universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Effect of Cryopreservation on the Microbiome of Plants“

verfasst von / submitted by

Gabriel Alexander Vignolle

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Pharmazie (Mag.pharm.)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 449
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Diplomstudium Pharmazie
Betreut von / Supervisor:	Univ.-Prof.Dr. Sergey B. Zotchev
Mitbetreut von / Co-Supervisor:	Dr. Andrea Kodym

List of abbreviations

A 1-7	*Artemisia laciniata* samples
ASV	amplicon sequence variant
BAP	Benzylaminopurin
bp	base pairs
GIT	gastro intestinal tract
S 1-7	*Solanum tuberosum* samples
kbp	kilo base pairs
M 1-7	*Mentha aquatic*a samples
MAK	Maximale Arbeitsplatz Konzentration
MCC	Matthews correlation coefficient
MS	Murashige and Skoog basal medium (Murashige T. et al., 1962)
nt	nucleotide
OUT	operational taxonomic unit
PVS	Plant vitrification solution
SOP	Standard operating procedure

Table of contents

Table of contents

1. Introduction

1.1. *In-vitro* cultivation of plants and cryopreservation

There are different approaches to the conservation of plant material. The classical way is to store seeds for extended periods. Although for some plant categories this poses a problem, as they do not produce seeds at all, or have no viable seeds. Vegetatively propagated plants are stored in living collections e.g in botanic gardens. Storage methods to ensure survival of commercial crops or endangered plants at the open can lead to the exposure of the stored plant material to e.g. droughts, frost, pathogens and pests. (Tahtamouni *et al.,* 2015) *In-vitro* cultivation followed by cryopreservation offers an alternative without the threat of disease or weather conditions.

To initiate *in-vitro* cultures, the plant material has to be surface-sterilized, usually in a hypochlorite solution. This surface-sterilized plant material is then taken into culture. Many different protocols exist for various plant species and different *in-vitro* techniques have been proven to be reliant. Slow growth storage is a technique used to reduce the growth of plant material under *in-vitro* conditions to allow midterm storage of valuable germplasm. This is achieved by modifying the ingredients of the culture medium, reduction of light intensity and temperature. This prolongs the intervals between subculturing and these established cultures could be stored for months up to five years, depending on the species. (Tahtamouni *et al.,* 2015; Ozudogru *et al.,* 2010)

Cryopreservation enables long-term germplasm storage and can be an adequate alternative for some species. It avoids having to subculture at regular intervals. Cryopreservation is the preservation of explants from *in-vitro* cultures in liquid nitrogen (-196°C). At this temperature all biological processes within the cell are stopped, this makes the material storable for theoretically unlimited time without genetic variation occurring. To attain efficacious cryopreservation, several factors have to be considered. The danger of lethal intra-cellular ice crystals forming while freezing has to be prevented. To avoid the formation of ice-crystals the cytosol of the cryopreserved cells have to be vitrified. Vitrification is used to describe a metastable solid physical state of water solutions at ultra-low temperature. Instead of crystallizing in an ordered way, the water molecules shift into an amorphous state where no crystallization occurs, transitioning the system into a glassy state. The vitrification of cells can be achieved by administrating highly concentrated vitrification solutions, such as PVS3 (Plant Vitrification Solution 3). (Nishizawa *et al.,* 1993) These solutions act as an antifreeze medium containing cryoprotectants and shield the treated cells from intra-cellular-crystal forming. Cryoprotectants are anti freezing agents such as glycol or DMSO. (Benson, 2008)

Different protocols and techniques have been developed to achieve successful cryopreservation: slow freezing method (considered as the classical approach); pre-culture dehydration; encapsulation-dehydration and the vitrification method.

The droplet vitrification technique, which was used in this research, is variation of the vitrification method based on highly concentrated vitrification solutions containing cryoprotectants. It consists of three different phases: loading phase, dehydration and unloading phase. The loading phase involves the treatment of the plant material with cryoprotectants i.e. diluted vitrification solutions. The dehydration is accomplished by exposure of the explants to Plant Vitrification Solutions (PVS) that contain high concentrations of sugars. The treated plant material can then be placed into small droplets of PVS on aluminum strips thus facilitating their handling before being transferred into cryovials. These cryovials are then plunged into liquid nitrogen and can be stored thereafter at -80°C or in liquid nitrogen for indefinite time. The unloading phase is used to re-establish the cryopreserved plant materials and start regeneration of the plant; this phase can be seen as part of the cryopreservation technique, because without it the stored material could not be regenerated. Unloading consists of a washing step with solutions with a lower sucrose concentration than the PVS. This cryopreservation technique has been successfully applied to many different plant species such as *Mentha aquatica*, thyme, potato and *Artemisia laciniata*. (Benson, 2008; Ozudogru *et al.,* 2010; Tahtamouni *et al.,* 2015; Kodym *et al.*, 2018)

1.2. Pathogen eradication via cryopreservation

Cryopreservation can be used for purposes other than germplasm conservation and long-term storage. One such biotechnological use is cryotherapy, a rather recent development for eliminating pathogens in infected plants. It can be seen as complementary or as a substitute to classical virus eradication techniques, e.g. thermotherapy and meristem culture. (Engelmann, 2004) Especially viral diseases constitute a growing threat to many economically important crops, because they cannot be treated chemically. Commercial production of crops depends on certain unique traits and therefore cultivars which are vegetatively propagated from cuttings or by grafting end up being infected. The cultivation of virus-free plants is of outmost importance to the sustainable production of crops. (Wang *et al.*, 2018)

Thermotherapy is achieved by heat-treatment of the plants over several weeks through growing them at high temperatures (40-50°C depending on the plant species). Thermotherapy alone is insufficient for elimination of heat-stable viruses and might not be suitable for all plant species. It is usually combined with meristem culture based on the concept of pathogen-free meristems. The

ability of plants to regenerate from such small explants can be regarded as a crucial factor, as well as employing skilled technicians for the time consuming step of meristem excision. The utilization of cryopreservation in the pathogen eradication in plants results in greater regeneration rates and higher percentages of obtained virus-free plants than meristem cultures. (Engelmann, 2004) Cryotherapy targets infected differentiated cells of apices, which contain higher amounts of water and can form intra-cellular ice-crystals, while the meristematic pathogen-free cells have a more concentrated cytosol, have smaller vacuoles and survive the freezing process. This selective cell destruction is used in cryotherapy to obtain pathogen-free plantlets. Furthermore the plant tissue is maintained juvenile throughout the cryotherapy steps, demonstrating another advantage over the classical pathogen eradication techniques. (Engelmann, 2004; Wang *et al.,* 2018; Ozudogru *et al.,* 2010)

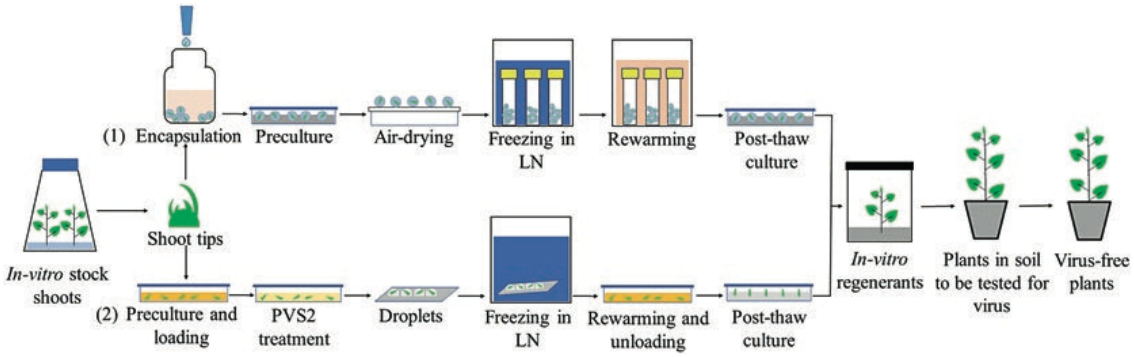


Figure 1.1 (Wang *et al.*, 2018): Overview of the main steps in cryopreservation used as cryotherapy (1) encapsulation-dehydration, (2) droplet-vitrification. LN = liquid nitrogen.

Cryotherapy is realized in several steps described and illustrated in figure 1.1 by Wang *et al.* (2018): First, the infected plant material is taken into *in-vitro* culture. After a propagation cycle, shoot tips are excised and prepared with different techniques such as droplet-vitrification or encapsulation-dehydration for cryogenic freezing in liquid nitrogen. The excised plant material is treated for a short time in liquid nitrogen. The ultra-cooled explants are rewarmed and placed on regeneration medium. The regenerated plantlets are hardened, by slowly acclimatizing them to *ex-vitro* conditions, and transferred to greenhouse conditions. The final step is the control of the plants by virus indexing, to verify the success of the procedure. (Wang M. *et al.,* 2018) Not only viruses can be eradicated through cryotherapy, but it has also been used successfully for the eradication of bacteria, although to date only one report has been published, namely eradication of *Candidatus Liberobacter asiaticus* from different citrus species. (Ding *et al.,* 2008)

1.3. Endophytes in *In-vitro* culture

Plant endophytes are microorganisms living within living plant tissue across all or a part of their life cycles, without causing disease symptoms. (Orlikowska *et al.,* 2017) Microorganisms and plants share a symbiotic evolution and development, putting them in close relationship with each other. It is widely accepted, that microorganisms such as non-pathogenic bacteria (*Azospirillum brasilense* improves water usage of wheat*, Pseudomonas mendocina* increases the drought resistance of lettuce*, Burkholderia phytofirmans* modifies the carbohydrate metabolism of grapevine increasing its adaptation to cold weather conditions) live within an interdependent relationship with their host plant and attribute to the fitness and growth of them through the production of secondary metabolites and the activation of otherwise silenced genes. (Orlikowska *et al.,* 2017; Porras-Alfaro *et al.,* 2011; Lee *et al*., 2017)

The *in-vitro* cultivation of plants is based on the premises of creating pathogen-free cultures; strictly these should not be called "axenic" (cultures free of all biological contaminations) but should be labeled as "aseptic" (free of detectable contaminations by pathogens). Within these aseptic explants there are still microorganisms residing within the tissues and cells in a dormant state. Until recently, the presence of bacteria in *in-vitro* cultures was always considered as negative factor for the cultures themselves. Now, we are aware that certain bacteria can be beneficial and have positive effects e.g. on the micropropagation, rooting and growth of plants. (Thomas *et al.,* 2008; Orlikowska *et al.,* 2017) These endophytic microorganisms may be latent and only become visually detectable after many subculturing steps and appear only under certain stressful conditions such as elevated temperature, long-term storage, changing of the medium composition and cryopreservation. (Thomas *et al.,* 2008)

Endophytic microorganisms can start to overproliferate under *in-vitro* conditions. Given their intra- and inter-cellular nature it is not possible to eradicate them via surface sterilization. Some of these endophytes have beneficial roles in *ex-vitro* conditions, such as *Pseudomonas fluorescens* and *Azospirillum brasilense*, but in *in-vitro* conditions they may inhibit growth of shoots or overgrow the plantlets. (Orlikowska *et al.,* 2017) The *in-vitro* environment with nutrient-rich media and the lack of epiphytic microorganisms give endophytes the chance to grow beyond their specific niche and even kill their host. The source of contamination is therefore difficult to determine often it is considered as external contamination, coming from inadequate surface sterilization or unsuccessful sterile practices during subculturing. An evidence of contamination can be obtained through visual examination, exhibited by cloudy spots in the media or the formation of bacterial colonies around the plant. However, microorganisms, either endophytic or contaminating, which do not grow on plant culture media, may not be detectable via visual inspection (Reed *et al*., 1995). On the other

hand, placing aseptic tissue explants from the plantlets onto media that supports bacterial growth can make dormant bacteria proliferate. Recent studies demonstrated the occurrence of bacterial endophytes as ubiquitous in *in-vitro* cultures, including plant-cell suspension cultures. (Thomas *et al.,* 2008; Orlikowska *et al.,* 2017; Massimo *et al*., 2015; Porras-Alfaro *et al.,* 2011)

1.4. Microbiome

Epiphytic and endophytic microorganisms colonizing a plant are called the microbiome and referred to as the "second genome" (Berg *et al*., 2014) giving the microbiome the same level of importance as the genome. Therefore plants, like other eukaryotes, can be described as superorganisms, meta-organisms, pan-organisms or holobionts and their microbiome as an expansion of their genotype and furthermore their phenotype. (Orlikowska *et al.,* 2017; Berg *et al*., 2014) The microbiome can vary from cultivable bacteria and fungi to obligate symbionts and uncultivable microorganisms, which do not manifest themselves phenotypically. This interwoven mutualistic partnership between microbiome and the host can lead to positive synergistic effects such as extended life and resistance to pathogens. The composition of the microbiome is different from individual to the next. Under *in-vivo* living conditions the plant is colonized through the rhizosphere, leaves, flowers, stems and cotyledons, in sum every part in contact with the exterior. Moreover, microorganisms can be transmitted over generations, vertically from parents to their progeny or horizontally through vegetative multiplication. (Lee *et al*., 2017; Ludwig-Müller, 2015; Mhlongo *et al*., 2018; Massimo *et al*., 2015; Porras-Alfaro *et al.,* 2011; Lutenberg, 2015; Riedel *et al.,* 2014)

1.4.1. Uncultivated Microorganisms

Since the 1800s the classical or conventional approach to studying microbiomes has been the isolation and cultivation of single bacteria or fungi. After the microorganisms were taken into culture they could be studied and since the 1990s their genome or parts of it (for example the 16S rRNA gene) could be sequenced with classical Sanger DNA sequencing. The obvious drawback of this technology is that it depends on clones of the microorganisms, and therefore only cultivable microorganisms can be studied, and is furthermore prone to clone bias (Riedel *et al.,* 2014). Cloning bias is the predisposition for certain regions in the genome to be cloned less frequently than others during sequencing, and thus those regions are less likely to be entirely sequenced. This bias results in a lower than anticipated read coverage (number of reads at one position) in this region, which produces gaps in the sequence. The clearest cause of cloning bias is adenine-thymine (A-T) richness,

which affects cloning rates in the Sanger sequencing method. The even greater weakness of this approach is the big proportion of uncultivable microorganisms, which cannot be identified this way. Thomas *et al.* (2008) have revealed, that the change of media composition could make what seem first uncultivable bacteria cultivable. (Massimo *et al*., 2015; Xiao *et al.,* 2014; Porras-Alfaro *et al.,* 2011; Thomas *et al.,* 2008) However, neither the correct composition of a microbiome nor the exact abundance of its members can be identified with this method. For the study of the complex variety of microbiomes, amplicon sequencing methods and the Next-generation sequencing (NGS) methods have proven of utmost importance. They made the characterization of highly diverse microbiomes possible, such as various plants, the human intestinal system or the composition of microorganisms living on the skin. (Lutenberg, 2015; Xiao *et al.,* 2014)

1.4.2. NGS-Methods in amplicon based genomics

Next-generation sequencing (NGS) technologies have augmented the research of all sciences that utilize the analysis of DNA. NGS has opened the possibility to produce large amounts of sequencing data (up to giga- and tera bp) to a drastically lower price and higher precision than conventional technologies, such as Sanger sequencing. One of these research areas, which has profited from the NGS technologies, is the analysis of microbial communities by the so-called amplicon sequencing. This targeted metagenomic approach is undertaken with the polymerase chain reaction (PCR)-assisted amplification of the part of bacterial 16S rRNA gene , and the fungal internal transcribed spacer (ITS, spacer DNA between the small-subunit ribosomal RNA (rRNA) and large-subunit rRNA genes) using specific primers . (Knief, 2014; Porras-Alfaro *et al.,* 2014; Singer *et al.,* 2016) Sequencing of these amplicons, representing mixtures of PCR products originating from microbiome, followed by bioinformatics analyses provide unprecedented insight into the compositions of bacterial and fungal communities associated with the host.

Illumina MiSeq, one of the NGS technologies most often used in amplicon-based microbiome studies , involves a two-step PCR. ; The first PCR step is the amplification of the region of interest with specialized Illumina primers, which needs to be between 400-900 bp in length. After the first amplification specialized adaptors are ligated to both ends of the amplicon. The second PCR is conducted on a lane of a flow cell made out of glass via bridge-PCR, after hybridization (attachment) of the amplicons to surface bound oligonucleotides with their specialized adaptors. It is called bridge-PCR because the attached DNA makes a bridge to nearby surface bound oligonucleotides, hybridizes to them with their adaptors and is amplified by the polymerase in this conformation. Each single bound amplicon is amplified into a clonal cluster through bridge-PCR. The result of this is a glass slide, called flow cell, dotted with single stranded DNA. (Figure 1.2)

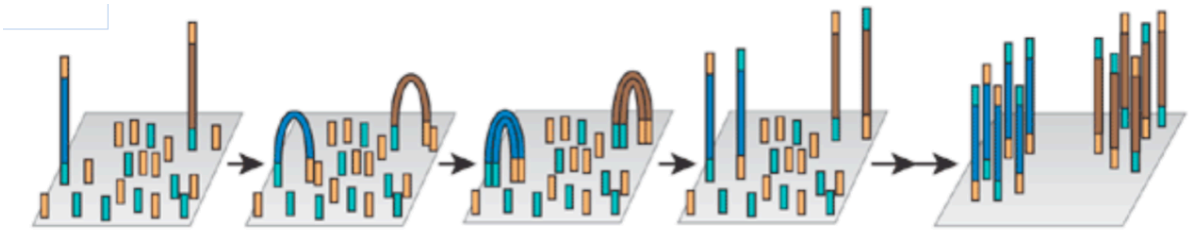


Figure 1.2 (Shendure *et al.*, 2008): graphical illustration of the bridge-PCR procedure.

The sequencing method is similar to Sanger sequencing and is achieved by reversible terminator sequencing by synthesis with modified end-blocked fluorescent nucleotides (dNTP). To initiate the first sequencing cycle primers hybridizing to the free specialized adaptors, all four labeled reversible terminator nucleotides (dNTP) and DNA polymerase are added to the flow cell. The end-cap or terminator at the 3` end blocks the polymerization, therefor only a single nucleotide can be added. The four differently labeled nucleotides (dNTP) are flushed over the flow cell at the same time. The detection of the nucleotides is achieved by differently colored fluorophors. The fluorescent emission to determine the first nucleotide of each cluster is optically detected after laser excitation. Before the polymerase can attach the next nucleotide the end-cap at the 3` terminus and the flourophore is removed. After detection and end-cap removal the polymerase can attach the next modified end-blocked fluorescent nucleotide. (Figure 1.3) The quality of the base calling is reduced at the end of the sequenced read, mainly due to substitutions introduced by the polymerase.

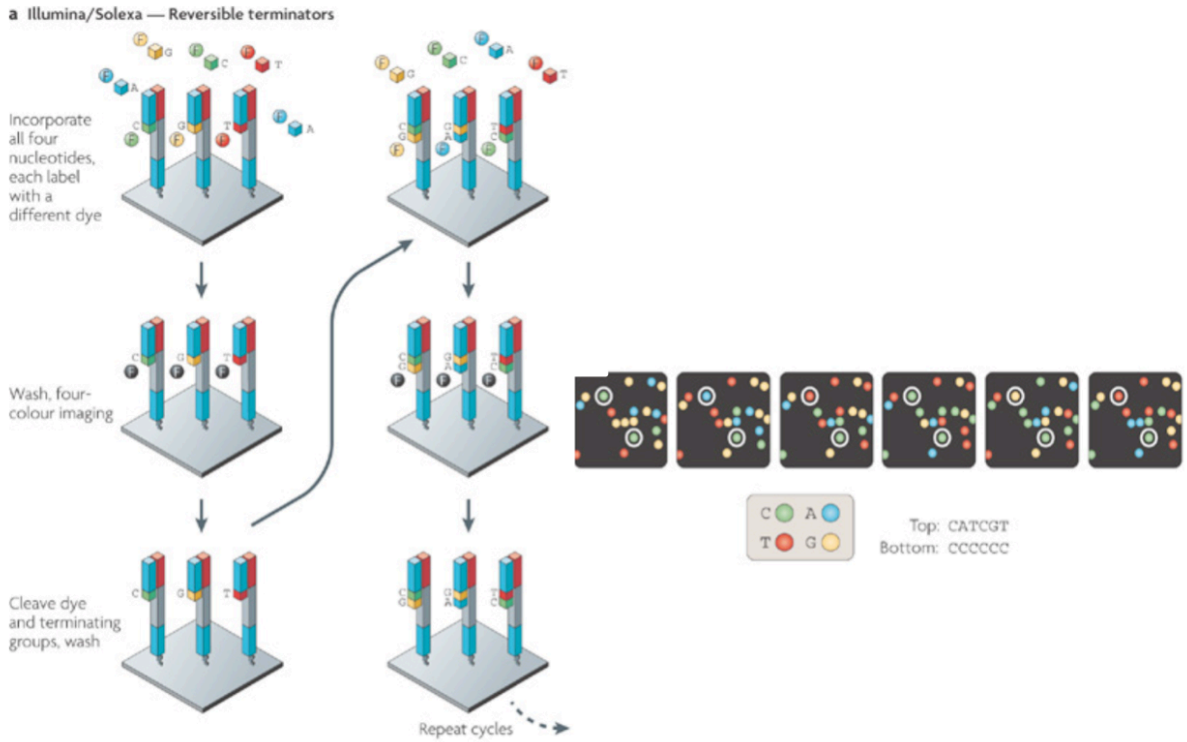


Figure 1.3 (Mardis, 2008): basic graphical illustration of the first cycle of Illumina sequencing. To the right six images taken after each individual cycle, with two examples below on how the base sequence is determined.

The recorded signal becomes more ambiguous (more than one color in one clonal cluster) and thusly of lower quality. The quality is registered for each base together with the sequence of one clonal cluster. (Knief, 2014; Zoll *et al.,* 2016; Xiao *et al.,* 2014) Illumina Miseq allows the sequencing from both ends of the amplicon. For each individual forward read a corresponding reverse read is assigned. This is achieved through introduction of specific forward and reverse primer annealing sites and matching indices to match the reads after sequencing. Paired-end reads are used to improve overall sequence quality of amplicons. The resulting paired-end reads are reads that were sequenced from both ends of the same amplicon. This makes the studies of microbial communities independent from cultivation and eliminates the clone bias.

The ca 1.5 kb-long 16S rRNA gene encoding rRNA of the small ribosomal subunit is relatively conserved within bacterial genera, and is often used for the taxonomic identification. This gene contains nine different hypervariable regions (V1-V9), providing species-specific signature sequences, flanked by highly conserved ones, which can be used for design of PCR primers with near universal bacterial specificity. Even if less informative than full-length 16S rDNA sequences, the higher coverage of shorter reads per sample or many samples per run compensate for this downside. The 16S rRNA gene is used as a marker gene sequence to determine the taxonomy of bacteria based on the similarity of the sequence to a taxonomically classified 16S gene. Taxonomic marker genes can furthermore be used to build phylogenetic trees. Phylogenetic trees are branching diagrams showing the relationships among marker genes and are used to illustrate the relatedness of the taxonomic marker genes to each other. In case of unrooted trees no ancestral sequence is needed or inferred. Reference trees are phylogenetic trees in which annotated taxonomic marker genes are introduced to determine the relatedness of unknown marker genes to these. The great challenge posed to the exponentially growing nucleotide databases is correctly placing the vast amounts of unclassified novel taxa, based on short reads, accurately into a given reference tree, without cultivated annotated representatives. (Singer *et al.*, 2016; Xiao *et al.,* 2014)

The internal transcribed spacer (ITS) region is used as barcode for the classification of fungi to the genus level, analogous to the 16S rDNA for bacteria. (Porras-Alfaro *et al.,* 2014) The length of the entire fungal ITS region is approximately 600-750 bp long. In there, two hypervariable regions are located (ITS1, ITS2) which flank the highly preserved 5.8S rDNA. The ITS region has been used according to the Consortium for the Barcode of Life for the estimation and taxonomic identification of fungal communities and uncultured taxa. (Zoll *et al.,* 2016; Deshpande *et al*., 2016)

The generated sequences of the ITS and 16S rDNA regions are processed *in-silico* with different bioinformatics tools. The usual format of these reads are *.fasta (without the quality of base calling) and *.fastq (with the quality of base calling) files as well as *.bam and *.sam files.

1.4.3. Bioinformatics tools used for sequence analyses

The improvements in DNA sequencing methodologies arrived before the ability to comprehensively analyze the generated data, making bioinformatics a bottleneck in microbiome studies. The general problem with large datasets is the computational time and power needed to analyze them. The computational resources do not increase in performance as fast as the data generation, computational power to solve a problem increases in a constant proportion and sequence data sizes increase exponentially.

