Enzymes involved in the chitin cycle are classified in glycosyl hydrolase families (De Tender et al., 2019). The enzyme chitin deacetylase is able to deacetylate chitin, which results in chitosan (Figure 3, Figure 4) and acetate. Chitosan is soluble in water and dilute acidic solutions, at a pH suitable for cellular life (Ramirez et al., 2010). Chitosan can then be further degraded by chitosanase and glucosaminidase enzymes to N-glucosamines (De Tender et al., 2019).

Chitin can also be degraded through another pathway, by way of endochitinases and β-1,4- N-acetylglucosaminidases (Figure 4). Endochitinase cuts chitin chains at internal β-1,4-bonds. This

reaction generates multiple smaller chitin chains, and often the dimer product di-acetylchitobiose and the trimer chitotriose. Exochitinases such as β-1,4- N-acetylglucosaminidases are able to split these small multimers into the monomeric N-acetylglucosamine (GlcNAc) (De Tender et al., 2019; Ilangumaran et al., 2017).

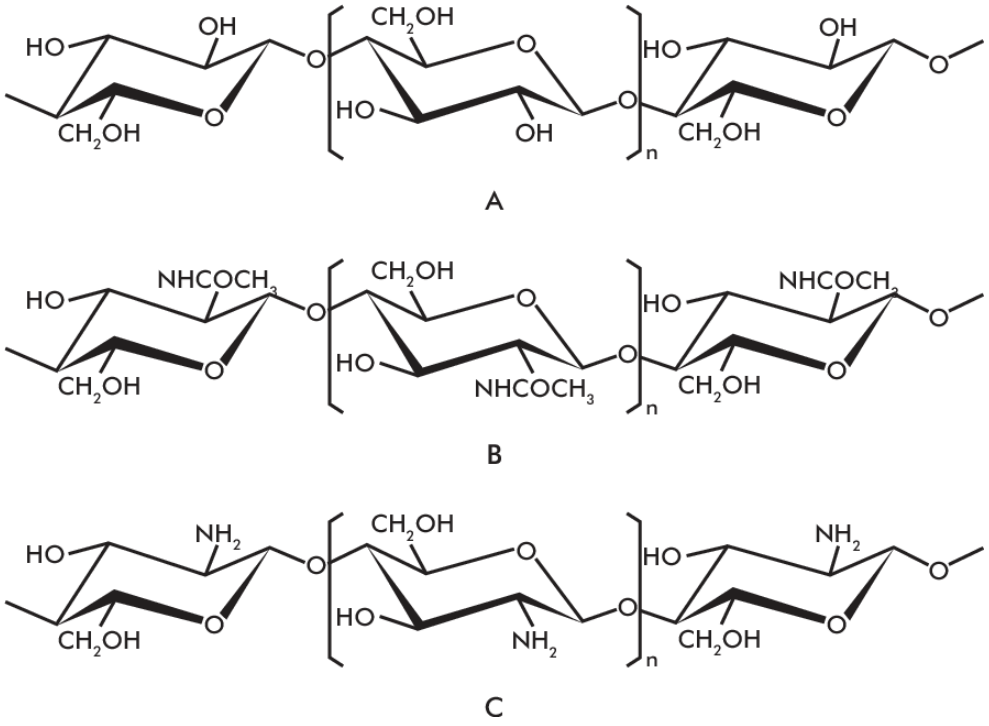


Figure 3: Structure of A: Cellulose, B: Chitin, and C: Chitosan, Ramirez et al., 2010

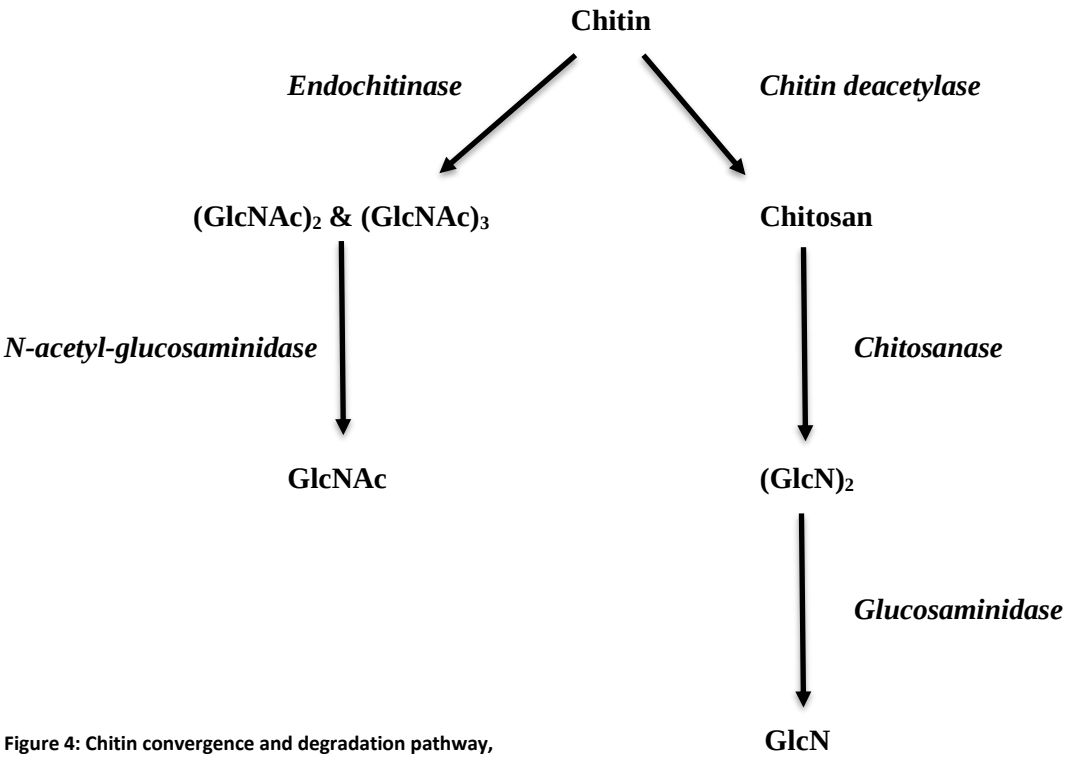


Figure 4: Chitin convergence and degradation pathway, adapted from De Tender et al., 2019

2.2 Effect in soil

Chitin addition to soil is known to improve soil quality, plant growth and plant resilience (Sharp, 2013). El Hadrami et al. (2010) have shown that chitin and chitosan are responsible for the accumulation of phytoalexins and proteinase inhibitors, and activation of other plant protection mechanisms such as lignin synthesis and callose synthesis. Ramirez et al. (2010) have shown that addition of chitin to soil facilitates the growth of plant growth promoting mycorrhizas and *Rhizobium* species, and that the growth of chitinolytic microorganisms, that act as biological controls against common plant pests, is stimulated. Chitosan is a known Microbe Associated Molecular Pattern (MAMP) that triggers plant defense mechanisms against chitin-containing harmful organisms. Chitin and its derivates are also known to show antibacterial, antiviral, antifungal and nematicidal activity in plants (Ramirez et al., 2010). Chitosan's cationic properties can disrupt potassium signaling in bacterial pathogens, and can interfere with membrane integrity in vacuoles and other endomembrane organelles in fungal pathogens (Ilangumaran et al., 2017).

Changes in rhizosphere microbiology may also affect the plant physiology and its capacity to be colonized (Gryndler et al., 2003). The rhizosphere is defined as the soil around the roots that is influenced by the root (Hiltner, 1904). Debode et al. (2016) hypothesize that chitin addition to potting soil favors chitin-degrading microorganisms, among which certain plant growth promoting rhizobacteria (PGPR) and plant growth promoting fungi (PGPF), protecting the plant from harmful pathogens. They found that chitin addition in soil successfully increased yield, and had a decreasing effect on human pathogens, e.g. *Salmonella*, on the leaves. The rhizosphere microbiome changed towards a more favorable environment for lettuce cultivation.

Estimations are that yearly 170,000 tons of chitinous waste comes from global fish industry, because of the exoskeletons of crustaceans (Ramirez et al., 2010). Exploiting chitin for its plant growth promoting properties would be an environmentally friendly way of recycling this otherwise wasted resource, all the while cutting down on chemical crop protection products and fertilizers. In this thesis we investigate the effect that chitin has on the microbiome in the rhizosphere and bulk potting soil, and its plant growth promoting properties in lettuce. Debode et al. (2016) and De Tender et al. (2019) demonstrated that the presence of the fungus *Mortierella* highly increased in the lettuce rhizosphere due to chitin addition to potting soil.

3. *Mortierella*

We hypothesize that a *Mortierella hyalina* strain that we cultured in the lab is able to metabolize chitin and has a plant growth promoting effect on lettuce in potting soil. *Mortierella hyalina* is a fungus of the phylum *Mucuromycota*, subphylum *Mucuromycotina*, order *Mortierellales*, family *Mortierellaceae*. There is no whole genomic assembly data of *M. hyalina* in public databases.

3.1 Presence

Mortierella are soil dwelling saprophytes, able to decompose dead plant material. They can grow in deep humus layers of the bulk soil but also colonize the roots of the plants and thrive in the rhizosphere, where they are able to dominate the microbiome by releasing antibiotics (Li et al., 2017). Shen et al. (2014) and Xiong et al. (2014) have shown that colonization of plant roots by

Mortierella species is linked to suppression of *Fusarium wilt* disease in soil and a decreased disease rate in plants. *Mortierella* species are not known to be pathogenic towards plants or animals (Li et al., 2017).

3.2 Functions

Fungi of the subphylum *Mucuromycotina* are used as a model in biotechnology for their lipid metabolism and are an important industrial producer of lipids and lipid-derived products (Etienne et al., 2014). *Mortierella* are able to aid mycorrhizal fungi in phosphorus acquisition, and can metabolize toxic polycyclic aromatic hydrocarbons such as naphthalene, which are dangerous soil contaminants mostly found in heavy industrialized areas (Mukhopadhyay et al., 2017). Like many fungi, *Mortierella* is a producer of melanin, which contributes to more stable soil organic matter accumulation (Fernandez & Koide, 2014).

Mortierella elongata is able to synthesize the plant growth hormone indole-3-acetic acid (IAA) and stress hormone abscisic acid (ABA). IAA is the most common auxin in nature. It is necessary for root development and greatly improves crop yield. Inoculation with *M. elongata* and subsequent increase in IAA levels in the rhizosphere have resulted in an increase of harvested biomass in maize (Li et al., 2017). ABA is the most important plant stress hormone, and increased levels of ABA might indicate increased resistance to stress. With the ability to synthesize IAA and ABA, *M. elongata* is able to colonize the rhizosphere without alarming the innate defense mechanisms of the plant (Li et al., 2017; Spaepen et al, 2007; Yasuda et al., 2008). *M. hyalina* is able to colonize the *Arabidopsis thaliana* rhizosphere and facilitate phosphate uptake of the plant which results in a substantial increase in aboveground biomass. (Johnson et al., 2019)

Most importantly to this thesis, Kim et al. (2008) have shown that *Mortierella sp.* are able to enzymatically deacetylate chitin by releasing an extracellular chitin deacetylase into the environment.

3.3 Genome

Since the genome of *M. hyalina* is not yet publicly known, we need to estimate its genome size by looking at its close relatives. *M. alpina* has a genome size of 39.53 Mb and a GC content of 50.4% (Etienne et al., 2014). *M. elongata* has a genome size of 49.85 Mb and a GC content of 48.1% (DOE Joint Genome Institute, 2016). Note that even though these two species are of the same genus, there is a massive disparity in genome size. We know that gene loss is an important driving force in fungal evolution (Spanu et al., 2010), so this is not surprising, but it does complicate our efforts to make an estimation about the genome size of *M. hyalina*.

4. High Throughput Sequencing

4.1 Second Generation Sequencing

Second Generation Sequencing (SGS), or Next Generation Sequencing (NGS) is a hypernym for sequencing techniques that use massively parallel sequencing. Massively parallel sequencing distinguishes itself from the older Sanger sequencing, also known as first generation or dideoxy sequencing, in that millions of nucleotide fragments can be sequenced simultaneously, while the

Sanger method sequences only a single fragment. Because of this advantage, NGS enables us to sequence environmental samples that can contain the DNA of many of individuals (Illumina Inc., 2020).

NGS has a much higher sensitivity and coverage than Sanger sequencing (Shendure et al., 2008; Schuster, 2008), which has enabled researchers in the last decade to identify new mutations correlated with genetic diseases, and novel high-throughput techniques such as exome sequencing and RNA sequencing have been developed (Churko et al., 2013). NGS comprises many technologies, such as pyrosequencing, sequencing by synthesis, sequencing by ligation and ion semiconductor sequencing (Applied Biological Materials, 2015). In this thesis we used Illumina's sequencing by synthesis method. Because of its relatively long read length of 2x300 bp, we used the Illumina MiSeq for metabarcoding. The Illumina HiSeq3000 was used for whole genome shotgun sequencing, because of its higher coverage. Both methods are summarized in Table 1.

	MiSeq	HiSeq 3000
Read Length	Up to 2x300 bp	Up to 2x150 bp
Quality Scores	≥ 70% bases higher than Q30	≥ 75% of bases above Q30
Reads per run	44-50 million	2.1-2.5 billion
Run Time	Approx. 56 hours	< 1-3.5 days

Table 1: Illumina MiSeq and HiSeq 3000 specifications

4.2 Illumina sequencing

Illumina sequencing after library preparation (discussed below) comprises 2 steps. First clusters of DNA template are generated on the flow cell, then the clusters are sequenced by synthesis (Illumina Inc., 2016).

To generate clusters, the DNA library is immobilized on a flow cell by hybridization through the adaptor sequences (De Visscher, 2019). Then a complement of the DNA template is synthesized by a polymerase. The double stranded molecule is denatured, and the original template is washed away. The strands are subsequently hybridized through bridge amplification (Figure 5). After several rounds of bridge amplification, the reverse strands are cleaved and washed off, and the flow cell is inhabited by clusters of thousands of identical copies of the original DNA template. The 3' ends are blocked to prevent unwanted priming (Illumina Inc., 2016).

Now sequencing by synthesis begins. The first sequencing primer is extended to produce the first read. Fluorescently tagged deoxynucleotides are added to the flow cell, and in each cycle, 1 nucleotide is incorporated into the growing chain, based on the sequence of the template. The unbound nucleotides are then washed away, and the clusters are excited with a light source. The fluorescence signal from each cluster is detected and through its wavelength and intensity the incorporated nucleotide can be inferred. After each cycle, the fluorescent tag is cleaved off, so the next nucleotide may be incorporated in the chain. The length of the read is determined by the amount of cycles in this process. When the appropriate read length is obtained, the product is washed away, the first index is sequenced and the 3' ends are deprotected, which causes the template to fold over and the reverse strand to be synthesized, resulting in a double stranded bridge.

The forward strands are then washed off and the reverse strands (and the second index) are sequenced in an analogue way (Illumina Inc., 2016).

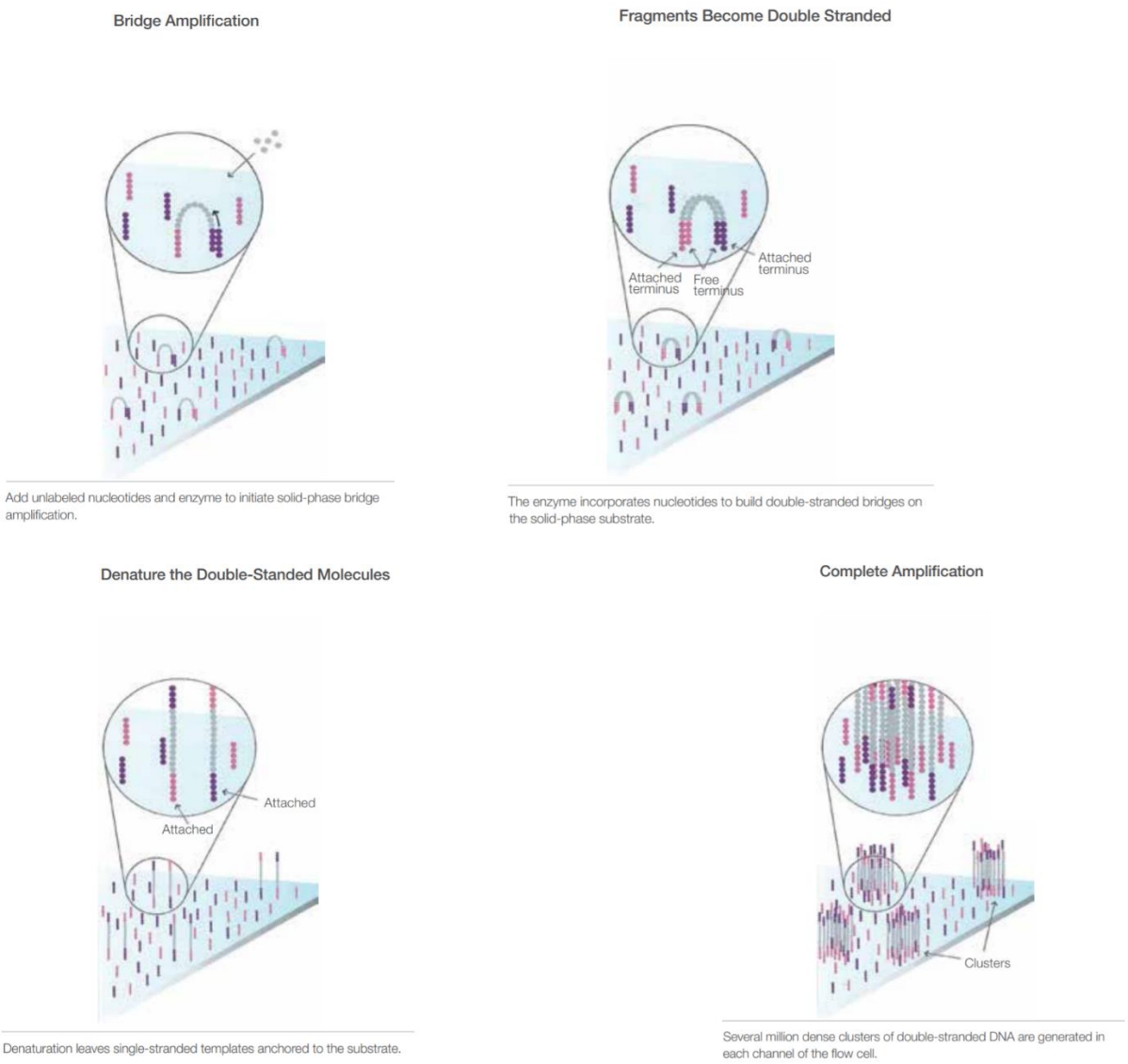


Figure 5: Bridge amplification, adapted from De Visscher (2019), Illumina (2010)

4.3 Metabarcoding

When considering samples that contain a community of microorganisms, the great plate anomaly (Staley and Konopka, 1985) states that there is a massive divergence between the numbers of colony forming units on a plate, and cell counts retrieved from microscopic examination (Renella et al., 2014). Therefore we lack immense amounts of data when studying soil and other ecosystems in traditional ways. Second generation sequencing and bioinformatics offer a solution to this challenge, as we can directly sequence the genetic information of all microorganisms present in the sample, including those strains that cannot be cultivated in laboratory conditions. The nucleic acids are extracted from the environmental sample and form the metagenome (Handelsman et al., 1998).

In metabarcoding, also referred to as amplicon sequencing, a PCR is performed on an environmental DNA sample, thus containing genetic information of the whole microbiome. PCR primers select a specific region in the genome, which is ubiquitously present among a certain clade, e.g. all bacterial or fungal sequences (ChunLab Inc., 2019). In bacterial metabarcoding we often choose the 16S rRNA, and for the identification of fungi we opt for the ITS rRNA. The power of 16S and ITS sequencing lies in this fact that the PCR primer can be developed to select a certain conserved region in order to capture specifically the desired genetic locus of an incredibly diverse group of organisms, yet the amplified region covers a hypervariable domain in the genome that allows us to accurately discriminate between organisms up to the genus level (Janda & Abbott, 2007). Also, for both barcodes, an extensive public database is available (SILVA & UNITE), which contains thousands of sequences linked to a species name (Quast et al., 2013; Nilsson et al., 2018.)

There are 2 PCR steps needed before Illumina MiSeq amplicon sequencing by synthesis (as described above) takes place, this is called library preparation (Figure 6). First there is an amplicon PCR step, where the genomic sequence in the sample is captured by a region of interest-specific primer, attached to an adaptor sequence. Next there is an index PCR step, where multiplex indices and Illumina sequencing adapters (P5 and P7, Figure 6) attached to an adaptor sequence are added to the mixture. This adaptor sequence targets the adaptor sequence attached to the primers. The resulting library consists of an amplified region of interest, that can attach to the Illumina MiSeq instrument through the P5 and P7 adapters, and that contains multiplex indices to discriminate between samples (Illumina Inc., 2013).

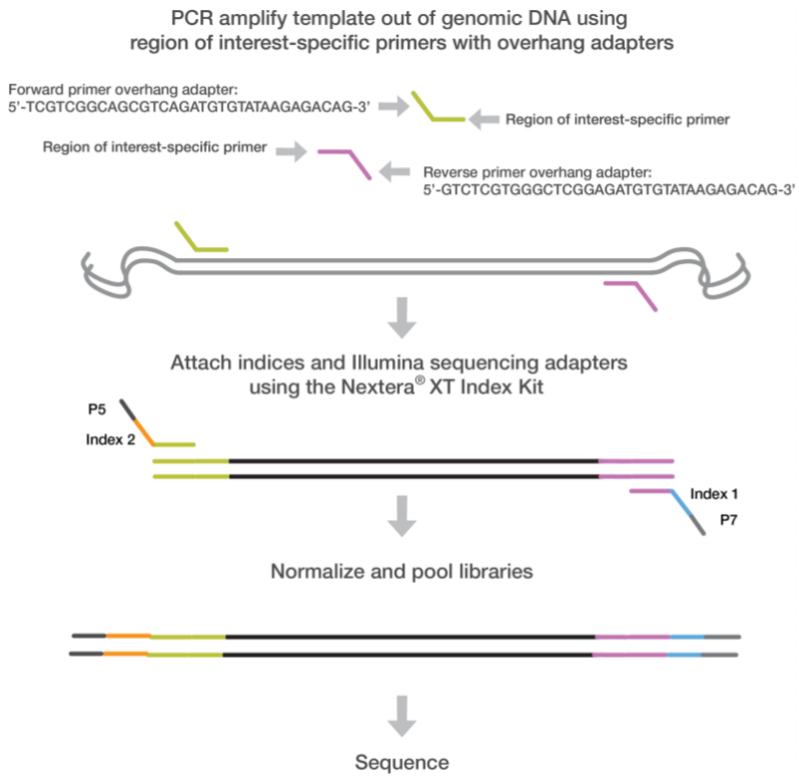


Figure 6: Library preparation for metabarcoding, reprinted from Support.Illumina.com, '16S Metagenomic Sequencing Library Preparation', 2013

Finally, after normalizing the samples with a fluorometric quantification method (Figure 6, Illumina Inc., 2013), sequencing by synthesis on the Illumina MiSeq instrument allows us to read the amplicon from both ends. Forward (F) and Reverse (R) reads are generated, but we need to make sure in the experimental setup that the amplicon is shorter than the read length of the sequencing technology used, so there exists an overlap from both read ends (Figure 7). By combining the information captured in the F and R reads we can accurately tell the sequence of the amplicon (Illumina Inc., 2013).

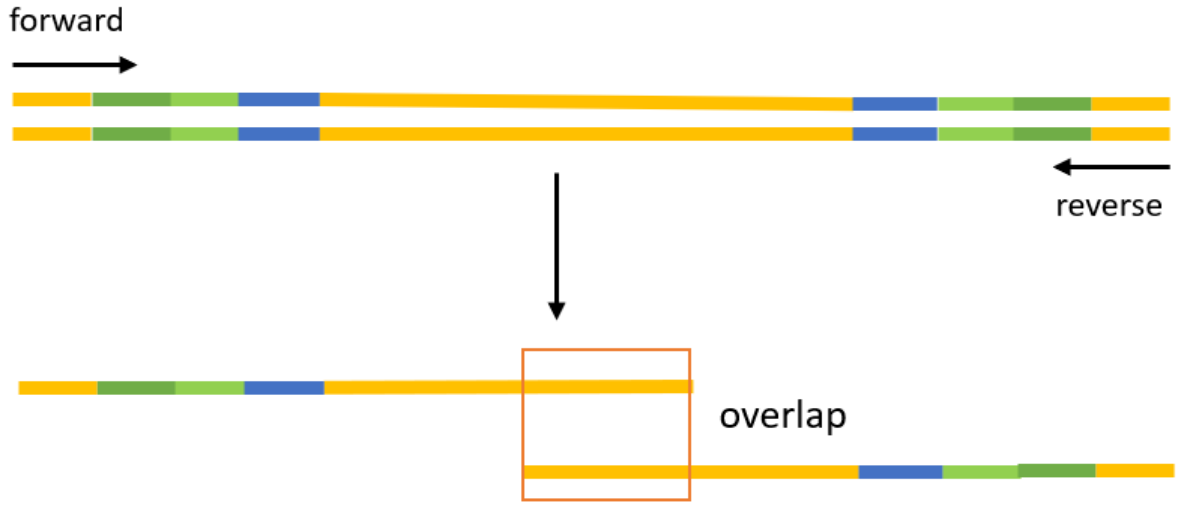


Figure 7: Overlap between forward and reverse reads allows for high-resolution sequencing

4.3.1 ITS/16S loci

Typically, ribosomal RNA (rRNA) is used for taxonomical classification. Ribosomes are essential to biological life as we know it, because of their role in protein synthesis (O'Connor et al., 2010). Certain ribosomal sequences have been conserved in evolution because the 3D structure is an important factor of ribosome functionality. Small differences in sequence identity can have a massive impact on 3D structure, which results in evolutionary conservation, and the rRNA is assumed to be nearly absent of horizontal gene transfer (Větrovský & Baldrian, 2013; Fox et al., 1977; De Visscher, 2019). These are the regions that will be targeted by the PCR primers. In between these conserved regions are highly variable bases, with little or no biological significance, which can freely and randomly undergo mutation. These regions can be exploited to identify interspecies differences, and we can assess a degree of evolutionary distance from these hypervariable regions (Fox et al., 1977).

The eukaryotic rRNA contains hypervariable Internally Transcribed Spacer, or ITS, regions (Figure 8). The multicopy nature of the ITS regions and high sequence variability makes them fit for phylogenetic analysis (Ghosh et al., 2019). Both ITS1 and ITS2 are a popular choice, in this study we targeted the ITS2 region for metabarcoding with the Illumina MiSeq (2x300 bp) because the primer is more conserved than in the ITS1 region. The length of the ITS regions can be variable between distinct taxa, so we need to take this into consideration during data analysis. The ITS2 region can vary from 100 up to 1200 bp, but most commonly occurs in the range of 200 to 500 bp (Yao et al., 2010).

Another popular target for metabarcoding of eukaryotes is the 18S rRNA region, yet this region lacks discriminative power in fungi, and the ITS regions have higher sequence variability on the genus level (Ghosh et al., 2019).

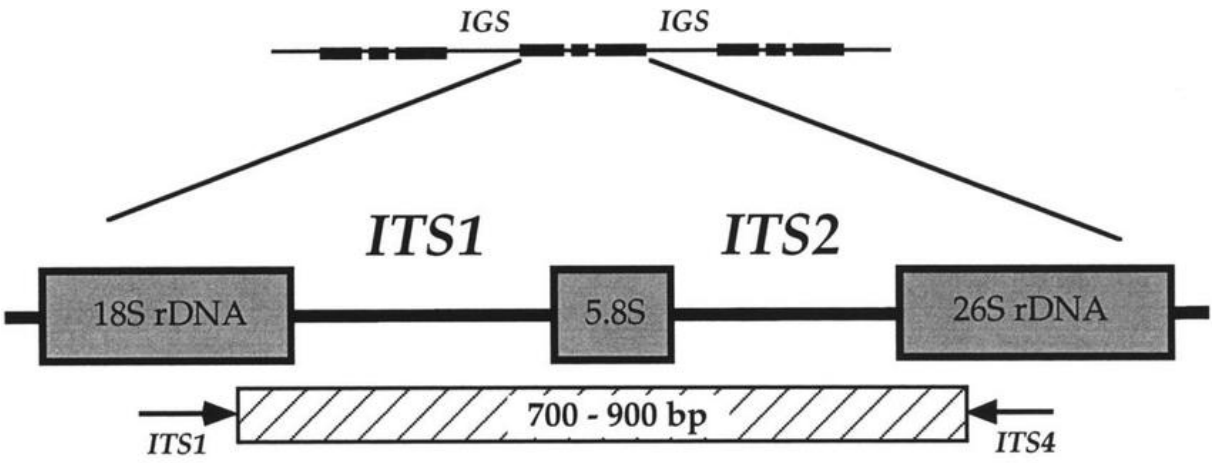


Figure 8: ITS rRNA region, adapted from Shaw et al., 2002

In bacteria the 16S rRNA is used for metabarcoding (Figure 9). The 16S region is around 1500 bp long (ChunLab Inc., 2019). We targeted the V3-V4 hypervariable region (420 bp), fit for analysis with the Illumina MiSeq (2x300 bp). The V3-V4 region shows highest nucleotide heterogeneity, and thus is most fit for a high-resolution discrimination on the genus level, as stated by Klindworth et al. (2013).

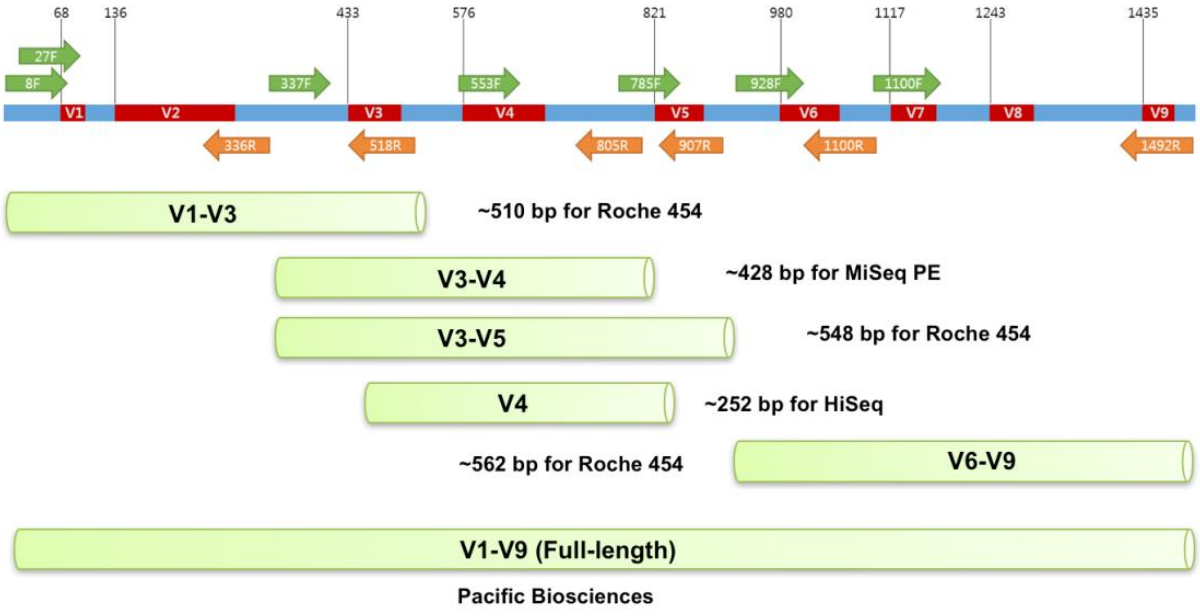


Figure 9: 16S rRNA region, ChunLab Inc., 2019

4.3.2 Data analysis

In classic metabarcoding data analysis, Operational Taxonomic Units (OTU's) are used to reduce technical error misinterpretation as biological variation. OTU's are clusters of amplified sequences

that differ less than a fixed dissimilarity threshold (often 3% sequence dissimilarity). The problem with OTU's however is that they under-utilize the quality of modern sequencing technology, with fine-scale variation and high-resolution sequencing (Callahan et al., 2016). Also, because OTU's are defined by comparing sequences to each other in a single dataset, the emerging OTU's are features of this dataset. For example, the sequences of the OTU's depend on relative abundances in the sampled microbial community. This introduces unwanted bias in the data, and makes it less fit to compare your data to external databases (Callahan et al., 2016).

In this thesis we use DADA2, a novel software to analyze metabarcoding data. DADA2 groups sequences into Amplicon Sequence Variants (ASV's), which is a more precise clustering unit than OTU's, and uses an error model that assumes that biological variation is more abundant than sequencing errors. Because of this assumption, ASV's with as few as only 1 difference in sequence identity, that would be grouped in the same OTU, can coexist as long as both sequences are adequately abundant. It is clear that this method is able to capture more biological variation and thus has better resolution as a method of community profiling, while the total amount of sequences in the sample is reduced due to errors being corrected (Callahan et al., 2016).

It is important to realize that neither OTU's, nor ASV's represent biological species. They are merely a concept used for sequence clustering, and it is very possible that multiple OTU's or ASV's correspond to the same species, or that sequences from the same species are classified in different clusters. We transform the sequence data into ASV's as a way of organizing the data by optimally utilizing the quality of Illumina MiSeq, and in further downstream analysis we will compare these ASV's to a database where we can assign taxonomy to each ASV.

4.4 Whole Genome Shotgun sequencing and genome assembly

Early methods to perform genome sequencing and assembly used a clone-based gene mapping approach. This approach meant splitting up the whole genome into smaller segments (around 40 to 200 Kb), and cloning them into separate Bacterial Artificial Chromosomes (BAC's). Then the BAC's are sequentially analyzed, typically by Sanger sequencing at the time, and the sequences are pasted together. This clone-based approach however is extremely tedious, slow, and labor intensive (Cook, 2019).

Whole Genome Shotgun (WGS) analysis provides a smart solution for this problem. WGS analysis is a faster way to perform genome assembly, and was used by Craig Venter of Celera Genomics in his quest to assemble the human genome before the publicly funded Human Genome Project could. It uses algorithms that are able to combine sequence reads by overlap into contigs and scaffolds, across the entire genome. This allows the full fragmentation of the genome without any cloning steps, which creates a shift from tedious laboratory work to a fast, computational analysis (Venter, 2001). While this analysis is certainly faster than the clone-based approach, Sanger sequencing technology was still used for gathering the data. The true revolutionary development in genome sequencing was the rise of SGS technologies, which has made WGS much faster, cheaper, and accessible (Koboldt et al., 2013).

WGS analysis is performed by first breaking up the DNA sample into many overlapping fragments, and sequencing all these fragments in parallel. These fragments are then algorithmically assembled into larger contigs and from read pairs in these contigs scaffolds are built (Figure 10). A read pair is a sequence read of which one end is assembled into one contig, and the other is assembled into another contig, effectively linking the two contigs together without the need for overlap with other sequence reads (Green E.D., 2001). On these scaffolds, certain marker genes, or genetic sequence-based landmarks, can be identified in order to map the contigs correctly to a reference genome that shows the same landmarks (Figure 10, red circles).

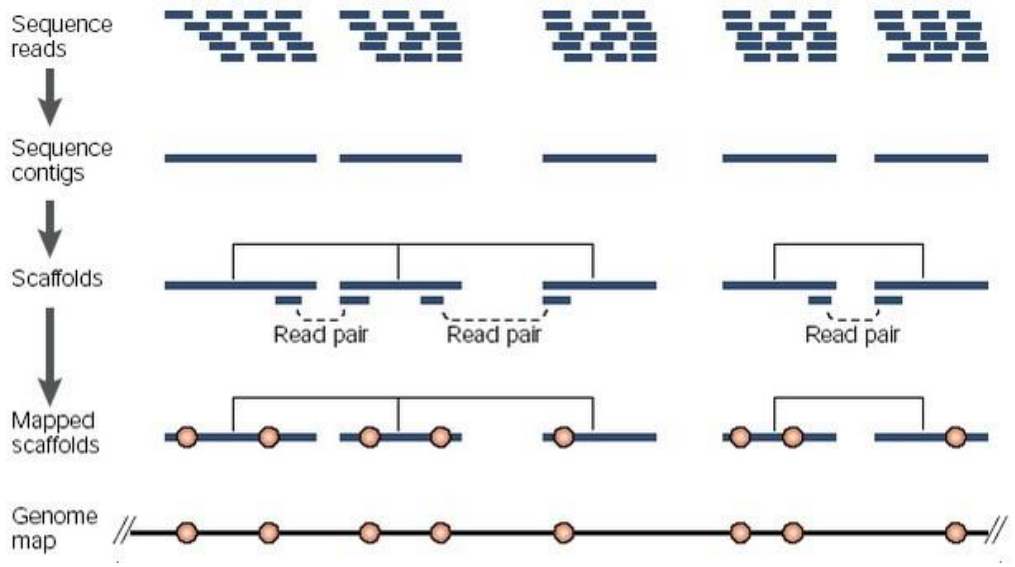


Figure 10: Overview of WGS analysis, adapted from Nature Reviews Genetics, 2001

There remains one major pitfall. Eukaryotic genomes contain many repetitive elements, caused by gene duplications and transposable elements, which complicates the assembly. To partially solve this issue, the genome was sequenced as paired end (2x150 bp) reads by Illumina HiSeq (Figure 11). Paired-end reads align less ambiguously to the reference sequence than single-end reads, especially in highly repetitive regions. The use of paired-end reads allows for the construction of high-quality, long contigs, able to identify INDELs (insertions & deletions) and inversions, where single-end data might provide confusing results (Illumina Inc., 2019). The use of paired-end data is essential in Eukaryotic genome assembly, where repeats are very common. Where a single-end sequence read might align to multiple identical positions in repetitive elements in the genome, the paired-end read provides extra information from the other end of the sequence read, which can help identify which repeat it should align to, provided the other end is located outside the repeat. By assembling large contigs of paired-end data we can increase the probability that large repeat areas are entirely crossed by the contig (Illumina Inc., 2019). The use of paired-end reads however does not guarantee that the repetitive mapping problem is solved. The quality of the assembly can be significantly improved by using longer read lengths.

