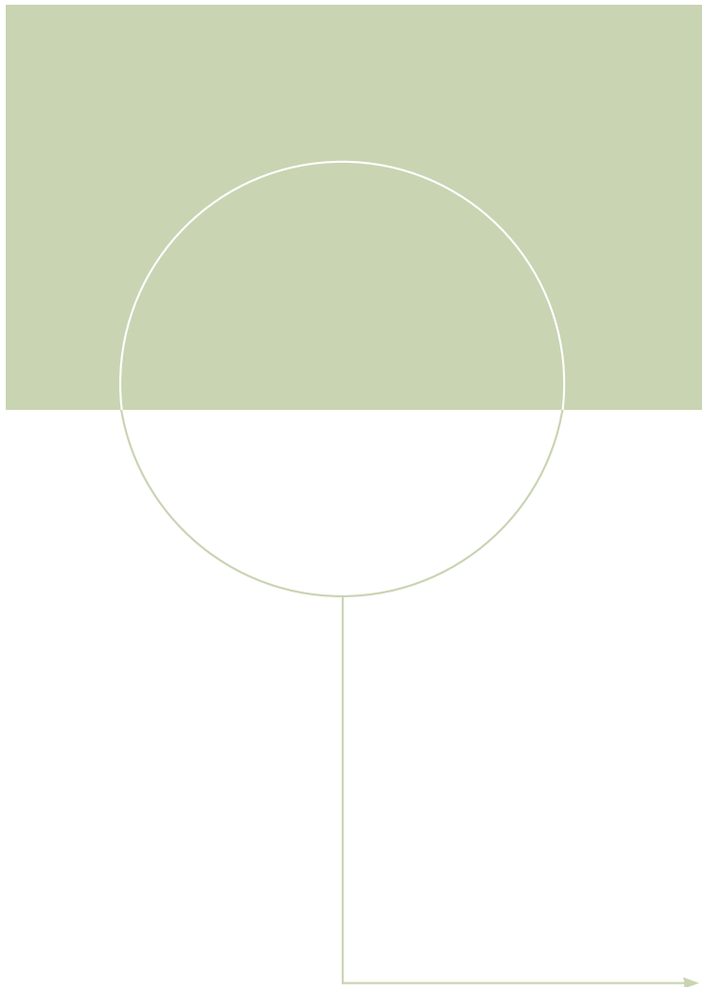


Master's thesis

2019

Lasse Topstad

**NTNU**
Norwegian University of
Science and Technology
NTNU University Museum
Department of Natural History

Lasse Topstad

# A multi-marker metabarcoding approach to study tardigrade diversity in forest ecosystems.

August 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

# A multi-marker metabarcoding approach to study tardigrade diversity in forest ecosystems.

Lasse Topstad

Biology (Master's Programme) Ecology, Behaviour, Evolution and Biosystematics
Submission date: August 2019
Supervisor: Torbjørn Ekrem

Norwegian University of Science and Technology
Department of Natural History

## Abstract

As other small members of the meiofauna, tardigrades are often neglected in ecological and environmental surveys. Occurring in all parts of the world, from deep marine sediments to alpine environments, tardigrades can play important roles in most ecosystems and should be incorporated in biomonitoring programs. Sampling of minute animals is, however, both tedious and time-consuming, impeding their inclusion in large scale ecological surveys. This study provides a step in bridging this gap by exploring the use of a multi-marker metabarcoding approach on environmental DNA samples. Samples of moss, lichens and litter were investigated by traditional morphology-based methods and metabarcoding and compared in terms of tardigrade diversity and community composition of the sampled microhabitats. By using locally constructed COI reference libraries, complemented by BOLD and GenBank sequences, metabarcoding in most samples detected more species of tardigrades than traditional methods. Additionally, metabarcoding detected the same community differences and microhabitat distribution patterns as traditional methods. In general, metabarcoding of litter eDNA samples was unreliable, with only one out of three markers consistently amplifying and detecting tardigrades. For its future use, the current lack of tardigrade reference sequences limits the taxonomic resolution of metabarcoding surveys. This impediment is easily overcome by adding barcodes of more species to the reference library, but can in its current state be partly circumvented by using multiple markers.

## Sammendrag

Som annen mikroskopisk fauna blir bjørnedyr ofte oversett i økologiske undersøkelser. Med sin kosmopolitiske utbredelse i alt fra sedimenter på havets dyp til høyalpine fjellområder, spiller bjørnedyr viktige roller i mange økosystemer og bør integreres i forvaltnings- og overvåkningsprogrammer. Men, identifisering av små dyr i en taksonomisk utfordrende gruppe er en svært tidkrevende og omstendelig prosess, noe som gjør at de som regel blir utelatt i større biologiske overvåkningsprosjekter. I min masteroppgave utforsker jeg om DNA-basert identifisering av bjørnedyr fra miljøprøver gir et godt bilde på deres artsmangfold, og om slik metodikk kan erstatte tradisjonell identifisering med morfologi og gjøre det enklere å inkludere bjørnedyr i overvåkingsprosjekter. Bjørnedyr i prøver av mose, lav og strø ble identifisert med metabarcoding og morfologi, og resultatene sammenliknet med hensyn på diversitet og samfunnsøkologi. Ved å bruke et lokalt COI-referansebibliotek, supplert med sekvenser fra BOLD og GenBank, identifiserte metabarcoding flere arter av bjørnedyr enn tradisjonelle metoder i de aller fleste av prøvene. Metabarcoding gav også de samme samfunnsøkologiske mønstrene i mose, lav og strø som identifisering med morfologi. Generelt gav ikke metabarcoding pålitelige resultater for strøprøver, hvor bare én av de tre markørene gav konsekvent identifisering. Den nåværende mangelen på referansesekvenser til mange bjørnedyrarter begrenser den taksonomiske oppløsningen ved metabarcoding. Begrensningen vil ikke elimineres før alle bjørnedyrarter får sin strekkode, men kan for øyeblikket delvis omkommes ved å bruke flere markører.

## Acknowledgements

I would like to express my deepest gratitude to my supervisor, Torbjørn Ekrem, for invaluable advice, fruitful discussions and constructive feedback, extending his support beyond what can be expected from a supervisor. Roberto Guidetti and the rest of the lab crew of the tardigrade lab at UNIMORE also deserves a huge thanks, as they welcomed me with open arms during my stay there and shared their expertise on tardigrades with me. The completion of my thesis would not have been possible without the helpful guidance of Markus Majaneva, providing me with insightful suggestions regarding the lab work and data processing. Aina Mærk Aspaas also have my gratitude for the effort she made, helping me barcode and photograph tardigrade specimen. Lastly, I would like to extend my sincere thanks to my friends and family, who have supported me and helped me with my manuscript. This dissertation was part of the NorBOL project "Tardigrades in Norwegian Forests", funded by the Norwegian Taxonomy Initiative.

# Contents

## Introduction

About 40% of the ice-free land surface on Earth is covered by forests (FAO, 2018). Represented by a variety of ecosystems, the biodiversity in such environments are generating indispensable ecosystem services of tremendous economic value (Balvanera et al., 2006; Hooper et al., 2005; Liang et al., 2016; Mori, Lertzman, & Gustafsson, 2017). The diverse composition of micro- and mesofauna inhabiting the soil of forests are the major biotic drivers underpinning the breakdown rate of litter and detritus, or conversely, the accumulation of soil organic matter (Swift, Heal, Anderson, & Anderson, 1979). These groups have in the last decades gained increasing attention, with many studies demonstrating their immense effect on resource cycling (Begon, Townsend, & Harper, 2005; Lillebø, Flindt, Pardal, & Marques, 1999; Swift et al., 1979; Wardle et al., 2004). Mapping such diversity is, however, often a more challenging task than for macrofauna and plants as it consists of many diverged, complex and species rich taxa. The sampling techniques often rely on hand-picking of individuals and morphology-based species delimitation, making it a time-consuming activity (Van Bezooijen, 2006). One such group, which still holds much unexplored diversity, is tardigrades (Tardigrada).

Tardigrades are hygrophilous, microscopic invertebrates inhabiting nearly all ecosystems on earth. The phylum currently holds about 1270 described species, mainly belonging to the two classes Heterotardigrada and Eutardigrada. These bilaterally symmetrical bodied micrometazoans resemble small bears in their appearance and behavior, often being recognized by the slow and inept motions of their four pairs of limbs when observed in a microscope. In the different trophic levels, various species of tardigrades are found as carnivores, herbivores, detritivores or bacterivores, demonstrating their broad range of interactions in the food web. They consume oligochaetes, nematodes, mites, collembola, plants, algae, bacteria, insects and other tardigrades (Hohberg & Traunspurger, 2005; Schmid-Araya, Hildrew, Robertson, Schmid, & Winterbottom, 2002), and have been observed to consume prey amounting up to 43% of their own biomass in less than 4 hours (Ramazzotti, 1962). As they often occur in high abundances (>1000 individuals per gram of microhabitat) , their influence in food web interactions can be far-reaching (Nelson, Guidetti, & Rebecchi, 2015). Whereas the distribution of many tardigrade species has received notable attention, little research has been done on the habitat specificity and preference of species, especially between different microhabitats. Terrestrial ecosystems often harbor populations of tardigrades in moss, lichen, soil, litter and stream microhabitats (Guidetti,

Bertolani, & Nelson, 1999; Ito, 1999; Jönsson, 2003), but to what extent these and other substrates need to be investigated to map an area's complete diversity is currently unknown. Collecting, filtering and processing samples for extraction of tardigrades is slow and exhausting as individuals and eggs must be hand-picked and mounted on microscopy slides for species identification. Furthermore, due to their simple body-plan, the taxonomy is mainly based on a limited set of morphological characters. This, and the fact that most species delimitation studies done before the 1980s are inaccurate in their species descriptions, has resulted in the presence of several species-complex groups and numerous cryptic species (Ramazzotti, 1962). The latter term is here defined as morphologically inseparable, but genetically different species (using a 97% similarity threshold). Integration of molecular tools has been proposed as an effective replacement for the exhaustive traditional methods (Bik et al., 2012; Deiner et al., 2017; Orgiazzi, Dunbar, Panagos, de Groot, & Lemanceau, 2015), and have in recent years been included in tardigrade studies (e.g. (Bertolani, Rebecchi, Giovannini, & Cesari, 2011; Cesari et al., 2019; Gąsiorek et al., 2016; Guidetti, Peluffo, Rocha, Cesari, & de Peluffo, 2013)). While also being more accurate for both species identification and detection of cryptic species (Blaxter, Elsworth, & Daub, 2004), the potential of these methods in biomonitoring and diversity surveys is immense (Orgiazzi et al., 2015; Stoeck et al., 2010; Taberlet, Roy, et al., 2012). With its emergence in 2003, DNA barcoding has grown to be an effective species identification tool (Hebert, Cywinska, Ball, & Dewaard, 2003; Orgiazzi et al., 2015). By targeting a DNA sequence consisting of a hypervariable region, flanked by two short, conserved regions, taxonomic information about the source species can be obtained. This is achieved by using pre-designed primers, which attach themselves to the short, conserved sites of the target DNA. PCR-based methods then read and extend from the primers along the anchored sequence and amplify the hypervariable region through repeated thermo-cycling steps. The result is exponentially increasing numbers of copies of the original target sequence. The new high-concentration DNA can then be sequenced to confidently reconstruct the original DNA barcode, allowing the comparison of the nucleotide assembly to a library of reference sequences with known identity. For single individuals, this method has been proven to be extremely valuable, serving as a key component in fields such as phylogenetics and phylogeography (Dayrat, 2005; Reitzel, Herrera, Layden, Martindale, & Shank, 2013; Schlick-Steiner et al., 2010). For large scale studies spanning whole taxa, this methodology, using single specimen DNA extracts, is not appropriate.

However, many of the techniques and concepts of DNA barcoding can be implemented in other molecular methods.

With the backbone of reference databases constructed by DNA barcoding, and the multifold increase in sequencing capabilities developed during the last decade, environmental DNA (eDNA) has become a valid option for biomonitoring (Orgiazzi et al., 2015; Taberlet, Bonin, Zinger, & Coissac, 2018). Environmental DNA is a somewhat diffuse term describing the mixture of DNA components found in environmental samples such as soil, sediments, water, feces or bulk samples (Ficetola, Miaud, Pompanon, & Taberlet, 2008; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). A broad range of processing options are available for such inventories, (e.g. filtering, sieving or grinding of material) dependent on the sampling technique used. With this methodology, it is possible to obtain comprehensive taxonomic libraries solely based on the extracted and processed eDNA inventories. However, eDNA consists of both extracellular and intracellular DNA, and any resulting inferences are dependent on the proficiency of the protocol in use (Majaneva, Diserud, Eagle, Hajibabaei, & Ekrem, 2018). These two categories of DNA have different origins and qualities and are therefore relevant to assess depending on what taxonomic groups are in study. As intracellular DNA originates from inhabitant organisms' tissue or living cells, it is often of superior quality compared to the normally degraded extracellular DNA derived from dead cells. When focusing on meso- and microfauna, the effect of filtering samples may not be as drastic as with larger organisms, as the organisms themselves can be caught or let through the filter based on the mesh size. In the case of tardigrades, a sieve of appropriate size allows the animals to go through the meshes together with most of the extracellular DNA, enabling an effective and simple protocol for collecting both live specimen and free-floating DNA components, while discarding larger particles. However, with the remarkable gripping ability of their hind legs, it is likely that a portion of the tardigrades cling on to the substrate during sieving. It is therefore expected that tardigrade DNA will be traceable in the sieve-caught inventory as well. As tardigrades constitute only a small fraction of the biomass, the vast majority of eDNA present in both the filtered and sieve-caught material is expected to stem from the numerous other organisms found in the environment. Furthermore, most of the eDNA is not suitable for species identification. DNA sequences are only taxonomically informative when they possess a hypervariable region flanked by two short, conserved regions (Hebert et al., 2003). A relatively novel method, termed metabarcoding, builds

on this principle by targeting either the total taxa or a subset of taxa in environmental samples by using universal primers, taxon specific primers or both. The output of such an approach is normally millions of sequences, and its conduct and processing require experience in ecology, molecular biology and bioinformatics, making it an interdisciplinary complex and challenging activity. As the output sequences are of unknown origin, they need to be blasted against a reference library of sequences with known taxonomic identity. In this context, micro- and mesofauna have received little attention, with relatively few barcode sequences being deposited in the three institutions of the International Nucleotide Sequence Database Collaboration (INSDC) (Bienert et al., 2012; Ramirez et al., 2014). With global collaborations affiliated to the IBOL initiative using the Barcode of Life Data Systems (BOLD) (Ratnasingham & Hebert, 2007), it is likely that the near future will see an increase in sequence deposition in public databases. This will greatly aid the metabarcoding approach and provide higher taxonomic resolution of mesofaunal diversity studies.

The aforementioned ability of metabarcoding to identify large scale inventories of organisms in an environment merely by the presence of their DNA has huge potential implications on biomonitoring (Orgiazzi et al., 2015). For many years, the scientific community believed that metabarcoding could possibly revolutionize and unify biomonitoring as one single toolbox. There have, however, been reported instances where the markers or protocols used cannot sufficiently identify the targeted taxa, even when known to be present (Bienert et al., 2012; Porazinska et al., 2010). In recent years, it has become clear that there currently exists no universal metabarcode that can track all organisms. It is therefore vital to thoroughly investigate the applicability of metabarcoding on all scales, both in regards of target taxon and experimental protocol of conduct. With this in mind, the goal of this study is to investigate the applicability of metabarcoding on the taxon Tardigrada by comparing the retrieved MOTUs to the findings of a traditional survey using morphology-based identification. In relation to this goal, we aim to assess the two methods' ability to capture tardigrade diversity and community compositional differences, both on a sample-to-sample basis and when comparing total inventories from different substrates.

# Methods and Material

## Field sampling

Field work was conducted 11th of August 2017 in Skråstadheia Nature reserve (58.19899°N, 7.99329°E, 28 m a.s.l.) in Southern Norway (Fig. 1). ~100cm$^2$ substrate samples were collected within a 50 m x 50 m plot by using a standardized container. Five samples of each moss, lichens and litter were collected in a stratified manner, yielding a total of 15 samples. The distance between samples was at least 3 meters away from each other to prevent gathering samples with overlapping populations and with DNA migration. Each sample was placed in a sterile plastic bag, marked with a unique tag and stored in a cooler for subsequent transport to the lab. All samples were dried in a fume hood and transferred back to their respective bags for storage.


