Bachelor Thesis B&M

DNA metabarcoding analysis of fungal diversity in soil of three vegetations in the Netherlands.

A study into the workflow of Naturalis for the ARISE soil project.

by

Sophie van Melis

Supervisor Naturalis Biodiversity Centre: Vincent Merckx
1st Supervisor Hogeschool Leiden: Ivo Horn
2nd Supervisor Hogeschool Leiden: Boet van Riel
January 2022

Abstract (EN)

Fungi can be spotted everywhere, the forest, a park, a grassland, or in the dunes. However, many fungi do not manifest in observable structures. They manifest in soil underground and have a significant impact on plant fitness and ecosystem functions. In 2017 it was estimated that at least 1.3 million species were still unidentified and only ~100.000 fungi have been characterized. There is no specific workflow to characterise the 'unknown' fungal species. Here we map the different fungi species in three vegetations with a soil sampling design and the metabarcoding method. By targeting the ITS2 region of extracted DNA and sequence the data with Illumina MiSeq technique, a standard workflow was set-up for identification of fungi diversity. We found that Ascomycota and Basidiomycota are the most common phyla in the park and the dunes. Other phyla detected were: Chytridiomycota, Glomeromycota, Monoblepharomycota, Mortierellomycota, and Rozellomycota. The diversity and variation of fungi is bigger in the park, compared to the dunes. Furthermore, the soil of the nature grassland likely contains inhibitors preventing accurate determinations of fungi. Our results demonstrate that many phyla are detectable in the soil with the workflow. We anticipate the use of the workflow to be an intermediate step for the research into fungal diversity in soil. For example, the removal of inhibitors from the soil without damaging DNA, or the difference in composition when seasonal changes occur, has yet to be researched.

Abstract (NL)

Schimmels kom je overal tegen, in het bos, een park, grasveld of in de duinen. Echter zullen niet alle schimmels in waarneembare structuren zichtbaar zijn. Ze manifesteren zich ondergronds en hebben een aanzienlijk effect op de gezondheid van een plant en de het ecosysteem. In 2017 waren er naar schatting nog minstens 1.3 miljoen soorten niet geïdentificeerd waren er maar ~100.000 schimmels beschreven staan. Er is geen specifieke aanpak beschikbaar die onbeschreven, en onbekende, schimmel soorten karakteriseert. Hier proberen we de schimmelsoorten in drie verschillende vegetaties in kaart te brengen met een design voor het nemen van samples en de metabarcoding methode. Door ons te richten op het amplificeren van het ITS2-gebied van geïsoleerd DNA en de gegevens te sequencen met de Illumina MiSeq methode, werd een standaard workflow opgezet voor de identificatie van schimmeldiversiteit. We ontdekten dat Ascomycota en Basidiomycota de meest voorkomende phyla zijn in het park en de duinen. Andere gedetecteerde phyla waren: Chytridiomycota, Glomeromycota, Monoblepharomycota, Mortierellomycota en Rozellomycota. De diversiteit en variatie van schimmels is in een park groter dan in de duinen. Verder bevat de bodem van het onbewerkte grasland waarschijnlijk inhibitors die identificatie van schimmelsoorten in de weg staan. Onze resultaten tonen aan dat veel phyla detecteerbaar zijn in de bodem met de metabarcoding workflow. We verwachten dat het gebruik van de workflow een tussenstap zal zijn voor het onderzoek naar schimmeldiversiteit in de bodem. Bijvoorbeeld, het verwijderen van inhibitors uit de bodem zonder DNA te beschadigen of het verschil in samenstelling bij seizoenswisselingen zijn stappen die getest kunnen worden.

Index

Introduction

Soil is a biodiverse environment; it contains many microbial communities [1]. Soil environments, only micrometres apart, can differ in their abiotic characteristics, and microbial community composition, therefore, be a diverse microbiome [1]. A group that is particularly important for soil biodiversity are the fungi [1]. The interaction between soil microbes, like fungi, and plants are important for the ecosystem, plant community, and the diversity [2].

Fungi play a major role in preserving the biodiversity. Many plants depend on the nutrient's fungi provide them. The exchange of nutrients of mutualists (fungi) are important for the heterotroph characterization of fungi [3], but the kingdom of Fungi also possess different strategies for survival [4]. They can interact as decomposers, mutualists, or as pathogens of plants and animals. The symbiosis between plant roots and fungi (referred to as mycorrhizae) enables plants to acquire nutrients and water in exchange for photosynthetically derived sugars [5]. Although a plants condition can change due to numerous biotic and abiotic factors [5], like pH, organic carbon concentration, nitrogen concentration, animal burrows, and water flow paths [1]. The absence of appropriate fungi can have a significant impact on the plant and alter the structure of a plant community or the plants fitness *[5,6]*.

Mycorrhizae fungi form a fungal network for communication and exchanging nutrients with trees and plants. The fungi hyphae can reach a length of hundreds of meters per gram of soil. *[5,8]*. The hyphae can extend to enter the cell of plants (Endomycorrhizae) or compose a Hartig net (inward growing hyphae) over the cortical cells of plants and grow in the extracellular space (Ectomycorrhizae) [3]. Microbial factors like life forms, nutritional strategies, and associations with other organisms can influence a plant. About 100.000 fungi have been characterized. In 2017 it was predicted that at least 1.3 million species were still unidentified [9]. 'Unknown' fungi species are largely uncharted, this might be because many fungi do not manifest in observable structures. Furthermore, many different species are hidden under established specie names. And lastly, geographic areas are still largely understudied [9]. With the development of high-throughput sequencing, we are now able to assess the diversity of soil fungi.

This aim of this study is to get insight into the biodiversity of the soil in the Netherlands and map the different fungi species with a sampling design of three vegetations near Leiden, a park, a nature grassland, and the dunes. Additionally, to learn for the benefit of the ARISE soil project and test and refine the metabarcoding workflow, focusing on the sampling strategy.

Method in short

To characterize and order fungi found in soil samples various techniques will be used. A checkerboard sampling design suggested by *Arita and Rodriguez (2002)* [10] will be used for sample collection, to study how fungal diversity differs in relation to sampling distance. For the DNA extraction, coated magnetic beads and the Kingfisher Flex purification robot will be used. A mix of PCR primers will amplify the inter transcribed spacer (ITS), ITS region 2, part of the universal barcode for fungi. After the clean-up, the adapters and identifying indexes will be added. With Miseq Illumina High-Throughput sequencing the samples will be sequenced, by Baseclear B.V (Leiden). A DADA2 ITS pipeline creates an amplicon sequence variant (ASV) table for data-analysis. The analysis of the data is supported by statistical tests.

Background and Principles

DNA extraction

KingFisher Flex robot. The KingFisher Flex is a magnetic particle processor designed for automated transfer and processing of particles. The purification system is designed with a magnetic head (or rod) and uses a disposable tip comb and plastic plates for processing. The KingFisher robot can isolate DNA, RNA, and proteins, and can be used for cell purification [11].

The KingFisher Flex has multiple advantages compared to an extraction kit. Firstly, any protocol can be created manually or uploaded, within the limits of the KingFisher Flex [11,12]. Second, the system is a fully automated system to yield high-speed purification [11,12]. Third, the magnetic head has 96 magnetic rods, because of that 96 samples can be processed at once. At last, it enables a more efficient wash due to the technology used; it moves the magnetic particles instead of the liquids. This way, cross contamination is limited to a minimum [12].

Beads. An important component in DNA extraction, using the KingFisher, are beads. Beads (from an extraction kit) are 'magnetic' particles with a smooth surface [11]. Beads exhibit magnetic properties when they are placed in a magnetic field but lose magnetism when they are removed from the field [11, 12].