Long reads have a great advantage over short reads because they enable us to better solve the mapping problem in repetitive regions: long reads can span entire repetitive regions and thus remove any ambiguity. Third Generation Sequencing (TGS) can provide reads averaging around 10 kb, albeit

with high per sequence error rates of up to 15%. Technologies such as PacBio and Nanopore sequencing can create these reads. Best practice would be to combine long and short reads into one assembly, since short reads have different error profiles, and they would be complementary to each other (Miller et al., 2017). This however is not performed in this thesis since our study is limited to an exploratory functional genomic analysis, and thus an exact mapping of the chromosomes is not required.

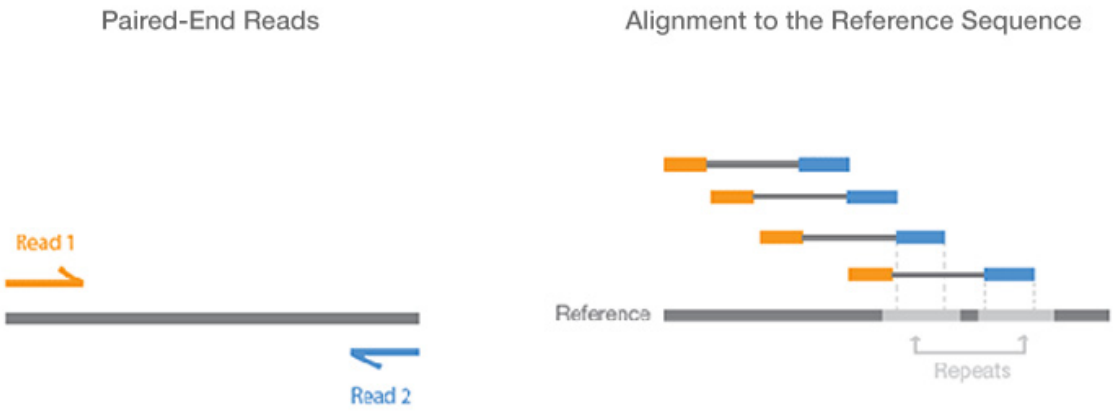


Figure 11: Paired-end sequencing, adapted from Illumina Inc., 2019

5. Aim of the study

A previous study at ILVO has shown that addition of chitin to potting soil has a plant growth promoting effect on lettuce, and that this effect is correlated with the presence of a fungus on the substrate (Barneveldt, 2019). In this thesis we wish to investigate the abundance of this fungus in the samples, and determine and quantify the effect of chitin on the bacterial and fungal community. We will explore the differences between microbial populations in the rhizosphere and the bulk potting soil, and to account for evolution in the microbial population over time, we will measure the effects at two timepoints, being after 4 and 8 weeks.

The aforementioned fungus is hypothesized to be *Mortierella hyalina*. To perform a functional genomic analysis of this fungus, we will be assembling its genome. The *M. hyalina* genome has never been assembled, so we do not have the possibility to map our reads against existing references. This however is not a problem, since the goal of our assembly is not to map the entire genome. We aim to identify genes that are relevant to this study, this being (1) genes that provide a selective advantage of *M. hyalina* over other soil organisms in the presence of chitin, and (2) genes that promote lettuce growth and stimulate lettuce's immune system. These genes can be identified from the list of contigs that the assembly software outputs, and in order to find these genes there is no need to sort the contigs into the right order with respect to the genome sequence. We will particularly look for chitinase and N-acetyl-glucosaminidase genes, because of their role in chitin degradation, and genes linked to the N-cycle, for they might be plant growth promoting by releasing highly nutritional nitrogen-containing compounds in the soil, which the plant may use to its advantage.

Outline

Chitin is the second most abundant biopolymer in nature, and has been the subject of many recent studies for its plant-growth promoting properties. The aim of this thesis is twofold. First, we will explore how the addition of chitin to the growth medium shifts the biodiversity of the lettuce microbiome, and second, we will identify the chitin-degrading role a *Mortierella* fungus.

Therefore, this thesis is organized into 3 chapters:

1) Metabarcoding of soil microbiome after chitin treatment
2) Phylogeny of the *Mortierella* strain based on ITS
3) Genome assembly and analysis of the *Mortierella* strain

In the first chapter we study the effect of a chitin treatment on the bacterial and fungal community of lettuce plants in potting soil. The treated samples were supplemented with chitin in the beginning of the experiment, the control samples were not. Samples were taken from 2 different locations, i.e. the rhizosphere and the bulk potting soil, and at 2 different timepoints, i.e. 4 and 8 weeks after inoculation.
In the molecular analysis of the bacterial samples, 16S V3-V4 metabarcoding was used. The fungal samples were subjected to ITS2 metabarcoding. From this data we aim to unveil what influence chitin has on the lettuce microbiome, i.e. which species are overrepresented in the chitinous samples.

Next, we sequenced the full ITS region of the apparent fungal mycelium on soil and rhizosphere samples treated with chitin and untreated control samples. From a previous experiment at ILVO, we know that the fungus present in the chitin samples is likely to be of the genus *Mortierella*. We aligned the sequences obtained from fungal cultures to all the *Mortierella* ITS sequences in the UNITE database (v022019), and created a couple of phylogenetic trees. From these trees we can see the evolutionary relationships between our fungus of interest and other known *Mortierella* species.

Lastly, we sequenced the whole genome of the *Mortierella* strain and constructed a *de novo* genome assembly. We then interrogated the genome assembly from a Top-Down and Bottom-Up perspective to identify chitinases, N-acetylglucosaminidases, N-cycling genes, plant growth-promoting genes and plant immune response stimulatory genes.

Methods

1. Metabarcoding of the bacterial and fungal community

1.0 Experimental setup

In a previous experiment (Figure 12), lettuce plants were grown in potting soil, either with or without addition of chitin (2%). To analyze the metagenome of these plants, samples were taken after 4 and 8 weeks, from the bulk potting soil and the rhizosphere (Barneveldt, 2019). Metabarcoding libraries were prepared at ILVO and were sent to Admera Health (NJ, USA) for sequencing. Here we will analyze the metabarcoding data. The experiment has 3 variables: Treatment, Location & Timepoint. Throughout this thesis, these variables will be written capitalized. Every variable has 2 possible values: no treatment (control)/chitin (CT/CH), bulk potting soil/rhizosphere (BS/rhizo), 4/8 weeks (T4/T8). This results in 2^3=8 experimental conditions. For each condition we have three biological repetitions, so we have the data of 24 ITS2 metabarcoding samples (fungi). In the V3-V4 metabarcoding dataset (bacteria), 1 experimental repeat was missing from the T8 BS CT group. The dataset we received from Admera Health is organized into forward and reverse reads, both having different quality scores (usually the quality of the reverse reads is lower).

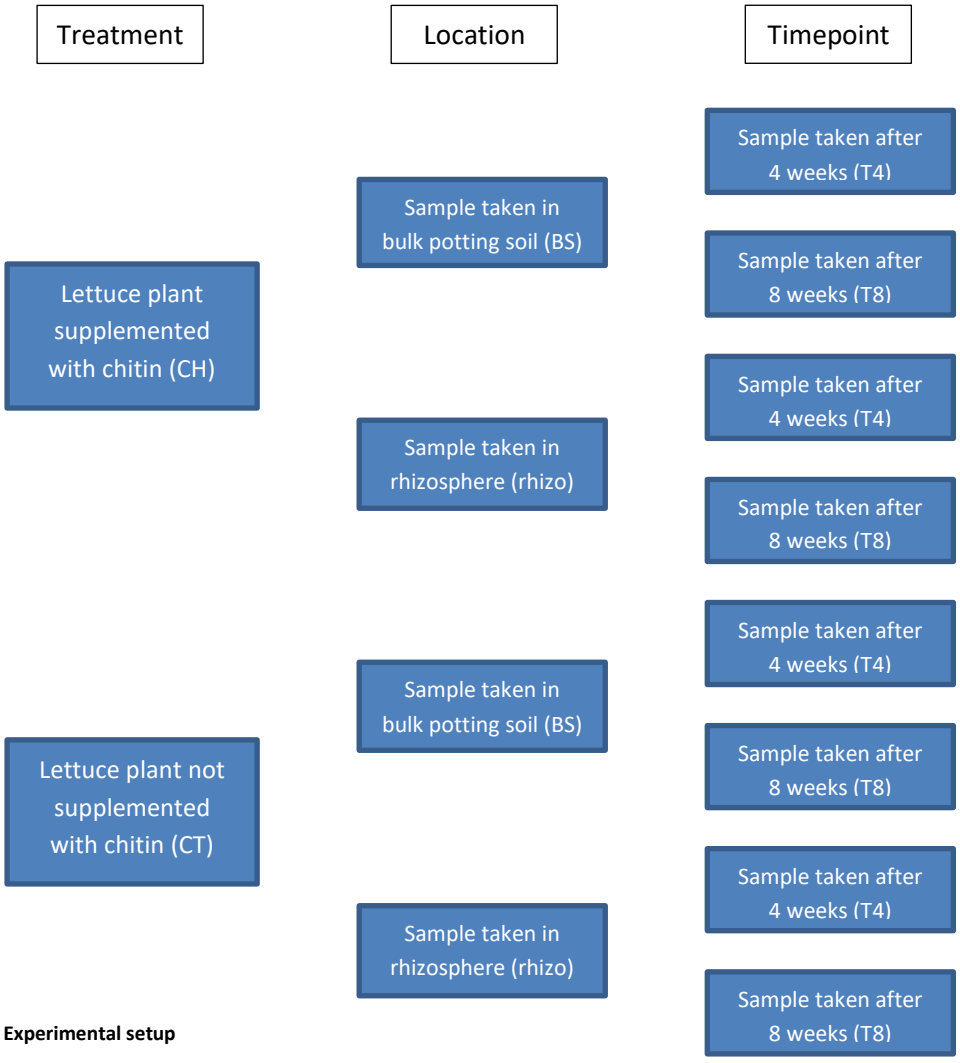


Figure 12: Experimental setup

1.1 Preprocessing

To generate ASV tables from the raw data, we will use the DADA2 software (Callahan et al., 2016). Before we can run the DADA2 pipeline, we first need to preprocess the data. Preprocessing includes the removal of primers and Quality Control (QC) of the sequence data. Removal of adapter sequences from the data was already performed by Illumina.

At every step where reads are discarded, we count the number of reads remaining in the dataset. This allows us to construct a graph showing the loss of data throughout the pipeline and assess whether we filtered too stringently in any step.

1.1.1 Primer removal

Bacteria

For the bacterial dataset we removed the primers in a Linux Command Line Interface (CLI) with Trimmomatic (Bolger, Lohse & Usadel, 2014). The forward primer was removed by TrimmomaticSE with specification HEADCROP:17. The reverse primer was removed by TrimmomaticSE with specification HEADCROP:21. We then evaluated the quality of the reads in R with the "PlotQualityProfile" method (DADA2). Everything in the metabarcoding chapter from here on out happens in an R environment.

Fungi

Due to the variable length of the ITS regions, the amplified region might be shorter than the read length in some species. This can result in the inclusion of the opposite primer in reverse complement form, also known as read-through. Thus in the fungal data we need to remove the primers in a more complex way, accounting for the possibility of read-through. We first created character vectors of all orientations of the primers, then we removed all reads containing ambiguous bases (already a filtering step in fact), because they can hinder the mapping of the primers and they are not allowed in the DADA2 algorithm. Then we mapped the primers to the reads to quantify the amount of read-through. The results of one sample provide us with enough information about the whole dataset, since all samples are assumed to contain the whole fungal community of the dataset, albeit in different relative abundances. Now we cut the primers in both orientations from the sequence data with Cutadapt (Martin, 2011). Afterwards we re-did the mapping step to see if the primers were indeed removed, and the quality of the reads was assessed with PlotQualityProfile.

1.1.2 Filtering & Trimming

Before we let DADA2 estimate an error model for the data, we filter out the low-quality reads.

Bacteria

We filtered and trimmed the reads using the "FilterAndTrim" method (DADA2). The forward and reverse reads were trimmed at 260 bp. All reads with ambiguous bases, all forward reads with more than 3 expected errors, and all reverse reads with more than 5 expected errors were filtered out. The filtering was less stringent on reverse reads than on forward reads because they are generally of lower quality, and the expected errors are calculated based on the quality scores. We also discarded

the reads that matched against the phiX genome, a viral control genome, commonly used as internal control during sequencing.

Fungi

Because of the variable length of the ITS regions, we will lose a lot of biological variation if we simply trim the reads to a fixed length (since all real reads with a length lower than the fixed length will be removed from the dataset). This requires some other parameters to be used for the fungal sequence filtering. Based on the quality control plot, we trimmed the 20 rightmost bases of the forward and reverse reads. The filtering parameters remain the same, and an additional condition is included to filter out all sequences smaller than 50 bp, as those will likely be sequencing artifacts and not real ITS regions.

1.1.3 Error Model, Sample Inference & ASV table

These steps are the core functionality of the DADA2 pipeline. They are the same for the bacterial and fungal data. First, the "learnErrors" method was used to create an error model from the data, then the estimated error rates were plotted, they can be found in the supplementary materials. With the "dada" method, the ASV's were inferred from the filtered data and the error rates. We then merged the forward and reverse read pairs, created the ASV tables, and removed chimeric sequences.

Chimeras are false reads that are formed from sequences that originate from different species. This can occur in a number of ways, for example when two sequences of different origin stick together during PCR and are amplified that way, or when forward and reverse reads from different origins were merged. The "removeBimeraDenovo" method from the DADA2 package can identify and remove chimeras if they can be reconstructed through an exact combination of a left- and right-segment of sufficiently more abundant reads.

The number of reads lost throughout the filtering pipeline was plotted for all 8 experimental conditions. The length of the reads in the ASV table was plotted.

1.1.4 Taxonomy

We assigned taxonomy to the ASV's with the assignTaxonomy method, a naïve Bayesian classifier built into DADA2. The algorithm requires a reference dataset as input for a training set and classifies the input ASV's in this framework. The bacterial taxonomy was classified using the Silva_v132 dataset as reference (Quast et al., 2013), and for the fungi we used the UNITE_v02022019 dataset (Nilsson et al., 2018).

1.2 Data Exploration

We use the phyloseq package (McMurdie & Holmes, 2014) to explore the data. Everything is performed on the fungal dataset and on the bacterial dataset.

1.2.1 Constructing the Phyloseq object

We read in the ASV table and the taxonomy table, both were generated in the DADA2 pipeline. The ASV table shows how many times the ASV (row) has been counted in the sample (column). The taxonomy table links each ASV (row) to a certain taxonomy (column).

First the ASV table was filtered to exclude ASV's which were not present at 2 cpm (counts per million) in at least 3 samples. Then the taxonomy table was subset to exclude the same low-count ASV's. The data was transformed to a matrix, and split into an ASV matrix (which contains the count data), a taxonomy matrix (which contains the taxonomy data) and a DNA sequence list (which contains the ASV's as DNA sequences), for compatibility with the "phyloseq" method, which creates a phyloseq-class object. The raw counts in the ASV matrix were recalculated to relative frequencies per sample. A sample dataframe was created to include in the phyloseq object. This dataframe contains the experimental conditions, which were encoded in the original filenames. The phyloseq object was then created from the ASV frequency matrix, the sample dataframe and the taxonomy matrix. This object does not yet contain the sequence information of the ASV's. The DNA sequence list was transformed to a DNAStringSet object (Biostrings package), and added to the phyloseq object with the "merge_phyloseq" method.

1.2.2 Diversity

We calculated the within-group (alpha) diversity and the between-group (beta) diversity. A group is defined as all replicates from the same experimental conditions.

The alpha diversity, which is a measure for richness within a group, was calculated as the Shannon index:

$$S = -\sum_{i=1}^{R} p_i * \ln(p_i)$$

Where S is the Shannon index, p_i is the relative proportion of sequences represented by i^{th} ASV, and R is the total number of ASV's. In biological context, p_i is the relative proportion of individuals of the i^{th} species/genus/family (or whatever the most specific taxonomy given to this ASV was), and R is the total number of species (Shannon, 1948).

The alpha diversity was then visualized by a richness plot showing the mean and standard error of the Shannon index, grouped per biological repeat.

The beta diversity, which is a measure for ecological distances between groups, was calculated as the Bray-Curtis dissimilarity:

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$$

Where BC_{ij} is the Bray-Curtis dissimilarity, C_{ij} is the sum of the lesser count for every ASV that the groups have in common, and S_i, S_j are the total number of ASV's in group i resp. j (Bray & Curtis, 1957).

A Bray-Curtis dissimilarity matrix of all pair-wise group combinations was calculated. We used multidimensional scaling (MDS) to display the high-dimensional data in a limited number of artificial dimensions. The ordination method used was Principal Coordinate Analysis (PCoA), and the input was

the Bray-Curtis dissimilarity matrix. This is all done in the "ordinate" method, so no explicit dissimilarity matrix object is obtained. PCoA plots were constructed for multiple combinations of dimensions on the axes.

A 4-D plot was constructed, in which the 3 most important dimensions of the PCoA are plotted on the x, y and z axes, and where the Shannon index is color-coded. This enables us to visualize both alpha- and beta diversity on 1 plot.

To get a clearer view on the taxonomy of the soil community, the data was agglomerated on the phylum and on the family level. Bar charts of the relative abundance of all phyla and the 10 and 30 most abundant families were plotted for every group.

1.3 Statistical Analysis

We use the vegan (Oksanen et al., 2019) and edgeR (Robinson et al., 2010) packages for the statistical analysis. Everything is performed on the fungal dataset and on the bacterial dataset. First, we will analyze the effect of the 3 experimental variables "Treatment", "Location" and "Timepoint" on the microbial community. Second, we will find out which ASV's are differentially abundant between treatment and control samples, and create a table which lists the abundance of all families in the bacterial & fungal communities in the sample groups.

1.3.1 Vegan

We want to determine whether there are factors that have a significant influence on the microbial community. The measure we use to compare microbial communities will be the beta-diversity (a Bray-Curtis dissimilarity matrix).

A Bray-Curtis dissimilarity matrix of all pair-wise group combinations was calculated. Remember, a group is defined as all replicates from the same experimental conditions. To identify factors with a significant effect on the microbial diversity, we will perform a Permutational Multivariate Analysis of Variance (PERMANOVA) test on the data. This test is an alternative to the Analysis Of Variance (ANOVA) test, where the assumption of normally distributed data is not made. Indeed, count data is characterized by a Poisson or Negative Binomial distribution, so the PERMANOVA is the right test.

However, the PERMANOVA test does require that the variance between all groups is equal, so this should first be checked. This was done with the combination of the "betadisper" and "anova" methods, which results in a multivariate analogue of Levene's test for homoscedasticity. Betadisper takes the Bray-Curtis matrix and group structure as input, calculates a median group centroid and outputs a betadisper-class object. This object contains the average distance of every group to the median and the eigenvalues for the PCoA axes. Then an ANOVA was performed on the betadisper object. If the variances of the groups show to be equal, only then can a PERMANOVA test be performed on the Bray-Curtis matrix with the "adonis2" method.

1.3.2 EdgeR

EdgeR is used to analyze differential abundance of ASV's. EdgeR was intended for use on gene transcription count data, for differential expression analysis, but the software works equally well on other types of count data that can be modelled with a Poisson or Negative Binomial distribution, such

as metabarcoding data. The starting object for this analysis is the filtered ASV count table (not the relative frequency!). This time, a phyloseq object was created from the count data, sample data and taxonomy table, and again the phyloseq object was merged with the DNA sequence data. From this phyloseq object a DGEList object was created. A design matrix and a contrast matrix were constructed. The contrasts will conduct 4 tests: the difference between the control group and the treatment group in every combination of the other two variables. A Generalized Linear Model (GLM) was fitted on the DGEList object and the design matrix, and the contrasts were used to conduct Likelihood Ratio Tests (LRT's) on the GLM. The output table shows the number of differentially abundant ASV's between the control and the treatment in the 4 conditions. Then the taxonomy of the differentially abundant ASV's in each condition was retrieved and the intersection was found.

This edgeR analysis was repeated twice: once with the phyloseq object (with count data) agglomerated on the phylum level, and once on the family level. The original edgeR analysis can be considered agglomerated on the ASV level. The output table on the phylum/family level then shows the number of differentially abundant phyla/families between the control and treatment in the 4 conditions.

We created output tables which list the mean relative abundance and standard error of all families (rows) in all grouped experimental conditions (columns) for both bacteria and fungi. These tables can be found in the supplementary materials.

Finally, we extracted all ASV's of the genus *Mortierella* from the fungal dataset and created a heatmap. The rows are the ASV's and the columns are the 24 samples. The color at the (i,j) position of the map represents the relative abundance of the i^{th} ASV in the j^{th} sample.

2. Identification of *Mortierella* strain using ITS phylogeny

2.1 Preprocessing

To take a deeper look into the phylogeny of the *Mortierella* that grew on the samples, we sequenced the full ITS region (Illumina MiSeq) and created 3 phylogenetic trees from the data.

We used the Sequence Scanner v1.0 software to perform QC on the .ab1 forward and reverse datafiles and eliminated the unreliable sequences, i.e. those that did not have a Continuous Read Length (CRL) of at least 300 and an average Phred score of at least 20. The trustworthy forward and reverse datafiles were merged and edited into consensus sequences in BioNumerics.

The dataset (Table 2) now contains 6 sequences, 3 of which were preprocessed by me ("Chitin7rh", "Control19rh" and "Chitin20rh"), and 3 more sequences that were preprocessed in a previous experiment at the ILVO ("Chitin1", "Chitin11" and "Chitin20rh"; Barneveldt, 2019). Note that the Chitin20rh sample is included twice in the dataset, once in the sequences we preprocessed, and once in the already preprocessed sequences. This is done as an internal control to check if our preprocessing was performed correctly. The "Control" samples were taken from a lettuce plant in potting soil without chitin, the "Chitin" samples were taken from a lettuce plant in potting soil with 2% chitin, the "rh" samples were taken from the rhizosphere and the samples without "rh" were taken from the bulk potting soil.

Strain	Full sample name in tree	Experimental conditions	Preprocessed when
Chitin1	"*Mortierella hyalina* Chitine 11 uit 181220"	Chitin treatment, potting soil sampling	Previous experiment
Chitin11	"*Mortierella hyalina* Chitine 1 uit 181220"	Chitin treatment, potting soil sampling	Previous experiment
Chitin20rh	"*Mortierella hyalina* Chitine 20 rh eigen controle uit 181220"	Chitin treatment, rhizosphere sampling	Previous experiment
Chitin20rh	"Chitin20rh"	Chitin treatment, rhizosphere sampling	This thesis
Chitin7rh	"Chitin7rh"	Chitin treatment, rhizosphere sampling	This thesis
Control19rh	"Controle19rh"	No chitin treatment, rhizosphere sampling	This thesis

Table 2: ITS phylogeny dataset

2.2 Alignment & Tree Generation

We then aligned these sequences to all ITS sequences of the genus *Mortierella* in the February 2019 UNITE (UNITEv022019) database with MUSCLE (Edgar, 2004). When constructing a neighbor-joining (NJ) phylogenetic tree, a distance matrix is computed from this multiple sequence alignment, and this distance matrix is used as input for the tree generating algorithm. The output of a tree generating algorithm is a Newick (.new) file. We created 3 phylogenetic trees.

First, we made a tree with the online tool of EBI[1] which uses a NJ algorithm, using default settings. The job is queued and one day later the .new file was ready for download.

The other trees were created in a Linux CLI with the fastphylo and fasttree packages. Fastphylo (Khan et al., 2013) uses the Fast Computation of Distance Estimators algorithm to compute a distance matrix with a very fast running time. It feeds this distance matrix to the Fast Neighbor Joining algorithm to create a phylogenetic tree. We used 100 bootstraps in this workflow.

Fasttree (Price et al., 2009) uses a maximum likelihood (ML) algorithm to create phylogenetic trees. In this method, evolutionary information is captured in a substitution model that scores the probability of certain mutations in the sequences. We used the Jukes-Cantor model, which assumes equal mutation rates.

The .new files were visualized with the ITOL online tool[2].

[1] https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/
[2] https://itol.embl.de/

3. Genome Analysis

3.1 *De novo* assembly

Whole genome shotgun sequencing was performed on a *Mortierella hyalina* sample by Illumina HiSeq 2x150 bp. We received the data in the FASTQ format, as forward and reverse read files.

After assessing the quality of the data with FastQC (Andrews, 2010), we employed the Shovill pipeline for genome assembly (Seeman, 2018). Shovill is an assembler that uses SPAdes, which is the gold standard for *de novo* genome assembly, at its core. SPAdes (Bankevich et al., 2012) is a De Bruijn graph-based assembler, which splits the sequences into overlapping kmers to connect contigs. The structure of the assembly depends greatly on the chosen kmer size. A short kmer size results in a graph which is overconnected, and does not reflect the way a chromosome is structured (Figure 13). A large kmer size results in lower overlap between kmers, and subsequently in an incredibly disconnected graph, again not reflective of the true chromosomal structure (Figure 14). SPAdes uses a range of kmer sizes, with which multiple graphs are built on top of one another, leading to a more highly connected graph when reaching the highest kmer size.

De novo genome assembly can take a long time and can get very complicated, so in order to decrease running time and increase quality, some other algorithms are included into the Shovill workflow, both before and after the SPAdes core. Shovill was initially intended for bacterial genome assembly, but can operate on small and haploid fungal genomes as well. Shovill includes adapter trimming in its workflow so we do not have to do this with Cutadapt or Trimmomatic.

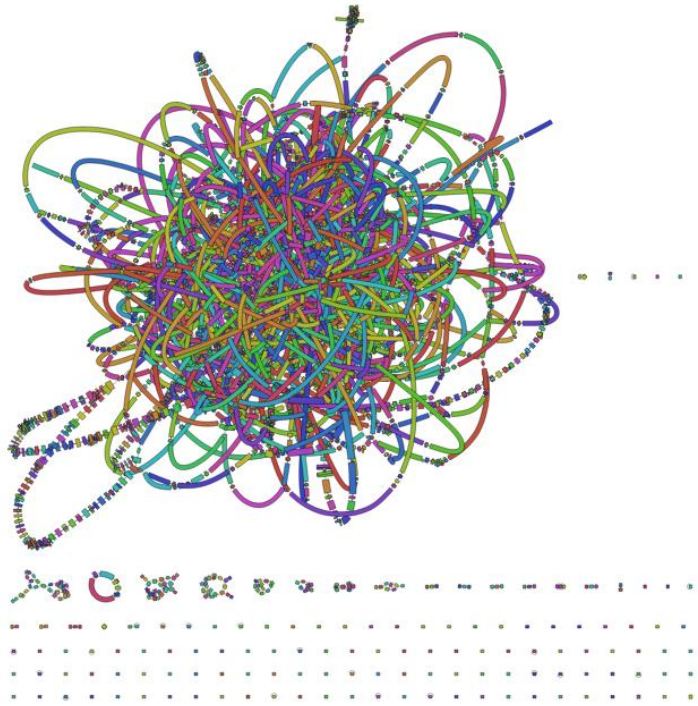


Figure 13: Overconnected De Bruijn Graph with small kmer size of 51. (https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)

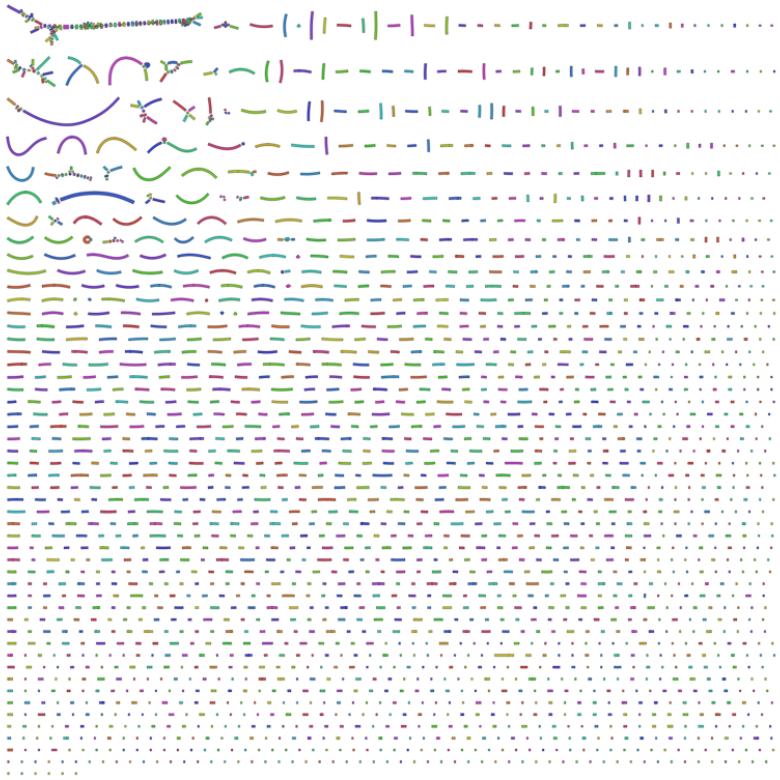


Figure 14: Disconnected De Bruijn Graph with large kmer size of 91. (https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)

The Shovill pipeline goes as follows:

First the genome size and read length are estimated from the reads by the “Mash” algorithm (Ondov et al., 2016). Then the FASTQ input files are reduced to a sensible depth, we used the default setting of x100. The downsampling of reads is important because if a certain sequencing error occurs too frequently due to high coverage (which is inevitable in high coverage data), it might be interpreted as a separate read, resulting in the formation of false contigs that end prematurely and do not connect to other contigs. The coverage of our data was approximately 183x, and thus needed to be downsampled. Now the adapters are trimmed from the reads, and kmer lengths 31, 51, 71, 91 and 111 were chosen from the read length distribution.

After the kmer range is chosen, the “Lighter” algorithm (Song et al., 2014) conservatively corrects sequencing errors in the reads. Then the merging of the PE reads is done by the “FLASH” algorithm (Magoc & Salzberg, 2011) before feeding the reads to SPAdes. Here, already 75.12% of reads are combined. Now the SPAdes assembly happens and contigs are created. When the assembly is complete, mapping errors are corrected by the combination of the “BWA-samtools” aligner (Li et al., 2009) and “Pilon” (Walker et al., 2014). The BWA-samtools aligner is used to map short reads to a reference genome. Here, the reads that fall outside the contigs are seen as the short reads, and the contigs are used as reference genome. Its output is a Binary Alignment Map (BAM) file of reads aligned to the genome FASTA file. Pilon then corrects Single Nucleotide Polymorphism (SNP) errors, insertions & deletions (indels) and larger gaps based on the alignment of the BAM file to the FASTA file. Finally, the contigs that are shorter than a given minimum length are removed, and a FASTA file of the assembled genome is returned (contigs.fa).

We set the minimum length of contigs to 500bp, which is a reasonable length that can contain (part of) a gene. The evaluation of the assembly was done with "QUAST" (Gurevich et al., 2013), which provides a simple and interactive user interface and several interesting plots.

3.2 Genome Annotation

We will tackle the genome annotation from two different perspectives: a Top-Down and a Bottom-Up approach. In the Top-Down approach we try to predict all genes in the *Mortierella* genome, retrieve their amino acid sequence and BLAST them against Uniprot to assign a putative function to the proteins. Then we will search for chitinases in the output. In the Bottom-Up approach we will treat our genome as a database. We BLAST a small number of curated fungal chitinases against our genome to find out if similar genes are present in *M. hyalina*. The major difference between both approaches is that the Top-Down approach will find a great number of genes, but some rare species-specific chitinase genes might be missed because the gene prediction algorithm does not recognize them. In the Bottom-Up approach, we can identify possibly novel chitinase genes by sequence comparison with curated chitinases directly to the genome. In this approach we are not dependent on the gene prediction algorithm since we directly search the genome.

3.2.1 Top-Down

We used AUGUSTUS (Stanke et al., 2004) to predict genes from the assembled genome. AUGUSTUS has been trained on a number of training sets to specifically predict genes of certain species. To run the software, we first need to determine which reference data set from the AUGUSTUS database is most similar to *Mortierella*. To do this, we checked the NCBI Taxonomy database and queried all fungal training sets in the database, and identified the organism that is taxonomically closest to *Mortierella*. This turned out to be *Rhizopus oryzae.* Note that even though out of these options *Rhizopus oryzae* most closely resembles *Mortierella*, it is only a relative up to the phylum level of *Mucuromycota*. This however is not an issue, as the AUGUSTUS website itself says that the software can deal with these kinds of distances, e.g. the human training model works well enough to predict genes of all mammals.

The output of AUGUSTUS is a Generic Feature Format (.gff) file with all predicted genes. Now the amino acid sequences are retrieved as a .aa FASTA file with the getAnnoFast.pl script of the AUGUSTUS pipeline. Note that not every predicted amino acid sequence in this file is a full length protein: many of these sequences do not start with M (for Methionine, the first amino acid in every protein). Because of the considerable size of the .aa file, Diamond BLASTp (a faster version of BLASTp; Buchfink et al., 2015) was used to query our predicted proteins against UniProt (Swiss-Prot + trEMBL; The UniProt Consortium, 2019). We used a local download of the November 2019 version of UniProt. From the query results we took the top 10 hits for every protein. This was done because we noticed that the top hit often was a *Mortierella elongata* protein, due to its resemblance to *M. hyalina*. This species however is not very well annotated either, which means that the proteins that might be chitinases are likely not annotated as chinitases. By looking at the 10 top hits, we are more likely to find well annotated chitinases from other species in the results. Note that UniProt contains both Swiss-Prot, a manually curated database, and trEMBL, a non-curated database (Bairoch & Apweiler, 2000). Querying a non-curated database implies that much redundant information is retrieved. This

means that many of the 10 top hits are likely a different entry of the same protein. This redundancy is a positive thing however, because now our chances of having a well-annotated entry among the top 10 hits are greatly increased.

3.2.2 Bottom-Up

We queried the Enzyme Commission number[3] for endochitinase (EC 3.2.1.14) on Swiss-Prot and selected only the fungal sequences. This resulted in a list of 42 curated fungal chitinases. With the "makeblastdb" command, we then constructed two databases from our *M. hyalina* data: a nucleotide database from the contigs.fa file, and a protein database from the .aa file, both created by the *de novo* assembly as described above. With tBLASTn & BLASTp we then queried the nucleotide & protein databases for the 42 curated fungal chitinases. BLASTp is a protein-protein alignment, while in tBLASTn the nucleotide sequences in the database are translated into amino acids in all 6 Open Reading Frames (ORF's) so the same protein-type query can be used. Note that in the tBLASTn output the same query protein will match multiple locations on the same contig. This is because not every exon exists in the same ORF, causing different hits for multiple exons of the same gene. This makes the output file somewhat more redundant, but presents no serious issue.

[3] https://enzyme.expasy.org/

Results

1. Metabarcoding of the bacterial and fungal community

1.1 Preprocessing

1.1.1 Filtering & Trimming

Before we trim the primers of the ITS2 sequences, we check how many reads contain the primers and in which orientation. This is summarized in Table 3. We see that the forward primer is present in the forward reads in its forward orientation, yet there is also a significant amount forward primer found in the reverse reads, in its reverse orientation. This indicates read-through, caused by the variable length of the ITS2 region. The same can be said for the reverse primer.

	Forward	Complement	Reverse	Reverse Complement
FWD.ForwardReads	99267	0	0	0
FWD.ReverseReads	0	0	0	13933
REV.ForwardReads	0	0	0	20735
REV.ReverseReads	100035	0	0	0

Table 3: Presence of primers in the ITS2 reads before primer removal

The internal primers that were caused by read-through have been effectively removed from the data, as is shown in Table 4.

	Forward	Complement	Reverse	Reverse Complement
FWD.ForwardReads	0	0	0	0
FWD.ReverseReads	0	0	0	0
REV.ForwardReads	0	0	0	0
REV.ReverseReads	0	0	0	0

Table 4: Presence of primers in the ITS2 reads after primer removal with cutadapt

The QC of the V3-V4 region (bacterial data) is shown in Figure 15. The quality of the forward reads stays high (>20) throughout the whole sequence, and the reverse read quality starts to drop below a PHRED of 30 around the 260 bp region. The V3-V4 region is about 420 bp in size, so if we trim both at 260 bp, with 2x260 = 520 bp in read length, we have a nice overlap of forward and reverse reads. We do not necessarily need to trim the forward reads at 260 bp, since the quality of the forward reads is still sufficient towards the end, yet due to the long read size in comparison to the amplified region, we can afford to trim the last few bases off the forward read to slightly improve sequence quality.