**Fig 1.** Map of sampling locality.

## Traditional sampling

The preparation protocol for tardigrade extraction consisted of homogenizing the 100cm^2 substrate and subsampling it into four equal parts (25% of the total biomass). The first part was used for traditional sampling, second for construction of reference sequences using single individual DNA barcoding, third for metabarcoding and fourth for backup. For the traditional methods, the subsamples of moss and lichen were weighed and immersed in 500mL of $dH_2O$ for 30 minutes. Next, the sample was rigorously shaken for 1 min before being poured through a 500 µm sieve into a measuring cylinder. After 45 minutes of decantation, the top 400 mL was removed, whereas the bottom 100mL with precipitate was transferred to a glass container. For soil samples, the subsamples were washed through a sieve stack with the top sieve and bottom sieve being 500 µm and 45 µm, respectively. This was done to filter out debris larger than 500 µm and smaller than 45 µm, while also catching all tardigrades and eggs in-between the sieves. The captured inventories were transferred to glass containers in equal volume to the moss and lichen extracts. Using a glass pipette, the extract was transferred to petri dishes and sought through using a stereo microscope (Leica MZ6). The search was conducted by moving the petri dish horizontally from left to right, working from top to bottom. This procedure was repeated for the full volume of the extract, and, for each petri dish, all tardigrades and eggs were harvested and mounted in Hoyer's medium on slides for species identification. To investigate the proficiency of the initial searches, the complete extracts of the first three samples were run through the searching protocol an additional time. The proficiency of one search was considered to be sufficiently high, as the reruns yielded less than 0.1% new individuals compared to the first search. The remaining samples were therefore only sought through once. After animals and eggs were collected and mounted, all slides were investigated for species identification using phase contrast and differential interference contrast. Using 630x and oil immersed 1000x magnifications, all individuals and eggs were identified using key taxonomic literature (Bingemer & Hohberg, 2017; Fontoura & Pilato, 2007; Kaczmarek & Michalczyk, 2017; Morek, Gąsiorek, Stec, Blagden, & Michalczyk, 2016; Pilato & Binda, 2010; Ramazzotti, 1962; Thorp & Covich, 2009).

## DNA barcoding

To facilitate higher taxonomic resolution when annotating the metabarcode sequences, a local reference library was constructed for the barcode fragment of the mitochondrial cytochrome c

oxidase I gene (COI). This was done by investigating the second 25% subsample of the original moss, lichen and litter samples to extract tardigrades and eggs for barcoding. All located specimens were individually mounted on slides in a drop of water and identified at 630x magnification. Adding more water to the slide allowed for the individuals to be recollected. To account for non-successful barcoding attempts, and to increase the probability of detecting cryptic species, 1-15 individuals of each morphospecies were kept depending on their availability, while excess individuals were discarded. The individuals and eggs selected for barcoding were washed and transferred to wells on 96 well plates. DNA was extracted from single specimen using the QuickExtract$^{TM}$ DNA Extraction Solution kit by Lucigen using a modified version of the manufacturer's protocol (Appendix E). These modifications were made to make the protocol more appropriate for microscopic organisms. Avoiding the exuviae in the bottom, 15µm of each DNA extract was transferred to new wells on a sequencing plate and sent to the Genomic Facility at the University of Guelph. DNA barcodes were retrieved using universal COI primers and were added to the local reference library. Vouchers were made by recovering the exuviae and mounting them in Hoyer's medium on microscopy slides. These are deposited in the scientific collections of the Department of Natural History (INH) at the NTNU University Museum, Trondheim, Norway.

**DNA extraction**

From the original samples, a 25% subsample was collected and processed through the same processing protocol as used for traditional sampling. The litter samples were run through the same protocol as moss and lichen. For all samples, the substrate from the sieve was scraped into collection tubes and processed with the DNeasy PowerMax Soil Kit from Qiagen following manufacturer's protocol. From the measuring cylinder, the top 400mL and the bottom 100mL (with precipitate) were separated and filtered through mixed cellulose ester (CN) filters connected to an electrical vacuum pump. The filters were then processed with the DNeasy PowerSoil Kit from Qiagen following manufacturer's protocol. From each sample, there were therefore three extracts (Fig 2), yielding a total of 45 extracts for the 15 samples. With 3 PCR replicates per extract, there were 135 PCR reactions per marker, totaling to 405 reactions for the three markers used.

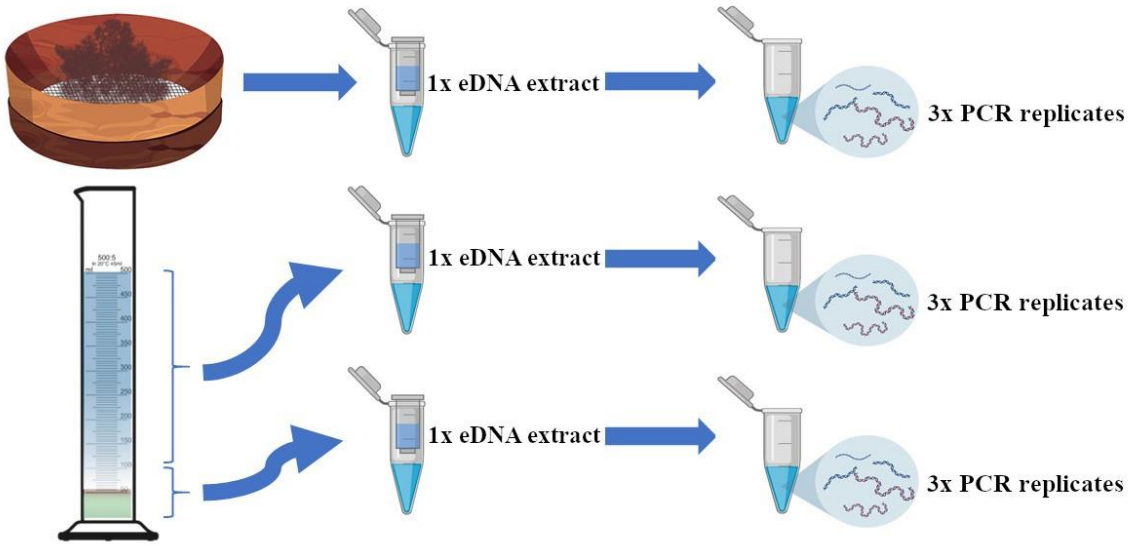



**Fig 2.** The DNA extraction process used on each sample. From each sample, three DNA extracts were collected. One from the sieved substrate, one from the top 400mL part of the measure cylinder and one from the bottom 100mL with precipitate. All markers were run on each eDNA extract.

## Library preparation

The initial PCR amplification was carried out using 2.5 µL 10× reaction buffer (200 mM Tris HCl, 500 mM KCl, pH 8.4), 0.2 µL Invitrogen's Platinum Taq polymerase (5 U/µL), 1 µL dNTP mix (10mM), 1 µL $MgCl_2$ (50mM) and 0.5 µL of each primer (10mM) with illumina adapters. 17.3 µL biology grade $H_2O$ and 2 µL template DNA was added to make the final volume 25 µL for each reaction. For amplification of the substrate samples, the primer and Taq volumes were increased to 1 µL and 0.3 µL, respectively.

Two fragments of the mitochondrial COI gene (hereafter COI-1 and COI-2) and one fragment of the V9 region of the 18S rRNA marker were amplified using the primers described in table 1. All primer pairs were inspected *in silico* in the AliView software (Larsson, 2014) to assess their specificity. The primers were modified *in silico* to account for observed variability in the tardigrade target sites.

Table 1. Primers with illumina tails used for amplifying the different markers in PCR run 1. Primer sequence parts are marked in bold.

| DNA fragment | Primer name | Primer direction | Primer sequence (5'-3') | Primer source |
|---|---|---|---|---|
| COI | BF2_mod_IL | Forward | TCGTCGGCAGCGTCAGATGTGTATAAG-AGACAG**GCNCCNGAYATRKCNTTYCC** | (Elbrecht & Leese, 2016)* |
| | BR2_mod_IL | Reverse | GTCTCGTGGGCTCGGAGATGTGTATAA-GAGACAG**TCDGGRTGNCCRAARAAYCA** | (Elbrecht & Leese, 2016)* |
| COI | BF2_mod_IL | Forward | TCGTCGGCAGCGTCAGATGTGTATAAG-AGACAG**GCNCCNGAYATRKCNTTYCC** | (Elbrecht & Leese, 2016)* |
| | TarR_IL | Reverse | GTCTCGTGGGCTCGGAGATGTGTATAA-GAGACAG**GGWARAATHARAATATADAC** | (Guil & Giribet, 2009)* |
| 18S rRNA | 18S_TAReuk454FW_IL | Forward | TCGTCGGCAGCGTCAGATGTGTATAAGAGA-CAG**CCAGCASCYGCGGTAATTCC** | (Stoeck et al., 2010)* |
| | 18S_TAReukREV3r_IL | Reverse | GTCTCGTGGGCTCGGAGATGTGTATAAGAG-ACAG**ACTTTCGTTCTTGATYRA** | (Stoeck et al., 2010)* |

*primer was modified to increase its specificity to tardigrades.

The PCR programs used for the different primers are found in table 2 and were chosen based on amplification achieved during pilot runs.

Table 2. PCR programs used for amplification of universal and tardigrade specific primers for COI and 18S.

| Step | COI universal | | | COI specific | | | 18S Universal | | |
|---|---|---|---|---|---|---|---|---|---|
| | Temperature [*C] | Time [min:sec] | Cycles | Temperature [*C] | Time [min:sec] | Cycles | Temperature [*C] | Time [min:sec] | Cycles |
| Initial denaturation | 94 | 03:00 | 1 | 94 | 03:00 | 1 | 94 | 03:00 | 1 |
| Denaturation | 94 | 00:40 | 35 | 94 | 00:40 | 35 | 94 | 00:40 | 25 |
| Annealing | 52 | 00:30 | 35 | 48 | 00:30 | 35 | 52 | 01:00 | 25 |
| Elongation | 72 | 00:40 | 35 | 72 | 00:40 | 35 | 72 | 00:40 | 25 |
| Final elongation | 72 | 02:00 | 1 | 72 | 02:00 | 1 | 72 | 02:00 | 1 |

The DNA extracts were amplified in three PCR replicas, where each PCR plate had 6 negative controls distributed as described by figure 3.

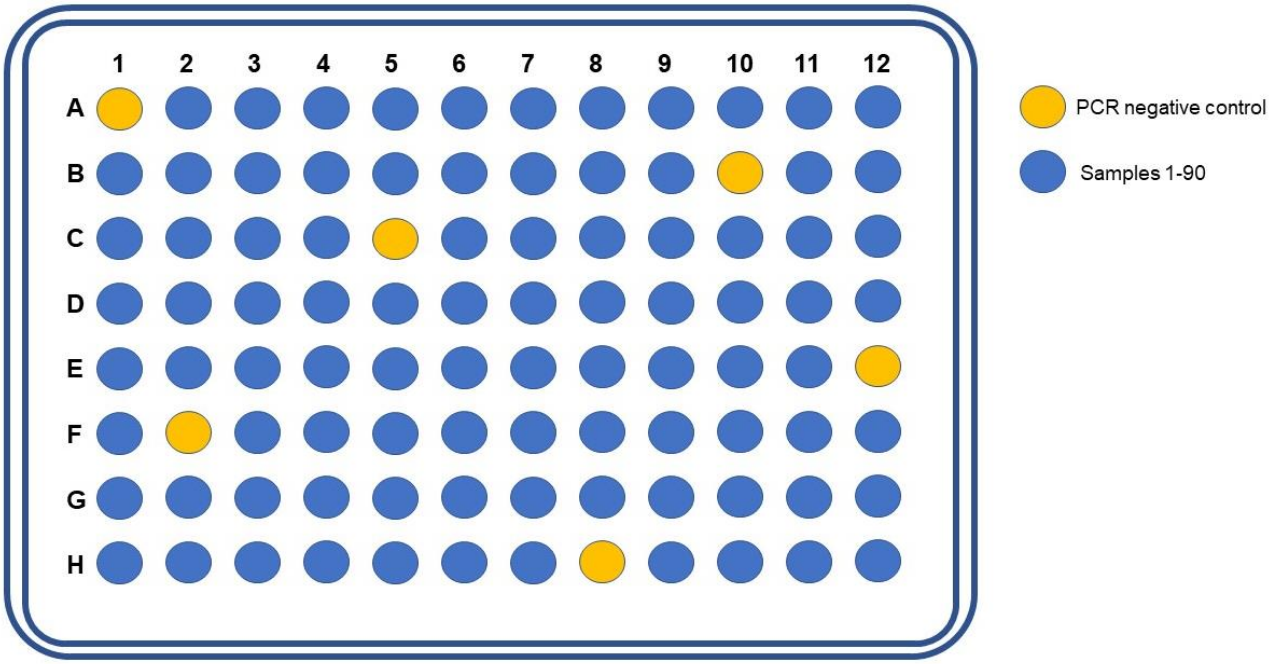



Figure 3. PCR plate layout used for all PCR runs.

The amplicons of the PCR reactions were assessed on a 1.2% agarose gel by electrophoresis, while a subset of samples was quantified using the dsDNA HS Assay kit in Qubit 2.0, following manufacturer's protocol. The DNA concentration of each marker was normalized by dilution and mixed with their respective sample on standard 96 well plates, so that each well contained equal concentrations of all markers from one PCR replicate per sample extract. Likewise, the negative controls from the different marker PCRs were pooled three by three. A second PCR was run with 10 cycles to dual index the illumina tailed amplicons, using Nextera XT indices (FC-131-1002, Illumina, San Diego, CA, USA). The indexed amplicons were then transferred to one of two pools. The first pool consisted of 66 sample extracts and 5 negative controls, and the second of 67 sample extracts and 4 negative controls – both pools being balanced in regards of sample-substrate composition. Using SPRI magnetic beads, the pools were cleaned following the protocol of (Fisher et al., 2011) using a bead:sample ratio of 0,55:1. The purified volume was

quality checked by dsDNA HS Assay kit in Qubit 2.0 and BioAnalyzer and was measured to a DNA concentration of 19,5 ng/µL and 16,4 ng/µL. Finally, each purified pool was used as template for constructing two libraries on standard flow cells using the 600 cycle V3 Illumina MiSeq sequencing kit (MS-102-3003).

**Data cleaning and filtering**

Sequences of the different markers were separated in mothur (Schloss et al., 2009) using an .oligos file including the primer sequences. The forward and reverse strands of the sequences where then merged in vsearch (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) using the following *–fastq_mergepairs* command. Subsequent filtering of low quality paired-end sequences was done using the commands *--fastq_filter, --fastq_maxee, --fastq_minlen, --fastq_maxlen* and *--fastq_maxns*. For all markers, max ambiguous nucleotides were set to 0, while max error rate used was estimated for each marker following the protocol of Rognes et al. (2016). Using the *summary.seqs* command in mothur, the lengths of the sequences of each marker, with quantiles, were calculated, This, in addition to inspecting the length of known tardigrade sequences, was used to determine the minimum and maximum sequence length parameters allowed during the filtering. Sequences fulfilling these criteria were relabeled by sample names based on their combination of tags and were kept as separate fasta files for each sample. These files were then merged in mothur before primer sequences were trimmed. Dereplication to unique sequences was done in vsearch with the commands *--derep_fulllength*, followed by de novo chimera detection using the command *--uchime_denovo,* with *--abskew* set to 5. Next, all non-chimera sequences were clustered using the swarm software (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2014), with parameters -d 1, -t 2, -z, -w and using the fastidious command. The original quality filtered sequences, excluding singletons, were then affiliated to the different swarms in preparation for the final blasting. These swarms were regarded as separate molecular operational taxonomic units (MOTUs). To assign taxonomic labels to these MOTUs, and to assess their validity as true MOTUs, reference sequences were used. For the 18S marker, all tardigrade sequences were downloaded from NCBI, while the COI reference sequences were downloaded from NCBI and BOLD. All marker reference databases were trimmed to reduce computational time, using the cutadapt unix command. This trimming was done by keeping only the sequences containing primer anchor-sites matching at least two thirds of the primer sequence. Furthermore, this allowed the retainment of the hypervariable regions

between the primers, while removing uninformative flanking regions. Taxonomic annotation of sequences of each marker was achieved by blasting MOTUs and their nested sequences to their respective reference databases using a 97% similarity threshold. This was done in vsearch with the commands *--usearch_global, --maxaccepts 0, --maxrejects 32,* and *--id 0.97*. The output of the blasts was converted to OTU tables, assigned with read numbers of DNA-species/MOTUs in each sample. Although negative controls contained raw read sequences, they were included in the bioinformatic pipeline and were found to contain no tardigrade sequences. To remove inconsistent PCR replicates, and to remove samples that yielded no tardigrade sequences, an R-script was run until no further replicates were removed (Appendix A). This was done by comparing the composition of read numbers and MOTUs between PCR replicates and discarding outlier replicates. The script yielding a retainment of 67% of the initial replicates. For each extract of the different samples, the mean value of its PCRs was calculated, resulting in 12, 12 and 11 quality extracts of moss, lichen and litter samples in the final OTU table. To define the presence of a MOTU in a sample, the lower threshold was set to 10 reads, meaning a MOTU had to be observed with at least 10 reads in a sample to be included in further analyses.