DNA amplification

Primers. The primers were designed for the ITS2 region, as suggested by *Tedersoo (2014)* [13]. A mix of multiple 5 forward primers were used to increase the chances of matching with all fungi species in the soil. These primers are considered as universal fungal primers and are commonly used in fungal community analyses.

ITS region. The Internal Transcribed Spacer (ITS) region of the ribosome encoding genes is a commonly used marker for many fungal groups [14, 15]. The ITS region is a conserved region, the variability is sufficient to distinguish between closely related species [16].

Amplifying the complete ITS region (ITS1 and ITS2) is suggested to have biases against species with longer amplicons. Producing shorter amplicons can solve the bias [15]. ITS region 1 and 2 vary in amplicon length. Sequence variability of ITS2 region is useful with a fast and specific diagnosis of fungi-DNA-extracts. PCR, using fungi-specific primers, focused on conserved sequences of the 5.8S- and 28S-ribosomal DNA (rDNA) results in the amplification of ITS2 [16, 17].

Illumina sequencing. Sequencing by synthesis is a next generation sequencing (NGS) technique. A volume of one of the four nucleotides required for DNA synthesis is added to all the wells. If the bases on the template strand is complementary to the nucleotide, the nucleotide joins the growing strand, releasing PP_i (pyrophosphate). The release of PP_i, per nucleotide, causes a flash of light that is recorded [18].

Bioinformatics. The DADA2 ITS pipeline workflow is a rapid and accurate sample inference from amplicon data [19]. The pipeline works with Illumina-sequenced paired-end fastq files (input), and it creates an amplicon sequence variant (ASV) table, providing records of the number of times each amplicon sequence variant was observed in each sample. Compared to operational taxonomic unit (OTU) tables, ASV tables have a higher-resolution analogue resulting in analysing biological differences of even 1 or 2 nucleotides and fewer false-positive sequence variants [19, 20].

Method

Study sites. The locations for sampling were chosen based on the vegetation of the soil. The three locations were: a park, a nature grassland, and the coastal dunes near Naturalis, Leiden, the Netherlands. Location A: a park, named 'Leidse Hout' (52.176954, 4.477630). Location B: a nature reserve with grassland characteristics 'Lentevreugd' (52.163225, 4.391914). Location C: the dunes 'Berkheide' (52.164047, 4.392443) (appendix A).

Sampling design. The design for sample collection was suggested by *Arita and Rodriguez (2002)* and described by *Gavito et al. (2019)* [10, 21] (Figure A). The total plot can be divided into three similar scale plots (S1-S3), connecting in a diagonal arrangement. The largest scale is the 80 x 80 m plot (A0), divided into three extra spatial scales, divided into three additional scales. Plot A0 consists of 4 subplots of 40 x 40 m (A1=1600 m2), 16 subplots of 20 x 20 m (A2=400 m2), or 64 subplots of 10 x 10 m (A3=100m2). A sample was taken in the centre of 32 of the 64 subplots, following a checkerboard design.

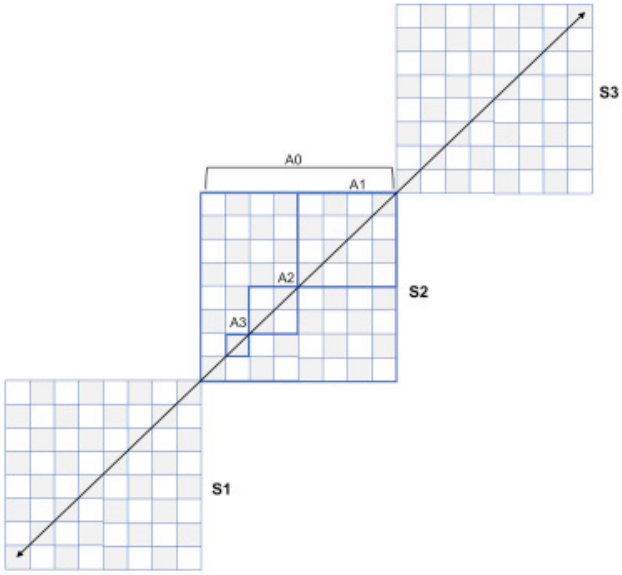


Figure A: the squared sample collection design, suggested by *Gavito*.
The plot has three subplots (S1 – S3) showing a checkerboard design. The measurements of the largest scale are 80 x 80 m and has four scales (A1=1600 m2, A2=400 m2, A3=100m2). The sampling square was triplicated in a diagonal arrangement. A sample was taken in every shaded square of A3 [10, 21]

Design sampling tubes. Tubes for sampling were developed out of plastic PVC-tubes, with a diameter of 3,2cm. Tubes were sawn with an electric tool to the length of 17 cm, with a slanting (Fig. B). The edge in the tube design made it possible to pierce through hard layers of soil.

Figure B: sampling tubes
Homemade sampling tubes of plastic PVC-tubes. Length 15cm on the short side, and 17cm on the long side, Ø 3,2cm. The tube was sawn with a slanting, the slanted side pierced the hard top layer of soil and pushed through until 10cm.

Sample collection. Samples for DNA extraction were obtained from the specific locations (appendix A). Following a checkerboard design, a sample was taken in the centre of 32 of the 64 subplots (Fig. A). The tube was hammered down 10 cm of the soil layer and the soil was subsequently placed in a Ziploc bag. The Ziploc bag contained a tag to label the samples and given a register number. The samples were stored in the fridge, at 4 degrees Celsius, until further use.

DNA extraction. DNA was extracted with the MagAttract Powersoil DNA KF kit (QIAGEN) [22] and the KingFisher Flex System (ThermoFisher) [23]. Changes were made to the manual extraction of the kit. The samples of each location were extracted separately to avoid contamination between sites. The positive control came from the Naturalis collection (TH9240) and was identified as the *Lactarius sp.* (taken in Guyana). The speed and time of the centrifuge step was adjusted to the limits of the centrifuge rotor. Following the Tissuelyser II (QIAGEN), the plates were centrifuged at 2250xg for 9 minutes. The same speed was used after adding the IR Solution. Avoiding the residual pellet, no more than 450µL supernatant of each sample was transferred into a KingFisher 96-DeepWell plate. The remaining supernatant was put in storage (4 degrees Celsius) in a second KingFisher Deep Well plate. The ClearMag Beads (Zorb Reagent) was resuspended properly. The ClearMag Beads was added to the ClearMag Binding Solution and mixed well. To ensure uniform distribution of the beads to each well, the pipette reservoir was put on a plate shaker (IKA MS3). The solution was added to each well containing the lysate. Three wash plates were prepared with the ClearMag Wash Solution and RNase free water was used as an elution buffer.

Loading the KingFisher. The KingFisher was pre-set with the correct protocol, named: KF_Flex_MoBio_PowerMag_RNA_DNA_96DW. Before proceeding with the DNA amplification, the concentration and purity of DNA was confirmed by measuring the concentration with the Trinean DropSense 96, and DropQuant v1.5 following the user manual [24].

Primers. Primers were designed for the specific use of the amplification of the ITS2 region. Following the suggestions of *Tedersoo (2014),* a mix of five forward primers (ITS3NGS1-5) and one reverse primer (ITS4NGS) was used to amplify the ITS2-region (Table A) [13]. The amplification creates a ±500bp product, the tail for the MiSeq index included.

Table A: '5-to'3 sequence of ITS3NGS1-5 and ITS4NGS primers, and matching organisms. The product (ITS2 region) is 385bp, ±500bp with the tail for the Miseq index [13].