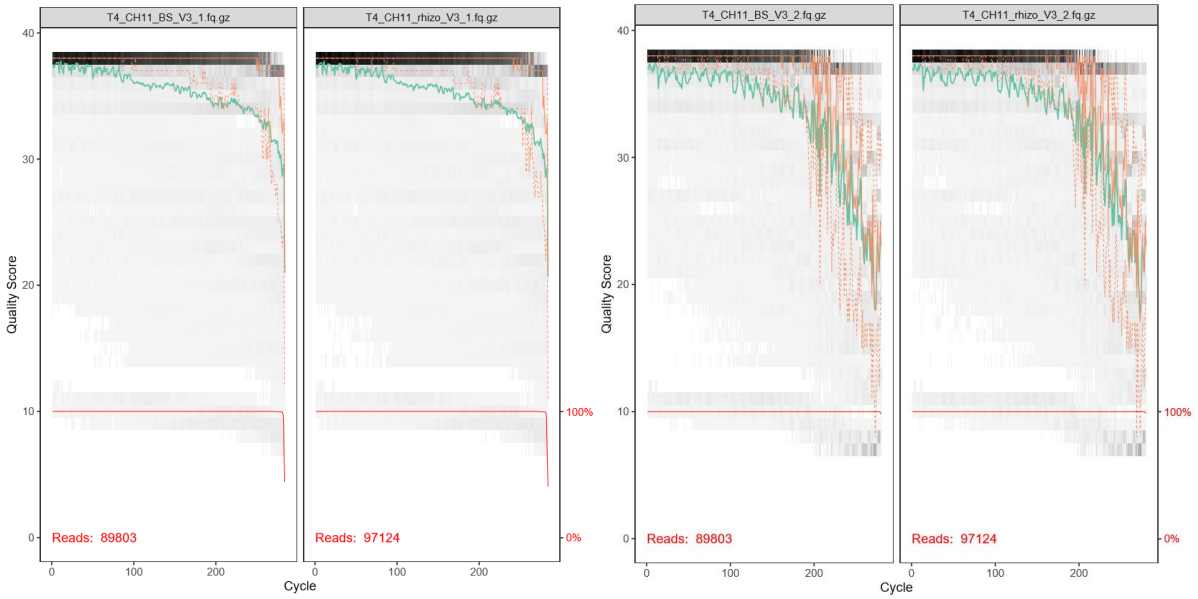


Figure 15: PlotQualityProfile of bacteria: PHRED score from 0 (low quality) to 40 (excellent quality)
Left: Forward reads of first two samples; Right: Reverse reads of first two samples

The QC of the ITS2 region (fungal data) is shown in Figure 16. The quality of both the FWD and REV reads starts to drop below a PHRED of 25 around the final 20 bases. For the fungal data, we certainly want to avoid trimming too much in this step because of the variable length in the ITS region, so we deem a PHRED of 25 to be good enough. In comparison with the V3-V4 data, the ITS2 data is of slightly lower quality, which was expected due to the variable length in the ITS region. Yet, a minimal PHRED score of 25 is sufficient for the forthcoming data analysis.

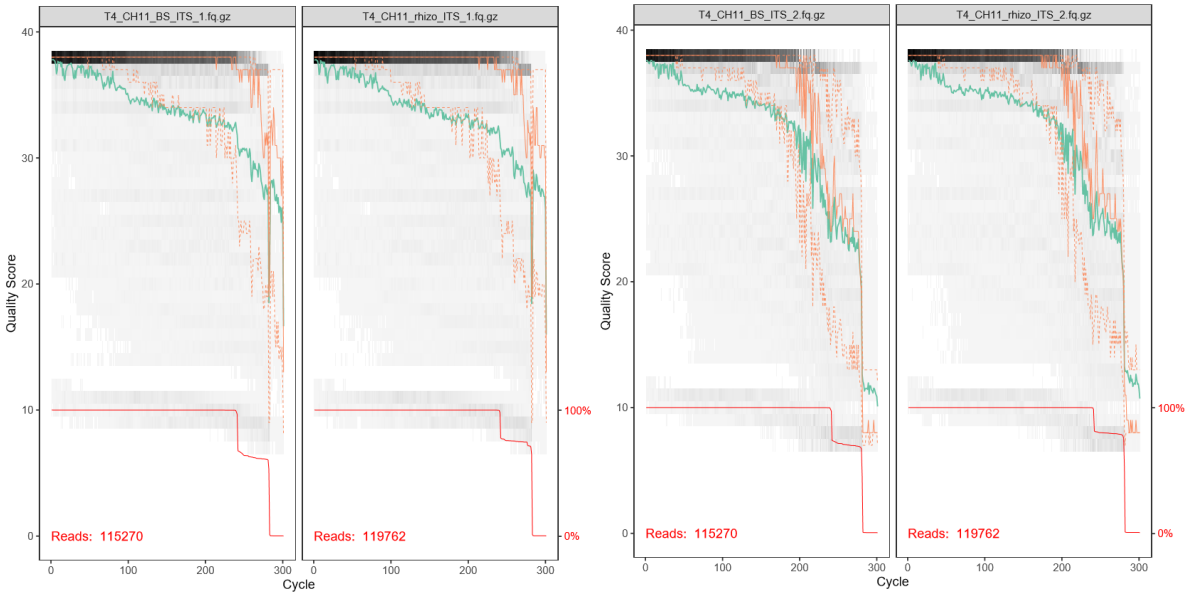


Figure 16: PlotQualityProfile of fungi: PHRED score from 0 (low quality) to 40 (excellent quality)
Left: Forward reads of first two samples; Right: Reverse reads of first two samples

1.1.2 Error Model & Sample Inference

Now we will perform a few exploratory tests. Since all samples are assumed to be similar in structure, we used the first bacterial sample as a representative for the full dataset in the first test.

Test for roughly how many ASV's get inferred from the input:

dadaFs[1]: 1761 sequence variants were inferred from 25547 input unique sequences.
dadaRs[1]: 1409 sequence variants were inferred from 36037 input unique sequences.

This shows us that however there are more unique sequences in the reverse reads, we find fewer sequence variants. This is likely due to the lower quality of the reverse reads, which introduces more errors in the error model. This decreases the total number of sequence variants because more bases in the sequences were considered erroneous, and were corrected by the error model.

In a test on the full bacterial data we found that 547 out of 6055 ASV's were marked as chimeras. Chimeras represent 9.03% of all sequence variants, yet only 1.27% of all reads in the table originate from chimeras. Thus, chimeras are often sequence variants that are supported by only few reads.

1.1.3 ASV Tables

The resulting ASV tables show the number of inferred ASV's (rows) in the samples (columns). The dimensions are 5508x23 (bacteria) and 2481x24 (fungi). We present the first 5 rows and columns of the bacterial table for context (Table 5). Note that the row names are abbreviations of the ASV's, since the ASV's real sequence is more than 400 characters long. Also note that the experimental conditions are encoded in the sample names: T4/8 means timepoint = 4/8 weeks, CH[##]/CT[##] means treatment = chitin/control [sample number], BS/rhizo means location = bulk soil/rhizosphere and V3 indicates that it is the V3-V4 region.

	T4_CH11_BS_V3	T4_CH11_rhizo_V3	T4_CH5_BS_V3	T4_CH5_rhizo_V3	T4_CH8_BS_V3
TCGAGA…	598	563	281	618	208
TGGGGA…	1252	702	942	205	705
TGGGGA…	1010	837	619	527	823
TAGGGA…	1000	886	516	324	616
TGGGGA…	2273	1421	2028	756	2080

Table 5: First 5 rows and columns of the bacterial ASV table. Rows are ASV's, columns are samples

1.1.4 Filtering results

Figure 17 displays the length of the ASV's that passed the filtering steps. The left plot clearly shows that the ITS region is extremely variable in length compared to the V3-V4 region (right plot), which only shows 3 peaks, at 400, 420 and 427 bp. This makes sense, since the amplified region, including primers and adapters which were removed is ~460 bp in length.

We kept track of the number of reads lost during the filtering pipeline (Figure 18). There are no apparent differences in loss of reads between experimental conditions. The biggest drop in reads can be seen after the first filtering step. This is indeed the most stringent step, where the low-quality reads are removed. About 85% of bacterial reads and 65% of fungal reads survive this first step. The fungal data remains at around 65% reads kept during the rest of the pipeline, indicating that the very

stringent first filtering step removed most of the bad reads. In the bacterial data, there is again a slight drop in read count after the merging of the reads. This means the bacterial data contained quite a few reads that did not overlap adequately to merge together. Overall around 80% of bacterial reads survive the filtering. The higher percentage of bacterial reads remaining in the dataset as compared to fungal reads is likely due to the variable length of the ITS region, which makes the data less uniform, and introduces sequencing errors, ultimately resulting in a need for more stringent filtering parameters.

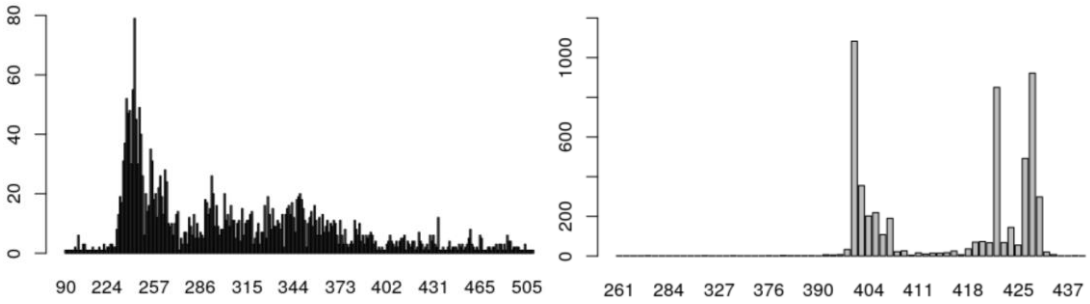


Figure 17: Length of the ASV's that were retained after all filtering
y-axis: number of ASV's, x-axis: length of ASV's.
Left plot: fungal ITS2 ASV's, right plot: bacterial V3-V4 ASV's.

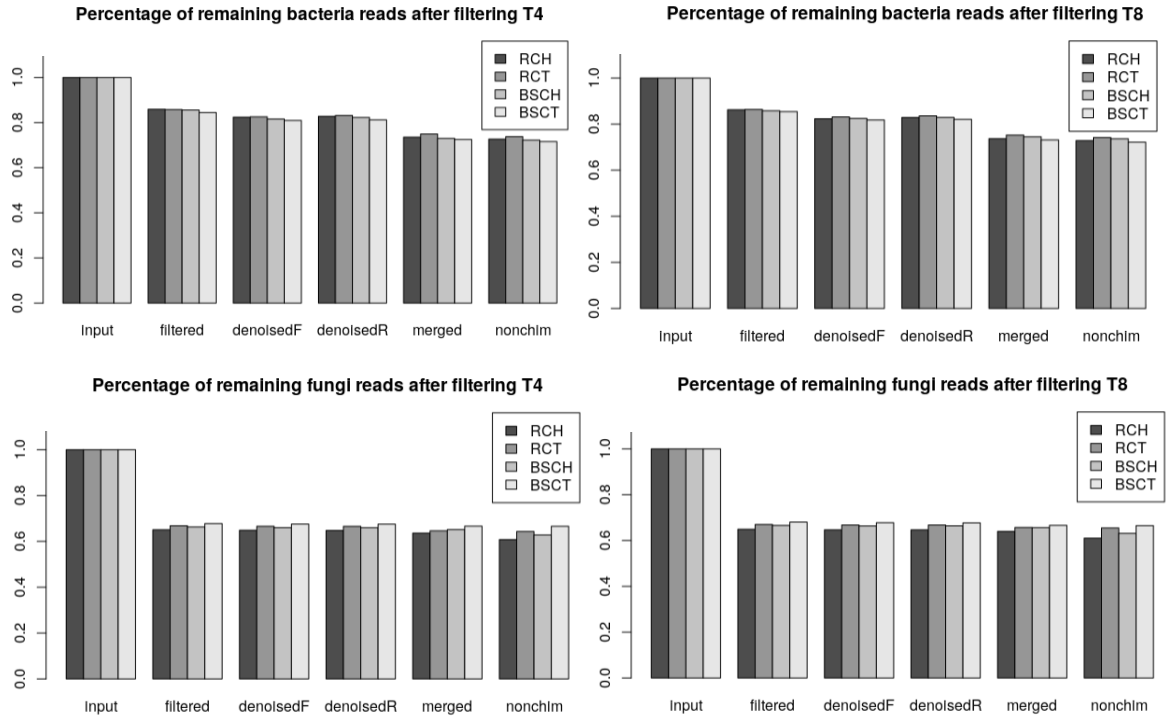


Figure 18: Percentage of reads that are kept after each filtering step
R: rhizosphere, BS: bulk potting soil, CH: chitin, CT: control, T4: 4 weeks, T8: 8 weeks.
Input: raw reads, filtered: filterAndTrim, denoised(F/R): error model correction (forward/reverse), merged: F/R merging, nonchim: chimera filtering.

1.2 Data Exploration

1.2.1 Bacteria

We plotted the alpha diversity as the Shannon index of the sample groups to get a first look on the effect of the experimental conditions on the bacterial biodiversity in the samples (Figure 19). The spread of data points appears to be more or less random. No clear effect of any parameter on alpha diversity emerges from the data in the bacterial dataset. Note the giant error bar in the "T8 BS CT" group. This is due to the group consisting of only 2 samples, instead of the usual 3, and those 2 samples' alpha diversity was very much not alike, causing a high standard error.

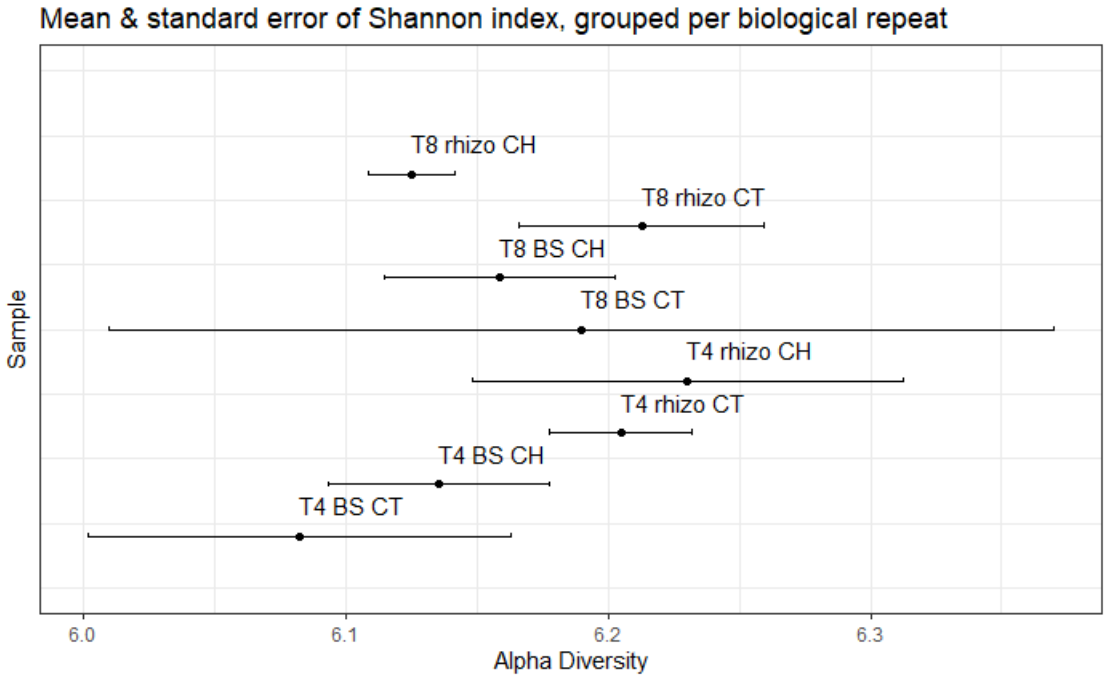


Figure 19: Alpha diversity of bacteria
The dot indicates the group mean Shannon index; the error bar represents the group standard error.

We then plotted the beta diversity as various combinations of PCoA dimensions from a Bray-Curtis dissimilarity matrix to see if certain groups are separated on the basis of their between-group diversity, i.e. dissimilarity (Figures 20-23). These figures have an exploratory purpose; they show us what to expect from the data, but bear little statistical significance. Later on, in Results section *1.3.1: PERMANOVA*, we will perform a statistical analysis on the Bray-Curtis dissimilarity matrices to try to prove a significant dissimilarity between groups.

There is a distinct separation between the control and chitin groups on the first PCoA dimension, and a reasonable separation between the rhizosphere and bulk potting soil groups in the chitin-treated samples on the second PCoA dimension. In the control samples, the separation based on location is less apparent (Figure 20).

Figure 21 shows the same data plotting on the same dimensions as Figure 20, but now the symbol represents Timepoint instead of Location. On these 2 dimensions, no clear separation based on Timepoint can be found.

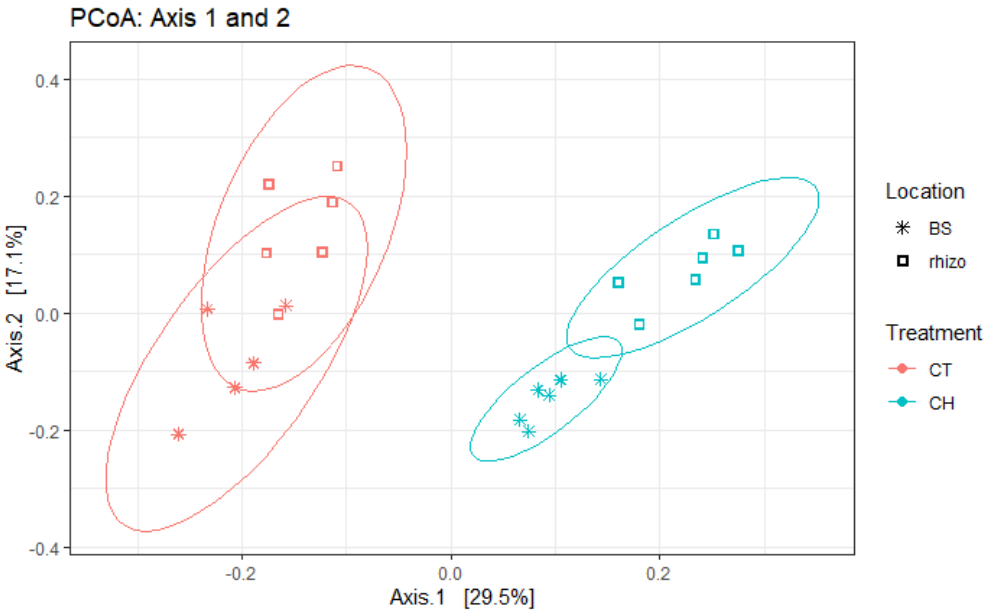


Figure 20: Beta diversity of bacteria on PCoA dimensions 1 & 2
Measurement: Bray-Curtis dissimilarity

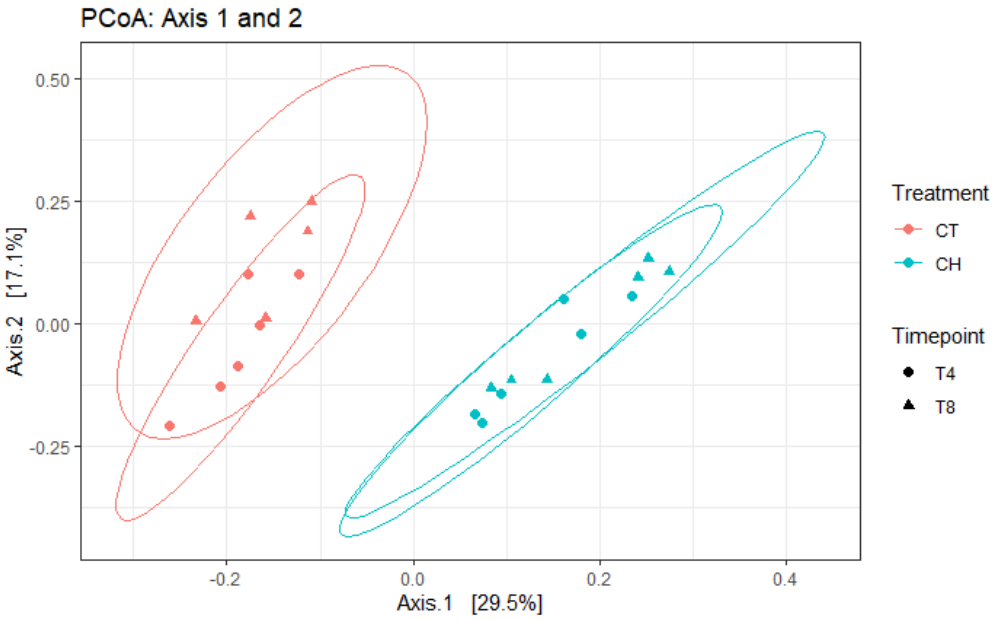


Figure 21: Beta diversity of bacteria on PCoA dimensions 1 & 2
Measurement: Bray-Curtis dissimilarity

When plotting the data on the first and fourth PCoA dimensions, we do notice a separation on Timepoint along the y-axis, both in control and chitin samples (Figure 22). Plotting on the second and fourth PCoA dimensions, still results in a reasonable separation based on Location and Timepoint (Figure 23).

A plot including the third PCoA dimension was made, but was not very informative. The third dimension did not seem to correspond with any experimental variable, and was not able to achieve any separation between groups (Supplementary Figure 11).

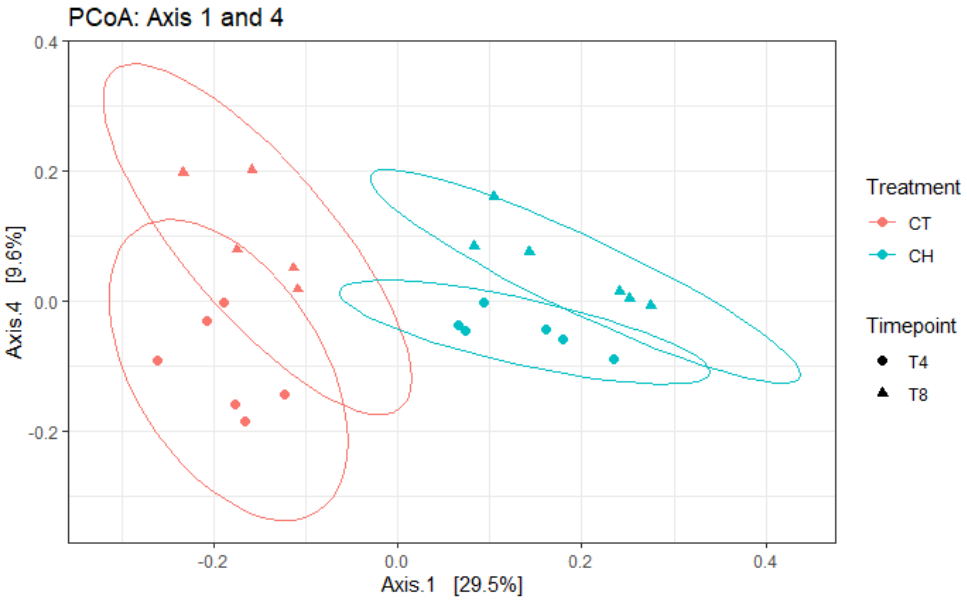


Figure 22: Beta diversity of bacteria on PCoA dimensions 1 & 4
Measurement: Bray-Curtis dissimilarity

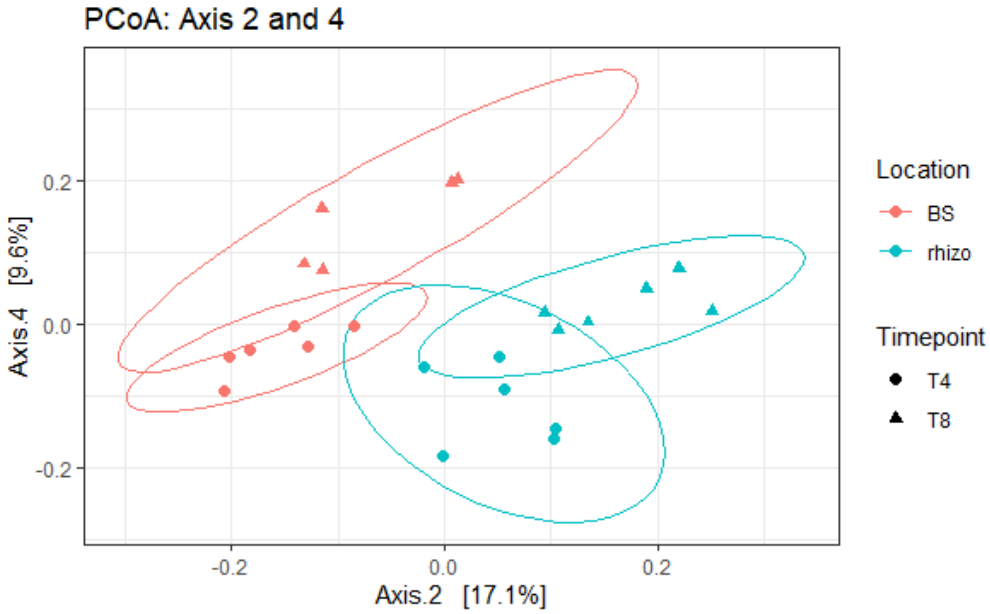


Figure 23: Beta diversity of bacteria on PCoA dimensions 2 & 4
Measurement: Bray-Curtis dissimilarity

We can draw several early conclusions from these beta diversity plots:

- The first PCoA dimension (29.5%) roughly corresponds to the Treatment effect.
- The second PCoA dimension (17.1%) roughly corresponds to the Location effect.
- The fourth PCoA dimension (9.6%) roughly corresponds to the Timepoint effect.
- Most of the variance in the data is not caught in a single PCoA dimension; a reasonable separation between experimental conditions can be achieved while excluding the most important first dimension.

Nothing stands out when plotting the bacterial alpha and beta diversity on a single 4D plot (Figure 24). All datapoints seem to be spread randomly. Keep this in mind when looking at the fungal 4D plot.

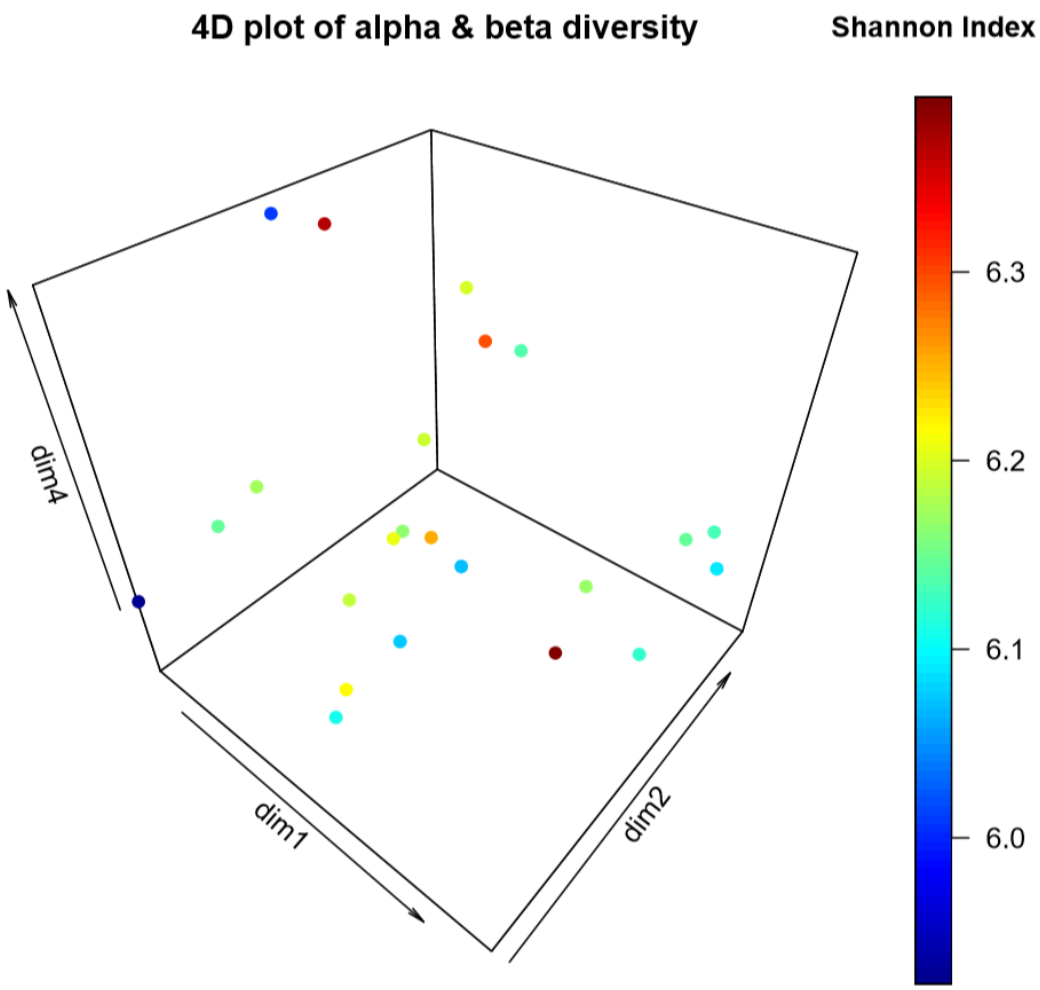


Figure 24: 4D plot of bacterial alpha and beta diversity
Beta diversity is shown on the spatial axes, alpha diversity is color coded.

Abundance plots of the top 10 and 30 phyla and families were made. There is an immense biodiversity in the bacterial soil community, which makes the bar charts harder to read. Only one of them is shown here (counts aggregated on phylum level), the others can be found in the

supplementary materials. There is a slight decrease in *Verrucomicrobia* in the presence of chitin, and a slight increase in *Bacteroidetes* (Figure 25), but these differences are negligible compared to the effect that chitin has on the fungal community (Results section 1.2.2: *Fungi*). The *Patescibacteria* are more abundant after 8 weeks, and the *Actinobacteria* appear to be more abundant in bulk potting soil than in rhizosphere.

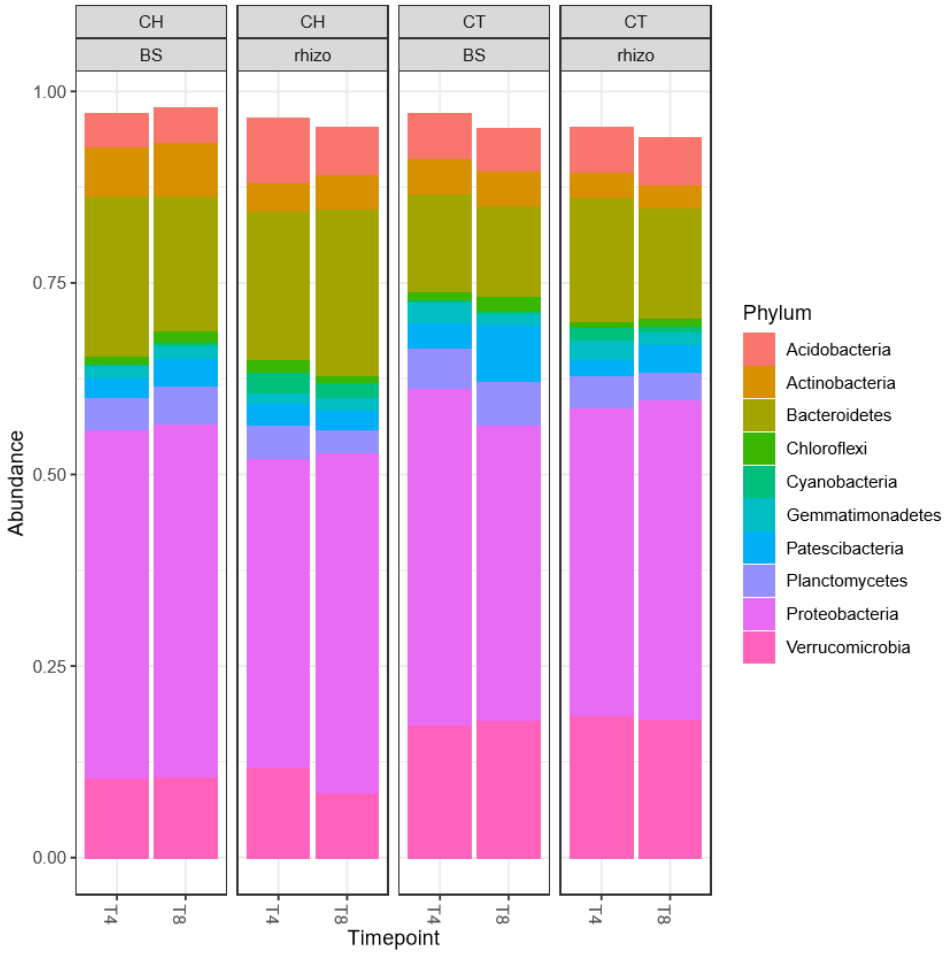


Figure 25: Top 10 phyla in the bacterial samples

1.2.2 Fungi

Again, we plotted the alpha diversity as the Shannon index of the sample groups to get a first look on the effect of the experimental conditions on the fungal biodiversity in the samples (Figure 26). There is a clear trend towards a lower biodiversity in the chitin-treated samples. This is likely due to the fact that much of the biodiversity is lost when certain species thrive in a chitin-rich environment. By comparing the alpha diversity of the fungi and the bacteria (Figure 19), we see that the chitin treatment has a much greater effect on the fungi than on the bacteria. From Figure 26 we can also see that in the fungal control samples the bulk potting soil has a higher biodiversity than the rhizosphere, likely because of certain secretions by the plant's root system that hinder the growth of certain species, causing them to disappear from the rhizosphere, and stimulate growth of certain species, causing them to thrive and outcompete others.

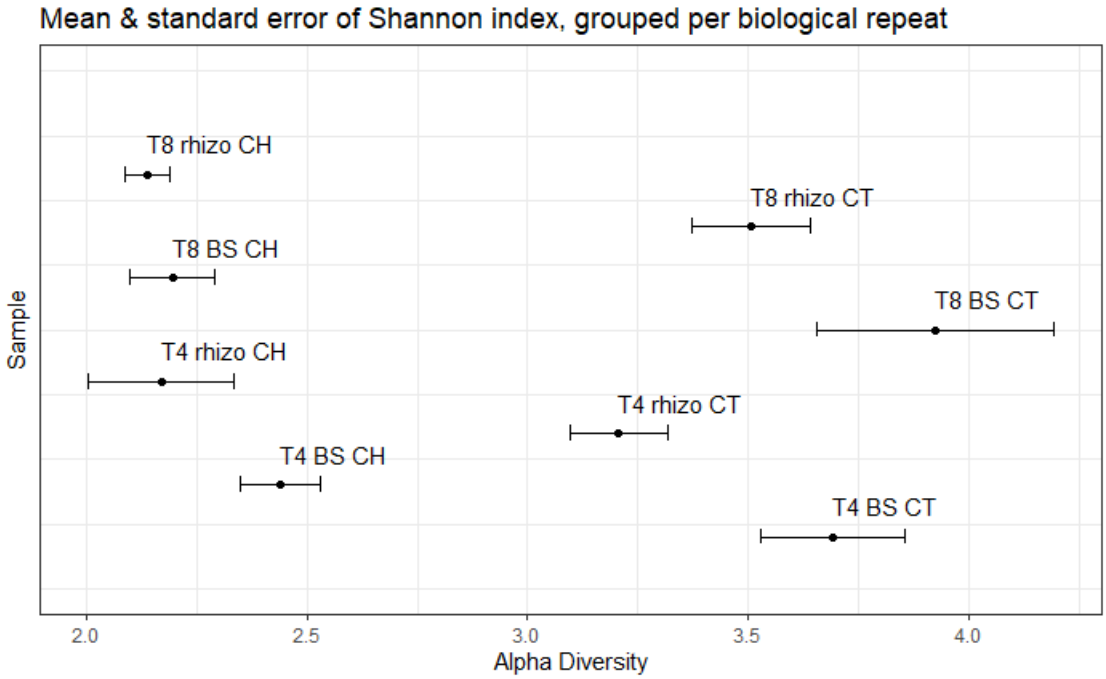


Figure 26: Alpha diversity of fungi
The dot indicates the group mean Shannon index; the error bar represents the group standard error.

This shift in fungal biodiversity in the chitin-treated samples is also very apparent from the beta diversity plots. Note once again that the observations made here will be subjected to statistical testing in Results section *1.3.1: PERMANOVA*. There is a distinct separation between the control and chitin groups on the first PCoA dimension (Figure 27). This dimension captures 58.2% of the variance in the data, and has high discriminatory power. We notice an apparent separation based on Location in the control group (red samples) on both the first and the second PCoA dimension (Figure 27). Indeed, earlier in the alpha diversity plot we saw that there is a higher biodiversity in the bulk potting soil of the control samples. In the chitin-treated group however, the samples are near indistinguishable based on Location, due to the fact that the chitin treatment had such a massive impact on the fungal community. No Timepoint effect is captured in the first 2 PCoA dimensions (Figure 28).

