Molecular identification of Collembola and their fungal associates

STEN ANSLAN

Dissertationes Biologicae Universitatis Tartuensis 326
STEN ANSLAN

Molecular identification of Collembola and their fungal associates
Dissertation was accepted for the commencement of the degree of Doctor philosophiae in Botany and Mycology at the University of Tartu on August 31st, 2017 by the Scientific Council of the Institute of Ecology and Earth Sciences, University of Tartu.

Supervisor: Dr. Leho Tedersoo, University of Tartu, Estonia

Opponent: Dr. Andy F.S. Taylor, The James Hutton Institute, Craigiebuckler, Aberdeen, Scotland, UK

Commencement: Room 1019, 14a Ravila Street, Tartu, on 24 November 2017 at 10.15 a.m.

Publication of this thesis is granted by the Institute of Ecology and Earth Sciences, University of Tartu.

ISSN 1024-6479
ISBN 978-9949-77-578-1 (print)
ISBN 978-9949-77-579-8 (PDF)

Copyright: Sten Anslan, 2017

Tartu University Press
www.tyk.ee
CONTENTS

LIST OF ORIGINAL PUBLICATIONS ....................................................... 6
INTRODUCTION.......................................................................................... 7
MATERIAL AND METHODS ..................................................................... 10
   Sampling sites and study design ........................................................... 10
   Molecular analysis ............................................................................... 13
   Data analysis ....................................................................................... 15
RESULTS AND DISCUSSION ............................................................... 17
CONCLUSIONS ....................................................................................... 21
SUMMARY ............................................................................................... 22
SUMMARY IN ESTONIAN ....................................................................... 24
ACKNOWLEDGEMENTS ......................................................................... 26
REFERENCES .......................................................................................... 27
PUBLICATIONS ....................................................................................... 35
CURRICULUM VITAE ............................................................................. 105
ELULOOKIRJELDUS .............................................................................. 107
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications that are referred in the text by their Roman numerals:


Author’s contribution to each publication:

I. performed sampling, molecular analysis and statistics, wrote the manuscript with contributions from a coauthor.

II. participated in conceiving the study idea and design, performed sampling, molecular analysis and statistics, wrote the manuscript with contributions from coauthors.

III. participated in conceiving the study idea, programmed the software with contributions from coauthors, designed the software, performed software evaluation, wrote the manuscript with contributions from coauthors.

IV. participated in conceiving the study idea and design, performed sampling, molecular analysis and statistics, wrote the manuscript with contributions from coauthors.
INTRODUCTION

Soil harbors a vast diversity of microbiota and fauna, therefore it has been referred to as the ‘poor man's tropical rainforest’ (Giller, 1996). Microbial communities in this diverse habitat play a fundamental role in several processes including soil formation and microclimate regulation (van der Heijden et al., 2008; Clemmensen et al., 2013). These processes are often mediated by soil fauna (Lavelle et al., 2006; Crowther et al., 2015). Most soil animals interact with microbes primarily through direct predator-prey interactions, which drives microbial abundance and composition (Crowther et al., 2013). Simultaneously, microbial parasites and biocontrol agents may also regulate faunal activity (Siddiqui & Mahmood, 1996; Klironomos & Hart, 2001). Moreover, as soil invertebrates are frequently in contact with fungal spores and bacteria, they may further shape microbial communities by dispersing viable propagules.

Collembola (springtails) are probably the most abundant hexapods on Earth that inhabit all continents, including Antarctica. They are particularly common members of soil communities, where their densities may reach up to 60,000 specimens per m² (Hopkin, 1997). These wingless hexapods are usually 0.2–5 mm in length and classified as members of soil mesofauna. Approximately 8000 species of Collembola have been described, but geographical undersampling and increasing evidence from molecular studies suggests much higher diversity (Emerson et al., 2011). The majority of springtails feed on fungi or decaying plant material (Hopkin, 1997). Their interactions with fungi may alter fungal communities, which result in changes in decomposition (Addison et al., 2003), nutrient cycling (McGonigle, 1995) and plant growth (Seres et al., 2007; Ngosong et al., 2014). Several studies have demonstrated, that besides active feeding on microbes, Collembola alter the microbiota by ecto- and endozoochory (Visser et al., 1987; Thimm et al., 1998; Dromph, 2003; Lilleskov & Bruns, 2005; Seres et al., 2007; Buse et al., 2014; Nakano et al., 2017).

A few grams of soil may potentially harbor hundreds of species of fungi, invertebrates, protists and prokaryotes (Gans et al., 2005; Tedersoo et al., 2016). Identification of these small soil organisms is hampered by the poor resolution in microscopic morphological characters. The sequence-based molecular identification tool, DNA barcoding, has enabled to overcome this short-fall. DNA barcoding of selected genes has become a popular and cost-efficient method that enables to assign taxonomic identity over all ontogenetic stages and sexes of any organisms, improving precision in determining species richness from local to global scales. Combined with high-throughput sequencing (HTS) methods, DNA barcoding provides large amounts of inventory data that enables to identify whole communities and record changes in biodiversity over space and time. This so-called metabarcoding approach (Taberlet et al., 2012) is commonly used to identify selected groups of micro- and macro-organisms based on a single or several taxonomic markers in combination. Information about the distribution of organisms is often used in environmental (Tedersoo et
al., 2014; Wilson et al., 2016) and palaeoecological surveys (Epp et al., 2012; Willerslev et al., 2014; Capo et al., 2016), monitoring diseases (Lohan et al., 2016), species invasions (Cicconardi et al., 2017), forensics (Pechal et al., 2014), etc.