Studies gathering genomic information on microbial communities unravel microbiomes with the use of phylogenetic/taxonomic marker genes (e.g. 16S rRNA gene). The produced sequence reads (raw sequencing result for one DNA target) have to be filtered (discarding sequence reads of overall low base quality), assembled (grouping unambiguously overlapping reads into contiguous sequence), clustered together (to create operational taxonomic units OTUs), chimeras have to be eliminated and the taxonomy has to be identified. Chimeras superficially increase the diversity of a bacterial community; they are hybrid products between multiple parent sequences created by the polymerase in the PCR-steps. (Haas *et al.,* 2011) OTUs are clusters of DNA sequences variants from a specific taxonomic marker gene, based on similarity, such as the 16S V3-V4 region. Every cluster (OTU) represents a taxonomic unit of a species or genus depending on the similarity threshold. Sequences are usually clustered to an OTU defined by a similarity threshold of 97%, if a sequence is 97% or above similar to an OTU then it is clustered to it. To increase the resolution of the OTUs a higher threshold can be adopted of 98% or 99%. However OTUs do not represent real sequences but pragmatic representations of sequences found in the data, in other words they approximate reality. Therefor to make OTUs comparable between different samples the sequences from all samples have to be pooled in one dataset. During the creation of OTUs it is possible to create sequences that are 97% similar to the sequences found in the data, but are not present as such. In other words, spurious OTUs can be created, that do not exist as exact sequence representation in the data. A large variety of tools exist for each step of sequence analyses, implemented in different software and evaluated by numerous studies (BLAST, UCLUST, USEARCH, CD-HIT, QUIIME, mothur, DADA2, Panda, FLASH, CAMSA …etc.). (Zhang *et al.,* 2013; Magoc *et al.,* 2011; Schloss *et al*., 2009; Edgar, 2017; Edgar *et al.,* 2015; Callahan *et al.* 2016; Lemos *et al.* 2017)

The bioinformatics tools used to analyze the amplicon sequencing data in this study are the state-of-the-art and have been applied and assessed in different studies (Probandt *et al*., 2017; Mhlongo *et al*., 2018):

1. FLASH v.2.2.00 to assemble the paired end reads. FLASH (Fast Length Adjustment of Short reads) is a computational tool to correctly overlap and merge paired-end reads. (Zhang *et al.,* 2013; Magoc *et al*., 2011)

2. Mothur v.1.39.5 (Schloss *et al*., 2009) is a comprehensive software package that gives the possibility to analyze data from NGS sequencing from different microbial communities. The results of its pipeline are sequences aligned to Operational Taxonomic Units (OTUs).

3. USEARCH v.10.0.240 (Edgar, 2017; Edgar, 2017; Edgar *et al.,* 2015) is a unique sequence analysis tool, offering search and clustering algorithms, with a very high throughput. This fast and robust software package results in sequences that are aligned to OTUs, which are taxonomically classified, similar to the mothur pipeline.

4. DADA2 v.1.8, is a software package (R Core Team, 2018) that creates amplicon sequence variants ASVs from sequencing data. ASVs are exact representations of the sequences found in the data. Therefor ASVs from different samples can be compared without the need of pooling the sequences from these samples. Furthermore ASVs can be used to resolve the taxonomy to species level because they represent exact sequences found in the data. DADA2 has the big advantage of inferring sequences exactly without introducing impurities into spurious OTUs and therefore resolving differences between sequences down to one nucleotide. DADA2 produces more real variants ASVs than other methods. (Callahan *et al.* 2016) The crucial advantage is that it uses more of the information within the data, incorporating error models from quality information, quantitative abundance and making exact sequences comparable between samples, whereas OTUs become only comparable if the samples are pooled together during OTU construction (because of the 97% similarity threshold during clustering). Consequently DADA2 can individually analyze each sample, increasing its performance and accuracy in comparison with mothur and USEARCH. Resolving partially the problem with very large data sets.

5. The RDP-classifier (Ribosomal Database Project) is a naïve Bayesian classifier used to accurately provide taxonomic assignments for sequences (for example 16S amplicon sequences), with confidence estimations for each assignment. A naïve Bayesian classifier uses the probability theory to classify sequences to a given taxonomical sequence and its algorithms make use of Bayes's theorem used for machine learning. Therefor the probability of an accurate classification increases with the size of the database used for classification. It

is a conglomerate of different machine learning algorithms that use statistical independence. (Wang *et al*. 2007)

These bioinformatics tools are used to analyze microbiomes such as those associated with plants. To analyze the low abundance endophytes within plants the epiphytes and latant pathogenic microorganisms have to be eradicated. This is possible via *in-vitro* plants, which are considered aseptic but may carry latent infections and endophytes. Endophytes are microorganisms that live within plants, without causing disease symptoms. (Orlikowska *et al.,* 2017) These endophytic microorganisms can become visible after many subculturing steps and appear only after certain stress factors such as cryopreservation. (Thomas *et al.,* 2008) Cryopreservation can be used for different purposes other than germplasm conservation and long-term storage. One such biotechnological use is cryotherapy, a rather recent development for eliminating viral pathogens and one bacterial pathogen (*Candidatus Liberobacter asiaticus*) in infected plants. (Engelmann, 2004; Ding *et al.,* 2008) Therefore changes in the microbiome of *in-vitro* plants undergoing cryopreservation could be expected. Until now, only cultivable bacteria and fungi from *in-vitro* plants were identified and characterized. In this study NGS (next generation sequencing) was applied to analyze the microbiome of *in-vitro* plants for the first time.

2. Aim of work

The aim of this study was to find out which bacteria and/or fungi are present in *in-vitro* plants before cryopreservation and after cryopreservation, were some bacteria or fungi lost, acquired and were there changes in their abundance.

The studied plants were *Artemisia laciniata* Willd. (Compositae / Asteraceae), *Mentha aquatica* L. (Lamiaceae) and *Solanum tuberosum* L. subsp. tuberosum L. (Solanaceae). *Artemisia laciniata* was chosen because it is a highly endangered species in Europe. *Mentha aquatica* was chosen because it hybridizes with *Mentha spicata* to produce healthy medicinally important *Mentha x piperita*. *Solanum tuberosum* was chosen because it holds a paramount position in agriculture and nutrition of the human population. Furthermore, the study-organisms belong to three different plant families giving a wide perspective on the addressed questions.

3. Materials and methods

3.1. Collection of micropropagation and cryopreservation samples

Micropropagation step: *In-vitro* plants from three different plant species, *Artemisia laciniata* Willd. (Compositae / Asteraceae), *Mentha aquatica* L. (Lamiaceae) and *Solanum tuberosum* L. (Solanaceae) before and after undergoing cryopreservation (Figure 1) were sampled. The *S. tuberosum* and *M. aquatica* samples were micropropagated at Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung IPK Gatersleben in Germany and *A. laciniata* at the University of Vienna. Stock plants were already established *in-vitro* at the onset of this study. *M. aquatica* and *S. tuberosum* had been vegetatively derived from one individual mother plant, *A. laciniata* on the other hand originated from one seed to ensure working with individual clones.

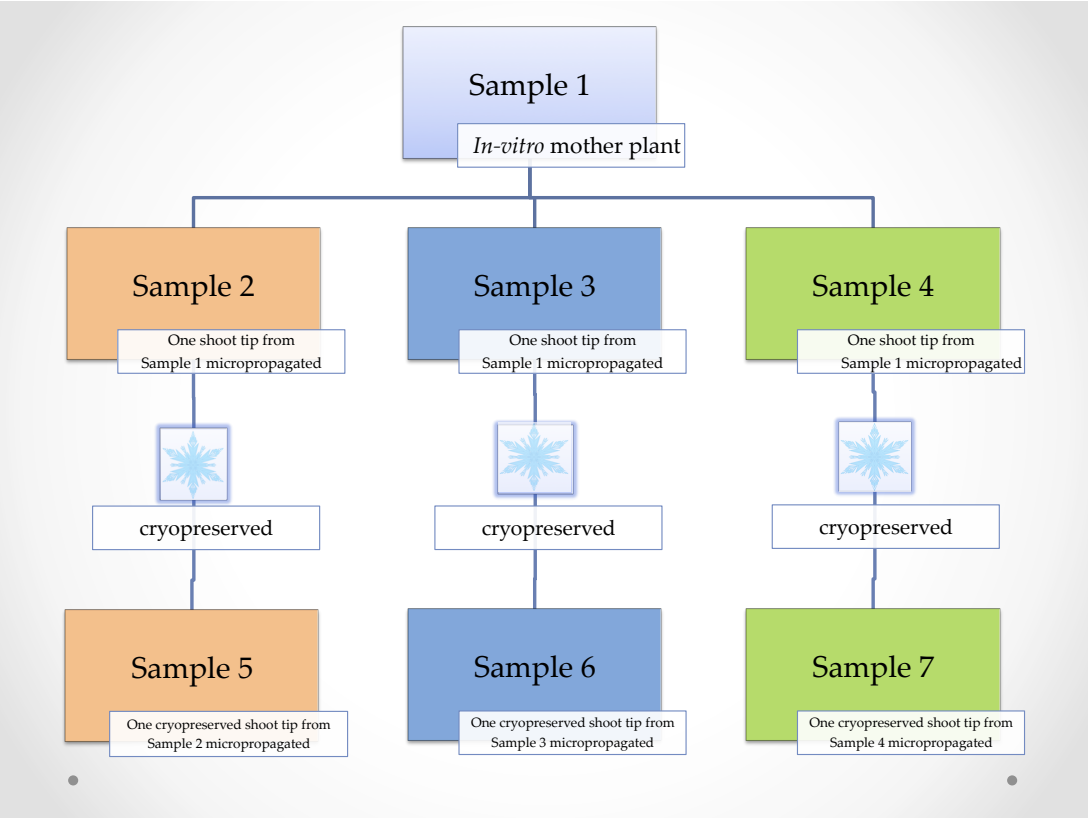


Figure 1 Procedure for the micropropagation and DNA sampling applied to *A. laciniata, S. tuberosum* and *M. aquatica*.

A. laciniata cultures were routinely subcultured every 5 to 8 weeks on MS-basal medium (Murashige T. *et al.*, 1962) with 0.5 μM BAP and 3% w/v sucrose in Magenta GA-7 plant culture boxes (Sigma-Aldrich) with Magenta B-caps (Sigma-Aldrich). (Kodym A. *et al.*, 2018) The plants were cultivated under an alternating temperature regime of 22/8°C and a 16 h photoperiod under 45 μmol m^{-1} s^{-1} produced by fluorescent lamps (Sylvania Grolux T8 58W). *S. tuberosum* cultures were subcultured every 4-6 weeks and maintained on MS-medium with 2% w/v sucrose, without vitamins

at 25°C and a 16h photoperiod (60 μmol m^{-1} s^{-1}). *S. tuberosum* underwent cold hardening at 5°C for one week prior to cryopreservation. *M. aquatica* cultures on MS-medium with 3% w/v sucrose underwent subculturing every 4-6 weeks at 25°C with cold hardening under an alternating temperature regime of 25/-1°C for 2 weeks prior to cryopreservation.

Cryopreservation step: Once sufficient material of sample 1 was available and showed no apparent bacterial growth, shoot tips were isolated. After further propagation, shoot tips were isolated (samples 2-4) and individually cryopreserved. The cultures were visually checked for contamination on a regular basis. The droplet vitrification method (Senula A. *et al*., 2007; Kodym A. *et al*., 2018) was taken as a standard.

For *A. laciniata* little shoot tips (2mm in length and one node) were excised and precultured overnight on filter paper in liquid MS medium containing 0.3 M sucrose at 25°C. After which they were transferred onto a filter paper with 2 ml loading solution (2 M glycerol + 0.4 M sucrose) for 20 min followed by incubation in 3 ml plant vitrification solution 3 (PVS3) (50% w/v sucrose + 50% w/v glycerol in liquid basal culture medium) (Nishizawa S. et al., 1993) for 2 h. Shoot tips were transferred into 4 μl droplets of PVS3 on aluminum foil stips. One strip with 1 droplet was placed in a cryovial (order ID 72.380 from Sarstedt, Nürnbrecht, Germany) and plunged into liquid nitrogen (-195.79°C). After a minimum of one hour in the container (model 26B, from Roth, Karlsruhe, Germany) the vials were rewarmed by submersing them into a water bath al 40°C for 40 s. The foil strips were then removed under sterile conditions in the laminar flow bench and the shoot tips incubated in 6 ml washing solution (1.2 M sucrose) for 20 min. Afterwards the tips were transferred to 60 mm Petridishes (10 explants per Petridish) with basal medium with 0.5 μM BAP and cultured in the dark at 25°C until regrowth appeared before being gradually moved to light. All solutions were prepared with liquid MS, pH adjusted to 5.8 and were autoclaved. Shoot tips with regrowth were singularly placed in Magenta GA-7 plant culture boxes (Sigma-Aldrich) on MS-basal medium with 0.5 μM BAP with Magenta B-caps (Sigma-Aldrich) for further cultivation, as described before. This was performed at the University of Vienna in Austria. The *S. tuberosum* and *M. aquatica* samples were prepared at IPK Gatersleben in Germany.

For *S. tuberosum,* shoot tips with the approximate length of 1 – 2 mm were excised and precultured for 20 to 24 h at 25°C in liquid MS medium containing 10% w/v sucrose (0.3 M) in petri dishes, with 2 ml liquid on filter paper. The shoot tips were transferred onto a filter paper with 2 ml loading solution (2 M glycerol + 0.4 M sucrose) for 20 min, followed by incubation in 2 ml plant vitrification solution 3, for 2 h at room temperature (30% w/v glycerol, 15% w/v ethylene glycol, 15% w/v dimethyl sulfoxide and 13.7% w/v sucrose in MS liquid medium). Shoot tips were transferred into 2 μl droplets of PVS3 on aluminum foil stips. Each strip had one droplet with one shoot tip each and was placed in individual cryovials and placed into liquid nitrogen for a minimum of one hour. The

vials were rewarmed by submersing them into a water bath at 40°C for 10 – 20 s. The foil strips were then removed under sterile conditions in the laminar flow bench and the shoot tips incubated in 6 ml washing solution (1.2 M sucrose) for 20 min. The washing steps were performed in multiwell plates to avoid cross contamination. The recovery medium used was basal MS medium with 30 g/L sucrose, 2.28 μM zeatin riboside, 2.85 μM IAA and 0.57 μM GA3 solidified with 10 g/l agar. Then they were cultured in the dark at 25°C until regrowth appeared (1 week) before moving them under low light (30 – 35 μmol m^{-1} s^{-1}).

For *M. aquatica* shoot tips of length ~1 mm were excised from axillary buds, as described by Senula *et al*. (2007) and precultured for 20 to 24 h at 25°C in liquid MS medium containing 10% w/v sucrose (0.3 M) in petri dishes, with 2ml liquid on filter paper. The shoot tips were transferred onto a filter paper with 2ml loading solution (2 M glycerol + 0.4 M sucrose) for 2 h, followed by incubation in 2 ml plant vitrification solution 2, for 20 min at room temperature. Shoot tips were transferred into 2 μl droplets of PVS3 on aluminum foil stips. Each strip had one droplet with one shoot tip each and was placed in individual cryovials and placed into liquid nitrogen for a minimum of one hour. The vials were rewarmed by submersing them into a water bath at 40°C for 10 – 20 s. The foil strips were then removed under sterile conditions in the laminar flow bench and the shoot tips incubated in 6ml washing solution (41% w/v sucrose) for 20 min. in multiwell plates. The recovery medium used was basal MS medium containing 3% w/v sucrose, 2.28 μM 6-(γ , γ -Dimethylallylamino)purine and 0.53 μM 1-Naphthaleneacetic acid (NAA) (pH 5.8). Then they were cultured in the dark at 25°C until regrowth appeared (2 days) before moving them under low light (30 – 35 μmol m^{-1} s^{-1}).

Sampling: When the various samples had been repeatedly subcultured so that sufficient material for DNA isolation was available it was removed from the glass jars under sterile conditions, weighed and placed in sterile 50 ml "Sterilin™" Eppendorf tubes. These closed tubes were then plunged in liquid nitrogen and kept there for at least one hour. After this step, the tubes were opened again and a sterile microfiber polyester net fabric was placed over the opening and fixed with a rubber band . The plant material in the tubes was lyophilized over night (for a minimum of 12 hours). After this the net fabric was removed and the tubes closed with their individual caps. The tubes were then stored in zip-lock bags filled with silica gel in a freezer at -20°C until further use.

3.2. Sample preparation

DNA-free sterile glass-beads were prepared to be able to grind the material in an adequate size for further DNA extraction. The glass-beads were filled in an Schott bottle (500 ml) and washed 3 times with hand warm deionized water (mixed well each time). Afterwards the bottle was filled with warm 10% Sodium-hypochlorite solution. The bottle was closed and after 20 min the solution was discarded and the content washed again 3 times with deionized water. Then the bottle with the washed glass-beads was autoclaved. These now DNA-free sterile glass-beads were placed in a sterile petri dish in the laminar airflow workbench for further drying before being transferred into the tubes with the lyophilized plant material. In the laminar flow, 3-5 glass-beads were placed in each tube. The glass-beads needed to be completely dry before adding them to the plant material, to avoid reactivation of enzymes during the grinding process. The tubes were vortexed for 1-5 min depending on the individual properties and the amount of the plant material. The more material was placed into a single tube, the longer the vortex process took (up to 5 min for a fully packed tube). If this was not enough for grinding the material to a fine powder the material was divided into new tubes (under sterile conditions) and vortexed again. The tubes were then stored in zip-lock bags filled with silica gel in a freezer at -20°C until further use.

3.3. DNA Extraction, purification and quality control

Under sterile conditions 3 ZR BashingBead™ Lysis Tubes (0.1 mm & 0.5 mm) from the Kit "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research were filled with 40 mg of grinded plant material per sample. Thus having three repetitions of each sample for DNA-extraction. Sterility was achieved by placing an appropriate scale into a sterile workbench, using autoclaved weighing paper and sterile utensils to transfer the material from the 50 ml tubes to the Lysis tubes from the Kit.

For the following steps it was of utmost importance to clean the used pipettes thoroughly with 70% Ethanol. Also all possible steps were performed in the laminar airflow workbench to avoid contamination of the samples, such as a fine-dust mask according to EN 149:2001 FFP1 was worn filtering dusts and aerosols up to 4 times the maximum workplace concentration (MAK value). Before beginning with the DNA Extraction all solutions and materials used therein were prepared, to minimize unnecessary waiting time and to maximize the efficiency of the DNA-extraction.

The "Genomic Lysis Buffer" was prepared in a ventilated fume hood. As a strong reducing agent, beta-mercaptoethanol was used to disrupt disulfide bonds and inactivate enzymes, such as desoxyribonuclases that would digest the DNA further during the extraction process (PubChem;

CID=1567, https://pubchem.ncbi.nlm.nih.gov). In the sterile workbench 200μl refrigerated deionized autoclaved water (4°C) and 750μl "Lysis Solution" from the Kit was added to each Lysis tube. After this step the samples were put in an icebox. The samples were then processed in the Precellys 24 tissue homogenizer from Bertin Tech. (France) (Figure 3) with following cycle specifications: 4200 rounds per minute (rpm) for 10 seconds with a cool down period of 30 seconds. The cycle was repeated 3 times and during the cool down period the samples were placed in an icebox and kept there for 5 minutes assuring that the samples did not reach a temperature above 30°C. This was checked with an IR-laser-thermometer. The measured Temperatures fluctuated between 4°C and 27°C (after one cycle). If the samples would reach 37°C or above this would reactivate enzymes and could cause a strong degradation of the DNA, most enzymes have a temperature optimum above 30°C (Peterson M. et al., 2007; Robinson P., 2015). The following extraction was preformed with the "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research (see Appendix 'DNA extraction'). The extracted DNA was cleaned with the "*DNA Clean & Concentrator* ™ – 5 Kit" from Zymo Research to obtain high quality DNA required for Illumina sequencing (see Appendix 'DNA purification').

After extraction and purification of the DNA its concentration was measured with a Nanodrop lite in ng μl^{-1} (Figure 4). The absorbance at 260 nm is used to calculate the concentration of the DNA. At A_{260} = 50 the concentration of dsDNA is 1 μg ml^{-1} with a path length of 1cm.

As a second step, analytical gel electrophoresis was performed to evaluate the degree of DNA degradation. To make the DNA visible during UV-light exposure with the GelDoc™, a device for the analysazions of gels, an intercalating fluorescent nucleic acid stain "Gel red" (IUPAC: 5,5'-(6,22-dioxo-11,14,17-trioxa-7,21-diazaheptacosane-1,27-diyl)bis(3,8-diamino-6-phenylphenanthridin-5-ium) iodide) was added directly to the agarose gels at preparation. (Huang Q. *et al.*, 2010) After a gel was cast and got solid, the comb was removed and the samples could be applied. First a reference 1 kb (kilo base) ladder was loaded in the first well of the 0,8% agarose gel then the sample DNA in the neighboring. The gels were run for 40 min at 100 V in TBE buffer (1x). (see Appendix 'gel-electrophoresis')

The specific cycling conditions used for the PCR targeting the 16S rRNA gene were: initial denaturation at 95°C for 2 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 53°C for 1 min, and extension at 72°C for 1 min; a final extension was preformed with 72°C lasting 10 min. Each PCR tube contained 30 μl deionized autoclaved water, 4 μl buffer "standard Taq buffer", 2 μl DMSO, 1 μl 799F forward primer (10 μM), 1 μl 1391R reverse primer (10 μM), 0,5 μl dNTP, 0,5 μl Taq polymerase and 1 μl sample DNA.

The specific cycling conditions used for the PCR targeting the ITS region were: initial denaturation at 95°C for 5min, followed by 40 cycles of denaturation at 95°C for 30sec, annealing at 50°C for 15sec, and extension at 72°C for 3min; a final extension was preformed with 72°C lasting

10min. For each sample mixture containing 30 µl deionized autoclaved water, 4 µl buffer "standard Taq buffer", 2 µl DMSO, 1 µl ITS 1-F forward primer (10 µM), 1 µl ITS 4 reverse primer (10 µM), 1 µl dNTP, 1 µl Taq polymerase was prepared. 39 µl of this master mix were placed in a PCR tube and 1 µl sample DNA was added. A gel electrophoresis was performed to evaluate the results of both PCRs.

Table 1.6: Specifications of the used primers for PCR controls.

Primer	**Primer sequence 5'-3'**	**Region / Position**	**References**
799F	AACMGGATTAGATACCCKG	V5-V6-V7	Chelius M. et al. 2001
1391R	GACGGGCGGTGWGTRCA	V5-V6-V7	Walker J. et al. 2007
ITS1-F	CTTGGTCATTTAGAGGAAGTAA	1723-1744	Gardens M. et al. 1993
ITS4	TCCTCCGCTTATTGATATGC	2390-2409	White T. et al. 1990

After passing these quality controls, the DNA was ready to be sequenced. 50 µl extracted DNA of each sample was shipped in a UV-sterilized 96 well plate at ambient temperature to Eurofins Genomics in Germany. The minimum concentration required was 50 ng μl^{-1}. The plate was sealed using a cap stripe. The ordered product was called "Microbiome Profiling by analysing 2 target regions".

3.4. Illumina MiSeq – facility

The sequencing was performed at Eurofins genomics using Illumina MiSeq according to the company's internal protocols. The specific primers used by the company are listed in Table 1.7. The sequenced amplicons were delivered as FASTQ files as a download with forward and reverse reads in separate files.

Table 1.7: Specifications of the used primers for the two-step PCR protocol performed by Eurofins genomics.

Primer	**Primer sequence 5'-3'**	**Region / Position**	**References**
forward 16S primer	CTACGGGNGGCWGCAG	V3-V4	Klindworth A. *et al.* 2013
reverse 16S primer	GACTACHVGGGTATCTAATCC	V3-V4	Klindworth A. *et al.* 2013
ITS3_KYO2	GATGAAGAACGYAGYRAA	2029-2046	Toju H. *et al.* 2012
ITS4	TCCTCCGCTTATTGATATGC	2390-2409	White T. *et al.* 1990
forward Illumina adapter sequence	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-forward primer target sequence		
reverse Illumina adapter sequence	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-reverse primer target sequence		

3.5. Bioinformatics tools

In a first step, FLASH v2.2.00 was used to assemble the paired end reads. (Zhang J. *et al.* 2013, Magoc T. *et al.* 2011) The tool mothur v.1.39.5. (Schloss P. *et al.* 2009) was used to create OTUs for the 16S V3V4 amplicons, USEARCH v.10.0.240 (Edgar R. C. 2017; Edgar R. C. 2017; Edgar R. C. *et al.* 2015) to create OTUs for the ITS2 amplicons and the DADA2 v.1.8 software package (Callahan B. J. *et al.* 2016) that models and corrects sequenced amplicon errors to create amplicon sequence variants ASVs for both the 16S and ITS amplicons.

3.6. Analysis pipeline

The computational resource used for analyzing the sequence data is located at the Life Science Compute Cluster (LiSC), which is part of the Computational Systems Biology Department of the University of Vienna. The LiSC is a medium-sized compute cluster, which keeps updated copies of relevant biological databases available for its users. It is equipped with an up-to-date bioinformatics software repository; therefore most software packages did not have to be installed prior to usage.

3.6.1. FLASH v.2.2.00

To assemble the forward and the reverse read the software FLASH v.2.2.00 (Zhang J. *et al.,* 2013; Magoc T. *et al.,* 2011) was used. The FLASH algorithm considers all possible overlaps at or above a minimum length between the paired reads and chooses the overlap that results in the lowest possible proportion of mismatched bases in the region, also known as mismatch density. (Zhang J. *et al.,* 2013; Magoc T. *et al*., 2011) FLASH furthermore computes a consensus sequence in the overlapping region by selecting the base with the higher quality value, breaking draws randomly. The minimum overlap size was set to 10 bp to reduce false-positive merges.

The dataset was uploaded via FileZilla v.3.32.0 (https://filezilla-project.org/) to the Life Science Compute Cluster (LISC) Cube Computational Systems Biology at the University of Vienna. The files were then uncompressed using `gunzip`. and sorted according to the sequenced region 16S V3-V4 and ITS2.

3.6.2. mothur v.1.39.5

Myself processed the raw 16S rDNA amplicons with mothur v.1.39.5. (Schloss P. *et al.* 2009) according to the MiSeq SOP (Probandt *et al.* 2017; Kozich *et al.* 2013) with modifications and adaptations for our own dataset, without external help. The specific method for the 16S V3V4 region is described in the appendix. The Sequences were classified using the non-redundant Ref SILVA database release 132. With these steps the output files to analyze the composure of the bacterial microbiome had been created:

```
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
fasta
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
opti_mcc.0.03.cons.taxonomy
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
opti_mcc.shared
```

They showed that certain OTUs were observed a definite number of times as well as the percentage of how well the sequences classified to this OTU. The Taxonomy is ordered as follows: domain-*regio*; pylum-*pylum*; class-*classis*; order-*ordo*; family-*familia* and genus-*genus*.