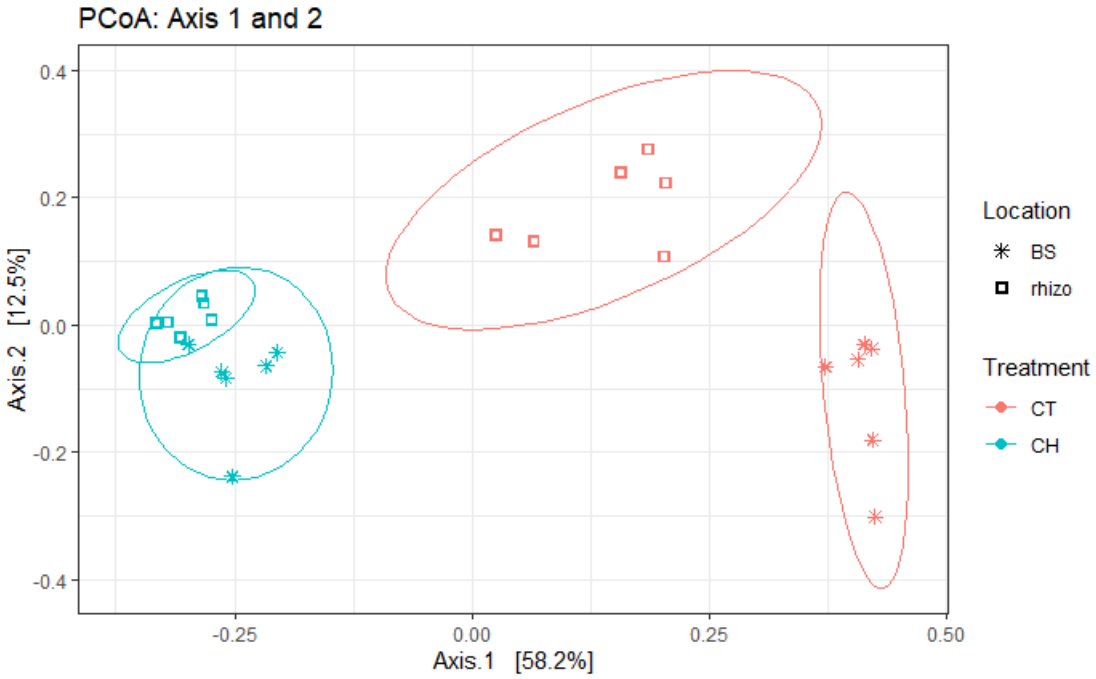


Figure 27: Beta diversity of fungi on PCoA dimensions 1 & 2
Measurement: Bray-Curtis dissimilarity

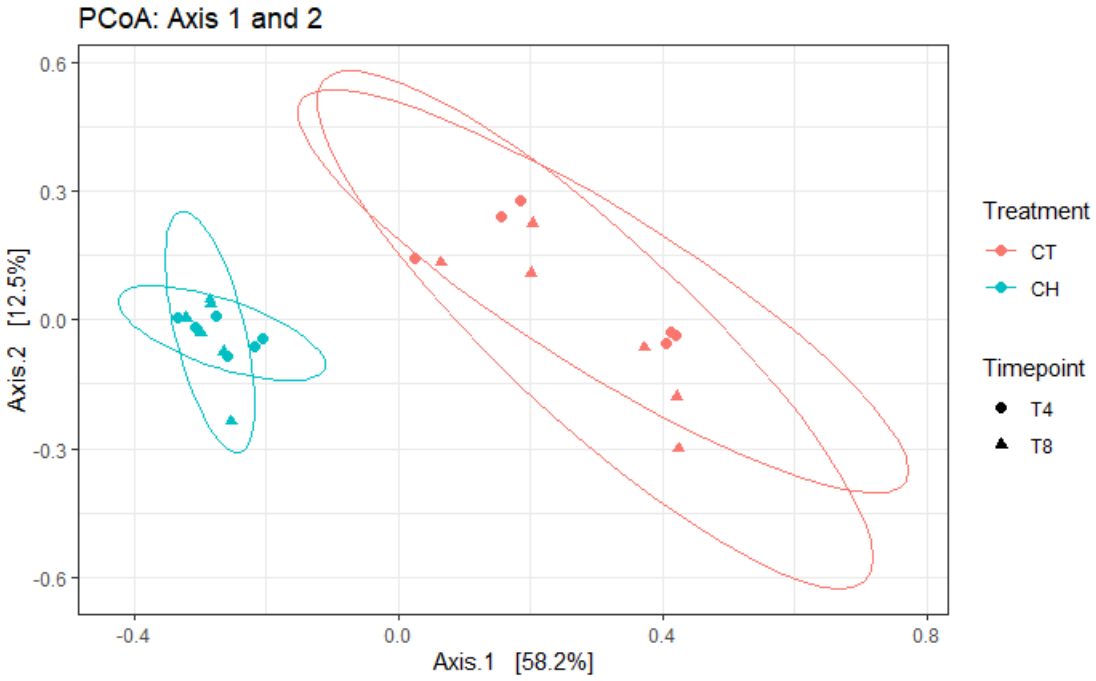


Figure 28: Beta diversity of fungi on PCoA dimensions 1 & 2
Measurement: Bray-Curtis dissimilarity

There is a slight separation in the data based on Location over the second dimension, and little to no separation based on Timepoint (Figure 29). Plotting on the first and third dimension results in a clear separation based on Treatment. Only in the control group are these dimensions able to slightly separate the samples based on Timepoint (Figure 30).

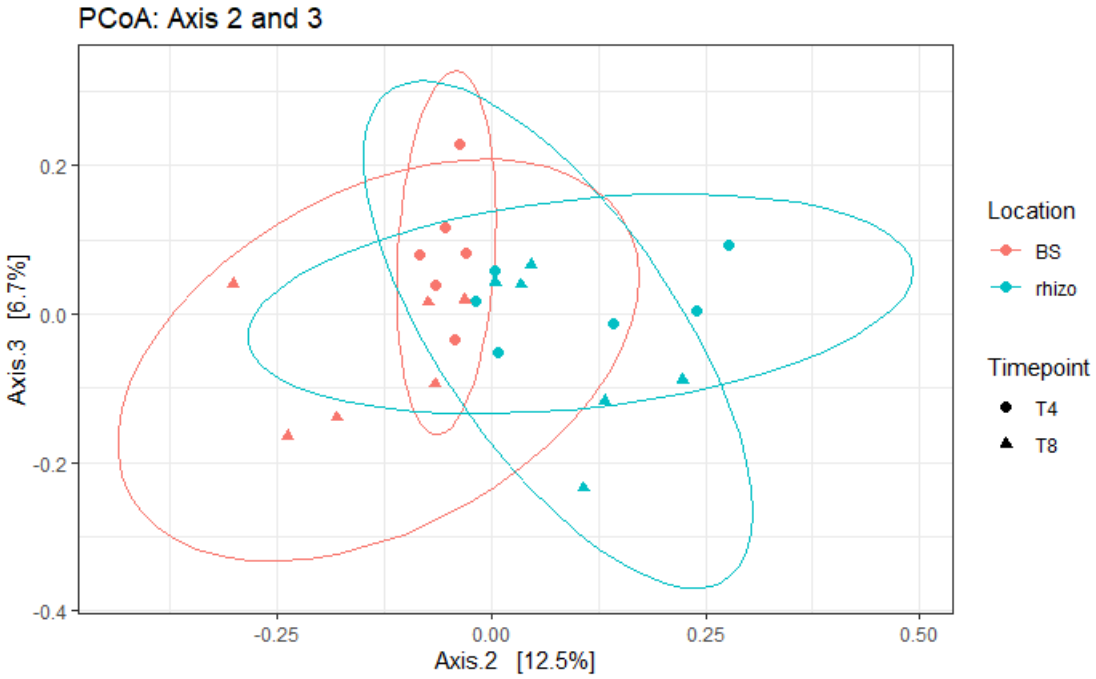


Figure 29: Beta divesity of fungi on PCoA dimensions 2 & 3
Measurement: Bray-Curtis dissimilarity

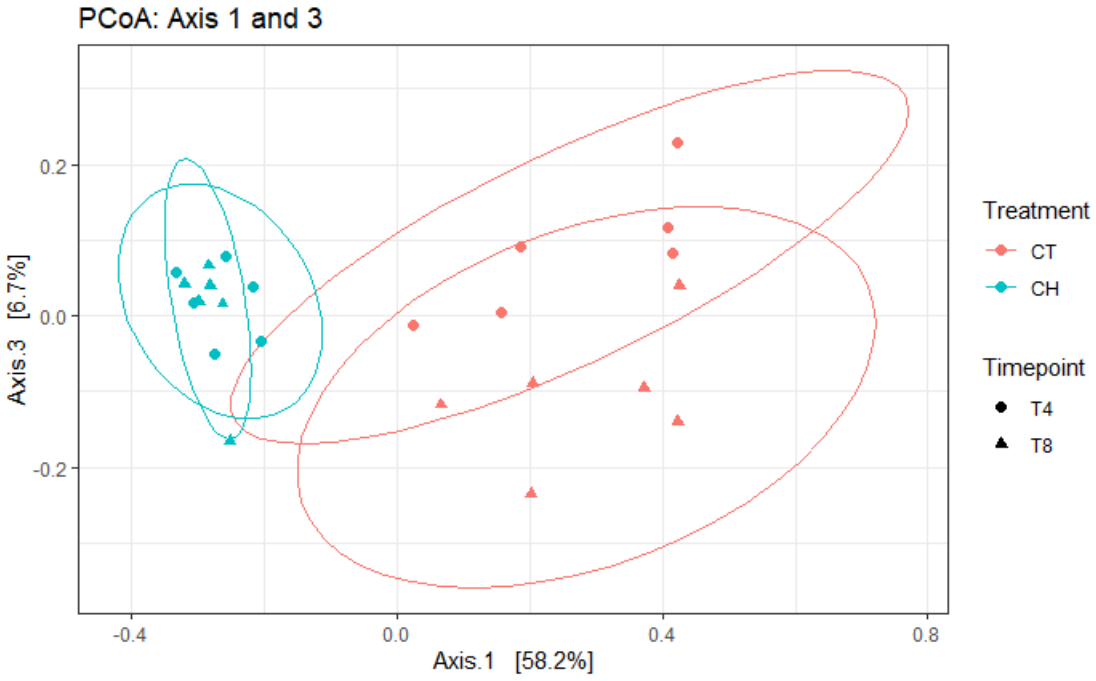


Figure 30: Beta divesity of fungi on PCoA dimensions 1 & 3
Measurement: Bray-Curtis dissimilarity

We draw several conclusions from these beta diversity plots:

- The first PCoA dimension (58.2%) clearly carries the Treatment effect. It also includes some of the variance based on Location in the control samples.
- The second PCoA dimension (12.5%) roughly corresponds to the Location effect.

- The third PCoA dimension (6.7%) roughly corresponds to the Timepoint effect.
- Compared to the first PCoA dimension of the bacteria (29.5%), the first PCoA dimension of the fungi has doubled in size (58.2%). This suggests that most of the variance in the fungal data can be explained by a single factor, being the chitin treatment, while the variance in the bacterial data stems from a more complicated combination of factors. The chitin effect on the fungal community is so large that for the most part it overshadows the other experimental conditions in the chitin-treated samples.

The fungal 4D plot is quite interesting (Figure 31). Notice that (apart from one outlier) all samples with low alpha diversity (blue hue) are grouped closely together on the spatial axes. As we learned from the alpha diversity graph, these are the samples treated with chitin. The control samples with higher alpha diversity are spatially distributed over a larger volume. This means that their more diverse fungal microbiome differs a lot from sample to sample, while the fungal microbiomes of the chitin-treated samples all differ only a slight bit in terms of beta diversity (apart from the one outlier). This is due to the incredible increase in presence of one particular fungus which outcompetes most other fungi in a chitin-rich environment: *Mortierella*. We see no such effect in the bacterial 4D plot (Figure 24), which means that in our data there was no single bacterium that was able to outcompete all other bacteria in a chitinous environment.

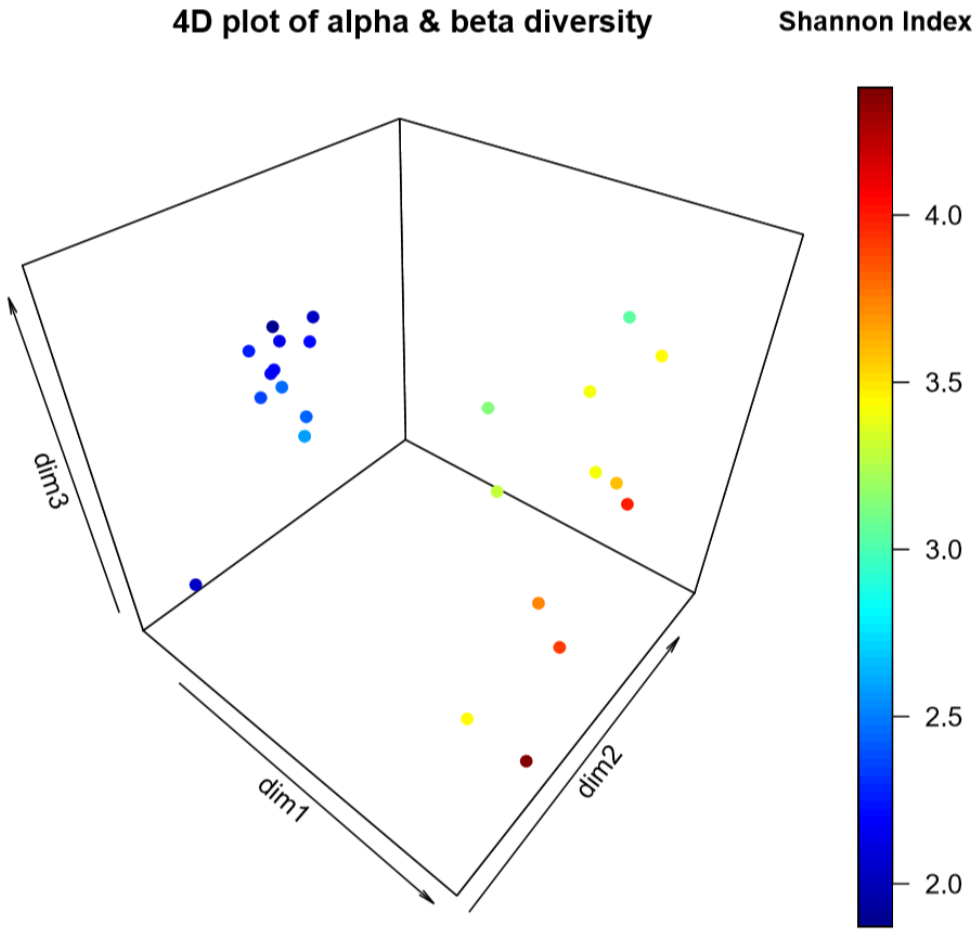


Figure 31: 4D plot of fungal alpha and beta diversity
Beta diversity is shown on the spatial axes, alpha diversity is color coded.

Abundance plots of the top 10 and 30 phyla and families were made (Figures 32 & 33). Only 2 of them are shown here, the other one can be found in the supplementary materials (note that the top 10 and top 30 fungal phyla are the same graph, since in the top 10 graph an abundance of 100% is reached with 7 phyla).

Figure 32 shows that 99% of fungal biodiversity can be captured in 3 phyla: *Ascomycota, Basidiomycota* and *Mortierellomycota.* The chitin-treated samples show a great increase in *Mortierellomycota* as compared to the control samples, where *Ascomycota* are more abundant. The abundance of *Mortierellomycota* is lower after 8 weeks than after 4 weeks. This may be explained by the fact that the chitin was added in the beginning of the experiment, and that after 8 weeks this resource may be depleted to some degree.

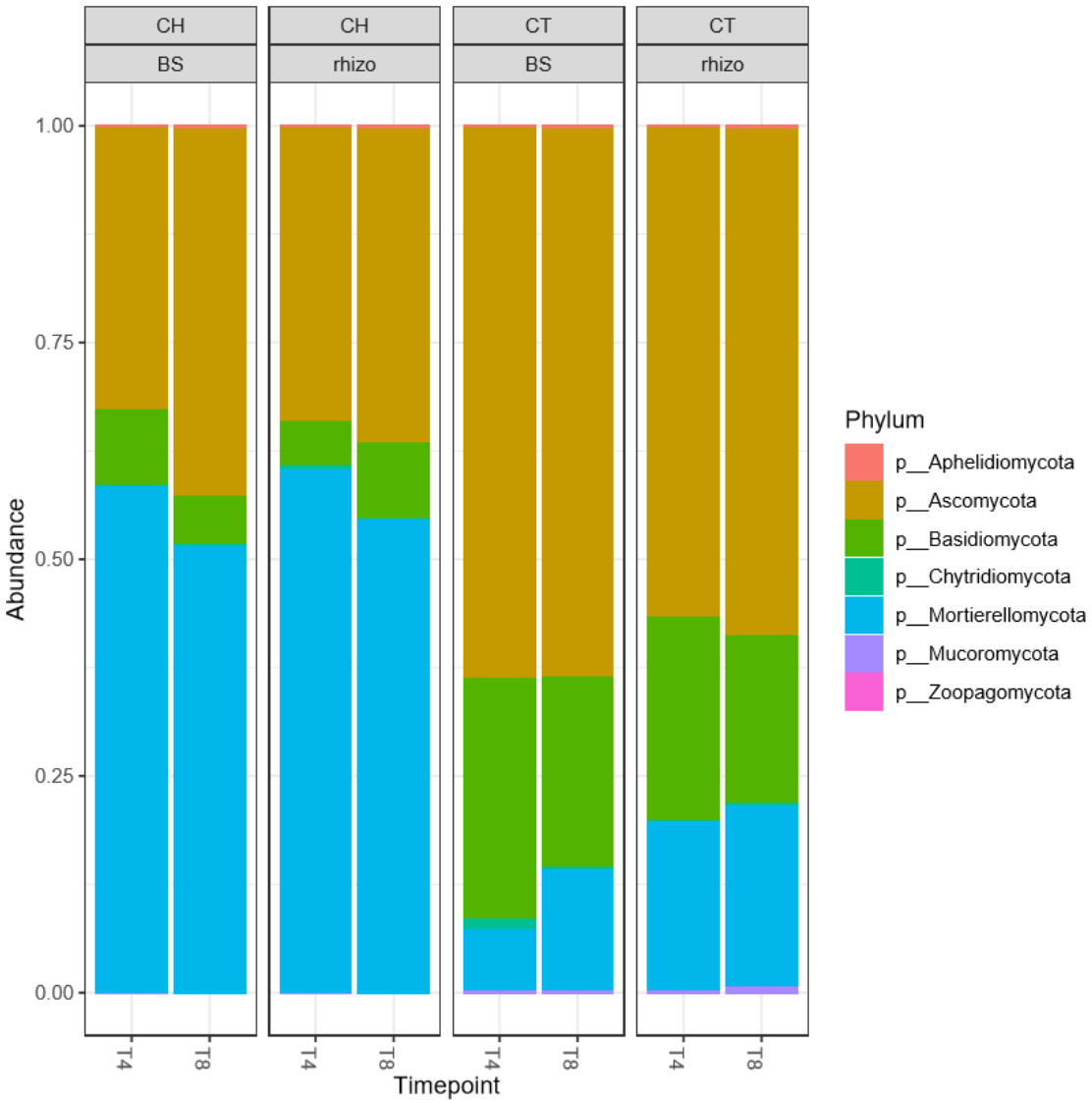


Figure 32: Top 10 phyla in the fungal samples

The *Mortierellaceae* are extremely abundant in the chitin-treated samples, both in rhizosphere and bulk potting soil, and after both 4 and 8 weeks (Figure 33). Due to the incredible abundance of *Mortierellaceae* in these samples, there is very low biodiversity, and an abundance of nearly 100% is obtained by plotting the top 10 families. The *Mortierellaceae* are also more abundant in the control rhizosphere than in the control bulk potting soil, suggesting that even without chitin addition, this fungus may have a close interaction with the plant's root system.

The bulk potting soil without chitin clearly has the most diverse fungal microbiome of all experimental conditions, particularly after 8 weeks, as can be seen by the bars not being nearly as close to 100% abundance as the others (Figure 33). This makes sense, since the fungi in the bulk potting soil have minimal interaction with the plant's root system, and without these external factors that can either stimulate or hinder growth, many more diverse fungi can establish a niche.

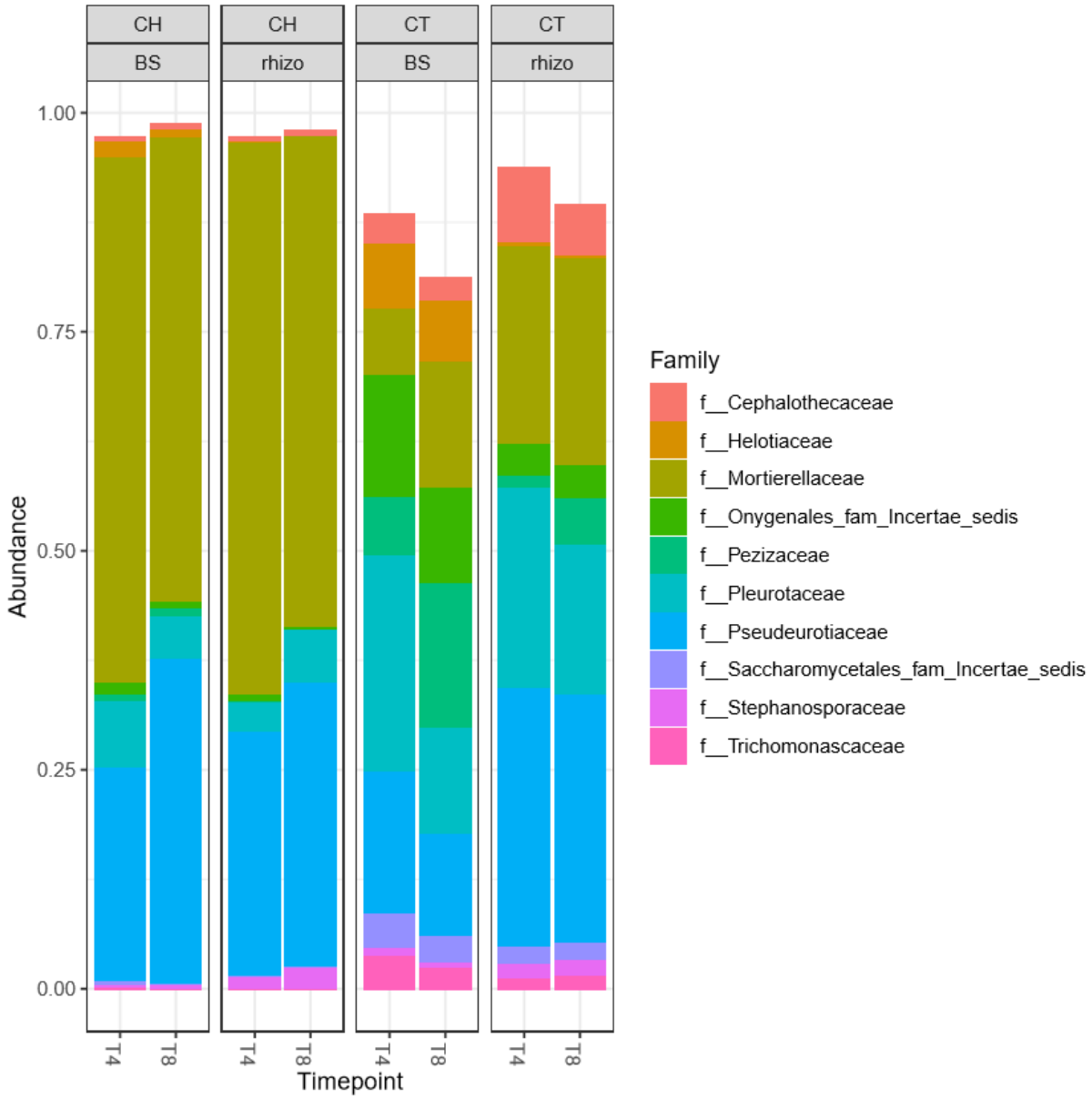


Figure 33: Top 10 families in the fungal samples

1.3 Statistical Analysis

1.3.1 PERMANOVA

Here we perform a statistical analysis on the Bray-Curtis dissimilarity matrices that contain the distance between groups. Remember, a group is defined as all replicates from the same experimental conditions. We conduct a PERMANOVA test, which is an alternative to the ANOVA test without the assumption of normally distributed data. With this test we will be able to prove or disprove the conclusions drawn from the beta diversity plots in Results sections *1.2.1: Bacteria* and *1.2.2: Fungi.*

We first tested the null hypothesis that the variances between all groups are equal (Tables 6 & 7). These tests show that the group variances within the bacterial dataset (Table 6) and fungal dataset (Table 7) are equal, since the P values are insignificant.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	7	0.0045455	0.00064936	0.4543	0.8522
Residuals	15	0.0214406	0.00142937		

Table 6: Analysis of Variance table on the betadisper object, BACTERIA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groups	7	0.068492	0.0097846	2.05	0.1111
Residuals	16	0.076367	0.0047729		

Table 7: Analysis of Variance table on the betadisper object, FUNGI

We are now allowed to perform PERMANOVA analysis on the Bray-Curtis dissimilarity matrices (Tables 8 & 9). The bacterial PERMANOVA shows that Treatment, Location and Timepoint all have a significant effect on the beta diversity at P = 0.05 (Table 8). This was predicted in Figures 20-23, where we saw a significant separation on the basis of Treatment on the first PCoA dimension (29.5%), Location on the second PCoA dimension (17.1%) and Timepoint on the fourth PCoA dimension (9.6%).

There also is a significant interaction effect between Treatment:Location, and Treatment:Timepoint at P = 0.05 (Table 8), which indicates that the Location and Timepoint effects both differ significantly between control and chitin-treated samples.

	Df	SumOfSqs	R^2	F	Pr(>F)	Signif. code
Treatment	1	0.67440	0.27584	15.4173	0.001	***
Location	1	0.38688	0.15824	8.8444	0.001	***
Timepoint	1	0.24673	0.11318	6.3262	0.001	***
Treatment:Location	1	0.13774	0.05634	3.1488	0.008	**
Treatment:Timepoint	1	0.18736	0.07663	4.2832	0.002	**
Location:Timepoint	1	0.07848	0.03210	1.7940	0.072	.
Treatment:Location:Timepoint	1	0.04718	0.01930	1.0786	0.367	
Residual	15	0.65615	0.26837			
Total	22	2.44491	1			

Table 8: Permutation tests for adonis under reduced model, BACTERIA

Signif. codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, ' ': 1

The fungal PERMANOVA shows that Treatment and Location have a significant effect on the beta diversity at P = 0.05, while Timepoint does not (Table 9). Indeed, in Figures 27-30 we clearly saw the effect of the chitin Treatment on the first PCoA dimension (58.2%), and the Location effect on the second PCoA dimension (12.5%). Little to no separation between groups was achieved when plotting on the third PCoA dimension (6.7%), which roughly corresponded to the Timepoint effect. From the PERMANOVA analysis, we now know on a confidence level of P = 0.05 that there was no significant Timepoint or Treatment:Timepoint effect on the fungal community (P = 0.066).

There is a significant interaction effect between Treatment:Location at P = 0.05 (Table 9), which indicates that the effect of the sampling Location significantly differs between control and chitin-treated samples.

	Df	SumOfSqs	R^2	F	Pr(>F)	Signif. Code
Treatment	1	1.8529	0.51839	38.9630	0.001	***
Location	1	0.4344	0.12153	9.1348	0.002	**
Timepoint	1	0.1182	0.03307	2.4857	0.066	.
Treatment:Location	1	0.2163	0.0605	4.5475	0.014	*
Treatment:Timepoint	1	0.1130	0.03162	2.3768	0.066	.
Location:Timepoint	1	0.0394	0.01101	0.8278	0.440	
Treatment:Location:Timepoint	1	0.0394	0.01100	0.8265	0.457	
Residual	16	0.7609	0.21287			
Total	23	3.5743	1			

Table 9: Permutation tests for adonis under reduced model, FUNGI
Signif. codes: '***': 0.001, '**': 0.01, '*': 0.05, '.': 0.1, ' ': 1

1.3.2 GLM LRT

We created a GLM from the count data and conducted LRT's for differential abundance testing between control and chitin-treated samples in every sampling condition. We performed these tests on the phylum and family levels (Tables 10-13). We then took the intersection of the differentially abundant phyla and families in all conditions to see which bacteria and fungi were differentially abundant between treatment and control, regardless of other conditions (timepoint, type of soil).

	T4 BS CT/CH	T4 rhizo CT/CH	T8 BS CT/CH	T8 rhizo CT/CH
-1	1	2	2	1
0	5	3	4	5
+1	1	2	1	1

Table 10: Differentially abundant phyla: FUNGI
-1: phylum with significant negative log FC
0: not differentially abundant phylum
+1: phylum with significant positive log FC

Intersection Table 10: Phylum *Mortierellomycota* (+1) and *Mucuromycota* (-1)

	T4 BS CT/CH	T4 rhizo CT/CH	T8 BS CT/CH	T8 rhizo CT/CH
-1	0	0	1	0
0	75	74	73	74
+1	1	2	2	2

Table 11: Differentially abundant families: FUNGI
-1: family with significant negative log FC
0: not differentially abundant family
+1: family with significant positive log FC

Intersection Table 11: Family *Mortierella*ceae (+1)

	T4 BS CT/CH	T4 rhizo CT/CH	T8 BS CT/CH	T8 rhizo CT/CH
-1	3	5	4	6
0	24	21	20	19
+1	2	3	5	4

Table 12: Differentially abundant phyla: BACTERIA
-1: phylum with significant negative log FC
0: not differentially abundant phylum
+1: phylum with significant positive log FC

Intersection Table 12: Phylum *Tenericutes* (+1) and *Fibrobacteres* (-1)

	T4 BS CT/CH	T4 rhizo CT/CH	T8 BS CT/CH	T8 rhizo CT/CH
-1	21	20	22	23
0	142	141	129	132
+1	17	19	29	25

Table 13: Differentially abundant families: BACTERIA
-1: family with significant negative log FC
0: not differentially abundant family
+1: family with significant positive log FC

Intersection Table 13: Family *Chitinibacteraceae* (+1)*, Sphingobacteriaceae* (+1)*, Streptomycetaceae* (+1)*, Cellvibrionaceae* (+1)*, Rhizobiaceae* (+1)*, Cytophagaceae* (+1)*, Fibrobacteraceae* (-1)*, Xanthomonadaceae* (+1)*, Paludibacteraceae* (-1)

We see from the intersections of Tables 10 & 11 that in the presence of chitin, only the *Mortierellaceae* family, *Mortierellomycota* phylum is significantly more abundant in the fungal community. This is in line with the observations in Figures 32 & 33 and with the hypothesis that the growth of a *Mortierella* strain is stimulated by chitin treatment. In Table 13 we see that after 8 weeks, many more families of bacteria are differentially abundant between treated and non-treated samples, as compared to the same samples at 4 weeks, especially in the bulk potting soil. This suggests that the chitin addition may have a long-term effect on the bacterial potting soil diversity.

1.3.3 Heatmap

From the LRT's we now know that *Mortierella* is the differentially overabundant fungus in the chitinous samples. To check which species is responsible for the increased abundance, we plotted all ASV's of genus *Mortierella* into a heatmap (Figure 34). Two *Mortierella* ASV's in particular have an outspoken increase in abundance in the chitin-treated samples, namely ASV 2 and 3. We looked up the taxonomy that was assigned to these ASV's, and found that they are *Mortierella hyalina* species,

which strengthens our hypothesis that this specific fungus was overrepresented in the chitin-treated samples.

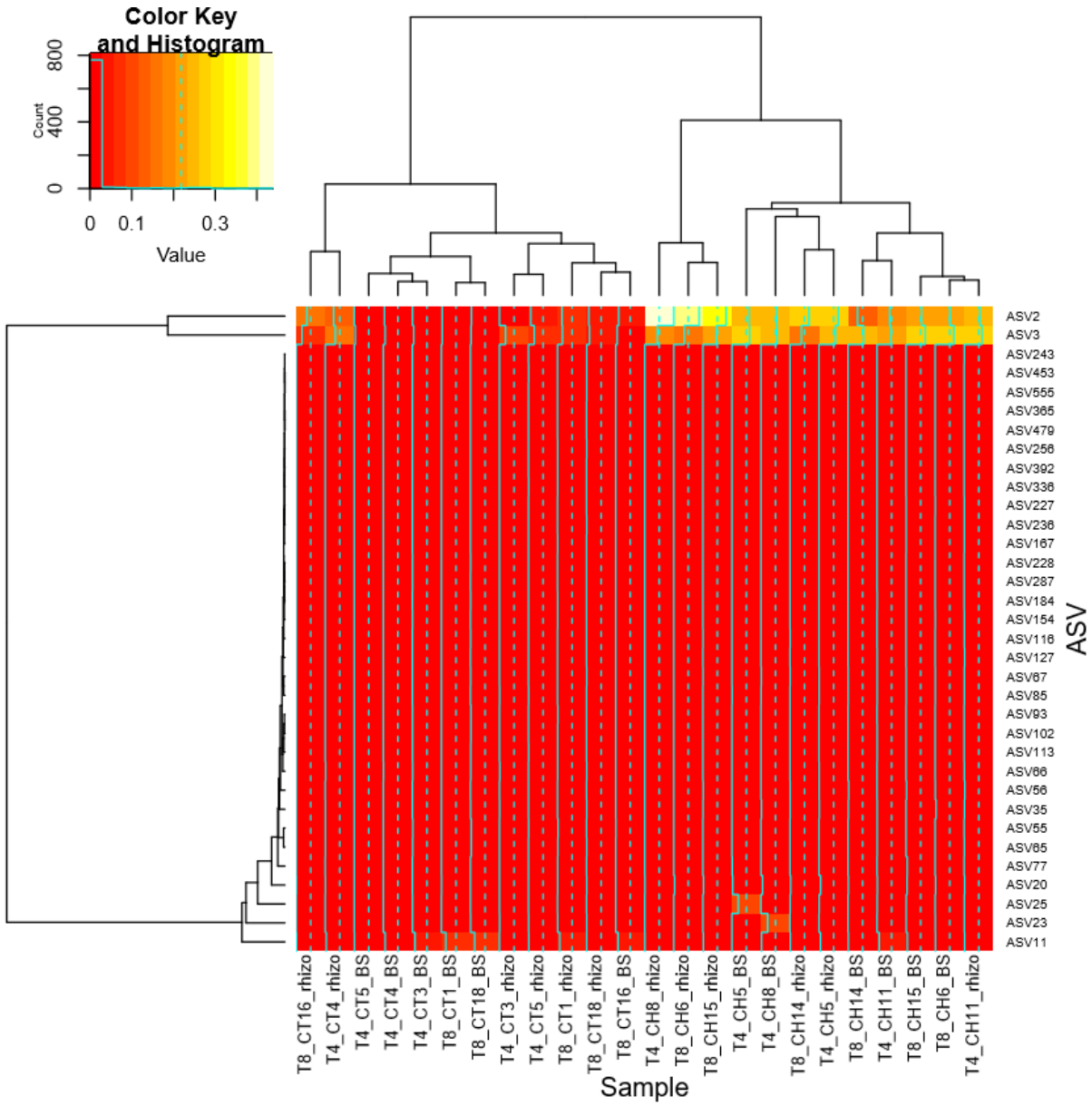


Figure 34: Heatmap of all ASV's taxonomically linked to genus *Mortierella*
Rows: ASV's, columns: samples, colour indicates abundance of ASV in sample

Note that this taxonomy is based on a small piece of the ITS region (metabarcoding: ITS2), and thus drawing conclusions up to the species level is quite overconfident. Therefore, in order to accurately determine the species taxonomy and confirm the hypothesis that the overabundant fungus is *Mortierella hyalina*, the strain should be isolated and the full ITS region should be sequenced, which is what we did in the next section.

2. Identification of *Mortierella* strain using ITS phylogeny

2.1 Phylogenetic Analysis

We sequenced the full ITS region of isolated mycelium samples from chitin-treated potting soil, and a control plant pot. To confirm that the overrepresented fungus in the chitin-treated samples is the species *Mortierella hyalina*, we created phylogenetic trees comparing our sequenced samples to all *Mortierella* sequences in the UNITE database (Figures 35-37). The strains we used for constructing the phylogenetic trees together with their full sample names as they appear in the trees are summarized in Table 2 (copy from Methods section 2.1: Preprocessing). The full-length trees are quite large and can be found in the supplementary materials. Here we zoom in on certain aspects of the trees that catch the eye.

Strain	**Full sample name in tree**	**Experimental conditions**	**Preprocessed when**
Chitin1	"*Mortierella hyalina* Chitine 11 uit 181220"	Chitin treatment, potting soil sampling	Previous experiment
Chitin11	"*Mortierella hyalina* Chitine 1 uit 181220"	Chitin treatment, potting soil sampling	Previous experiment
Chitin20rh	"*Mortierella hyalina* Chitine 20 rh eigen controle uit 181220"	Chitin treatment, rhizosphere sampling	Previous experiment
Chitin20rh	"Chitin20rh"	Chitin treatment, rhizosphere sampling	This thesis
Chitin7rh	"Chitin7rh"	Chitin treatment, rhizosphere sampling	This thesis
Control19rh	"Controle19rh"	No chitin treatment, rhizosphere sampling	This thesis

Table 2: ITS phylogeny dataset (copy from Methods section 2.1: Preprocessing)

Firstly, notice that "*Mortierella hyalina* Chitine 20 rh eigen controle uit 181220" and "Chitin20rh", our controls for the preprocessing, are always located in close proximity to one another (Figure 35, 36, Supplementary Figures 4,5,6). This tells us that the preprocessing was done correctly, since these sequences originate from the same sample, but one of them was preprocessed in this thesis, and the other was preprocessed by a different researcher in a previous experiment at ILVO.

In Figure 35 we see that the chitin11 sample is grouped with 2 *Mortierella hyalina* samples from the UNITE database (AY157494 and KC922119). These samples are colored red. The chitin20rh and chitin1 samples are also grouped with a *Mortierella hyalina* sample (HQ630355) and with an unidentified *Mortierella sp*. sample (which is likely *Mortierella hyalina*) from the UNITE database. These samples are colored orange.

In Figure 36 we applied the same coloring to these samples, including another UNITE *M. hyalina* (JX898566) in the orange group. We clearly see that the chitin20rh, chitin11 and chitin1 samples fall into a clade of *Mortierella hyalina*.