**Statistical analysis**

The Bray-Curtis dissimilarity index is often used to quantify how correlated sites are. The index is defined by

$$BC_{jk} = 1 - \frac{2\sum_{i=1}^{p}\min(N_{ij}, N_{ik})}{\sum_{i=1}^{p}(N_{ij} + N_{ik})}$$

where $\min(N_{ij}, N_{ik})$ is the observed minimum number of individuals of a common species between the two communities i and j, and $(N_{ij} + N_{ik})$ is the total number of the species in both communities. With an index of 0, the communities are identical, while an index of 1 means there are no common species. As it is a dissimilarity index, it does not qualify for statistics working with distance metrices. However, when expressed as

$$BC_{jk} = \frac{1}{2}\sum_{i=1}^{p}|N_{ij} - N_{ik}|$$

it becomes a measure of distance between communities, and thereby takes the structure of a matrix. This permits the use of multidimensional scaling methods (MDS), which preserves the relative distances between communities when going from n-dimensional to two-dimensional space. Such methods have their strength in how they can depict similarity or dissimilarity in species composition between two or more communities

**Principal coordinate analysis (PCoA)**

To ease the interpretation of the hyperdimensional dissimilarity matrices (Bray Curtis matrices), multidimensional scaling was used to reduce the number of dimensions to a more understandable configuration. This was done using the vegdist() and cmdscale() commands of the vegan package (Oksanen et al., 2007) in R (R Development Core Team, 2013). These commands approximated new n-dimensional coordinates for the original dissimilarity distances. The PCoA run on the matrices were done in two dimensions (k=2), yielding the largest eigenvalues (appendix D, Fig D2). To measure the fit of the approximated coordinates, $R^2$ was determined by calculating the squared correlation between the original distance matrix and the PCoA point coordinates.
The significance of an effect of substrate was evaluated by PermANOVA on the clusters of sample measurements of each substrate. A permutation test was run to assess the homogeneity in dispersion of measurements within each substrate category. These were run using the adonis and betadisper commands from the vegan package (Oksanen et al., 2007), with the dispersion subsequently being tested for significance using the permutest command with 1000 permutations.

## Results

Overall, traditional methods and the multi-marker metabarcoding approach detected similar species inventories (Fig. 4). To allow merging of records between different markers when more than one cryptic variant was detected, such records were elevated to genus level. Complete species lists for each method can be found in appendix A. As metabarcoding quantifies abundance by number of sequences, while traditional methods use number of individuals, the measurements were not scalable between the two methods. Nevertheless, clear trends were observed in the relative abundance of species between the methods. When converted to species proportions, the data of metabarcoding and traditional methods were highly correlated (0.80).


Figure 4. Heatmap of the different methods and their species detection. In cases where more than one variant of a species was recorded, identifications were elevated to genus level to allow comparison of results from the 18S and COI markers. Blank cells indicate no detection. Red asterisk marks species lacking reference sequence for that marker.

With traditional methods, 3788 tardigrades were recorded and identified to 40 morpho-species, spanning 2 classes, 5 orders, 7 families and 19 genera. The abundances and distribution of species were variable between samples and are reported for each substrate in table 3. Barcodes were successfully retrieved for 24 of the species, yielding 32 BINS.

Table 3. Average number of species and abundances in moss, lichen and litter samples, reported with the variability between samples of each substrate type. All samples contained tardigrades, ranging from 35 to 480 individuals.

| | Moss | Lichen | Litter |
|---|---|---|---|
| Mean number of species | 12.3, SD=3.5 | 8.0, SD=3.2 | 10.8, SD = 2.0 |
| Mean number of individuals | 272, SD=135.3 | 176, SD=131.3 | 61.5, SD=19.3 |

Coalescing the samples of each substrate type resulted in large areas of overlap in species presence between the different substrates (Fig. 5). Only 3 species were found exclusively in litter, while the numbers of moss and lichen were 6 and 9, respectively.

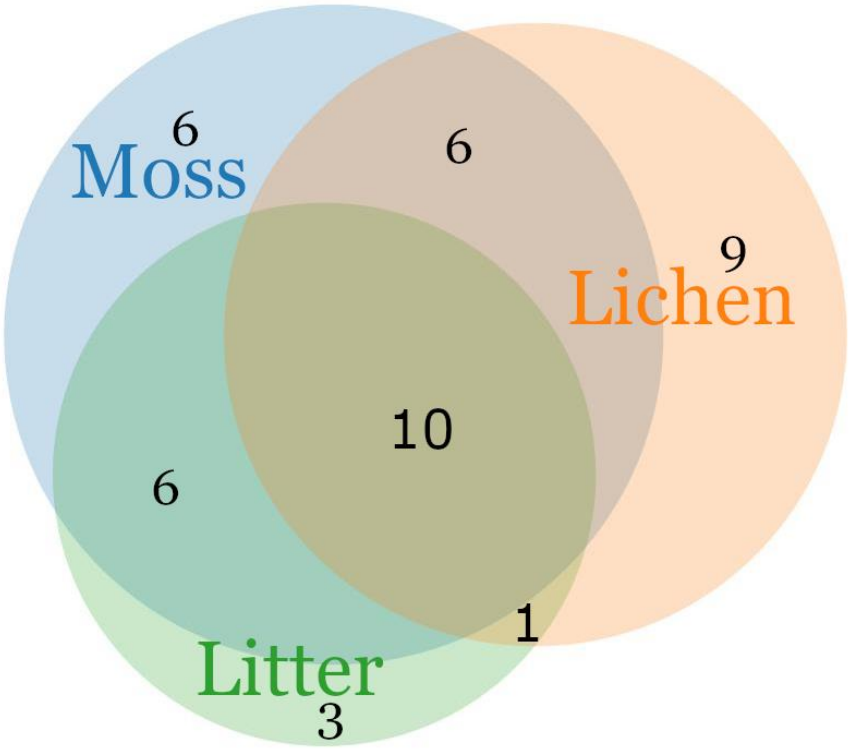



Figure 5. Venn diagram of the number of tardigrade species recorded exclusively in and shared between the sampled substrates using traditional methods.

For the metabarcoding data, the sequencing resulted in 16440661 raw reads. After quality filtering, 10792600 sequences remained. The performance of the different markers in detecting tardigrade species were 31, 32 and 40 MOTU-matches above 97% similarity by the COI-1, COI-2 and 18S markers, respectively. Additionally, 28, 30 and 25 more MOTUs were assigned to Tardigrada for each marker, resulting in very similar total diversity estimates for all three markers. All three markers successfully detected a wide range of tardigrade species, spanning most major terrestrial families and genera (appendix A and F). Of the 24 locally COI barcoded species (32 BINs), all DNA-species and BINs except *Pseudechiniscus suillus* and one cryptic variant of the *Mesobiotus harmsworthi* group were retrieved by both COI markers. These two species were found by traditional methods in quite high abundances in several of the investigated samples, yet remained undetected by both COI markers. The 18S marker did, however, detect *Pseudechiniscus sp*. and *P. facettalis* - the latter belonging to the same species complex group as *P. suillus*. 18S also identified two *Mesobiotus harmsworthi* group species. For the species

complex *Macrobiotus hufelandi* group, identified as two morphotypes during traditional sampling, DNA barcoding of individuals of these two types revealed 6 distinct BINs. All of these were recovered by both COI markers.

When inspected at sample level performance, metabarcoding showed high variability in its ability to detect species, as shown in table 4. The three extracts of each sample yielded similar species composition and were averaged for each marker. In general, 18S retrieved the most species, compared to COI-1, COI-2, both COI combined and traditional sampling. In sample 213 and 237, more species were observed by traditional methods than by 18S. For the mitochondrial primers, COI-2, being the more tardigrade specific, performed better than the more universal COI-1 by detecting more DNA-species per sample. Both COI primers were unsuccessful in retrieving consistent PCR replicas for all extracts of the litter samples 233, 234, 236 and 237. Interestingly, the 18S marker was consistent for all these samples except 236 (which was measured to have a DNA concentration of ~0 for at least two of the extracts).

Table 4. Number of DNA-species and morpho-species detected in the investigated samples. Numbers given by metabarcoding markers for each sample are average number of MOTUs per PCR matching reference sequences above the 97% similarity threshold. Inconsistent PCR replicates were excluded from the calculations. Morpho-species sampled by traditional methods are species recorded by at least 1 individual or egg in the sample. The – symbol indicates samples that were excluded from analysis due to inconsistent PCR replicates.

| Sample ID | COI-1 | COI-2 | COI total | 18S | Trad |
|---|---|---|---|---|---|
| 212 | 11 | 13 | 15 | 15 | 10 |
| 213 | 7 | 10 | 17 | 17 | 19 |
| 215 | 8 | 11 | 13 | 24 | 13 |
| 216 | 5 | 6 | 7 | 14 | 14 |
| 219 | 10 | 13 | 13 | 15 | 8 |
| 221 | - | 7 | 7 | 13 | 3 |
| 222 | 8 | 9 | 17 | 15 | 9 |
| 224 | 12 | 12 | 13 | 19 | 13 |
| 226 | 10 | 15 | 18 | 18 | 8 |
| 230 | 5 | 7 | 7 | 10 | 7 |
| 233 | - | - | - | 13 | 8 |
| 234 | - | - | - | 10 | 14 |
| 236 | - | - | - | - | 11 |
| 237 | 10 | 11 | 16 | 6 | 9 |
| 240 | - | 8 | 9 | 21 | - |

Concatenating the species lists of the three markers, while merging mutual species found by more than one marker, yielded a total of 49 DNA-species. These included 85.7% of the morpho-species detected by morphology (Fig. 6), while detecting 7 new species not recorded by traditional methods, in addition to cryptic variants. Of these, three have previously been recorded from Norway, while four are new records. One was a match to *Acutuncus antarcticus*, a species endemic to Antarctica. Of the species exclusively found by traditional sampling, 5 out of 7 species lacked reference sequences. Furthermore, of the two species represented in the reference database, one was found by traditional sampling as single individuals in two samples, while the other was a cf. level identification (*Echiniscus* cf. *testudo*).

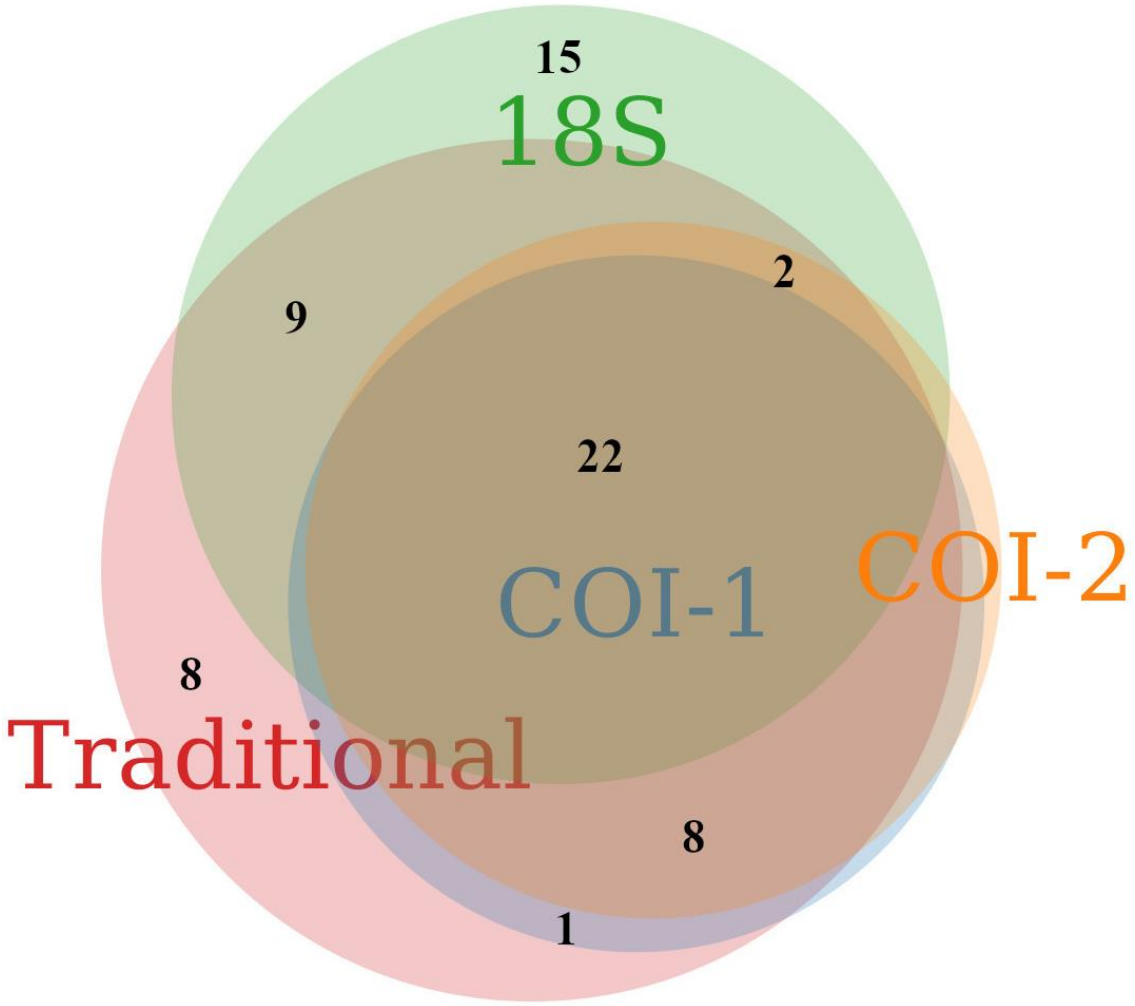



Fig. 6. Venn diagram of the number of species identified by the different methods. The overlap in recorded species between the traditional morphological identification and the three markers used for metabarcoding. MOTUs assigned to Tardigrada below the 97% threshold are not included in the figure.

Based on the BC-distances, both methods recorded moss, lichens and litter to host dissimilar tardigrade communities (Fig 7). Litter samples contained the most distinct and consistent composition of tardigrade species by both methods. Moss and lichens, although being significantly different in their composition, showed an area of overlap. This overlap was observed by a single lichen sample (226) containing very similar composition of species as the moss samples. For both methods, there were detected no statistically significant difference in the dispersion of samples between the three substrates (p-values= 0.025 and 0.09). However, lichen samples contained the highest variability in composition, followed by moss samples. Overall,

less of the variance in sample distribution was explained in the metabarcoding PCoA (54% by both axes) than in the PCoA based on traditional data (71% and 74% for axis 1 and 2). This variance represents how well the PCoA maintains the relationship between samples when going from n dimension to 2 dimensions, and thereby indicates that more information was lost for the metabarcoding data than traditional data. For both analyses, the approximated sample coordinates were concordant with the true similarity distances between the samples ($R^2$=0.91 for traditional data and $R^2$=0.77 for metabarcoding data). Furthermore, although there were slight differences in the distribution of samples between the methods, the major patterns were similar (Fig 7).

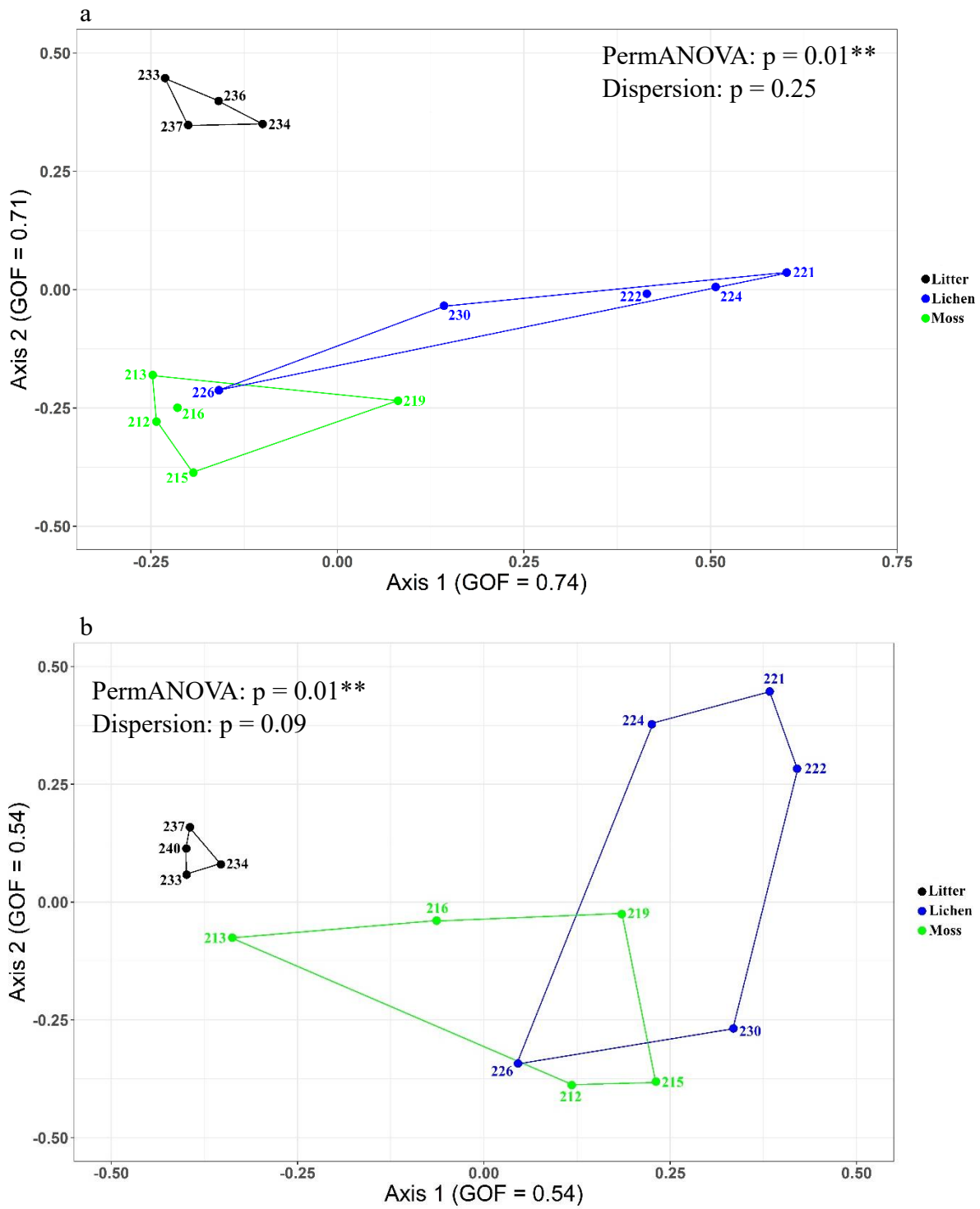



Figure 7. PCoA of the Bray-Curtis distances on relative frequencies of species detected by a) traditional methods and b) metabarcoding of the 18S marker. The p-values for the differences between clusters and dispersion within clusters are listed in the right corner.