3.6.3. USEARCH v.10.0.240

Myself then continued data analysis with a different program for the ITS2 amplicons. The original sequences, after sorting and uncompressing with the software FLASH v.2.2.00 (Zhang J. *et al.,* 2013, Magoc T. *et al.,* 2011), were processed with USEARCH v.10.0.240 (Edgar R. C., 2017; Edgar R. C., 2017; Edgar R. C. *et al.,* 2015). (Hamad I. *et al*., 2017) The Sequences were classified using the UNITE reference database release 01.12.2017. The specific method for the ITS2 region is described in the appendix.

3.6.4 DADA2 v1.8

Both amplicons, 16S V3V4 Region and ITS2, were processed and analyzed following the DADA2 pipeline with minor adaptations for our dataset. (Callahan B. J. *et al.,* 2016) The input sequences were the de-multiplexed paired end reads as fastq files. The final table did not contain OTUs, as in MOTHUR or USEARCH, but amplicon sequence variants (ASVs). The commands were performed in R version 3.5.0 (R core Team, 2018). The specific method for the creating the ASVs is described in the appendix.

The databases used to assign a taxonomy to the OTUs and ASVs were for the 16S amplicons the RDP 16S rRNA training set 16 (Wang Q. *et al*. 2007), the SILVA database (Quast C. *et al.*, 2012) and the rRNA_typestrains/prokaryotic_16S_ribosomal_RNA database (NCBI Resource Coordinators, 2012). It reached down to genus level for the 16S V3V4 region within the DADA2 pipeline. The 16S ASVs were identified again with the RDP Naïve Bayesian rRNA Classifier version 2.11 to the RDP 16S rRNA training set 16 and a confidence threshold of 85% as well as using the BLASTn 2.2.29+ algorithm with the rRNA_typestrains/prokaryotic_16S_ribosomal_RNA database. The ITS amplicons were assigned to a taxonomy by using the UNITE database (UNITE Community, 2017) and the Warcup Fungal ITS trainset 2 (Deshpande V. *et al*., 2016). The ITS2 ASVs were classified with the UNITE database general release 01.12.2017 within the DADA2 pipeline. The ITS2 ASVs were re-identified with the RDP Naïve Bayesian rRNA Classifier Version 2.11 to the Warcup Fungal ITS trainset 2 and the UNITE Fungal ITS trainset with a confidence threshold of 85% and using the BLASTn 2.2.29+ algorithm in the ultra condensed output mode. The ASVs were assigned to a specific taxonomy by aligning them to individual databases within the DADA2 pipeline. The SILVA reference database v128 train set was used to assign the ASVs from the 16S V3V4; the UNITE database v.7.2 was used for the ITS2 ASVs. This outputted a general taxonomy, to confirm this assignment the ASVs were aligned again with a different assignment tool to different databases.

The 16S ASVs were aligned to the RDP 16S rRNA training set 16 using the RDP Naïve Bayesian rRNA Classifier version 2.11 and to the rRNA_typestrains/prokaryotic_16S_ribosomal_RNA database using the BLASTn 2.2.29+ algorithm. (Wang Q. *et al*. 2007) Where the taxonomy varied from each other the taxonomy with the higher % identity was chosen. 88.63% were classified as mitochondria or chloroplasts and were consequently removed from the dataset.

The ITS2 ASVs were aligned to the Warcup Fungal ITS trainset 2 using the RDP Naïve Bayesian rRNA Classifier and the UNITE Fungal ITS trainset using the BLASTn 2.2.29+ algorithm. The classifications led to the same taxonomical assignment of the ASVs. As for the 16S ASVs using the different databases had the aim to assign taxonomy with high accuracy and confidence.

Further analysis and visualizations were done in R v.3.5.0 with the package phyloseq v.1.24.2, ggplot2

v3.0.0, readxl and ape v.5.2. See Appendices for further details.

4. Results

In-vitro plants from three different plant species, *Artemisia laciniata* Willd. (Compositae / Asteraceae), *Mentha aquatica* L. (Lamiaceae) and *Solanum tuberosum* L. (Solanaceae) before and after undergoing cryopreservation (Figure 1.1) were sampled. The *S. tuberosum* and *M. aquatica* samples were micropropagated at Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung IPK Gatersleben in Germany and *A. laciniata* at the University of Vienna. Stock plants were already established *in-vitro* at the onset of this study. *M. aquatica* and *S. tuberosum* had been vegetatively derived from one individual mother plant, *A. laciniata* on the other hand originated from one seed to ensure working with individual clones.

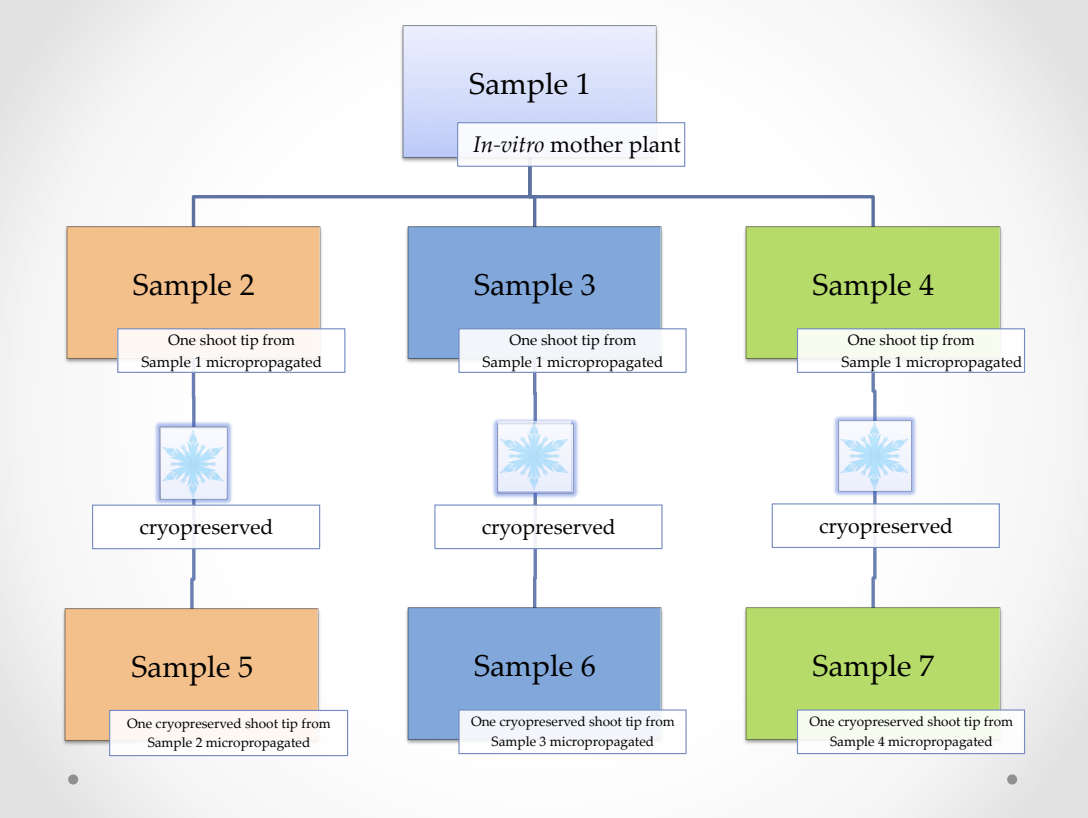


Figure 1.1 Procedure for the micropropagation and DNA sampling applied to *A. laciniata, S. tuberosum* and *M. aquatica*.

When the various samples had been repeatedly subcultured so that sufficient material for DNA isolation was available it was removed from the glass jars under sterile conditions, lyophilized, weighed and placed in sterile 50 ml tubes. Before grinding the dry material was weighed under sterile conditions (Table 1.1) to determine the amount of available material.

Table 1.1 Weight of lyophilized plant material per tube.

Sample	**Total lyophilized weight in mg**		
	Artemisia laciniata	*Mentha aquatica*	*Solanum tuberosum*
Sample 1	580 mg	220 mg	160 mg
Sample 2	460 mg	220 mg	120 mg
Sample 3	650 mg	230 mg	150 mg
Sample 4	470 mg	228 mg	145 mg
Sample 5	320 mg	250 mg	140 mg
Sample 6	910 mg	240 mg	160 mg
Sample 7	890 mg	223 mg	245 mg
Average	611.4 mg	230.1 mg	160 mg

4.1. DNA Extraction, purification and quality control

The DNA from the sampled *in-vitro* plant material before and after cryopreservation was extracted and purified. (Chapter 3.3; Appendix) The first step after extraction and purification of the DNA was to measure its concentration with Nanodrop lite in ng μl^{-1}. In the tables 4.1, 4.2, 4.3 and 4.4 the A_{260}/A_{280} ratio is used to indicate protein contamination and purity of the sample. Pure DNA samples have a ratio of greater than or equal to 1.8.

Table 4.1: The calculated concentration of the DNA in ng μl^{-1} measured for *Artemisia laciniata*. Further the A_{260}/A_{280} ratio and the absorbance at 260 nm. The second number 1.1, 1.2 and 1.3 stands for the three individual extractions from each sample.

Sample	ng μl^{-1}	A_{260}/A_{280}	A_{260}
A 1.1	133.1	1.8	2.662
A 1.2	100.8	1.8	2.016
A 1.3	167.2	1.8	3.344
A 2.1	91.9	1.8	1.838
A 2.2	106.7	1.8	2.134
A 2.3	97.1	1.79	1.942
A 3.1	97.1	1.8	1.943
A 3.2	84.2	1.8	1.684
A 3.3	107.4	1.8	2.148
A 4.1	101.5	1.8	2.031
A 4.2	267.5	1.72	5.35
A 4.3	115.4	1.77	2.309
A 5.1	119.7	1.81	2.393
A 5.2	115.9	1.8	2.318
A 5.3	94.7	1.8	1.894
A 6.1	101.8	1.8	2.037
A 6.2	93.3	1.84	1.865
A 6.3	110.2	1.85	2.203
A 7.1	115	1.8	2.3
A 7.2	109.9	1.79	2.198
A 7.3	114.1	1.8	2.283

Table 4.2: The calculated concentration of the DNA in ng μl^{-1} measured for *Mentha aquatica*. Further the A_{260}/A_{280} ratio and the absorbance at 260 nm. The second number 1.1, 1.2 and 1.3 stands for the three individual extractions from each sample.

Sample	ng μl^{-1}	A_{260}/A_{280}	A_{260}
M 1.1	109.9	1.81	2.199
M 1.2	113.2	1.81	2.264
M 1.3	104.5	1.81	2.089
M 2.1	128.8	1.8	2.575
M 2.2	124.4	1.81	2.487
M 2.3	115.6	1.8	2.312
M 3.1	118.2	1.81	2.364
M 3.2	124.4	1.8	2.488
M 3.3	103.4	1.81	2.069
M 4.1	129.4	1.81	2.587
M 4.2	112.2	1.8	2.244
M 4.3	180.4	1.79	3.609
M 5.1	113.8	1.8	2.276
M 5.2	124.9	1.8	2.497
M 5.3	119.3	1.8	2.386
M 6.1	118.8	1.81	2.376
M 6.2	132.1	1.81	2.641
M 6.3	120.5	1.81	2.41
M 7.1	113.4	1.81	2.269
M 7.2	118.9	1.81	2.379
M 7.3	122.6	1.8	2.452

Table 4.3: The calculated concentration of the DNA in ng μl^{-1} measured for *Solanum tuberosum*. Further the A_{260}/A_{280} ratio and the absorbance at 260 nm. The second number 1.1, 1.2 and 1.3 stands for the three individual extractions from each sample.

Sample	ng μl^{-1}	A_{260}/A_{280}	A_{260}
S 1.1	103	1.81	2.06
S 1.2	104.7	1.81	2.094
S 1.3	95.1	1.81	1.903
S 2.1	101.9	1.81	2.038
S 2.2	126.7	1.82	2.535
S 2.3	94.5	1.81	1.89
S 3.1	122.9	1.81	2.458
S 3.2	125.2	1.81	2.504
S 3.3	116.7	1.81	2.334
S 4.1	99.7	1.81	1.993
S 4.2	99.6	1.81	1.991
S 4.3	107.7	1.81	2.153
S 5.1	95.7	1.83	1.914
S 5.2	101	1.81	2.019
S 5.3	88.7	1.81	1.773
S 6.1	106.7	1.79	2.133
S 6.2	90.7	1.8	1.814
S 6.3	107.7	1.81	2.154
S 7.1	98.9	1.82	1.977
S 7.2	98	1.8	1.96
S 7.3	111.5	1.81	2.23

Table 4.4: The values measured after pooling (p) the three extractions from each sample. The concentrations of *Artemisia laciniata* (A), *Mentha aquatica* (M) and *Solanum tuberosum* (S).

Sample	ng μl^{-1}	A_{260}/A_{280}	A_{260}
A 1 p	130.9	1.78	2.617
A 2 p	109.6	1.79	2.132
A 3 p	93.7	1.82	1.873
A 4 p	119.4	1.8	2.389
A 5 p	118.7	1.79	2.374
A 6 p	104.2	1.8	2.085
A 7 p	98.5	1.82	1.971
M 1 p	117.1	1.8	2.341
M 2 p	134.3	1.8	2.686
M 3 p	123.5	1.8	2.471
M 4 p	139.4	1.8	2.789
M 5 p	129.1	1.8	2.583
M 6 p	135.1	1.81	2.702
M 7 p	126.6	1.8	2.533
S 1 p	105.6	1.81	2.072
S 2 p	96.9	1.81	1.937
S 3 p	129	1.81	2.58
S 4 p	104.8	1.81	2.096
S 5 p	95.4	1.81	1.907
S 6 p	105.9	1.8	2.118
S 7 p	106.8	1.81	2.136

All extractions yielded a high amount of DNA with at least 93.7 ng μl^{-1} and were of high quality. The second step was to check the extracted DNA via gel electrophoresis. The DNA was fragmented during the extraction process and lay mainly in a span of 8 kbp – 3 kbp. A 1kb (kilo base) DNA ladder was added in the first well. (Figures 5-9) A DNA-ladder is a set of DNA standards with defined lengths to identify the size of the sample DNA run parallel on a gel during electrophoresis. It uses the principal that molecular weight is inversely proportional to the length of the path completed through the gel. (Figure 9.1)

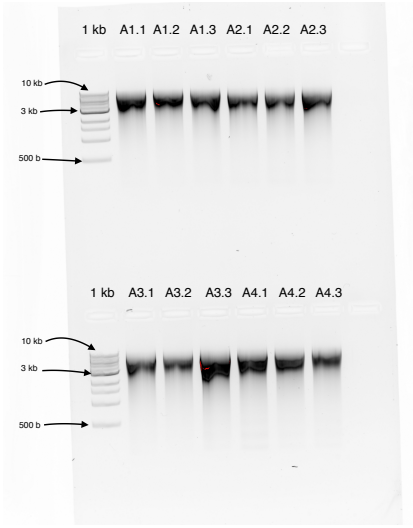


Figure 5 DNA of samples A1.1-A4.3

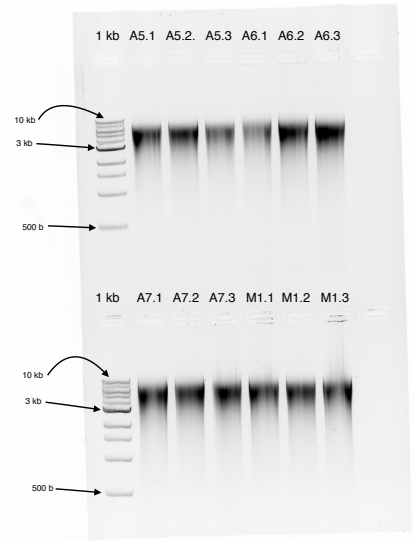


Figure 6 DNA of samples A5.1-7.3 and M1.1-1.3

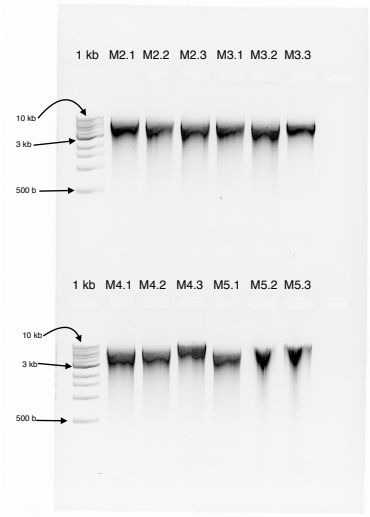


Figure 7 DNA of samples M2.1-5.3

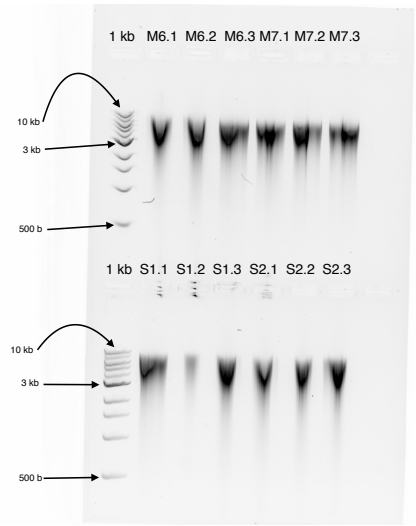


Figure 8 DNA of samples M6.1-7.3 and S1.1-2.3

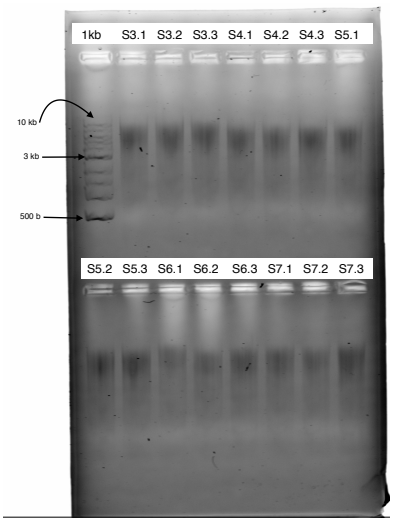


Figure 9 DNA of samples S3.1-7.3

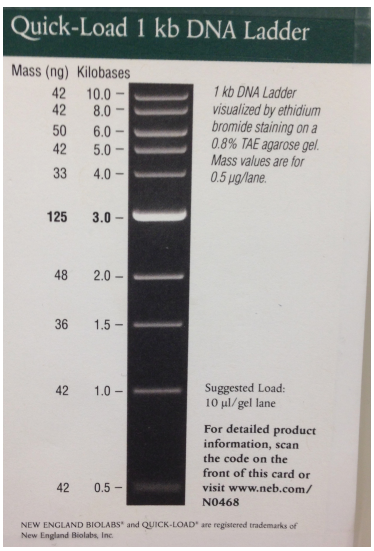


Figure 9.1 DNA-ladder used in all gels.

Next I performed PCR on these verified DNA samples. Two PCR reactions were conducted: The first targeting the 16S rRNA gene of bacteria. This was done to see if actually any bacteria were present in the *in-vitro* plants. The first problem arising was, that chloroplasts hold the same 16S rRNA gene, this biases under normal circumstances the outcome of this experiment. For that reason primers were used, that ideally reduce the contamination with chloroplast 16S rDNA. Beckers B. *et al*. (2016) have shown in their research that the usage of a 799F forward primer and 1391R reverse primer significantly reduces the co-amplification of chloroplast DNA (Table 1.6). These primers were used to amplify the 16S region of bacteria in the samples. The cycling conditions were first adopted from Beckers B. *et al*. (2016) and then adapted to yield best possible results in trial runs before being used with the actual samples. All samples had a positive amplification of the 16S rDNA with an expected 1kb amplicon size. (Figure 10) In the *M. aquatica* samples there were two to three bands visible.

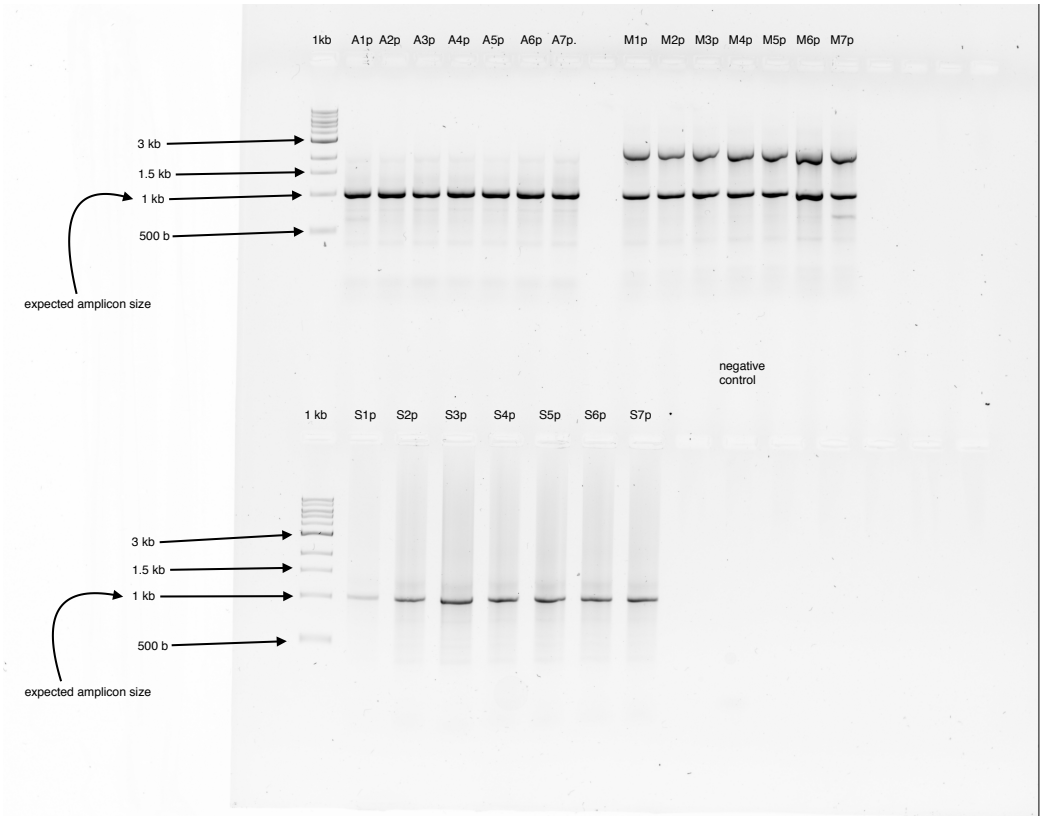


Figure 10 The PCR products of the 16S rDNA amplification with the primer combination799F and 1391R for samples A1p-S7p (p standing for pooled, meaning A1.1-1.3 pooled together to A1p)

The second PCR targeted the ITS region of ascomycetes and basidiomycetes, as well as giving an insight into general fungi. This was performed to see if there are any kinds of fungi present in the *in-vitro* samples. For the PCR ITS 1-F as forward primer and ITS 4 as reverse primer were used (Table 1.6). The PCR product results in representing an over span of the ITS1 region, the 5.8S region and the ITS2 region of the fungal DNA. (Toju H. *et al*., 2012) All samples showed a positive amplification of the ITS region with an expected 600 – 750 bp amplicon size. (Figure 11)

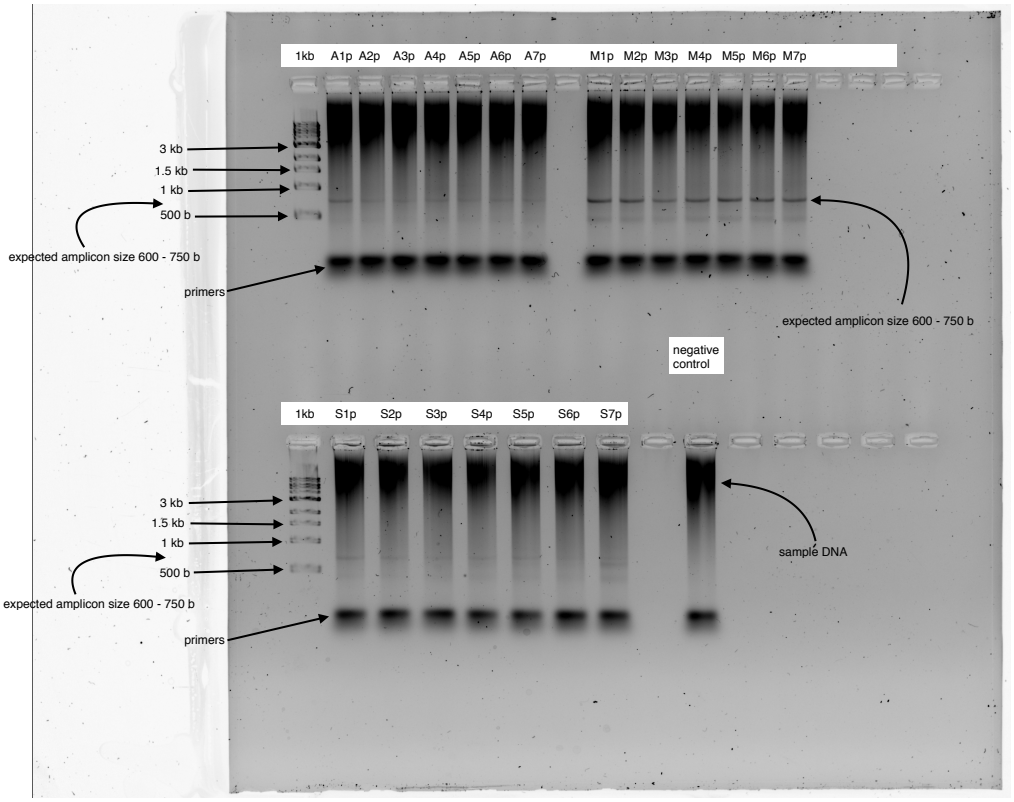


Figure 11 The PCR products of the over-span of the ITS1 region, the 5.8S region and the ITS2 region for each individual sample (A1p-S7p) performed with the primer combination ITS 1-F forward and ITS 4 reverse. The smear of high MW (above 3 kb) sample DNA is clearly visible because the gel was loaded with the whole amount of PCR product (5 μl) and the picture taken was darkened to make the amplicons visible.