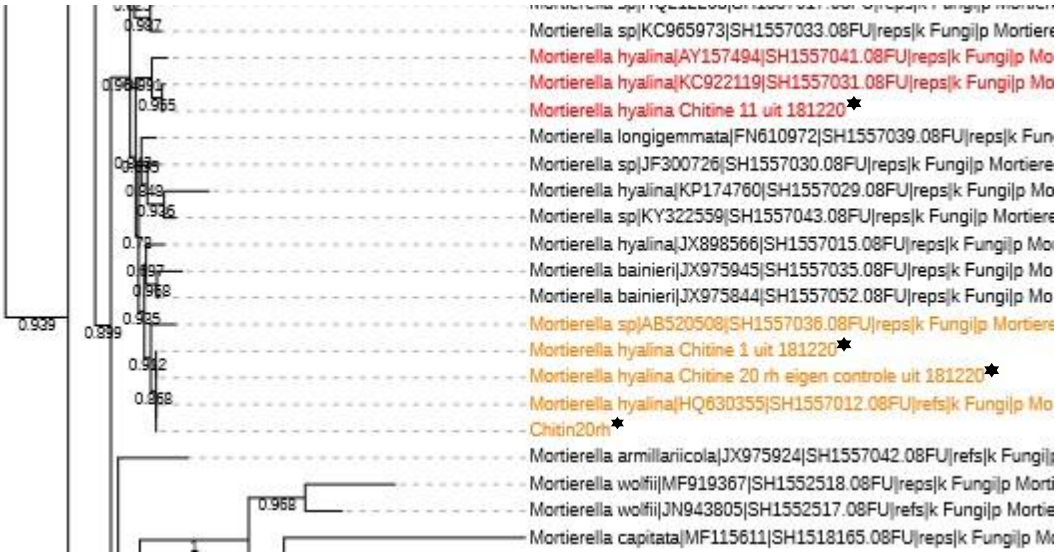


Figure 35: Fasttree ML tree, close-up on *Mortierella hyalina*

The first cluster of *Mortierella hyalina* that contains at least 1 sample from our dataset is colored red, the second cluster is colored orange.
Sequences from our dataset are marked with a star (*).
Numbers shown on nodes are bootstrap values.

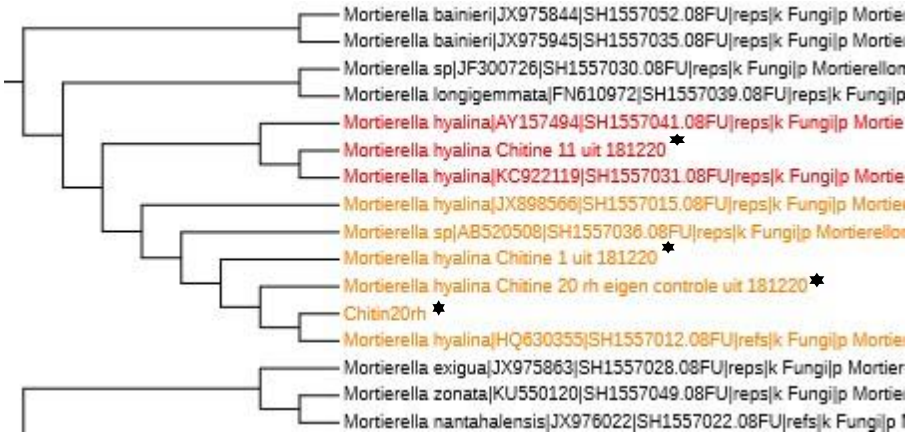


Figure 36: Fastphylo NJ tree, close-up on *Mortierella hyalina*

The same red and orange clusters from Figure 35 are identified.
Sequences from our dataset are marked with a star (*).

Figure 37 shows the control19rh and chitin7rh sample, colored green. These two samples fall all the way at the end of the tree, which indicates that they have the least similarity to any of the *Mortierella* species in the database. This occurs in every tree we built. To get a clearer view on how vast the trees are and how isolated these two samples are, take a look at the full trees in Supplementary Figures 9-11. We knew that the control19rh sample was probably not going to be *Mortierella*, since it came from potting soil without chitin. The chitin7rh sample however was hypothesized to be *Mortierella hyalina*, yet it does not group together with the other samples. A BLAST search shows that these two samples are indeed not *Mortierella*. Control19rh is *Umbelopsis isabellina*, from the family *Umbelopsidaceae*, and chitin7rh is *Pseudogymnoascus pannorum*, from the family *Pseudeurotiaceae*.

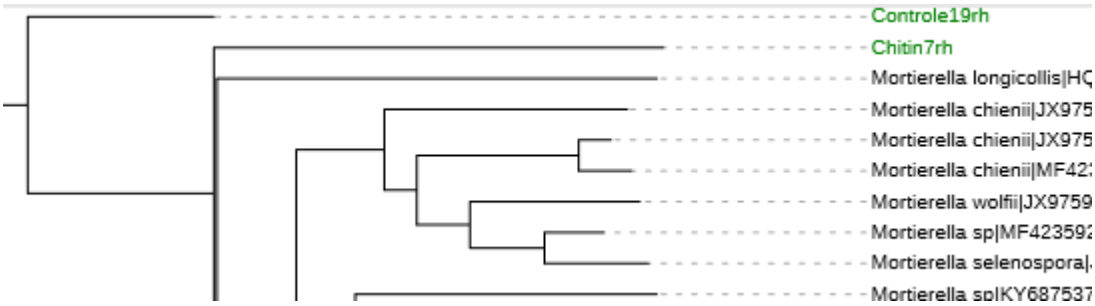


Figure 37: EBI NJ tree, close-up on the control sequence
Sequences from our dataset are colored green.

We looked up the *Umbelopsidaceae* and *Pseudeurotiaceae* families in Supplementary Table 2. *Pseudeurotiaceae* is the first entry in the table, *Umbelopsidaceae* is the third to last entry. The *Umbelopsidaceae* are more abundant in the control conditions than in the chitin-treated conditions, and the *Pseudeurotiaceae* are more abundant in the chitin-treated conditions than in the control conditions, except for the T4 rhizosphere condition where they are slightly more abundant in the control group.

3. Genome Analysis

3.1 De novo assembly

We first perform QC on the .fastq files with FastQC. We see that the Illumina HiSeq data is of exquisite quality, with the average phred score being over 30 everywhere (Figure 38, top plots). The "Per base sequence content" plots (Figure 38, bottom plots) show that the reads are randomly distributed among the genome (the entire genome is sequenced) so the position in a read should not have any influence on the base being found. In this circumstance the plot should be horizontal at the average GC/AT content, which can be seen from base 15 onward. Bases 1-15 are an Illumina adapter that must be trimmed.

At the end of the Shovill pipeline, we assessed the quality of the assembly with QUAST. The results are summarized in Table 14 and visualized in Figures 39-41. The final assembly contains 3059 contigs totalling 46.55 Mb. As we know from the literature study, the *Mortierella elongata* genome is 49.85 Mb in size and has a GC content of 48.1%. This is quite close to our *M. hyalina* genome size of 46.55 Mb, and GC content of 48.66% (Table 14).

The N50/75 value is the contig length such that 50/75% of bases in the assembly can be produced by contigs that are larger than or equal to this length (Table 14). This is also shown in Figure 40.

The L50/75 value is the minimum number of contigs that produce 50/75% of bases in the assembly, or the number of contigs of length at least N50/75 (Table 14).

Note that the QUAST summary also includes mismatch statistics that indicate that there were 0 mismatches between the assembly and the reference genome (Table 14). This is of course not applicable to our data, for we ran a *de novo* assembly, which does not use a reference genome, hence no mismatches can occur.

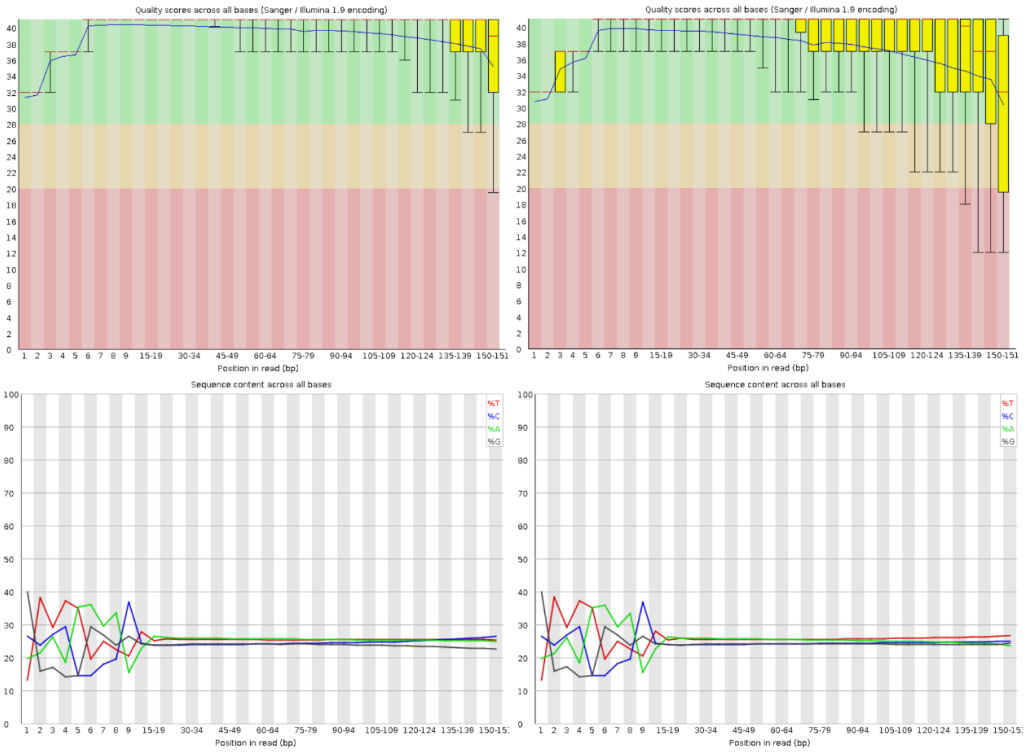


Figure 38: FastQC results
Top: Per base sequence quality plot
Bottom: Per base sequence content plot
Left: FWD reads, right: REV reads

# contigs	3059
# contigs (>= 0 bp)	3059
# contigs (>= 1000 bp)	2430
# contigs (>= 5000 bp)	1549
# contigs (>= 10000 bp)	1173
# contigs (>= 25000 bp)	617
# contigs (>= 50000 bp)	241
Largest contig	307113
Total length	46548106
Total length (>= 0 bp)	46548106
Total length (>= 1000 bp)	46103374
Total length (>= 5000 bp)	44051764
Total length (>= 10000 bp)	41350290
Total length (>= 25000 bp)	32244269
Total length (>= 50000 bp)	19000076
N50	40678
N75	20480
L50	336
L75	736
GC (%)	48.66
Mismatches	
# N's	0
# N's per 100 Kb	0

Table 14: QUAST summary of assembly

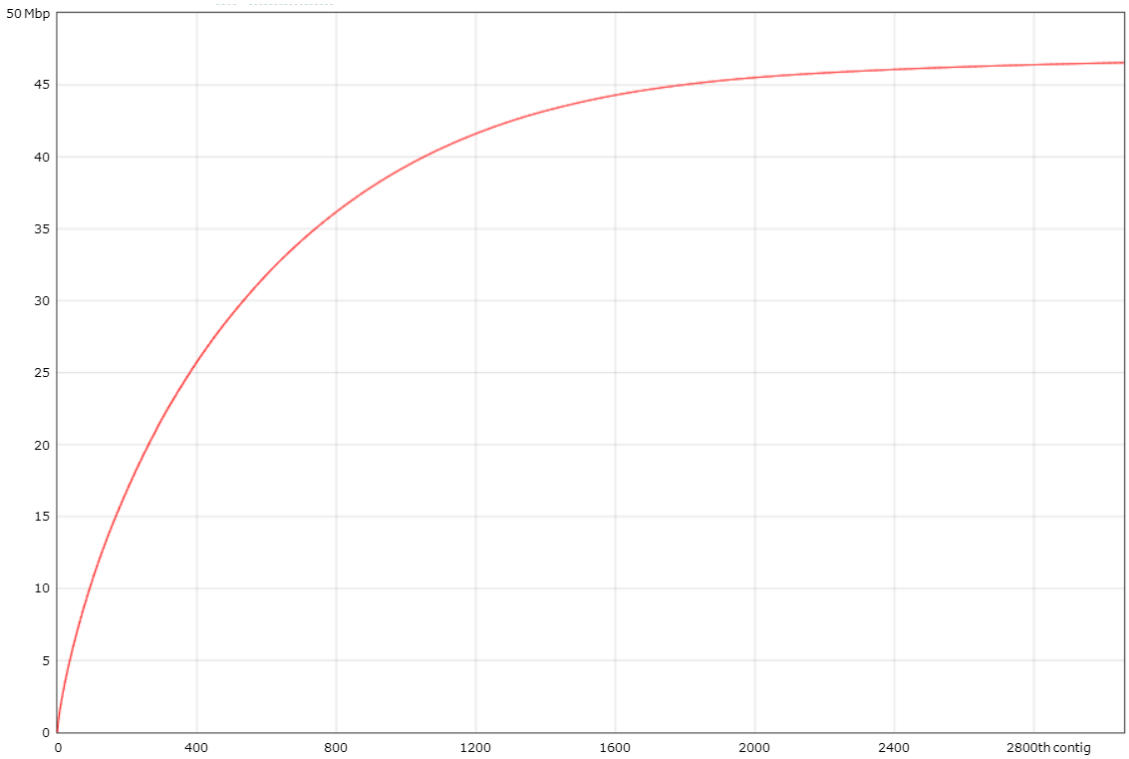


Figure 39: Cumulative length plot

x-axis shows the total number of contigs, sorted from largest to smallest

y-axis shows the total genome size in Mb

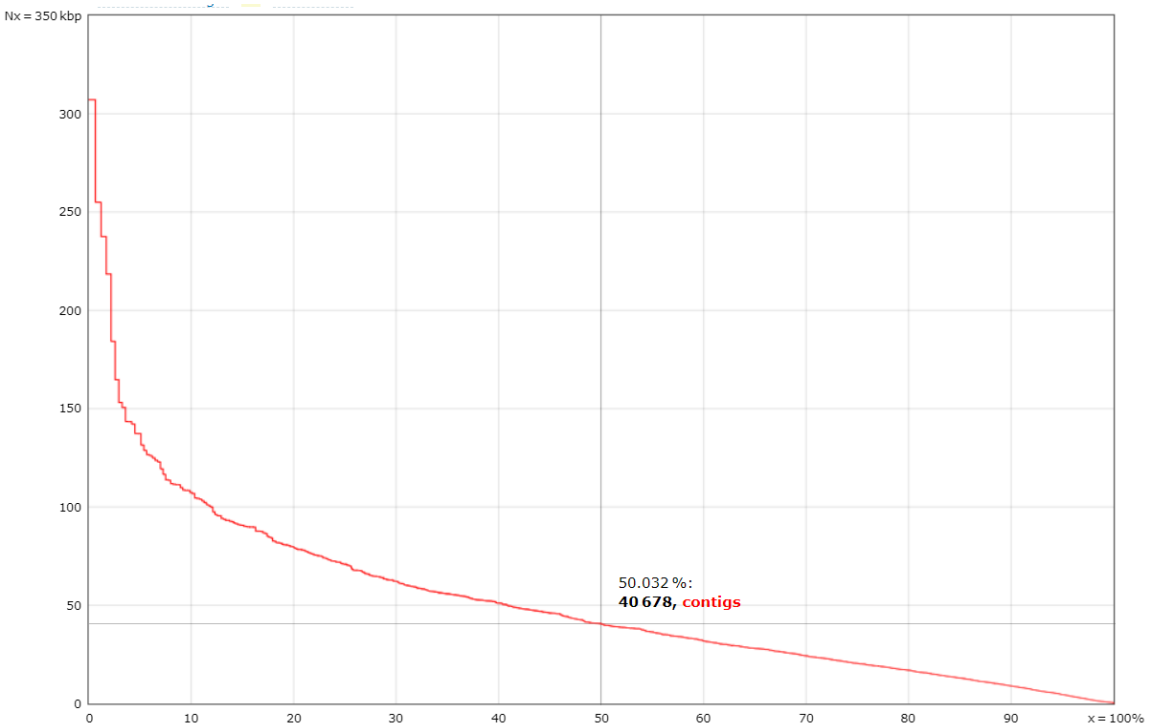


Figure 40: Nx plot

x-axis shows the percentage of contigs with a size greater than the value on the y-axis

y-axis shows the size of contigs in Kb

The value indicated on the graph is the N50

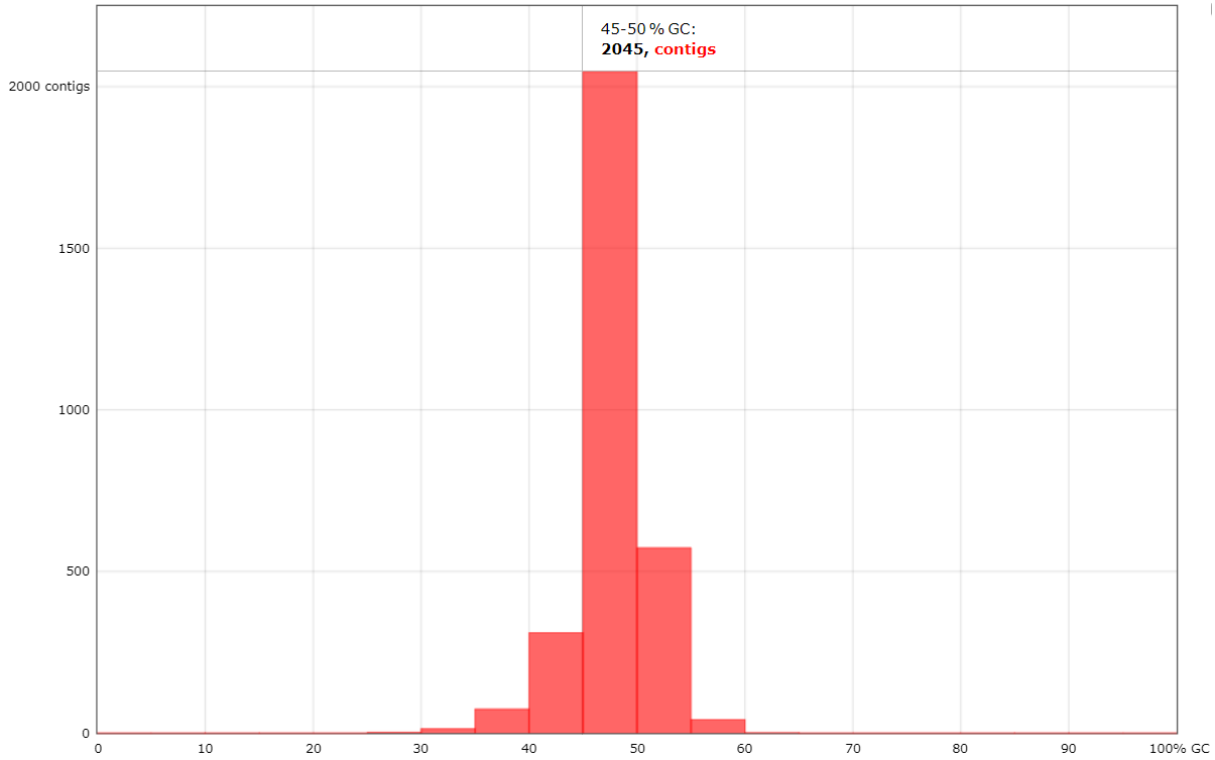


Figure 41: GC bar graph
x-axis shows the % GC content in 5% windows
y-axis shows the number of contigs
The value indicated on the graph shows that roughly 2/3 of the contigs fall into the 45-50% GC range.

The cumulative length plot (Figure 39) shows that a total length of 45 Mb is reached around 1800 contigs, meaning that the 1200 smallest contigs only account for 3% of the information in the assembly. This suggests that our decision to delete contigs smaller than 500 bp to reduce assembly errors is justified, since the smallest contigs really account for nearly no extra information in the assembly and would only add noise.

The GC bar graph (Figure 41) shows that the GC content of the contigs is normally distributed around the 48.66% GC value.

3.2 Genome Annotation

The workflow for genome annotation is summarized in Figure 42. In the Top-Down workflow we used AUGUSTUS to predict all genes in the *Mortierella hyalina* genome. We then translated the genes into amino acids and used Diamond BLASTp to query Uniprot for similar proteins. This resulted in a 21 MB-sized file with the 10 best matches in Uniprot for every predicted protein in the *Mortierella hyalina* genome. In the Bottom-Up workflow we transformed the *Mortierella hyalina* assembled genome into a database, and queried 42 curated fungal chitinases (EC 3.2.1.14) against it. With BLASTp we queried the chitinases against a database of the putative proteome generated in the AUGUSTUS pipeline, and with tBLASTn we queried the chitinases against the contigs generated in the Shovill pipeline. This brought forth 2 files: a 39 KB-sized file for the BLASTp output, and an 84 KB-sized file for the tBLASTn output. We will start with the analysis of the Bottom-Up files.

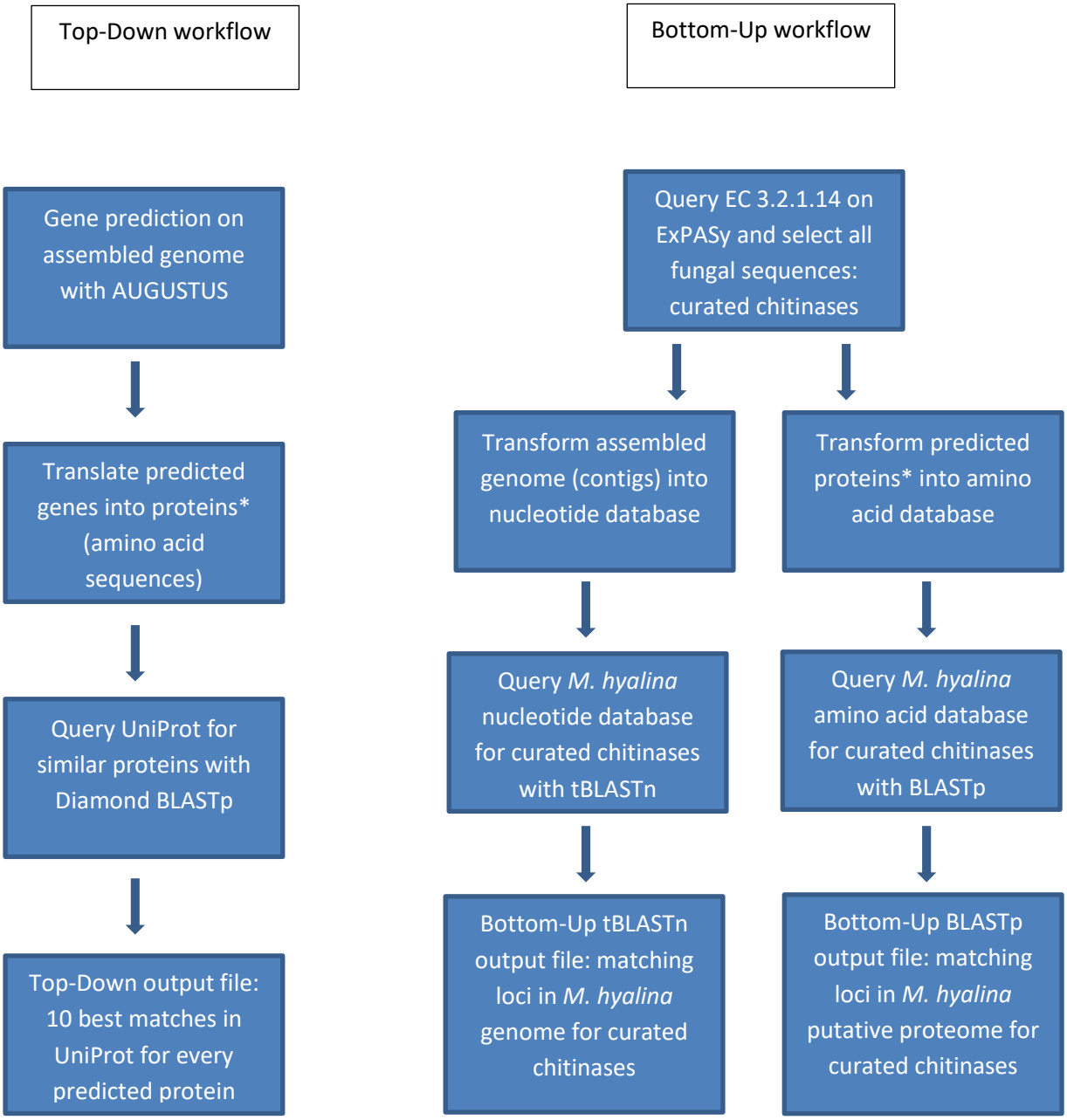


Figure 42: Summary of Top-Down and Bottom-Up genome annotation workflow
The star (*) indicates that the Bottom-Up workflow uses this file from the Top-Down workflow, therefore the Top-Down workflow should be executed first

First a redundancy check was needed in the Bottom-Up BLASTp and tBLASTn output files. There are two reasons for redundancy in the data: 1) many of the chitinases in the query will match the same locus in the genome, because of their sequence similarity & 2) every query chitinase can match multiple loci in the genome due to gene duplications. After filtering the BLASTp output for E-value < 1.00e-2, we found that 37 query curated chitinases matched one or more of 15 subject proteins in the genome. By performing the same filtering on the tBLASTn output, we found that the same 37 query chitinases matched on 11 contigs in the genome, suggesting that either multiple chitinases in the *M. hyalina* genome are located on the same contig in the assembly, or that the tBLASTn query did

not capture some of these genes with the same confidence. By consulting the .gff file we located all 15 subject proteins called in the BLASTp query on a set of 15 different contigs, in which all 11 contigs called in the tBLASTn query are contained.

The Bottom-Up files only show information on the sequence comparison to the curated chitinases and the locus of the matching gene. To unveil the identity of the inferred chitinases we need to match the locus to the Top-Down file. This is done in Table 15. From the filtered output files the subject loci that matched one or more of the curated chitinases were retrieved. They are listed in the first column of Table 15. We then queried the Top-Down output file for these loci, to see if the results match. The second column shows the gene name that the Top-Down BLAST matched the inferred chitinase with. The third column shows the contig on which the genes were found.

sSeqID	Top-Down identification	Contig	Length (AA)
g7031.t1	Glycoside hydrolase family 18	contig00296*	764
g6612.t1	Class III Chitinase	contig00268*	453
g14063.t1	Glycoside hydrolase family 18	contig01279*	446
g3147.t1	Endochitinase	contig00089*	717
g876.t1	Glycoside hydrolase family 18	contig00016*	522
g14196.t1	Glycoside hydrolase family 18	contig01325*	542
g12165.t1	Glycoside hydrolase family 18	contig00849*	483
g3857.t1	Chitotriosidase	contig00120*	448
g8506.t1	Glycosyl hydrolase family 18	contig00411*	437
g6192.t1	Chitotriosidase	contig00242*	504
g5430.t1	Chitinase	contig00199*	968
g10170.t1	Uncharacterized protein	contig00577	366
g12022.t1	Uncharacterized protein	contig00823	1820
g2215.t1	Transcription elongation factor	contig00056	2395
g3818.t1	Uncharacterized protein/ TOG domain-containing protein	contig00117	1416

Table 15: Inferred proteins in the *Mortierella hyalina* genome assembly that matched the curated chitinases in the BLASTp query
sSeqID: subject sequence identification as shown in the BLASTp output
Top-Down identification: name of the gene that matched this locus in the Top-Down query
Contig: contig identification as shown in the .gff file, contigs that were found in the tBLASTn output are marked with a *
Length: length of the protein in amino acids

In Table 15 many of the chitinases are referred to as “glycoside hydrolase family 18”, which we know are enzymes that hydrolyse the glycosidic bond between carbohydrates, and fungal endochitinases are indeed grouped into GH family 18. This shows that many, but not all, matches from the Bottom-Up query can be found in the Top-Down query. Subject sequences g10170.t1 and g12022.t1 only matched an uncharacterized protein of *Mortierella elongata*. These chitinases would thus not have been found with only a Top-Down search, which proves the importance of a multi-perspective approach to genome annotation. Subject sequences g2215.t1 and g3818.t1 matched with a transcription elongation factor and TOG domain-containing protein. The latter was also classified as “uncharacterized” in *M. elongata*. This transcription elongation factor thus bears a structural resemblance to curated chitinases, and the TOG-domain might be found in an unannotated chitinase gene, definitely interesting topics for follow up research.

Table 16 shows other genes that we found in the *Mortierella hyalina* genome. These genes are mainly plant growth-promoting genes, plant immune system stimulating genes, genes linked to the N-cycle and other chitin degradation-related genes that will be elaborated on in the Discussion.

sSeqID	Top-Down identification	Contig	Length (AA)
g5146.t1	Nitrogen assimilation transcription factor nit-4	contig00183	1074
g1039.t1	Nitrogen assimilation transcription factor nit-4	contig00020	1064
g17.t1	Nitrogen assimilation transcription factor nit-4	contig00001	1450
g11441.t1	Nitrogen assimilation transcription factor nirA	contig00734	349
g761.t1	Auxin efflux carrier	contig00014	395
g7242.t1	Auxin efflux carrier	contig00310	564
g7243.t1	Auxin efflux carrier	contig00310	557
g8788.t1	Auxin efflux carrier	contig00436	540
g8815.t1	Auxin efflux carrier	contig00439	774
g3584.t1	PR-1 like protein	contig00107	201
g3585.t1	PR-1 like protein	contig00107	188
g3586.t1	PR-1 like protein	contig00107	478
g4737.t1	PR-1 like protein	contig00161	222
g6549.t1	PR-1 like protein	contig00264	238
g6550.t1	PR-1 like protein	contig00264	239
g8757.t1	PR-1 like protein	contig00433	219
g8251.t1	Alpha-N-acetylglucosaminidase	contig00387	957
g8252.t1	Alpha-N-acetylglucosaminidase	contig00387	1038
g8450.t1	Endo-beta-N-acetylglucosaminidase	contig00406	986
g11085.t1	Alpha-N-acetylglucosaminidase	contig00682	833
g2392.t1	Chitin deacetylase CDA1	contig00062	438
g2555.t1	Putative chitin deacetylase	contig00068	375
g4254.t1	Chitin deacetylase	contig00137	155
g4259.t1	Chitin deacetylase	contig00137	249
g4958.t1	Chitin deacetylase	contig00173	264
g8040.t1	Putative chitin deacetylase	contig00370	366
g8642.t1	Chitin deacetylase CDA1	contig00423	252
g8643.t1	Chitin deacetylase CDA1	contig00423	386
g9717.t1	Chitin deacetylase CDA2	contig00528	355
g15468.t1	Chitin deacetylase	contig02062	151

Table 16: Other proteins found in the *Mortierella hyalina* genome assembly

sSeqID: subject sequence identification

Top-Down identification: name of the gene that matched this locus in the Top-Down query

Contig: contig identification as shown in the .gff file

Length: length of the protein in amino acids

Discussion

This concluding section is organized into 4 parts. First, we discuss the results of the metabarcoding analysis and the genome assembly, then we will feed these new insights back to the application in horticulture, and finally we will look at future perspectives.

1. Metabarcoding of the bacterial and fungal community

We performed metabarcoding to study the bacterial and fungal microbiome of the bulk potting soil and rhizosphere of lettuce plants, grown with or without chitin addition to the potting soil. The chitin treatment had a massive impact on the biodiversity of the fungal community, which plummeted because of the enormous enrichment of one particular fungus, which we identified as *Mortierella hyalina* based on the ITS sequence. Both fungal and bacterial biodiversity of the bulk potting soil was higher than that of the rhizosphere in non-treated samples. This observation is backed by Hartmann et al. (2009), who reported that microbial communities are less diverse in the rhizosphere of plants, due to the "biased rhizosphere" concept: plant root secretions have a selective effect on specific microorganisms that exert plant-growth promoting or biocontrol effects. The different composition of the rhizosphere is responsible for a higher amount of N-cycling, which can lead to more nitrogen being available for the plant, with all due plant growth-promoting consequences (De Tender et al., 2016). We do notice in our data that the biased rhizosphere effect is somewhat diminished in the fungal community of the chitin-treated samples. This can be explained by the fact that *Mortierella* dominated the fungal community in the presence of chitin, with relative abundances of over 50%, which minimizes the effect of other factors on the biodiversity (Figures 32, 33).

The richness of the bacterial community was also deeply affected by the treatment, sampling location and sampling timepoint, but these effects were harder to visualize. The reason for this is that there is an enormous biodiversity of bacteria in potting soil. Soil is considered the most diverse ecosystem on earth (Curtis et al., 2002), and while potting soil is less diverse than normal soil, it still has a huge biodiversity compared to other ecosystems such as the rhizosphere or the human microbiome, which is estimated to have only 10% of the diversity of a soil microbiome (Blum et al., 2019). This large biodiversity of bacteria in potting soil complicates the analysis; hence the bacterial abundance plots were not as straight-forward to interpret as their fungal counterparts. Many bacteria were influenced in different ways by the chitin treatment and the plant's secretions in the rhizosphere, but these shifts in biodiversity are difficult to capture in abundance plots. There was no evidence for a huge increase in abundance of 1 particular bacterial taxon due to chitin treatment.

We performed a PERMANOVA analysis to statistically prove the dissimilarity in both bacterial and fungal diversity between sampling groups. We will begin by discussing the results of the fungal community. The statistical evidence showed a significant effect of Treatment and Location on the fungal community, and a significant interaction effect between Treatment & Location on the fungal biodiversity at $P = 0.05$. This proves that the fungal community composition of the chitin-treated group is different from the non-treated groups, both in bulk potting soil (Treatment effect) and rhizosphere (Treatment:Location effect), and also that the fungal community composition of the rhizosphere of the non-treated group is different from the bulk potting soil of the non-treated group

(Location effect). Both Treatment and Treatment:Location effects are explained by the increased abundance of *Mortierella* because of the chitin treatment, regardless of sampling location.

The Location effect is accredited to the biased rhizosphere concept, which can be seen in Figures 32, 33. We again notice that the difference in community composition is mainly due to the *Mortierellomycota/Mortierellaceae*, that are more abundant in the rhizosphere than in the bulk potting soil, both in chitin-treated and control conditions. Observe the reversed behavior of the Timepoint effect in control and chitin-treated conditions. In the chitin-treated samples, the relative abundance of *Mortierella* decreases over time. This can be explained by the fact that the chitin feedstock starts to run out, and the immense levels of *Mortierella* in the soil and rhizosphere are no longer sustainable. Yet, in the control samples, *Mortierella* abundance increases over time, indicative of an underlying selection towards the *Mortierellomycota/Mortierellaceae*. This may be attributed to the ever changing nature of soil microbiomes (Turner et al, 2017). Soil communities are incredibly diverse and dynamic. The rise of the *Mortierellomycota/Mortierellaceae* in the control conditions may well diminish again over time, but we cannot know because no more samples were taken after 8 weeks. Note that these last observational comments are based on the abundance plots; the Timepoint effect just fell short of being statistically significant on the fungal community composition at P = 0.05, as both P(Timepoint) and P(Treatment:Timepoint) were equal to 0.066.

The PERMANOVA on the bacterial data showed that all studied experimental variables had a significant effect on the bacterial community, and the interaction effects between Treatment & Location, and Treatment & Timepoint were also significant at P = 0.05. The Location effect in the bacterial data can again be attributed to the biased rhizosphere. It has been known for years that bacteria in the rhizosphere co-operate and form an ecosystem that benefits both the bacteria and the plant host (Barea et al., 2005).

The chitin-treated groups are different from the non-treated groups, both in bulk potting soil (Treatment effect) and rhizosphere (Treatment:Location effect). The chitin treatment thus did have a significant influence on the bacterial community, yet there was not 1 particular taxon that stood out. Instead, it changed the relative abundances of many present bacteria in a broader sense. The bacterial community also changes over time, both in control and chitin-treated conditions (Timepoint and Treatment:Timepoint effects). The significant Timepoint effect is indicative of the influence that bacteria have on each other. Fluctuations in nutrients and biochemical composition of soil due to the metabolism and secretions of bacteria cause the microbiome to continuously change (Turner et al., 2017). The combination effect between Treatment and Timepoint shows that the bacterial community composition of chitinous soil evolves in a fundamentally different way than that of normal potting soil.

In the EdgeR analysis we took a deeper look at the effect of the chitin treatment on the bacterial and fungal community composition by determining which phyla & families were differentially abundant in all sampling conditions. By taking the intersection of the differentially abundant phyla/families in every condition, we could deduce which bacteria and fungi were most affected by the chitin treatment, regardless of sampling timepoint or location. The fungal response to the chitin treatment was evident; only *Mortierella* was overabundantly present in the chitin-treated samples. The

bacterial response to the chitin treatment was more nuanced, as we found 6 bacterial families with a significant positive log fold-change (*Chitinibacteraceae, Sphingobacteriaceae, Streptomycetaceae, Cellvibrionaceae, Rhizobiaceae* and *Cytophagaceae*), and 3 families with a significant negative log fold-change (*Fibrobacteraceae, Xanthomonadaceae, Paludibacteraceae*). These results are partly in line with those of Debode et al. (2016), who found that *Sphingobacteriaceae*, *Streptomycetaceae*, *Cytophagaceae,* and *Bradyrhizobiaceae* (which are closely related to *Rhizobiaceae*) were differentially overabundant in chitin-treated soil samples.

The fact that *Mortierella* is differentially overabundant in all chitin-treated sample groups, combined with the evidence in the heatmap (Figure 34) and phylogenetic trees (Figures 35, 36) that the most abundant *Mortierella* species in the data was *Mortierella hyalina* confirms our hypothesis that a chitin treatment significantly increases the abundance of *Mortierella hyalina*. In the next section we will tackle how *Mortierella hyalina* is able to thrive in these conditions and why it might prove to be beneficial for crop growth and resilience, by sequencing and deciphering its genetic code.