## Discussion

With its ability to identify cryptic species and species elusive to handpicking-methods, metabarcoding will in many cases detect higher species richness than traditional methods (Taberlet et al., 2018). This was confirmed for tardigrades, as the number of DNA-species detected by metabarcoding exceeds the number of morpho-species detected by traditional methods. The unlabeled MOTUs assigned to Tardigrada below the sequence similarity threshold, coupled with the incoherence of reference databases indicate that the species numbers obtained are underestimated. The use of multiple markers facilitated resolute detection of tardigrade species as the shortcomings of one marker's reference library were often complemented by the other markers. This approach did, however, complicate the comparison of species lists for traditional and metabarcoding data. As tardigrades are known to contain high level of cryptic diversity, it is difficult to synonymize cryptic variants detected by different markers. To circumvent this challenge, the taxonomic resolution for such species groups were elevated to a higher taxonomic level. This approach discards information, but ensures comparison of truly equivalent taxa. For assessing the community composition, no species were merged, as only data from the 18S marker was used, due to its higher consistency and more complete reference library compared to COI. For the PCoA analysis, metabarcoding was able to successfully highlight differences in species assemblage between moss, lichen and litter substrates, identifying the same patterns as traditional methods. The lower goodness of fit for the metabarcoding data, both regarding variance explained by axis 1 and 2, and the $R^2$, is expected. This data has a more complex structure, as it encompasses more species, as well as cryptic variants considered as one single species by morphology. Larger, more complex datasets, will always be harder to compress from n dimensions to 2 dimensions by the nature of how complexity of data works (Binder, 1983).

Comparing the results of the two methods is not as straight forward as it seems, as both methods have their pros and cons aside from how they capture tardigrade diversity. Although tedious and time-consuming, the activity of hand-picking individual tardigrades yields quality species data, including the abundance of adults, larvae and eggs. The taxonomy of tardigrades is, however, a complex topic as it in many cases require the presence of both adult specimen and eggs for confident species identification (Kaczmarek & Michalczyk, 2017; Ramazzotti, 1962). This is often not achievable, either due to eggs not being present, or the eggs being too inconspicuous to

be detected. Additionally, the primitive morphological state of these minute organisms has produced several groups of species with similar appearance (Cesari et al., 2019; Fontoura & Pilato, 2007; Kaczmarek & Michalczyk, 2017). As traditional sampling is solely based on morphology, species are often identified to species complex group level. Accompanied by DNA barcoding, tardigrade studies have in the last decade been able to at least partly circumvent this issue (Bertolani et al., 2011; Blaxter et al., 2004; Jørgensen, Møbjerg, & Kristensen, 2007). Unfortunately, such single-individual barcoding approaches are practically impossible to apply in large scale tardigrade studies, as they often include thousands of individuals. In this aspect, metabarcoding has a major advantage as it is able to detect thousands of species in a single run (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). However, this data does not contain abundance of said species, only the relative proportion of their DNA. Additionally, these proportions are normally biased, meaning inferences made on such data might not reflect the true species proportions (Taberlet et al., 2018). Within metabarcoding runs, these biases should affect all samples similarly. As the observed patterns in community composition were similar for both methods used in our study, these biases seem to have negligible effects on the inferences. Generally, traditional surveys including species abundances are considered to be less biased, but several factors argue against this. Such surveys will lack information on cryptic species, which are common in tardigrades. Additionally, the species tables are likely skewed, as some species are more conspicuous than others, meaning a higher proportion of individuals of these species will be detected. Lastly, it is generally accepted that some species will always remain undetected (Connor & McCoy, 1979). As the traditional sampling method used in this study was very thorough, we tried to limit such biases in the data, but they are definitely still present to some extent. Additionally, as the orientation of the animals on the microscopy slide is crucial for observing taxonomic traits, a portion of our specimen could not be identified to species level due to being compressed, twisted or destroyed. If the cover slip is not applied carefully, the transparent specimen disappears to the outside of the coverslip by drifting along as the mounting medium is compressed by the glass. These movements can change the orientation of the specimen, affecting the visibility of taxonomic characters. The consequence is uncertain species identifications, circumvented by reducing the identification from species level to higher taxonomic level, be it genus-, subfamily- or family level. Depending on how such ambiguous data are treated in downstream analyses, they can lead to biased inferences (Olsgard, Somerfield,

& Carr, 1997). In this study, several species complex group- and genus level identifications were made. Such data are informative when assessing community composition and microhabitat differences, but can generate false similarity between communities (Bartels & Nelson, 2007; Olsgard et al., 1997). As several species appear morphologically similar, when they in reality are genetically different, they are treated as one species by traditional morphology-based methods. If these cryptic variants are exclusively found in different samples, the recorded data will count them as the same species occurring in all samples. This is likely to be the case for many tardigrade studies, including this one, causing the analyses done on sample differences to suffer from biased estimates towards more similar composition of species. As significantly distinct clusters of samples of all three substrates were still obtained, it seems like this bias can be neglected. However, such an event can also occur within-substrate and not between, meaning that samples of one substrate type appear more similar than what they really are. This will reduce within-cluster variance, while maintaining or decreasing the between-cluster variance, resulting in an altered ratio between components of variation used in the PermANOVA and dispersion analyses. Inferences based on these estimates will be biased towards more dissimilar communities (Olsgard et al., 1997). As many confidently identified morpho-species were observed in two or all substrates, such an event is considered unlikely. Conversely, this overlap highlights that tardigrade species often occupy several types of microhabitat. Relating this to the community similarity analysis reveals that although tardigrade species inhabit different substrates, their abundance in those substrates are highly different, often being negatively correlated. Whether this is due to altered abiotic conditions between substrates, different competitive pressure, or some other factors, remains unclear.

Several authors have reported patchiness as a common trait for tardigrade populations (Bartels & Nelson, 2007; Degma, Katina, & Sabatovičová, 2011; Meyer, 2006). Patchy populations increase the variability between measurements, requiring more samples to be investigated to obtain a complete inventory of a substrate's diversity (Meyer, 2006). Although the volume of samples investigated in this study was large, only five samples of each substrate type was investigated. The accumulation curves of our sampling effort clearly indicate that we have not reached the asymptotic level considered as sufficient sampling effort (Thompson & Withers, 2003) (Appendix D, fig. D1). This means that although clear clusters were observed for each substrate, we cannot conclude that the observed pattern constitutes true compositional differences, as it

could be an artifact of randomness due to insufficient sampling. Nevertheless, as our focus is on comparing the ability of metabarcoding and traditional sampling to capture tardigrade diversity, reaching the asymptote on the accumulation curve is not required. The captured diversity of different substrates by the two methods, and the methods' ability to assess the similarity still holds, regardless of how the true patterns of community similarity are. Interestingly, several authors have investigated the effect of microhabitat on the composition of tardigrades species, all of which have detected no significant effect (Guidetti et al., 1999; Ito, 1999; D. R. Nelson & Bartels, 2007). These studies, most of which have larger sampling efforts than our study (although smaller sample volume), indicate that the observed pattern of the dissimilarity analysis is an artifact of random noise. However, as the observed pattern is highly significant, it should not be disregarded completely, but rather investigated further. As there exist numerous species of moss and lichen in Skråstadheia, the observed distributional patterns of samples may be due to different host species being collected. Upon investigation of what mosses and lichens were collected, the samples often differed in their host species. This raises the question whether the increased variability in tardigrade species composition in moss and lichen samples is due to the different hosts or if it is true variability, independent of host species. The lower variation between samples of litter compared to moss and lichen will in this case be expected, as litter is a more homogenous habitat.

Several protocols for tardigrade sampling have been reported in the literature (Bartels & Nelson, 2006; Dastych, 1985; Ramazzotti & Maucci, 1983; Sands, Convey, Linse, & McInnes, 2008; Stelzer, 2009). The mechanical approach of Bartels and Nelson (2006), and the density gradient approach of Sands et al. (2008) are likely less time-consuming for collecting tardigrade specimen than our study, but will not produce inventories appropriate for methods using eDNA. Our study utilizes an approach based on decantation, in which objects heavier than water, such as tardigrades, are separated from water after removing large particles. The method successfully discards most debris larger than 500 µm while concentrating tardigrades and eggs into a ~100 mL volume. Whether this method is more accurate than other methods is difficult to assess, but as most unwanted particles are removed, the number of objects able to hide tardigrades is reduced. This should increase the conspicuousness of tardigrades, yielding superior detection rates. Additionally, as the method is based on immersing the sample in water, the whole sample can be processed using DNA-based filtration methods. This enables comparability between

traditional and molecular approaches, allowing us to evaluate the proficiency of both methods on the same samples. This statement is based on the assumption that the different subsamples contain equal composition of tardigrades after homogenization.

Metabarcoding is clearly a viable tool for tardigrade diversity assessment, but is currently dependent on the usage of more than one marker, as a single marker approach will suffer from incoherence of reference databases. Developing these reference databases by generating barcodes of more tardigrade species will increase the quality of species data recovery. This is likely to be the case for all three markers used in this study. Although species tables detected by 18S and COI are difficult to combine, especially for overlapping identifications of species complex groups, combined results yield more complete species inventories. As our study observed high levels of cryptic diversity, overlapping identifications (species identified by both markers, but not necessarily the same species) were merged to one as a conservative approach. This means that as a worst-case scenario, our estimates are biased towards lower species richness. By using total MOTU numbers, any inference made will include artificial MOTUs (i.e. erroneous estimates) that were not removed during the filtering pipeline (*de novo* chimera detection, quality filtering, singleton removal etc.). These spurious MOTUs can only be completely circumvented by using reference libraries, as they are highly unlikely to match a true reference sequence (Brown et al., 2015). Interestingly, the number of MOTUs assigned to Tardigrada was similar for all three markers, indicating either few, or similar, numbers of erroneous MOTUs for all markers. The slightly higher MOTU recovery by the COI-2 marker than COI-1 is likely explained in their different specificity to tardigrade sequences. As the COI-1 primers are universal, they target and amplify other taxa as well, meaning a deeper sequencing depth is needed to obtain similar numbers to the more specific COI-2. Artificial MOTUs are marker specific and should occur at a rate given by the intraspecific variability and region similarity between species (Edgar, Haas, Clemente, Quince, & Knight, 2011). Thus, as the numbers are similar for all three markers, it indicates reliable species richness estimates. This is further supported by the conservative read threshold used to define a species presence, as spurious sequences should occur in few numbers (Edgar et al., 2011) and are thereby discarded.

The proficiency of metabarcoding in detecting tardigrade diversity is dependent on its ability to amplify sequences originating from widely diverged species. Amplification is often biased due to

primer mismatch and will consequently skew the read numbers obtained for the different MOTUs (Polz & Cavanaugh, 1998). The *P. suillus* species not recovered by COI-1 or COI-2 is likely due to primer bias, as it was observed in high numbers during traditional sampling. This is further supported by the fact that DNA barcoding efforts of this species was unsuccessful in most cases during the construction of a local reference library. As *P. facettalis* was recovered by 18S, it could be a wrongly annotated reference sequence of *P. suillus*. However, Cesari et al. (2019) recently barcoded individuals of *P. suillus* and *P facettalis*, identifying high intraspecific variability, but no clear subdivision between the specimen. It is therefore likely that the incongruence between the metabarcoding data and morphology data for these species is explained in their lack of a barcode gap. The *P. facettalis* detection, also indicates that the primer bias observed in COI does not apply to 18S, further supporting the use of multiple markers. The same case might apply to the *Mesobiotus harmsworthi* group, which was detected by 18S and traditional methods, but not COI-1 and COI-2. Despite these two false negatives, both COI and the 18S markers independently detected species spanning most families of terrestrial tardigrades. Hence, metabarcoding should also be applicable to species of tardigrades not encountered in our study. However, not all metabarcoding records can be trusted. The *Acutuncus antarcticus* recorded by the 18S marker is likely not found in Norway and is therefore considered a false positive. Blasting its sequence in NCBI showed a high similarity (98.6%) to other *Acutuncus antarcticus* sequences. Interestingly, the blast also revealed a 97.3% similarity to both *Mesocrita revelata* and *Calohypsibius sp.* both previously recorded in Norway (Gąsiorek et al., 2016; Meier, 2017; this study). This highlights a possible scenario of barcode overlap, or alternatively, a wrongly annotated reference sequence. Either way, this emphasizes the danger of blindly trusting results inferred from downloaded reference sequences.

Another challenging task is to determine the lower read threshold for when to consider a species as present. Depending on the number of tag- and index jumps, a threshold must be selected to mitigate the inclusion of false positives (Taberlet et al., 2018). However, as singleton species are often represented by few reads, setting the threshold is a trade-off between exclusion of true positives and inclusion of false positives. As we included no positive control in our study, and consequently had no way to estimate tag jump rates, we chose a strict threshold by discarding all MOTUs observed by less than 10 reads. Our threshold seems conservative, as it removed several MOTUs with close matches to sequences in our reference library. Interestingly, three of these

species were not detected by the traditional sampling, and have neither been recorded in Norway previously (Meier, 2017). This raises the question whether they are true positives or not. They could be singleton species elusive to the traditional sampling, and as Norway has not been thoroughly investigated for tardigrade species, they could be rare encounters not detected in previous studies. Two of the species excluded by the read threshold were also found during traditional sampling (*Platicrista angustata* and *Hebesuncus conjugens*), represented by a few individuals. Nevertheless, the minimum 10 read criterion was kept as we considered exclusion of false positives more important than inclusion of all rare species when comparing the two methods. A metabarcoding approach using fusion primers instead of PCR-based tagging will reduce the number of tag jumps, allowing a lower exclusion threshold, which in turn will facilitate inclusion of low read species detections (Elbrecht & Leese, 2015).

The combined output of the three markers identified most of the species recorded by traditional sampling, all while suffering from the exclusion of true positives. Again, this means the proposed approach is conservative, but successful. However, this comparison was based on total diversity captured by the two methods. For sample-on-sample comparisons, species numbers observed by metabarcoding were more variable. For moss and lichen samples, PCR yields were consistent and, in all cases, identified more species than traditional sampling. These comparisons were done excluding unassigned Tardigrada MOTUs, meaning the true number of species is likely higher. However, for litter samples, metabarcoding output was highly inconsistent. These samples either contained no amplicon after PCR or had deviating PCR replicates leading to exclusion of these samples from further analysis. Several adjustments were attempted (diluted DNA, more PCR cycles, increased concentration of polymerase, primer and $MgCl_2$) all of which were unsuccessful. Only the 18S marker was able to produce consistent inventories for these samples (but was unsuccessful for 236), and the recovered species numbers were often lower than those obtained from traditional sampling. Litter is known to contain PCR inhibiting components (Griffiths, de Groot, Laros, Stone, & Geisen, 2018; Miller, Bryant, Madsen, & Ghiorse, 1999), meaning even if eDNA is successfully extracted from the sample, no or little amplification occurs during PCR. As all litter samples contained tardigrade specimen, although generally in lower abundances than moss and lichen samples, they should contain extractable tardigrade eDNA. Furthermore, the higher consistency of PCR replicates of 18S, indicates that the disparity is either due to the inapplicability of COI markers on low concentration of

tardigrade DNA, inhibiting conditions during PCR, or simply random stochasticity made visible by the low number of replicates per extract. In fact, tardigrade species were detected in all litter samples by all three markers. The PCR replicates of the excluded litter samples deviated too much in their composition, resulting in the samples being excluded as there was no way to decide which replicates to trust. Following Taberlet et al. (2018) and Zinger et al. (2019), one should decide whether to take the mean of all technical replicates for a sample, or to take the mean only of the consistent replicates. We chose the latter, as the eDNA extracts often had low DNA concentration, and thereby had variable amplification success during PCR. For each sample, we wanted to reconstruct the composition of species as similar to the true composition as possible, and unsuccessful PCR replicates would in this case heavily influence the sample estimates. Arguably, this approach led to the exclusion of some samples, but yielded quality inventories of the remaining ones.