Traditionally used DNA markers include the Cytochrome c Oxidase subunit I (COI), rRNA internal transcribed spacer (ITS) and rRNA 18S gene for animals, fungi and protists, respectively (Hebert et al., 2003; Schoch et al., 2012; Pawlowski et al., 2016). Accordingly, soil animal and microbial communities are typically studied separately using different markers and/or taxonomic group-specific primers. Studying soil animals and microbes simultaneously would improve our understanding about their relationships in food webs of ecosystems and boost data acquisition in biodiversity research. However, finding suitable high-coverage primers for particular environmental applications is one of the major challenges of metabarcoding (Coissac et al., 2012). To detect several eukaryotic groups from environmental samples by performing a single PCR reaction, the 18S rRNA gene lacks species-level resolution in most eukaryote groups (Bik et al., 2012; Tang et al., 2012), whereas the COI marker is not entirely adoptable for environmental applications (Ficetola et al., 2010; Deagle et al., 2014; Horton et al., 2017) due to the lack of highly conserved primers sites. Similarly to COI, the ITS region, particularly ITS2 subregion has a high differentiation success rate among a wide range of eukaryotes (Yao et al., 2010). Especially mycologists have long used ITS region for low taxonomic level phylogenetics, barcoding and metabarcoding approaches. The length of the ITS2 marker is sufficiently short to allow amplification of slightly degraded DNA (Yao et al., 2010) as the remains or traces of animals can be detected (Coissac et al., 2012). Thus, use of the ITS markers would potentially enable to identify a wide range of eukaryotes simultaneously from environmental samples. For the first step, there is a need for validating the desirable barcoding region and primers and compiling the appropriate reference database for metabarcoding.

The simultaneous study of soil faunal and microbial communities could be also conducted by identifying the specific animal-associated microbes. Although several studies have addressed the spatial structure of soil invertebrate-microbe associations (e.g. Visser et al., 1987; Varga et al., 2002; Greif & Currah, 2007; Buse et al., 2014), only a few have examined the temporal aspect of these interactions. The habitat of soil faunal communities, in particular litter layer, is exposed to fluctuating environmental variables such as temperature and moisture, which affects the composition of microbial communities (Rousk & Bååth, 2011). Thus, the structure of soil food webs is not fixed (Hassall et al., 2006), especially across temporal scales. Seasonal fluctuations in the fungal particle consumption by epigeic collembolans have been already noted by Anderson & Healey (1972). Using stable isotope analysis, Potapov et al. (2014) demonstrated the seasonal variation in collembolans’ diet. These studies, however, have been performed within a single year, which disallows to distinguish seasonality from random temporal change.

8
So far, the invertebrate-microbe association studies have been mostly restricted to microscopic identification of spores or cultures or indirect proxies such as stable isotopes. Visual analysis of small particles and propagules often lacks sufficient resolution for accurate differentiation between fungal species, whereas culture-based methods may underestimate the diversity as many mutualistic and parasitic fungi are very difficult to culture or require specific culturing techniques. Molecular approaches such as HTS enable to overcome biases associated with paucity of morphological characters or unculturability. This method has thus become increasingly popular for identification of animal host-associated microbial communities (Clements et al., 2014; Poulsen et al., 2014; Yun et al., 2014).

The development of high-throughput molecular identification methods has greatly improved our understanding about microbial communities. However, taking advantage of its benefits requires a bioinformatics expertise and careful consideration of the bioinformatics workflow before the application (Majaneva et al., 2015), especially when new methods, protocols or markers are used. Although there are multiple pipelines for processing HTS data (e.g. Schloss et al., 2009; Caporaso et al., 2010; Edgar, 2010), most of these have been optimized for the prokaryote 16S rRNA gene. Applying the tools that are developed for processing metabarcoding data from prokaryotes is not straightforward for eukaryote metabarcoding data, especially that of the ITS region (Gweon et al., 2015). Thus, the lack of bioinformatics expertise may pose a serious bottleneck for metabarcoding studies, which outlines the necessity of appropriate bioinformatics skills or user-friendly tools with graphical view and flexible bioinformatics options.

Because the ITS2 marker has a great potential for simultaneous use in metabarcoding surveys of multiple taxa, one of the aims of this thesis was to evaluate its relative performance for barcoding one of the most abundant soil faunal group – the Collembola (I). Furthermore, this thesis addresses the spatial and temporal structure of Collembola-associated fungal communities as based on culturing and HTS (II, IV). To simplify the HTS analyses, one of the objectives was compilation of a user-friendly and flexible program for bioinformatics analysis of custom high-throughput amplicon sequencing data (III). In this thesis, the following hypotheses were postulated:

1) the ITS2 marker provides sufficient resolution for species delimitation of Collembola (I);
2) HTS outperforms culturing in recovering fungal taxonomic groups associated with gut contents and body surfaces of Collembola (II);
3) fungal communities on the body surface of collembolans differ by season (II, IV);
4) the seasonal changes of richness and structure of the Collembola-associated fungal communities are consistent across years (IV);
5) fungal communities in gut content samples of Collembola differ between species due to dietary preferences (II, IV).
MATERIAL AND METHODS

Sampling sites and study design

For the Collembola barcoding study (I), we collected litter and soil samples (~500 cm³) throughout the vegetation periods of 2012 and 2013 from different sites in Estonia (Figure 1; specific locations are described in study I). Individuals of Collembola were extracted using a Tullgren funnel method and stored in 96% ethanol for morphological and genetic analyses. Specimens were morphologically identified according to Fjellberg (1998; 2007).

Figure 1. Red dots denote sampling sites of Collembola in Estonia, study I.