2. Genome Annotation

One would expect of a fungus that thrives in a chitin-rich environment that it would be able to degrade chitin into GlcNAc, its basic compound, which can be utilized for anabolism. Enzymes with this functionality are generally referred to as chitinases.

On top of the many chitinolytic GH family 18 enzymes that were found in the *Mortierella hyalina* assembly through a Bottom-Up query (Table 15), other chitinolytic enzymes were retrieved from the assembly in the Top-Down output (Table 16). Examples are N-acetylglucosaminidase and endochitinase, which enable the degradation of chitin to GlcNAc, and chitin deacetylase, which facilitates chitin conversion into chitosan. Chitosan is known for its plant-growth-promoting properties. Salachna & Zawadzińska (2014) have shown that regardless of the molecular weight of the administered chitosan polymer, chitosan-supplemented freesia plants grew larger and faster. Rahman et al. (2018) studied the effect of chitosan treatment to strawberry plants, and found a significant increase in fruit yield and plant health.

In the Top-Down pipeline output, we queried the genome for plant growth promoting genes, plant immune system stimulating genes, and genes linked to the N-cycle (Table 16). We identified multiple auxin efflux carrier genes, that might play part in nitrogen fixation in the rhizosphere (Ng et al., 2015), and PR-1 genes, that encode salicylic-acid responsive pathogen-related signaling molecules, indicative of systemic acquired resistance (Kojima et al., 2013). We also found nitrogen assimilation transcription factor *nirA* and nitrogen assimilation transcription factor *nit-4*, which are actively transcribed genes for N-cycling. Previous studies have reported on the chitin-degrading properties of *Mortierella sp.* (Kim et al., 2008) and on the increased relative abundance of *Mortierellomycota* in chitin- or GlcNAc-enriched soil (Zegeye et al., 2019), however, little is known about the actual molecular pathways that *Mortierella* uses in its N-cycling role in ecosystems.

This thesis presents a standardized pipeline for discovering many more enzymes in the *Mortierella hyalina* genome. By running a Swiss-Prot query with a different EC number you can retrieve a new list

of curated fungal enzymes. This list can then be the starting point of a new Bottom-Up query through the assembled genome.

The combination of a Top-Down and Bottom-Up approach to genome annotation presents another method to identify potential new chitinases or N-cycling genes (Figure 42). We could use a protein that was found in the putative proteome from the Top-Down AUGUSTUS pipeline as the query for another Bottom-Up BLASTp/tBLASTn in the databases of the *Mortierella hyalina* assembly. This enables us to identify genes that are similar in structure to a gene of interest, but that were not identified as such by the initial Top-Down BLASTp, due to a lack of annotation in related species. As a proof of concept we ran this pipeline on the *nirA* gene, that was found only once in the Top-Down query. A quick look into the results of the Bottom-Up query reveals multiple sequences similar to the *nirA* gene found in the initial Top-Down query of the *Mortierella hyalina* assembly.

3. Importance in horticulture & conclusion

Plant Growth-Promoting Fungi (PGPF) have been shown to be capable of producing phytohormones and reprogramming plant gene expression to control pathogens (Hossain et al., 2017). In a previous study at the ILVO, strawberry plants were infected with *Botrytis cinerea* and gene expression levels were compared between chitin-treated and control plants (De Visscher, 2019). They found that 482 genes were differentially expressed upon infection of chitin-treated plants, whereas control plants only differentially expressed 132 genes. The gene expression seen here might be influenced by a PGPF that was abundant in the chitinous conditions. Tagawa et al. (2010) showed that *Mortierella sp.* are able to protect potato plants against potato scab pathogens.

So, is our *Mortierella hyalina* strain plant growth-promoting? Well, the data shows that the chitin treatment dramatically increased the abundance of *Mortierella hyalina* in the rhizosphere, where it would be able to act as a PGPF. De Tender et al. (2019) found that lettuce growth promotion in the presence of chitin can be accredited to the degradation of chitin to GlcNAc and ammonium. Furthermore, we know that chitosan, a chitin derivative, may act as a MAMP to directly induce gene expression in plants (Vijayalakshmi et al., 2018). Yet other studies show that the Induced Systemic Resistance (ISR) in plants grown in chitinous soil is linked to the favorable environment that chitin creates for chitinolytic organisms that act as biocontrol agents for plant diseases (Ramirez et al., 2010). Of course all explanations can be true, and all can be linked to our *Mortierella hyalina* strain.

We have shown that *Mortierella hyalina* is able to 1) degrade chitin into GlcNAc, a very nutritious component which stimulates growth, 2) convert chitin into chitosan, the popular MAMP responsible for increased immune response in plants and 3) partake in N-cycling. These properties suggest that *Mortierella hyalina* is a fungus that thrives in chitin-rich environments, where it can function as a biocontrol agent for horticultural crops and plays part in nitrogen cycling, which is another step towards a fertilizer- and pesticide-free future.

4. Future Perspectives

In a follow-up metabarcoding study, I recommend using more than 3 biological replicates. In this thesis we found the Timepoint and Treatment:Timepoint effect of the fungal community to be statistically insignificant at P = 0.05, with a P-value of 0.066. However, this value is quite close to the significance threshold, and thus a similar study with more biological replicates might either confirm our findings, or result in a significant change in the fungal community over time. Sun et al. (2017) have found that, compared to temporal changes in bacterial soil beta diversity, fungal soil communities change at a far lower rate. They hypothesize that this might result from higher growth rates for bacteria, a higher tolerance to environmental changes for fungi, and a greater influence of vegetation on fungal communities.

The last two hypotheses particularly strike my interest. A greater influence of vegetation on fungal communities would indeed be able to stabilize the fungal community over short periods of time, such as in our experiment where we only took measurements up to 8 weeks, and where the vegetation was only 1 lettuce plant. An interesting future study can include later timepoints, where the stabilization in fungal diversity due to vegetation might diminish over time, and different (combinations of) vegetation, which will certainly influence the fungal microbiome. As such it may be tested whether *Mortierella hyalina* remains the dominant species in the hosts rhizosphere in chitin-rich environments, or whether there are certain plant hosts whose root exudates might inhibit growth of *Mortierella*, and another fungus may emerge as the dominant species in a chitin-rich rhizosphere.

The second hypothesis from Sun et al. (2017) also presents an interesting take. If fungi have a higher tolerance to environmental changes, this might be a promising prospect for stabilizing the microbial soil community in rapidly changing climates, or intensively cultivated farmland. A study of the effect of chitin addition to (potting) soil on the fungal community in different climatic conditions might offer insights into the role chitin-degrading fungi play in stabilizing a microclimate under stress, and new plant-growth-promoting and/or plant climatic-stress-relieving properties of chitin-degrading fungi may arise.

A final proposal for future research is to delve deeper into the new *Mortierella hyalina* genome assembly presented in this thesis, and to identify molecular mechanisms and pathways that can explain the ecological role of *Mortierella hyalina* as a nitrogen cycling saprotroph.

References

Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler, (2013), 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics* 29 (8): 1072-1075.

Andrews, S. (2010). 'FastQC: A Quality Control Tool for High Throughput Sequence Data [*Online*].' Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Applied Biological Materials (2015), '1) Next Generation Sequencing (NGS) - An Introduction', https://www.youtube.com/watch?v=jFCD8Q6qSTM&t=330s

Armstrong S. & Clough J. (2009), 'Crop protection chemicals', *Royal Society of Chemistry*, Bracknell, UK, https://edu.rsc.org/feature/crop-protection-chemicals/2020121.article

Bairoch, A., & Apweiler, R. (2000). 'The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.' *Nucleic acids research*, 28(1), 45–48. https://doi.org/10.1093/nar/28.1.45

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.' *Journal of computational biology : a journal of computational molecular cell biology*, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barea J.M. et al. (2005), 'Microbial co-operation in the rhizosphere', *Journal of Experimental Botany,* Volume 56, Issue 417, July 2005, Pages 1761–1778.

Barneveldt R. (2019), 'De nieuwe "Surf en Turf": krabbenafval als groeistimulator van sla', ILVO, 9820 Merelbeke

Bergen D. (2013), 'Agro-ecologie – Een nieuwe kijk op landbouw', Beleidsdomein Landbouw en Visserij, afdeling Monitoring en Studie, Brussel.

Blum, W., Zechmeister-Boltenstern, S., & Keiblinger, K. M. (2019). 'Does Soil Contribute to the Human Gut Microbiome?' *Microorganisms*, 7(9), 287. https://doi.org/10.3390/microorganisms7090287

Bray JR.& Curtis JT. (1957), 'An ordination of upland forest communities of southern Wisconsin.' *Ecological Monographs,* 27:325-349.

Brown S. et al. (2004), 'Insensitivity to the fungicide fosetyl-aluminum in California isolates of the lettuce downy mildew pathogen, Bremia lactucae.' *Plant Dis* 88(5):502–508.

Bruce J. Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K. Young, Ashlee M. Earl (2014), 'Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.' *PLoS ONE* 9(11): e112963. doi:10.1371/journal.pone.0112963

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

Callahan B. et al (2016), 'DADA2: High resolution sample inference from Illumina amplicon data.' *Nat. Methods*, July ; 13(7): 581–583

ChunLab Inc. (2019), '16S rRNA and 16S rRNA Gene' https://help.ezbiocloud.net/16s-rrna-and-16s-rrna-gene/

Churko, J. et al. (2013), 'Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases.', *Circ Res.*, 7;112(12):1613-23.

Cook, L. (2019), 'Genetics and Biotechnology', *Ed Tech Press*, page 233

Curtis TP, Sloan WT, Scannell JW. (2002), 'Estimating prokaryotic diversity and its limits.', *Proc Natl Acad Sci* USA. 99:10494–9.

Damalas, C. A., & Koutroubas, S. D. (2016), 'Farmers' Exposure to Pesticides: Toxicity Types and Ways of Prevention', *Toxics*, 4(1), 1

Davet P., Martin C. (1993), 'Resistance of Sclerotinia minor isolates to cyclic imides in lettuce field soils of Roussillon, France.' *J Phytopathol,* 138(4):331–342

De Visscher, J. (2019), 'Effect of biochar and chitin on plant defense and rhizosphere microbiome of strawberry', *Ghent University*

Debode J. et al. (2016), 'Chitin Mixed in Potting Soil Alters Lettuce Growth, the Survival of Zoonotic Bacteria on the Leaves and Associated Rhizosphere Microbiology.' *Frontiers in Microbiology*, April Vol. 7 Article 565.

Ecobichon D.J. (2001), 'Pesticide use in developing countries', *Toxicology*, 160:27–33

Edgar R. C. (2004), 'MUSCLE: multiple sequence alignment with high accuracy and high throughput.' *Nucleic acids research,* 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Etienne, Kizee A et al. (2014), 'Draft Genome Sequence of Mortierella alpina Isolate CDC-B6842.' *Genome announcements* vol. 2,1 e01180-13. 23 Jan. 2014

European Crop Protection Association (ECPA) 2014, 'Pesticide use and food safety'.

European Food Safety Authority (EFSA) (2019), 'The 2017 European Union report on pesticide residues in food.', *EFSA journal*, vol. 17 issue 6.

Farm Services (2017), 'Lettuce', Department of Environment and Primary Industries, Melbourne, Victoria, August 2003, reviewed September 2013, http://agriculture.vic.gov.au/agriculture/horticulture/vegetables/vegetables-a-z/lettuce

Fernandez C.W. & Koide R.T. (2014), 'Initial melanin and nitrogen concentrations control the decomposition of ectomycorrhizal fungal necromass.' *Soil Biology & Biochemistry*, 77, 150-157

Fox, G. et al. (1977), 'Classification of methanogenic bacteria by 16S ribosomal RNA characterization', *Proc Natl Acad Sci USA*, 74

Gellynck X., De Pelsmaeker S., Lambrecht E. & Vandenhaute H. (2017), 'De impact van cosmetische kwaliteitseisen op Voedselverlies.' Casestudie Vlaamse sector groenten en fruit, ordered by the Department of Agriculture and Fisheries, Brussels.

Ghosh, A. et al. (2019), 'Metagenomic Analysis and its Applications', *Encyclopedia of Bioinformatics and Computational Biology*, Volume 3; 184-193

Gooday G.W. (1990), 'The ecology of chitin degradation.', *Adv Micro Ecol* 1990;11:387-419

Green E. D. (2001), 'Strategies for the systematic sequencing of complex genomes.' *Nature Reviews Genetics* 2, 580.

Gryndler, M. et al. (2003), 'Chitin stimulates development and sporulation of arbuscular mycorrhizal fungi.' *Appl. Soil Ecol.* 2003, 22, 283–287.

Handelsman J. et al. (1998), 'Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products.' *Chem Biol.* Oct;5(10):R245-9.

Hartmann, A., Schmid, M., Tuinen, D.v. et al. (2009), 'Plant-driven selection of microbes.', *Plant Soil* 321, 235–257.

Hayes T.B. et al. (2010), 'Atrazine induces complete feminization and chemical castration in male African clawed frogs (Xenopus laevis).' *PNAS*, March 1

Hossain, Md & Sultana, Farjana & Islam, Shaikhul. (2017). Plant Growth-Promoting Fungi (PGPF): Phytostimulation and Induced Systemic Resistance. 10.1007/978-981-10-6593-4_6.

Ilangumaran G. et al. (2017), 'Microbial Degradation of Lobster Shells to Extract Chitin Derivatives for Plant Disease Management' *Frontiers in Microbiology*, Volume 8, Article 781

Illumina Inc. (2013), '16S Metagenomic Sequencing Library Preparation', https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Illumina Inc. (2016), 'Illumina Sequencing by Synthesis', https://www.youtube.com/watch?time_continue=5&v=fCd6B5HRaZ8&feature=emb_logo

Illumina Inc. (2019), 'Advantages of paired-end and single-read sequencing' https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html

Illumina Inc. (2020), 'Key differences between next-generation sequencing and Sanger sequencing', https://www.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html

Janda, J. & Abbott, S. (2007), '16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls', *Journal of Clinical Microbiology*, 45 (9) 2761-2764

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Johnson J.M. et al. (2019), 'The beneficial root-colonizing fungus Mortierella hyalina promotes the aerial growth of Arabidopsis and activates calcium-dependent responses which restrict Alternaria brassicae-induced disease development in roots', *Mol Plant Microbe Interact*. Mar;32(3):351-363.

Kamerling, H. (2007), 'Comprehensive Glycoscience From Chemistry to Systems Biology', *Elsevier Science*, ISBN 978-0-444-51967-2

Khan, M.A., Elias, I., Sjölund, E. et al. (2013), 'Fastphylo: Fast tools for phylogenetics.' *BMC Bioinformatics* 14, 334. https://doi.org/10.1186/1471-2105-14-334

Kim Y. et al. (2008), 'Enzymatic deacetylation of chitin by extracellular chitin deacetylase from a newly screened Mortierella sp. DY-52.' *Journal of microbiology and biotechnology*. 18. 759-66.

Klindworth, A. et al. (2013). 'Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.*', Nucleic Acids Res* 41(1)

Koboldt DC. et al. (2013), 'The next-generation sequencing revolution and its impact on genomics.' *Cell,* 155(1), 27–38. https://doi.org/10.1016/j.cell.2013.09.006

Kojima H. et al. (2013), 'Involvement of the salicylic acid signaling pathway in the systemic resistance induced in Arabidopsis by plant growth-promoting fungus Fusarium equiseti GF19-1.' *J Oleo Sci.* 62(6):415-26.

Lamine C. et al. (2010) 'Reducing the dependence on pesticides: a matter of transitions within the whole agri-food system.', 9th European IFSA Symposium.

Larsen, A. E., Gaines, S. D., & Deschênes, O. (2017). 'Agricultural pesticide use and adverse birth outcomes in the San Joaquin Valley of California.', *Nature communications*, 8(1), 302.

Li F. et al. (2017), '*Mortierella elongata's* roles in organic agriculture and crop growth promotion in a mineral soil', *Land Degrad Dev*. 2018;29:1642–1651

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009), 'The Sequence alignment/map (SAM) format and SAMtools.' *Bioinformatics*, 25, 2078-9.

Maertens E., Deuninck J. en D'hooghe J. (2014), 'Rentabiliteits- en kostprijsanalyse sla. Resultaten van bedrijven uit het Landbouwmonitoringsnetwerk', Beleidsdomein Landbouw en Visserij, afdeling Monitoringen Studie, Brussel.

Mario Stanke, Rasmus Steinkamp, Stephan Waack and Burkhard Morgenstern (2004). "AUGUSTUS: a web server for gene finding in eukaryotes" *Nucleic Acids Research*, Vol. 32, W309-W312

McMurdie and Holmes (2013), 'Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data'. *PLoS ONE*. 8(4):e61217

Michael C.R. Alavanja, Jane A. Hoppin, Freya Kamel (2004), 'Health Effects of Chronic Pesticide Exposure: Cancer and Neurotoxicity', *Annual Review of Public Health* 25:1, 155-197

Miller, J. et al. (2017), 'Hybrid assembly with long and short reads improves discovery of gene family expansions', *BMC Genomics*; 18: 541.

Morgan N. Price, Paramvir S. Dehal, Adam P. Arkin (2009), 'FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix.' *Molecular Biology and Evolution*, Volume 26, Issue 7, July 2009, Pages 1641–1650, https://doi.org/10.1093/molbev/msp077

Mukhopadhyay S. et al. (2017), 'Changes in polycyclic aromatic hydrocarbons (PAHS) and soil biological parameters in a revegetated coal mine spoil.' *Land Degradation & Development*, 28, 1047-1055

Nannipieri P, Pietramellara G and Renella G (2014), *Omics in Soil Science*, Caister Academic Press, Norfolk, UK, 198 p.

Ng, J. L., Perrine-Walker, F., Wasson, A. P., & Mathesius, U. (2015). 'The Control of Auxin Transport in Parasitic and Symbiotic Root-Microbe Interactions.' *Plants* (Basel, Switzerland), 4(3), 606–643.

Nilsson RH. et al. (2018), 'The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications.' *Nucleic Acids Research.* DOI: 10.1093/nar/gky1022

O'Connor, C. M. & Adams, J. U. *Essentials of Cell Biology*. Cambridge, MA: NPG Education, 2010.

Oerke E.C. et al. (2006), 'Crop losses to pests', *J. Agricult. Sci.,* 144, 31-43.

Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. (2016), 'Mash: fast genome and metagenome distance estimation using MinHash' *Genome Biol*. Jun 20;17(1):132. doi: 10.1186/s13059-016-0997-x.

Platteau J., Lambrechts G., Roels K., Van Bogaert T., Luypaert G. & Merckaert B. (eds.) (2019), 'Challenges for Flemish agriculture and horticulture, Agriculture Report 2018, Summary', Department of Agriculture and Fisheries, Brussels.

Quast C. et al. (2013), 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.' *Nucleic Acids Research.* 41 (D1): D590-D596.

Rahman M, Mukta JA, Sabir AA, Gupta DR, Mohi-Ud-Din M, Hasanuzzaman M, et al. (2018), 'Chitosan biopolymer promotes yield and stimulates accumulation of antioxidants in strawberry fruit.' *PLoS ONE* 13(9): e0203769. https://doi.org/10.1371/journal.pone.0203769

Ramirez M.A. et al. (2010), 'Chitin and its derivates as biopolymers with potential agricultural applications.' *Biotecnologia Aplicada* 2010; Vol.27, No.4

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010), 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.' *Bioinformatics (Oxford, England),* 26(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Salachna, Piotr & Zawadzińska, Agnieszka. (2014). 'Effect of chitosan on plant growth, flowering and corms yield of potted freesia.' *Journal of Ecological Engineering*. 10.12911/22998993.1110223.

Sambo P. et al (2019). 'Hydroponic Solutions for Soilless Production Systems: Issues and Opportunities in a Smart Agriculture Perspective' *Front Plant Sci.,* 10: 923

Schindler D.W. et al. (2004), 'Over fertilization of the World's Freshwaters and Estuaries', *University of Alberta Press*, p. 1.

Schuster, S. (2008), 'Next-generation sequencing transforms today's biology.', *Nat Methods*, 5(1):16-8

Seemann T. (2018), 'Shovill: Assemble bacterial isolate genomes from Illumina paired-end reads.' *GitHub,* https://github.com/tseemann/shovill.

Shan Sun, Song Li, Bethany N. Avera, Brian D. Strahm, Brian D. Badgley (2017), 'Soil Bacterial and Fungal Communities Show Distinct Recovery Patterns during Forest Ecosystem Restoration', *Applied and Environmental Microbiology*, 83 (14) e00966-17; DOI: 10.1128/AEM.00966-17

Shannon CE. (1948), 'A Mathematical Theory of Communication', *The Bell System Technical Journal*, 37(3): 379-423.

Sharp R.G. (2013), 'A Review of the Applications of Chitin and Its Derivatives in Agriculture to Modify Plant-Microbial Interactions and Improve Crop Yields' *Agronomy* 2013, 3, 757-793

Shaw, Arthur & McDaniel, Stuart & Werner, Olaf & Ros-Espín, Rosa. (2002). "Invited essay: New frontiers in bryology and lichenology. Phylogeography and phylodemography." *Bryologist.* 105. 373-383.

Shen Z.Z. et al. (2014), 'Banana *Fusarium wilt* disease incidence is influenced by shifts of soil microbial communities under different monoculture spans.' *Microbial Ecology*, 1-12

Shendure, J. et al. (2008), 'Next-generation DNA sequencing.', *Nat Biotechnol*., 26(10):1135-45.

Song, L., Florea, L. & Langmead, B. (2014). 'Lighter: fast and memory-efficient sequencing error correction without counting.' *Genome Biol* 15, 509. https://doi.org/10.1186/s13059-014-0509-9

Spaepen S. et al. (2007), 'Indole-3-acetic acid in microbial and microorganism-plant signaling.', *FEMS Microbiol Rev,* 31(4):425-48.

Spanu P.D. et al. (2010), 'Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism.' *Science*. 2010;330:1543–6.

Staley J.T., Konopka A. (1985), 'Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats.' *Annu Rev Microbiol* 39:321-46.

T. Magoc and S. Salzberg (2011), 'FLASH: Fast length adjustment of short reads to improve genome assemblies', *Bioinformatics* 27:21 2957-63.

Tagawa M. et al. (2010), 'Isolation and characterization of antagonistic fungi against potato scab pathogens from potato field soils', *FEMS Microbiology Letters* 305(2):136-42

The UniProt Consortium (2019), 'UniProt: a worldwide hub of protein knowledge' *Nucleic Acids Res.* 47: D506-515

Turner S. et al. (2017), 'Microbial Community Dynamics in Soil Depth Profiles Over 120,000 Years of Ecosystem Development', *Front. Microbiol.* (8);874.

Uehling J. et al. (2017), 'Comparative genomics of Mortierella elongata and its bacterial endosymbiont Mycoavidus cysteinexigens.' *Environ Microbiol*. 2017 Aug;19(8):2964-2983.

Venter, C. et al. (2001), 'The Sequence of the Human Genome', *Science*, Vol. 291, Issue 5507, pp. 1304-1351

Větrovský T, Baldrian P (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS ONE 8(2): e57923.

Virginie Barrière, François Lecompte, Philippe Nicot, Brigitte Maisonneuve, Marc Tchamitchian, et al. (2014), 'Lettuce cropping with less pesticides. A review.', *Agron. Sustain. Dev.*, 34:175–198.

Xiong W. et al. (2014), 'Distinct roles for soil fungal and bacterial communities associated with the suppression of vanilla *Fusarium wilt* disease.' *Soil Biology & Biochemistry*, 107, 198-207

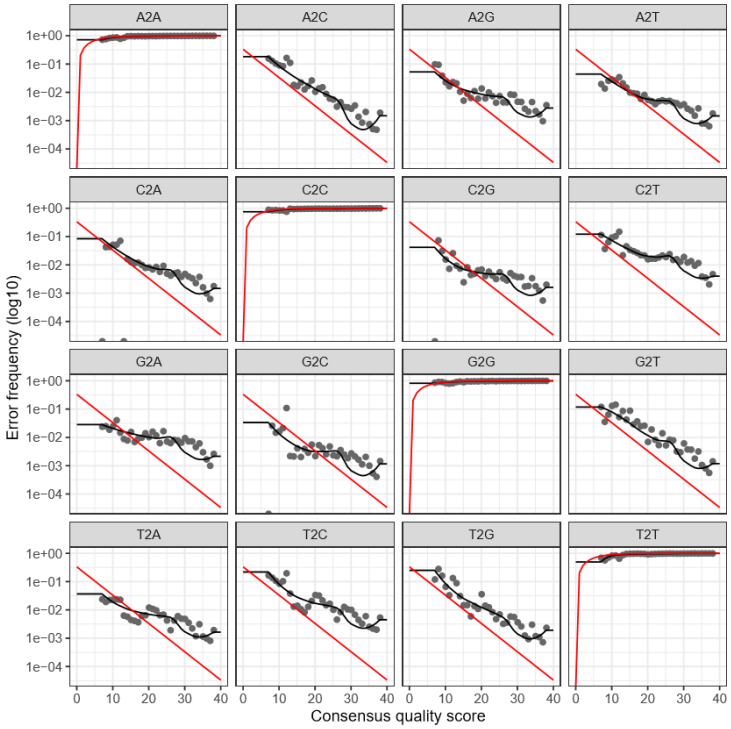
Yao, H. et al. (2010), 'Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals.*'* *PLoS ONE* 5(10)

Yara Vlaandingen B.V. (2019), 'Pure Nutrient Info #2 – Voorkomen van nitraatuitspoeling', Vlaardingen, Nederland, https://www.yara.nl/gewasvoeding/pure-nutrient/info-02-uitspoeling/uitspoeling/

Yasuda M. et al. (2008), 'Antagonistic interaction between systemic acquired resistance and the abscisic acid-mediated abiotic stress response in Arabidopsis.' *Plant Cell,* 20(6):1678-92

Zegeye EK. et al. (2019), 'Selection, succession, and stabilization of soil microbial consortia.' *mSystems* 4:e00055-19.https://doi.org/10 .1128/mSystems.00055-19.

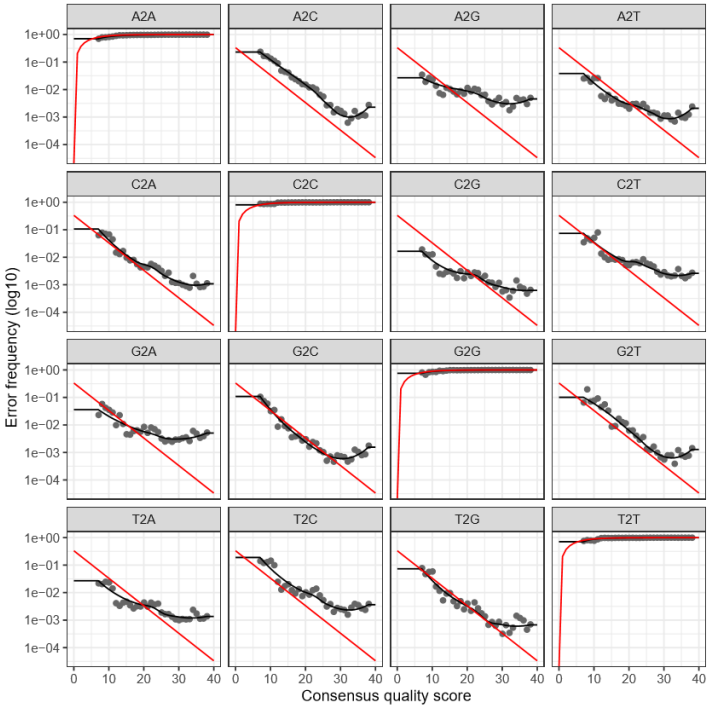
Supplementary Materials



Supplementary Figure 1: Error profile, bacteria, forward reads

Red line: expected errors

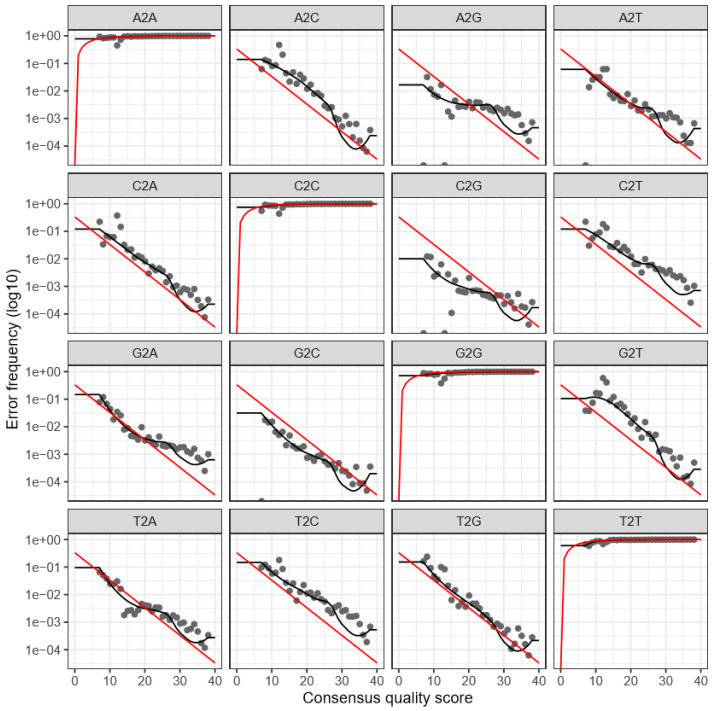
Black dots & line: true errors



Supplementary Figure 2: Error profile, bacteria, reverse reads

Red line: expected errors

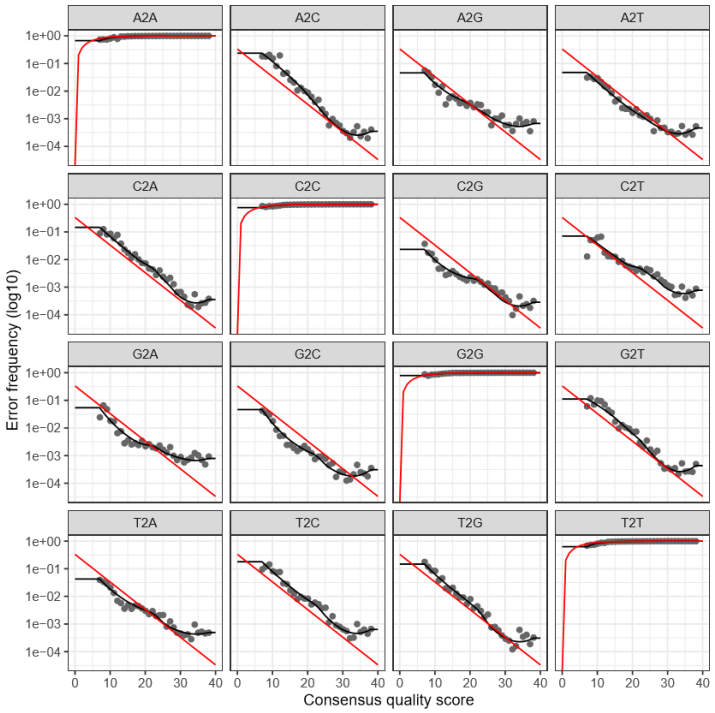
Black dots & line: true errors



Supplementary Figure 3: Error profile, fungi, forward reads

Red line: expected errors

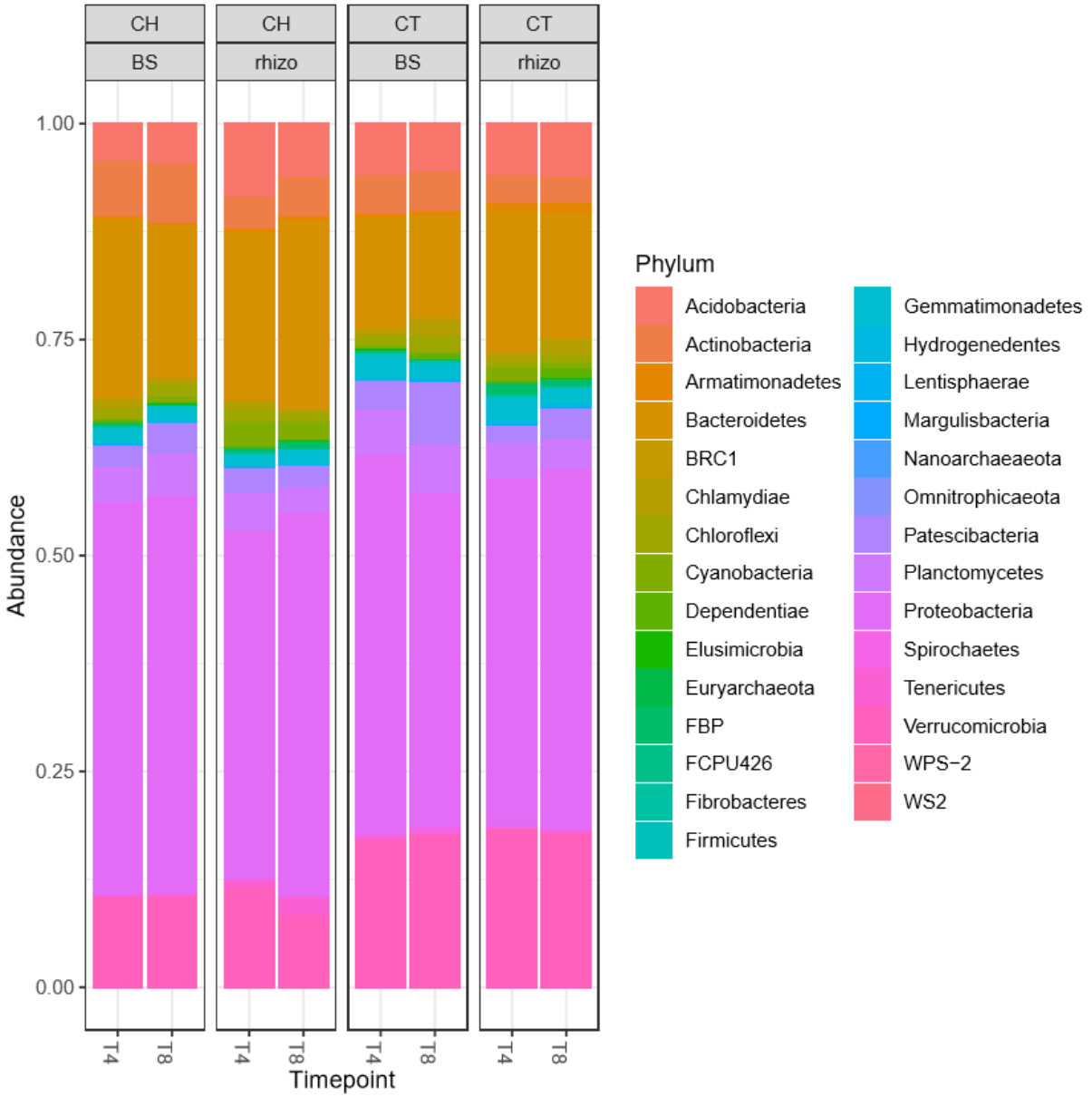
Black dots & line: true errors



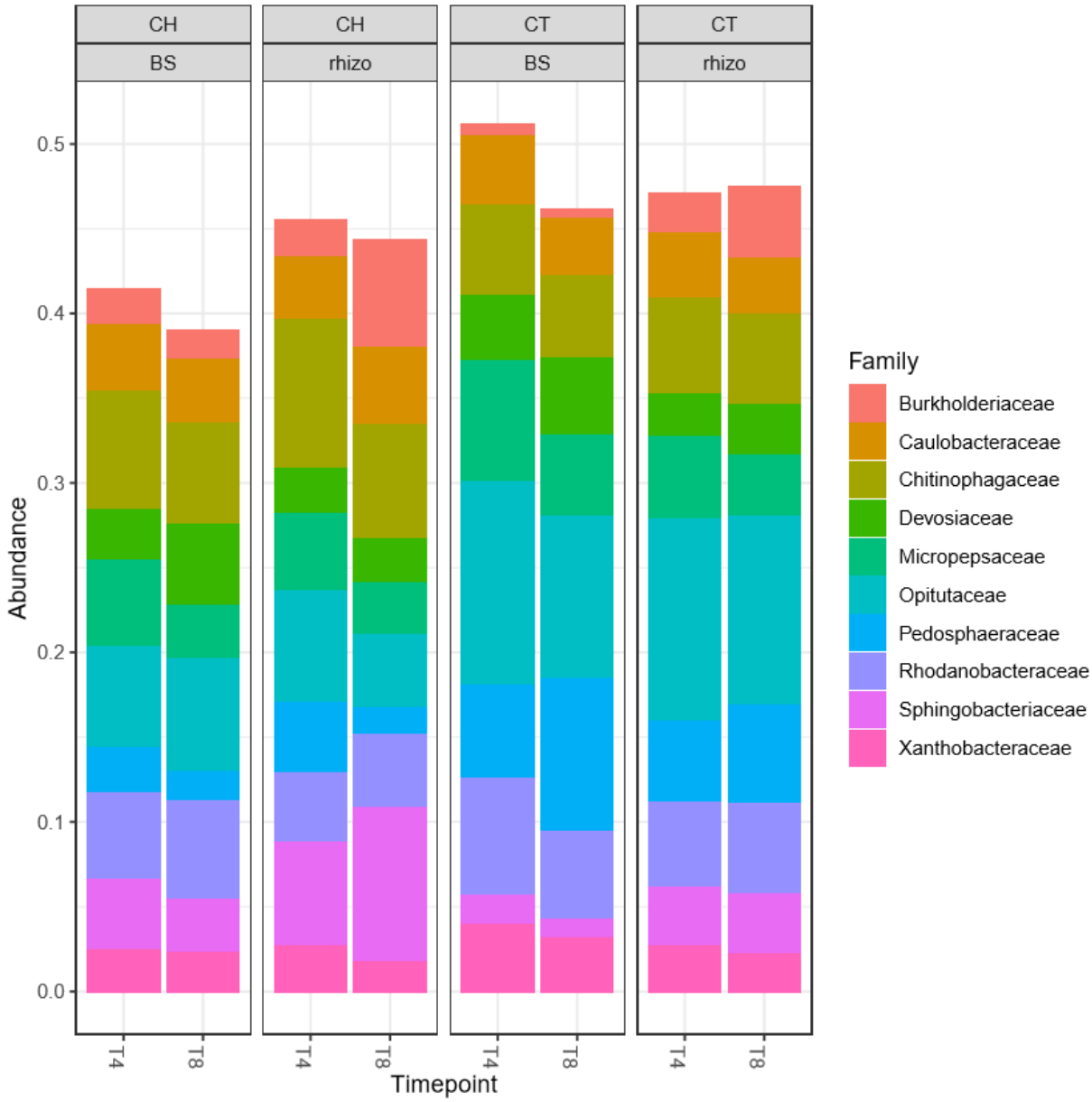
Supplementary Figure 4: Error profile, fungi, reverse reads