The species numbers obtained by metabarcoding on these samples include all cryptic variants weighed the same as any other species. Accordingly, the species numbers are expected to be higher than morphologically sampled numbers, even if some species remain undetected. Furthermore, the data used in these comparisons were based on number of MOTUs matching reference sequences above the used 97% threshold. As only about 60% of the MOTUs assigned to Tardigrada matched a reference sequence above the threshold, several species present in the samples were likely excluded from the analysis as the reference library lacked the species' reference sequences. No attempts were made on lowering the similarity threshold in order to assign taxonomic annotations on higher taxonomic levels within Tardigrada. The high occurrence of unlabeled MOTUs indicates that on sample level, metabarcoding suffers from the same shortcomings as for total diversity assessment; incomplete reference databases prevent the detection of certain species. These MOTUs are obviously present, as they are detected and assigned to Tardigrada. Such MOTUs are informative and should not be discarded completely, although caution must be taken when interpreting the results. In public databases, many sequences of low quality are deposited. These include rotifer sequences labeled as 'Tardigrada', and 'Tardigrada environmental sample' sequences (i.e. unknown sequences already matched to a database before being deposited). Before blasting, such sequences had to be removed from the reference library to avoid their inclusion in the diversity tables as unnamed MOTUs. For the community similarity analysis on the metabarcoding data, all MOTUs were included to infer on

the true composition of tardigrade MOTUs, and not only reference-based species records. Using alpha diversity measures (e.g. total species/MOTU richness) should be avoided as they are often overestimated (Jeunen et al., 2019). The beta diversity, i.e. the sample-wise comparison of species, can be evaluated within the same study, as all samples were normalized and are confounded by the same conditions and biases (Taberlet et al., 2018). Clearly, any amplification bias will skew the abundance of species, meaning the sequence data should be used with caution. The Bray Curtis distances on such data will to some degree suffer from these biases as they are based on species abundances or proportions. When used for within-study comparisons of beta diversity, such methods are still viable due to the same reasons stated above. The clustering of samples of different substrates showed, based on the BC-distances, very similar patterns as the communities sampled by morphological methods. Interestingly, the metabarcoding data indicated that two of the moss samples, which were distinctly different by traditional sampled data, contained very similar composition of tardigrade species as the litter samples. Which of the methods best describe the true pattern is impossible to say based on our data, but clearly there is a discrepancy between the two methods for these samples. The explanation likely lies in the number of PCR replicates used in the study. As only three replicates were used, there is a chance that two outlier PCR replicates containing spurious amplicons were similar by chance, and was thereby included in further analysis. As nearly 33% of all PCR replicates were inconsistent, and thereby excluded, such an event is not unlikely. Inconsistency between PCR replicates of eDNA is a common trait (Deiner et al., 2017). By using more PCR replicates, the probability of including such spurious measurements drastically decreases. It would also likely result in higher consistency for the litter samples that were excluded during PCR pruning. Following Taberlet et al. (2018), up to 9 replicas should be included in eDNA studies using soil samples. Although successful for most samples, our three replicates resulted in the exclusion of several samples, meaning information on these samples is lost. As tardigrades are known to have a high rate of singleton species (Bartels & Nelson, 2007), the use of more replicates is likely to result in a higher capture rate of rare taxa. For the 8 species detected by traditional methods, but not metabarcoding, most were singletons or rare species. Furthermore, rare species are more unlikely to be encountered in traditional surveys, and thereby have a lower probability of being barcoded. Thus, it is likely that at least some of the singletons elusive to metabarcoding remain hidden as unlabeled MOTUs due to their lack of reference sequence. An effort of assigning barcodes to

more tardigrade species will help solve this challenge, and as DNA based methods are on the rise within the field, the near future is likely to see more tardigrade species deposited in public databases.

## Conclusion

Traditional methods are time-consuming and often limited in their taxonomic resolution and are accordingly not suitable for large scale ecological studies. Metabarcoding overcomes these issues by its rapid conduct and its ability to identify species regardless of morphology and specimen developmental stage. Metabarcoding of tardigrades is, however, currently suffering from the incoherence of reference databases. A multi-marker approach partly circumvents this issue, but complicates the interpretation process. This is especially apparent when trying to synonymize species in the presence of cryptic variants. Nevertheless, metabarcoding of eDNA grants superior detection of tardigrade species, as well as capture of differences in community composition between host substrate. Looking beyond tardigrades, the approach used in this study should be applicable to most micro- and mesofauna, requiring only minor adjustments to the protocol.

# References


Balvanera, P., Pfisterer, A. B., Buchmann, N., He, J. S., Nakashizuka, T., Raffaelli, D., & Schmid, B. (2006). Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology letters, 9*(10), 1146-1156.

Bartels, P. J., & Nelson, D. R. (2006). A large-scale, multihabitat inventory of the Phylum Tardigrada in the Great Smoky Mountains National Park, USA: a preliminary report. *Hydrobiologia, 558*(1), 111-118.

Bartels, P. J., & Nelson, D. R. (2007). An evaluation of species richness estimators for tardigrades of the Great Smoky Mountains National Park, Tennessee and North Carolina, USA. *66*(1s), 104-110.

Begon, M., Townsend, C. R., & Harper, J. L. (2005). *From individuals to ecosystems*: Blackwell Publishers Hoboken.

Bertolani, R., Rebecchi, L., Giovannini, I., & Cesari, M. (2011). DNA barcoding and integrative taxonomy of Macrobiotus hufelandi CAS Schultze 1834, the first tardigrade species to be described, and some related species. *2997*(19), 19-36.

Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J., & Taberlet, P. (2012). Tracking earthworm communities from soil DNA. *21*(8), 2017-2030.

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *27*(4), 233-243.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 279-292.

Bingemer, J., & Hohberg, K. (2017). An illustrated identification key to the eutardigrade species (Tardigrada, Eutardigrada) presently known from European soils.

Blaxter, M., Elsworth, B., & Daub, J. (2004). DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *271*(suppl_4), S189-S192.

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., . . . Jumpponen, A. (2015). Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecology, 13*, 221-225.

Cesari, M., Montanari, M., Kristensen, R. M., Bertolani, R., Guidetti, R., & Rebecchi, L. (2019). An integrated study of the biodiversity within the Pseudechiniscus suillus–facettalis group (Heterotardigrada: Echiniscidae). *Zoological Journal of the Linnean Society*. doi:10.1093/zoolinnean/zlz045

Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist, 113*(6), 791-833.

Dastych, H. (1985). West Spitsbergen Tardigrada. *Acta Zoologica Craviensia, 28*, 169-214.

Dayrat, B. (2005). Towards integrative taxonomy. *Biological journal of the Linnean society, 85*(3), 407-417.

Degma, P., Katina, S., & Sabatovičová, L. (2011). Horizontal distribution of moisture and Tardigrada in a single moss cushion. *49*, 71-77.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., . . . De Vere, N. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *26*(21), 5872-5895.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*(16), 2194-2200.

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS one, 10*(7), e0130324.

Elbrecht, V., & Leese, F. (2016). Development and validation of DNA metabarcoding COI primers for aquatic invertebrates using the R package" PrimerMiner". *4*, e2044v2041.

FAO. (2018). The State of the World's Forests 2018. *Forest pathways to sstainable development*.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. J. B. l. (2008). Species detection using environmental DNA from water samples. *4*(4), 423-425.

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., . . . Ambrogio, L. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *12*(1), R1.

Fontoura, P., & Pilato, G. J. Z. (2007). Diphascon (Diphascon) faialense sp. nov. a new species of Tardigrada (Eutardigrada, Hypsibiidae) from the Azores and a key to the species of the D. pingue group. *1589*(1), 47-55.

Gąsiorek, P., Stec, D., Morek, W., Zawierucha, K., Kaczmarek, Ł., Lachowska-Cierlik, D., & Michalczyk, Ł. (2016). An integrative revision of Mesocrista Pilato, 1987 (Tardigrada: Eutardigrada: Hypsibiidae). *Journal of Natural History, 50*(45-46), 2803-2828.

Griffiths, B., de Groot, G. A., Laros, I., Stone, D., & Geisen, S. (2018). The need for standardisation: Exemplified by a description of the diversity, community structure and ecological indices of soil nematodes. *Ecological indicators, 87*, 43-46.

Guidetti, R., Bertolani, R., & Nelson, D. R. (1999). Ecological and Faunistic Studies on Tardigrades in. *238*, 215-223.

Guidetti, R., Peluffo, J. R., Rocha, A. M., Cesari, M., & de Peluffo, M. C. (2013). The morphological and molecular analyses of a new South American urban tardigrade offer new insights on the biological meaning of the Macrobiotus hufelandi group of species (Tardigrada: Macrobiotidae). *47*(37-38), 2409-2426.

Guil, N., & Giribet, G. (2009). Fine scale population structure in the Echiniscus blumi-canadensis series (Heterotardigrada, Tardigrada) in an Iberian mountain range—When morphology fails to explain genetic structure. *3*(51), 606-613.

Hebert, P. D., Cywinska, A., Ball, S. L., & Dewaard, J. (2003). Biological identifications through DNA barcodes. *270*(1512), 313-321.

Hohberg, K., & Traunspurger, W. (2005). Predator–prey interaction in soil food web: functional response, size-dependent foraging efficiency, and the influence of soil texture. *41*(6), 419-427.

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., . . . Naeem, S. (2005). Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological monographs, 75*(1), 3-35.

Ito, M. (1999). Ecological distribution, abundance and habitat preference of terrestrial tardigrades in various forests on the northern slope of Mt. Fuji, central Japan. *238*(3-4), 225-234.

Jeunen, G. J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., . . . Gemmell, N. J. (2019). Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecology and evolution, 9*(3), 1323-1335.

Jönsson, K. I. (2003). Population density and species composition of moss-living tardigrades in a boreo-nemoral forest. *26*(3), 356-364.

Jørgensen, A., Møbjerg, N., & Kristensen, R. M. (2007). A molecular study of the tardigrade Echiniscus testudo (Echiniscidae) reveals low DNA sequence diversity over a large geographical area. *66*(1s), 77-83.

Kaczmarek, Ł., & Michalczyk, Ł. (2017). The Macrobiotus hufelandi group (Tardigrada) revisited. *Contributions to Zoology, 4363*(1), 101.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics, 30*(22), 3276-3278.

Liang, J., Crowther, T. W., Picard, N., Wiser, S., Zhou, M., Alberti, G., . . . Pretzsch, H. (2016). Positive biodiversity-productivity relationship predominant in global forests. *Science, 354*(6309), aaf8957.

Lillebø, A. I., Flindt, M. R., Pardal, M. Â., & Marques, J. C. (1999). The effect of macrofauna, meiofauna and microfauna on the degradation of Spartina maritima detritus from a salt marsh area. *20*(4), 249-258.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *2*, e593.

Majaneva, M., Diserud, O. H., Eagle, S. H., Hajibabaei, M., & Ekrem, T. (2018). Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *2*, e26664.

Meier, T. (2017). *A checklist of Norwegian Tardigrada* (Vol. 37).

Meyer, H. A. (2006). Small-scale spatial distribution variability in terrestrial tardigrade populations. *558*(1), 133-139.

Miller, D., Bryant, J., Madsen, E., & Ghiorse, W. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol., 65*(11), 4715-4724.

Morek, W., Gąsiorek, P., Stec, D., Blagden, B., & Michalczyk, Ł. J. C. t. Z. (2016). Experimental taxonomy exposes ontogenetic variability and elucidates the taxonomic value of claw configuration in Milnesium Doyère, 1840 (Tardigrada: Eutardigrada: Apochela). *85*(2).

Mori, A. S., Lertzman, K. P., & Gustafsson, L. (2017). Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology, 54*(1), 12-27.

Nelson, Guidetti, & Rebecchi. (2015). Thorp and Covich's freshwater invertebrates. *1*, 347-380.

Nelson, D. R., & Bartels, P. J. (2007). "Smoky Bears"—Tardigrades of Great Smoky Mountains National Park. *Southeastern Naturalist, 6*(sp2), 229-239.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., & Oksanen, M. J. (2007). The vegan package. *10*, 631-637.
Olsgard, F., Somerfield, P. J., & Carr, M. R. (1997). Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. *Marine Ecology Progress Series, 149*, 173-181.
Orgiazzi, A., Dunbar, M. B., Panagos, P., de Groot, G. A., & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biology and Biochemistry, 80*, 244-250.
Pilato, G., & Binda, M. G. J. Z. (2010). Definition of families, subfamilies, genera and subgenera of the Eutardigrada, and keys to their identification. *2404*(1), 1-52.
Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol., 64*(10), 3724-3730.
Porazinska, D. L., GIBLIN-DAVIS, R. M., Esquivel, A., Powers, T. O., Sung, W., & Thomas, W. K. J. M. E. (2010). Ecometagenetics confirm high tropical rainforest nematode diversity. *19*(24), 5521-5530.
R Development Core Team. (2013). R: A language and environment for statistical computing.
Ramazzotti, G. (1962). *Il Phylum Tardigrada* (Vol. 14): Ise Cnr.
Ramazzotti, G., & Maucci, W. (1983). II Phylum Tardigrada. III. Redizione riveduta e aggiornata. (II Phylum Tardigrada. 3rd edited and updated edition.). *Memorie dell'Istituto Italiano di Idrobiologia, 41*.
Ramirez, K. S., Leff, J. W., Barberán, A., Bates, S. T., Betley, J., Crowther, T. W., . . . Steenbock, C. (2014). Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *281*(1795), 20141988.
Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes, 7*(3), 355-364.
Reitzel, A., Herrera, S., Layden, M., Martindale, M., & Shank, T. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology, 22*(11), 2953-2970.
Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *4*, e2584.
Sands, C. J., Convey, P., Linse, K., & McInnes, S. J. (2008). Assessing meiofaunal variation among individuals utilising morphological and molecular approaches: an example using the Tardigrada. *BMC ecology, 8*(1), 7.
Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology, 55*, 421-438.
Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Robinson, C. J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *75*(23), 7537-7541.
Schmid-Araya, J., Hildrew, A., Robertson, A., Schmid, P., & Winterbottom, J. (2002). The importance of meiofauna in food webs: evidence from an acid stream. *83*(5), 1271-1285.
Stelzer, C. P. (2009). Automated system for sampling, counting, and biological analysis of rotifer populations. *Limnology and oceanography: Methods, 7*(12), 856-864.
Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D., BREINER, H. W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *19*, 21-31.
Swift, M. J., Heal, O. W., Anderson, J. M., & Anderson, J. (1979). *Decomposition in terrestrial ecosystems* (Vol. 5): Univ of California Press.
Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*: Oxford University Press.
Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. J. M. e. (2012). Environmental DNA. *21*(8), 1789-1793.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology, 21*(8), 2045-2050.
Taberlet, P., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., & Choler, P. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *21*(8), 1816-1820.
Thompson, G. G., & Withers, P. C. (2003). Effect of species richness and relative abundance on the shape of the species accumulation curve. *Austral Ecology, 28*(4), 355-360.
Thorp, J. H., & Covich, A. P. (2009). *Ecology and classification of North American freshwater invertebrates*: Academic press.

Van Bezooijen, J. (2006). *Methods and techniques for nematology*: Wageningen University Wageningen, The Netherlands.
Wardle, D. A., Bardgett, R. D., Klironomos, J. N., Setälä, H., Van Der Putten, W. H., & Wall, D. H. (2004). Ecological linkages between aboveground and belowground biota. *304*(5677), 1629-1633.
Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., . . . Deagle, B. E. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular ecology, 28*(8), 1857-1862.