After passing these quality controls, the raw DNA was ready to be sequenced. The samples were sent to a sequencing facility to obtain the sequences to analyze the microbiome of the *in-vitro* plants. 50 μl extracted DNA of each sample was shipped in a UV-sterilized 96 well plate at ambient temperature to Eurofins Genomics in Germany. The minimum concentration required was 50 ng μl^{-1}. The plate was sealed using a cap stripe. The ordered product was called "Microbiome Profiling by analysing 2 target regions".

4.2. Illumina MiSeq – received data

Table 4.5 and 4.6 provides various statistics of the sequenced reads from the 16S V3V4 amplicons and ITS2 amplicons from the extracted DNA originating from the individual *in-vitro* plant samples sorted by sample and amplified region. These reads passed the default Illumina filter procedure (chastity filter). Illumina MiSeq performs an internal quality filtering called chastity filter it is defined as the ratio of the clone cluster (Chapter 1.4.2) with the highest light intensity divided by the sum of the highest light intensity and second highest light intensity clusters. Clusters pass the filter if no more than 1 base call (recognizing a correct base in the sequencing process) has a chastity value below 0.6 in the first 25 cycles (the first 25 bases called). This process removes the reads with lowest quality from the analysis. "Yield (Kbp)": describes the number of bases called in kilobase. "%Q30": represents the percentage of bases with quality scores of at least 30, this infers a base call accuracy of 99.9%. The Q-score is a prediction of the likelihood of a wrong base call (recognizing a correct base in the sequencing process).

Table 4.5: Fastq processing results. 16S V3V4 amplicons. (p standing for pooled, meaning A1.1-1.3 pooled together to A1p)

Sample	Read Pairs	Yield (Kbp)	%Q30	Mean Q
A 1 p	105318	59925	86.17	34.18
A 2 p	97715	55599	86.80	34.37
A 3 p	76559	43562	86.71	34.33
A 4 p	83081	47273	86.06	34.15
A 5 p	91314	51957	88.68	34.97
A 6 p	65825	37454	86.07	34.15
A 7 p	82813	4712	86.34	34.23
S 1 p	68185	38797	87.38	34.58
S 2 p	63873	36343	87.20	34.52
S 3 p	98807	56221	87.18	34.50
S 4 p	82884	4716	87.33	34.55
S 5 p	85115	4843	87.56	34.61
S 6 p	126486	7197	86.79	34.39
S 7 p	78211	44502	87.85	34.69
M 1 p	93637	53279	87.20	34.52
M 2 p	81847	4657	86.32	34.27
M 3 p	83312	47404	87.65	34.66
M 4 p	78512	44673	87.86	34.71
M 5 p	80589	45855	87.12	34.51
M 6 p	53067	30195	88.22	34.81
M 7 p	64020	36427	87.25	34.55

Table 4.6: Fastq processing results. ITS2 amplicons. (p standing for pooled, meaning A1.1-1.3 pooled together to A1p)

Sample	Read Pairs	Yield (Kbp)	%Q30	Mean Q
A 1 p	90391	50618	86.15	34.24
A 2 p	85544	47904	86.04	34.22
A 3 p	109058	61072	86.16	34.25
A 4 p	163384	91495	85.43	34.03
A 5 p	78909	44189	86.08	34.22
A 6 p	117953	66053	86.38	34.32
A 7 p	98769	5531	86.11	34.25
S 1 p	76004	42562	81.28	32.79
S 2 p	118647	66442	81.18	32.75
S 3 p	84254	47182	81.33	32.81
S 4 p	62360	34921	80.88	32.69
S 5 p	63512	35566	81.17	32.75
S 6 p	78193	43788	80.30	32.49
S 7 p	79699	44631	81.22	32.75
M 1 p	59769	3347	79.32	32.19
M 2 p	83120	46547	79.96	32.38
M 3 p	80348	44994	79.15	32.14
M 4 p	65295	36565	79.63	32.29
M 5 p	75565	42316	79.13	32.12
M 6 p	95121	53267	79.69	32.30
M 7 p	119	66	40.94	18.62
Total sequencing run	3626065	2045608	84.08	33.56

4.3. Bioinformatics

I received the sequencing data from Eurofins genomics as a downloadable directory. (Chapter 4.2) This data needed to be analyzed to obtain insight into the microbiome of *in-vitro* plants. (Chapter 1.4) The bioinformatics tools used to analyze the data and gain this insight were:

- FLASH v.2.2.00 to assemble the paired end reads
- mothur v.1.39.5 to create OTUs for the 16S V3V4 amplicons
- USEARCH v.10.0.240 to create OTUs for the ITS2 amplicons
- DADA2 v.1.8 that models and corrects sequenced amplicon errors to create amplicon sequence variants ASVs for both the 16S and ITS amplicons

4.3.1. FLASH v.2.2.00

The results of the merging of the overlapping paired-end reads with FLASH v.2.2.00 are shown in table 4.7 and 4.8. Merging is needed to obtain overall high quality reads. On average 96.19% of all pair reads were merged together with a minimum overlap size of 10 bp, to reduce false positive merges. The mode length for the 16S V3V4 region amplicons resulted in 404 bp for all samples. The mode length is the length of the sequence, which appears most often in a group of sequences. The mode length is a better way to observe the amplicon size because the mean length is distorted by wrongly merged reads (to long or to short). The mode length of the ITS2 region amplicons showed a different length according to the derived DNA, 368 bp for *A. laciniata* derived DNA, 352 bp for *S. tuberosum* derived DNA and 381 bp for *M. aquatica* derived DNA. Already at this stage the sample M7 deviated from the other samples having the lowest number of total pairs for the ITS2 region. The ITS2 region is only a part of the ITS region and therefor shorter than the whole ITS region which consists of the ITS1 region, the 5.8rRNA gene and the ITS2 region but can still be used to discriminate the diversity of the fungal microbiome. (Hamad *et al.,* 2017; Chapter 1.4.2)

Table 4.7: Read merging with FLASH v.2.2.00. The total pairs combined in percent are shown next to the inputted pairs per sample. The mean and mode lengths is the length of the merged read pairs in bp. (p standing for pooled, meaning A1.1-1.3 pooled together to A1p)

Sample ITS2 amplicons	Total pairs	Percent combined	Mean lengths	Mode lengths
A 1 p	90391	98.3%	367	368
A 2 p	85544	98.3%	367	368
A 3 p	109058	98.3%	367	368
A 4 p	163384	98.1%	368	368
A 5 p	78909	98.3%	367	368
A 6 p	117953	98.5%	367	368
A 7 p	98769	98.4%	367	368
S 1 p	76004	97.4%	352	352
S 2 p	118647	97.3%	352	352
S 3 p	84254	97.4%	354	352
S 4 p	62360	97.4%	352	352
S 5 p	63512	97.3%	352	352
S 6 p	78193	97.1%	352	352
S 7 p	79699	97.2%	352	352
M 1 p	59769	95.8%	381	381
M 2 p	83120	96.2%	379	381
M 3 p	80348	96.0%	381	381
M 4 p	65295	95.9%	381	381
M 5 p	75565	95.7%	381	381
M 6 p	95121	95.8%	381	381
M 7 p	119	38.5%	368	368

Table 4.8: Read merging with FLASH v.2.2.00. The total pairs combined in percent are shown next to the inputted pairs per sample. The mean and mode lengths is the length of the merged read pairs in bp. (p standing for pooled, meaning A1.1-1.3 pooled together to A1p)

Sample 16S amplicons	Total pairs	Percent combined	Mean lengths	Mode lengths
A 1 p	105318	96.9%	405	404
A 2 p	97715	97.2%	403	404
A 3 p	76559	96.9%	403	404
A 4 p	83081	97.2%	403	404
A 5 p	91314	97.1%	403	404
A 6 p	65825	97.1%	404	404
A 7 p	82813	96.7%	404	404
S 1 p	68185	98.5%	403	404
S 2 p	63873	98.3%	403	404
S 3 p	98807	98.4%	403	404
S 4 p	82884	98.3%	403	404
S 5 p	85115	98.4%	403	404
S 6 p	126486	98.2%	403	404
S 7 p	78211	98.5%	403	404
M 1 p	93637	98.3%	403	404
M 2 p	81847	98.2%	403	404
M 3 p	83312	98.6%	403	404
M 4 p	78512	98.5%	403	404
M 5 p	80589	98.3%	403	404
M 6 p	53067	98.6%	403	404
M 7 p	64020	98.5%	403	404

4.3.2. mothur v.1.39.5

The bacterial diversity and abundance of the 21 samples was estimated with the mothur pipeline. The mothur pipeline was fed with the raw reads described in 4.2. 158 unique OTUs were found after alignment to the SILVA database (release 132) and after excluding chloroplast, mitochondrial, archaea and eukaryote OTUs. The remaining OTUs were realigned with the RDP Naïve Bayesian rRNA Classifier Version 2.11 to the RDP 16S rRNA training set 16 and a confidence threshold of 85%. The classifier spotted 78 chloroplast OTUs (49.37%) from those classified by the mothur pipeline with the SILVA database as unassigned phyla or unknown bacterial phyla. Therefore leaving 80 OTUs ascribed to the bacterial phyla. All of these OTUs did not surpass the 0.1% threshold and were under the 0.1% cut off. They were therefore discarded as background noise.

4.3.2.1. Bacteria present in the *in-vitro* plant cultures

No observation was above the 0.1% cut off. Sample M7 was contaminated and therefore excluded from further analysis. The assumption being that an unclassified bacterium *Microbacteriaceae* had overgrown the sample based on a first analysis. (See Appendix)

4.3.3. USEARCH v.10.0.240

The fungal diversity and abundance of the 21 samples was estimated with the USEARCH pipeline. 38 unique OTUs were found after alignment to the UNITE database v.7.2. These OTUs were identified again with the RDP Naïve Bayesian rRNA Classifier Version 2.11 to the Warcup Fungal ITS trainset 2 with a confidence threshold of 85%. 31.76% of the OTUs classified as part of the fungal domain, the rest was discarded as stemming from amplification of plant genomic DNA fragments. Therefore leaving 8 OTUs ascribed to the *Ascomycota* phylum and 4 to uncultured fungal phyla.

4.3.3.1. Fungi present in the *in-vitro* plant cultures

The different genera detected within the classified phyla were *Fusarium* (8.33%), *Mycochaetophora* (8.33%), *Penicilium* (8.33%) and 75% classified as uncultured fungi. The composition of each individual sample is shown in figure 13.

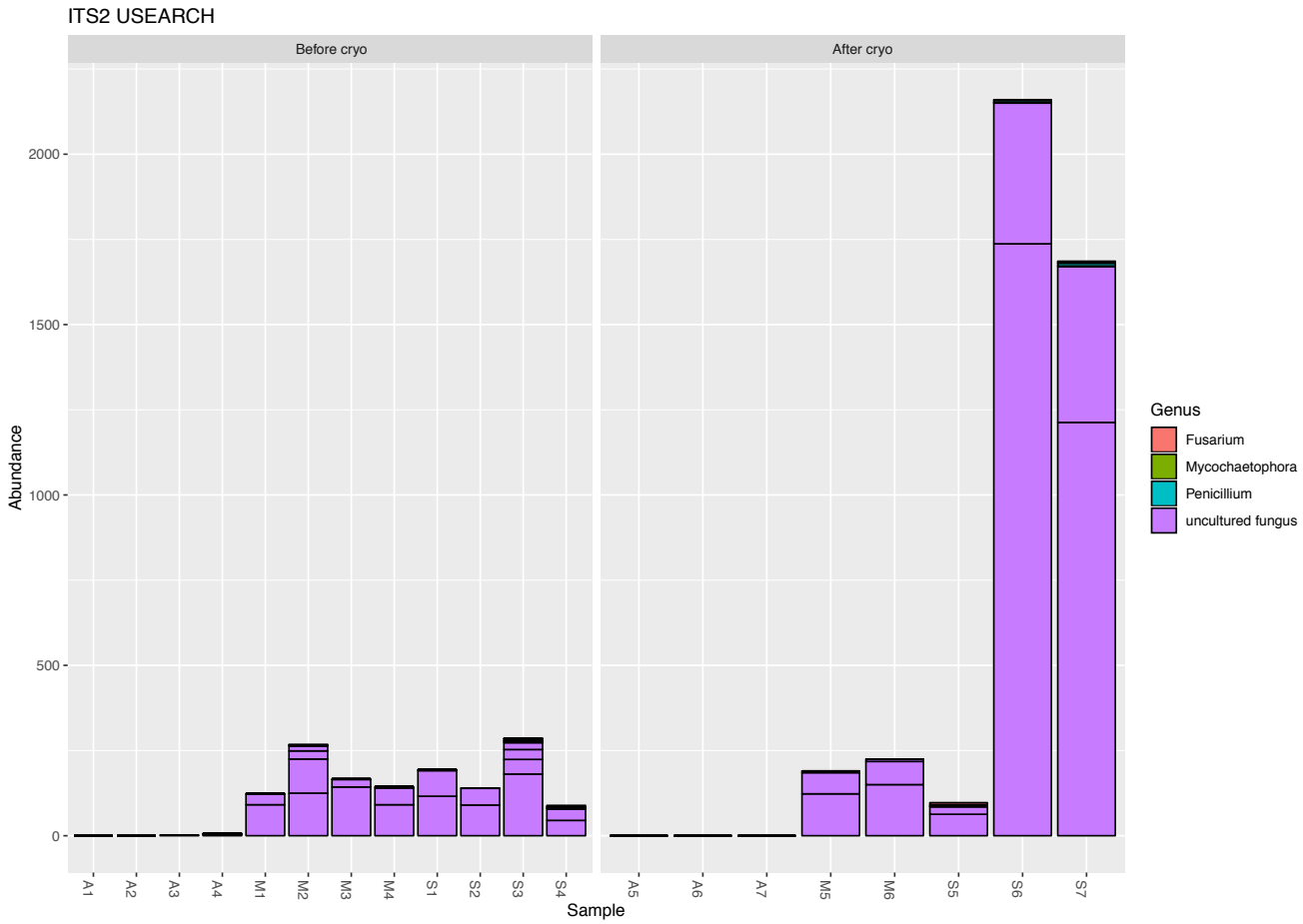


Figure 13 The composition of each individual sample according to the USEARCH pipeline. The Abundance is the read number that clustered into that specific OTU and displayed on the y-axis. The samples were furthermore divided into the groups "Before cryo" and "After cryo" to better visualize differences in the abundance and composition of the microbiome after the cryopreservation procedure; the 0.1% cut off was not applied in the visualization of the data.

There were no differences in the composition of the OTUs before and after cryopreservation. The abundance of the reads clustering in the uncultured fungi OTUs in sample S6 and S7 was substantially higher than in S3 and S4. There was no apparent difference between the samples M1-M6, neither in their composition nor in their abundance. All the *A. laciniata* samples show no signs of fungal OTUs above background.

4.3.4. DADA2 v.1.8

The DADA2 pipeline was used as well because of the big advantage of inferring sequences exactly and therefore resolving differences between sequences down to one nucleotide. DADA2 produces more real variants ASVs than other methods. (See 1.4.3) The fungal and bacterial diversity and abundance of the 21 samples was inferred with the DADA2 pipeline. The visualized quality

profiles (visualization of the base call accuracy of all reads in one .fastq file) of each two forward and two reverse reads for the 16S V3V4 region and the ITS2 region respectively showed expected results for Illumina Miseq sequenced reads, it is needed to decide to what length the sequences shall be trimmed. (Figure 14 and 15) As expected for Illumina Miseq sequenced reads the quality of the base calling decreased at the end of the read beginning approximately at the 200th bp, mainly due to substitutions. (Knief C., 2014; Zoll J. *et al.,* 2016; Xiao L. *et al.,* 2014) This happened because the polymerase ligated a wrong nucleotide during the sequencing steps. This led to the recorded signal from the differently colored fluorophors attached to the nucleotides becoming more ambiguous and thusly of lower quality. Therefore the ends of the reads were cut off and excluded from further analysis, this process is called trimming. From the quality profiles the trimming length was chosen to keep the median quality score Q of the reads above 30, this infers a base call (recognizing a correct base in the sequencing process) accuracy of 99.9%. The Q-score is a prediction of the likelihood of a wrong base call. For the 16S V3V4 region the read length was trimmed to 280 bp and 230 bp for the forward and reverse reads respectively. The ITS2 region reads were trimmed to 280 bp for the forward reads and 220 bp for the reverse reads.

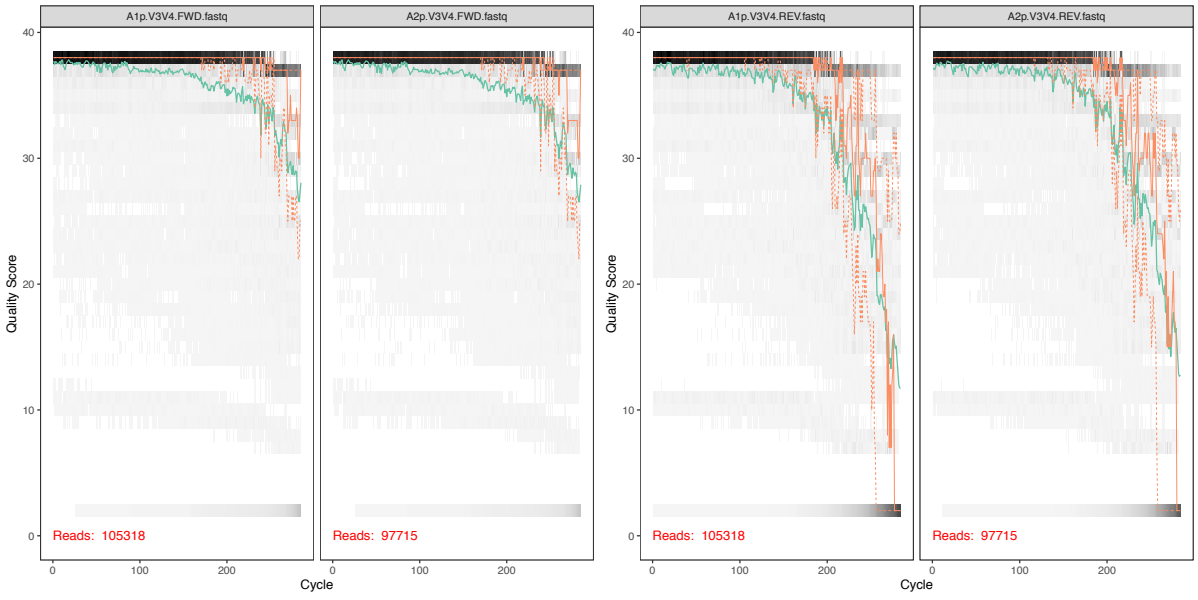


Figure 14 The visualized quality profiles of two 16S V3V4 forward reads and the corresponding reverse reads. In gray scale is a heat map of the frequency of each quality score at each base position. The green line shows the median quality score at each position; the orange lines represent the quartiles of the quality score distribution. Quartiles are the quality score values that divide all values at that position into four equal quarters. The dotted line represents therefore the lower 25% of the quality values and the orange line the higher 25% quality score values at that certain position.

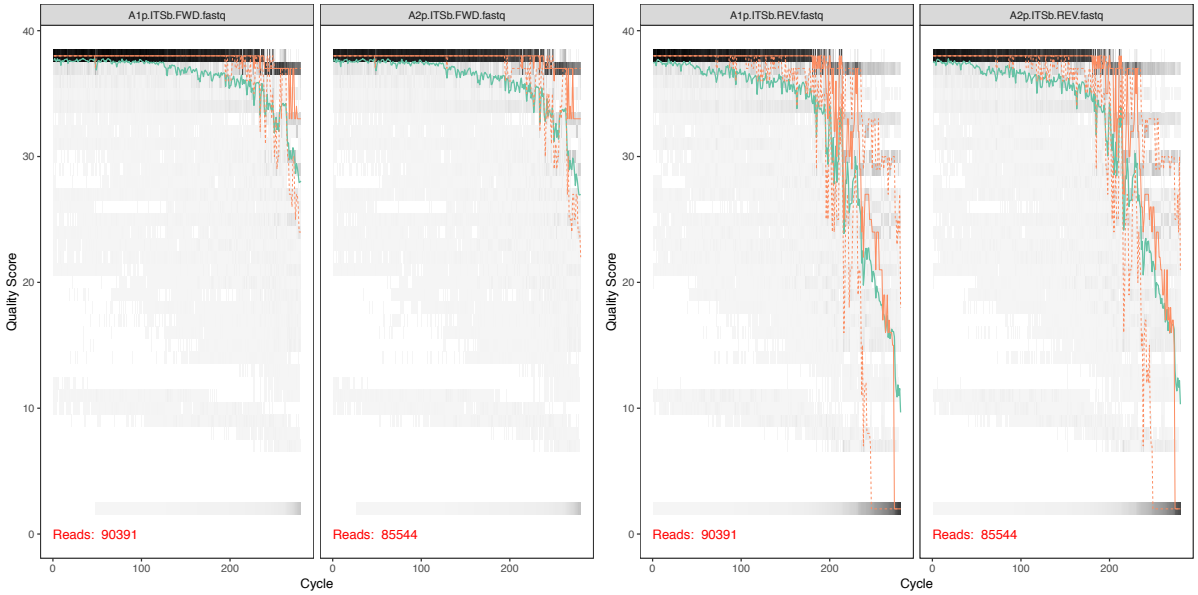


Figure 15 The quality profiles of two ITS2 forward reads and the corresponding reverse reads. In gray scale is a heat map of the frequency of each quality score at each base position. The green line shows the median quality score at each position; the orange lines represent the quartiles of the quality score distribution. Quartiles are the quality score values that divide all values at that position into four equal quarters. The dotted line represents therefore the lower 25% of the quality values and the orange line the higher 25% quality score values at that certain position.

The expected error rates (Figure 16 and 17) were used to filter the sequences with the "*filterAndTrim()*" command for the filtered and trimmed 16S V3V4 amplicons and ITS2 amplicons individually. The estimated error rates were a good fit to the observed rates and the error rates dropped with increased quality as anticipated for both amplicon sets.

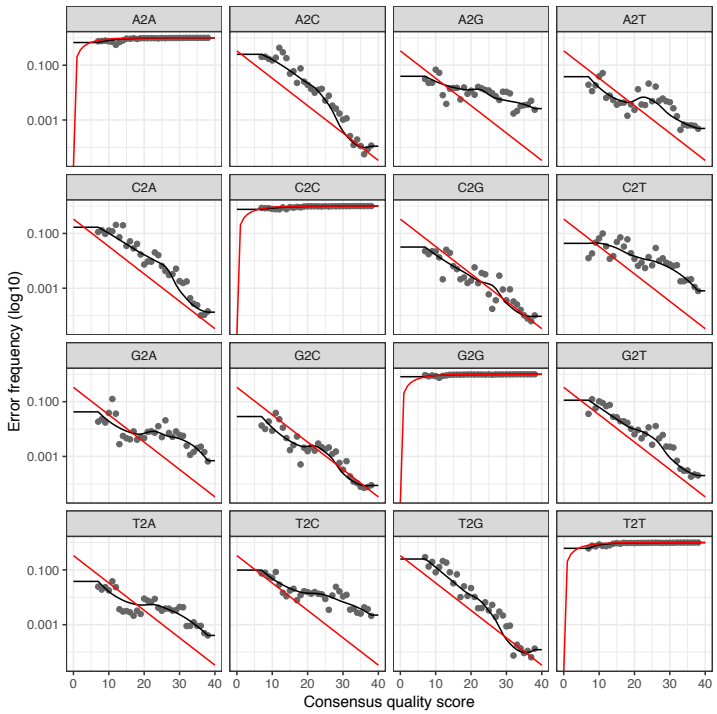


Figure 16 The error rates for each possible transition (A -> C, A -> G, A-> T, etc.) of the filtered and trimmed forward 16S V3V4 reads. Points are the observed error rates for each consensus quality score. The black line represents the estimated error rates after convergence of the machine-learning algorithm. (Chapter 1.4.3) The red line illustrates the error rates expected under the nominal definition of the Q-score (also called Phred quality score and logarithmically linked to the error frequency).

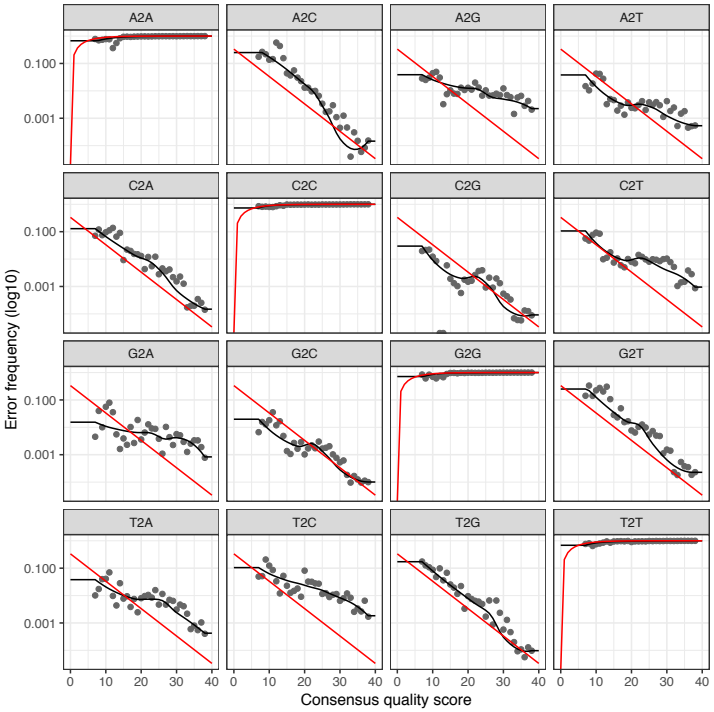


Figure 17 The error rates for each possible transition (A -> C, A -> G, A-> T, etc.) of the filtered and trimmed forward ITS2 reads. Points are the observed error rates for each consensus quality score. The black line represents the estimated error rates after convergence of the machine-learning algorithm. (Chapter 1.4.3) The red line illustrates the error rates expected under the nominal definition of the Q-score (also called Phred quality score and logarithmically linked to the error frequency).