The studies addressing the spatial and temporal structure of Collembola-associated fungal communities (II, IV) were carried out in two circular 2500 m² forested plots in Estonia (Ilmatsalu and Kardla, Tähtvere Comm.). The Ilmatsalu sampling site was comprised of only Norway spruce (Picea abies (L.) H. Karst). The Kardla sampling site included ca. 5% of silver birch (Betula pendula Roth) in addition to spruce. These studies included three epigeic Collembola species – Entomobrya nivalis Linnaeus 1758, Orchesella flavescens Bourlet 1839 and Pogonognathellus longicornis Tullberg 1871 – that are abundant on the forest floor of mixed and coniferous forests in North Europe. These species are considered as epigeic microorganism (and plant material) consumers, but consume a lower proportion of organic detritus as compared to endogeic species (Potapov et al., 2016). Specimens were collected throughout the vegetation period from May to September in 2013–2015 (study II included
year 2013, study IV included 2013–2015). At least four specimens (max. 9 specimens) of each Collembola species were collected during each sampling period with a single-animal aspirator (Figure 2). The specimens were instantly treated with chloroform to prevent their excretion of gut contents and prepared for molecular (II, IV) and culturing (II) analysis on the day of collection. In addition, study IV included 22 soil samples in order to characterize the local fungal communities from both sampling sites during each collection period (except June 2013). Soil samples were collected and processed following Tedersoo et al. (2014).

Figure 2. Collection aspirator with Eppendorf tube for a single Collembola individual. The inflow tube (transparent tube) was cleaned with ethanol (96%) after suction of each specimen.

The bioinformatics platform, PipeCraft, with graphical user interface for analyzing the HTS data was built using Gambas (3.8.4), Python (2.7) and bash programming language on Ubuntu 14.04 (III). PipeCraft has been built on the Docker container (www.docker.com) in which all the software and dependencies have already been installed and configured to process raw data produced by Illumina, Ion Torrent, Roche 454 and Pacific Biosciences sequencing platforms. It links a number of third-party applications (Table 1) in which the users may choose the most suitable options for their specific needs. PipeCraft is publicly available through PlutoF system (https://plutof.ut.ee/#/datacite/10.15156%2FBIO%2F587450) without the need of creating user account.
Table 1. List of third-party applications that are incorporated in PipeCraft, study III.

<table>
<thead>
<tr>
<th>Analysis step</th>
<th>Program</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create circular consensus</td>
<td>pbccs</td>
<td>2.0.2</td>
<td>github.com/PacificBiosciences/unanimity</td>
</tr>
<tr>
<td>sequences</td>
<td>bax2bam</td>
<td>0.08</td>
<td>github.com/PacificBiosciences/pitchfork</td>
</tr>
<tr>
<td>Assemble paired-end sequences</td>
<td>PANDAseq</td>
<td>2.10</td>
<td>Masella <em>et al.</em> (2012)</td>
</tr>
<tr>
<td></td>
<td>FLASH</td>
<td>1.2.11</td>
<td>Magoc &amp; Salzberg (2011)</td>
</tr>
<tr>
<td></td>
<td>vsearch</td>
<td>1.11.1</td>
<td>Rognes <em>et al.</em> (2016)</td>
</tr>
<tr>
<td>Quality filtering</td>
<td>mothur</td>
<td>1.36.1</td>
<td>Schloss <em>et al.</em> (2009)</td>
</tr>
<tr>
<td></td>
<td>OBItools</td>
<td>1.2.9</td>
<td>Boyer <em>et al.</em> (2016)</td>
</tr>
<tr>
<td></td>
<td>vsearch</td>
<td>1.11.1</td>
<td>Rognes <em>et al.</em> (2016)</td>
</tr>
<tr>
<td>Demultiplexing</td>
<td>OBItools</td>
<td>1.2.9</td>
<td>Boyer et al. 2016</td>
</tr>
<tr>
<td></td>
<td>mothur</td>
<td>1.36.1</td>
<td>Schloss <em>et al.</em> (2009)</td>
</tr>
<tr>
<td>Chimera filtering</td>
<td>vsearch</td>
<td>1.11.1</td>
<td>Rognes <em>et al.</em> (2016)</td>
</tr>
<tr>
<td>Gene extraction</td>
<td>ITS extractor</td>
<td>1.0.11</td>
<td>Bengtsson-Palme <em>et al.</em> (2013)</td>
</tr>
<tr>
<td></td>
<td>V-Xtractor</td>
<td>2.1</td>
<td>Hartmann <em>et al.</em> (2010)</td>
</tr>
<tr>
<td></td>
<td>Metaxa2</td>
<td>2.1</td>
<td>Bengtsson-Palme <em>et al.</em> (2016)</td>
</tr>
<tr>
<td></td>
<td>mothur (cut</td>
<td>1.36.1</td>
<td>Schloss <em>et al.</em> (2009)</td>
</tr>
<tr>
<td></td>
<td>sequences)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td>CD-HIT</td>
<td>4.6</td>
<td>Fu <em>et al.</em> (2012)</td>
</tr>
<tr>
<td></td>
<td>vsearch</td>
<td>1.11.1</td>
<td>Rognes <em>et al.</em> (2016)</td>
</tr>
<tr>
<td></td>
<td>swarm</td>
<td>2.1.8</td>
<td>Mahé <em>et al.</em> (2015)</td>
</tr>
<tr>
<td></td>
<td>mothur</td>
<td>1.36.1</td>
<td>Schloss <em>et al.</em> (2009)</td>
</tr>
<tr>
<td></td>
<td>usearch</td>
<td>8.1.1861</td>
<td>Edgar (2010)</td>
</tr>
<tr>
<td>Taxonomy assignment</td>
<td>BLAST+</td>
<td>2.2.28+</td>
<td>Camacho <em>et al.</em> (2009)</td>
</tr>
<tr>
<td></td>
<td>mothur</td>
<td>1.36.1</td>
<td>Schloss <em>et al.</em> (2009)</td>
</tr>
</tbody>
</table>
Molecular analysis