# Appendix

## Appendix A. Species lists

Table A1. Species observed by traditional sampling, identified by morphology. The species with successfully retrieved barcodes are marked in bold.

| | | |
|---|---|---|
| *Adropion arduifrons* | *Echiniscus sp. 2* | ***Macrobiotus cf. hufelandi*** |
| *Adropion scoticum scoticum* | *Echiniscus sp. 3* | ***Macrobiotus hufelandi group*** |
| ***Adropion prorsirostre*** | *Echiniscus sp. 4* | *Mesobiotus coronatus* |
| *Astatumen trinacriae* | *Echiniscus loxophthalmus* | ***Mesobiotus montanus*** |
| ***Calohypsibius ornatus*** | ***Fractonotus caelatus*** | ***Mesobiotus sp.*** |
| ***Diphascon pingue pingue*** | *Hebesuncus conjugens* | *Mesobiotus sp.1 (Egg)* |
| ***Diploechiniscus oihonnae*** | ***Hypsibius cf. convergens*** | *Milnesium sp. 1* |
| *Echiniscus arctomys group* | ***Hypsibius scabropygus*** | ***Milnesium sp*** |
| *Echiniscus cf. testudo* | *Hypsibius sp.1* | ***Minibiotus intermedius*** |
| ***Echiniscus spiniger*** | ***Isohypsibius sattleri group*** | ***Murrayon dianeae*** |
| ***Echiniscus spinulosus*** | *Isohypsibius sp. 1* | *Pilatobius oculatus oculatus* |
| ***Echiniscusm merokensis merokensis*** | *Itaquascon placophorum* | *Platicrista angustata* |
| ***Echiniscus quadrispinosus*** | *Macrobiotus furcatus* | ***Pseudechiniscus suillus group*** |

Table A2. Species detected by metabarcoding.

| | | |
|---|---|---|
| *Acutuncus sp.* | *Echiniscus trisetosus* | *Macrobiotus sp.* |
| *Adropion prorsirostre* | *Fractonotus caelatus* | *Mesobiotus harmsworthi group* |
| *Adropion scoticum* | *Hebesuncus conjugens* | *Mesobiotus insanis* |
| *Astatumen trinacriae* | *Hypsibiidae indet* | *Mesobiotus sp.* |
| *Calohypsibius ornatus* | *Hypsibius convergens* | *Mesocrista revelata* |
| *Calohypsibius sp.* | *Hypsibius cf. convergens* | *Microhypsibius bertolanii* |
| *Adropion sp.* | *Hypsibius scabropygus* | *Milnesium sp* |
| *Diphascon pingue sp. 1* | *Hypsibius sp.* | *Milnesium tardigradum* |
| *Diphascon pingue sp. 2* | *Isohypsibius sattleri* | *Minibiotus gumersindoi* |
| *Diphascon sp.* | *Itaquascon placophorum* | *Minibiotus intermedius* |
| *Diploechiniscus oihonnae* | *Macrobiotus hufelandi group 1* | *Minibiotus sp.1 TM* |
| *Echiniscus merokensis* | *Macrobiotus hufelandi group 2* | *Mixibius saracenus* |
| *Echiniscus quadrispinosus* | *Macrobiotus hufelandi group 3* | *Murrayon dianeae BIN 1* |
| *Echiniscus sp. 1* | *Macrobiotus hufelandi group 4* | *Murrayon dianeae BIN 2* |
| *Echiniscus sp. 2* | *Macrobiotus hufelandi group 5* | *Parachaela indet* |
| *Echiniscus sp. 3* | *Macrobiotus hufelandi group 6* | *Platicrista angustata* |
| *Echiniscus sp. 4* | *Macrobiotus scoticus* | *Pseudechiniscus facettalis* |
| *Echiniscus spiniger* | *Macrobiotus sp.* | *Pseudechiniscus sp.* |

## Appendix B. R-script for removing outlier PCR replicates

```
tag_bad_pcr=function(samples, counts, plot=TRUE){

counts=decostand(counts, method="hellinger")

#Transform count.

bc=aggregate(counts, by=list(factor(as.character(samples))), mean)

#grouping lines so that PCR correspond to the same sample. Then apply the
mean function for these. Remove the first column, and instead use it to name
the rows.

bc.name=as.character(bc[,1])

bc=bc[-1]

rownames(bc)=bc.name

bc=bc[as.character(samples),]

d=sqrt(rowSums((counts-bc)^2))

names(d)=as.character(samples)

#Eucledian distance between center (mean) and PCR replicate value

d.m=mode(d)

d.sd=sqrt(sum((d[d<=d.m]-d.m)^2)/sum(d<=d.m))

#estimate SD

d.max=aggregate(d, by=list(factor(as.character(samples))), max)

#Identify outlier PCRs by which one has the max eucledian distance to the
other PCRs of the sample

d.max.names=d.max[,1]

d.max=d.max[,2]

names(d.max)=d.max.names

d.max=d.max[as.character(samples)]

d.len=aggregate(d,

by=list(factor(as.character(samples))), length)

# For each PCR, count how many we have and keep if consistent (more explained
further down)

d.len.names=d.len[,1]

d.len=d.len[,2]

names(d.len)=d.len.names

d.len=d.len[as.character(samples)]

keep=((d<d.m+(d.sd*2))|d!=d.max)&d.len>1

selection=data.frame(samples=as.character(samples), distance=d,
maximum=d.max, repeats=d.len, keep=keep, stringsAsFactors=FALSE)
```

```
#Keep if it is not the highest PCR and if we got more than 1 PCR remaining in
the end.

rownames(selection)=rownames(counts)

attributes(selection)$dist.mode=d.m

attributes(selection)$dist.sd=d.sd

if(plot){

hist(d)

abline(v=d.m,lty=2,col="green")

abline(v=d.m+(d.sd*2),lty=2,col="red")

}

return(selection)

}

#Plot the output to visualize each iteration to see when we only have
consistent replicates left.
```

## Appendix C: DNA extraction table

Table C1. DNA concentration and quality estimates of samples measured with NanoDrop

| Sample | Nuleic Acid Conc. | A260 | A280 | 260/280 | 260/230 |
|---|---|---|---|---|---|
| 212 | 6.5 | 0.1 | 0.1 | 1.6 | 2.1 |
| 213 | 56.4 | 1.1 | 0.6 | 1.8 | 1.9 |
| 215 | 3.4 | 0.1 | 0.0 | 1.2 | 1.9 |
| 216 | 80.3 | 1.6 | 0.9 | 1.8 | 1.9 |
| 219 | 23.7 | 0.2 | 0.3 | 1.1 | 2.1 |
| 221 | 2.5 | 0.0 | 0.1 | 0.8 | 1.2 |
| 222 | 1.8 | 0.0 | 0.1 | 0.8 | 445.4 |
| 224 | 17.3 | 0.3 | 0.2 | 1.7 | 4.1 |
| 226 | 4.4 | 0.1 | 0.1 | 1.6 | 1.3 |
| 230 | 16.0 | 0.3 | 0.2 | 1.5 | -9.7 |
| 233 | 9.8 | 0.2 | 0.1 | 2.0 | 2.0 |
| 234 | 29.2 | 0.6 | 0.6 | 1.2 | 1.4 |
| 236 | 10.4 | 0.2 | 0.5 | 0.9 | 0.3 |
| 237 | 22.7 | 0.5 | 0.3 | 1.6 | 1.6 |
| 240 | 12.2 | 0.2 | 0.2 | 1.0 | 0.9 |

## Appendix D. Testing the assumptions of PermANOVA, dispersion test and PCoA

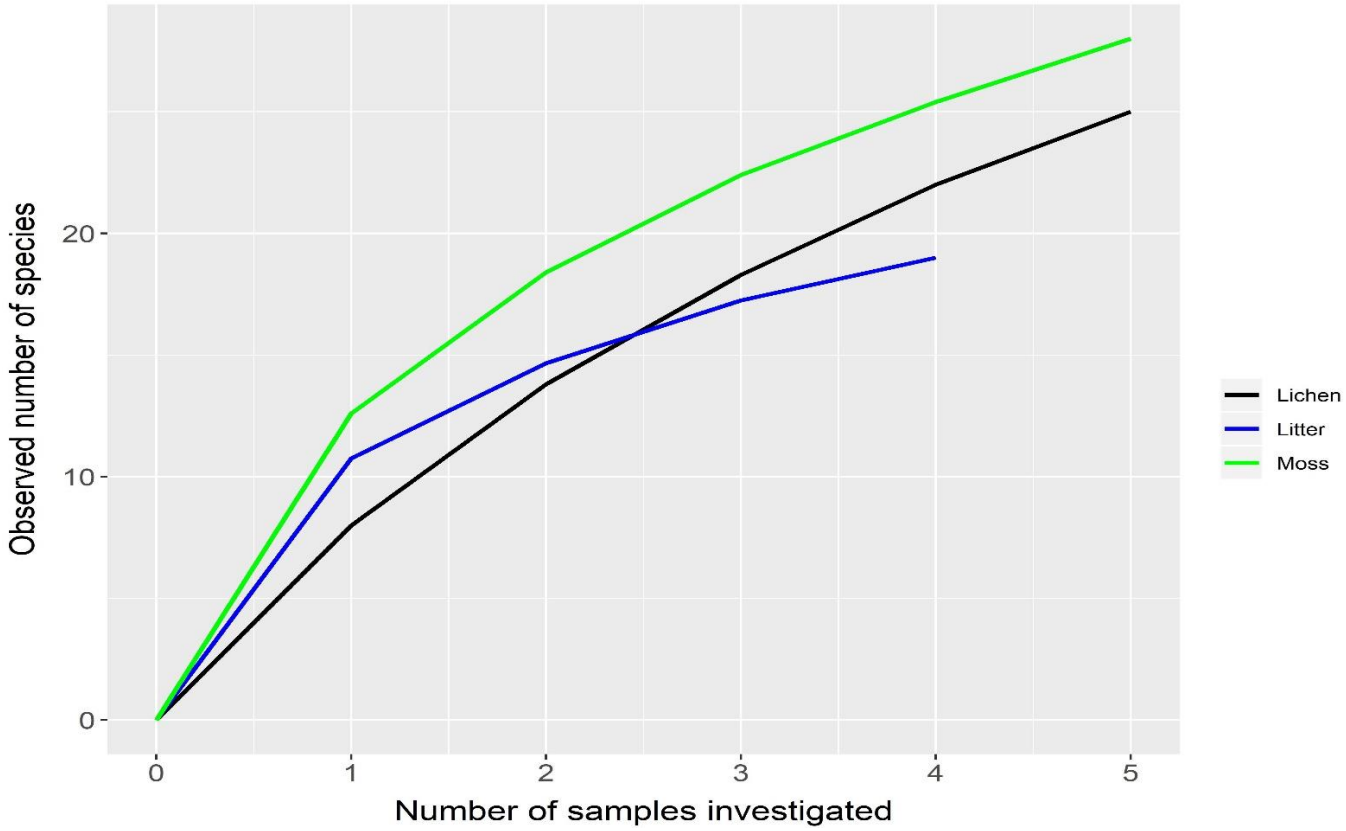



Figure D1. Species accumulation curves for tardigrade species recorded by traditional, morphology-based sampling.

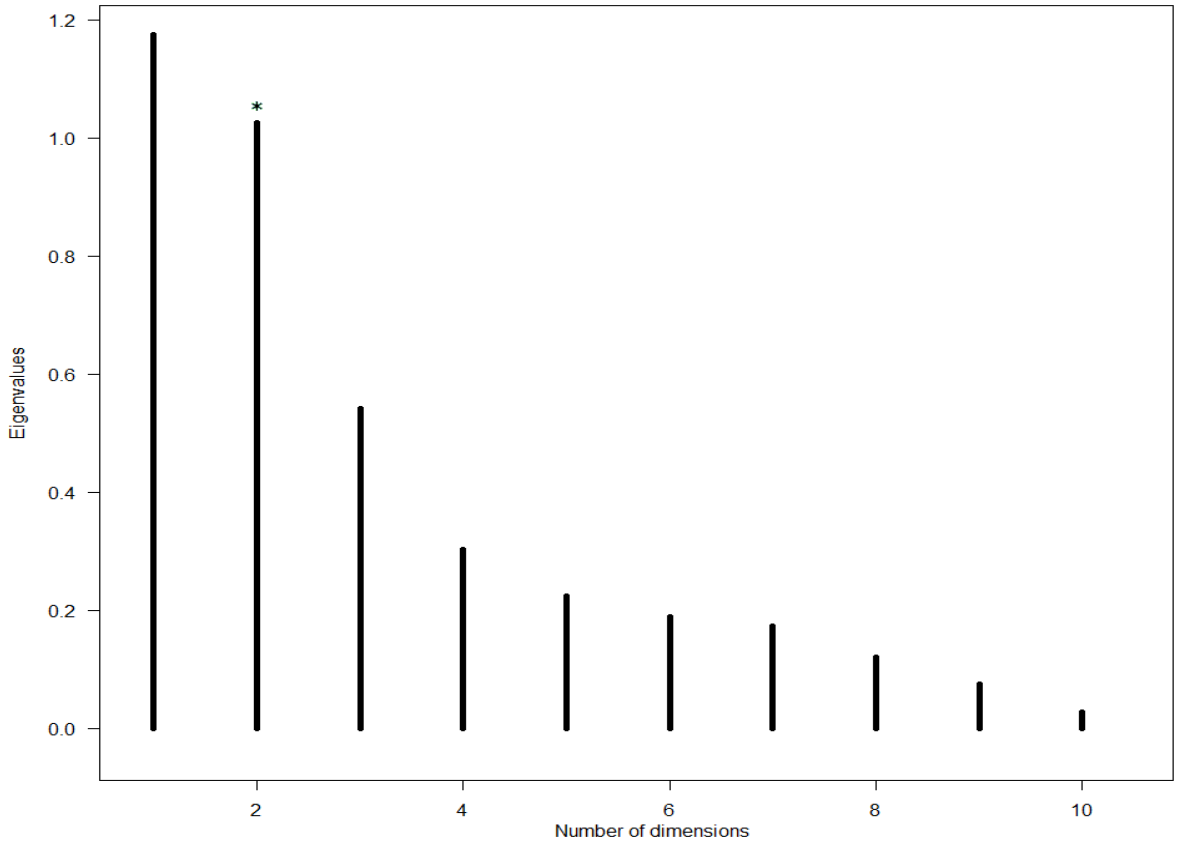



Figure D2. Calculated eigenvalues of the number different dimensions available to the PCoA plot used for the 18S marker metabarcoding data in this study. Asterisk marks the dimension used in the analysis.

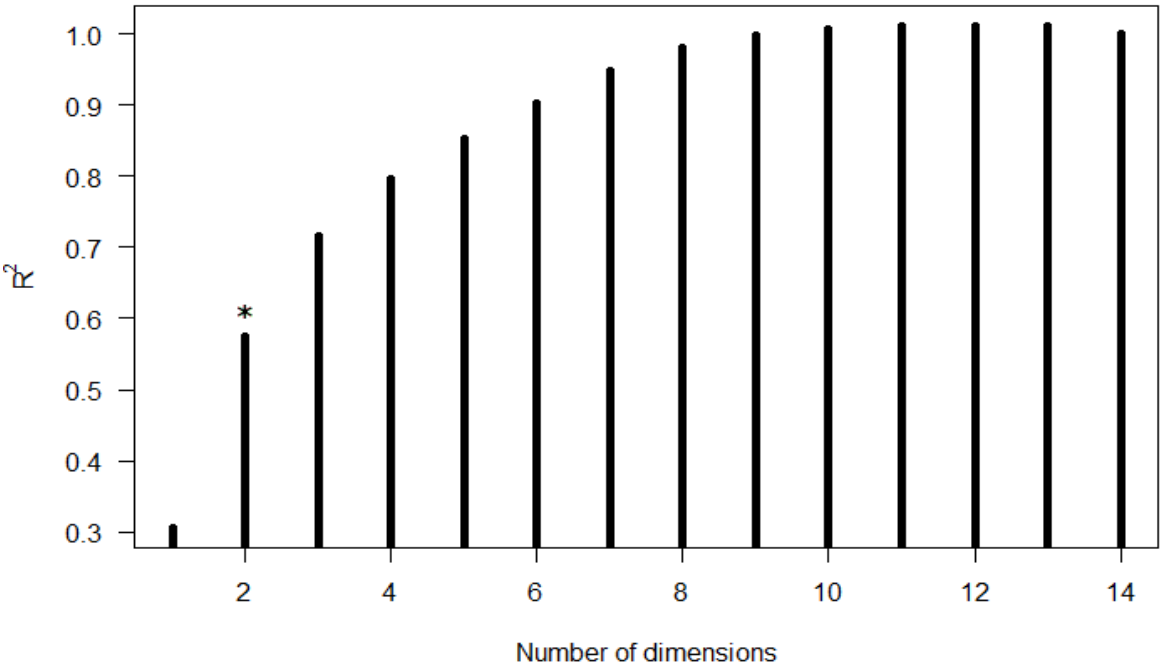


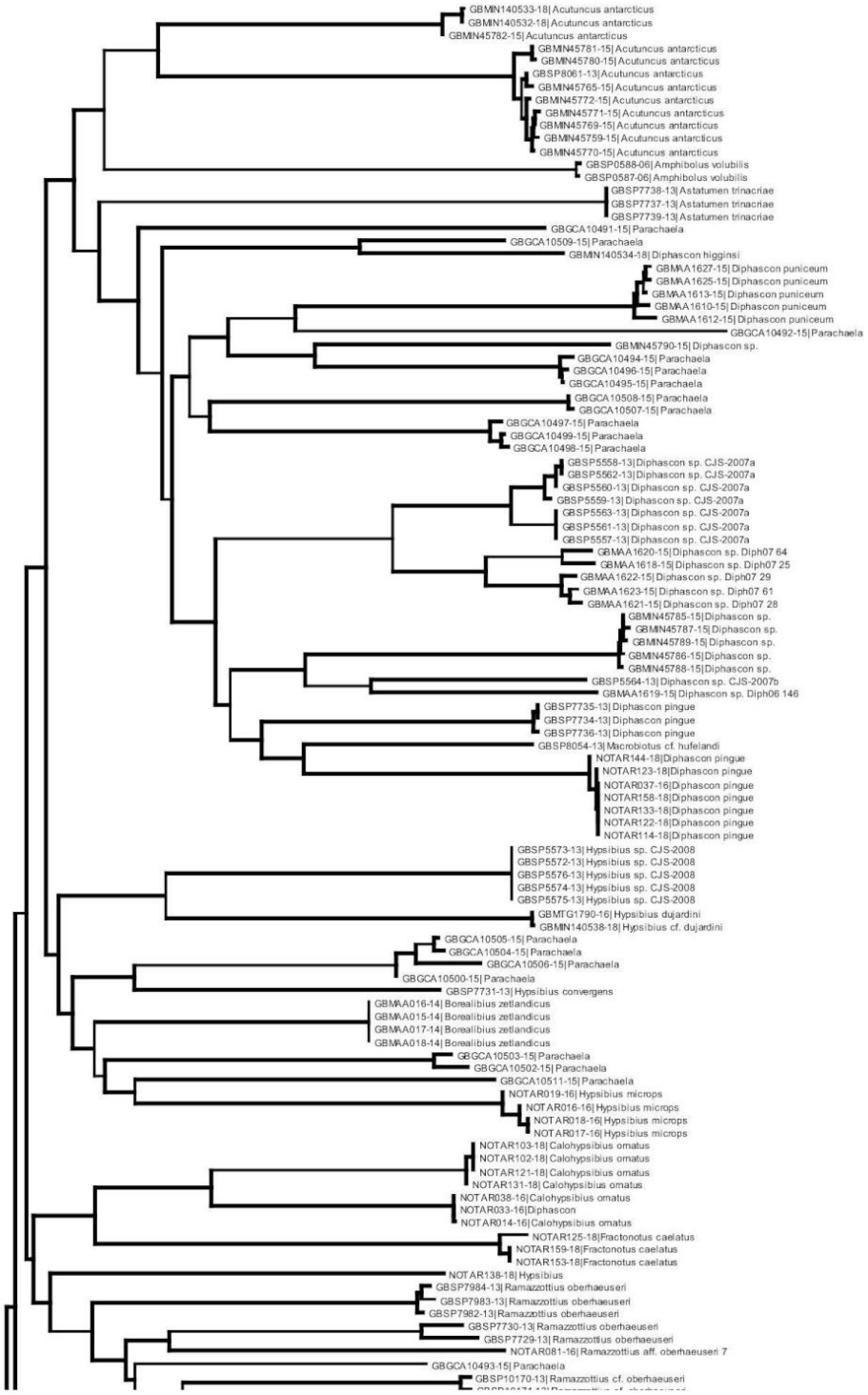

Figure D3. Calculated goodness of fit values for the different dimensions available to the PCoA plot used for the 18S marker metabarcoding data in this study. Asterisk marks the dimension used in the analysis.