Name	Sequence (primer)	Matching taxa
ITS3NGS1 (fwd)	CTAGACTCGTCATCGATGAAGAACGCAG	95% of all fungi
ITS3NGS2	CTAGACTCGTCAACGATGAAGAACGCAG	Chytridiomycota
ITS3NGS3	CTAGACTCGTCACCGATGAAGAACGCAG	Sebacinales p. parte
ITS3NGS4	CTAGACTCGTCATCGATGAAGAACGTAG	Glomeromycota
ITS3NGS5	CTAGACTCGTCATCGATGAAGAACGTGG	Sordariales p. parte
ITS4NGS (rev)	TCCTSCGCTTATTGATATGC	>99% fungi, plants, and most protists

PCR. The PCR was carried out using a 10x dilution of the DNA extract, based on results from the gel-electrophoresis (1,5% agarose, 1x TAE, and 8µL/80mL SYBR Safe). PCR was performed with a T100 Thermal cycler (BIORAD). The reaction mixture (20µL final volume) consisted of 1x PCR buffer (TaqMan Environmental Master Mix 2.0, ThermoFisher), 10pMol/µL of Forward primer (containing five forward primers, equimolar concentration), 10pMol/µL of Reverse Primer, MilliQ (water), and 2µL of DNA template.
The following cycling parameters were used for amplification: Initial denaturation at 95°C for 10 minutes; 35 cycles of denaturation at 95°C for 15 seconds, annealing at 50°C for 30 seconds, extension at 72°C for 40 seconds, and a final extension at 72°C for 10 minutes. The quality of the PCR product was checked by E-Gel (E-Gel™ 96 wells with SYBR™ Safe DNA Gel Stain, 2% agarose, Invitrogen) following the manufacturer's protocol [25].

Before they were labelled with sample-unique labels, the PCR products were cleaned with Macherey-Nagel NucleoMag C-Beads (MN Beads), using a magnetic extractor stamp. The MN-Beads clean-up was carried out following the standard Naturalis protocol. The amount of beads used was 0.9 times the PCR products' volume. The purified PCR product was transferred to a PCR plate and was ready for adding adapters and labels with a second PCR.

Sample labelling. The purified PCR products were labelled with three different label sets of MiSeq Nextera XT labels. The reaction mixture (20µL final volume) consisted of 1x PCR buffer (Taqman Environmental Master Mix, ThermoFisher) and MilliQ. The labels, 10pMol/µL of NXT_S, 10pMol/µL of NXT_N, and 3µL of DNA template, were added directly to the PCR plate. The following cycling parameters were used for amplification: Initial denaturation at 95°C for 10 minutes; 8 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 1 minute, extension at 72°C for 30 seconds, and a final extension at 72°C for 7 minutes. After adding the labels and adapters the product was cleaned with the MN-Beads clean-up, using the standard amount of beads (0.9 times the sample volume) The quality and concentration of each sample was validated with the Fragment Analyser, following manufacturer's protocol [26].

Pooling. After labelling each sample with an individual barcode, the samples were pooled and normalised manually, to a final volume of 20µL. To ensure that each sample is equally represented in the end pool, Excel (Microsoft) was used to calculate the concentration and volume, the lowest concentration determined the volume added to the end pool. The concentration of the end pool was determined with the 4150 Tapestation (Agilent) [27].

Sequencing. The sequencing of the final pool was done by Baseclear B.V, Leiden on an Illumina MiSeq sequencing platform. Baseclear requires 20µL of a ≥ 5nm / pool to be able to sequence the amplicons [28].

Bioinformatics. The DADA2 ITS pipeline workflow (v1.8) was used in Rstudio to create an amplicon sequence variant (ASV) table [19, 29]. Taxonomy was assigned to the output ('ITS sequence variants') using the UNITE database (v8.3) and the implementation of the naive Bayesian classifier method. The sequences were filtered according to the threshold value (100%) for similarities, which means when one nucleotide in the sequence cannot be assigned to any of the existing ASVs it became a new group, ASV. A key component in the DADA2 ITS pipeline was the identification and removal of primers from the reads (in any orientation).

Data-analysis. The ASV table created by the DADA2 pipeline was used to analyse the data with MicrobiomeAnalyst.ca [31]. Using the 'Marker Data Profiling' option, a comprehensive, statistical, and visual analysis on the DADA2 pipeline output were completed. The analysis was done to assess the difference between microbial communities or to visualise the taxonomic composition of microbial communities with alpha-diversity analysis and beta-diversity analysis. The species richness was determined with the use of an index, Chao1 [32]. The analysis was supported by statistical tests, like Permanova and T-tests.

Results

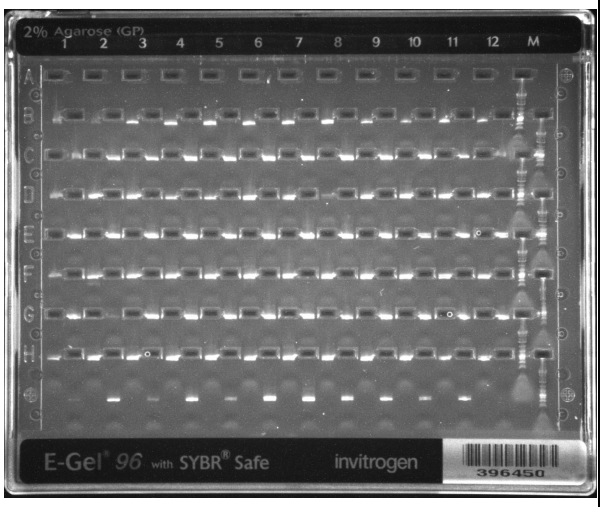
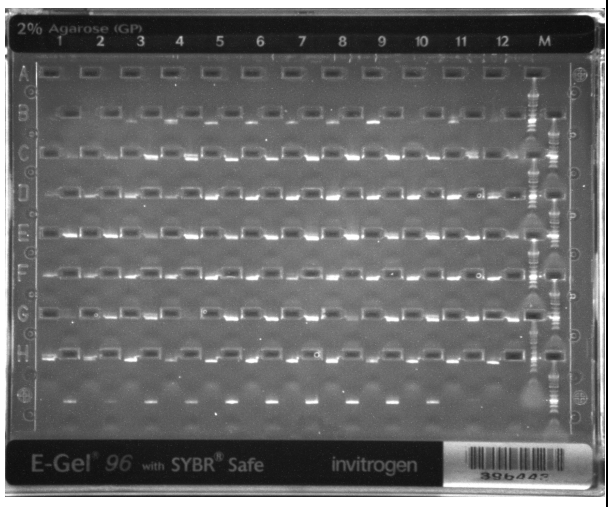
The workflow of the metabarcoding method for fungi diversity starts with sampling of soil.

Sampling. The purpose of sampling was to collect samples following an intense scale / plot to gain insight on the diversity in soil. Three different locations were chosen based on the specific characteristics of the soil, such as a park, a nature grassland, and the dunes. A total of 285 samples were taken at the specific locations (appendix A).

A control step of the experiment is to test the presence of DNA in all the samples and the success of the DNA extraction, this was done with E-gel96.

PCR product analysis. The dilution of the PCR-product (for further processing) was determined with agarose gel-electrophoresis. To check if DNA extraction for all samples was successful, an E-Gel96 was performed (Table B). Location A and C showed bands at the 500bp marker (Table B). On the E-Gel96 of location B no bands were visible in any dilution.

Table B. E-Gel96 results for location A and location C, 10x dilution of the PCR-product. Positive control (position A1) and negative control (position H12) included in the E-Gel96. The marker (M) is the 100bp, bands visible at 500bp.

Location A	Location C
	

The success rate of the E-Gel96 for location A is 96,8% (91/94 samples showed bands) and location C is 95,7% (90/94 samples showed bands).

After the labelling of adapters and the clean-up the extract of each sample was pooled. A concentration of 8,0 nmol/pool was used for Illumina sequencing at Baseclear. After sequencing, the raw data (Illumina-sequenced fastq-files) was put through the DADA2 pipeline in Rstudio and could it be analysed.