To test the relative performance of ITS2 subregion for identification purposes of Collembola (I), the genomic DNA of collembolans was extracted in a lysis buffer (0.8 M Tris-HCl, 0.2 M (NH₄)₂SO₄, 0.2% w/v Tween-20; Solis BioDyne, Tartu, Estonia) using a proteinase K method. Specimens were incubated (100 ml of lysis buffer and 2.5 ml of proteinase K) at 56 °C for 24 h and at 98 °C for 15 min. To validate the usefulness of ITS marker for barcoding Collembola, the study I included also COI and large subunit of rDNA (LSU) markers, which have proven efficient to separate among certain species of Collembola (Hogg & Hebert, 2004; Porco et al., 2012). Primers used for amplification of ITS, COI and LSU regions are listed in Table 2. Detailed descriptions of the PCR conditions and sequencing are given in study I.

The analysis of fungal communities associated with three epigeic Collembola species were carried out separately for body surface and gut content (II, IV). The body surface samples included the extracted appendages (legs, antennae, body hairs or scales). For the analysis of gut contents, the same specimen without appendages was surface sterilized with 1.5% sodium hypochlorite (NaOCl) for 15 sec and subsequently rinsed in sterilized water for 1 min. Surface sterilization of the body degraded most of the body hairs and scales and thus minimized potential cross-contamination between the body surface and gut content. Further, the head capsule was removed and gut content extracted. The specimens were prepared in a biosafety cabinet with laminar flow. For culturing purposes in study II, these samples were placed into 1.5 ml Eppendorf tubes filled with 50 ml sterile water. Tubes were vortexed for 10 sec, followed by pipetting the liquid onto Petri dishes with Modified Melin-Norkrans (MMN) culture medium (Marx, 1969; 2% agar, 0.1% glucose, penicillin 200 mg/ml, kanamycin 20 mg/ml) and dispersing it with a spatula. The Petri dishes were stored at 20 °C for 30 days. For identification purposes, culturable fungi were subjected to genomic DNA extraction in a lysis buffer using the proteinase K method as described above. The ITS2 region was amplified using the primers 5.8SF and ITS4 (Table 2). For the direct molecular analysis in studies II and IV, the appendages and gut contents of each Collembola specimen were subjected to DNA extraction by use of Power-Soil®DNA Isolation Kit (MoBio, Carlsbad, CA, USA) following manufacturer's instructions. The samples were prepared as described above. The fungal ITS2 region was targeted using primers gITS7 and ITS4ngs (Table 2), the latter being tagged with 10–11 base unique identifiers (specified in study IV). Thermocycler conditions and sequencing are described in study II.
Table 2. Primers used in studies I, II and IV. Reg. denotes amplification region.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Primer name</th>
<th>Sequence 5’ – 3’</th>
<th>Reference</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COI</strong></td>
<td>LCO1490</td>
<td>GGTCAACAAATCATAAAGA TATTGG</td>
<td>Folmer et al. (1994)</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>HCO2198</td>
<td>TAAACTTCAGGGTGACCAA AAAATCA</td>
<td>Folmer et al. (1994)</td>
<td>I</td>
</tr>
<tr>
<td><strong>ITS</strong></td>
<td>ITS5</td>
<td>GGAAGTAAAAGTCGTAACA AGG</td>
<td>White et al. (1990)</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>ITS4</td>
<td>TCCTCCGCTTATGATATGC</td>
<td>White et al. (1990)</td>
<td>I, II</td>
</tr>
<tr>
<td></td>
<td>5.8SF</td>
<td>ATGCATCGATGAAGACGC</td>
<td>Martin &amp; Rygiewicz (2005)</td>
<td>I, II</td>
</tr>
<tr>
<td></td>
<td>ITS7-Coll</td>
<td>GTGAACCTGCAAGACATG</td>
<td>Anslan &amp; Tedersoo (2015)</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>ITS4-Coll</td>
<td>GCTAAAATTTAGCGGTAATC</td>
<td>Anslan &amp; Tedersoo (2015)</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>gITS7</td>
<td>GTGARTCATCGARTCTTTG</td>
<td>Ihrmark et al. (2012)</td>
<td>II, IV</td>
</tr>
<tr>
<td></td>
<td>ITS4ngs</td>
<td>TCCTSCGCTTATGATGC</td>
<td>Tedersoo et al. (2014)</td>
<td>II, IV</td>
</tr>
<tr>
<td><strong>LSU</strong></td>
<td>CTB6</td>
<td>GCATATCAAAGCGGAGG</td>
<td>unite.ut.ee/primers.php</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>TW13</td>
<td>GGTCGGTGTTCAGGACG</td>
<td>T.J. White, unpublished</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>C1’</td>
<td>ACCCGCTGAATTTAGCAT</td>
<td>D’Haese (2002)</td>
<td>I</td>
</tr>
</tbody>
</table>
Data analysis

Quality check and manual trimming of sequences obtained from Collembola specimens (I) and cultured fungi (II) were performed using Sequencher 5.1 (Gene Codes Corporation, Ann Arbor, USA). Bioinformatics for the ITS data in study II and IV were performed using the early unpublished version of PipeCraft and PipeCraft v1.0 software (III), respectively. The detailed options are specified in papers II and IV. After the bioinformatics processing, operational taxonomic units (OTUs) were separated based on 97% sequence similarity threshold and these were further checked and filtered based on BLASTn search values, positive and negative controls to remove putative contaminants, non-fungal OTUs, potential artefacts and tag-switching errors (Nguyen et al., 2015; Schnell et al., 2015). OTUs with 75, 80, 85, 90, 95 and 97% similarity were considered to represent phylum, class, order, family, genus and species level, respectively (Tedersoo et al., 2014). Based on taxonomic assignments, we parsed fungal OTUs into trophic guilds using FUNGuild (Nguyen et al., 2016).