The number of the reads that passed each individual step of the accurate sample inference pipeline with single nucleotide resolution are displayed in table 4.7 for the 16S V3V4 reads and table 4.8 for the ITS2 reads. From the tables the percentage of average passed reads could be calculated, 79.63% for the amplified 16S V3V4 region and 64.45% for the amplified ITS2 region. Furthermore, the track of the number of passed reads showed evidence that sample M7 had been overgrown by a bacterium which resulted in no passed reads for the ITS2 region and was therefore excluded from the analysis. With the exception of the filtering step, which was kept stringent, there was no step in which the majority of reads were lost.

Table 4.7: Track of the number of 16S V3V4 reads that passed each step in the DADA2 pipeline.

16S V3V4 reads							
Sample	Input	Filtered	denoised F	denoised R	merged	Non chimeric	% passed
A 1 p	105318	84419	84396	84395	84045	77264	73.36 %
A 2 p	97715	79340	79318	79328	79178	77888	79.71 %
A 3 p	76559	61986	61981	61975	61861	60996	79.67 %
A 4 p	83081	66551	66536	66544	66067	65050	78.30 %
A 5 p	91314	78203	78185	78199	77871	76689	83.98 %
A 6 p	65825	52774	52744	52749	52643	51307	77.94 %
A 7 p	82813	66508	66480	66497	66343	65348	78.91 %
S 1 p	68185	56447	56433	56434	56251	55019	80.69 %
S 2 p	63873	52710	52700	52700	52478	51380	80.44 %
S 3 p	98807	81634	81616	81628	80912	77281	78.21 %
S 4 p	82884	68409	68390	68393	68123	66761	80.54 %
S 5 p	85115	71055	71050	71040	70818	69495	81.65 %
S 6 p	126486	103595	103582	103580	103221	101308	80.09 %
S 7 p	78211	65981	65967	65969	65765	64427	82.37 %
M 1 p	93637	77309	77281	77299	76930	75026	80.12 %
M 2 p	81847	65864	65800	65849	64826	61835	75.55 %
M 3 p	83312	68349	68324	68313	68002	66334	79.62 %
M 4 p	78512	65596	65574	65580	65217	63404	80.75 %
M 5 p	80589	66107	66089	66097	65787	64124	79.57 %
M 6 p	53067	44984	44965	44963	44773	43627	82.21 %
M 7 p	64020	52217	52180	52175	51725	50327	78.61 %

Table 4.7: Track of the number of ITS2 reads that passed each step in the DADA2 pipeline.

ITS2 reads							
Sample	Input	Filtered	Denoised F	Denoised R	merged	Non chimeric	% passed
A1p	90391	72253	72248	72246	72211	71080	78.63 %
A2p	85544	66945	66932	66934	66886	65810	76.93 %
A3p	109058	86846	86840	86841	86794	85440	78.34 %
A4p	163384	128222	128188	128189	128030	126138	77.20 %
A5p	78909	62935	62932	62929	62891	61931	78.48 %
A6p	117953	94517	94458	94511	94399	92967	78.81 %
A7p	98769	78343	78337	78339	78301	77093	78.05 %
K1p	76004	46882	46856	46869	46073	46057	60.59 %
K2p	118647	73006	72978	72989	72056	72040	60.71 %
K3p	84254	51900	51869	51879	51280	50604	60.06 %
K4p	62360	34971	34942	34943	34498	34490	55.30 %
K5p	63512	39359	39328	39333	38695	38687	60.91 %
K6p	78193	44365	44329	44338	43289	43286	55.35 %
K7p	79699	49543	49439	49522	48189	48185	60.45 %
M1p	59769	35025	35011	35020	34242	32574	54.49 %
M2p	83120	50435	50362	50414	49486	46890	56.41 %
M3p	80348	45818	45799	45809	44839	43019	53.54 %
M4p	65295	39331	39325	39312	38450	36023	55.16 %
M5p	75565	44821	44815	44818	43808	40889	54.11 %
M6p	95121	58088	57998	58073	56752	52689	55.39 %
M7p	119	6	6	1	0	0	/

These passed reads or amplicon sequence variants (ASVs) were assigned to a taxonomy with different databases using different classifiers (Chapter 3.6.4). This led to 44 ASVs from which 88.63% classified as chloroplasts or mitochondria and were therefore removed from the dataset. Chimeras were thought to consist of two different mother strands and on this premises they are detected by the DADA2 pipeline with high confidence. But the possibility of triple chimeras remains. Such triple chimeras form if two strands overlap from both ends one middle strand that forms a loop; the polymerase therefore duplicates a part of the first strand, the loop of the middle strand and a part of the third overlapping strand. (Gogvadze E. *et* al., 2005) These triple chimeras are usually very rare and under the background cut off of 0.1%. In our case they had to be considered because of the low amount of ASVs and the low amount of sequences that clustered to these. To detect these chimeras the ASVs were cut in 3 equal length sequences and classified to detect chimeras that were not recognized by the DADA2 algorithm. If the 3 sequences from one ASV were classified to the same taxonomy the ASV was kept in the dataset if not it was discarded as a triple chimera. Two ASVs were spotted as chimeras and therefore excluded from the 16S dataset, leaving 5 ASVs ascribed to the bacterial phyla and none as unknown.

The ITS2 ASVs were classified with different databases using different classifiers (See 3.6.4). This led to 85 ASVs from which 87.05% classified as plant derived and were removed from the

dataset. As for the 16S ASVs the remaining ITS ASVs were cut in three equal length sequences to detect triple chimeras. If the 3 sequences from one ASV were classified to the same taxonomy the ASV was kept in the dataset if not it was discarded as a triple chimera. Leaving 11 ASVs ascribed to fungal phyla.

4.3.4.1 Bacteria present in the *in-vitro* plant cultures

The 5 ASVs ascribed to the bacterial phyla did not surpass the 0.1% threshold and were under the 0.1% cut off. They were therefore discarded as background noise. Sample M7 was contaminated and therefore eliminated from further analysis. The contamination could be classified as *Curtobacterium* from the family *Microbacteriaceae*.

None of the observations were above background, which is typically set at 0.1% of the total amount of processed reads. There were no apparent different observations, nor statistical differences between samples before and after cryopreservation, if the background cut off of 0.1% was used. (Table 4.9) None of the observations could be seen as repeatable because they are all under the 0.1% cut off. The analysis of the bacterial microbiome of *in-vitro* plants based on 16S V3V4 ASVs resulted in no observable microbiome. (Table 4.9)

4.3.4.2 Fungi present in the *in-vitro* plant cultures

The 11 fungal ASVs were classified into the *Ascomycota* (18.18%) phylum and uncultured fungi (81.81%) phyla. The different genera detected within the classified phyla were *Fusarium* (9.09%), *Penicillium* (9.09%) and 81.81% classified as uncultured fungus. (Figure 19)

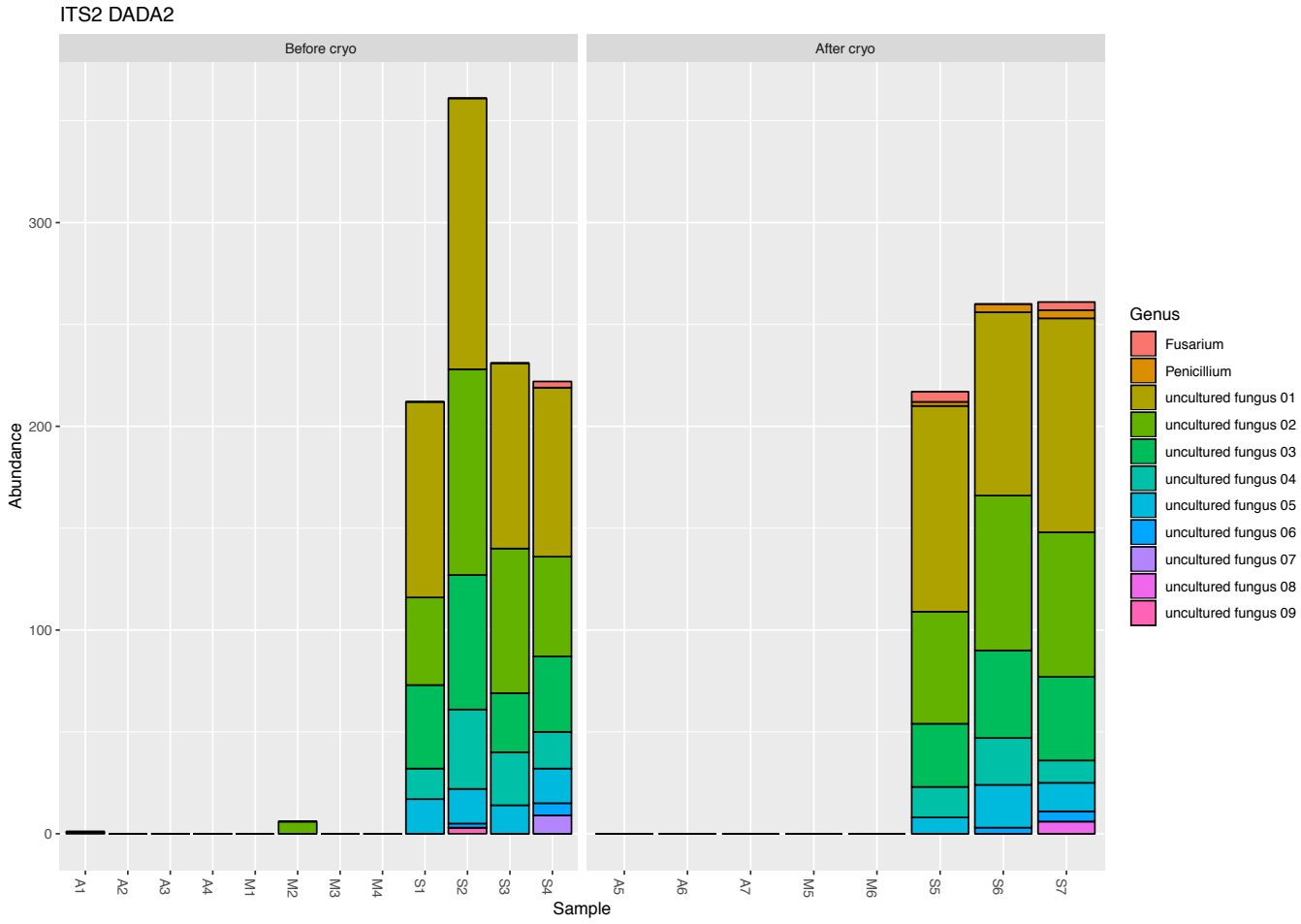


Figure 19 The composition of fungal microbiome in each individual sample according to DADA2 pipeline.. The Abundance is the read number that clustered into that specific ASV and displayed on the y-axis. The samples were furthermore divided into the groups “Before cryo” and “After cryo” to better visualize differences in the abundance and composition of the microbiome after the cryopreservation procedure; the 0.1% cut off was not applied in the visualization of the data.

Relevant ASV abundances and compositions have only been found in the samples from *S. tuberosum*. The composition of the fungal microbiome of the S1-4 samples did not statistically vary from the S5-7 samples. As displayed in table 4.8, the mean number of reads that clustered to their specific ASV did not vary statistically before or after cryopreservation. Neither the composition, nor the abundance of the fungal microbiome varied significantly before and after cryopreservation.

Table 4.8: The mean number of reads that clustered to their specific ASV and the corresponding standard deviation (SD) and coefficient of variation (CV) or relative standard variation for the samples S1-7, split into the groups before and after cryopreservation.

	Before cryopreservation			After cryopreservation		
ASV Genus	Mean	SD	CV	Mean	SD	CV
uncultured fungus 01	100.75	±22.15	±21.99%	98.66	±7.76	±7.87%
uncultured fungus 02	66	±26.25	±39.78%	67.33	±10.96	±16.29%
uncultured fungus 03	43.25	±15.96	±36.91%	38.33	±6.42	±16.77%
uncultured fungus 04	24.5	±10.72	±43.77%	16.33	±6.50	±39.84%
uncultured fungus 05	16.25	±1.50	±9.23%	14.33	±2.51	±17.54%

To reduce the differences in ASV diversity and abundance between the samples alpha diversity was measured. (Figure 20) Alpha diversity indices are mathematical measures that provide more information about the composition of a microbial community because they take relative abundance, richness and distribution of different species into account. It is a way to quantify diversity. Alpha diversity indices Shannon and Simpson showed that there were no significant differences within the S1-7 group nor the A1-7 or M1-7, therefore no differences in diversity can be observed. The groups A, S and M compared to each other showed that only S had any alpha diversity values due to the fact, that there was no kind of fungal microbiome found in the *A. laciniata* and *M. aquatica* samples. Furthermore the Chao 1 as an alpha diversity estimator, that attempts to extrapolate from available observations the total number of ASVs in the community, was plotted. (Figure 20) Chao 1 is a non-parametric estimator (not involving any assumptions) that estimates the number of species present in the samples based on the abundance of ASVs to which single reads cluster. The estimator indicated, that the total number of ASVs for the *S. tuberosum* samples lie between 5 and 8. It was used see if my analysis captured the main proportion of the diversity of the fungal microbiome and if there were any differences in diversity after cryopreservation compared to before cryopreservation.

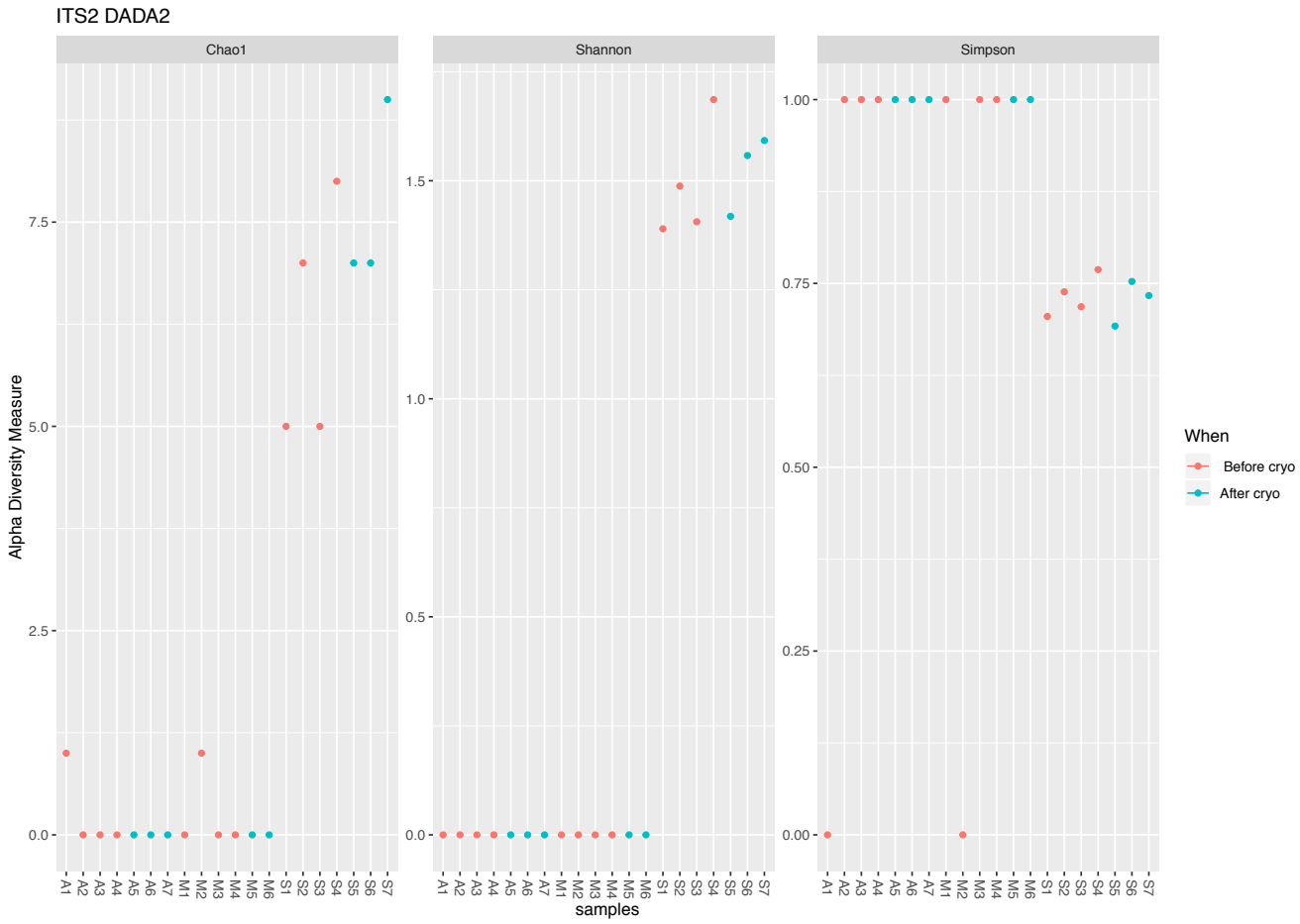


Figure 20 The Alpha diversity measurements for all samples. Chao 1 as an alpha diversity estimator that attempts to extrapolate from the available observations the total number of ASVs within the samples. The Shannon and Simpson index are plotted to show the ASV diversity between the samples. The red dots represent the samples before cryopreservation whereas the blue dots represent the samples after cryopreservation.

The phylogenetic analysis of the obtained fungal endophyte ASVs reveals two clades of the uncultured fungi. (Figure 21) One clade contained the ASVs of uncultured fungi 01, 02, 04, 05, 06 and the second clade the ASVs of uncultured fungi 03, 07, 08 and 09. It furthermore showed the rather high phylogenetic distance from the two *Ascomycota* ASVs *Penicillium* and *Fusarium* and it illustrated the abundance of the ASVs within each sample. The command to build the tree used the minimum distance between two sequences. (Chapter 1.4.2) Therefore, the shorter the branch connecting two ASVs was the lower the genetic distance from each other. Furthermore a tree was built using the minimum linkage approach including *Penicillium deleae*, *Fusarium annulatum*, *Puccinia chrysanthemi* and *Hydnum rufescens* ITS2 sequences. (Appendices) The two Basidiomycota (*Puccinia chrysanthemi* and *Hydnum rufescens*) clustered together, the *Penicillium* sequences clustered and the *Fusarium* sequences clustered together. Therefore it could be assumed that the uncultured fungus clades do not belong to the Ascomycota and Basidiomycota phyla, which include the most fungal species detected by conventional culture methods (Riedel C. *et al.,* 2014).

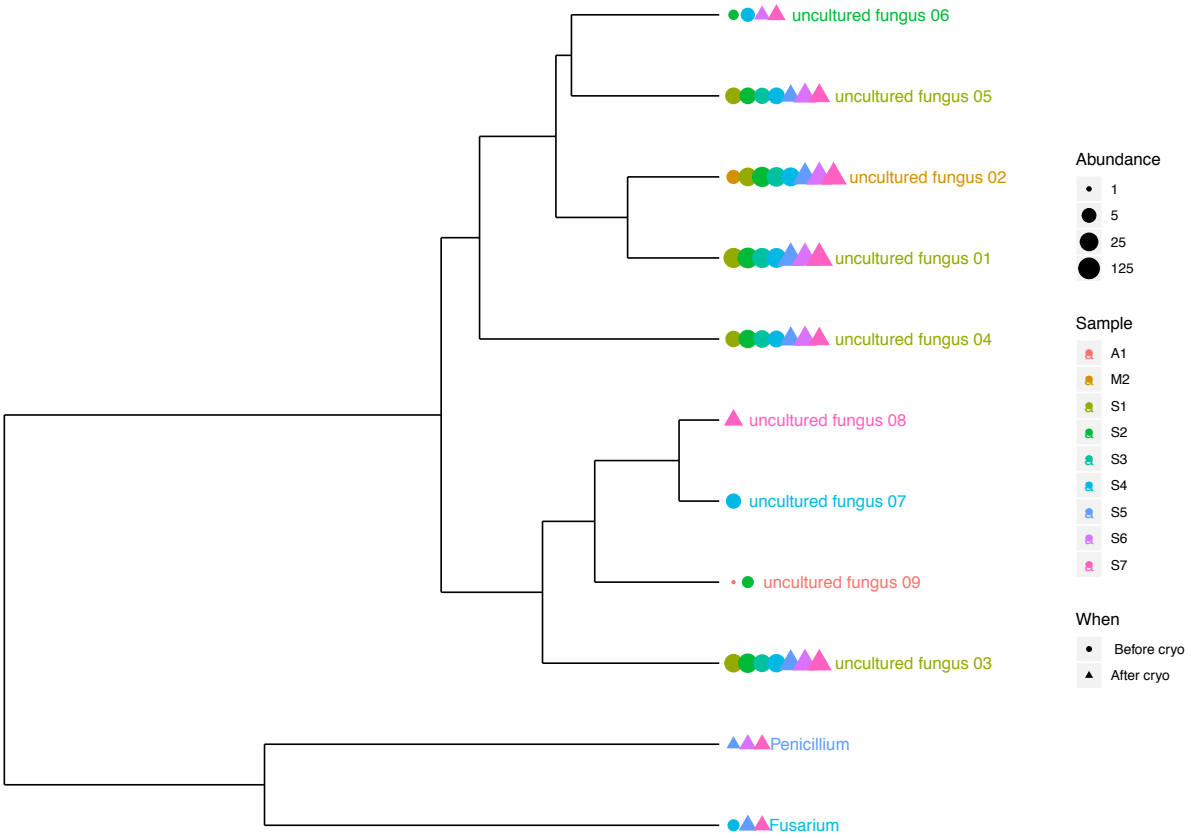


Figure 21 Phylogenetic analysis of the obtained fungal endophyte ASVs. The different colors of the objects (triangle and circle) indicate the sample in which they were found. The different forms of the objects indicate if they were samples from before or after cryopreservation. Circles indicate before cryopreservation, triangles indicate after cryopreservation. The size of the objects indicates the abundance of the fungal endophyte within its sample. The length of the branches between each ASV indicates their genetic distance from each other. The tree depicts the result of a minimum linkage approach; this is equivalent with single linkage and is used to find the closest connected components of a graph. It uses the minimum distance between two sequences to build the tree.

The analysis did not find any detectible differences before and after cryopreservation within the fungal microbiome. The uncultured fungi ASVs were spread over all the *S. tuberosum* samples equally, keeping their composition and abundance within variation. According to this analysis, the *M. aquatica* and *A. laciniata* samples could be considered free from evident fungi, because none of the observations were above the 0.1% background cut off. (Table 4.9) The appearance of *Penicillium* after cryopreservation is probably the result of contamination during cryopreservation and/or regeneration. The detection of *Fusarium* on the other hand could also be the result of cross contamination during the DNA extraction or could be a carry-over contamination during the Illumina run in the Eurofins facility. (Laurence M. *et al.* 2014) The possibility of small fragmented DNA in the air, falling into the tubes and contaminating the extracted DNA can also be considered as a potential

contamination path and would explain the occurrence of the below background cut off (0.1%) fungal ASVs in sample A1 (*A. laciniata*) and M2 (*M. aquatica*). (Porter-Jordan K. *et al.,* 1990)

4.3.5. Combined and compared - The microbiome of *in-vitro* cultivated plants before and after cryopreservation

The mothur and DADA2 pipeline did not differ in the outcome, there is no detectable change in the bacterial microbiome before and after cryopreservation. Although the analysis did confirm the difference of the pipelines as such showing that mothur created spurious OTUs (Chapter 1.4.3) that did not really represent the reality. DADA2 on the other hand resulted in an ASV, which was only found in the M7 sample and could be spotted as contamination and with 97% likelihood identified as *Curtobacterium* from the family *Microbacteriaceae*. Whereas according to the mothur pipeline this OTU was denoted as unclassified *Microbacteriaceae*.

The USEARCH and DADA2 pipeline resulted in slightly different abundances and compositions of the fungal microbiome. The DADA2 pipeline could not find the OTUs found by USEARCH within the *M. aquatica* samples. Both pipelines found *Penicillium* and *Fusarium* individually in very low abundance, making it highly likely to be present within the samples. Both pipelines suggested, that except for the change in abundance in the USEARCH pipeline, there were no observable changes in the fungal microbiome of *in-vitro* cultivated plants after cryopreservation.