Diversity of fungi. The fungal diversity is based on the number of reads determined the specific locations of the sampling plot by the DADA2 pipeline. The number of times an amplicon sequence variant was observed can be calculated and visualised.

There are 2.581.549 reads sequenced, of which are 2.112.745 fungi reads (81.84%), and 82.094 reads (3,18%) are unidentified. The reads are merged into 9.297 amplicon sequence variants

(ASVs), of which 5.040 are fungi ASVs (54,21%) and 1.802 are unidentified ASVs (19,38%). With a 100% threshold for the confidence of taxonomic identification, the nine phyla determined are Ascomycota, Basidiomycota, Chytridiomycota, Glomeromycota, Monoblepharomycota, Mortierellomycota, Mucoromycota, Rozellomycota, and Zoopagomycota. At location A is Ascomycota the most common phylum, 65,51%, followed by Basidiomycota, 29,00% (Figure C). The remaining 5,49% is distributed over eight more phyla: Chytridiomycota, Glomeromycota, Monoblepharomycota, Mortierellomycota, Mucoromycota, Rozellomycota, and Zoopagomycota.

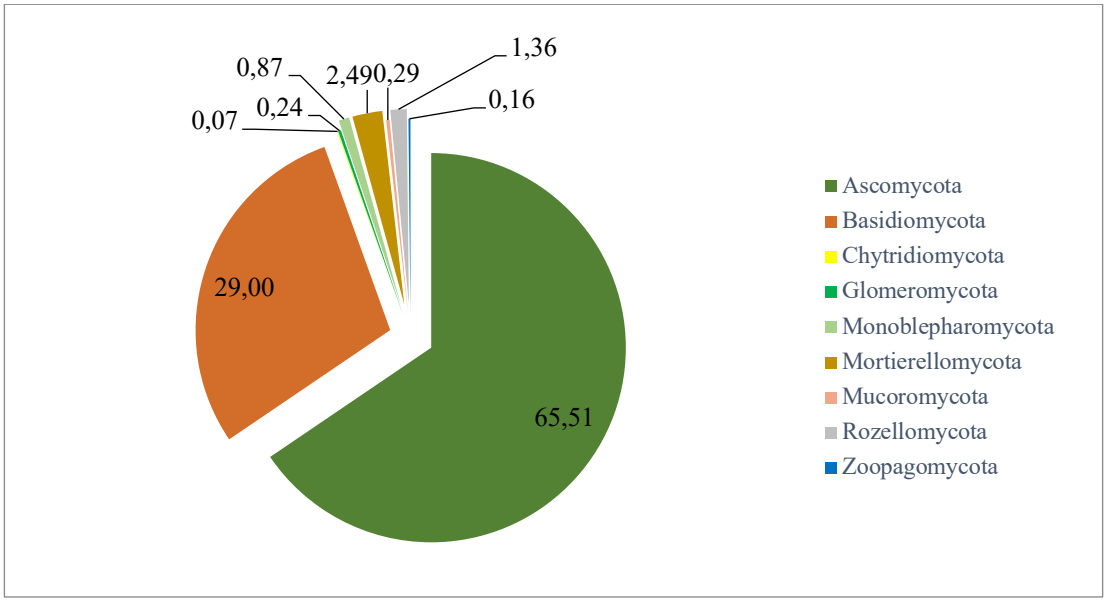


Figure C. The fungi diversity (amount of reads in %) of location A. Ascomycota (65,51%) and Basidiomycota (29,00%) are the most common phyla. Chytridiomycota (0,07%), Glomeromycota (0,24%), Monoblepharomycota (0,87%), Mortierellomycota (2,49%), Mucoromycota (0,29%), Rozellomycota (1,36%), and Zoopagomycota (0,16%), make-up the remaining phyla. 0,04 %

At location C is Ascomycota the most common phylum, 95,46%, followed by Basidiomycota, 3,50% (Figure D). The remaining 1,04% is divided over five phyla: Chytridiomycota, Glomeromycota, Monoblepharomycota, Mortierellomycota, and Rozellomycota.

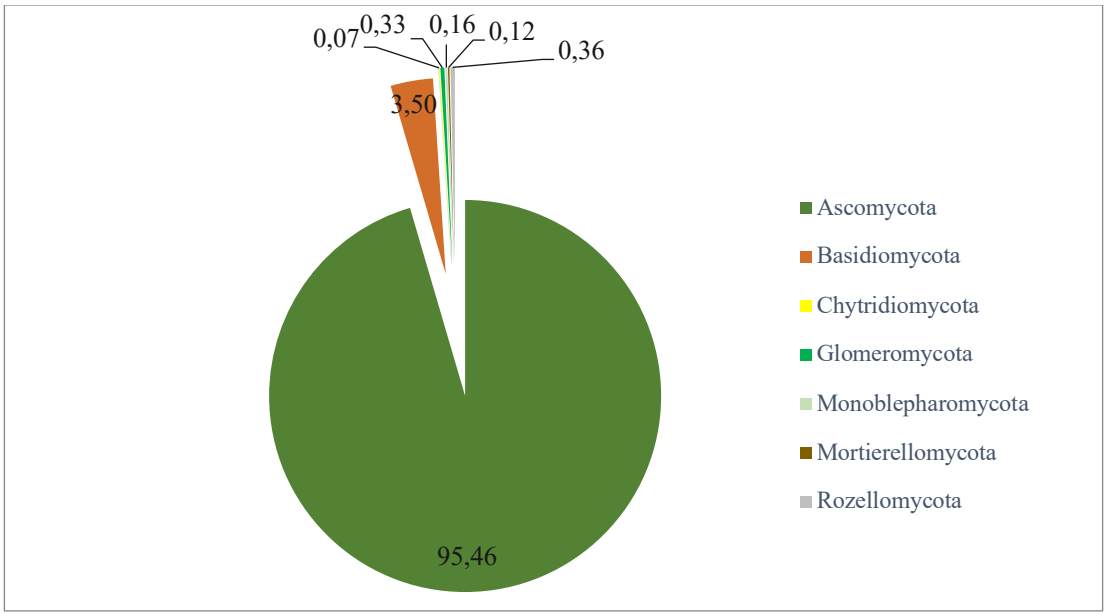


Figure D. The fungi diversity (amount of reads in %) of location C. Ascomycota (95,46%) and Basidiomycota (3,50%) are the most common phyla. Chytridiomycota (0,07%), Glomeromycota (0,24%), Monoblepharomycota (0,87%), Mortierellomycota (2,49%), Mucoromycota (0,29%), Rozellomycota (1,36%), and Zoopagomycota (0,16%), make-up the remaining phyla.

At location A and location C, Ascomycota was the most common phyla followed by Basidiomycota. To take a closer look into the fungi diversity and variation per sample within a location, additional analysis was done supported by statistical tests.

Diversity analysis. The mean diversity of species in different sites or habitats within a local scale was calculated (Figure E). The analyses, alpha-diversity analysis, compares location A (orange) and Location C (green) and the diversity of each sample. The spread of the dots shows the diversity and variation within each location. The significant difference was calculated with a T-test, p-value = 4,805e-07.