The Collembola DNA sequences in study I were aligned using MAFFT 7 (Katoh et al., 2002) and edited in Seaview 4.4.2 (Gouy et al., 2010). The ITS sequences were trimmed to include only the ITS2 subregion without the flanking genes by using the ITS2 database annotation tool (Keller et al., 2009). LSU sequences were manually trimmed to include only the D1–D2 variable domains. The sequences of COI were trimmed to 606 bp because of the poor quality of 5’ and 3’ ends of some COI sequences. Intraspecific and interspecific sequence divergence was calculated based on Levenshtein distance (raw distance, equally weighing indels and substitutions) and Kimura two-parameter distance (K2P) using usearch 7.0.1090 (Edgar, 2010) and MEGA6 (Tamura et al., 2013), respectively. Neighbor-Joining trees of COI, LSU and ITS2 were conducted to illustrate clustering of Collembola species (conducted in MEGA6).

To minimize the effect of inconsistencies in sequencing depth across samples in HTS studies, a subsampling procedure was performed to the depth of 1100 sequences using mothur (II; Schloss et al., 2009) or the number of sequences (log-transformed) per sample was included in the analysis as a covariate (IV). The differences in recovering fungal richness between HTS and culturing was tested using Mann-Whitney U test (II). The effect of season (fixed factor with levels May, June, August and September), year (fixed factor with levels 2013, 2014, 2015), Collembola species (fixed factor with levels E. nivalis, O. flavescens, P. longicornis) and site (random factor with levels Ilmatsalu and Kardla) on log-transformed fungal OTU richness were tested using factorial ANOVA (II) or ANCOVA (IV) followed by Tukey HSD tests. To address the relative importance of these factors on fungal community structure, PERMANOVA analysis (Anderson, 2005) was performed on Bray-Curtis similarity matrix based on Hellinger transformed data. Because PERMANOVA is sensitive to differences in multivariate dispersion among
groups (Clarke & Gorley, 2006), PERMDISP was used to test whether there are differences in the within-group multivariate dispersion among groups (IV). The analyses were performed using PRIMER6 software (Clarke & Gorley, 2006). Additionally, Mantel tests were performed to determine whether the dissimilarity of fungal community composition (FCC) was related to temporal distance using R v.3.2.2 (IV; R-Core-Team, 2015). OTUs that occurred only once across the dataset were excluded. Additionally, in study IV, for the analysis of gut content samples, OTUs that occurred in less than 5% of the gut samples were removed to account for potential environmental contamination in detecting the diets by molecular means (de la Cadena et al., 2017). Outliers were screened and removed based on non-metric multidimensional scaling (NMDS). After removing non-significant terms with the smallest mean square, only significant factors were included in the final model. Statistical significance was considered at $\alpha=0.05$. Monte Carlo P levels ($P_{MC}$) were used in case of low permutation levels (<100).
RESULTS AND DISCUSSION

The Collembola barcoding study (I) included 33 species from 10 families. For COI, LSU and ITS2 markers, the study respectively included 162, 154 and 162 high-quality sequences (range: 2–13 specimens per species) to determine barcoding gaps. The intraspecific variability of collembolan COI sequences was in the range of 0–10.7% and the barcoding gap lied between 10.7 and 13.9% (Levenshtein distance). As an exception, interspecific variability of *Isotoma viridis* (Bourlet, 1839) and *Isotoma riparia* (Nicolet, 1842) remained in the boundaries of intraspecific variability (10.1–10.7%). There was no intraspecific variability in LSU sequences, whereas interspecific variability ranged from 0.7% to 28.7%. The intraspecific ITS2 sequence variability of Collembola specimens was in the range of 0.0–3.7%. All ITS2 interspecific differences were greater than 6% (6.4% to ca. 60%) and thus the barcoding gap ranged between 3.7 and 6.4% sequence dissimilarity for these Estonian specimens of selected species. Thus, all tested markers proved to be useful for species identification purposes. I anticipate, however, that a distinct gap may represent artefacts of insufficient sampling across taxa and geographical scales (Bergsten *et al.*, 2012).

As the ITS2 marker has been used for metabarcoding studies for a long time and it has a high species-level discrimination power across eukaryotes (Yao *et al.*, 2010), it could serve as a potential marker for targeting several eukaryotic groups simultaneously from environmental samples. The study I revealed that ITS2 marker is a viable alternative to COI and LSU for identification purposes of Collembola species, thus we used its application in a metabarcoding study to detect several eukaryotic groups from soil samples (Tedersoo *et al.*, 2016). The latter study utilized the information about the barcoding gap (dissimilarity threshold, 6%) as established in study I. Compared with the 3% dissimilarity threshold, the number of Collembola OTUs were reduced 3-fold and the environmental effects in tree diversity experiments on Collembola richness and composition became more strongly supported. Moreover, we found that soil C/N ratio was the strongest predictor of taxonomic richness of Collembola OTUs, with a positive effect (Tedersoo *et al.*, 2016), indicating that Collembola spp. tend to favor soils rich in humus and degraded litter. Globally, higher C/N ratio corresponds to greater fungal biomass (Fierer *et al.*, 2009). This suggests that the higher resource (fungal abundance) availability may drive richness and abundance of the predominately fungivorous Collembola (Wardle, 2002). These trophic relationships outline the importance of studying soil animals and microbes simultaneously to be able to learn more about their relationships in the context of food chain and boost data acquisition in biodiversity research.