Table 4.9: The percentages of sequences clustering to their specific amplicon sequence variant (ASV) in relation to the non-chimeric passed reads from the DADA2 pipeline per individual sample. The green shaded areas of the table indicate the percentages of sequences from the mother plants. The rose shaded areas of the table indicate the percentages of sequences from samples before cryopreservation. The blue shaded areas of the table indicate the percentages of sequences from samples after cryopreservation. The first four ASVs are derived from the 16S amplicons and the represented genera are part of the bacterial kingdom. Separated by an engorged line are the ASVs derived from the ITS2 amplicons and the represented genera are part of the fungal kingdom. M7 has been excluded because it had been overgrown by *Curtobacterium sp.*

Genera	Sample																		
	A 1	A 2	A 3	A 4	A 5	A 6	A 7	S 1	S 2	S 3	S 4	S 5	S 6	S 7	M 1	M 2	M 3	M 4	M 5	M 6
% Bosea	0	0	0	0	0	0.019	0	0	0	0	0	0	0	0	0	0	0	0	0	0
% uncultured bacillus	0	0	0	0.003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
% unclassified Paenibacillaceae	0.085	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
% unclassified Paenibacillaceae	0	0	0	0	0	0	0	0	0	0.002	0	0	0	0	0	0	0	0	0	0
% Fusarium	0	0	0	0	0	0	0	0	0	0	0.008	0.012	0	0.008	0	0	0	0	0	0
% Penicillium	0	0	0	0	0	0	0	0	0	0	0	0.005	0.009	0.008	0	0	0	0	0	0
% uncultured fungus 01	0	0	0	0	0	0	0	0.208	0.184	0.179	0.240	0.261	0.207	0.217	0	0	0	0	0	0
% uncultured fungus 02	0	0	0	0	0	0	0	0.093	0.140	0.140	0.142	0.142	0.175	0.147	0	0.012	0	0	0	0
% uncultured fungus 03	0	0	0	0	0	0	0	0.089	0.091	0.057	0.107	0.080	0.099	0.085	0	0	0	0	0	0
% uncultured fungus 04	0	0	0	0	0	0	0	0.032	0.054	0.051	0.052	0.038	0.053	0.022	0	0	0	0	0	0
% uncultured fungus 05	0	0	0	0	0	0	0	0.036	0.023	0.027	0.049	0.020	0.048	0.029	0	0	0	0	0	0
% uncultured fungus 06	0	0	0	0	0	0	0	0	0.002	0	0.017	0	0.006	0.010	0	0	0	0	0	0
% uncultured fungus 07	0	0	0	0	0	0	0	0	0	0	0.026	0	0	0	0	0	0	0	0	0
% uncultured fungus 08	0	0	0	0	0	0	0	0	0	0	0	0	0	0.012	0	0	0	0	0	0
% uncultured fungus 09	0.001	0	0	0	0	0	0	0	0.004	0	0	0	0	0	0	0	0	0	0	0

5. Discussion

With the dawn of NGS (Next generation sequencing) the unraveling of the composition of highly diverse microbiomes, from sand grains to the human gut, has taken place. This study was to our knowledge the first to use this technology for the analysis of *in-vitro* cultivated plants. *In-vitro* plant cultures were viewed as "aseptic" (free of optically detectable contaminations). Recent studies founded on a cultivation-based approach proved the occurrence of bacterial endophytes as omnipresent in *in-vitro* cultures. (Orlikowska T. *et al.,* 2017; Massimo N. *et al*., 2015; Porras-Alfaro A. *et al.,* 2011) The cultivation-independent, NGS-based approach to detect all possible bacteria and fungi used in this study should therefore detect more genera than culture-based approaches. The objective behind this work was to evaluate the diversity of the microbiome of *in-vitro* cultivated plants before and after cryopreservation using an amplicon-based metagenomic approach. The use of NGS-based metagenomic sequencing to evaluate the microbiome from various holobionts is becoming more common. (Fonseca *et al.*, 2018; Hammad *et al.,* 2017; Berg *et al.*, 2014; Burns *et al.*, 2015; Chelius *et al.*, 2001; Knief *et al.*, 2014; Liaqat *et al.*, 2016; Massimo *et al.*, 2015; Riedel *et al.*, 2014)

The DNA extractions produced a consistent average amount of 114.12 ± 14.47 ng μl^{-1} DNA sheared to 8 kbp – 3 kbp fragments in all samples, thus proving the robustness and reproducibility of the procedure. Furthermore, all samples showed a positive amplification of the 16S and ITS region in the PCR using specific primers. In the *M. aquatica* samples there were two to three bands visible (MW of approximately 2 kb). This might have happened because of tandem duplication of the 16S rRNA gene in chloroplasts / mitochondria or a dominating bacterial species. This fact was not further investigated, because the following sequencing was performed with different primers for the V3-V4 region. It was assumed that the amplification of the specific DNA fragment would indicate the presence of bacteria and fungi. As expected, the specific 16S primers also amplified chloroplast and mitochondrial DNA and the ITS primers genomic plant DNA.

Second generation NGS, such as the applied Illumina Miseq, are based on a PCR protocol for the signal amplification. (Knief C., 2014; Zoll J. *et al.,* 2016; Xiao L. *et al.,* 2014) The yielded paired-end reads had a mean Q-score of 33.56 inferring a base call accuracy of 99.956%. In comparison to other studies that used an amplicon-based metagenomic approach to describe the microbiome of plants the sequencing depth was approximately the same or higher (average of 62214 merged paired-end reads per sample per region). (Fonseca *et al*., 2018; Manter *et al.*, 2010; Appendix) The merging of the paired-end reads with FLASH (Zhang J. *et al.* 2013, Magoc T. *et al.* 2011) outputted a consistent mode length for the 16S V3V4 amplicons for all samples. The mode length is the length of the sequences, which appears most often in a group of sequences. This was expected, since the

prokaryotic 16S V3V4 region does not fluctuate a great deal in length. The mode length of the ITS2 amplicons on the other hand was different according to the derived DNA. This reflected the variable length of the internal transcribed spacer and indicated a plant specific length. This meant that the primers primarily amplified plant genomic DNA, we can therefore conclude, that the mode length of the ITS2 amplicons was plant–specific in this study.

To analyze the composition and the specific abundance of each genus the attained amplicons had to be clustered together into operational taxonomic units or in the case of the DADA2 (Callahan B. J. *et al.* 2016) pipeline into amplicon sequence variants (ASV). DADA2 inferred sequences exactly and resolved differences between sequences down to one nucleotide. Because of this and the comparison of the pipelines (Chapter 4.3.5) the results from the DADA2 pipeline were used to analyze the composition of the microbiome of the *in-vitro* plants. (Figure 18) The taxonomy was determined by aligning the ITS2 ASVs to the Warcup Fungal ITS trainset 2 using the RDP Naïve Bayesian rRNA Classifier and the UNITE Fungal ITS trainset using the BLASTn 2.2.29+ algorithm, using the different databases had the aim to assign taxonomy with high accuracy and confidence. Those ASVs classified as plant ITS2 regions were excluded from the analysis of the microbiome.

The 16 S V3V4 ASV from the presumed contamination could be successfully aligned to *Curtobacterium sp.* from the family *Microbacteriaceae* with 97% likelihood. (Chapter 4.3.4) *Curtobacterium sp.* is a gram-stain positive strictly aerobic chemoorganotrophic Actinobacterium associated with plant infections. (Saddler G. S. *et al.*, 2015; Kodym A. *et al.*, 2018) Below 0.1% cut off, 4 16S V3V4 ASVs were observed. (Appendix) None of the observations could be seen as repeatable, except for the contamination, because all observations were under the 0.1% cut off. Therefore, the analysis of the bacterial microbiome of *in-vitro* plants based on 16S V3V4 ASVs in this study resulted in no detectable microbiome. Furthermore, there were no differences between the samples before and after cryopreservation. (Table 4.9)

Nine of the eleven ITS2 ASVs classified as uncultured fungi and were only significantly represented in the *S. tuberosum* samples. (Figure 19, Table 4.9) Therefore, the natural communities of fungi in the *in-vitro S. tuberosum* are largely composed of uncultured fungal taxa. The composition and individual abundance of each ASV from this uncultured fungal microbiome does not significantly vary from before and after cryopreservation. (Table 4.8) This revealed some highly preserved endophytes within the species *S. tuberosum*. The remaining two ASVs were below 0.1% cut off and recognized as *Fusarium sp.* and *Penicillium sp.* The appearance of *Penicillium sp.* after cryopreservation is probably the result of contamination during cryopreservation and/or regeneration. The possibility of small fragmented DNA in the air, falling into the tubes and contaminating the extracted DNA can also be considered as a potential contamination path and

would explain the occurrence of the below background cut off (0.1%) fungal ASVs in sample A1 (*A. laciniata*) and M2 (*M. aquatica*). (Porter-Jordan K. *et al.,* 1990)

DNA can be extracted from living fungi and from their spores; this DNA does not vary in its sequences. (Mhlongo M. *et al*., 2018) Because of this fact it was not possible to determine from the given data alone if the uncultured fungi are only present as metabolic inactive spores or metabolic active mycelium within the *in-vitro* plants. Because of the aseptic cultures, the sterile work practice and the reoccurrence of the uncultured fungi within the *S. tuberosum* before and after cryopreservation, we could assume that these are indeed highly preserved endophytes. Given the unchanged abundances throughout the *S. tuberosum* samples we could furthermore assume that the fungi are metabolic active in the plants. Additionally we could positively say, that these uncultured fungi are present in or surrounding healthy meristematic pathogen-free cells, because only these cells would survive the cryopreservation step. (Wang M. *et al.,* 2018)

The phylogenetic analysis of the obtained fungal ASVs used the minimum linkage approach to determine the phylogenetic distance between the given sequences. (Edgar R. C., 2017; Figure 21) The culture-independent approach revealed evidence for the presence of an unclassified fungal phylotypes in *S. tuberosum*. A more in depth phylogenetic analysis of the given uncultured fungal ASVs showing the evolutionary relationships among a wide range of different fungal species could conclusively better place them in a reference tree. This was not performed in this study because the aim was to assess the diversity of the microbiome of *in-vitro* cultivated plants before and after cryopreservation.

The comparison of the diversity and richness of bacterial OTUs among different studies varies among plant species and living conditions. (Fonseca *et al*., 2018) Manter *et al.*, (2010) analysis of the bacterial root microbiome of 12 *in-vivo Solanum spp.* revealed an average of 477 bacterial OTUs (represented by 1000 sequences per sample). They found 260 different genera of bacteria living as endophytes. (Manter *et al.*, 2010) Our study on the other hand could show that *in-vitro S. tuberosum* contained no bacterial ASVs above background cut off. This gives further evidence for the correlation between the living conditions and the composition of the bacterial microbiome. Fonseca *et al*., (2018) found that the main microbial community's residing in Eucalyptus roots under different conditions were those conducting biological nitrogen fixation, giving the Eucalyptus the ability to grow in soil without nitrogen. In our study we found a presumably essential fungal microbiome within *in-vitro S. tuberosum* even if its purpose remains unknown because of the lack of a more in depth phylogenetic and biochemical analysis, we can assume that its highly conserved nature before and after cryopreservation is due to some sort of advantage they provide to the host *in-vitro* cultivated plant.

6. Conclusion

Until now, only cultivable bacteria and fungi from *in-vitro* cultivated plants have been identified and characterized. In the present study, for the first time next generation sequencing (NGS) was used to capture the microorganisms present within *in-vitro* plants and to analyze the influence of cryopreservation on the microbiome. The plants studied belonged to three different families (Asteraceae, Lamiaceae and Solanaceae), *Artemisia laciniata* was chosen because it is a highly endangered species in Europe. *Mentha aquatica* was chosen because it hybridizes with *Mentha spicata* to produce medicinally important *Mentha x piperita* L. *Solanum tuberosum* was chosen because it holds a paramount position in agriculture and nutrition of the human population.

- The NGS analysis of the *A. laciniata* and *M. aquatica* samples could not detect any kind of fungal or bacterial microbiome above 0.1% cut off.

- *S. tuberosum* showed no observations for bacterial microbiome above 0.1% cut off.

- The analysis of *in-vitro* cultivated plants revealed highly preserved uncultured fungal endophytes belonging to an un-described fungal phylotype within *S. tuberosum* as well as below 0.1% cut off *Fusarium sp.* and *Penicillium sp.*.

- Based on the results, which may be highly influenced by the methodologies used in this study, we could suggest, that *A. laciniata* and *M. aquatica* do not rely on a microbiome to survive under *in-vitro* conditions.

- There was no evidence of significant differences in abundance or composition within the members of fungal microbiome of *S. tuberosum* before and after cryopreservation. It can be assumed, that cryopreservation has no effect on these highly preserved endophytic fungi.

- Because of the scarcity of endophytes, plants coming from *in-vitro* cultures could be excellent starting material for *in-vivo* studies on the role of endophytes. The *in-vitro* plants could receive artificially assembled microbiomes in different abundances and compositions. These plants could then be tested for survival, robustness, stability of the administered microbiome, produced secondary metabolites etc.

Conclusion

The technology used in this study appeared to be suitable to answer the posed questions. What could be improved in the future studies is the reduction of amplified chloroplast reads for the 16S amplicons by using primers that have lower affinity to chloroplast and mitochondrial DNA, or using an additional blocking primer.

7. Abstract

In-vitro cultivated plants, despite being considered aseptic, may carry latent infections and endophytes,microorganisms that live within plants without causing disease symptoms. Until now, only cultivable bacteria and fungi from *in-vitro* cultivated plants have been identified and characterized. Next generation sequencing (NGS) has revolutionized all studies based on DNA sequencing. In this study NGS was applied to analyze the microbiome of *in-vitro* plants for the first time.

Endophytic microorganisms often stay dormant and become visible only after several subculturing steps and after exposure to stress, such as cryopreservation. The technique of cryopreservation is mainly used for the purpose of long-term storage of germplasm. Another biotechnological use is cryotherapy, a rather recent development for eliminating viral pathogens and bacterial pathogens in infected plants. These facts provided the basis for the hypothesis of this study, namely that changes can be expected in the composition and abundance of the microbiome of *in-vitro* cultivating plants undergoing cryopreservation.

The first aim of this study was to find out which bacteria and/or fungi are present in *in-vitro* plants before and after cryopreservation. The second aim was to find out if some bacteria or fungi were lost, acquired and if there were changes in their abundance.

DNA was extracted from *in-vitro* plants derived from three different plant species, *Artemisia laciniata* Willd. (Compositae / Asteraceae), *Mentha aquatica* L. (Lamiaceae) and *Solanum tuberosum* L. (Solanaceae) before and after undergoing cryopreservation. 21 samples (seven samples per species, two sets before cryopreservation and one set after) were analyzed to give an insight into the fungi and bacteria residing within the *in-vitro* plants. The internal transcribed spacer 2 (ITS2) region was used as barcode for the classification to the genus level of fungi, and the 16S V3-V4 region for bacteria. Both were amplified and then sequenced on an Illumina Miseq. The generated sequences of the ITS2 and 16S V3V4 region were processed *in-silico* using different bioinformatics tools and pipelines.

The NGS analysis of the *A. laciniata* and *M. aquatica* samples showed no kind of microbiome above 0.1% cut off. *S. tuberosum* showed no observations for bacteria above 0.1% cut off. The analysis of *in-vitro* plants revealed highly preserved uncultured fungal endophytes within *S. tuberosum* belonging to an un-described fungal phylotype. There was no evidence to significant changes in abundance or composition within the microbiome before and after cryopreservation. It can be assumed, that cryopreservation has no effect on highly preserved endophytes such as those detected within *S. tuberosum.*

8. Zusammenfassung

In-vitro Pflanzen können, obwohl sie als aseptisch betrachtet werden, latente Infektionen und Endophyten, Mikroorganismen, die in Pflanzen leben, tragen, ohne Krankheitssymptome zu verursachen. Bisher wurden nur kultivierbare Bakterien und Pilze aus *In-vitro* Pflanzen identifiziert und charakterisiert. Next Generation Sequencing (NGS) hat alle auf DNA-Sequenzierung basierenden Studien revolutioniert. In dieser Studie wurde erstmals NGS zur Analyse des Mikrobioms von *In-vitro* Pflanzen eingesetzt.

Endophytische Mikroorganismen bleiben oft inaktiv und werden erst nach mehreren Subkultivierungsschritten und nach Belastung, beispielsweise durch Kryokonservierung, sichtbar. Die Technik der Kryokonservierung wird hauptsächlich zum Zweck der Langzeitlagerung verwendet. Eine weitere biotechnologische Anwendung ist die Kryotherapie, eine relativ junge Entwicklung zur Beseitigung von viralen Pathogenen und bakteriellen Pathogenen in infizierten Pflanzen. Diese Tatsachen bildeten die Grundlage für die Hypothese dieser Studie, nämlich dass Änderungen in der Zusammensetzung des Mikrobioms von *in-vitro* Pflanzen, die sich einer Kryokonservierung unterziehen, erwartet werden können.

Das erste Ziel dieser Studie war es herauszufinden, welche Bakterien und / oder Pilze vor und nach der Kryokonservierung in *In-vitro* Pflanzen vorhanden sind. Das zweite Ziel bestand darin, herauszufinden, ob einige Bakterien oder Pilze verloren gegangen waren, erworben wurden und ob sich Veränderungen in ihrer Abundanz zeigten.

DNA wurde aus *In-vitro* Pflanzen extrahiert, die von drei verschiedenen Pflanzenarten, *Artemisia laciniata* Willd, (Compositae / Asteraceae), *Mentha aquatica* L. (Lamiaceae) und *Solanum tuberosum* L. (Solanaceae) vor und nach der Kryokonservierung stammen. 21 Proben (sieben Proben pro Art, zwei Sätze vor der Kryokonservierung und eine danach) wurden analysiert, um einen Einblick in die Pilze und Bakterien zu erhalten, die sich in den *In-vitro* Pflanzen befinden. Die internal transcribed Spacer 2 (ITS2) wurde als Barcode für die Klassifizierung der Gattungsebene der Pilze und die 16S V3-V4-Region für Bakterien verwendet. Beide wurden amplifiziert und dann auf einem Illumina Miseq sequenziert. Die erzeugten Sequenzen der ITS2- und 16S-V3V4-Region wurden *in-silico* mit verschiedenen Bioinformatik-Tools und Pipelines verarbeitet.

Die NGS-Analyse der Proben von *A. laciniata* und *M. aquatica* zeigte keine Art von Mikrobiom über 0,1%. *S. tuberosum* zeigte keine Beobachtungen für Bakterien mit einer über 0,1%. Die Analyse von *In-vitro* Pflanzen zeigte in *S. tuberosum* hoch konservierte nicht kultivierte Pilzendophyten, die zu einem nicht beschriebenen Pilz-Phylotyp gehören. Es gab keine Hinweise auf signifikante Änderungen der Abundanz oder Zusammensetzung im Mikrobiom vor und nach der

Kryokonservierung. Es kann davon ausgegangen werden, dass die Kryokonservierung keinen Einfluss auf hoch konservierte Endophyten hat, wie sie in *S. tuberosum* nachgewiesen wurden.

9. References

Beckers B., Op De Beeck M., Thijs S., Truyens S., Weyens N., Boerjan W., Vangronsveld J. (2016) Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. *Front. Microbiol.* 7: 650.

Benson E. E. (2008) Cryopreservation Theory. In: Reed B. M. (eds.), *Plant Cryopreservation: A Practical Guide*, Springer International Publishing, New York, S. 15-32

Berg G., Grube M., Schloter M., Smalla K. (2014) Unraveling the plant microbiome: looking back and future perspactives. *Front. Microbiol*. (5) 148:

Boufridi A., Quinn R. J. (2017) Harnessing the Properties of Natural Products. *Annu. Rev. Pharmacol. Toxicol.* 2018, 58: 17.1-17.20.

Burns J. H., Anacker B. L., Strauss S. Y., Burke D. J. (2015) Soil microbial community variation correlates most strongly with plant species identity, followed by soil chemistry, spatial location and plant genus. *AoB PLANTS*: Open Access article.

Callahan B. J., McMurdie P. J., Rosen M. J., Han A. W., Johnson A. J. A., Holmes S. P. (2016) DADA2: High resolution sample inference from Illumina amplicon data. *Nat. Methods* 13 (7): 581-583.

Chelius M. K., Triplett E. W. (2001) The diversity of archaea and bacteria in association with the roots of *Zea mays* L. *Microb. Ecol.* 41: 252-363.

CSID:1512, http://www.chemspider.com/Chemical-Structure.1512.html (accessed 07:30, Mar 17, 2018).

Deshpande V., Wang Q., Greenlield P., Charleston M., Porras-Alfaro A., Kuske C. R., Cole J. R., Midgley D. J., Tran-Dinh N. (2016) Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* 108 (1): 1-5.

Ding F., Fin S., Hong N., Zhong Y., Cao Q., Yi G., Wang G. (2008) Vitrification-cryopreservation, an efficient method for eliminating *Candidatus Liberobacter asiaticus*, the citrus Huanglongning pathogen, from in vitro adult shoot tips. *Plant Cell Rep.* 27: 241-250

References

Edgar R. C. (2017) Updating the 97% identity threshold for ribosomal RNA OTUs. *bioRxiv* preprint.

Edgar R. C. (2017) SEARCH_16S: A new algorithm for identifying 16S ribosomal RNA genes in contigs and chromosomes. *bioRxiv* preprint.

Edgar R. C., Flyvbjerg H. (2015) Error filternig, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* 31 (21): 3476-3482.

Engelmann F. (2009) Use of biotechnologies for conserving plant biodiversity, *Acta Hortic*. 812: 63-82.

Engelmann F. (2004) Plant Cryopreservation: Progress and Prospects, *In Vitro Cell. Dev. Biol. – Plant* 40: 427-433

Fonseca E. D. S., Peixoto R. S., Rosado A. S., Baliero F. C., Tiedje J. M., Rachid C. T. (2018) The Microbiome of Eucalyptus Roots under Different Management Conditions and Ist Potential for Biological Nitrogen Fixation. *Microb Ecol*., 75 (1): 183-191.

Gardes M., Bruns T.D. (1993) ITS primers with enhanced specificity for Basidiomycetes: application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2: 113–118.

Gogvadze E. V., Buzdin A. A., Sverdlov E. D. (2005) Multiple template switches on LINE-directed reverse transcription: The most probable formation mechanism fort he double and triple chimeric retroelements in mammals. *Bioorganicheskaya Khimiya*, 31 (1): 74-81.

Haas B. J., Gevers D., Earl A. M., Feldgarden M., Ward D. V., Giannoukos G., Ciulla D., Tabbaa D., Highlander S. K., Sodergen E., Methè B., DeSantis T. Z., The Human Microbiome Consortium, Petrosino J. F., Knight R., Birren B. W. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* 21: 494-504.

Hamad I., Ranque S., Azhar E. I., Yasir M., Jiman-Fatani A. A., Tissot-Dupont H., Raoult D., Bittar F. (2017) Culturomics and Amplicon-based Metagenomic Approaches for the Study of Fungal Population in Human Gut Microbiota. *Nature Scientific Reports* 7: 16788.

References

Hambeck M. (2017) Cryopreservation – dealing with bacterial infections, Diploma Thesis, Univ. Wien.

Huang Q., Baum L., Fu WL. (2010) Simple and practical staining of DNA with GelRed in agarose gel electrophoresis. *Clin. Lab*. 56: 149-152.

Klindworth A., Pruesse E., Schweer T., Peplies J., Quast C., Horn M., Glöckner O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research.* 41 (1)

Knief C. (2014) Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Front. Plant. sci.* 5 Article 216

Kodym A., Senula A., Temsch E. M., Hood-Nowotny R., Schumacher F., Sekurova O. N., Zotchev S. B., Kiehn M. (2018) Micropropagation and cryoconservation of the endagered plant species *Artemisia laciniata* (Asteraceae). unpublished Manuscript

Laurence M., Hatzis C., Brash D. E. (2014) Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes. *PLoS One* 9 (5): e97876

Lee C., Kim S., Li W., Bang S., Lee H., Lee H., Noh E., Park J., Bang W., Shim S. (2017) Bioactive secondary metabolites produced by an endophytic fungus *Gaeumannomyces sp.* JS0464 from a maritime halophyte *Phragmites communis*. *The Journal of Antibiotics* 70: 737-742.

Lemos L. N., Morais D. K., Tsai S. M., Roesch L., Pylro V. (2017) Bioinformatics for Microbiome Research: Concepts, Strategies, and Advances. In: Pylro V., Roesch L. (eds.), *The Brazilian Microbiome*, Springer International Publishing, Cham, Heidelberg, New York, Dordrecht, London, S. 111-123.

Liaqat F., Eltem R. (2016) Identification and characterization of endophytic bacteria isolated from in vitro cultures of peach and pear rootstocks. *Biotech J*. 6: 120.

Ludwig-Müller J. (2015) Plants and endophytes: equal partners in secondary metabolite production? *Biotechnol Lett.* 37 (7): 1325-34.

References

Lutenberg B. (2015) Introduction to Plant-Microbe Interactions. In: Lutenberg B. (eds.), *Principles of Plant-Microbe Interactions*, Springer International Publishing, Cham, Heidelberg, New York, Dordrecht, London, S. 1-5

Magoc T., Salzberg S. L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27 (21): 2957-2963.

Maira O., Alexander M., Vargas T. E. (2010) Micropropagation and Organogenesis of *Anthurium andreanum* Lind cv Rubrum. In: Jain S. M., Ochatt S. J. (eds.), *Protocols for In Vitro Propagation of Ornamental Plants, Methods in Molecular Biology*, vol. 589, Humana Press, Springer, Heidelberg, New York, Dordrecht, London, S. 3-14

Manter D.K., Delgado J.A., Holm D.G., Strong R.A. (2010) Pyrosequencing reveals a highly diverse and cultivar-specific bacterial endophyte community in potato roots. *Microb. Ecol.* 60 (1): 157-166

Mardis ER. (2008) Next-generation DNA sequencing methods. *Annu Rev. Genomics Hum. Genet*. 9: 387-402.

Massimo N. C., Devan M. M., Arendt K. R., Wilch M. H., Riddle J. M., Furr S. H., Steen C., U'Ren J. M., Sandberg D. C., Arnold A. E. (2015) Fungal Endophytes in Aboveground Tissues of Desert Plants: Infrequent in Culture, but Highly Diverse and Distinctive Symbionts. *Microb. Ecol.* 70 (1): 61-76.

Matthews B. W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta* 405: 442-451

Mhlongo M., Piater L., Madala N., Labuschagne N., Dubery I. (2018) The Chemestry of Plant-Microbe Interactions in the Rhizosphere and the Potential for Metabolomic to Reveal Signaling Related to Defense Priming and Induced Systemic Resistance. *Front. Plant Sci.* 9: 112.

Müller-Esterl W. (2018) Biochemie. 3. Auflage, Springer-Verlag, Frankfurt, Berlin, S. 3-703

Murashige T., Skoog F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant* 15: 473-497.

Nagarajan N., Pop M. (2013) Sequence assembly demystified. *Nature Genetics 14:* 157 – 168.