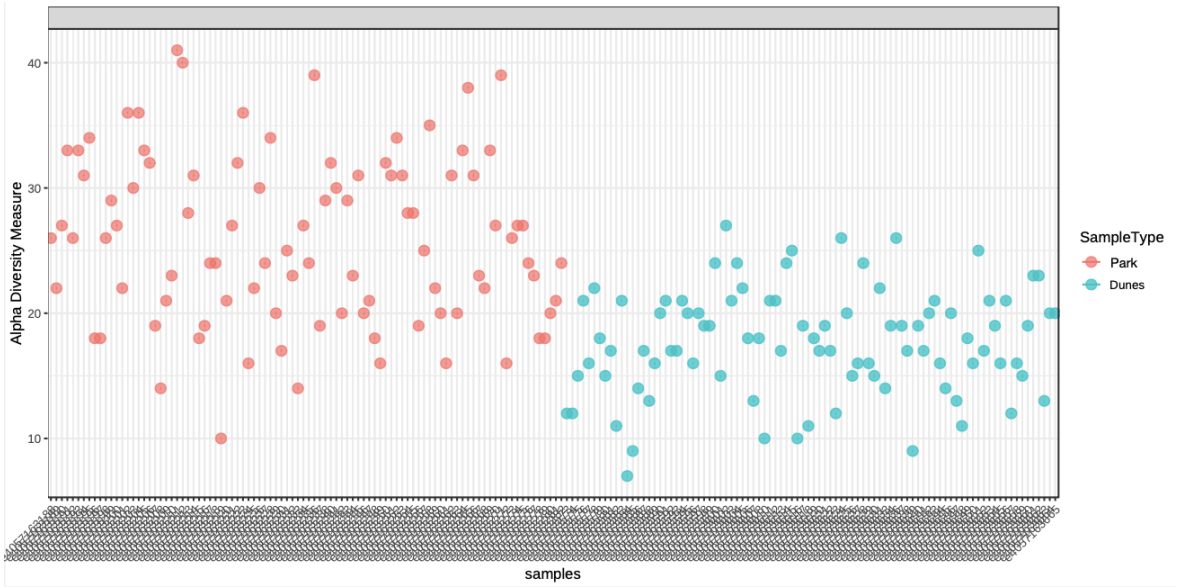


Figure E. Alpha-diversity analysis of Location A (orange) and Location C (green). Each dot corresponds to a location. The x-as shows the sample barcode and the y-as is the number of ASVs observed. The confidence calculated with T-test, P-value 4,805e-07.

The range of the dots for location A (orange) is wider and the alpha-diversity measure had a higher value, compared to location C (green).

To measure the change in diversity of species from location A (orange) compared to location C (green) was done with a Beta-diversity analysis. Every dot corresponds with a sample within the location and are clustered when the ASVs are similar (Figure F). Statistical analysis was performed with Permanova, the p-value is < 0,001.

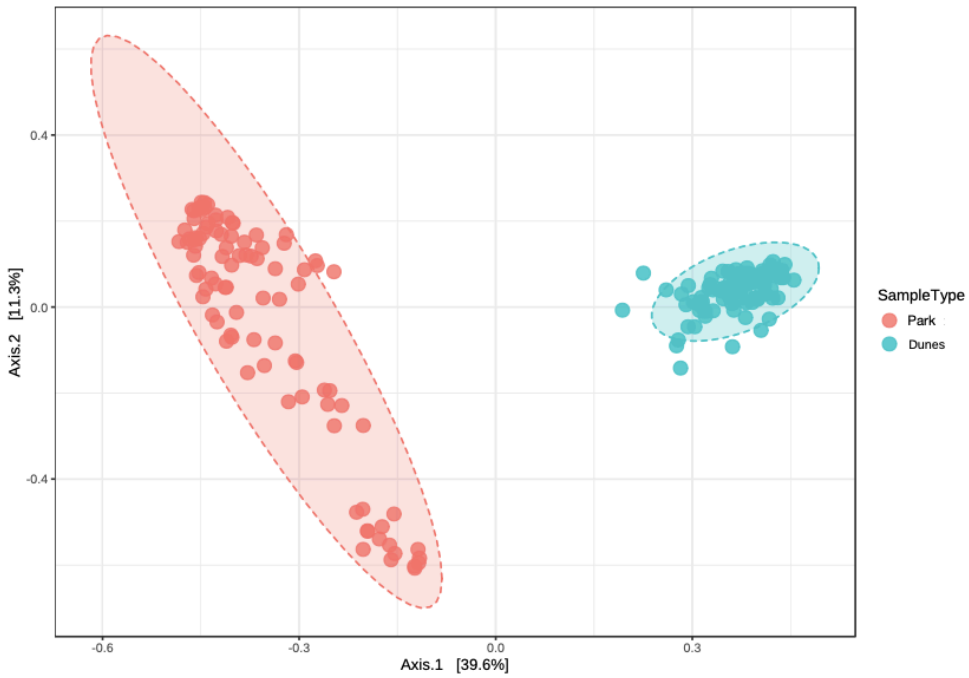


Figure F. Beta-diversity analysis of location A (orange) and location C (green). Each dot corresponds with a sample within the sample type. The similarities between sample and ASVs, within a location, are clustered. Statistical test for significant difference: Permanova, p-value < 0,001.

All the samples with similar sequences are clustered together, within their location. The sequences of location A and location C are significantly different.

Diversity analysis per sample location. To visualise the diversity and variation of phyla per sample in the number of reads, a stacked bar plot was done per location. The stacked bars show which phyla are detected in each sample for location A (figure G) and location C (Figure H) by actual abundance.

On the x-as the sample location (assigned following appendix A) was plotted for the specific location and on the y-as the actual abundance (in number of reads). In 83% of all the samples for location A, is Ascomycota the phyla with the most reads per sample. In 17% has Basidiomycota more reads per sample. For location C, 100% of the samples contain more Ascomycota.