To further characterize Collembola-fungal associations, we studied fungal communities associated with common epigeic Collembola species on the forest floor (II, IV). Collembolans’ appendages and gut samples were plated on solid agar medium in Petri dishes, which resulted in successful fungal propagule
Germination in 48 (32.0%) and 34 (22.7%) dishes (150 in total), respectively (II). Of the 84 total spore germination events, 46 fungal OTUs were detected based on ITS2 sequence similarity threshold of 97%. In contrast to culturing, the HTS data of Collembola appendages and gut samples detected 1277 fungal OTUs (97% similarity threshold of ITS2 sequences). As expected, the HTS method revealed much greater fungal richness compared with culturing (II). However, the Collembola-associated fungal OTU accumulation curves exhibited no signs of reaching an asymptote, indicating potentially much higher richness. Consequently, our following study (IV) with more Collembola samples in two extra years detected over 2800 Collembola-associated fungal OTUs. Both of these HTS studies (II, IV) revealed that the largest part of the Collembola-associated fungal OTUs was assigned to Ascomycota, followed by Basidiomycota. Amongst identifiable trophic groups, Collembola were mostly associated with saprotrophic fungi, with a few taxa belonging to ectomycorrhizal fungi, plant pathogens and entomopathogens.

Studies II and IV revealed that the richness and structure of fungal communities associated with Collembola body surface exhibited weak seasonal variation. However, these seasonal changes were largely unpredictable across years due to some annual effects and a significant year × season interaction term (IV). Top layers of the forest floor are exposed to environmental variables such as temperature and moisture that fluctuate nocturnally and across seasons. Thus, the uppermost soil layers exhibit highest temporal variability of fungal communities (Voríšková et al., 2014; Bahram et al., 2015). The studied epigeic Collembola species dwell mostly in the litter layer, thus, these patterns suggest that their fungal associations may represent the available mycota rather than seasonality-driven preference per se. Besides the fungal community changes arising from seasonal fluctuations, we found that Collembola-associated fungal community composition and richness on body surface as well as in guts exhibited annual changes (IV). Several studies have demonstrated that the fungal community composition may exhibit annual turnovers (Izzo et al., 2005; Haňáčková et al., 2015; Matsuoka et al., 2016; Mundra et al., 2016). Investigating the fungal succession in the litter of Picea abies over several years, Haňáčková et al. (2015) found that time was the most significant factor affecting the structure of the fungal community, i.e. fungal community composition exhibited significant temporal distance decay of similarity. Accordingly, the soil samples from the immediate environment of collected collembolas in our study (IV) exhibited significant temporal distance decay of similarity, which was also evident considering Collembola-associated fungal communities. Although we cannot determine the underlying mechanisms, these successional changes may reflect annual fluctuation of environmental conditions such as moisture and temperature that may alter the competitive balance among soil inhabiting fungal species (Rousk & Bååth, 2011). The temporal distance decay of collembolan-associated fungi probably reflects shifted feeding habits related to the succession of the litter fungal community. These results emphasize the
importance of considering temporal scales in determining the diets of soil invertebrates that dwell in habitat with fluctuating resource availability.

Despite seemingly irregular foraging behavior, collembolans are reactive to the quality of their environment and may exhibit specific preferences for food sources (Varga et al., 2002; Jorgensen et al., 2003; Nakamori & Suzuki, 2005; Men'ko et al., 2006; Rotheray et al., 2011; Chauvat et al., 2014). We found that the soil samples from collembolans’ habitat were characterized by similar proportion of saprotrophic and mycorrhizal fungi, whereas considerably more saprotrophic than mycorrhizal fungi were detected from the gut samples (IV). This suggests the collembolans’ preference for saprotrophs over mycorrhizal fungi, which has been also proposed by Malmstrom & Persson (2011) and Potapov & Tiunov (2016). However, we did not detect fungal dietary differences among the studied Collembola species (II, IV). Collembolans are able to switch their food sources (Chahartaghi et al., 2005; Rotheray et al., 2009; Krab et al., 2013) and generally benefit from mixed diet (Scheu & Folger, 2004). Moreover, microbivorous soil animals live in close contact with their food resources and simultaneous consumption of various microorganisms may be simply inevitable (Scheu & Folger, 2004). Thus, diet specialization is often difficult to detect in natural conditions of heterogeneous habitats with fluctuation of available resources, especially among species that dwell in the same microhabitats, thus sharing same resources. The inability to detect diet partitioning with our HTS analysis does not exclude the possibility that they have species-specific impact on the associating fungal communities (Tordoff et al., 2008; Fujii et al., 2016) by promoting or suppressing certain taxa through dispersal and feeding.