National Center for Biotechnology Information. PubChem Compound Database; CID=1567, https://pubchem.ncbi.nlm.nih.gov/compound/1567 (accessed Mar. 17, 2018).

NCBI Resource Coordinators (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 41: 8-20.

Nishizawa S., Sakai A., Amano Y., Matsuzawa T. (1993) Cryopreservation of asparagus (*Asparagus* officinalis L.) embryogenic suspension cells and subsequent plant regeneration by vitrification. *Plant Sci* 91: 67-73.

Orlikowska T., Nowak K., Reed B. (2017) Bacteria in the plant tissue culture environment. *Plant Cell Tiss Cult.* 128: 487-508.

Ozudogru E. A., Previati A., Lambardi M. (2010) In Vitro Conservation and Cryopreservation of Ornamental Plants. In: Jain S. M., Ochatt S. J. (eds.), *Protocols for In Vitro Propagation of Ornamental Plants, Methods in Molecular Biology*, vol. 589, Humana Press, Springer, Heidelberg, New York, Dordrecht, London, S. 303-324

Peterson M., Daniel R., Danson M., Eisenthal R. (2007) The dependence of enzyme activity on temperature: determination and validation of parameters. *Biochem J.* 402(Pt2): 331-337.

Porras-Alfaro A., Bayman P. (2011) Hidden Fungi, Emergent Properties: Endophytes and Microbiomes. *Annu. Ev. Phytopathol.* 49: 291-315.

Porras-Alfaro A., Liu K., Kuske C. R., Xie G. (2014) From Genus to Phylum: Large-Subunit and Internal Transcribed Spacer rRNA Operon Regions Show Similar Classification Accuracies Influenced by Database Composition. *Appl. Enviro. Micro.* (80) 3: 829-840.

Porter-Jordan K., Rosenberg E. I., Keiser J. F., Gross J. D., Ross A. M., Nasim S., Garrett C. T. (1990) Nested polymerase chain reaction assay for the detection of cytomegalovirus overcomes

false positives caused by contamination with fragmented DNA. *Journal of Medical Virology.* 30: 85-91.

Probandt D., Eickhorst T., Ellrot A., Amann R., Knittel K. (2017) Microbial life on a sand grain: from bulk sediment to single grains. *ISME J.* 197: 1-11.

PubChem Compound Database; CID=1567, https://pubchem.ncbi.nlm.nih.gov/compound/1567 (accessed Mar. 17, 2018)

Qzast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner F. O. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web- based tools. *Nucleic Acids Research* 14: 590-596.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (23.04.2018)

Reed B. M. and Tanprasert P. (1995) Detection and control of bacterial contaminants of plant tissue cultures. A review of recent literature, *Plant tissue culture and Biotechnology*, 1(3), 137-142.

Riedel C. U., Schwiertz A., Egert M. (2014) The Stomach and Small and Large Intestinal Microbiomes. In: Marchesi J. R. (eds.), *The Human Microbiota and Microbiome*, CABI, London, S. 1-20.

Robinson P. (2015) Enzymes: principles and biotechnological applications. *Essays Biochem.* 59: 1-41.

Saddler G. S., Guimaraes P. M. (2015) Curtobacterium. *Bergey's Manual of Systematics of Archaea and Bacteria*. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118960608.gbm00096 (05.12.2018)

Schloss P. D., Westcott S. L., Ryabin T., Hall J. R., Hartmann M., Hollister E. B., Lesniewski R. A., Oakley B. B., Parks D. H., Robinson C. J., Sahl J. W., Stres B., Thallinger G. G., Van Horn D. J., Weber C. F. (2009) Introducing mothur: Open-Source, Platform-Indipendent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 75 (23): 7537.

References

Senula A., Keller J., Anduijav T., Yohannes T. (2007) Cryopreservation of cold-acclimated mint (Mentha spp.) shoot tips using a simple vitrification protocol. *CryoLetters* 28(1): 1-12.

Shendure J., Ji H. (2008) Next-generation DNA sequencing. *Nature Biotech 26*: 1135 – 1145.

Singer E., Bushnell B., Choleman-Derr D., Bowman B., Bowers R. M., Levy A., Gies A., Cheng J., Copeland A., Klenk H., Hallam S. J., Hugenholtz P., Tringe S. G., Woyke T. (2016) High-resolution phylogenetic microbial community profiling. *ISME.* 10: 2020-2032.

Thatamouni R. W., Shibli R. A., Abdallat A. M., Al-Qudah T. S., Younis L., Al-Baba H. and Ruwaiei H. (2015) *In vitro* conservation and cryopreservation of medicinal and aromatic plants: a review, *Jor. J. Agric. Sc.* 11: 147-159.

Thomas P., Swarna G.K., Roy P.K., Patil P. (2008) Identification of culturable and originally non-culturable endophytic bacteria isolated from shoot tip cultures of banana cv. Grand Nain. *Plant Cell Tiss Organ Cult*. 93: 55-63.

Toju H., Tanabe A., Yamamoto S., Sato H. (2012) High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. PLoS ONE 7(7): e40863

UNITE Community (2017) UNITE general FASTA release. Version 01.12.2017.

Walker J. J., Pace N. R. (2007) Phylogenetic composition of rocky mountain endolithic microbial ecosystems. *Appl. Environ. Microbiol.* 73: 3497-3504.

Wang M., Chen L., Zhang Z., Blystad D., Wang Q. (2018) Cryotherapy: A Novel Method for Virus Eradication in Economically Important Plant Species. In: Loyola Vargas V. M., Ochoa-Alejo N. (eds.), *Plant Cell Culture Protocols, Methods in Molecular Biology*, vol. 1815, Humana Press, Springer, Heidelberg, New York, Dordrecht, London, S. 257-268

Wang Q., Garrity G. M., Tiedje J. M., Cole J. R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73: 5261-5267.

References

White T.J., Bruns T., Lee S., Taylor J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR protocols a guide to methods and applications, *Academic Press*, San Diego: 315–322.

Xiao L., Qin J., Shen D., Liang C., Chen W., Liu C., Wang J. (2014) Next-generation Sequencing Methods to Investigate the Human Microbiome. In: Marchesi J. R. (eds.), *The Human Microbiota and Microbiome*, CABI, London, S. 147-181

Zhang J., Kobert K., Flouri T., Stamatakis A. (2013) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. 30 (5): 614-620.

Zoll J., Snelders E., Verweij P. E., Melchers W. J. G. (2016) Next-Generation Sequencing in the Mycology Lab. *Curr. Fungal. Infect Rep.* 10: 37-42.

10. Acknowledgments

First I would like to thank my supervisor Univ.-Prof. Dr. Sergey B. Zotchev for accepting me as his diploma student and for having the possibility of accomplishing my diploma thesis in his amiable working group. I am thankful for the trust he placed in me and for his innovative and informative insights.

Next, I would like to thank my co-supervisor Dr. Andrea Kodym for giving me the opportunity to realizing my diploma thesis under her supervision. I have learned a lot from her and I am truly thankful for having such a competent, fast working guide. I highly appreciate her support and encouragement in all matters. I furthermore am grateful for all the time she took to discuss my questions and concerns, when there was no easy answer, she found specialists that could remove all obstacles.

I would also like to thank my second co-supervisor Dr. Martina Oberhofer for teaching me practical methods, giving me advices trough bottlenecks and all the invigorating discussions that I had with her.

I would like to thank Dr. Craig Herbold for his indispensible insights in the study of microbial communities and Univ.-Prof. Dr. Thomas Rattei for giving me access to the LiSC and thusly making the analysis possible. I would like to thank Ass. Prof. Ovidiu Paun for his support, insight and everything I have learned from him in Bioinformatics.

My thanks also go to Ass. Prof. Dr. Christoph Wawrosch, Dr. Olga Sekurova, and Jaime Felipe Guerrero Garzón, MSc for their helpful hints and support they gave me. Furthermore, I would like to give thanks to Mag. Silvia Löbsch for her technical support. I would like to thank all my colleagues at the Department of Pharmacognosy for the enjoyable working atmosphere.

Last but not least I would like to thank my family and friends for their support and motivation.

Appendix

Mothur v.1.39.5 – pipeline

First the fastq files were divided into fasta and quality files into an own directory for further analysis. Then a stability.files table was created containing the name of the sample and their matching forward and reverse reads, spaced by a tab. This file was used to assemble the reads using the quality score from the fastq files; the reads and its reverse complement must have a quality above 25 to be considered real:

```
mothur >
fastq.info(inputdir=/home/user/vignolle/NGS_in_vitro/raw_data/16S_merged,
outputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fastq=A1p.V3V4.merged.fastq)

mothur > make.contigs(file=stability.files, processors=8)
```

After this step a summary file of each fasta file was created to have an overview of the assembled sequences:

```
mothur >
summary.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta= stability.trim.contigs.fasta)
```

The summary files contained the information that was used to specify the details for the implementation of the screening command that removes any sequences with ambiguous bases and anything longer than 404 bp. Clearly some reads did not assemble (these were kept for further analysis), and some were longer than the actual amplified site, these wrongly assembled reads were then removed:

```
mothur >
screen.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta= stability.trim.contigs.fasta, maxambig=0, maxlength=404)
```

Because in general it is computationally wasteful to align the same thing many times and it is expected that the files contain many sequence duplicates, a command was chosen to merge these duplicates. If two sequences were identical, they were considered duplicates of each other and merged:

```
mothur >
unique.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
```

```
fasta= stability.trim.contigs.good.fasta)
```

After this a table was created in which the number of unique sequences was counted:

```
mothur >
count.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
name= stability.trim.contigs.good.fasta)
```

Then the summary command was rerun with the *.unique.* fasta file as input:

```
mothur >
summary.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
count= stability.trim.contigs.good.count_table, fasta=
stability.trim.contigs.good.unique.fasta)
```

The sequences were now ready to be aligned. They were aligned to the non-redundant Ref SILVA database. Using the whole database would take a lot of computational time because the sequences would be aligned to the whole annotated genome of for example *E. coli*. This is not useful therefore the database was trimmed down to only the significant sites. First the database was downloaded and unzipped:

```
[.]$ wget -N https://www.arb-
silva.de/fileadmin/arb_web_db/release_132/Exports/SILVA_132_SSURef_Nr99_tax
_silva_full_align_trunc.fasta.gz
```

Before the database was modified the exact location of the V3-V4 region was selected. The used 16S amplicon PCR Forward Primer =

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG

16S amplicon PCR Reverse Primer =

5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

to amplify the V3-V4 region were needed. (Klindworth A. *et* al. 2013) Then the *E. coli* 16S rDNA sequence was downloaded. This sequence was trimmed with the primers. Important was, that the reverse complement of the reverse primer sequence was needed. This content was saved to a fasta file and aligned with the downloaded unzipped SILVA database:

```
mothur > align.seqs(fasta=ecoli_V3V4.fasta, reference=
SILVA_132_SSURef_Nr99_tax_silva_full_align_trunc.fasta)

mothur > summary.seqs(fasta=ecoli_V3V4.align)
```

The output of `summary.seqs` indicated the starting position 6388 and end position 25319. These coordinates were used for the modification of the database.

```
mothur >
pcr.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta= SILVA_132_SSURef_Nr99_tax_silva_full_align_trunc.fasta, start=6388,
end=25319, keepdots=F)

mothur >
rename.file(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
input= SILVA_132_SSURef_Nr99_tax_silva_full_align_trunc.fasta,
new=silva.V3V4.fasta)
```

Now the database was ready for further processing. The sequences could be aligned to it without wasting computational time. Mothur offers three choices for the search of the template sequence. The default search option by kmer was chosen. The kmer size was set to 8. The needleman algorithm was chosen for the alignment method. This penalizes the same amount for opening and extending a gap in the alignment. The threshold of 0.50 means that if 50% of the bases were removed in the alignment the alignment is categorized as insufficient and with the flip parameter set to true the reverse complement sequence was tried. The better alignment was then reported:

```
mothur >
align.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta= stability.trim.contigs.good.unique.fasta,
reference=silva.V3V4.fasta, processors=2)
```

After this step `summary.seqs` was re-run. Some sequences had a different start and end position, these deviants from the mode positions happened because of deletions or insertions at the ends of the alignments. This indicated a very poor alignment and is often due to unspecific amplification. Then `screen.seqs` command was used again to make sure the reads overlap the same region, as well as setting the maximum homopolymer length to 8, because the database doesn't include anything with a stretch of 9 or more of the same base in a row, and again `summary.seqs` .

```
mothur > screen.seqs(fasta=A1p.V3V4.merged.good.unique.align, count=
stability.trim.contigs.good.count_table, summary=
stability.trim.contigs.good.unique.summary, start=40, end=17052,
maxhomop=8)
```

Now the sequences overlapped the same alignment coordinates and they could be filtered to remove the overhang at both ends. As well as removing the gap characters ("-"), without losing any information:

```
mothur >
filter.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta= stability.trim.contigs.good.unique.good.align, vertical=T, trump=.)
```

The `unique.seqs` command was re-run because this may have created some redundancy among the sequences. In order to further de-noise the sequences they were pre-clustered for up to two differences between sequences. Generally one difference for every 100 bp of sequence is allowed, this would mean four in this case. But due to the fact that these guidelines have been developed for a higher error prone dataset, I have chosen to allow one difference for every 200 bp of sequence. The following command sorted the sequences by abundance going from most to least abundant. If their difference is within 2 nt of each other they got merged.

```
mothur >
pre.cluster(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline
fasta= stability.trim.contigs.good.unique.good.filter.unique.fasta, count=
stability.trim.contigs.good.unique.good.filter.count_table, diffs=2)
```

As many errors as possible were eliminated. It was time to tackle the next problem and to reduce possible chimeras by using the VSEARCH algorithm, which was an evolution from the Chimera Slayer. Chimeras wrongly increase the diversity of a bacterial community, although they are only hybrid products between multiple parent sequences created by the polymerase in the PCR-steps. (Haas B. *et al.,* 2011) These sequences had to be filtered and extracted from the dataset. The sequences were referenced to the most abundant sequences and if a sequence was flagged as chimeric it was only removed in the sample in which it was found and not in all samples (dereplicate=t). The default way to do this was to remove this sequence from all samples, nevertheless a less aggressive approach was chosen:

```
mothur >
chimera.vsearch(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline
fasta=
stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta,
count=
stability.trim.contigs.good.unique.good.filter.precluster.count_table,
dereplicate=t)

mothur >
remove.seqs(inputdir=/proj/in_vitro_microbiome/16S_V3V4/mothur_Pipeline,
fasta=
stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta,
accnos=
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vse
arch.accnos, count=
```

```
stability.trim.contigs.good.unique.good.filter.unique.precluster.count_tabl
e)
```

The final steps before the actual OTU based analysis was to extract unwanted sequences from the dataset. These were 16S rDNA fragments from chloroplast and mitochondrial sequences and wrongly amplified regions from Archaea or 18S rDNA. As well as removing "unknown" classifications, that couldn't be ordered into one of the domains by the classifier. A Bayesian classifier was used. (Wang Q. *et al.,* 2007)

```
mothur >
classify.seqs(inputdir=/home/user/vignolle/NGS_in_vitro/16S_V3V4/mothur_Pip
eline, fasta=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta
, count=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.count
_table, reference=silva.nr_v132.align, taxonomy=silva.nr_v132.tax,
cutoff=80)
```

After these steps the undesirables could be removed from the datasets, as well as creating an updated taxonomy summary file, that gave an overview of these exclusions.

```
mothur > remove.lineage(fasta=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta
, count=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.count
_table, taxonomy=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.nr_v1
32.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-
Eukaryota); summary.tax(taxonomy=current, count=current)
```

The following commands clustered the sequences into OTUs. The first command determined uncorrected pairwise distances between aligned DNA sequences. This approach did not store these distances into RAM but printed them directly into a file. Only distances below 0.03 were saved. Such a high value was chosen in order to be able to create a *.dist file to run the clustering command afterwards. These high distances were tolerated, even if a cutoff of 0.005 would have been the actual choice for further clustering.

```
mothur > dist.seqs(fasta=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
fasta, cutoff=0.03)
```

Now that the distance matrix was generated, it could be used to assign sequences to OTUs. Mothur implemented different clustering methods:

- Nearest neighbor: Each of the sequences within an OTU is at most X% distant from the most similar sequence in the OTU.
- Furthest neighbor: All of the sequences within an OTU are at most X% distant from all of the other sequences within the OTU.
- Average neighbor: This method uses the average distance between all sequences within an OTU and a sequence clustered to it must be at most X% distant from it. This is a compromise between the two algorithms above.
- Opti: The OTUs are assembled using metrics to determine the quality of clustering. The metric options available are clustering by Matthews correlation coefficient, sensitivity, specificity, true positives + true negatives, false positives + false negatives, … etc.

In the following clustering step the OTUs were assembled using the Matthews correlation coefficient (MCC) due to the fact that it could establish the quality of binary classification problems, because it included the balance of the four confusion matrix categories (false positive, false negative, true positive, true negative). The value of MCC reaches from -1 to 1. (Matthews B., 1975) The chosen parameter to select initial randomization for the Opticluster method was “singleton”. This way each sequence was randomly assigned to its own OTU and then they were clustered together.

```
mothur > cluster(column=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
dist, count=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
count_table, cutoff=0.20)
```

The consensus taxonomy for each OTU could be determined with the classify.otu command:

```
mothur > classify.otu(list=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
opti_mcc.list, count=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.
count_table, taxonomy=
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.nr_v1
32.wang.taxonomy, label=0.20)
```

USEARCH v.10.0.240 – pipeline

The first step was to filter out low quality reads. This was achieved by setting a maximum on the expected number of errors (E). The expected number or errors is the average number of errors in a large collection. E was set to 1, because generally the most probable number of errors of a filtered read is zero. (Edgar R. C. *et al.,* 2015)

```
[vignolle@accelerator ITS2_Fungi]$ /apps/usearch/10.0.240/usearch -
fastq_filter A1p.ITSb.merged.fastq -fastq_maxee 1.0 -fastaout
A1p.ITSb.merged.filter.fasta

usearch v10.0.240_i86linux32, 4.0Gb RAM (65.9Gb total), 8 cores
(C) Copyright 2013-17 Robert C. Edgar, all rights reserved.
http://drive5.com/usearch

License: herbolc9@univie.ac.at

00:04 70Mb    100.0% Filtering, 93.2% passed
     89840  Reads (89.8k)
      6118  Discarded reads with expected errs > 1.00
     83722  Filtered reads (83.7k, 93.2%)
```

Then the filtered samples were pooled together for further analysis.

```
[vignolle@accelerator usearch_pipeline]$ cat A1p.ITSb.merged.filter.fasta
A2p.ITSb.merged.filter.fasta A3p.ITSb.merged.filter.fasta
A4p.ITSb.merged.filter.fasta A5p.ITSb.merged.filter.fasta
A6p.ITSb.merged.filter.fasta A7p.ITSb.merged.filter.fasta
K1p.ITSb.merged.filter.fasta K2p.ITSb.merged.filter.fasta
K3p.ITSb.merged.filter.fasta K4p.ITSb.merged.filter.fasta
K5p.ITSb.merged.filter.fasta K6p.ITSb.merged.filter.fasta
K7p.ITSb.merged.filter.fasta M1p.ITSb.merged.filter.fasta
M2p.ITSb.merged.filter.fasta M3p.ITSb.merged.filter.fasta
M4p.ITSb.merged.filter.fasta M5p.ITSb.merged.filter.fasta
M6p.ITSb.merged.filter.fasta M7p.ITSb.merged.filter.fasta
>In_vitro.pooled.fasta
```

The sequences within the pooled samples file were de-replicated. In this step the sequences were compared on each position with each other, if they were identical over the full length they got de-replicated and relabeled with “Uniq” and the size annotations was added to the output sequence label.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
fastx_uniques In_vitro.pooled.fasta -fastaout In_vitro.unique.fasta -
sizeout -relabel Uniq

00:06 592Mb   100.0% Reading In_vitro.pooled.fasta
```

```
00:08 826Mb   100.0% DF
00:08 840Mb  1200455 seqs, 255649 uniques, 219792 singletons (86.0%)
00:08 840Mb  Min size 1, median 1, max 153964, avg 4.70
00:11 676Mb   100.0% Writing In_vitro.unique.fasta
```

The next step constructed a set of OTU representative sequences from the de-replicated reads. It used the UPARSE-OTU algorithm. Its goal was to identify a set of OTU representative sequences that had <97% pair-wise sequence identity with each other; chimeras were disregarded and all non-chimeric input sequences should match at least one OTU with >= 97% identity. Chimeras were also filtered by this command; it worked better than using a reference-based chimera filtering after clustering.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
cluster_otus In_vitro.unique.fasta -otus In_vitro.otus.fasta -uparseout
In_vitro.uparse.txt -relabel Otu

00:02 47Mb    100.0% 38 OTUs, 32 chimeras
```

To be able to build an OTU table the sequences of the samples had to be relabeled with the sample name.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
fastx_relabel A1p.ITSb.merged.filter.fasta -prefix A1_ -fastaout
A1p.ITSb.relabeled.fasta -keep_annots
```

Then the fasta files were concatenated to build the set of sequences that would match the OTUs.

```
[vignolle@accelerator usearch_pipeline]$  cat A6p.ITSb.relabeled.fasta
M3p.ITSb.relabeled.fasta A1p.ITSb.relabeled.fasta K5p.ITSb.relabeled.fasta
A7p.ITSb.relabeled.fasta M4p.ITSb.relabeled.fasta A2p.ITSb.relabeled.fasta
K6p.ITSb.relabeled.fasta K1p.ITSb.relabeled.fasta M5p.ITSb.relabeled.fasta
A3p.ITSb.relabeled.fasta K7p.ITSb.relabeled.fasta K2p.ITSb.relabeled.fasta
M6p.ITSb.relabeled.fasta A4p.ITSb.relabeled.fasta M1p.ITSb.relabeled.fasta
K3p.ITSb.relabeled.fasta M7p.ITSb.relabeled.fasta A5p.ITSb.relabeled.fasta
M2p.ITSb.relabeled.fasta K4p.ITSb.relabeled.fasta > In_vitro.pooled.fasta
```

The OTU table was created from the relabeled and filtered sequences before dumping singletons and low-abundance sequences. This was done because many of these sequences would map to an OTU and therefore increase sensitivity.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
otutab In_vitro.pooled.fasta -otus In_vitro.otus.fasta -otutabout
In_vitro.otutab.txt -mapout In_vitro.map.txt

00:00 40Mb    100.0% Reading In_vitro.otus.fasta
00:00 6.4Mb   100.0% Masking (fastnucleo)
```

```
00:00 7.2Mb   100.0% Word stats
00:00 7.2Mb   100.0% Alloc rows
00:00 7.2Mb   100.0% Build index
00:12 116Mb   100.0% Searching, 100.0% matched
83717 / 83722 mapped to OTUs (100.0%)
00:12 116Mb  Writing A1p.otutab.txt
00:12 116Mb  Writing A1p.otutab.txt ...done.
```

The next step aligned the file containing the OTUs to the reference UNITE database.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
usearch_global In_vitro.otus.fasta -db
utax_reference_dataset_01.12.2017.fasta -id 0.8 -strand both -alnout
In_vitro.otu.08.aln -uc In_vitro.otu.08.uc

00:01 90Mb    100.0% Reading utax_reference_dataset_01.12.2017.fasta
00:02 57Mb    100.0% Masking (fastnucleo)
00:03 58Mb    100.0% Word stats
00:03 58Mb    100.0% Alloc rows
00:06 211Mb   100.0% Build index
00:06 332Mb   100.0% Searching, 94.7% matched
```

All trees were created with the cluster_agg command. It uses the minimum linkage approach, equivalent with single linkage and finds the closest connected components of a graph. It used the minimum distance between two sequences to build the tree.

```
[vignolle@accelerator usearch_pipeline]$ /apps/usearch/10.0.240/usearch -
cluster_agg In_vitro.otus.fasta -treeout In_vitro.otus.tree -linkage min

00:00 40Mb    100.0% Reading In_vitro.otus.fasta
00:00 7.2Mb   100.0% Word stats
00:00 7.2Mb   100.0% Alloc rows
00:00 7.4Mb   100.0% Build index
00:00 83Mb    100.0% Distance matrix/usort
00:00 83Mb    100.0% Building tree
```

The OTU table In_vitro.otutab.txt and In_vitro.otu.08.aln were then manually combined into a taxonomy table:

```
#OTU ID   Domain    Phylum    Class     Order     Family    Genus     Species
Otu2      Plantae   Plantae   Plantae   Plantae   Plantae   Plantae   Plantae
Otu31     Plantae   Plantae   Plantae   Plantae   Plantae   Plantae   Plantae
Otu22     Plantae   Plantae   Plantae   Plantae   Plantae   Plantae   Plantae
```

DADA2 v1.8 – pipeline

The first step was to load the DADA2 package into the R environment.

```
library(dada2); packageVersion("dada2")
```

Then the path to the sequences was loaded into an R-object.

```
path <- "/proj/in_vitro_microbiome/dada2/16S_V3V4/fastq_files"
```

It was important to check the kind of format of the filenames to define the sample names correctly. Forward and reverse fastq filenames had the format: SAMPLENAME.V3V4.FWD.fastq and SAMPLENAME.V3V4.REV.fastq or for the ITS2 region SAMPLENAME.ITSb.FWD.fastq and SAMPLENAME.ITSb.REV.fastq.

```
fnFs <- sort(list.files(path, pattern=".V3V4.FWD.fastq", full.names = TRUE))
fnRs <- sort(list.files(path, pattern=".V3V4.REV.fastq", full.names = TRUE))
sample.names <- sapply(strsplit(basename(fnFs), "[.]"), `[`, 1)
```

To inspect the quality of the forward and reverse reads it was visualized by plotting. The first two samples of the dataset were enough to get an overview of the dataset.

```
plotQualityProfile(fnFs[1:2])
plotQualityProfile(fnRs[1:2])
```

This information provided the length where the reads were truncated. A directory for the filtered reads was created and their file-path defined for further placement.