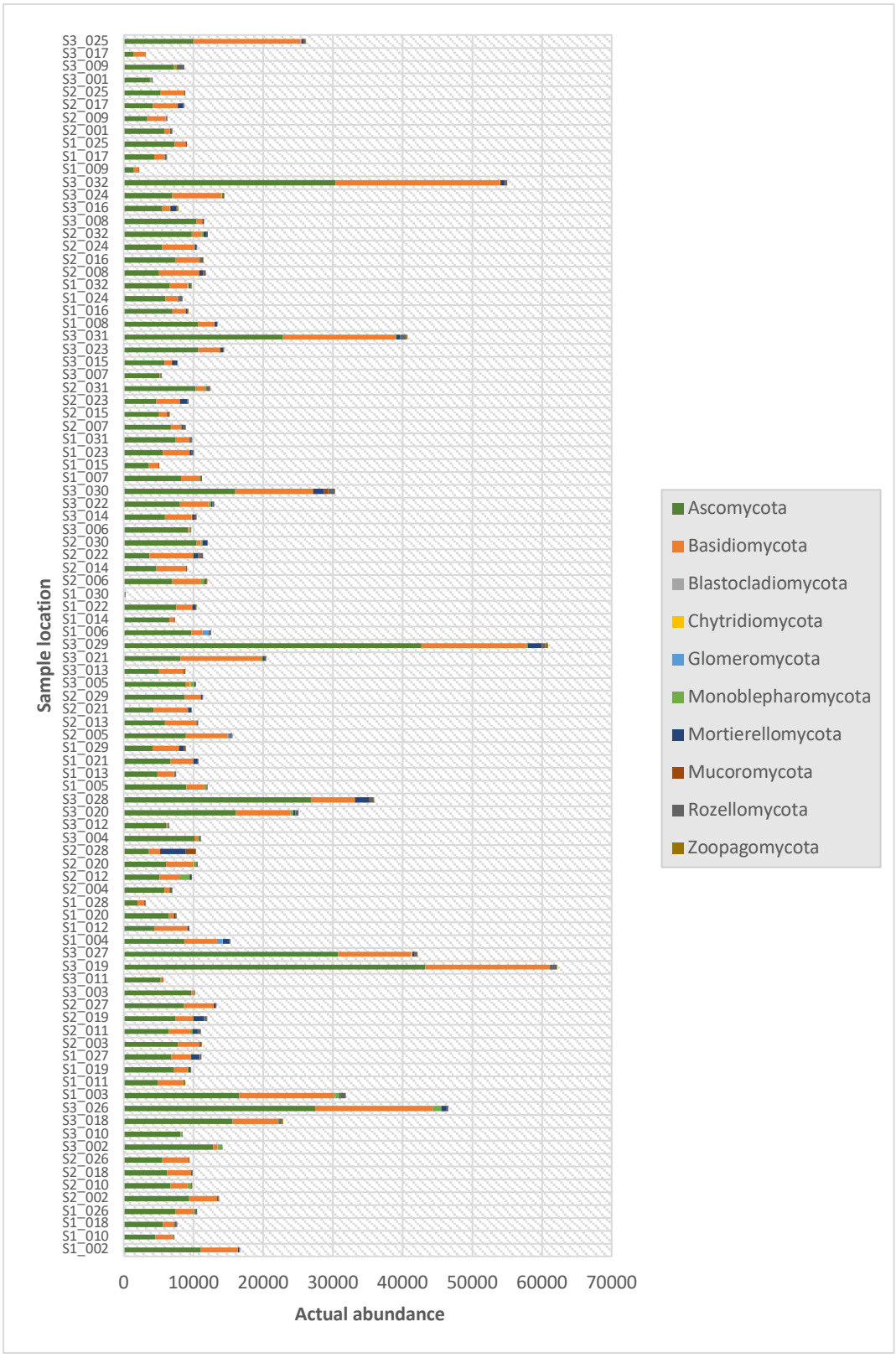


Figure G. The fungi variation per sample in actual abundance, in number of reads, for location A. Ascomycota (green) and Basidiomycota (orange) are present in all samples.

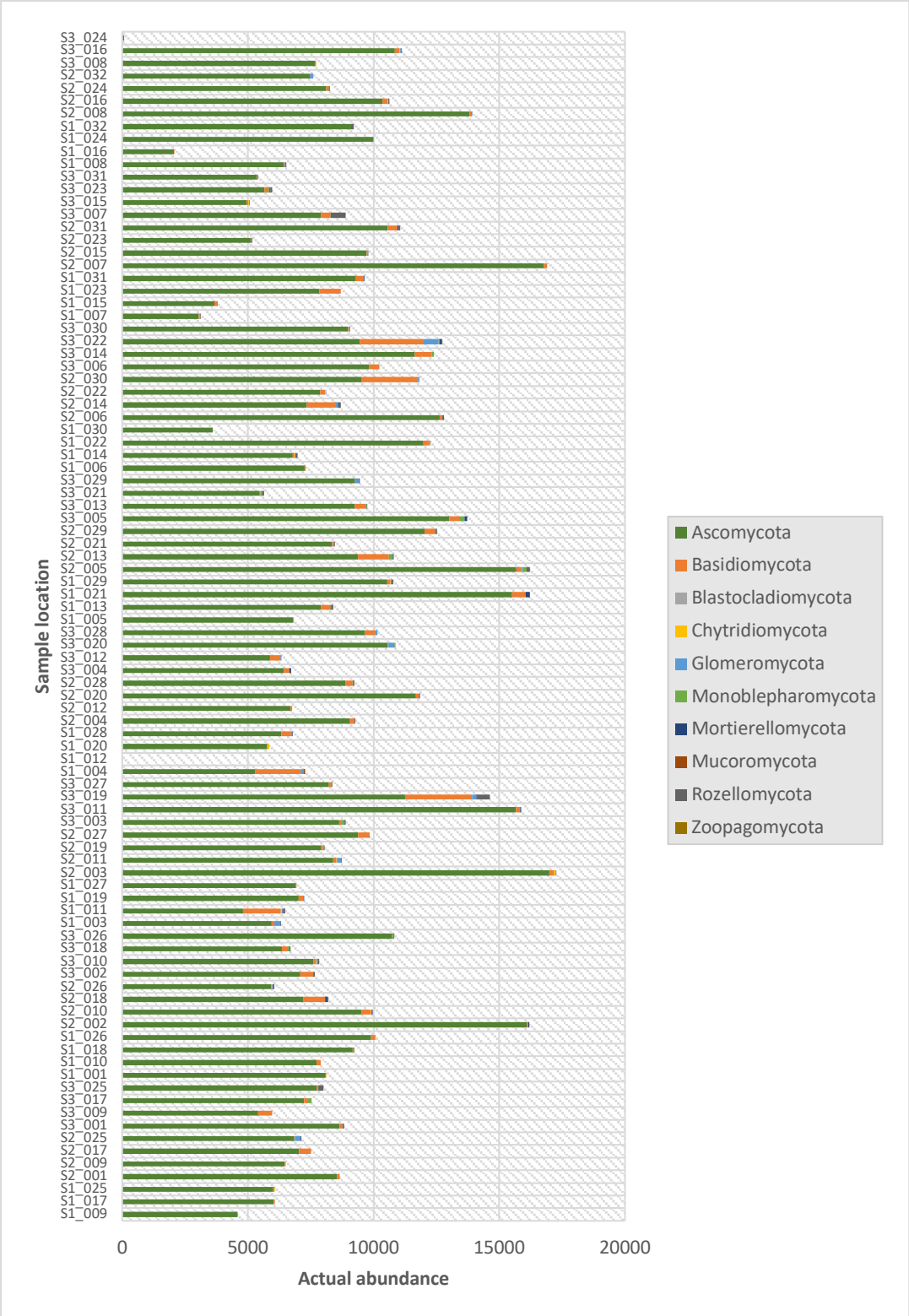


Figure H. The fungi variation per sample in actual abundance (in number of reads) for location C. Ascomycota (green) and Basidiomycota (orange) are present in all samples.

The variation and existence of fungi phyla in soil can differ if looked at the location, compared to samples individually.

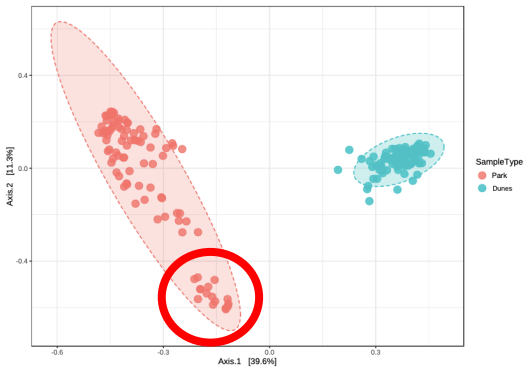
Discussion and conclusion

The identification of fungi community composition can possibly restore the ecological quality of sites. Therefore, research into the fungal communities needs to valid and without biases, because diversity can depend on the sequencing depth, the sample size, and the DNA extraction method [1, 33]. A reason why research with metabarcoding into samples collected from the environment has the potential to identify unidentified species [34]. In this study, we focused on getting insight into the biodiversity of a park (location A), a nature grassland (location B), and the dunes (location C), with a metabarcoding method. The samples were obtained with a sampling design, suggested by *Arita and Rodriguez (2002)*. Targeting the ITS2 region, specific for fungi, we used a cocktail of forward primers, following *Tedersoo (2014)*, to amplify the extracted DNA from soil. We used Illumina MiSeq sequencing, the DADA2 ITS2 pipeline, and the UNITE database to sequence, edit and determine the fungi phyla.

Diversity. The metabarcoding workflow generated 2.112.745 reads that could be determined as fungi, all the reads assigned to fungi are merged into a total of 5.040 ASVs. Primarily Ascomycota and Basidiomycota were detected in the soil in the park and the dunes, making up over 94% of all reads to be determined as the two phyla. Similar findings were presented by the study researching soil biodiversity by *Buée et al. in 2009* [35], the study found Ascomycota and Basidiomycota to be the most detected phyla out of soil samples with a metabarcoding method. Ascomycota is in the park and the dunes the most determined phyla. In the park, over 65% of the reads are Ascomycota. However, looking closer into composition of samples individually, Ascomycota was not the most detected phylum in all samples.