Although collembolans use fungi as a main food source, their antagonistic interactions may also provide certain benefits to the grazed fungi. In particular, their trophic relationships may contribute to the dispersal of viable fungal propagules by means of ecto- and endozoochory (Visser et al., 1987; Klironomos & Moutoglis, 1999; Dromph, 2001; Seres et al., 2007; Nakano et al., 2017). The culturing treatment in study II revealed that 32.0% of body surface and 22.7% of gut content samples resulted in successful fungal propagule germination of mostly saprotrophic taxa. The absence of collembolans may result in significant reduction in litter colonization frequencies of saprobic fungi (Lussenhop, 1992), suggesting the contribution of collembolans in dispersing propagules, thus promoting fungal colonization. Because of their feeding preference, collembolans seem to mainly associate with saprobic fungi (Malmstrom & Persson, 2011; Potapov & Tiunov, 2016), yet our studies detected various mycorrhizal, especially ectomycorrhizal fungi on body surface and in guts of Collembola (II, IV). Although fungal spores can be wind-dispersed over long distances (Horton, 2017), a high concentration of the ectomycorrhizal fungal spores falls within near proximity of the fruit body (Li, 2005; Galante et al., 2011). The basidiospores of many ectomycorrhizal fungal species exhibit high degree of ornamentation, which could aid spore dispersal via attaching to arthropods (Halbwachs et al., 2015). Therefore, vectors such as soil invertebrates may be important for reaching into suitable microhabitats, where compatible host is
available (Lilleskov & Bruns, 2005). Because Collembola lack external cavities (mycangia) and internal sacs (sporothecae) to carry microbial propagules, their role in dispersal on body surfaces can be regarded as passive. Nevertheless, the role of these microarthropods in promoting or suppressing microbial colonization and diversity through assisted dispersal and feeding may have a great importance for ecosystem processes through shaping the microbial communities (Crowther et al., 2011), providing inoculum for decomposition (Visser et al., 1987; Wardle, 2006) and root colonization by mycorrhizal fungi (Klironomos & Moutoglis, 1999; Lilleskov & Bruns, 2005; Seres et al., 2007).

The HTS data for studies II and IV were analyzed using PipeCraft platform, which represents a wide compilation of open source bioinformatics software and custom scripts, and analysis over a user-friendly graphical interface (III; Figure 3). This platform, which converts data formats and links the analytical procedures of several public tools enabled fast and efficient HTS data analysis. In PipeCraft, users are able to customize the pipeline by selecting the most suitable tools and options to process raw data from Illumina sequencing platform as well as from Ion Torrent and Roche 454 second-generation sequencing platforms. The latest version includes scripts and routines to run both raw and fastq-formatted output of the third-generation sequencing platform Pacific Biosciences (PacBio). Because the software and all the dependencies are compiled in a single downloadable image, users are free of struggle to set up multiple tools for bioinformatics analyses. Using PipeCraft analysis pipeline, our team recently published one of the first methodological studies about the use of PacBio sequencing for identification of fungi and other eukaryotes based on various rRNA genes and their combinations with fragment length of 250–2500 bp (Tedersoo et al., 2017). As a scientific novelty, such long barcodes enable combining variable and conserved regions of markers to provide both accurate species delimitation and taxonomic placement of divergent sequences.

Figure 3. Illustration of the graphical interface of PipeCraft.
CONCLUSIONS

The following main conclusions and working hypothesis can be inferred from my thesis:

- The ITS2 barcoding marker provides sufficient resolution for discriminating among Collembola species (I); thus, ITS2 may serve as an alternative barcode to COI and LSU markers to identify Collembola. The advantage of this alternative barcode is its potential to be used in metabarcoding analyses for identifying several eukaryotic groups to species level.

- High-throughput sequencing (HTS) of collembolans’ appendages and guts recovered considerably more fungal OTUs compared with the culturing method (II). HTS data demonstrated that collembolans are associated with much higher diversity of fungi than previously anticipated (II, IV).

- Fungal communities on the body surface and in guts of epigeic, litter layer dweller collembolans exhibited seasonal (II, IV) and annual variation (IV). However, the seasonal changes were unpredictable across years (IV).

- We did not detect any fungal dietary differences among the studied epigeic Collembola species, which suggests that they possess relatively opportunistic feeding behavior (II, IV).

- With the use of our compiled high-throughput amplicon sequencing data analysis platform, PipeCraft (III), the bioinformatics processes were efficient and fast for the analysis of fungal ITS2 sequences in soil and Collembola-associated samples.
Microbial and faunal communities are highly diverse in soils where they play fundamental roles in several ecosystem processes. The identification of small soil organisms has benefited from the development of molecular methods that enable identification of single species to whole communities. However, the members of soil communities (e.g. animals and microbes) are typically studied separately due to different expertise, research interests and methods. Studying soil animals and microbes simultaneously would improve our understanding about their relationships in food webs of ecosystems and boost data acquisition in biodiversity research. However, finding suitable high-coverage primers for simultaneous identification various eukaryotes from environmental samples is one of the major challenges of metabarcoding. The simultaneous study of soil faunal and microbial communities has been conducted by identifying the specific animal associated microbes. However, the majority of studies have been mainly restricted to microscopy and culturing techniques, which have biases associated with paucity of morphological characters or unculturability. Moreover, we know little about the temporal aspects of soil faunal-microbial associations since most of the studies have addressed only the spatial structure of these associations. In this thesis, I examined the usefulness of the rDNA ITS2 subregion for identification purposes of Collembola, because of its potential for simultaneous use in metabarcoding surveys of multiple taxa. Furthermore, this thesis addresses the spatial and temporal structure of Collembola-associated fungal communities as based on culturing and high-throughput sequencing (HTS). To simplify the HTS data analyses, one of the objectives of this thesis was the compilation of a user-friendly and flexible platform for bioinformatics analysis of custom high-throughput amplicon sequencing data. In ecological case studies, the following hypothesis were postulated: 1) the ITS2 marker provides sufficient resolution for species delimitation of Collembola; 2) HTS outperforms culturing in recovering fungal taxonomic groups associated with gut content and body surface of Collembola; 3) fungal communities on the body surface of collembolans differ by season; 4) the seasonal changes of richness and structure of the Collembola-associated fungal communities are consistent across years; 5) fungal communities in gut content samples of Collembola differ between species due to dietary preferences.