Red line: expected errors

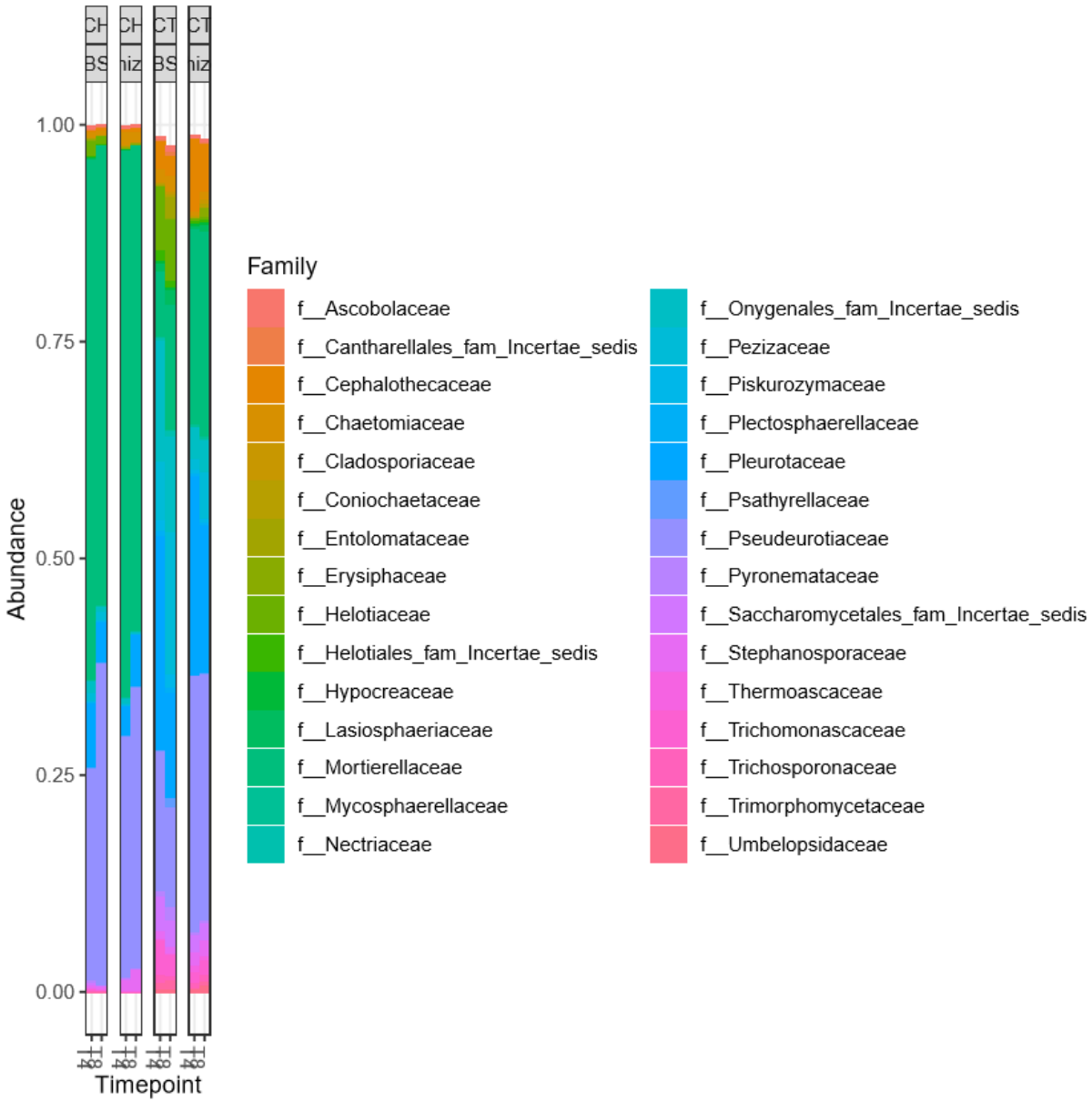
Black dots & line: true errors



Supplementary Figure 5: Top 30 phyla in the bacterial samples



Supplementary Figure 6: Top 10 families in the bacterial samples



Supplementary Figure 7: Top 30 families in the fungal samples

Phylum	Family	T4_CT_BS_mean	T4_CT_BS_stdDev	T4_CH_BS_mean	T4_CH_BS_stdDev	T8_CT_BS_mean	T8_CT_BS_stdDev	T8_CH_BS_mean	T8_CH_BS_stdDev	T4_CT_rhizo_mean	T4_CT_rhizo_stdDev	T4_CH_rhizo_mean	T4_CH_stdDev	T8_CT_rhizo_mean	T8_CT_rhizo_stdDev	T8_CH_rhizo_mean	T8_CH_rhizo_stdDev
Acidobacteria	Blastocatellaceae	0.00762	0.00267	0.01365	0.00086	0.01842	0.00116	0.01798	0.00174	0.0176	0.000888	0.01929	0.00445	0.00697	0.000089	0.03045	0.00301
Acidobacteria	Solibacteraceae_.Subgroup_3.	0.01997	0.00347	0.0089	0.00063	0.02173	0.00244	0.00809	0.0011	0.01669	0.001548	0.01387	0.00389	0.01884	0.001754	0.00733	0.00103
Acidobacteria	Thermoanaerobaculaceae	0.00117	0.00121	0.00215	0.00119	0.00084	0.00089	0.00101	0.00084	0.00022	0.000243	0.00224	0.00116	0.00316	0.003412	0.00091	0.00013
Acidobacteria	Holophagaceae	0.00171	0.00066	0.00105	0.00016	0.00055	0.00026	0.00089	0.00032	0.00159	0.000254	0.00064	0.00009	0.00064	0.000177	0.00073	0.00012
Acidobacteria	Acidobacteriaceae_.Subgroup_1.			0.0038	0.00048	0.00171	0.00014	0.00189	0.00042	0.00047	0.00051	0.00327	0.000938	0.00169	0.00059	0.00222	0.000048	0.00118	0.00037
Acidobacteria	Koribacteraceae	0.00048	0.00018	0.00041	0.00041	0.00023	0.00024	0.00024	0.00022	0.00038	0.000154	0.00087	0.0001	0.00019	0.000274	0.0001	0.00017
Actinobacteria	Streptomycetaceae	0.0001	0.00017	0.01376	0.00864	0	0	0.01037	0.00223	0.00023	0.000246	0.00959	0.00335	0.00019	0.000274	0.01142	0.0013
Actinobacteria	Microbacteriaceae	0.00816	0.00182	0.0146	0.0027	0.00442	0.00056	0.01728	0.00104	0.00569	0.001756	0.00691	0.00249	0.00784	0.001051	0.00871	0.00091
Actinobacteria	Nocardioidaceae	0.00509	0.00022	0.00739	0.0007	0.00501	0.00228	0.00909	0.00155	0.0053	0.000868	0.00562	0.00081	0.00497	0.000059	0.0106	0.00025
Actinobacteria	Cellulomonadaceae	0.00712	0.00127	0.0047	0.00061	0.00355	0.00016	0.00342	0.00035	0.00297	0.000917	0.00142	0.00086	0.00903	0.001634	0.00159	0.00013
Actinobacteria	Iamiaceae	0.00436	0.00102	0.00497	0.00146	0.00588	0.00187	0.00816	0.00127	0.00693	0.002259	0.00301	0.001	0.00855	0.001229	0.00615	0.00058
Actinobacteria	Sporichthyaceae	0.00124	0.00012	0.00041	0.00026	0.00301	0.00069	0.00015	0.00026	0.00223	0.000556	0.00075	0.00016	0.00118	0.000215	0.00143	0.00059
Actinobacteria	Micrococcaceae	0.00188	0.00089	0.00158	0.0005	0.0007	0.00023	0.00262	0.00201	0.00106	0.000631	0.00125	0.00088	0.00097	0.000242	0.00084	0.00043
Actinobacteria	Microtrichaceae	0.00146	0.00035	0.00145	0.00049	0.00064	0.00003	0.00168	0.00013	0.00065	0.000275	0.00029	0.00007	0.0018	0.000575	0.00096	0.00025
Actinobacteria	Frankiaceae	0.00109	0.00027	0	0	0.00091	0.00033	0.00039	0.00067	0.00117	0.000097	0.0002	0.00035	0.00074	0.000347	0.00023	0.00041
Actinobacteria	Mycobacteriaceae	0.00155	0.00057	0.00216	0.00044	0.00133	0.00047	0.00211	0.00026	0.00185	0.000152	0.00064	0.0004	0.0019	0.000114	0.00114	0.00013
Actinobacteria	Solirubrobacteraceae	0.00048	0.00016	0.00128	0.00035	0.00016	0.00016	0.00131	0.00021	0.00006	0.000103	0.00037	0.00026	0.00069	0.000636	0.00076	0.00026

Actinobacteria	Ilumatobacteraceae	0.00055	0.00032	0.00067	0.00059	0.00061	0.00022	0.00101	0.00021	0.00069	0.00026	0.00059
0.00029	0.00082	0.000041	0.0012	0.00021								
Actinobacteria	Intrasporangiaceae	0.00026	0.00025	0.00029	0.00026	0	0	0.00052	0.0009	0	0	0
0	0.00014	0.000201	0.00019	0.00017								
Actinobacteria	Micromonosporaceae	0.00017	0.00015	0.00061	0.00019	0	0	0.00119	0.00153	0.00002	0.000034	0.00042
0.00019	0.00017	0.000028	0.00006	0.00011								
Actinobacteria	Brevibacteriaceae	0	0	0	0	0.00059	0.0003	0	0	0.00002	0.000042	0
0	0	0	0.00041	0.00004								
Actinobacteria	Demequinaceae	0.00006	0.0001	0.00006	0.0001	0.00006	0.0001	0.00014	0.00025	0.00019	0.000186	0.00008
0.00013	0.00027	0.000104	0	0								
Actinobacteria	Pseudonocardiaceae	0	0	0	0	0.00012	0.00011	0.0002	0.00035	0	0	0.00004
0.00007	0	0	0.00004	0.00007								
Actinobacteria	Acidothermaceae	0.00022	0.00007	0.0002	0.00017	0	0	0.00051	0.00029	0.00002	0.000033	0
0	0	0	0	0	0							
Actinobacteria	Gaiellaceae	0.00006	0.0001	0	0	0	0	0.00008	0.00014	0	0	0
0	0.00006	0.000091	0	0								
Armatimonadetes	Fimbriimonadaceae	0.00409	0.00186	0.00276	0.00082	0.01224	0.00292	0.00296	0.00088	0.00945	0.004714	0.00498
0.00098	0.00443	0.002465	0.00549	0.00102								
Armatimonadetes	Chthonomonadaceae	0	0	0	0	0.00006	0.00005	0.00002	0.00003	0.00007	0.000063	0
0	0.00004	0.000062	0	0								
Bacteroidetes	Microscillaceae	0.02231	0.00306	0.02487	0.00607	0.0201	0.00322	0.03121	0.00554	0.02561	0.004617	0.01963
0.00727	0.02965	0.002035	0.02111	0.00341								
Bacteroidetes	Sphingobacteriaceae	0.01763	0.00243	0.04142	0.00201	0.03587	0.01021	0.03106	0.00428	0.03431	0.009573	0.06024
0.01979	0.01077	0.000061	0.09118	0.00891								
Bacteroidetes	Chitinophagaceae	0.05325	0.00483	0.07026	0.0077	0.05373	0.00891	0.05965	0.00522	0.0566	0.003064	0.08816
0.00498	0.04893	0.011788	0.06745	0.00092								
Bacteroidetes	NS9_marine_group	0.00795	0.00068	0.01296	0.00115	0.00459	0.00076	0.00813	0.00298	0.00942	0.001227	0.00892
0.00065	0.00446	0.001683	0.0064	0.00186								
Bacteroidetes	Flavobacteriaceae	0.01057	0.00175	0.03452	0.01174	0.01829	0.00554	0.02318	0.00178	0.022	0.005156	0.02124
0.0044	0.01536	0.003111	0.02151	0.00692								
Bacteroidetes	Spirosomaceae	0.00034	0.00039	0.00407	0.00064	0.00533	0.00141	0.00519	0.00204	0.00459	0.002828	0.0075
0.00453	0.00173	0.000407	0.01728	0.00317								
Bacteroidetes	Cytophagaceae	0.00043	0.00025	0.00681	0.00313	0.00005	0.00005	0.01129	0.01217	0.00026	0.000067	0.00132
0.00059	0.00046	0.000086	0.00066	0.00024								
Bacteroidetes	env.OPS_17	0.0097	0.00188	0.01327	0.00144	0.00454	0.00119	0.00502	0.00083	0.0072	0.000502	0.00609
0.00218	0.01014	0.000175	0.00371	0.00074								
Bacteroidetes	Crocinitomicaceae	0.00632	0.00136	0.00846	0.00145	0.00247	0.00153	0.008	0.00457	0.00399	0.00113	0.00412
0.00123	0.00553	0.000976	0.00486	0.00097								
Bacteroidetes	Weeksellaceae	0.00002	0.00004	0.00115	0.00114	0.0047	0.0043	0.00324	0.00203	0.00109	0.000528	0.00224
0.00106	0.0007	0.000528	0.00435	0.00017								

Bacteroidetes	Paludibacteraceae	0.0039	0.00237	0.00076	0.00022	0.00132	0.00021	0.00082	0.00022	0.00164	0.000295	0.00038
0.00036	0.00421	0.000361	0.00011	0.00012								
Bacteroidetes	AKYH767	0.0005	0.00049	0.00019	0.00021	0.00042	0.00022	0.00043	0.00033	0.00016	0.000139	0.00079
0.00065	0.00076	0.000205	0.00043	0.00032								
Bacteroidetes	Saprospiraceae	0.00041	0.00049	0.0001	0.00011	0.00092	0.00061	0.00008	0.00007	0.00018	0.000161	0.00112
0.00119	0.00092	0.00031	0.00056	0.0005								
Bacteroidetes	BSV26	0.00067	0.00026	0.00013	0.00012	0.00049	0.00029	0.00006	0.00005	0.0008	0.000222	0.00006
0.00006	0.00032	0.000234	0.0002	0.00007								
Bacteroidetes	NS11.12_marine_group		0.00098	0.00022	0.00184	0.00018	0.00056	0.00005	0.00098	0.00013	0.00085	0.000093
0.00085	0.00094	0.00085	0.000112	0.00065	0.00019							
Bacteroidetes	FFCH9454	0.00014	0.00024	0	0	0.00045	0.00059	0	0	0.00049	0.000841	0
0	0.00018	0.000248	0	0								
Bacteroidetes	KD3.93	0.00099	0.00057	0	0	0.00039	0.00026	0.00026	0.00013	0.00045	0.00033	0.00003
0.00005	0.00136	0.000278	0	0								
Bacteroidetes	Bacteroidetes_vadinHA17		0.00047	0.00021	0.00038	0.0001	0.00039	0.00006	0.00033	0.0002	0.00027	0.00024
0.00006	0.00005	0.00115	0.000628	0.00008	0.00007							
Bacteroidetes	Cyclobacteriaceae	0	0	0	0	0.00006	0.00007	0	0	0	0	0.00033
0.00003	0.00009	0.000122	0.00011	0.00002								
Bacteroidetes	Prolixibacteraceae	0.00002	0.00004	0.00008	0.00007	0.00002	0.00004	0	0	0.00002	0.000043	0.00003
0.00006	0	0	0.00002	0.00004								
Bacteroidetes	LD.RB.34	0.00006	0.00011	0.00004	0.00007	0.00007	0.00007	0	0	0	0	0
0	0	0	0	0								
Bacteroidetes	Cryomorphaceae	0.00002	0.00003	0.00002	0.00004	0	0	0.00002	0.00004	0.00003	0.000055	0
0	0	0	0	0								
Chlamydiae	cvE6	0.0034	0.00153	0.00669	0.0053	0.01619	0.00847	0.00458	0.00345	0.00795	0.003493	0.00613
0.00166	0.017	0.004635	0.00425	0.00104								
Chlamydiae	Simkaniaceae	0.00088	0.00031	0.00221	0.00174	0.00111	0.00027	0.00155	0.00079	0.00076	0.000342	0.00183
0.00113	0.00144	0.000918	0.00134	0.00075								
Chlamydiae	Parachlamydiaceae	0.00076	0.00025	0.00066	0.00049	0.00064	0.00028	0.00067	0.0001	0.00093	0.001041	0.00029
0.0001	0.00142	0.000304	0.00013	0.00012								
Chloroflexi	A4b	0.00197	0.00069	0.00212	0.00042	0.00345	0.00062	0.00225	0.00118	0.00171	0.001138	0.00401
0.00153	0.00566	0.00206	0.00132	0.00032								
Chloroflexi	Caldilineaceae		0.00109	0.00034	0.00108	0.00043	0.00048	0.00018	0.00133	0.00088	0.00057	0.000141
0.00081	0.00034	0.00083	0.000373	0.0008	0.00023							
Chloroflexi	JG30.KF.AS9		0.00106	0.00039	0.00147	0.00071	0.00038	0.00022	0.00149	0.00138	0.00056	0.000289
0.00293	0.00109	0.00131	0.00161	0.00035	0.00023							
Chloroflexi	Anaerolineaceae		0.00052	0.00028	0.00031	0.00009	0.00044	0.00014	0.00047	0.00023	0.00028	0.000055
0.00031	0.00015	0.0012	0.000031	0.00031	0.00006							
Chloroflexi	JG30.KF.CM45		0.00043	0.00023	0.00098	0.00054	0.00076	0.00021	0.00206	0.00044	0.00024	0.000221
0.00059	0.00026	0.00099	0.000187	0.00105	0.00014							

Chloroflexi	Roseiflexaceae	0.00006	0.00005	0.00014	0.00013	0.00005	0.00008	0.00008	0.00015	0	0	
0.00007	0.00003	0.00044	0.000071	0.00003	0.00005							
Chloroflexi	Ktedonobacteraceae	0.00015	0.00007	0.00015	0.0001	0.00002	0.00004	0.00049	0.00015	0	0	
0	0	0.00016	0.000228	0.00003	0.00005							
Chloroflexi	Thermomicrobiaceae	0.00005	0.00009	0.00012	0.00011	0	0	0	0	0	0	
0.00006	0.0001	0.0001	0.000139	0.00004	0.00007							
Cyanobacteria	Phormidiaceae	0.00006	0.0001	0.00026	0.00044	0	0	0.00008	0.00008	0		
0	0	0	0	0	0	0						
Cyanobacteria	Nodosilineaceae	0.00004	0.00006	0.00007	0.00011	0	0	0.00002	0.00004	0		
0	0	0	0	0	0	0						
Dependentiae	Vermiphilaceae	0.00049	0.00023	0.00091	0.00085	0.00989	0.00967	0.0006	0.00038	0.00044		
0.000398	0.00038	0.00021	0.00477	0.004679	0.00107	0.00089						
Dependentiae	UBA12409	0.00114	0.00193	0.00264	0.00282	0.00223	0.00173	0.00113	0.00062	0.00182		
0.002928	0.00356	0.00435	0.0021	0.001116	0.00093	0.00021						
Euryarchaeota	Methanosarcinaceae	0.0001	0.00011	0	0	0	0	0	0	0		
0	0	0	0.00006	0.000091	0	0						
Fibrobacteres	Fibrobacteraceae	0.00214	0.0006	0.0007	0.00009	0.00175	0.00109	0.00025	0.00016	0.00254		
0.000962	0.00021	0.00009	0.00294	0.000193	0.0001	0.00009						
Firmicutes	Paenibacillaceae	0.00047	0.00053	0.00083	0.00061	0.0007	0.00018	0.00016	0.00017	0.00066		
0.000375	0.00139	0.00092	0.00111	0.000444	0.00013	0.00017						
Firmicutes	Heliobacteriaceae	0.00006	0.00011	0	0	0.00012	0.00011	0	0	0.00006		
0.000099	0	0	0	0	0	0						
Gemmatimonadetes	Gemmatimonadaceae		0.02653	0.00429	0.01445	0.00049	0.01586	0.00328	0.0166	0.00241		
0.02772	0.005954	0.01447	0.00142	0.01641	0.000573	0.01715	0.00214					
Gemmatimonadetes	Longimicrobiaceae	0.00432	0.00138	0.00288	0.00031	0.00354	0.00054	0.00157	0.0003	0.00215		
0.000423	0.00111	0.0001	0.00294	0.000329	0.00132	0.00056						
Hydrogenedentes	Hydrogenedensaceae	0.00272	0.00049	0.0028	0.00094	0.00542	0.00132	0.0031	0.0009	0.00491		
0.000765	0.00191	0.00022	0.00417	0.000851	0.00362	0.00066						
Patescibacteria	Saccharimonadaceae	0.00092	0.00027	0.00168	0.00027	0.00034	0.00025	0.00328	0.00216	0.00042		
0.000139	0.00143	0.00032	0.00035	0.000088	0.0016	0.00035						
Planctomycetes	Pirellulaceae	0.01741	0.00205	0.02345	0.00076	0.0152	0.00256	0.03253	0.00269	0.01143		
0.001172	0.01767	0.00399	0.02926	0.001867	0.01778	0.00236						
Planctomycetes	WD2101_soil_group	0.0171	0.00267	0.00773	0.00104	0.00916	0.00118	0.01166	0.00127	0.01501		
0.001917	0.01517	0.00548	0.0157	0.002945	0.00795	0.00099						
Planctomycetes	CPla.3_termite_group		0.01365	0.00219	0.00447	0.00014	0.00606	0.00015	0.00147	0.00041		
0.01094	0.002812	0.00603	0.00206	0.00784	0.000652	0.00168	0.00031					
Planctomycetes	Schlesneriaceae	0.0013	0.00044	0.00137	0.00018	0.00145	0.00054	0.00145	0.00038	0.00174		
0.000362	0.00148	0.00021	0.00114	0.000153	0.00201	0.00048						
Planctomycetes	Rubinisphaeraceae	0.00258	0.00026	0.00316	0.00056	0.00094	0.00038	0.00356	0.0014	0.00192		
0.000322	0.00218	0.00068	0.00156	0.000195	0.00198	0.00039						

Planctomycetes	Phycisphaeraceae	0.00088	0.00071	0.00031	0.00025	0.0016	0.00031	0.00042	0.00033	0.00171	0.000783	0.00242
0.00093	0.00127	0.000521	0.00074	0.00022								
Planctomycetes	Gemmataceae	0.00053	0.00043	0.00043	0.00051	0.00084	0.00035	0.00088	0.00084	0.00038	0.000198	0.00153
0.00059	0.00105	0.000585	0.00053	0.00032								
Planctomycetes	Isosphaeraceae	0.00026	0.00009	0.00059	0.00011	0.00025	0.00004	0.00026	0.00032	0.0004	0.00008	0.00024
0.00021	0.00012	0.00017	0.00021	0.00019								
Planctomycetes	Tepidisphaeraceae	0.0001	0.00009	0	0	0.00006	0.0001	0.00025	0.00007	0.00015	0.000004	0.00007
0.00007	0	0	0.00011	0.00011								
Proteobacteria	Rhodanobacteraceae	0.06913	0.01268	0.05151	0.00459	0.053	0.00836	0.05811	0.00945	0.05021	0.007274	0.04098
0.00626	0.05168	0.007363	0.04281	0.00349								
Proteobacteria	Devosiaceae	0.03864	0.00406	0.03	0.00235	0.02942	0.00135	0.04785	0.00807	0.02489	0.001677	0.02641
0.00682	0.04503	0.001433	0.02593	0.00157								
Proteobacteria	Chitinibacteraceae	0	0	0.04547	0.00419	0	0	0.01061	0.00972	0.00052	0.000901	0.04888
0.0388	0	0	0.00666	0.00478								
Proteobacteria	Dongiaceae	0.02876	0.00228	0.02697	0.00322	0.01978	0.00073	0.02348	0.00258	0.01861	0.002513	0.01981
0.00278	0.02451	0.005963	0.01567	0.00085								
Proteobacteria	Caulobacteraceae	0.04019	0.00396	0.03891	0.00307	0.03306	0.00238	0.03775	0.00324	0.03821	0.000595	0.03666
0.00348	0.03366	0.005599	0.04534	0.00582								
Proteobacteria	Pseudomonadaceae	0.00009	0.00015	0.0107	0.00992	0.02472	0.01803	0.00802	0.00661	0.02006	0.00082	0.00733
0.0056	0.00044	0.000163	0.04289	0.01673								
Proteobacteria	Solimonadaceae	0.00642	0.00362	0.0079	0.0023	0.01298	0.00368	0.00728	0.00116	0.0132	0.006117	0.00933
0.00233	0.00598	0.001074	0.0135	0.00554								
Proteobacteria	Burkholderiaceae	0.00587	0.00234	0.01965	0.01254	0.04024	0.03668	0.01636	0.00232	0.02196	0.002659	0.02043
0.00529	0.00473	0.000965	0.06258	0.01067								
Proteobacteria	Micropepsaceae	0.07165	0.0102	0.05042	0.00729	0.03602	0.0043	0.03139	0.00724	0.04913	0.004827	0.04566
0.00358	0.04817	0.002163	0.03079	0.00195								
Proteobacteria	Phaselicystidaceae	0.01099	0.00245	0.00724	0.00039	0.00474	0.00086	0.00999	0.00094	0.01087	0.001442	0.00694
0.00074	0.00893	0.000055	0.0066	0.0015								
Proteobacteria	Hyphomonadaceae	0.01562	0.00705	0.0106	0.00065	0.00707	0.00133	0.01228	0.00225	0.01001	0.001461	0.0144
0.00528	0.0163	0.006574	0.00924	0.00088								
Proteobacteria	Xanthobacteraceae	0.04021	0.00171	0.02572	0.00474	0.02304	0.00261	0.02424	0.00269	0.02821	0.003209	0.02833
0.00665	0.03276	0.003633	0.01866	0.00286								
Proteobacteria	BIrii41	0.02021	0.00359	0.01617	0.00052	0.00575	0.00235	0.02307	0.00495	0.00566		
0.001618	0.00784	0.00298	0.02532	0.000414	0.00514	0.00114						
Proteobacteria	Sphingomonadaceae	0.00895	0.00078	0.01076	0.00044	0.02554	0.00161	0.01621	0.00212	0.0181	0.001062	0.01944
0.00144	0.01254	0.000024	0.02805	0.00259								
Proteobacteria	Cellvibrionaceae	0.00035	0.00022	0.01218	0.01461	0.0001	0.00009	0.04372	0.02937	0.00028	0.000104	0.0069
0.00194	0.00015	0.000033	0.01365	0.01818								
Proteobacteria	Rhizobiaceae	0.0057	0.00125	0.01866	0.00494	0.00604	0.00088	0.02806	0.00202	0.00485	0.001232	0.01215
0.00174	0.00491	0.000628	0.01887	0.00119								

Proteobacteria	Sneathiellaceae	0.00389	0.00051	0.00458	0.00025	0.00313	0.0003	0.00445	0.00064	0.00385	0.000105	0.00429
0.00065	0.00331	0.000878	0.00617	0.0009								
Proteobacteria	SC.I.84	0.00626	0.00071	0.00417	0.00051	0.00472	0.00044	0.00484	0.0013	0.00682	0.00223	0.00541
0.00054	0.00383	0.000617	0.00535	0.00106								
Proteobacteria	Reyranellaceae	0.00656	0.00081	0.00446	0.00058	0.01367	0.0021	0.00497	0.00059	0.01167	0.003354	0.00663
0.00065	0.00692	0.001185	0.00997	0.00124								
Proteobacteria	Hyphomicrobiaceae	0.00288	0.00017	0.00315	0.00109	0.00144	0.00081	0.0051	0.00105	0.00234	0.000319	0.00359
0.00123	0.00195	0.000804	0.00273	0.00037								
Proteobacteria	Acetobacteraceae	0.0024	0.00054	0.00261	0.00023	0.00509	0.00061	0.0023	0.00037	0.00459	0.000939	0.00372
0.001	0.00291	0.000024	0.0041	0.00066								
Proteobacteria	Bdellovibrionaceae	0.01545	0.00265	0.01402	0.00226	0.00907	0.00165	0.00993	0.00032	0.01164	0.002805	0.00767
0.00096	0.01479	0.001999	0.0062	0.00142								
Proteobacteria	Rhizobiales_Incertae_Sedis	0.01484	0.00493	0.00962	0.002	0.0108	0.00149	0.01007	0.00359	0.01077	0.001871	0.01689
0.00899	0.01579	0.005859	0.00813	0.00182								
Proteobacteria	Beijerinckiaceae	0.00505	0.0002	0.0042	0.00064	0.0021	0.00136	0.00545	0.00204	0.00296	0.000884	0.00184
0.00093	0.00275	0.000173	0.00277	0.00047								
Proteobacteria	Diplorickettsiaceae	0.00184	0.00028	0.00442	0.0014	0.01717	0.00217	0.00243	0.00076	0.00774	0.00251	0.00459
0.00129	0.00607	0.002348	0.00904	0.00331								
Proteobacteria	TRA3.20	0.00158	0.00092	0.00105	0.00003	0.00396	0.00086	0.00197	0.00062	0.00205	0.000467	0.00279
0.00054	0.00251	0.000551	0.00505	0.00052								
Proteobacteria	P3OB.42	0.00241	0.00063	0.00082	0.00009	0.0014	0.00035	0.0008	0.00067	0.00253	0.00021	0.0005
0.00029	0.00198	0.000639	0.00038	0.0001								
Proteobacteria	Nitrosomonadaceae	0.00639	0.002	0.00754	0.00167	0.00582	0.00098	0.00787	0.00187	0.00435	0.000571	0.00968
0.00172	0.00746	0.000089	0.0088	0.00098								
Proteobacteria	Methylophilaceae	0.00017	0.00004	0	0	0.00731	0.00223	0.00034	0.00003	0.00634	0.001004	0.00272
0.00096	0.00095	0.000426	0.00595	0.00047								
Proteobacteria	Mitochondria	0.00053	0.0002	0.0003	0.00014	0.00179	0.00031	0.00145	0.00078	0.0019	0.000377	0.0021
0.0011	0.00079	0.000466	0.00155	0.00035								
Proteobacteria	Legionellaceae	0.00262	0.00122	0.00197	0.00013	0.00197	0.00042	0.00098	0.00036	0.00373	0.002138	0.00153
0.00052	0.00174	0.000206	0.00091	0.00032								
Proteobacteria	Xanthomonadaceae	0.00421	0.0005	0.00787	0.0002	0.00382	0.00161	0.00756	0.00207	0.00491	0.000953	0.00834
0.00259	0.00252	0.000339	0.01248	0.00065								
Proteobacteria	Unknown_Family	0.00151	0.00162	0.00078	0.00062	0.00136	0.00054	0.00162	0.00168	0.00062	0.00042	0.00432
0.00287	0.00194	0.001939	0.00185	0.00035								
Proteobacteria	Haliangiaceae	0.00367	0.00085	0.00301	0.00055	0.00131	0.00006	0.0024	0.00046	0.00167	0.000169	0.00285
0.00185	0.00349	0.001152	0.00197	0.00051								
Proteobacteria	Neisseriaceae	0.00061	0.00061	0.00088	0.00153	0.00409	0.00431	0.00008	0.00015	0.00204	0.001402	0.00063
0.0011	0.00174	0.001725	0.00027	0.00028								
Proteobacteria	Labraceae	0.00141	0.00039	0.00044	0.00021	0.00311	0.00044	0.00069	0.00036	0.00214	0.000199	0.00015
0.00013	0.00169	0.000022	0.00075	0.00035								

Proteobacteria Methylococcaceae 0.00125 0.00072 0.00077 0.00043 0.00031 0.00006 0.00164 0.00087 0.00035
0.000038 0.00021 0.00012 0.00069 0.00004 0.0004 0.00012

Proteobacteria Steroidobacteraceae 0.00093 0.00032 0.00032 0.00033 0.00209 0.00115 0.00096 0.00025 0.00107
0.000183 0.00019 0.0002 0.00299 0.000544 0.00036 0.00032

Proteobacteria Micavibrionaceae 0.00112 0.00083 0.00165 0.00058 0.00106 0.00046 0.00102 0.00023 0.00077
0.000214 0.00073 0.00013 0.00248 0.000198 0.00068 0.00009

Proteobacteria Magnetospiraceae 0.00036 0.00004 0.00026 0.00033 0.00145 0.00031 0.0005 0.00007 0.00118
0.000311 0.00028 0.00027 0.00034 0.000103 0.00107 0.00011

Proteobacteria Blfdi19 0.00179 0.00261 0.00047 0.00024 0.00032 0.00017 0.00103 0.00095 0.00017
0.000296 0.00067 0.00067 0.00134 0.000412 0.00067 0.00041

Proteobacteria Bacteriovoracaceae 0.00057 0.00031 0.00097 0.00042 0.00054 0.00016 0.00044 0.00008 0.00115
0.000575 0.00068 0.00036 0.00064 0.000222 0.00044 0.00005

Proteobacteria Polyangiaceae 0.0006 0.0003 0.00066 0.00025 0.00014 0.00005 0.00026 0.00033 0.00024
0.000416 0.0016 0.00037 0.00045 0.00027 0.00057 0.00016

Proteobacteria Parvibaculaceae 0.00055 0.00053 0.00063 0.00003 0.00116 0.00061 0.00054 0.00005 0.00147
0.000643 0.00114 0.00099 0.00124 0.000705 0.00073 0.00013

Proteobacteria Rhodobacteraceae 0.00008 0.00007 0.00014 0.00014 0.00162 0.00066 0.00024 0.00024 0.00058
0.000285 0.00017 0.00015 0 0 0.00051 0.0002

Proteobacteria Sandaracinaceae 0.00109 0.00068 0.00122 0.00049 0.00033 0.00028 0.00231 0.00058 0.00047
0.000195 0.00262 0.00121 0.00129 0.000858 0.00216 0.00022

Proteobacteria Hydrogenophilaceae 0.00063 0.00025 0.00024 0.00024 0.00089 0.00022 0.00013 0.00013 0.00096
0.000397 0.00023 0.00008 0.00054 0.000024 0.00027 0.00009

Proteobacteria Oligoflexaceae 0.00193 0.00067 0.00234 0.00066 0.00187 0.00047 0.00141 0.00035 0.00112
0.000347 0.0007 0.00012 0.00591 0.002852 0.00084 0.00034

Proteobacteria Geobacteraceae 0.00049 0.00008 0.00026 0.00015 0.0004 0.00007 0.00017 0.00003 0.00032
0.000122 0.0001 0.00009 0.00075 0.000148 0.00005 0.00009

Proteobacteria Methyloligellaceae 0.00036 0.00032 0.00029 0.00028 0.00017 0.00018 0.00013 0.00023 0.0001
0.000166 0.00097 0.00055 0.00044 0.000623 0.00017 0.00029

Proteobacteria Rhodospirillaceae 0.00095 0.00024 0.00028 0.00025 0.00121 0.00014 0.00017 0.00029 0.00076
0.000218 0.00009 0.00016 0.00091 0.000418 0.00013 0.00022

Proteobacteria SM2D12 0.00087 0.00045 0.00072 0.00044 0.00072 0.00036 0.00034 0.00009 0.00075
0.000165 0.00057 0.00007 0.00076 0.000297 0.00067 0.00006

Proteobacteria Nannocystaceae 0.00052 0.00058 0.00023 0.00014 0.00031 0.00017 0.00053 0.00026 0.00021
0.000104 0.00065 0.00014 0.0011 0.000753 0.00017 0.00008

Proteobacteria Archangiaceae 0.0007 0.00047 0.0005 0.00033 0.00028 0.00021 0.0006 0.00018 0.00036
0.000103 0.0005 0.00075 0.00036 0.000011 0.00032 0.00006

Proteobacteria Coxiellaceae 0.00009 0.00016 0.00024 0.00018 0.00195 0.00064 0.00008 0.00013 0.00068
0.000351 0.00052 0.00038 0.00026 0.000027 0.00023 0.00039

Proteobacteria Acetobacterales_Incertae_Sedis 0.00007 0.00012 0 0 0.0006 0.00003 0.00012 0.00013
0.0008 0.000406 0.00018 0.00019 0.00025 0.000172 0.00022 0.00004