Diversity analysis. The diversity can also be defined by the mean diversity of phyla in different sites, figure E. A wider spread of the dots in location A, compared to location C, can be concluded that the park has more diversity and variation in the soil. The difference between locations is significant, following the statistical analysis with a T-test, p-value = 4,805e-07.
Assessing the differences between communities, locations (and samples) is done with the beta-diversity analysis, figure F. Samples with similar sequences are clustered together [36]. Thus, the sequences of the park have many similarities and is significantly different compared to samples taken in the dunes. The cluster for the park is bigger, which also suggests more diversity in the soil [36]. Interestingly, when determining the similarities of sequences within the park, a second clustering at the bottom formed (figure F). When determining the original location of the samples (16 samples), all samples could be traced back to a field situated in the park (table C). Suggesting the correct usage of the workflow and data-analyses if these sequences are clustered together, based on similarity.

Table C. Column C1 shows the beta-diversity analysis graph with a second clustering within location A (circled red), in column C.2 the photographic evidence is shown of a field in the park surrounded with large trees.

Column C.1	Column C.2
	

Inhibitors in soil samples. A third vegetation was used in this research: location B, a nature grassland, following the exact workflow (extraction method with KingFisher Flex, and amplification targeting the ITS2-region) as location A and location C. However, when checking the presence of DNA with gel-electrophoresis no bands, DNA, were visible (Figure I). The template DNA for amplification were undiluted extract of soil samples and a 10 times dilution.

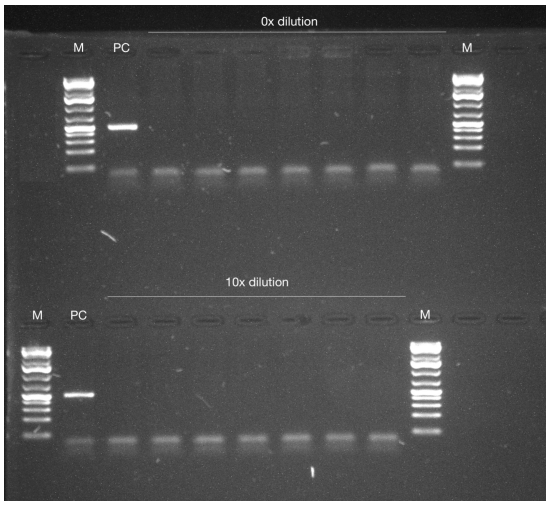


Figure I. The gel-electrophoresis gel of location B. Undiluted samples (top row) and a 10 times dilution (bottom row) of DNA extraction were tested.

The absence of DNA in samples of location B after gel-electrophoresis could have multiple explanations. The first explanation tested was the failure of DNA extraction, however this seemed improbable because DNA from soil samples of location A and Location C was extracted with the same method. Therefore, could reason two, PCR inhibitors in the soil, to be more likely. Soil contains enzymes that could manipulate DNA, like humic substances (humic acids) [37]. These substances are easily co-purified with DNA and are difficult to remove without additional treatments to obtain DNA useful for PCR [38]. A reason why humic

substances are hard to remove from soil, separate from DNA, is the similarities in molecular structure. Both DNA, and humic substances are long-chain-molecules and carry a negative charge [39] and the chemical-characteristics are similar to the phosphate groups of the backbone of DNA [40].

Suggestions for improvement.
DNA extraction. After evaluation of the agarose Gel-electrophoresis of samples from the nature reserve (location B), the protocol was altered. The DNA-extraction was done with silica-coated magnetic beads from *Oberacker et al. (2019)* and buffers from *Sellers et al. (2018)* to isolate DNA [41, 42]. However, the combination of protocols needs more improvement before it could be used for official analysis of soil samples. In research of Davis et al. (2009), the importance of the correct DNA-extraction method is emphasised. A major obstacle for accuracy of microbiome amplicon-based sequencing is efficient DNA extraction from samples that differ in biomass content [43].

Species accumulation curve. In this study, multiple analyses were used to determine the diversity and variation of locations or individual samples. However, to determine the species richness of a location an accumulation curve, or rarefaction curve, could be done on the current data [44]. The curve allows researchers to assess and compare diversity between populations or to evaluate the benefits of additional or fewer sampling to get an overview of diversity in the soil [44]. If the sequence depth is not deep to reach a plateau, we could consider to re-sequence these samples to increase sequence depth. It helps in deciding if the dataset should be rarefied or excluding samples from downstream analysis [31].

Expending the database. Fungal metabarcoding struggles with identification of operational taxonomic units (OTUs) or amplicon sequence variants (ASVs) to any taxonomic lineage beyond the kingdom or phylum level [45]. The determination of species is as good as the database used as reference. In this study, over 80.000 reads and 1.802 ASVs could not be determined. Expanding the reference databases for fungal communities is emphasised in *Buée, et al. (2009)*. The urgent need for more sequence databases is a relevant conclusion, the analysed sequences corresponded only 73% with only 26 identified taxa [35]. This is an extension of the theory that fungi species are largely uncharted [9].

Follow-up research.
Soil microbiome has an influence on the health of plants [1]. The properties of soil, like pH, texture, and organic carbon concentration, vary when soil is affected by factors. The main factors for variation of soil formation are, namely climate, organisms, parent material, and time [1]. Soil conditions are highly variable, but it has been indicated that climate is an important factor for the global distribution of soil fungi [1, 46]. For a follow-up study, the difference in soil composition or fungal diversity when weather conditions change over time could be researched. The study could see if soil composition and fungi diversity changes when seasons, or weather changes.

References

Photo on title page: ttps://www.nrc.nl/nieuws/2021/08/08/het-wood-wide-web-als-we-de-bomen-verliezen-verliezen-we-onszelf-a4054088