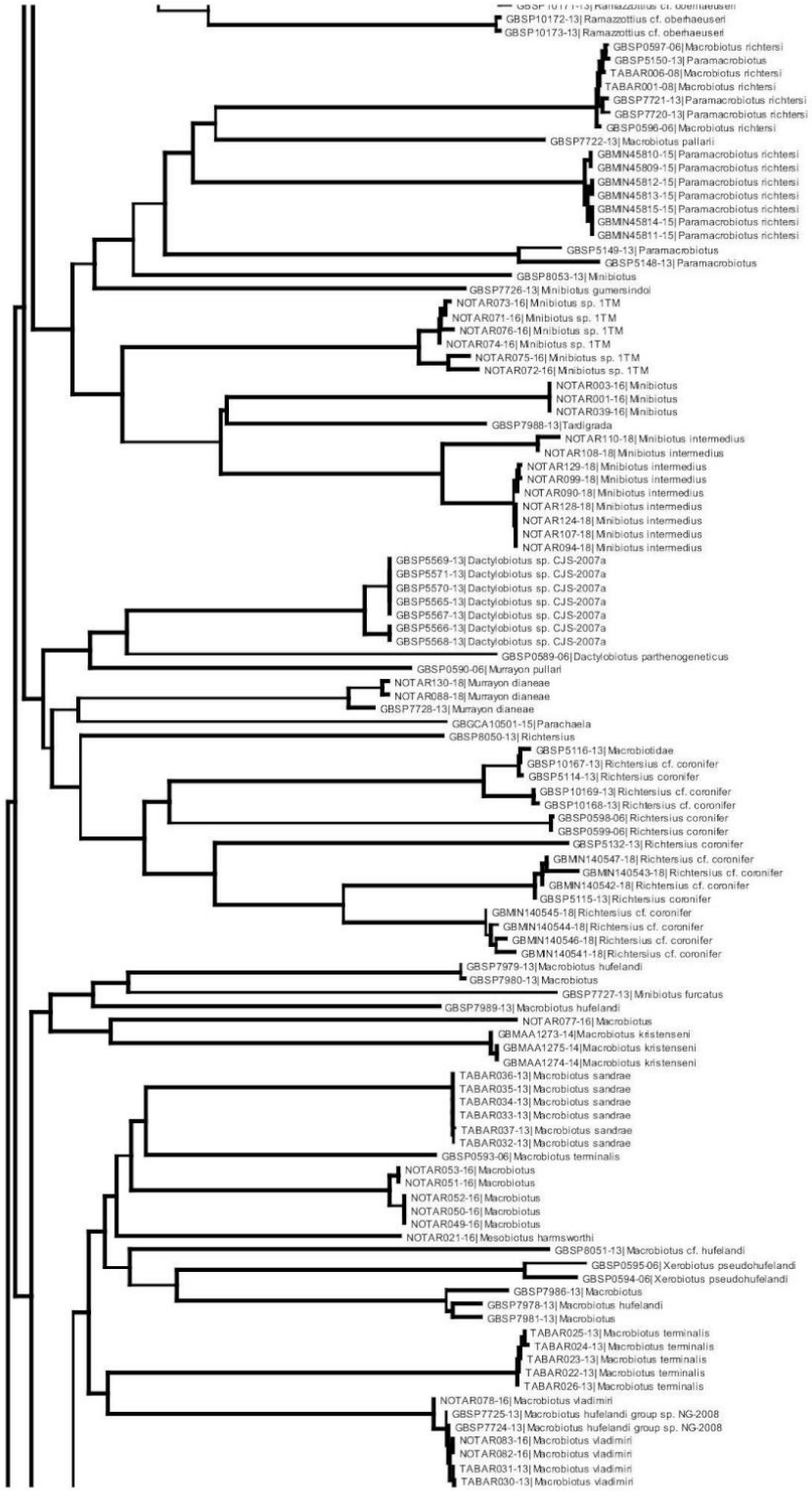
**Appendix E. DNA extraction protocol for barcoding of tardigrades**

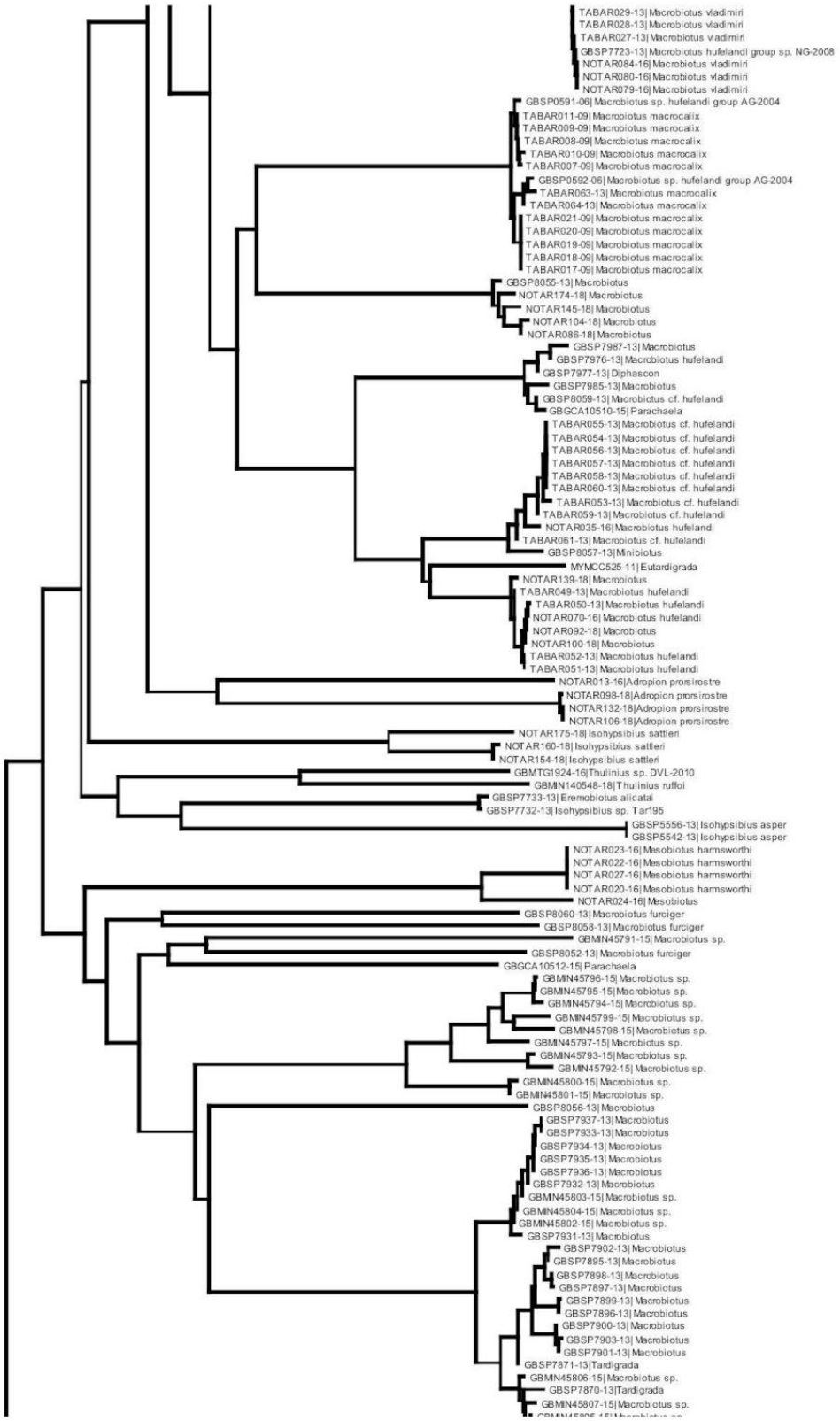
Protocol for DNA extraction of single tardigrades using QuickExtract™ DNA Extraction Solution kit by Lucigen

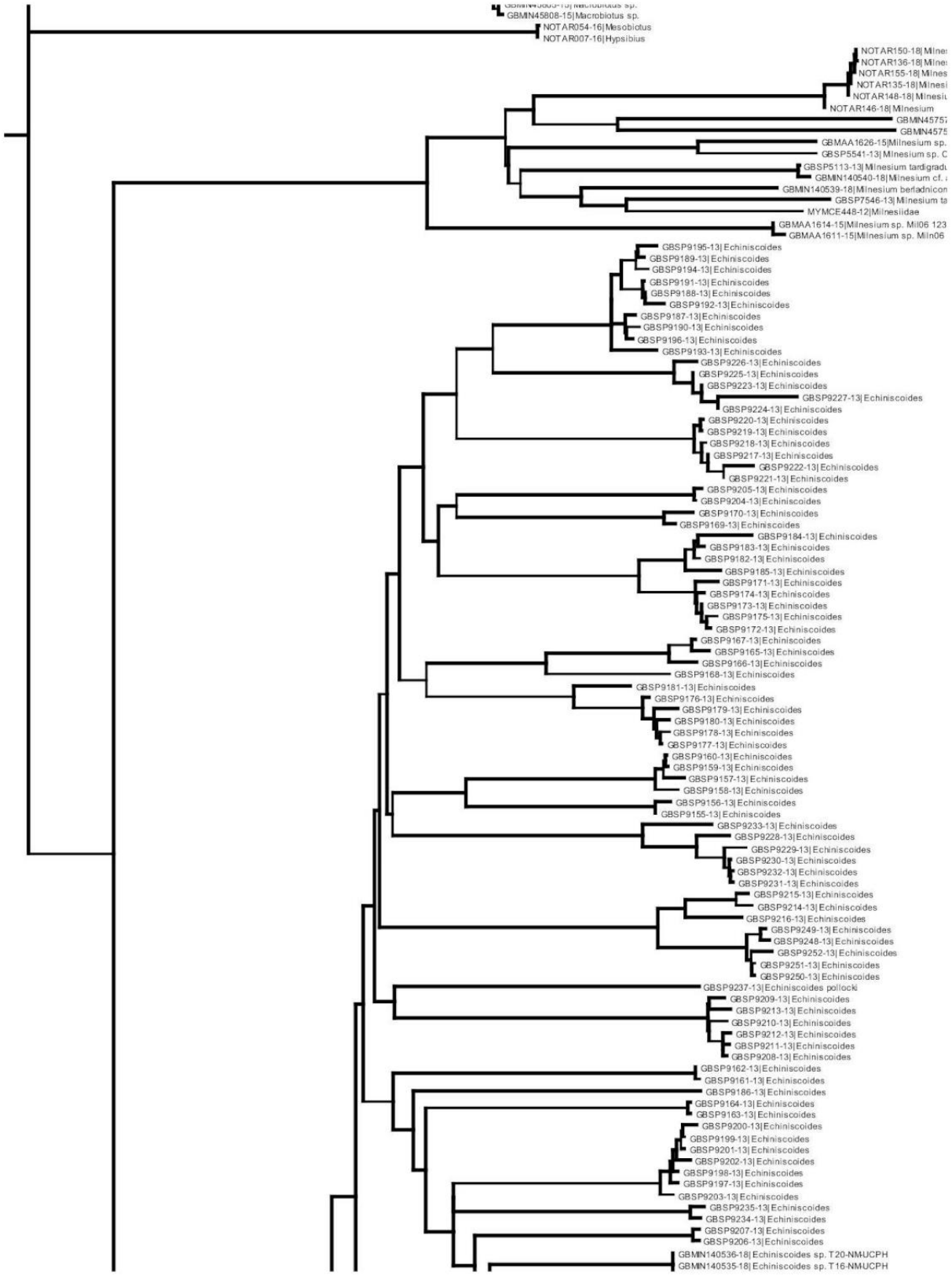
1. Sort tardigrades in water, place single specimen on temporary slides for initial identification and photography in microscope.
2. Transfer tardigrade by pipette into pcr-tube.
3. Add 70 µl QuickExtract.
4. Vortex well and spin down.
5. Incubate in room temperature for 2 hrs.
6. Incubate in 65 degrees C for 15 min (pcr-machine), vortex every 5 min.
7. Spin down.
8. Incubate in 98 degrees C for 2 min.
9. Pipet 60 µl extract into new, sterile pcr tube; carefully only from top to avoid exuviae at bottom -> store at -20 degrees C.
10. Add ca. 70 µl $H_2O$ to tube with exuviae and mix well with pipette to wash skin.
11. Transfer water with exuviae to petri dish -> search for skin and mount on microscope slide in Hoyer.