The main results and conclusions are the following: 1) the ITS2 barcoding marker provides sufficient resolution for discriminating among Collembola species; 2) HTS outperformed the culturing method in terms of recovering Collembola-associated fungal species, and it revealed that collembolans are associated with much higher diversity of fungi than previously anticipated; 3) the Collembola-associated fungal richness and community structure exhibited significant variation in different temporal scales, which probably reflects the succession of the litter fungal community; 4) diet specialization among the
studied epigeic Collembola species was not evident, suggesting that these arthropods possess relatively opportunistic feeding behavior; 5) the compiled high-throughput amplicon sequencing data analysis platform enabled efficient bioinformatics workflow for the analysis of fungal ITS2 amplicons in soil and Collembola-associated samples.
Doktoritöö peamised tulemused ja järelused on järgmised: 1) ITS2 lõik omab piisavat liikidevahelist erinevust, et eristada hooghännaliste liike; 2) hooghännalistega seotud seeneliikide tuvastamine on tõhusam mass-sekve neerimise meetodiga, mis tõi esile, et hooghännalised on seotud palju rohke mate seeneliikidega kui seni tuvastatud; 3) tulenevalt seenekoosluste suhtes sionist on hooghännalistega seotud seenekoosluste struktuur ja liigirikkus mõjutatud nii sesoonist kui aastast; 4) töös kasutatud hooghännaliste liikide vahel ei tuvastatud toitumiseelulistse Osas; 5) koostatud HTS andmete töötlemise programm võimaldab kiiret ja tõhusat DNA järjestuste töötlust.
ACKNOWLEDGEMENTS

During my studies, especially when it started on collembolans, my mother has reminded me several times how I was always more fascinated who crawls under the loose bark of trees and in forest litter than picking chanterelles, while mushrooming with my family. I am very grateful to my supervisor, Leho Tedersoo, who has guided me throughout the university studies so far and thanks to whom the childhood fascination observing the “hidden world” of small animals has got a scientific output in combination with more interest in “chanterelles” as well. I thank Mohammad Bahram who has been always open to share his knowledge, especially introducing me to the bioinformatics, which boosted my further development in that area. I thank Urmas Kõljalg for the overall support and the enthusiasm towards biodiversity informatics, which has aided the management of biodiversity information. I am thankful to my colleagues for all the assistance and support: Kessy Abarenkov, Rasmus Puusepp, Ingrid Liiv, Sergei Põlme, Indrek Hiiesalu, Eveli Otsing, Helery Harend, Heidi Tamm, Jane Oja, Kadri Põldmaa, Irma Zettur and many others! I thank Toomas Tammaru for the feedback and suggestions to improve the text of this thesis. Special thanks to my family for the all-time encouragement and to my dearest Kertu, who has accepted my long days at work and no summer vacation during the preparation of this thesis.

Financial support was provided by Estonian Science Foundation grants (9286, PUT0171, PUT1317, PUT1399, EMP265), IUT20-30, MOBERC, ECOL-CHANGE and FIBIR.
REFERENCES


29


33
CURRICULUM VITAE

Name: Sten Anslan
Date of Birth: 25.03.1987
Citizenship: Estonian
Address: Institute of Ecology and Earth Sciences, 14a Ravila St, 50411, Tartu, Estonia
Phone: +372 5837 2084
E-mail: sten.anslan@gmail.com

Education:
1994–2006  Väike-Maarja Gymnasium
2006–2011  University of Tartu, B.Sc. student, Ecology and Biodiversity Conservation
2011–2013  University of Tartu, M.Sc. student, Botany and Mycology
2013 –  University of Tartu, Ph.D. student

Publications:


ELULOOKIRJELDUS

Nimi: Sten Anslan
Sünniaeg: 25.03.1987
Kodakondsus: Eesti
Aadress: Ökoloogia ja Maateaduste Instituut, 14a Ravila St, 50411, Tartu, Eesti
Telefon: +372 5837 2084
E-mail: sten.anslan@gmail.com

Hariduskäik:
1994–2006 Väike-Maarja Gümnaasium
2006–2011 Tartu Ülikool, B.Sc, ökoloogia ja elustiku kaitse
2011–2013 Tartu Ülikool, M.Sc, bioloogia
2013 – Botaanika osakond, Ökoloogia ja Maateaduste Instituut, Tartu Ülikool, doktorant

Publikatsioonid:


52. **Ebe Sild.** Impact of increasing concentrations of O₃ and CO₂ on wheat, clover and pasture. Tartu, 1999, 123 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
77. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
80. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
87. **Aare Abro.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N₂ fixation in some Estonian lakes. Tartu, 2004, 111 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation \( \text{pheBA} \) operon in *Pseudomonas putida*. Tartu, 2005, 126 p.


108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.


111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.


113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn4652 into the upstream region of \( \text{pheBA} \) operon in *Pseudomonas putida*. Tartu, 2005, 106 p.


118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.


123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisiae*. Tartu, 2006. 113 p.


154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.


158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.


188. **Kersti Kangro.** The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.


119


258. **Villu Soon.** A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.


270. **Tiina Talve.** Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.

271. **Mehis Rohtla.** Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.


275. **Pratyush Kumar Das.** Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p.


301. **Kristiina Mark.** Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.


309. **Liis Leitsalu.** Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.