Proteobacteria	Rickettsiaceae	0.00059	0.00086	0.00098	0.00122	0.00035	0.00044	0.00049	0.00012	0.00074	0.000967	0.00025
0.00025	0.00003	0.000046	0.00051	0.00032								
Proteobacteria	Vulgatibacteraceae	0.00032	0.00028	0.00034	0.00016	0.00011	0.0002	0.00008	0.00014	0.0001	0.00018	0.00116
0.00027	0.00023	0.000319	0.00032	0.00002								
Proteobacteria	KD3.10	0.00014	0.00024	0	0	0.00072	0.00042	0.00003	0.00005	0.00022	0.000112	0
0	0.00062	0.000007	0.00007	0.00009								
Proteobacteria	Gallionellaceae	0.00075	0.00033	0.00008	0.00008	0.00047	0.00038	0.00037	0.00015	0.00036	0.000034	0
0	0.00061	0.000361	0.00019	0.00017								
Proteobacteria	Methylomonaceae	0.00035	0.00034	0.00122	0.0007	0.00062	0.00036	0.00148	0.00019	0.00006	0.000103	0.0003
0.00029	0.00127	0.000246	0.00024	0.00016								
Proteobacteria	A21b	0.00021	0.00022	0.00017	0.00015	0.00005	0.00009	0.0002	0.00006	0.00009	0.000152	0.00083
0.00036	0.00028	0.000088	0	0								
Proteobacteria	AB1	0.0001	0.00006	0.00008	0.00007	0.00021	0.00024	0.00002	0.00004	0.00047	0.000205	0.00012
0.00013	0.00015	0.000002	0.00003	0.00005								
Proteobacteria	URHD0088	0.00033	0.0002	0.00022	0.00006	0.00009	0.00016	0.00022	0.00008	0.00013	0.000142	0
0	0	0	0.00006	0.0001								
Proteobacteria	Pseudohongiellaceae	0.00023	0.0002	0.00048	0.00042	0.00011	0.00005	0.00023	0.00008	0.00048	0.000173	0.00022
0.00027	0.00016	0.000044	0.00021	0.00018								
Proteobacteria	Paracaedibacteraceae	0.00025	0.00022	0.00014	0.00012	0.0001	0.00009	0.00001	0.00002	0.0007	0.000326	0.00009
0.00009	0	0	0.00001	0.00002								
Proteobacteria	Sulfuricellaceae	0.0004	0.00019	0	0	0.00019	0.00025	0	0	0.00023	0.000165	0
0	0	0	0	0	0							
Proteobacteria	Azospirillaceae	0.00006	0.0001	0.00002	0.00004	0.00035	0.00005	0	0	0.00008	0.000131	0
0	0.00037	0.000026	0	0								
Proteobacteria	Beggiatoaceae	0.00025	0.00022	0.00009	0.00008	0.00005	0.00008	0.0001	0.00009	0.00011	0.000106	0.00003
0.00006	0.00004	0.000062	0	0								
Proteobacteria	KF.JG30.B3	0.00018	0.00016	0.00012	0.0001	0.00005	0.0001	0.00011	0.00009	0.00006	0.000103	0.00064
0.00022	0.00036	0.000164	0	0								
Proteobacteria	Syntrophobacteraceae	0.0001	0.00018	0.00004	0.00007	0.00009	0.00016	0.00004	0.00007	0.00009	0.000097	0
0	0.00015	0.000059	0	0								
Proteobacteria	Desulfarculaceae	0	0	0.00002	0.00003	0.00003	0.00006	0	0	0.00003	0.000055	0.00058
0.00029	0.00029	0.000257	0.00014	0.00016								
Proteobacteria	Amb.16S.1323	0	0	0.00014	0.00024	0	0	0.00014	0.00025	0	0	0.00011
0.00019	0	0	0	0								
Proteobacteria	Magnetospirillaceae	0	0	0	0	0.00038	0.00006	0	0	0.0001	0.000175	0
0	0	0	0.00001	0.00002								
Proteobacteria	Enterobacteriaceae	0	0	0	0	0	0	0	0	0.0002	0.000208	0
0	0	0	0.00012	0.0002								
Proteobacteria	Alteromonadaceae	0	0	0	0	0.00003	0.00006	0	0	0	0	0.00027
0.00046	0	0	0.00003	0.00006								

Proteobacteria	Inquilinaceae	0	0	0	0	0.00031	0.00031	0.00003	0.00005	0	0	0.00019
0.00019	0	0	0.00004	0.00007								
Proteobacteria	Porticoccaceae	0	0	0	0	0.00004	0.00007	0	0	0	0.00002	0.00003
0.00033	0.000226	0	0									
Proteobacteria	Midichloriaceae	0.00003	0.00005	0	0	0.0001	0.0001	0	0	0.00002	0.000042	0
0	0.00022	0.00012	0	0								
Proteobacteria	WC3.116	0	0	0	0	0.0002	0.00007	0	0	0	0	0
0.00013	0.000186	0	0									
Proteobacteria	Kaistiaceae	0	0	0	0	0.00004	0.00007	0.0001	0.00011	0	0	0
0.00005	0.000076	0	0									
Proteobacteria	Moraxellaceae	0.00008	0.00014	0	0	0	0	0	0	0	0	0.00015
0.00018	0	0	0	0								
Proteobacteria	A0839	0	0	0	0	0	0	0.00004	0.00007	0	0	0.00014
0.00018	0	0	0.00003	0.00006								
Spirochaetes	Spirochaetaceae	0.00453	0.00113	0.00145	0.00047	0.00199	0.0007	0.00101	0.00025	0.00231	0.00024	0.00037
0.00048	0.00791	0.001512	0.00015	0.00011								
Spirochaetes	Leptospiraceae	0.00059	0.00028	0.00057	0.00028	0.00051	0.00024	0.00035	0.00007	0.00058	0.000306	0.00043
0.0001	0.00083	0.000066	0.00012	0.00014								
Verrucomicrobia	Pedosphaeraceae	0.05525	0.02228	0.02605	0.00435	0.05832	0.0142	0.01725	0.00246	0.04768	0.014751	0.04196
0.00846	0.0895	0.020503	0.01601	0.00035								
Verrucomicrobia	Opitutaceae	0.11942	0.02194	0.06017	0.00711	0.11155	0.0093	0.06656	0.00186	0.1193	0.012225	0.06622
0.00561	0.09601	0.003348	0.04289	0.00647								
Verrucomicrobia	Verrucomicrobiaceae	0.00912	0.00149	0.01056	0.00168	0.00891	0.00186	0.00861	0.00142	0.00886	0.002194	0.00806
0.00044	0.01336	0.000658	0.00555	0.00051								
Verrucomicrobia	Chthoniobacteraceae	0.00828	0.00185	0.00357	0.00032	0.00958	0.00297	0.00333	0.00074	0.01838	0.005775	0.00547
0.00117	0.0055	0.000191	0.00508	0.00036								
Verrucomicrobia	Rubritaleaceae	0.00286	0.00066	0.0122	0.00286	0.01071	0.00098	0.02141	0.01041	0.01322	0.00159	0.01549
0.00596	0.00342	0.001977	0.02737	0.00406								
Verrucomicrobia	Puniceicoccaceae	0.00163	0.00141	0.00108	0.00084	0.00504	0.00293	0.00023	0.0003	0.00089	0.000661	0.00215
0.0007	0.00655	0.004519	0.00032	0.00002								
Verrucomicrobia	Terrimicrobiaceae	0.00114	0.00018	0.00068	0.00048	0.00217	0.0006	0.00071	0.00067	0.00246	0.00092	0.001
0.0003	0.0014	0.000187	0.00043	0.00011								
Verrucomicrobia	Xiphinematobacteraceae	0.00022	0.00028	0.00031	0.00004	0.00156	0.00053	0.0001	0.0001	0.00128	0.000801	0.00141
0.00066	0.00079	0.000036	0.00076	0.00021								
Verrucomicrobia	Methylacidiphilaceae	0.00009	0.00015	0	0	0	0	0	0	0.00008	0.000131	0.00022
0.00006	0	0	0.00002	0.00004								
NA	X67.14	0.00977	0.00263	0.00962	0.00245	0.00238	0.0008	0.01013	0.0008	0.0032	0.001401	0.00468
0.00142	0.00558	0.00065	0.003	0.00028								
NA	X0319.6G20	0.00579	0.00048	0.00496	0.00103	0.00267	0.0005	0.00308	0.00058	0.00444		
0.000563	0.00384	0.0009	0.00347	0.000673	0.0019	0.00059						

NA	X37.13	0.00237	0.0011	0.00251	0.00094	0.00122	0.00002	0.00141	0.00021	0.00445	0.000382	0.00264
	0.00117	0.00069	0.00027	0.00142	0.00008							
NA	X01D2Z36	0	0	0.00045	0.00047	0	0	0.00061	0.00028	0.00001	0.000026	0.0002
	0.00018	0	0	0.00018	0.0002							
NA	Unknown_Family.1	0	0	0	0	0	0	0.00009	0.00008	0	0	
	0.00006	0.00005	0	0	0.00017	0.00017						

Supplementary Table 1: Bacteria output table, mean relative abundance and standard error per group

Phylum	**Family**	**T4_CT_BS_mean**	**T4_CT_BS_stdDev**	**T4_CH_BS_mean**	**T4_CH_BS_stdDev**	**T8_CT_BS_mean**					
	T8_CT_BS_stdDev	**T8_CH_BS_mean**	**T8_CH_BS_stdDev**	**T4_CT_rhizo_mean**	**T4_CT_rhizo_stdDev**	**T4_CH_rhizo_mean**					
	T4_CH_stdDev	**T8_CT_rhizo_mean**	**T8_CT_rhizo_stdDev**	**T8_CH_rhizo_mean**	**T8_CH_rhizo_stdDev**						
Ascomycota	Pseudeurotiaceae	0.16168	0.0059	0.24542	0.09321	0.11851	0.04513	0.37358	0.11665	0.29731	
	0.04286	0.27837	0.0181	0.28234	0.06496	0.325412	0.04677				
Ascomycota	Onygenales_fam_Incertae_sedis	0.14069	0.04447	0.01394	0.00408	0.1085	0.01458	0.00788	0.00338		
	0.03626	0.00294	0.00736	0.00375	0.03698	0.00325	0.002324	0.0002			
Ascomycota	Pezizaceae	0.06666	0.06588	0.00799	0.00551	0.15694	0.15814	0.00901	0.00206	0.0132	0.00611
	0.00119	0.00133	0.05436	0.01084	0.000477	0.00018					
Ascomycota	Cephalothecaceae	0.03081	0.01544	0.00304	0.00136	0.02387	0.01146	0.00407	0.00379	0.08468	
	0.0282	0.00398	0.00129	0.05592	0.00401	0.004883	0.00151				
Ascomycota	Helotiaceae	0.07525	0.01232	0.01764	0.00403	0.07215	0.04937	0.00902	0.00186	0.00438	
	0.00083	0.0007	0.00048	0.00348	0.00028	0.000445	0.00003				
Ascomycota	Trichomonascaceae	0.03849	0.01094	0.00263	0.00074	0.02504	0.00786	0.0018	0.00034	0.0129	
	0.00402	0.00139	0.0009	0.01579	0.00483	0.000978	0.00021				
Ascomycota	Saccharomycetales_fam_Incertae_sedis	0.03941	0.00737	0.00335	0.00091	0.03099	0.00928	0.00159			
	0.00078	0.0202	0.00592	0.00228	0.0014	0.02025	0.0036	0.001481	0.00013		
Ascomycota	Pyronemataceae	0.00637	0.01091	0.00294	0.00463	0.01555	0.02016	0.00053	0.00092	0.00267	
	0.00379	0.0005	0.00045	0.00403	0.00644	0.000472	0.00054				
Ascomycota	Chaetomiaceae	0.0159	0.01567	0.00547	0.00249	0.01898	0.01679	0.00552	0.00493	0.00519	
	0.00082	0.01596	0.01322	0.00925	0.00411	0.012044	0.00231				
Ascomycota	Plectosphaerellaceae	0.00621	0.00161	0.00103	0.00119	0.00602	0.00326	0.00069	0.00108	0.00304	
	0.00064	0.00057	0.00031	0.00313	0.00098	0.000376	0.00023				
Ascomycota	Helotiales_fam_Incertae_sedis	0.01131	0.00369	0.00143	0.00055	0.00809	0.00053	0.00092	0.00049		
	0.00205	0.00025	0.00031	0.00027	0.00334	0.00153	0.000195	0.00013			
Ascomycota	Erysiphaceae	0	0	0	0	0.00114	0.00198	0.00009	0.00013	0	
	0	0	0	0.01068	0.01775	0.000102	0.00003				
Ascomycota	Lasiosphaeriaceae	0.00998	0.01214	0.00333	0.00547	0.01664	0.02495	0.0004	0.00065	0.00275	
	0.00223	0.00043	0.00041	0.00716	0.0096	0.000354	0.00041				
Ascomycota	Thermoascaceae	0.00245	0.00116	0.00015	0.00013	0.00222	0.00081	0.00012	0.00021	0.00669	
	0.00371	0.00011	0.00019	0.00461	0.00127	0.000138	0.00015				

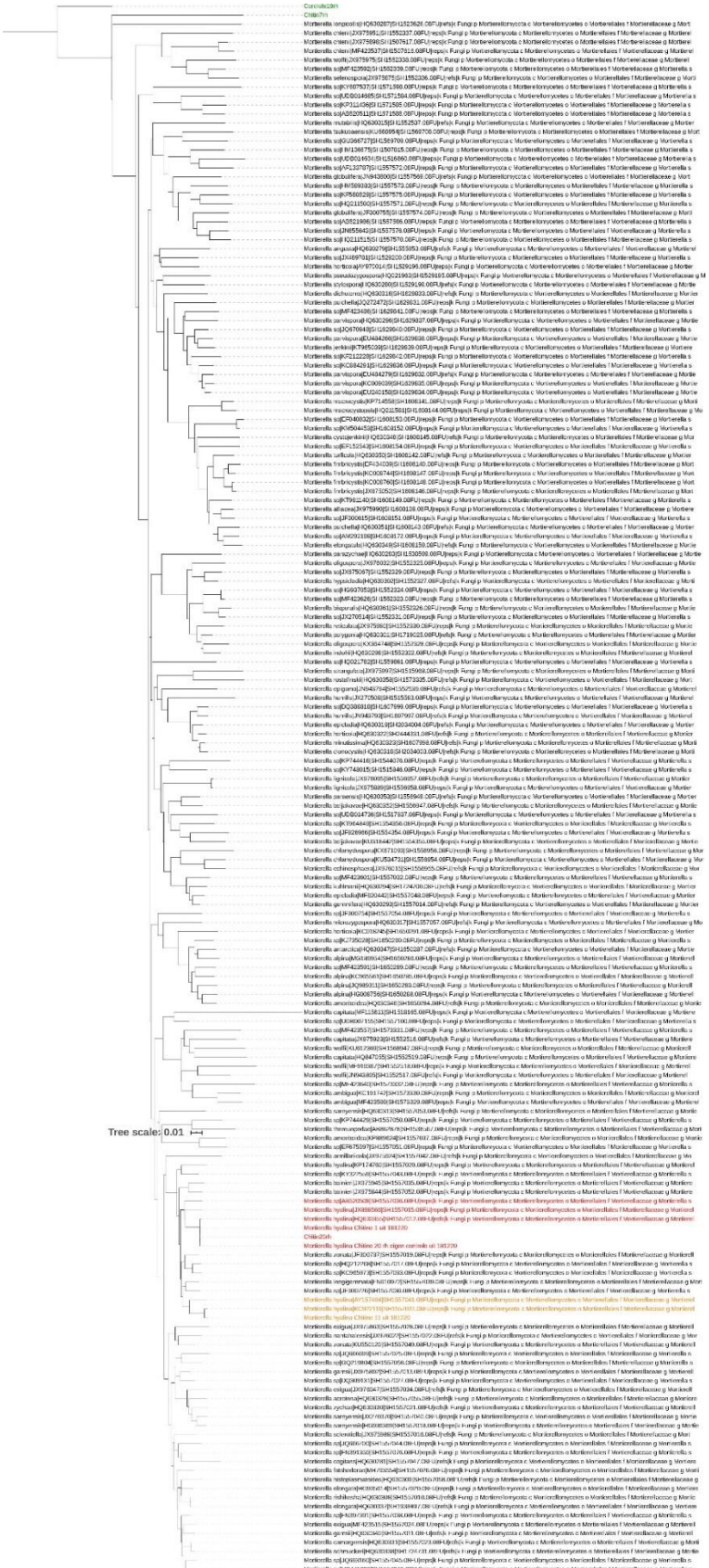
Ascomycota	Hypocreaceae	0.00295	0.00096	0.00032	0.00008	0.00289	0.00047	0.00014	0.00012	0.00284	0.00137	0.00038	0.00014
0.00339	0.00064	0.000325	0.00016										
Ascomycota	Mycosphaerellaceae	0.00185	0.00152	0.00038	0.00056	0.00098	0.00117	0.00012	0.00021	0.00209	0.00214	0.00086	0.00122
0.00198	0.00343	0.000093	0.00012										
Ascomycota	Cladosporiaceae	0.00075	0.00129	0.00038	0.00059	0.0018	0.00135	0.00001	0.00002	0.00121	0.00118	0.00046	0.00047
0.00484	0.00348	0.000218	0.00038										
Ascomycota	Nectriaceae	0.00288	0.0038	0.00119	0.00181	0.00673	0.01024	0	0	0.001	0.00122	0.00037	0.0004
0.00203	0.00291	0.000201	0.00017										
Ascomycota	Coniochaetaceae	0.00223	0.00175	0.00074	0.00036	0.00335	0.00454	0.00018	0.00025	0.00083	0.00028	0.00009	0.00008
0.00359	0.00307	0.000089	0.00015										
Ascomycota	Ascobolaceae	0.00032	0.00055	0.0003	0.00051	0.00557	0.00833	0	0	0.00046	0.00042	0.00008	0.00009
0.00086	0.00128	0.000084	0.00015										
Ascomycota	Trichocomaceae	0.00071	0.00047	0.00002	0.00003	0.00235	0.00238	0.00002	0.00003	0.0027	0.00363	0.00012	0.00003
0.00099	0.0008	0.000141	0.00021										
Ascomycota	Sordariaceae	0.00011	0.00019	0.00006	0.00007	0.00242	0.00163	0.00009	0.00016	0.00038	0.00065	0	0
0.0003	0.00033	0.000136	0.00015										
Ascomycota	Melanommataceae	0.00036	0.00062	0.00009	0.00016	0.00197	0.00287	0	0	0.0002	0.00017	0.00003	0.00005
0.00093	0.00161	0	0										
Ascomycota	Pleosporaceae	0.00136	0.00189	0.00021	0.00029	0.00075	0.00125	0	0	0.00019	0.0002	0.00073	0.00076
0.00055	0.00057	0.00016	0.0002										
Ascomycota	Symbiotaphrinaceae	0.00042	0.00037	0	0	0.00047	0.00029	0	0	0.00102	0.00113	0.00004	0.00004
0.0016	0.0015	0	0										
Ascomycota	Didymellaceae	0.00007	0.00012	0.00009	0.00011	0.00039	0.00047	0.00003	0.00006	0.00069	0.00103	0.0001	0.00008
0.00061	0.0003	0.000089	0.00013										
Ascomycota	Microdochiaceae	0.00115	0.001	0.00006	0.00011	0.00053	0.00053	0.00006	0.0001	0.00021	0.00018	0	0
0.00017	0.00029	0	0										
Ascomycota	Clavicipitaceae	0.00014	0.00025	0.00023	0.00039	0.00104	0.00124	0	0	0.00027	0.00023	0.00005	0.00008
0.00015	0.00025	0.00003	0.00005										
Ascomycota	Leotiaceae	0.00183	0.00062	0.00013	0.00008	0.00077	0.00093	0.00009	0.00008	0.00017	0.0003	0.00009	0.00013
0.00017	0.00016	0	0										
Ascomycota	Magnaporthaceae	0.00063	0.00109	0.0003	0.00048	0.00098	0.00086	0	0	0.00034	0.00059	0.00001	0.00002
0.00103	0.00178	0.000039	0.00005										
Ascomycota	Orbiliaceae	0.00012	0.00021	0	0	0.00037	0.00018	0	0	0.00027	0.00008	0.00004	0.00007
0.00066	0.00048	0.00002	0.00003										
Ascomycota	Myxotrichaceae	0.00054	0.00026	0.00007	0.00007	0.00103	0.00129	0.00002	0.00003	0.00049	0.00006	0.00005	0.00009
0.00059	0.00019	0.000034	0.00003										
Ascomycota	Aspergillaceae	0.00105	0.00062	0.00014	0.00014	0.00104	0.00031	0.00004	0.00004	0.00144	0.00141	0.00002	0.00003
0.00344	0.00246	0.000009	0.00002										
Ascomycota	Vibrisseaceae	0.00079	0.0008	0.00016	0.00016	0.00042	0.00037	0.00005	0.00006	0.00003	0.00006	0.00005	0.00008
0.00008	0.00014	0	0										

Ascomycota	Hyaloscyphaceae	0.00081	0.00078	0	0	0.00073	0.00037	0.00002	0.00003	0.00009	0.00016
0.00003	0.00004	0	0	0	0						
Ascomycota	Phaeosphaeriaceae	0	0	0.00009	0.00013	0.00139	0.00147	0	0	0.00008	0.00014
0	0	0.00026	0.00029	0	0						
Ascomycota	Microascaceae	0.00043	0.00052	0.00021	0.00036	0.00147	0.00074	0	0	0.00021	0.00019
0.0001	0.00017	0.00027	0.00014	0	0						
Ascomycota	Myrmecridiaceae	0.00052	0.00091	0	0	0.00035	0.00061	0	0	0.00008	0.00014
0	0	0	0	0	0						
Ascomycota	Didymosphaeriaceae	0.0002	0.00034	0.00012	0.00021	0.00057	0.00017	0	0	0.00004	0.00006
0.00004	0.00006	0.00041	0.00071	0	0						
Ascomycota	Dermateaceae	0.00086	0.00076	0	0	0.00044	0.00047	0.00002	0.00003	0.00015	0.00014
0	0	0.00058	0.00046	0	0						
Ascomycota	Trematosphaeriaceae	0.00016	0.00027	0	0	0.00054	0.00094	0	0	0.00005	0.00008
0	0	0	0	0	0						
Ascomycota	Niessliaceae	0	0	0.00017	0.0003	0.00059	0.00077	0	0	0	0
0	0	0.00007	0.00006	0	0						
Ascomycota	Sordariales_fam_Incertae_sedis	0.00014	0.0002	0.00005	0.00004	0.00034	0.00014	0.00004	0.00005	0.00001	0.00002
0.0001	0.0001	0.00017	0.00015	0	0						
Ascomycota	Saccharomycetaceae	0	0	0.00003	0.00006	0.00042	0.00043	0	0	0.00018	0.00031
0	0	0.00005	0.00008	0	0						
Ascomycota	Xylariales_fam_Incertae_sedis	0	0	0.00005	0.00009	0	0	0.00002	0.00004	0	0
0	0	0	0.00024	0.00042	0	0					
Ascomycota	Hypocreales_fam_Incertae_sedis			0.00014	0.00014	0.00004	0.00007	0.00025	0.00034	0	0
0	0	0	0	0.00004	0.00007	0	0				
Ascomycota	Periconiaceae	0	0	0.00001	0.00002	0.00009	0.00015	0	0	0.00011	0.00018
0	0	0.00005	0.00008	0	0						
Ascomycota	Phaffomycetaceae	0.00012	0.0002	0	0	0.00003	0.00005	0	0	0	0
0	0	0.00004	0.00006	0	0						
Basidiomycota	Pleurotaceae	0.24865	0.0791	0.07478	0.01996	0.12311	0.04312	0.047	0.02129	0.2311	0.04432
0.03444	0.00355	0.17097	0.03168	0.06011	0.01775						
Basidiomycota	Stephanosporaceae	0.00913	0.00838	0.00346	0.00097	0.00693	0.00511	0.00331	0.00149	0.01691	0.01102
0.01298	0.00881	0.01832	0.01424	0.023864	0.00284						
Basidiomycota	Piskurozymaceae	0.0121	0.00151	0.0016	0.00059	0.01587	0.00359	0.00098	0.00068	0.00433	0.00144
0.00059	0.00035	0.00411	0.00144	0.000459	0.00006						
Basidiomycota	Trichosporonaceae	0.00944	0.00208	0.00112	0.00018	0.00445	0.00068	0.00108	0.00057	0.00694	0.00304
0.00062	0.00019	0.00799	0.00105	0.001148	0.00018						
Basidiomycota	Trimorphomycetaceae			0.00842	0.00169	0.0009	0.00009	0.01069	0.00495	0.00052	0.00028
0.00175	0.00061	0.00025	0.00043	0.00447	0.00097	0.000226	0.00015				
Basidiomycota	Entolomataceae	0.00018	0.00031	0.00272	0.00334	0.02351	0.04072	0.00033	0.00047	0.00035	0.00021
0.00247	0.00167	0.00028	0.00019	0.001711	0.00247						

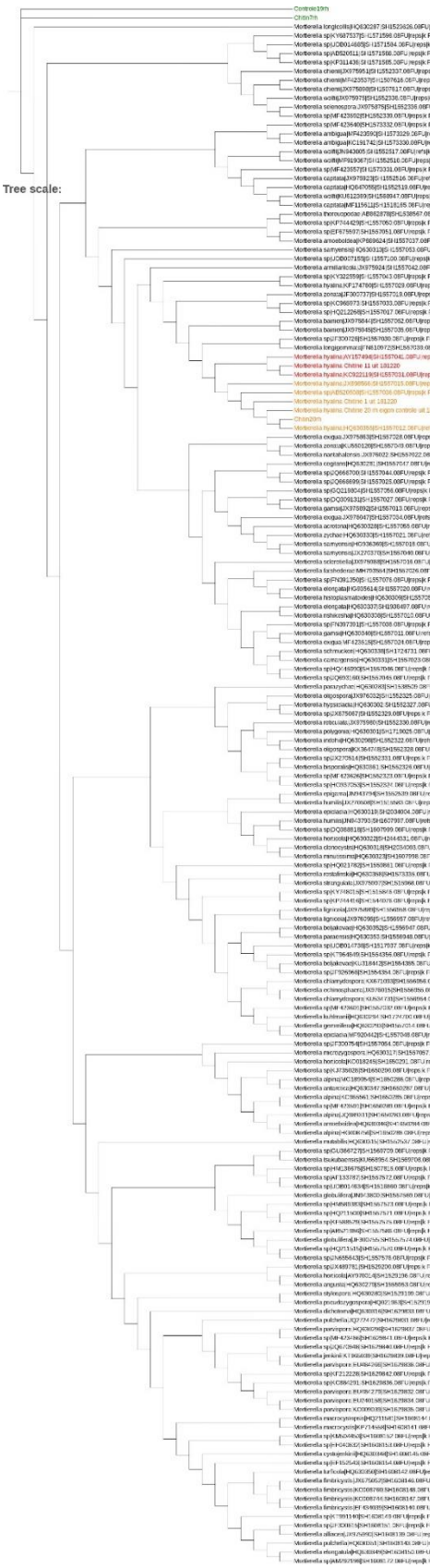
Basidiomycota	Psathyrellaceae	0.00089	0.00153	0.00018	0.00009	0.01035	0.0122	0.00002	0.00004	0.00051	0.00046	0.00045	0.00057
0.00136	0.00152	0.000059	0.00005										
Basidiomycota	Cantharellales_fam_Incertae_sedis	0.00244	0.00422	0.00124	0.00199	0.00367	0.00382	0.00042	0.00074	0.00106	0.00137		
0.00011	0.00018	0.00152	0.00264	0.000213	0.00017								
Basidiomycota	Cortinariaceae	0.00059	0.00073	0.00004	0.00004	0.00034	0.00038	0	0	0.0017	0.00003	0.00017	0.00003
0.00186	0.00119	0.000216	0.0001										
Basidiomycota	Erythrobasidiales_fam_Incertae_sedis	0.0002	0.00013	0.00002	0.00003	0.00038	0.00034	0	0	0.00009	0.0001		
0.00008	0.00014	0.00031	0.00012	0.000014	0.00002								
Basidiomycota	Agaricaceae	0	0	0	0	0.00026	0.00023	0	0	0.00079	0.00137	0	0
0.00008	0.00013	0.000068	0.00012										
Basidiomycota	Ustilaginaceae	0.00029	0.00045	0.00006	0.0001	0.00047	0.00065	0	0	0	0	0	0
0.00028	0.00048	0.000025	0.00004										
Basidiomycota	Chrysozymaceae	0.00029	0.00026	0.00001	0.00002	0.00019	0.00018	0	0	0.00015	0.00017	0.00003	0.00005
0.00012	0.0002	0.00001	0.00002										
Basidiomycota	Sporidiobolaceae	0	0	0.00005	0.00009	0	0	0	0	0	0	0.00033	0.00057
0.0001	0.0001	0.000029	0.00005										
Basidiomycota	Ceratobasidiaceae	0.00016	0.00028	0	0	0.00036	0.00062	0	0	0	0	0	0
0	0	0.000029	0.00003										
Basidiomycota	Chionosphaeraceae	0	0	0.00026	0.00032	0.00012	0.0002	0	0	0	0	0.00003	0.00005
0.00004	0.00007	0.000025	0.00004										
Basidiomycota	Exidiaceae	0.00037	0.00064	0	0	0	0	0	0	0.00009	0.00015	0.00003	0.00005
0	0	0											
Basidiomycota	Strophariaceae	0	0	0.00003	0.00006	0.00007	0.00011	0	0	0.00007	0.00013	0	0
0.00033	0.00058	0	0										
Basidiomycota	Bolbitiaceae	0	0	0.00003	0.00005	0	0	0	0	0.00017	0.0003	0	0
0.00018	0.00031	0.000024	0.00004										
Basidiomycota	Leucosporidiaceae	0.00017	0.00011	0	0	0	0	0	0	0.00017	0.00022	0	0
0	0	0	0										
Basidiomycota	Filobasidiaceae	0.00003	0.00005	0	0	0	0	0	0	0	0	0.00002	0.00004
0	0	0.000024	0.00004										
Chytridiomycota	Spizellomycetaceae	0.00026	0.00046	0	0	0.00003	0.00006	0.00014	0.00014	0.00038	0.00051	0.00005	0.00009
0	0	0.000195	0.00021										
Chytridiomycota	Rhizophydiales_fam_Incertae_sedis	0	0	0.00005	0.00009	0.00063	0.00109	0	0	0	0	0	0
0	0	0.00008	0.00013	0	0								
Chytridiomycota	Rhizophlyctidaceae	0	0	0.00005	0.00009	0.00051	0.00089	0	0	0.00004	0.00007	0	0
0	0	0	0										
Mortierellomycota	Mortierellaceae	0.07384	0.04196	0.59904	0.10927	0.14616	0.03574	0.52976	0.08546	0.22047	0.12399	0.62999	0.02952
0.23598	0.11255	0.560002	0.0532										
Mucoromycota	Umbelopsidaceae	0.00363	0.00224	0.00035	0.00006	0.00398	0.00107	0.00025	0.00008	0.00343	0.00148	0.00022	0.00032
0.00906	0.00483	0.000135	0.00006										

Mucoromycota	Mucoraceae		0.00009	0.00016	0	0	0.0002	0.00019	0	0	0.00039
0.00039	0.00002	0.00003	0.0011	0.0003	0	0					
Zoopagomycota	Piptocephalidaceae		0	0	0	0	0	0	0	0	0
0	0.00017	0.0002	0	0	0.000066	0.00011					

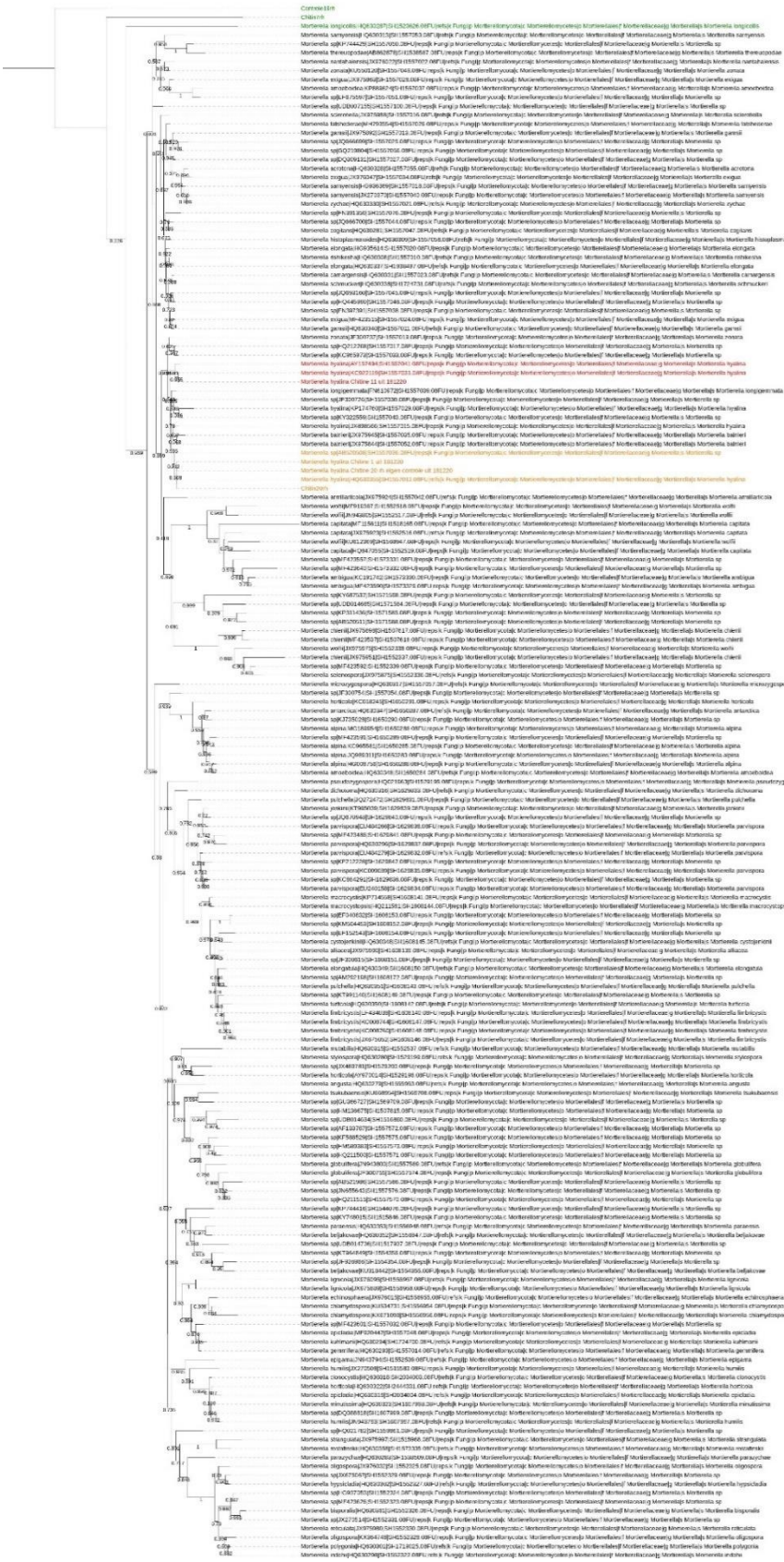
Supplementary Table 2: Fungi output table, mean relative abundance and standard error per group



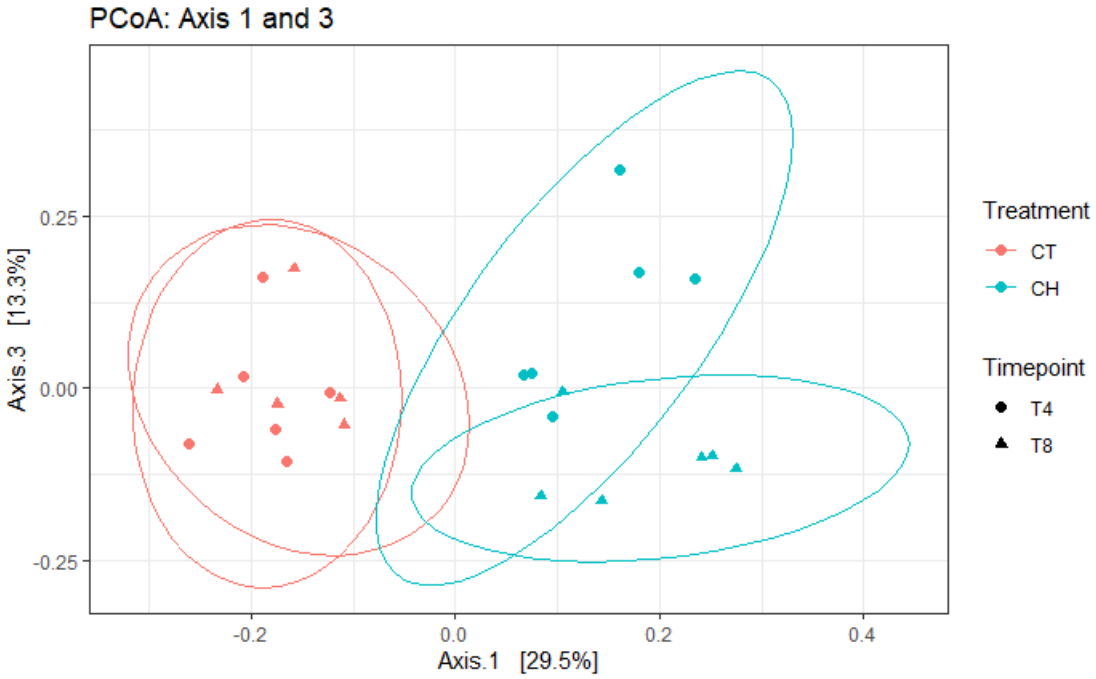
Supplementary Figure 8: EBI NJ phylogenetic tree



Supplementary Figure 9: Fastphylo NJ phylogenetic tree



Supplementary Figure 10: Fasttree ML phylogenetic tree



Supplementary Figure 11: Beta diversity of bacteria on PCoA dimensions 1 & 3
Measurement: Bray-Curtis dissimilarity