## Appendix F. Tardigrade phylogeny based on the mtDNA COI gene




GBSP10172-13| Ramazzottius cf. oberhaeuseri
GBSP10173-13| Ramazzottius cf. oberhaeuseri
GBSP0597-06| Macrobiotus richtersi
GBSP5150-13| Paramacrobiotus
TABAR006-08| Macrobiotus richtersi
TABAR001-08| Macrobiotus richtersi
GBSP7721-13| Paramacrobiotus richtersi
GBSP7720-13| Paramacrobiotus richtersi
GBSP0596-06| Macrobiotus richtersi
GBSP7722-13| Macrobiotus pallari
GBMN45810-15| Paramacrobiotus richtersi
GBMN45809-15| Paramacrobiotus richtersi
GBMN45812-15| Paramacrobiotus richtersi
GBMN45813-15| Paramacrobiotus richtersi
GBMN45815-15| Paramacrobiotus richtersi
GBMN45814-15| Paramacrobiotus richtersi
GBMN45811-15| Paramacrobiotus richtersi
GBSP5149-13| Paramacrobiotus
GBSP5148-13| Paramacrobiotus
GBSP8053-13| Minibiotus
GBSP7726-13| Minibiotus gumersindoi
NOTAR073-16| Minibiotus sp. 1TM
NOTAR071-16| Minibiotus sp. 1TM
NOTAR076-16| Minibiotus sp. 1TM
NOTAR074-16| Minibiotus sp. 1TM
NOTAR075-16| Minibiotus sp. 1TM
NOTAR072-16| Minibiotus sp. 1TM
NOTAR003-16| Minibiotus
NOTAR001-16| Minibiotus
NOTAR039-16| Minibiotus
GBSP7988-13|Tardigrada
NOTAR110-18| Minibiotus intermedius
NOTAR108-18| Minibiotus intermedius
NOTAR129-18| Minibiotus intermedius
NOTAR099-18| Minibiotus intermedius
NOTAR090-18| Minibiotus intermedius
NOTAR128-18| Minibiotus intermedius
NOTAR124-18| Minibiotus intermedius
NOTAR107-18| Minibiotus intermedius
NOTAR094-18| Minibiotus intermedius
GBSP5569-13| Dactylobiotus sp. CJS-2007a
GBSP5571-13| Dactylobiotus sp. CJS-2007a
GBSP5570-13| Dactylobiotus sp. CJS-2007a
GBSP5565-13| Dactylobiotus sp. CJS-2007a
GBSP5567-13| Dactylobiotus sp. CJS-2007a
GBSP5566-13| Dactylobiotus sp. CJS-2007a
GBSP5568-13| Dactylobiotus sp. CJS-2007a
GBSP0589-06| Dactylobiotus parthenogeneticus
GBSP0590-06| Murrayon pullari
NOTAR130-18| Murrayon dianeae
NOTAR088-18| Murrayon dianeae
GBSP7728-13| Murrayon dianeae
GBGCA10501-15| Parachaela
GBSP8050-13| Richtersius
GBSP5116-13| Macrobiotidae
GBSP10167-13| Richtersius cf. coronifer
GBSP5114-13| Richtersius coronifer
GBSP10169-13| Richtersius cf. coronifer
GBSP10168-13| Richtersius cf. coronifer
GBSP0598-06| Richtersius coronifer
GBSP0599-06| Richtersius coronifer
GBSP5132-13| Richtersius coronifer
GBMN140547-18| Richtersius cf. coronifer
GBMN140543-18| Richtersius cf. coronifer
GBMN140542-18| Richtersius cf. coronifer
GBSP5115-13| Richtersius coronifer
GBMN140545-18| Richtersius cf. coronifer
GBMN140544-18| Richtersius cf. coronifer
GBMN140546-18| Richtersius cf. coronifer
GBMN140541-18| Richtersius cf. coronifer
GBSP7979-13| Macrobiotus hufelandi
GBSP7980-13| Macrobiotus
GBSP7727-13| Minibiotus furcatus
GBSP7989-13| Macrobiotus hufelandi
NOTAR077-16| Macrobiotus
GBMAA1273-14|Macrobiotus kristenseni
GBMAA1275-14|Macrobiotus kristenseni
GBMAA1274-14|Macrobiotus kristenseni
TABAR036-13| Macrobiotus sandrae
TABAR035-13| Macrobiotus sandrae
TABAR034-13| Macrobiotus sandrae
TABAR033-13| Macrobiotus sandrae
TABAR037-13| Macrobiotus sandrae
TABAR032-13| Macrobiotus sandrae
GBSP0593-06| Macrobiotus terminalis
NOTAR053-16| Macrobiotus
NOTAR051-16| Macrobiotus
NOTAR052-16| Macrobiotus
NOTAR050-16| Macrobiotus
NOTAR049-16| Macrobiotus
NOTAR021-16| Mesobiotus harmsworthi
GBSP8051-13| Macrobiotus cf. hufelandi
GBSP0595-06| Xerobiotus pseudohufelandi
GBSP0594-06| Xerobiotus pseudohufelandi
GBSP7986-13| Macrobiotus
GBSP7978-13| Macrobiotus hufelandi
GBSP7981-13| Macrobiotus
TABAR025-13| Macrobiotus terminalis
TABAR024-13| Macrobiotus terminalis
TABAR023-13| Macrobiotus terminalis
TABAR022-13| Macrobiotus terminalis
TABAR026-13| Macrobiotus terminalis
NOTAR078-16| Macrobiotus vladimiri
GBSP7725-13| Macrobiotus hufelandi group sp. NG-2008
GBSP7724-13| Macrobiotus hufelandi group sp. NG-2008
NOTAR083-16| Macrobiotus vladimiri
NOTAR082-16| Macrobiotus vladimiri
TABAR031-13| Macrobiotus vladimiri
TABAR030-13| Macrobiotus vladimiri


TABAR029-13| Macrobiotus vladimiri
TABAR028-13| Macrobiotus vladimiri
TABAR027-13| Macrobiotus vladimiri
GBSP7723-13| Macrobiotus hufelandi group sp. NG-2008
NOTAR084-16| Macrobiotus vladimiri
NOTAR080-16| Macrobiotus vladimiri
NOTAR079-16| Macrobiotus vladimiri
GBSP0591-06| Macrobiotus sp. hufelandi group AG-2004
TABAR011-09| Macrobiotus macrocalix
TABAR009-09| Macrobiotus macrocalix
TABAR008-09| Macrobiotus macrocalix
TABAR010-09| Macrobiotus macrocalix
TABAR007-09| Macrobiotus macrocalix
GBSP0592-06| Macrobiotus sp. hufelandi group AG-2004
TABAR063-13| Macrobiotus macrocalix
TABAR064-13| Macrobiotus macrocalix
TABAR021-09| Macrobiotus macrocalix
TABAR020-09| Macrobiotus macrocalix
TABAR019-09| Macrobiotus macrocalix
TABAR018-09| Macrobiotus macrocalix
TABAR017-09| Macrobiotus macrocalix
GBSP8055-13| Macrobiotus
NOTAR174-18| Macrobiotus
NOTAR145-18| Macrobiotus
NOTAR104-18| Macrobiotus
NOTAR086-18| Macrobiotus
GBSP7987-13| Macrobiotus
GBSP7976-13| Macrobiotus hufelandi
GBSP7977-13| Diphascon
GBSP7985-13| Macrobiotus
GBSP8059-13| Macrobiotus cf. hufelandi
GBGCA10510-15| Parachaela
TABAR055-13| Macrobiotus cf. hufelandi
TABAR054-13| Macrobiotus cf. hufelandi
TABAR056-13| Macrobiotus cf. hufelandi
TABAR057-13| Macrobiotus cf. hufelandi
TABAR058-13| Macrobiotus cf. hufelandi
TABAR060-13| Macrobiotus cf. hufelandi
TABAR053-13| Macrobiotus cf. hufelandi
TABAR059-13| Macrobiotus cf. hufelandi
NOTAR035-16| Macrobiotus hufelandi
TABAR061-13| Macrobiotus cf. hufelandi
GBSP8057-13| Minibiotus
MYMCC525-11| Eutardigrada
NOTAR139-18| Macrobiotus
TABAR049-13| Macrobiotus hufelandi
TABAR050-13| Macrobiotus hufelandi
NOTAR070-16| Macrobiotus hufelandi
NOTAR092-18| Macrobiotus
NOTAR100-18| Macrobiotus
TABAR052-13| Macrobiotus hufelandi
TABAR051-13| Macrobiotus hufelandi
NOTAR013-16|Adropion prosirostre
NOTAR098-18|Adropion prosirostre
NOTAR132-18|Adropion prosirostre
NOTAR106-18|Adropion prosirostre
NOTAR175-18| Isohypsibius sattleri
NOTAR160-18| Isohypsibius sattleri
NOTAR154-18| Isohypsibius sattleri
GBMTG1924-16|Thulinius sp. DVL-2010
GBMIN140548-18| Thulinius ruffoi
GBSP7733-13| Eremobiotus alicatai
GBSP7732-13| Isohypsibius sp. Tar195
GBSP5556-13| Isohypsibius asper
GBSP5542-13| Isohypsibius asper
NOTAR023-16| Mesobiotus harmsworthi
NOTAR022-16| Mesobiotus harmsworthi
NOTAR027-16| Mesobiotus harmsworthi
NOTAR020-16| Mesobiotus harmsworthi
NOTAR024-16| Mesobiotus
GBSP8060-13| Macrobiotus furciger
GBSP8058-13| Macrobiotus furciger
GBMIN45791-15| Macrobiotus sp.
GBSP8052-13| Macrobiotus furciger
GBGCA10512-15| Parachaela
GBMIN45796-15| Macrobiotus sp.
GBMIN45795-15| Macrobiotus sp.
GBMIN45794-15| Macrobiotus sp.
GBMIN45799-15| Macrobiotus sp.
GBMIN45798-15| Macrobiotus sp.
GBMIN45797-15| Macrobiotus sp.
GBMIN45793-15| Macrobiotus sp.
GBMIN45792-15| Macrobiotus sp.
GBMIN45800-15| Macrobiotus sp.
GBMIN45801-15| Macrobiotus sp.
GBSP8056-13| Macrobiotus
GBSP7937-13| Macrobiotus
GBSP7933-13| Macrobiotus
GBSP7934-13| Macrobiotus
GBSP7935-13| Macrobiotus
GBSP7936-13| Macrobiotus
GBSP7932-13| Macrobiotus
GBMIN45803-15| Macrobiotus sp.
GBMIN45804-15| Macrobiotus sp.
GBMIN45802-15| Macrobiotus sp.
GBSP7931-13| Macrobiotus
GBSP7902-13| Macrobiotus
GBSP7895-13| Macrobiotus
GBSP7898-13| Macrobiotus
GBSP7897-13| Macrobiotus
GBSP7899-13| Macrobiotus
GBSP7896-13| Macrobiotus
GBSP7900-13| Macrobiotus
GBSP7903-13| Macrobiotus
GBSP7901-13| Macrobiotus
GBSP7871-13|Tardigrada
GBMIN45806-15| Macrobiotus sp.
GBSP7870-13|Tardigrada
GBMIN45807-15| Macrobiotus sp.


GBMIN45808-15|Macrobiotus sp.
NOTAR054-16|Mesobiotus
NOTAR007-16|Hypsibius
NOTAR150-18|Mlne
NOTAR136-18|Mlne
NOTAR155-18|Mlnes
NOTAR135-18|Mlnesi
NOTAR148-18|Mlnesiu
NOTAR146-18|Mlnesium
GBMIN4575
GBMIN4575
GBMAA1626-15|Mlnesium sp.
GBSP5541-13|Mlnesium sp. C
GBSP5113-13|Mlnesium tardigradu
GBMIN140540-18|Mlnesium cf.
GBMIN140539-18|Mlnesium berladnicon
GBSP7546-13|Mlnesium ta
MYMCE448-12|Mlnesiidae
GBMAA1614-15|Mlnesium sp. Mil06 123
GBMAA1611-15|Mlnesium sp. Mln06
GBSP9195-13|Echiniscoides
GBSP9189-13|Echiniscoides
GBSP9194-13|Echiniscoides
GBSP9191-13|Echiniscoides
GBSP9188-13|Echiniscoides
GBSP9192-13|Echiniscoides
GBSP9187-13|Echiniscoides
GBSP9190-13|Echiniscoides
GBSP9196-13|Echiniscoides
GBSP9193-13|Echiniscoides
GBSP9226-13|Echiniscoides
GBSP9225-13|Echiniscoides
GBSP9223-13|Echiniscoides
GBSP9227-13|Echiniscoides
GBSP9224-13|Echiniscoides
GBSP9220-13|Echiniscoides
GBSP9219-13|Echiniscoides
GBSP9218-13|Echiniscoides
GBSP9217-13|Echiniscoides
GBSP9222-13|Echiniscoides
GBSP9221-13|Echiniscoides
GBSP9205-13|Echiniscoides
GBSP9204-13|Echiniscoides
GBSP9170-13|Echiniscoides
GBSP9169-13|Echiniscoides
GBSP9184-13|Echiniscoides
GBSP9183-13|Echiniscoides
GBSP9182-13|Echiniscoides
GBSP9185-13|Echiniscoides
GBSP9171-13|Echiniscoides
GBSP9174-13|Echiniscoides
GBSP9173-13|Echiniscoides
GBSP9175-13|Echiniscoides
GBSP9172-13|Echiniscoides
GBSP9167-13|Echiniscoides
GBSP9165-13|Echiniscoides
GBSP9166-13|Echiniscoides
GBSP9168-13|Echiniscoides
GBSP9181-13|Echiniscoides
GBSP9176-13|Echiniscoides
GBSP9179-13|Echiniscoides
GBSP9180-13|Echiniscoides
GBSP9178-13|Echiniscoides
GBSP9177-13|Echiniscoides
GBSP9160-13|Echiniscoides
GBSP9159-13|Echiniscoides
GBSP9157-13|Echiniscoides
GBSP9158-13|Echiniscoides
GBSP9156-13|Echiniscoides
GBSP9155-13|Echiniscoides
GBSP9233-13|Echiniscoides
GBSP9228-13|Echiniscoides
GBSP9229-13|Echiniscoides
GBSP9230-13|Echiniscoides
GBSP9232-13|Echiniscoides
GBSP9231-13|Echiniscoides
GBSP9215-13|Echiniscoides
GBSP9214-13|Echiniscoides
GBSP9216-13|Echiniscoides
GBSP9249-13|Echiniscoides
GBSP9248-13|Echiniscoides
GBSP9252-13|Echiniscoides
GBSP9251-13|Echiniscoides
GBSP9250-13|Echiniscoides
GBSP9237-13|Echiniscoides pollocki
GBSP9209-13|Echiniscoides
GBSP9213-13|Echiniscoides
GBSP9210-13|Echiniscoides
GBSP9212-13|Echiniscoides
GBSP9211-13|Echiniscoides
GBSP9208-13|Echiniscoides
GBSP9162-13|Echiniscoides
GBSP9161-13|Echiniscoides
GBSP9186-13|Echiniscoides
GBSP9164-13|Echiniscoides
GBSP9163-13|Echiniscoides
GBSP9200-13|Echiniscoides
GBSP9199-13|Echiniscoides
GBSP9201-13|Echiniscoides
GBSP9202-13|Echiniscoides
GBSP9198-13|Echiniscoides
GBSP9197-13|Echiniscoides
GBSP9203-13|Echiniscoides
GBSP9235-13|Echiniscoides
GBSP9234-13|Echiniscoides
GBSP9207-13|Echiniscoides
GBSP9206-13|Echiniscoides
GBMIN140536-18|Echiniscoides sp. T20-NMUCPH
GBMIN140535-18|Echiniscoides sp. T16-NMUCPH

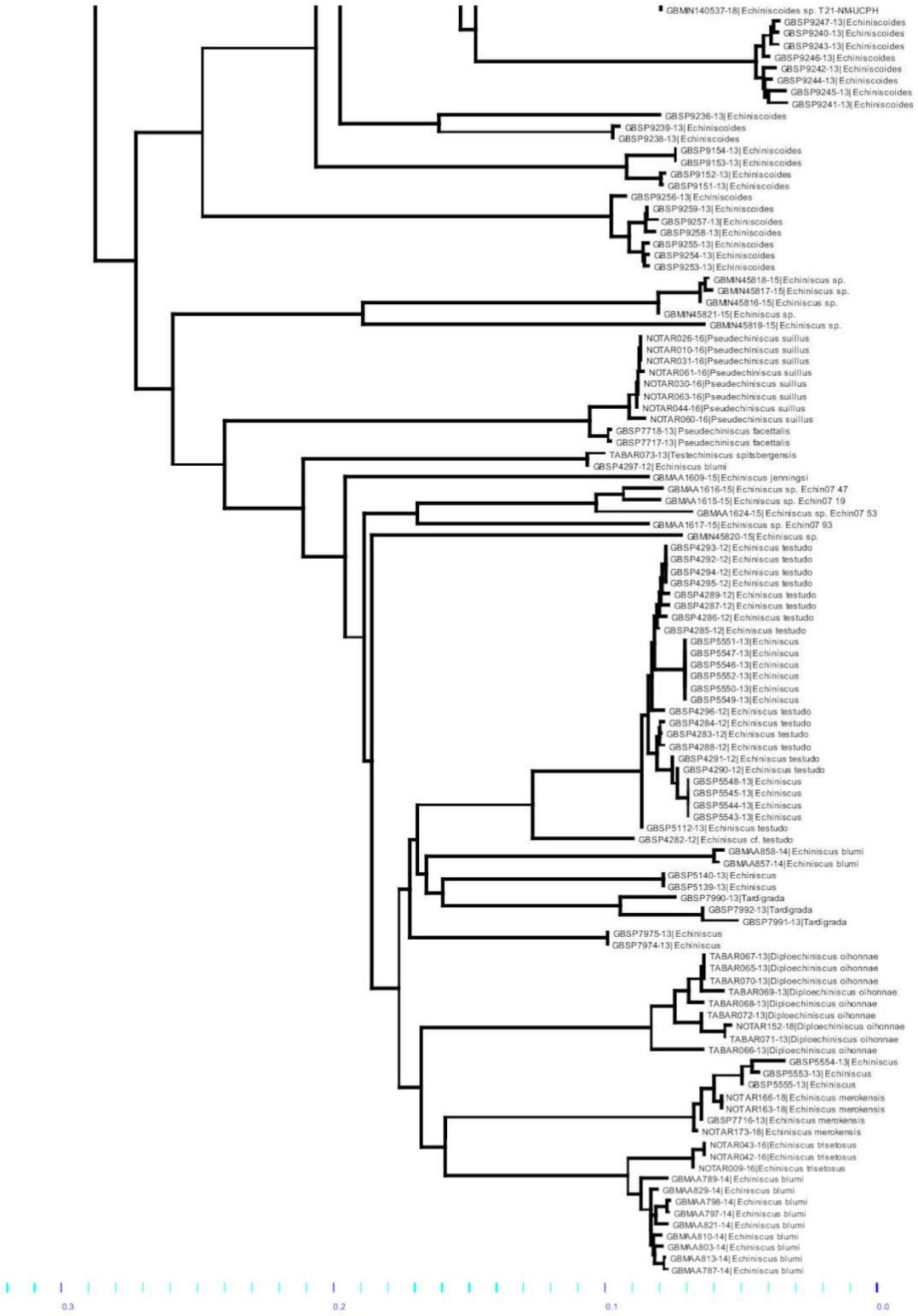



Figure F1. Neighbor joining phylogeny of the terrestrial tardigrade species and the genetic distance between them based on the COI mtDNA gene.