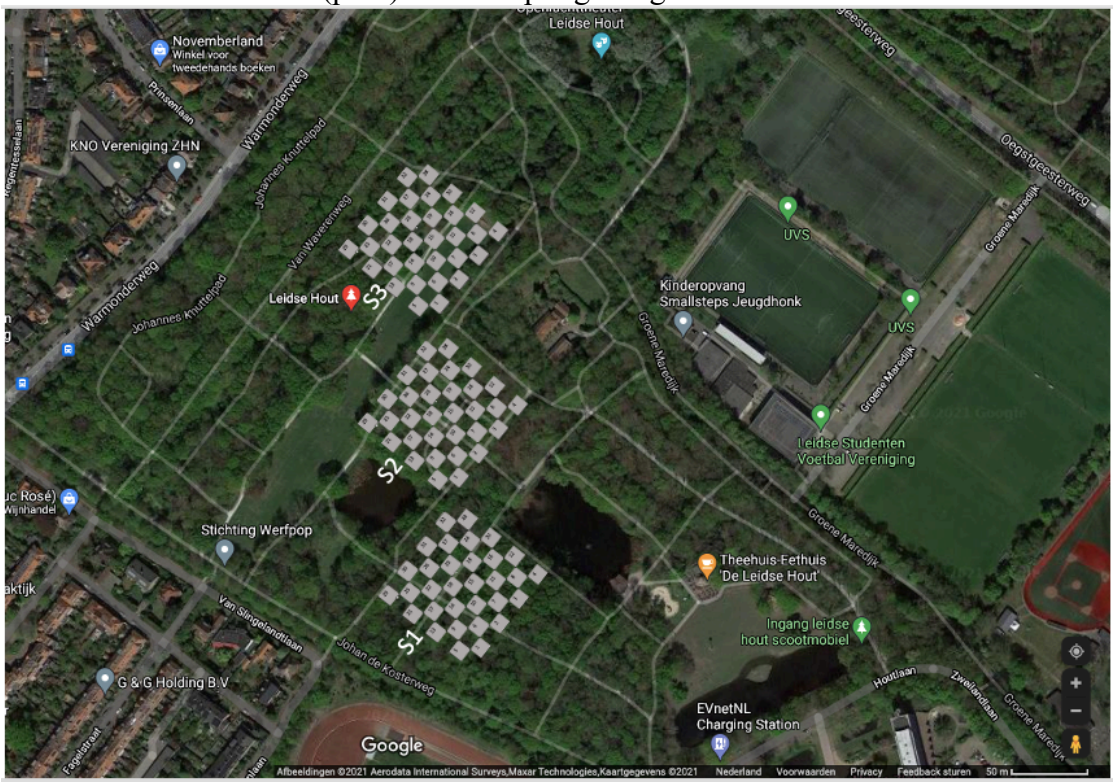
1. Fierer, N, "Embracing the unknown: disentangling the complexities of the soil microbiome", Nat Rev Microbiol 15, 579–590 (2017). https://doi.org/10.1038/nrmicro.2017.87
2. Sheldrake, M. "Entangled Life: How fungi make our worlds, change our minds, and shape our futures", 6th print, Atlas Contact, ISBN: 978-90-450-3614-4
3. Alberts, 2017, "Essential Cell Biology", 4th edition, GarlandScience, ISBN: 978-0-8153-4455-1
4. Nilsson, et al., "The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications", *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D259–D264
5. Peay et al., "Fungal Community Ecology: A hybrid Beast with molecular master", BioScience, 2008, Vol 58:9.
6. Weber, A. et al., "Thuja plicata exclusion in ectomycorrhiza-dominated forests: Testing the role of inoculum potential of arbuscular mycorrhiza fungi", Oecologia, 2005, 143: 148-156
7. Pauvert et al., "Bioinformatics matters: the accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline", Fungal Ecology, 2019, 41:23-33
8. Taylor AFS, Alexander I. "The ectomycorrhizal symbioses: Life in the real world", Mycologist, 2005, 19: 102-112.
9. Hawksworth DL. Et al. "Fungal Diversity Revisited: 2.2 to 3.8 Million Species" Microbiol Spectr. 2017 Jul;5(4). doi: 10.1128/microbiolspec.FUNK-0052-2016.
10. Arita, HT., Rodriguez, P., "Geographic range, turnover rate and the scaling of species diversity", Ecography, 2002, 25:541-550
11. https://www.thermofisher.com/nl/en/home/life-science/dna-rna-purification-analysis/automated-purification-extraction/kingfisher-systems/features.html, used December 2021.
12. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0019870_KingFisherFlex_UG.pdf, used December 2021
13. Tedersoo, L., et al. Fungal biogeography. Global diversity and geography of soil fungi. Science. 2014 1256688. https://doi.org/10.1126/science.1256688
14. Ihrmark, K. et al., "New primers to amplify the fungal ITS2 region – evaluation by 454-seqeuncing of artificial and natural communities", FEMS Microbiol. Ecol., 2012, 82:666-667
15. Bellemain, E. et al., "ITS as an environmental DNA barcode for fungi: an in-silico approach reveals potential PCR biases", BMC Microbiol, 2010, 10: 189.
16. Chen S, Yao H, et al., "Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species", PLoS ONE, 2010, 5(1): e8613. https://doi.org/10.1371/journal.pone.0008613
17. Turenne et al., "Rapid identification of fungi by using the ITS2 genetic region and an automated fluorescent capillary electrophoresis system", J Clin Microbiol. 1999 Jun;37(6):1846-51. doi: 10.1128/JCM.37.6.1846-1851.1999.
18. http://en.biomarker.com.cn/platforms/illumina, used January 2022
19. DADA2 https://benjjneb.github.io/dada2/ITS_workflow.html, used January 2022
20. Advantage of using DADA2. https://benjjneb.github.io/dada2/index.html, used January 2022
21. Gavito et al., "Local-scale spatial diversity patterns of ectomycorrhizal fungal communities in a subtropical pine-oak forest", Fungal Ecology, 2019, 42
22. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FD21035~.pdf, used December 2021
23. https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/magattract-powersoil-dna-isolation-kit/, used August 2021
24. http://www.caliperls.com/assets/028/8998.pdf, used August 2021
25. https://www.thermofisher.com/order/catalog/product/G720802#/G720802, used August 2021
26. https://www.agilent.com/cs/library/usermanuals/public/Fragment_Analzyer_system_manual_D0002110.pdf?elqTrackId=6524ffdebb444d93bc8b3aafae0b8ce4&elqaid=3243&elqat=2, used November 2021

27. https://explore.agilent.com/Software-Download-TapeStation-Systems, used November 2021
28. https://www.baseclear.com/genomics/next-generation-sequencing/illumina-sequencing/, used December 2021
29. https://www.rstudio.com/about/, used January 2022
30. https://unite.ut.ee/, used January 2022
31. https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/resources/tutorials/MDP_update.pdf, used December 2021
32. https://www.cd-genomics.com/microbioseq/the-use-and-types-of-alpha-diversity-metrics-in-microbial-ngs.html, used January 2022.
33. inceogˆlu, Ö. et al., "Effect of DNA extraction method on the apparent microbial diversity of soil", Appl. Environ. Microbiol., 2010, 76:3378-3382
34. Ruppert et al., "Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systemic review in methods, monitoring, and applications of global eDNA.
35. Buée et al., "454-pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity", New Phytol., 2009, 184:449-456
36. Walters, KE. et al., "Alpha-, beta-, and gamma-diversity of bacteria varies across habitats", PLoS One. 2020;15(9):e0233872.
37. Braid, MD. et al., "Removal of PCR inhibitors from soil DNA by chemical flocculation", Journal of Microbiological Methods, 2003, 52:389-393
38. Romanowski, G. et al., "Persistence of free plasmid DNA in soil monitored by various methods, including a transformation assay", Appl. Environ. Microbiol., 1992, 58:3012-3019
39. Cheng, WP. et al., "Effect of phosphate on removal of humic substances by aluminum sulfate coagulant", J. Colloid and Interface Sci., 2004, 272:153-157
40. Dong, D. et al., "Removal of humic substances from soil DNA using aluminium sulfate", Journal of Microbiological Methods, 2006, 66:217-222
41. Oberacker et al.(2019), Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid manipulation. PLOS Biology,17(1), https://doi.org/10.1371/journal.pbio.3000107
42. Sellers et al. "Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types", Metabarcoding and Metagenomics 2:e24556. (2018) https://doi.org/10.3897/mbmg.2.24556
43. Davis et al., "Improved yield and accuracy for DNA extraction in microbiome studies with variation in microbial biomass", www.BioTechniques.com, 2019, Vol. 66:6.
44. Deng C, et al., "Applications of species accumulation curves in large-scale biological data analysis", *Quant Biol*. 2015;3(3):135-144. doi:10.1007/s40484-015-0049-7
45. Nilsson, RH. et al., "The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications", *Nucleic Acids Research*, 2019, 47:D1:D259–D264, https://doi.org/10.1093/nar/gky1022
46. Větrovský, T., Morais, D., Kohout, P. *et al.* GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Sci Data* **7,** 228 (2020). https://doi.org/10.1038/s41597-020-0567-7

Appendix

Appendix A. The specific locations and the sample location.

Location 1: Leidse Hout (park) with sampling design to scale.



Location 2, Lentevreugd (nature reserve) with sampling design to scale

Location 3, Berkheide (dunes) with sampling design to scale