```
filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path, "filtered", paste0(sample.names, "_R_filt.fastq.gz"))
```

The standard filtering parameters were used: maxN=0 (DADA2 does not require Ns in the sequences), truncQ=2, rm.phix=TRUE and maxEE=2 (expected errors). The expected error rate as a filtering parameter is more efficient than simply averaging quality scores. It was imperative to consider the amplicon data, because it was important to keep the length long enough so they could overlap and at the same time the quality scores needed to be high enough, especially for the in

length-varying ITS2 region.

```
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(280,230),
                     maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
                     compress=TRUE, multithread=TRUE)
head(out)
```

DADA2 uses a parametric error model. This following method learns the specific error model from the data itself by altering the inference of sample composition and estimation of the error rates until they meet on a mutually coherent solution. The initial random guess for this machine-learning problem was the maximum possible error rate in the given data based on the premise that the most abundant sequence was correct and all the rest are errors.

```
errF <- learnErrors(filtFs, multithread=TRUE)
errR <- learnErrors(filtRs, multithread=TRUE)
```

The following command visualized the estimated error rates:

```
plotErrors(errF, nominalQ=TRUE)
plotErrors(errR, nominalQ=TRUE)
```

This combined all identical sequences to “unique sequences” with a matching “abundance” equal to the number of reads of that particular sequence. Redundant comparison was eliminated and therefore computational time reduced. The de-replication was performed with the following command:

```
derepFs <- derepFastq(filtFs, verbose=TRUE)
derepRs <- derepFastq(filtRs, verbose=TRUE)
```

Then the derep-class objects were named with the sample names.

```
names(derepFs) <- sample.names
names(derepRs) <- sample.names
```

The following command inferred true sequence variants from the de-replicated unique sequences for each sample, which was achieved with the core sample inference algorithm of the DADA2 pipeline. Each sample was processed individually by default. However it was possible to increase sensitivity to low abundance sequence variants across multiple samples by pooling. Two types of pooling were possible in the DADA2 package. dada(…, pool=TRUE) was a standard pooling, in which all samples were pooled together. dada(…, pool=”pseudo”) was a pseudo-pooling, in which samples were processed individually after sharing information among samples, approximating pooled sample inference in linear time. Pseudo-pooling was chosen because it gave a more accurate description of

ASVs at very low frequencies (1-5 reads per sample) as well as detecting contaminants across many samples.

```
dadaFs <- dada(derepFs, err=errF, pool="pseudo", multithread=TRUE)
dadaRs <- dada(derepRs, err=errR, pool="pseudo", multithread=TRUE)
dadaFs[[1]]
dadaRs[[1]]
```

Then the forward reads were merged with the corresponding reverse read to form the merged "contig" sequences. The sequences were merged only if they overlapped with at least 12 bases.

```
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
head(mergers[[1]])
```

After this step the ASV table could be constructed.

```
seqtab <- makeSequenceTable(mergers)
dim(seqtab)
```

The distribution of sequence lengths was examined and the amplicons cut to size if necessary. With the variable ITS2 amplicons this step was not performed.

```
table(nchar(getSequences(seqtab)))
seqtab2 <- seqtab[,nchar(colnames(seqtab)) %in% seq(401,410)]
```

Chimeras had to be removed; the core DADA2 method corrected substitutions and indel errors but not chimeras. The precision of ASVs made it easier to recognize and eliminate them.

```
seqtab.nochim <- removeBimeraDenovo(seqtab2, method="consensus",
multithread=TRUE, verbose=TRUE)
dim(seqtab.nochim)

sum(seqtab.nochim)/sum(seqtab2)

getN <- function(x) sum(getUniques(x))
track <- cbind(out, sapply(dadaFs, getN), sapply(dadaRs, getN),
sapply(mergers, getN), rowSums(seqtab.nochim))

colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR",
"merged", "nonchim")
rownames(track) <- sample.names
head(track)
```

Thereafter it was possible to assign the taxonomy. The SILVA Database v128 was used for the 16S V3V4 ASVs and the UNITE Database v.7.2 was used for the ITS2 region.

```
taxa <- assignTaxonomy(seqtab.nochim,
"/proj/in_vitro_microbiome/dada2/16S_V3V4/fastq_files/silva_nr_v128_train_s
et.fa.gz", multithread=TRUE, tryRC=TRUE)

taxa.print <- taxa
rownames(taxa.print) <- NULL
head(taxa.print)

library(xlsx)
write.xlsx(taxa, "ITS2_taxa.xlsx")
write.xlsx(track, "ITS2_track.xlsx")
write.xlsx(seqtab.nochim, "ITS2_seqtab.xlsx")
```

SOP: DNA Extraction

With "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research
(Gabriel A. Vignolle, 2017, modified and based on Martina Oberhofer)

1. Being a time sensitive experiment, try to minimize and optimize your workflow by preparing your Solutions and Materials before starting the SOP. Always label all your Tubes.

2. Prepare the "**Genomic Lysis Buffer**" mixed with beta-mercaptoethanol to a final dilution of 0,5% (V/V) in a separate sterile tube. Only prepare as much "**Genomic Lysis Buffer**" mixed with beta-mercaptoethanol as needed. You will find beta-mercaptoethanol in the Poison cupboard. For preparation work in a ventilated Workbench due to toxic fumes.

3. Add up to 100mg of Material to be analyzed (e.g. *Mentha sp.*) into a ZR BashingBead™ Lysis Tube (0,1mm & 0,5mm).

4. Add 200 μl of autoclaved refrigerated Water (4°C) and 750 μl "**Lysis Solution**" vortex before usage (Room temperature) to the Tube.

5. Proceed with an adapted "SOP: grinding-bashing" for your Material. Use the Precellys 24 lysis and homogenization from Bertin Tech. (Important: Do not let your Material reach 37°C! Cool it down before and during processing in an icebox. Check the Temperature with a IR-Laser-Thermometer)

6. Centrifuge the ZR BashingBead™ Lysis Tubes in a microcentrifuge at 14.000 rpm for 1 Minute. Set the microcentrifuge to 20°C

   After this step you will work mostly with open Tubes. It is advised to wear a Mask to reduce contamination of your DNA.

7. Break of the tip of a "Zymo-Spin™ IV Spin Filter (Orange Top)" and place in a "Collection Tube".

8. Transfer up to 400 μl supernatant from the ZR BashingBead™ Lysis Tubes to a "Zymo-Spin™ IV Spin Filter (Orange Top)"

9. Centrifuge at 10.000 rpm for 1 minute

10. Place a "Zymo-Spin™ IIC Column" in a "Collection Tube", label the Column.

11. Add 1200 μl (2x600 μl for better handling) of "**Genomic Lysis Buffer**" to the filtrate in the "Collection Tube" from step 7. After this point all Waste has to be collected separately airtight due to the beta-mercaptoethanol and disposed of in an appropriate way.

12. Transfer 800 μl of the mixture from Step 11 to the "Zymo-Spin™ IIC Column" in a "Collection Tube". Centrifuge at 14.000 rpm for 1 minute

13. Discard the flow through from the "Collection Tube" and repeat Step 12.

14. Place "Zymo-Spin™ IIC Column" in new "Collection Tube"

15. Add 200 μl "**DNA Pre-Wash Buffer**" to the "Zymo-Spin™ IIC Column" and centrifuge at 14.000 rpm for 1 minute

16. Discard flow through

17. Add 500 μl "**g-DNA Wash Buffer**" to the "Zymo-Spin™ IIC Column" and centrifuge at 14.000 rpm for 1 minute

18. Discard flow through

19. centrifuge at 14.000 rpm for 1 minute "empty"

20. Discard flow through

21. Place "Zymo-Spin™ IIC Column" in a clean 1,5 ml microcentrifuge tube and add 50 μl "**DNA Elution Buffer**" directly to column Matrix. Let it drop on it.

22. Wait for 5 minutes

23. Centrifuge at 14.000 rpm for 30 sec.

24. Your DNA is now in the 1,5 ml microcentrifuge tube.

SOP: DNA Purification

With "*DNA Clean & Concentrator* ™ – 5 Kit" from Zymo Research
(Gabriel A. Vignolle, 2017, modified and based on Martina Oberhofer)

1. Buffer Preparation:
   Add 26ml 95% ethanol to the 6 ml "**DNA Wash Buffer"** concentrate
   Add 104 ml 95% ethanol to the 24 ml "**DNA Wash Buffer"** concentrate

   It is advised to wear a Mask to reduce contamination of your DNA.

2. Add 350 μl "**DNA Binding Buffer**" to your 50 μl DNA (7:1) and mix briefly by vortexing

3. Place "Zymo-Spin™ Column" in new "Collection Tube"

4. Transfer the mixture to the "Zymo-Spin™ Column" and centrifuge at 14.000 rpm for 30 seconds

5. Discard flow through

6. Add 200 μl "**DNA Wash Buffer**" to the "Zymo-Spin™ Column" and centrifuge at 14.000 rpm for 30 seconds

7. Discard flow through

8. Add 200 μl "**DNA Wash Buffer**" to the "Zymo-Spin™ Column" and centrifuge at 14.000 rpm for 30 seconds

9. Discard flow through

10. centrifuge at 14.000 rpm for 30 seconds "empty"

11. Discard flow through

12. Place "Zymo-Spin™ Column" in a clean 1,5 ml microcentrifuge tube and add 50 μl "**DNA Elution Buffer**" from the "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research directly to column Matrix. Let it drop on it.

    DO NOT USE "**DNA Elution Buffer**" from the "*DNA Clean & Concentrator* ™ – 5 Kit"

13. Wait for 5 minutes

14. Centrifuge at 14.000 rpm for 30 sec.

15. Your purified DNA is now in the 1,5 ml microcentrifuge tube.

SOP: Gel-electrophoresis
(Gabriel A. Vignolle, 2017, based on Martina Oberhofer)

1. Making the Gel: (Schott Bottle 500ml)
    - Create a **0,8% Agarose Gel**

3.2 g	**Agarose**
400 ml	**TBE-Buffer**

    - Heat up the mixture in the Microwave and check for floating Agarose particles. Do not close the cap tightly!
    - Be careful and take the Gel out and mix it to be sure it is without particles (hold against the light)
    - When the Gel is completely clear, do not let it cool down anymore.
    - Add 20 μl of "**Gel red**" to your Gel (carcinogen!)
    - Store the Gel in a oven and do not let it cool under 60°C

2. Pouring the Gel:
    - Prepare a suitable form adapted to your needs
    - Check that your form is in level

- Measure at the combs how thick your Gel should be and pour your hot Gel into the form
- Keep in mind that the thickness of your gel should be in proportion to the Volume you want to place in each individual Well
- Let it cool down until it is hard enough to remove the comb without the Wells collapsing (ca. 20min depending on the thickness of your Gel)

3. Loading the Gel:
    - Place your hardened Gel in a suited Gel-electrophoresis camber (filled with TBE-Buffer reuse for max. 3 times)
    - Pipette 1 μl Ladder in the first Well (e.g. 1kb-Ladder)
    - Mix your DNA with Loading dye in a Multiwell-plate that is kept on ice (2 μl DNA with 1 μl Loading dye) to reduce evaporation
    - Pipette your DNA (2,5 μl) mixed with Loading dye in the remaining Wells (one DNA-Sample per Well)

4. Running the Gel:
    - Set your Gel-electrophoresis camber manually to 100V
    - Your DNA will run from – to +
    - On the – Pol there are going to be bubbles ascending if turned on
    - The Wells of your Gel should be closest to the - Pol
    - Depending on your demands 3-4 cm running-length are sufficient (Running Time: ca. 20-40min)

5. Make a photo of your Gel with a suited apparatus (UV-light exposure)

6. Dispose of your Gel in a specifically labeled Trash (due to "Gel red" being a carcinogen)

SOP: grinding-bashing of lyophilized *in-vitro* Plant material
With ZR BashingBead™ Lysis Tube (0,1mm & 0,5mm)
from "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research
(Gabriel A. Vignolle, 2017)

This SOP includes steps 3. 4. and 5. of the **SOP: DNA Extraction**

1. The *in*-vitro Plant Material should be lyophilized in 50 ml sterile "Sterilin™" Eppendorf tubes or an equivalent.

2. Prepare DNA-free sterile Glass-beads:
    - Put your Glass-beads in a Schott Bottle 500ml
    - Wash them with deionized hand warm Water 3 times, be sure to mix it well
    - Then fill your Bottle with deionized hand warm Water to cover your Glass-beads then add Sodium-hypochlorite solution
    - Close the Bottle and wait 20 min.
    - Wash the Glass-beads with deionized hand warm Water 3 times, be sure to mix it well
    - Autoclave the Bottle with the Glass-beads

3. Place 3-4 Glass-beads under sterile conditions (Sterile workbench) into a Eppendorf tube together with your lyophilized *in-vitro* Plant material.

   Attention! Your Glass-beads must be completely dry before putting them into a tube! Let them dry out in a sterile petri dish in the sterile workbench!

4. Vortex your tubes for 5 min

5. If your material has not been sufficiently crushed, prolong the vortexing to 8 min

6. If your material has not been sufficiently crushed, split up the Sample in two 50 ml sterile "Sterilin™" Eppendorf tubes or an equivalent, repeat steps 3. And 4. with both tubes

7. Under sterile conditions weigh up to 100 mg (60-80 mg is better to handle) of your material and transfer it into a ZR BashingBead™ Lysis Tube (0,1mm & 0,5mm)
   - This is achieved by placing a appropriate scale into a sterile workbench
   - Use autoclaved weighing paper
   - Use sterile utensils to transfer your material
   - Be careful your material is volatile and electrostatic
   - To avoid contamination as good as possible take all precautions necessary

8. After this point, try to minimize and optimize your workflow by preparing your Solutions and Materials before continuing. Always label all your Tubes.

9. Clean your pipette thoroughly with ethanol 70% before usage

10. In the sterile workbench, Add 200 μl of autoclaved refrigerated Water (4°C) and 750 μl "**Lysis Solution**" from "*Quick*-DNA™ Fungal/Bacterial Miniprep Kit" from Zymo Research, vortex before usage (Room temperature) to the Tube.

11. Use the Precellys 24 lysis and homogenization from Bertin Tech. (Important: Do not let your Material reach 37°C! Cool it down before and during processing in an icebox. Check the Temperature with a IR-Laser-Thermometer)

12. Place your Tubes facing each other in the Precellys 24

13. Use following specifications:

    4200 – 1x10 – 030

14. After the first cycle place your Tubes immediately back in your ice-box

15. Wait for 5 min

16. Use following specifications:

    4200 – 1x10 – 030

17. After the second cycle place your Tubes immediately back in your ice-box

18. Wait for 5 min

19. Use following specifications:

    4200 – 1x10 – 030

20. After the third cycle place your Tubes immediately back in your ice-box

21. Proceed with step 6. of the **SOP: DNA Extraction**

SOP: PCR of the 16s rDNA

(Gabriel A. Vignolle, 2017, based on Martina Oberhofer and Bertfried Salem)

1. Prepare the Master Mix per Sample: Add in Order (from larger Volume to smaller)

1	H_2O (autoclaved)	30 μl
2	Buffer "Standard Taq Buffer"	4 μl
3	DMSO	2 μl
4	Prim F (10μM)	1 μl
5	Prim R (10μM)	1 μl
6	dNTP	0,5 μl
7	Taq	0,5 μl
	TOTAL	39 μl

Increase by 10% after calculation per Sample (4 Samples +10% … prepare for 5)
Preheat Buffer and vortex
Add all components by immersing the Pipette and pipette up and down (3x)
Take Taq out quickly, is liquid because it is stored in Glycerol, and only when needed, store back in freezer ASAP

When finished Vortex the Master Mix

2. Keep the Master Mix on Ice

3. Place PCR tubes in an suited tube holder stored at -20°C (changes color from blue to white when it warms up) label Tubes before usage

4. Add 39 μl Master Mix in the PCR Tubes

5. Add 1 μl of your DNA Sample and pipette up and down (3x)

6. Place PCR tubes into the Mastercycler (Eppendorf Nexus X “) (smaller holes)

7. User: admin Pin: 1234

8. Menu – Users – Martina – 16s Amplification – 64 or 32 wells – start

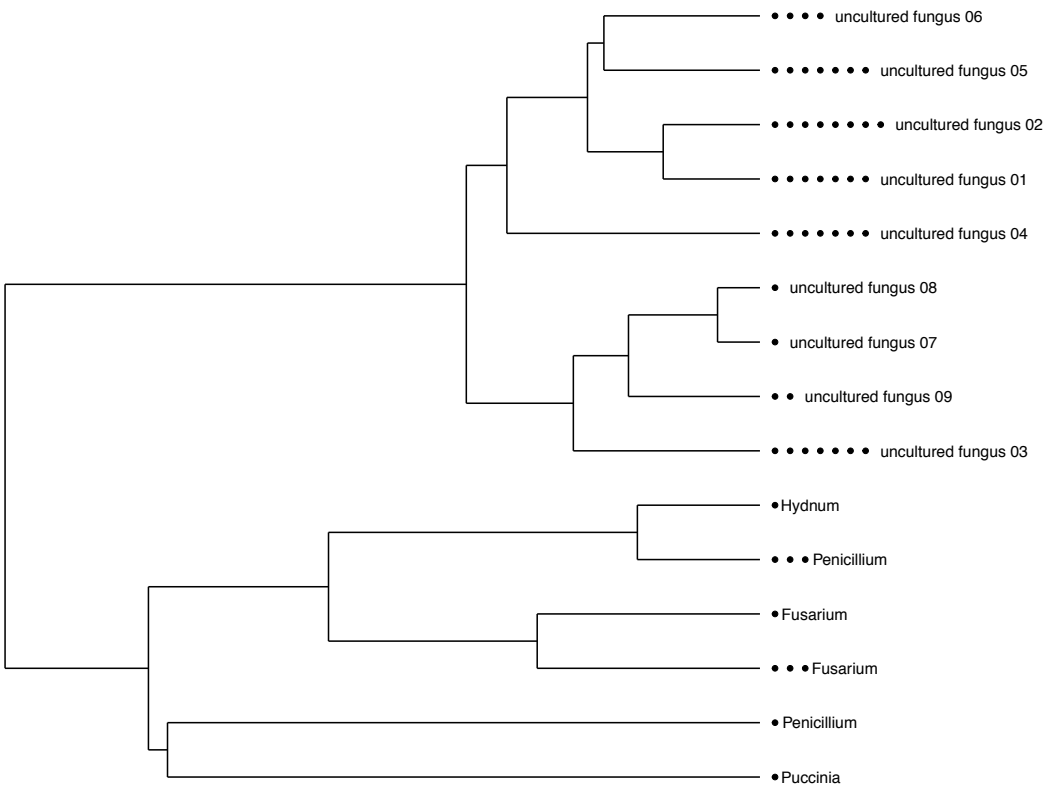
9. Control cycling conditions for applied Primerset:
(Calculate annealing Temperature if needed)

   27F / 1492R:
   95°C 2:00 ; 30x(95°C 00:30 ; 60°C 00:30 ; 72°C 3:00) 72°C 5:00

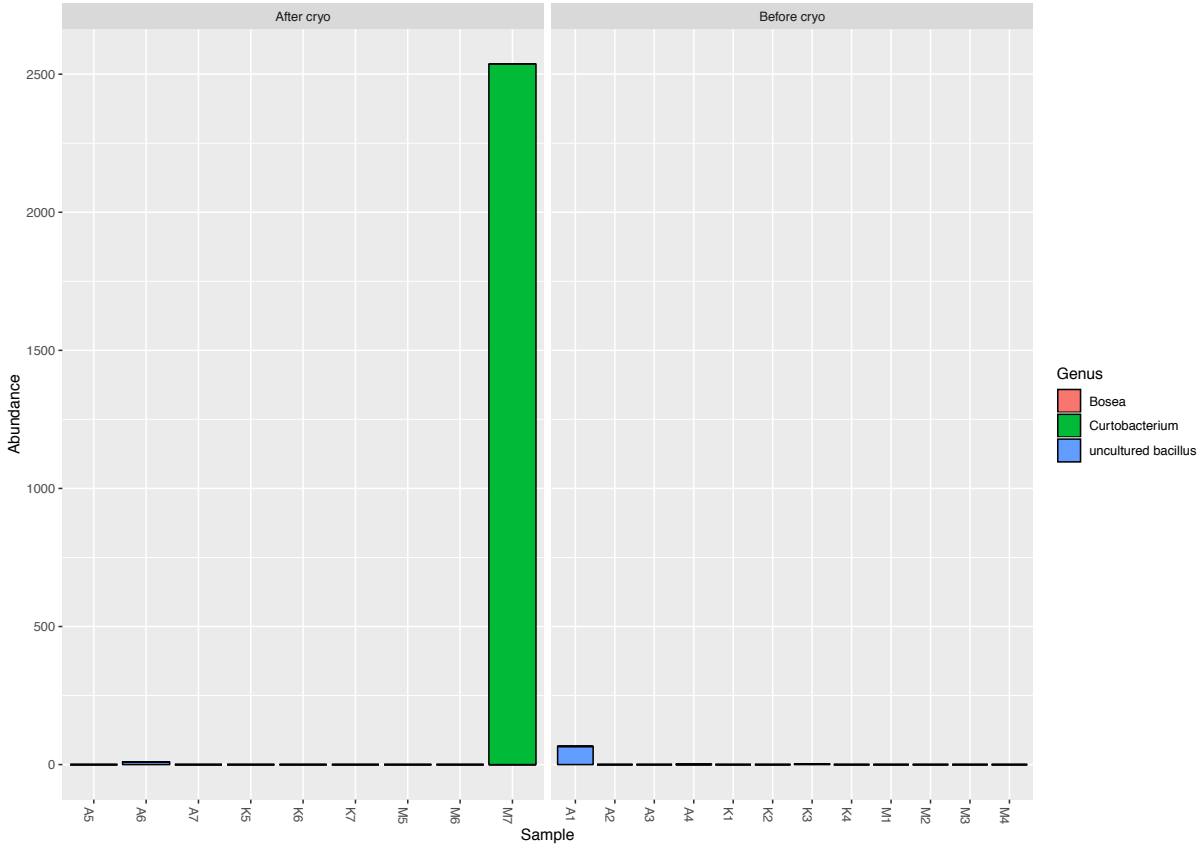
   799F / 1391R:
   95°C 2:00 ; 30x(95°C 00:30 ; 55°C 00:30 ; 72°C 3:00) 72°C 5:00

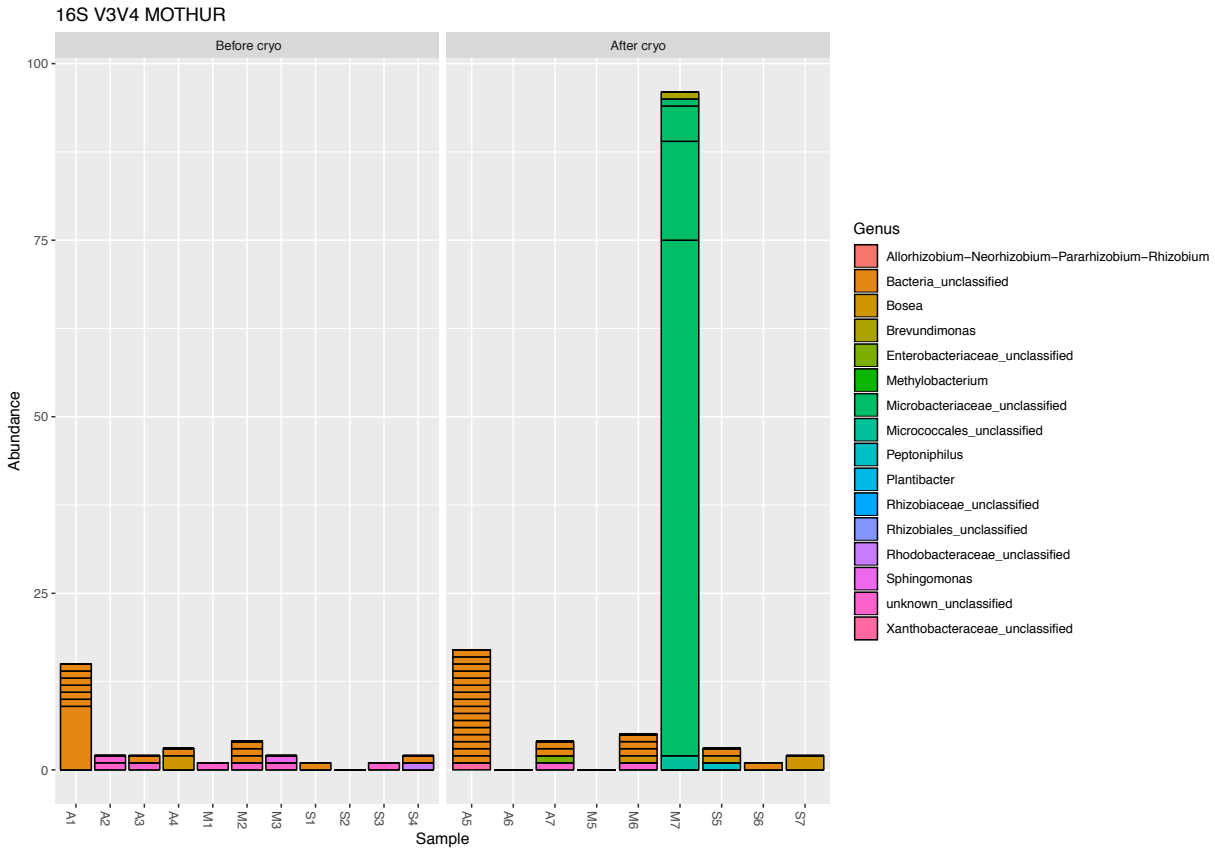
10. Finishing time is displayed

11. Tubes will be cooled down to 4°C after finishing



16S V3V4 DADA2





Rarefaction curve: