

New approaches for investigating microbial food  
webs across ecosystems

**Inaugural-Dissertation**

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

**M. Sc. Jule Freudenthal**

aus Thedinghausen

Köln, 2025

---

---

**Berichterstatter:**

Prof. Dr. Michael Bonkowski, *Terrestrische Ökologie, Universität zu Köln*

PD Dr. Kathrin Lampert, *Molekulare Ökologie, Universität zu Köln*

Prof. Dr. Jens Boenigk, *Allgemeine Botanik, Universität Duisburg Essen*

**Vorsitz der Prüfung:**

Prof. Dr. Tim Mansfeldt, *Bodengeographie/Bodenkunde, Universität zu Köln*

**Beisitzer:**

PD Dr. Kenneth Dumack, *Terrestrische Ökologie, Universität zu Köln*

**Tag der mündlichen Prüfung:** 12.03.2025

Diese Arbeit wurde im Rahmen der DFG-geförderten Schwerpunktprojekts „**Taxon-omics**: New approaches to discovering and naming biodiversity (SPP 1991)“ durchgeführt.

---

## ABSTRACT

Microbial communities drive essential ecosystem processes such as organic matter decomposition and nutrient cycling. However, the complex dynamics of microbial food webs, including the role of trophic interactions and environmental factors, are not yet fully understood, as disentangling biotic and abiotic factors remains challenging.

This thesis explores microbial food webs across various ecosystems, ranging from wastewater treatment plants and maize rhizospheres to alpine soils and forest canopies. Using metatranscriptomics or metabarcoding with group-specific primers, the complex interplay of biotic and abiotic factors that drive microbial diversity, community composition, and function are uncovered. In addition, this thesis facilitates the exploration and integration of functional traits for protists by providing a functional trait database for Amoebozoa - a widespread and dominant protist group - as well as introducing workflows for the investigation and comparison of physiological traits of individual protist taxa based on *de novo* transcriptomes.

We show that microbial food webs are strongly shaped by predation. In wastewater treatment plants, predation by protists and microscopic metazoans facilitated the removal of parasites. In the maize rhizosphere, predation by protists drove prokaryote community turnover, along with plant immune responses, root zones, or the effects of root manipulations. However, biotic interactions are not limited to predation. On canopy bark surfaces, microbial community assembly was shaped by an interplay of biotic and abiotic factors, specifically by competition between bacteria and fungi, symbioses between algae and fungi, bark topology, and environmental conditions. Furthermore, abiotic factors partly influenced microbial communities indirectly through biotic interactions. For example, seasonal changes affected predator communities in alpine soils and wastewater treatment systems, which, in turn, shaped the prey communities through selective predation pressure. In addition, analyses of the functional traits of protists revealed: First, variations in Amoebozoa and Cercozoa communities across ecosystems affected not only the taxonomic composition but also the functional composition. Second, even the physiological traits of individual protist taxa, including closely related strains, exhibit remarkable variation.

Collectively, these findings highlight the central role of biotic interactions in structuring microbial communities and emphasize the advantages of functional traits and holistic, molecular-based approaches for studying microbial communities. The insights into the complexity of microbial food webs, combined with the established methodologies and tools, will allow future studies to deepen our understanding of the astonishing diversity of microorganisms – particularly of protists.

---

# TABLE OF CONTENT

<b>General Introduction .....</b>	<b>1</b>
The microbial food web.....	1
Exploring microbial food webs across ecosystems .....	2
Molecular approaches to comprehensively assess microbial communities.....	3
Trait-based approaches to explore the functional diversity of protists.....	4
Aims, hypothesis, and outline of the thesis .....	6
<b>Contributions to the Chapters of the Thesis .....</b>	<b>7</b>
<b>Chapter 1: Microeukaryotic gut parasites in wastewater treatment plants: diversity, activity, and removal .....</b>	<b>9</b>
<b>Chapter 2: Root cap is an important determinant of rhizosphere microbiome assembly.....</b>	<b>38</b>
<b>Chapter 3: Microeukaryotic predators shape the wastewater microbiome.....</b>	<b>79</b>
<b>Chapter 4: Biotic interactions explain seasonal dynamics of the alpine soil microbiome .....</b>	<b>98</b>
<b>Chapter 5: Algae-fungi symbioses and bacteria-fungi co-exclusion drive tree species-specific differences in canopy bark microbiomes .....</b>	<b>128</b>
<b>Chapter 6: A novel protistan trait database reveals functional redundancy and complementarity in terrestrial protists (Amoebozoa &amp; Rhizaria) .....</b>	<b>155</b>
<b>Chapter 7: Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics .....</b>	<b>169</b>
<b>Conclusive Summary and General Discussion .....</b>	<b>223</b>
Biotic and abiotic factors shaping the microbial food web across ecosystems .....	223
Unveiling the functional diversity of protists .....	225
<b>General References .....</b>	<b>227</b>
<b>Acknowledgement .....</b>	<b>233</b>
<b>Erklärung zur Dissertation .....</b>	<b>234</b>

---

# GENERAL INTRODUCTION

## The microbial food web

Microbial food webs are essential for ecosystem processes as they drive organic matter decomposition, nutrient cycling, and energy flow. Microbial food web models illustrate how biodiversity is structured across trophic levels, describe the energy and matter fluxes within the community, and facilitate exploring the relationships between community structure (e.g., biomass, stability, and composition) and ecosystem functions. Thus, studying microbial food webs is essential to improve our understanding of ecosystem processes and potentially predict ecological responses (Dormann et al. 2018).

Historically, microbial food web models divided nutrient flows into bacterial and fungal energy channels (Hunt et al., 1987; Moore and Hunt, 1988). de Ruiter et al. (1998, 1995, 1993) used microbial food web models based on bacterial and fungal energy channels to link energy flow to ecosystem stability, biodiversity, or nutrient cycling. Additionally, the impact of microbial predators such as collembola, mites, nematodes, and protists was analyzed, highlighting their importance by estimating, for example, that amoebae (protists) and bacterivorous nematodes account for approximately 70% of nitrogen mineralization (de Ruiter et al., 1993). However, these early microbial food web models relied on a rather small prey spectrum of predators, considering protists and bacterivorous nematodes as the primary consumers of the bacterial energy channel and microarthropods and mycophagous nematodes as the main consumers of the fungal energy channels.

Recent studies showed that the microbial food webs are more complex than previously described. Potapov et al. (2021) questioned the size-based trophic hierarchy, i.e., that energy is channeled from smaller to larger organisms, by showing that the trophic positions are often independent of body size, particularly in terrestrial ecosystems. Moreover, the functional diversity of microbial taxa, including their trophic positions, appears to be more complex than previously proposed. Accordingly, Johnke et al. (2017), Petters et al. (2021), and Groß et al. (2023) highlighted predatory bacteria as one of the main consumers of other bacteria, as well as yeasts and filamentous fungi. Furthermore, predatory fungi have been shown to consume amoebae (protists), including slime molds, and even nematodes and rotifers (microscopic Metazoa, Michel et al., 2014; Pajdak-Stós et al., 2016; Rahman et al., 2023). However, in particular, the high diversity of protists challenged microbial food web models based on bacterial and fungal energy channels (Geisen, 2016).

Protists (microbial eukaryotes) form an essential part of microbial food webs in virtually all ecosystems on Earth (Bonkowski et al., 2019; Keeling et al., 2014; Singer et al., 2021). They represent the majority

## General Introduction

---

of the eukaryotic diversity (Adl et al., 2019), forming a paraphyletic group that spans all supergroups of the eukaryotic tree of life (Burki et al., 2020). Protists exhibit tremendous morphological diversity and evolutionary divergence (Adl et al., 2019; Geisen et al., 2020). It has been confirmed that bacterivorous protists indeed strongly shape the prokaryotic community, but their influence on the prokaryotic community differs due to selective feeding (Flues et al., 2017; Gao et al., 2019; Glücksman et al., 2010; Rosenberg et al., 2009). However, the protist diversity exceeds bacterivorous protists and further includes autotrophs, mixotrophs, saprotrophs, eukaryvorous, omnivorous, parasites and pathogens (Bonkowski et al., 2019; Geisen et al., 2016). Phototrophic protists (primary producers) contribute substantially to carbon fixation in aquatic (Falkowski, 2002) and terrestrial systems (Jassey et al., 2022), forming the basis of microbial food webs (Baldrian, 2017; Bringel and Couée, 2015; Thapa and Prasanna, 2018). Conversely, some pathogenic protists, such as the Oomycota (e.g., *Phytophthora infestans*), exhibit worldwide ecological and economic impacts, as evidenced by the potato late blight in the 1840s, causing food shortages in Europe (Mizubuti and Fry, 2006). Among predatory protists, several studies revealed a wide feeding range comprising fungi, other protists, and microfauna (Dumack et al., 2019; Estermann et al., 2023; Geisen et al., 2015b, 2016; Hess and Suthaus, 2022), including taxa with highly specialized feeding strategies. For example, Hess and Suthaus (2022) revealed that Vampyrellids (Rhizaria) exhibit specialized feeding strategies, such as "protoplast feeders" that infiltrate or extract protoplasts. Estermann et al. (2023) showed that *Cryptodifflugia oviformis* (Amoebozoa) perforates fungal cell walls and extracts the cellular contents, whereas Geisen et al. (2015b) demonstrated that *Cryptodifflugia operculata* (Amoebozoa) preys on nematodes several-fold larger by pack hunting.

Overall, these studies exemplify that microbial food webs are far more complex than models based on bacterial and fungal energy channels suggest. However, the questions arise: How complex are microbial food webs; how can we explore trophic interactions, and how can we disentangle the influences of biotic and abiotic factors shaping microbial communities?

### Exploring microbial food webs across ecosystems

Microbial food web research has so far mostly focused on the effects of environmental factors such as soil type, host genotype, or geographical location (Arrigoni et al., 2018; Edwards et al., 2015; Laforest-Lapointe et al., 2016). Although environmental factors are essential for characterizing and predicting microbial food web dynamics, they only partially reflect their complexity.

Biotic interactions strongly influence microbial community composition but are much more difficult to assess. One way to investigate biotic interactions is through extensive experiments. For example, competition has been revealed as a major driver of the bacterial community composition (Romdhane

## General Introduction

---

et al., 2022). Further, predation has been demonstrated to limit competitive exclusion, i.e., the dominance of a few species that are better adapted to the environment, thereby increasing the diversity and evenness of the prey (Burian et al., 2022; Corno et al., 2008; Gauzens et al., 2016; Saleem et al., 2012), which has been linked to the resistance of ecosystems to environmental stress (Wittebolle et al., 2009).

Molecular methods facilitate complementary approaches for exploring microbial food webs. Co-occurrence network analysis, for example, allows extracting patterns from complex datasets and forming hypotheses about ecological relationships between microorganisms based on count data (Faust and Raes, 2012; Röttjers and Faust, 2018). In general, co-occurrence network inference assesses the co-occurrence pattern of taxa across multiple samples by quantifying the similarity of their distributions, generating a null model that represents the data without biotic interactions, evaluating the significance of the associations, and correcting for multiple testing (Röttjers and Faust, 2018). However, the inference of co-occurrence networks from molecular data faces several challenges related to the typical biases of count data derived from molecular sequencing studies, as well as confounding factors that arise when studying complex systems (Faust and Raes, 2012; Röttjers and Faust, 2018; Weiss et al., 2016). Accordingly, Weiss et al. (2016) showed that less than a third of all edges were shared between networks derived with different approaches. We can improve the robustness and validity of network analyses and reduce spurious edges (edges that do not reflect biotic relationships) by accounting for various confounding factors (Faust, 2021; Faust and Raes, 2012; Röttjers and Faust, 2018; Weiss et al., 2016): 1. the varying sequencing depth of samples causing varying counts of taxa across samples unrelated to their “true” abundance, 2. the sparsity (many zeros), as associations inferred from matching zeros should be avoided, given that zeros cannot only result from the actual absence of a taxon, 3. the compositionality, i.e., counts represent proportions rather than absolute “true” abundances, 4. the heterogeneity of the data, as increased heterogeneity frequently causes reduced precision. Consequently, the establishment of a workflow that accounts for all these potential pitfalls, in combination with the choice of network inference tool, is essential for robust network inference and thus, for meaningful insights into microbial food webs.

## Molecular approaches to comprehensively assess microbial communities

The innovation of high-throughput sequencing techniques opened new possibilities for investigating microbial food webs. Metabarcoding has become one of the most widely used sequencing methods for studying microbial diversity, frequently employed in combination with “universal eukaryotic primers” to target a wide range of eukaryotic microorganisms. However, there are important technical

## General Introduction

---

limitations to consider when studying in particular protist diversity using “universal eukaryotic primers”: First, a few protist taxa, such as the Alveolata, are overestimated whereas others such as the Amoebozoa, one of the prevalent protist taxa, are underestimated (Fiore-Donno et al., 2016; Geisen et al., 2015a; Lentendu et al., 2014). Thus, a biased representation of protist diversity is obtained. Second, the sequence data are often dominated by multicellular organisms, for example, animals and fungi (Baldwin et al., 2013; Dupont et al., 2016).

To overcome the limitations of metabarcoding using “universal eukaryotic primers”, either smaller group can be targeted with specific primers, facilitating high coverage and sequencing depth (Fiore-Donno et al., 2018; Jauss et al., 2020; Sapp et al., 2019). Alternatively, metatranscriptomics represents a promising approach for a comprehensive assessment of microbial food webs. Metatranscriptomics uses a PCR-free approach with random hexamers instead of fixed primers. Therefore, metatranscriptomics is not prone to primer-biased amplification errors (Bonkowski et al., 2019; Voss et al., 2019). Furthermore, metatranscriptomics only targets the metabolically active fraction of microbiomes (Geisen et al., 2015c) and also covers so-called “long-branch organisms”, i.e., species with highly divergent marker gene sequences, which are often missed by primer-based sequencing methods (Schuler et al., 2018). Despite these advantages, metatranscriptomics has rarely been applied to explore microbial food webs across ecosystems.

### Trait-based approaches to explore the functional diversity of protists

Trait-based approaches facilitate a deeper understanding of the functions and dynamics of microbial food webs and the abiotic and biotic factors shaping them, as species traits are directly influenced by selective pressures (Litchman et al., 2021). In aquatic ecosystems, Litchman et al. (2015) showed that distinct functional groups within phytoplankton influence biogeochemical cycles such as nitrogen, phosphorus, and silica. Consequently, changes in phytoplankton community structure might alter element cycling. In terrestrial ecosystems, Fiore-Donno et al. (2019) applied a recently compiled functional trait database for Cercozoa and Endomyxa (Rhizaria, protists, Dumack et al., 2019) to soil communities, revealing that environmental factors shape the functional community composition both spatially and seasonally. However, comprehensive trait-based studies of entire microbial food webs are restricted by reliably curated and easily accessible functional trait databases for protists, as well as the limited knowledge of the functional traits of individual protist species.

Transcriptomics can be a powerful tool for investigating the physiological traits of individual species (Gerbracht et al., 2022; Ribeiro et al., 2020), as well as the phylogenetic relationships between species (Tice et al., 2021), facilitating the exploration of the physiological traits of species in an evolutionary

## General Introduction

---

context. Molecular data, in general, have already greatly improved our understanding of the evolutionary relationships of protists. For example, phylogenetic analyses based on single to few genes provided valuable insights into the evolutionary relationships of the Cercozoa, a protist group with high morphological diversity that limits the identification of taxa based on morphology (Cavalier-Smith et al., 2018; Cavalier-Smith and Chao, 2003). However, phylogenies based on single to few genes often fail to resolve the deep nodes due to the limited phylogenetic signal (Cavalier-Smith et al., 2018, 2015). Multi-gene phylogenies based on transcriptomics, comprising hundreds of genes, have the potential to resolve these relationships (Tice et al., 2021).

Transcriptomics can also be used, as said, to examine the physiological traits of individual protist species. For example, Ribeiro et al. (2020) explored the metabolic flexibility of *Arcella intermedia* (Amoebozoa) based on *de novo* transcriptomes by reconstructing the carbohydrate and amino acid metabolism or analyzing the signaling complexity. Similarly, Gerbracht et al. (2022) used *de novo* transcriptomes of the algivorous *Orciraptor agilis* (Rhizaria) to identify the enzymes used for cell wall perforation. However, the transcriptome representation of protists is generally poor, especially for Rhizaria (Sibbald and Archibald, 2017), and workflows for comparative *de novo* transcriptomics across species without reference genomes are yet not well established.

## General Introduction

---

### Aims, hypothesis, and outline of the thesis

In this thesis, we aimed to establish novel approaches to explore microbial food webs across various ecosystems. In particular, we were interested in disentangling biotic and abiotic factors shaping the microbial communities. Additionally, we aimed to deepen the current knowledge of protists by exploring their phylogenetic relationships and functional diversity, even at strain level. We addressed three hypotheses:

1. Biotic factors strongly shape microbial food webs.
2. Biotic factors are as important as abiotic factors for microbial community assembly.
3. Protists exhibit high variability in their functional traits, even at the strain level. Thus, distinct groups are affected differently by biotic and abiotic factors.

To test these hypotheses, we first established workflows for exploring microbial food webs of different ecosystems using metatranscriptomics (Chapter 1) and for robust network inference that accounts for potential pitfalls associated with environmental sequencing data (Chapters 1 & 3), even for datasets derived from different sequencing approaches (Chapter 2). We explored biotic and abiotic factors shaping the microbial food webs of wastewater treatment plants (Chapters 1 & 3), the rhizosphere of maize (Chapter 2), alpine soils (Chapter 4), and canopy bark surfaces (Chapter 5). Additionally, we provided a functional trait database for Amoebozoa (Chapter 6), explored the phylogenetic relationships of Rhizaria using multi-gene phylogeny (Chapter 7), and established workflows for comparative *de novo* transcriptomics across species without reference genomes to explore the physiological traits of Thecofilosea (Chapter 7).

---

# CONTRIBUTIONS TO THE CHAPTERS OF THE THESIS

The contributions of the author of this thesis are described according to the Contributor Roles Taxonomy (CRediT, <https://credit.niso.org/>):

## Chapter 1:

**Freudenthal, J., Ju, F., Bürgmann, H., Dumack, K., 2022.** Microeukaryotic gut parasites in wastewater treatment plants: Diversity, activity, and removal. *Microbiome* 10, 27.

*Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing*

This publication is based on the master's thesis of Jule Freudenthal, submitted at the University of Cologne in 2020.

## Chapter 2:

Rüger, L., Ganther, M., **Freudenthal, J., Jansa, J., Heintz-Buschart, A., Tarkka, M.T., Bonkowski, M., 2023.** Root cap is an important determinant of rhizosphere microbiome assembly. *New Phytol* 239, 1434–1448.

*Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - review & editing*

## Chapter 3:

Heck, N., **Freudenthal, J., Dumack, K., 2023.** Microeukaryotic predators shape the wastewater microbiome. *Water Res* 242, 120293.

*Formal analysis, Investigation, Methodology, Validation, Writing - review & editing*

## Chapter 4:

Fiore-Donno, A.M., **Freudenthal, J., Dahl, M.B., Rixen, C., Urich, T., Bonkowski, M., 2024.** Biotic interactions explain seasonal dynamics of the alpine soil microbiome. *ISME Commun* 4, ycae028.

*Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - review & editing*

## Chapter 5:

**Freudenthal, J., Dumack, K., Schaffer, S., Schlegel, M., Bonkowski, M., 2024.** Algae-fungi symbioses and bacteria-fungi co-exclusion drive tree species-specific differences in canopy bark microbiomes. *ISME J*, wrae206. <https://doi.org/10.1093/ismejo/wrae206>. **Accepted manuscript.**

*Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing*

---

**Chapter 6:**

**Freudenthal, J., Schlegel, M., Bonkowski, M., Dumack, K., n.d.** A novel protistan trait database reveals functional redundancy and complementarity in terrestrial protists (Amoebozoa & Rhizaria). ***In preparation.***

*Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing*

**Chapter 7:**

**Freudenthal, J., Bonkowski, M., Engelmann, N.-J., Schlegel, M., Dumack, K., n.d.** Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics. ***In preparation.***

*Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing*

---

# CHAPTER 1

## Microeukaryotic gut parasites in wastewater treatment plants: diversity, activity, and removal

Jule Freudenthal<sup>1</sup>, Feng Ju<sup>2,3</sup>, Helmut Bürgmann<sup>4</sup> and Kenneth Dumack<sup>1</sup>

<sup>1</sup>Terrestrial Ecology, Institute of Zoology, University of Cologne, Zùlpicher Str. 47b, 50674, Köln, Germany

<sup>2</sup>Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, 310024, China

<sup>3</sup>Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, 310024, China

<sup>4</sup>Eawag, Swiss Federal Institute of Aquatic Science and Technology, 6047, Kastanienbaum, Switzerland

*Keywords: protists, parasite removal, water treatment, food web, metatranscriptomics, metagenomics*

### Abstract

#### **Background**

During wastewater treatment, the wastewater microbiome facilitates the degradation of organic matter, reduction of nutrients, and removal of gut parasites. While the latter function is essential to minimize public health risks, the range of parasites involved and how they are removed is still poorly understood.

#### **Results**

Using shotgun metagenomic (DNA) and metatranscriptomic (RNA) sequencing data from ten wastewater treatment plants in Switzerland, we were able to assess the entire wastewater microbiome, including the often neglected microeukaryotes (protists). In the latter group, we found a surprising richness and relative abundance of active parasites, particularly in the inflow. Using network analysis, we tracked these taxa across the various treatment compartments and linked their removal to trophic interactions.

#### **Conclusions**

Our results indicate that the combination of DNA and RNA data is essential for assessing the full spectrum of taxa present in wastewater. In particular, we shed light on an important but poorly understood function of wastewater treatment – parasite removal.

## 1 Background

The microbiome in wastewater treatment plants (WWTPs) includes not only prokaryotes but also eukaryotes: fungi, protists, and microscopic metazoans. Together, this wastewater community facilitates anaerobic denitrification and aerobic nitrification, as well as heterotrophic respiration and flocculation [1,2,3,4,5]. Specifically, the coupling of denitrification and nitrification reduces wastewater nitrogen, while aerobic and anaerobic heterotrophs help to degrade organic material [1, 4]. Microbial biomass growth leads to flocculation, enabling the separation of solids through sedimentation [2, 3].

Although the abovementioned functions are fulfilled by the concerted actions of the entire WWTP microbiome, the vast majority of WWTP surveys focus merely on the most numerous microbial entity in WWTPs – bacteria [6]. Protists, in particular, remain largely underappreciated, despite representing the majority of the eukaryotic fraction [7,8,9]. Although they are less numerous than bacteria, protists profoundly affect the community composition of their prokaryotic and eukaryotic prey [10,11,12]. In WWTPs, protists have been found to modulate the composition and biomass of the microbial community, thus affecting denitrification, nitrification, and flocculation, for instance by feeding on filamentous bacteria or loosely attached bacteria from flocs [2, 3, 13, 14].

In addition to their key role in regulating the WWTP microbial community, protists deserve more attention for another important reason: this microeukaryotic group includes many gut-associated taxa that are potentially harmful to humans and animals [12, 15]. The removal of these parasites, which include taxa such as *Giardia* and *Entamoeba*, is a key function of wastewater treatment [16,17,18,19]. However, little is known about the mechanisms involved, including the role of predation.

This knowledge gap is to a large degree due to the challenges involved in the taxonomic identification and enumeration of the main predators in wastewater – protists [1]. Apart from labor-intensive microscopy, primer-based metabarcoding is currently the most commonly used method to assess microbiomes in various environments, including sewage. However, this method is inevitably selective as there is no general primer that enables the assessment of all taxa present leading to contradictory results in protist assessments [20,21,22,23]. Another alternative is to use shotgun methods, which are primer-independent and thus suitable for assessing microbial communities in their entirety, including (parasitic) protists [24, 25]. Specifically, shotgun metagenomics (DNA-based) are used to determine microbial community composition and functional potential, while shotgun metatranscriptomics (RNA-based) provides a proxy for assessing microbial activity [26, 27]. So far, these promising methods have rarely been used to investigate microbial communities in sewage or WWTPs on a large scale, and if so, the data were not screened for protists.

The present study addresses this gap by analyzing a publicly available data set of shotgun metagenomic and metatranscriptomic data provided by Ju et al. [28], who sampled microbial communities in various

treatment compartments of WWTPs across Switzerland. This data set allowed us to assess the WWTP microbial community as a whole, including protists, and without a primer bias. Our specific objective was to identify protist taxa that are potential gut parasites, track their abundance and activity patterns across the consecutive WWTP compartments (from inflow to effluent), and screen for putative predator-prey interactions that could explain parasite removal during wastewater treatment.

## 2 Material and methods

We made use of the publicly available data sets from Ju et al. [28]. In brief, these authors sampled 12 WWTPs across Switzerland for DNA (shotgun metagenomics) and RNA (shotgun metatranscriptomics). At each facility, they sampled four compartments connected by continuous flow: sewage-inflow after screening and primary sedimentation (INF), denitrification bioreactor (DNF), nitrification bioreactor (NFC), and effluent after passing of the secondary clarifier (EFF). For details of the sampling process and metagenomic and metatranscriptomic sequencing, see Ju et al. [28]. As explained below, after initial data processing, we based our final analysis on data from 10 of the 12 WWTPs sampled.

### *2.1 Data processing*

We used Ju et al.'s [28] metagenomic data (DNA) to assess the WWTP community in terms of taxonomic composition, and their metatranscriptomic data (RNA) as a measure of metabolic and reproductive activity [26, 27]. We assessed the raw data via MG-RAST [29] and made use of the implemented MG-RAST prefiltering and ribosomal sequence calling. All statistical analyses and data visualizations described in the present paper, unless otherwise stated, were performed with the packages *ggpubr* v. 0.4.0 [30], *rstatix* v. 0.7.0 [31], *SpiecEasi* v. 1.1.0 [32], and *vegan* v. 2.5-7 [33] in R v. 3.6.2 [34]. All figures, except the networks, were produced using *ggthemes* v. 4.2.4 [35], *ggplot2* v. 3.3.5 [36], and *ggpubr* v. 0.4.0 [30]. The networks were visualized using *Cytoscape* v. 3.8.0 [37].

To identify prokaryotic taxa (bacteria and archaea) in the WWTP samples, we searched for sequence similarities in the SILVA data base [38]. Similarly, to identify eukaryotic taxa (protists, fungi, and microscopic metazoa), we searched the PR<sup>2</sup> data base [39]. Using BLASTN [40], we filtered the search results using an *e* value of  $1e^{-50}$  and a similarity threshold of  $\geq 80\%$ , keeping only the best hit. Given the limitations of the sequencing method (read length of  $\sim 150$  bp per fragment, limited sequencing depth, and sequencing of random fragments), we binned sequences at genus level, to avoid overestimation of microbial diversity in the data set [28, 41, 42]. Singletons were removed and putative contaminants, such as sequences derived from macroscopic animals, higher plants (Streptophyta), and chloroplasts, were excluded. For convenience, in this paper, we refer to the assessed communities as

## Chapter 1

---

“microbial communities,” although they also include microscopic fauna (gastrotrichs, nematodes, rotifers, and tardigrades).

Considering that WWTP microbial communities are affected by location-specific environmental and operational factors, and therefore cannot necessarily be treated as biological replicates, we screened the data for potential outliers [1]. To this end, we compared the microbial communities of the different WWTP locations by exploring multivariate dispersion (visualized by non-metric multidimensional scaling, NMDS, function metaMDS, package vegan, Supplementary Fig. 1 A) and beta diversity (function vegdist, package vegan, Supplementary Fig. 1 B). Bray-Curtis dissimilarity was calculated from the relative abundance data, i.e., the number of reads of each taxon was divided by the total number of reads of the respective sample. Significant differences in beta diversity were identified using unpaired two-sample Wilcoxon tests (function stat\_compare\_means, package ggpubr). Based on these results, WWTP location “FD” [28] was removed from subsequent analyses due to clear differences in beta diversity (Supplementary Fig. 1). In addition, location “BE” was removed because its design prevented the sampling of its denitrification bioreactor. Consequentially, our further analysis focused on 10 out of the originally 12 WWTPs sampled by Ju et al. [28].

For these ten locations, sequence data were subsampled (rarefied) to guarantee a similar sampling depth of ribosomal (marker) gene sequences across the entire range of DNA and RNA data, respectively. Prior to rarefaction, one RNA sample and two DNA samples were removed from the data set because of exceptionally low sequencing depth in ribosomal genes. Accordingly, the data were rarefied to a depth of 13,359 DNA and 13,812 RNA marker gene sequences per sample.

Furthermore, we evaluated the variation caused by sample processing, i.e., sequencing. When Ju et al. [28] sampled the WWTPs for the database used in our study, they collected one sample per compartment, except for WWTP location “ZR”, where an additional two replicates in the inflow compartment were subjected to sequencing to assess the technical variation. We evaluated this variation based on an NMDS plot made with relative abundance data transformed by Bray-Curtis dissimilarity (metaMDS function, package vegan; Supplementary Fig. 2). This analysis showed that variation caused by sequencing was low. For the remainder of our analyses, we kept only one of the three replicates mentioned above, to ensure comparability with the single samples taken from the other compartments at the different WWTP locations.

Rarefaction curves were calculated from count data using the function rarecurve (package vegan). With a total richness of 1947 and 1887 operational taxonomical units (OTUs) identified in the rDNA and rRNA data respectively, rarefaction curves showed sufficient saturation in sequencing (Supplementary Fig. 3). An overview of the number of reads and OTUs of prokaryotes, protists, fungi, and microscopic metazoans is provided in Supplementary Table 1.

## Chapter 1

---

Finally, functional traits were assigned to the taxa identified, using published reference databases [43,44,45,46,47,48]. Based on these trait databases, we classified the following taxa as parasites: (a) all protist genera associated with human and animal gut and/or feces and (b) all prokaryote, fungal, and microscopic-metazoan genera that include potentially pathogenic species to humans and animals. The poorly investigated and difficult to detect *Rosculus* and *Guttulinopsis*, two protistan taxa that are primarily known from feces of livestock for which evidence of a complete gut passage is yet missing, are here also considered as parasites [49]. An overview of the parasitic genera thus identified is given in Supplementary Table 2.

### *2.2 Area plots, line plots, and box plots*

To analyze microbial community changes across wastewater treatment compartments, we computed area plots of the 11 most abundant prokaryotic and eukaryotic orders, including both free-living and parasitic taxa. Differences between the treatment compartments were tested both in terms of community composition (rDNA) and activity (rRNA), using Permutational Multivariate Analysis of Variance (PERMANOVA, *adonis* function, package *vegan*). Changes in the total number of ribosomal sequences over time, i.e., across the consecutive compartments, were analyzed based on qPCR analysis for DNA sequencing and spiked internal standards for RNA sequencing (RIS) (Supplementary Fig. 4). Total abundances of eukaryotic ribosomal sequences were estimated based on the relative proportion of shotgun data in relation to the total abundances of qPCR prokaryotic ribosomal sequences. Differences between the total abundance in the inflow versus denitrification bioreactor, the denitrification versus nitrification bioreactor, and the nitrification bioreactor versus the outflow were tested using sign test (function *sign\_test*, package *rstatix*, Supplementary Table 3). Additionally, the relative abundance of selected parasitic protist taxa over time were visualized in line plots and tested for significant differences using sign test (function *sign\_test*, package *rstatix*), comparing their relative abundance in the inflow versus denitrification bioreactor, and the nitrification bioreactor versus the outflow (Supplementary Table 4). Finally, to evaluate differences between measurable presence and activity, we compared the relative abundance of rDNA versus rRNA reads. This was done for the most numerous orders within the community, across all compartments (Supplementary Table 5), as well as for the seven parasitic protist taxa mentioned above, focusing on the inflow compartment where they were most abundant (box plots). For the latter, differences between rDNA and rRNA relative abundance were determined by Sign test, not considering outliers (package *rstatix*).

### *2.3 Network inference*

Co-occurrence network analyses were performed to assess the complexity of correlations between free-living and parasitic taxa within the WWTP microbial community and draw inferences about the role of predation in parasite removal. Beforehand, we conducted two pre-processing steps to reduce indirect associations (spurious edges).

First, we reduced spurious edges caused by environmental factors. In network inference, it is a challenge to disentangle microbial associations reflecting ecological relationships—direct edges—from those induced by the environment—indirect edges [50,51,52,53]. To evaluate the influence of environmental factors on the WWTP microbial community, NMDS plots were computed for the WWTPs as a whole (Supplementary Fig. 5) as well as for the individual compartment types (Supplementary Fig. 6), for both rDNA and rRNA relative abundance data (function metaMDS, package vegan). Next, the environmental data measured by Ju et al. [28] (pH, dissolved oxygen, dissolved organic carbon, total nitrogen, total phosphorus, temperature, and hydraulic retention time) were fitted onto the ordinations using envfit (vegan). The resulting  $p$  values were corrected for multiple testing according to Benjamini & Hochberg [54]. Significant environmental vectors, scaled (multiplied) with their correlation value, were added to the NMDS plots. This analysis showed that environmental factors such as pH, total phosphorus (TP), and dissolved organic matter (DOC) correlated significantly with the diversity of the microbial community at the WWTP level (Supplementary Fig. 5), but not at the individual compartment level (with one exception; see Supplementary Fig. 6). Thus, spurious edges caused by environmental factors were minimized by conducting the network analysis at the compartment level.

Secondly, we reduced spurious edges caused by rare species. Since co-absence can yield artificially high correlation values that have no ecological meaning [50], we filtered the data for rare taxa. Thus, for each WWTP compartment type, we excluded taxa detected in fewer than seven samples (of  $N=10$  samples across WWTP locations). Metagenomic and metatranscriptomic data were processed separately.

Following these two pre-processing steps, network analysis was conducted by combining two methods, i.e., Sparse Correlations for Compositional data (SparCC) and Sparse and Compositionally Robust Inference of Microbial Ecological Networks (SPIEC-EASI), as suggested by Chen et al. [55]. SparCC accounts for compositionality using a correlation measure derived from Aitchison's variance of log-ratios [56], while SPIEC-EASI, in addition to accounting for compositionality, also reduces indirect edges by using sparse neighborhood or inverse covariance selection to infer correlations [32]. Since each method relies on different approaches to optimally filter noises and none performs across all data sets, we combined the two methods in an attempt to improve the prediction accuracy [55, 57, 58].

## Chapter 1

---

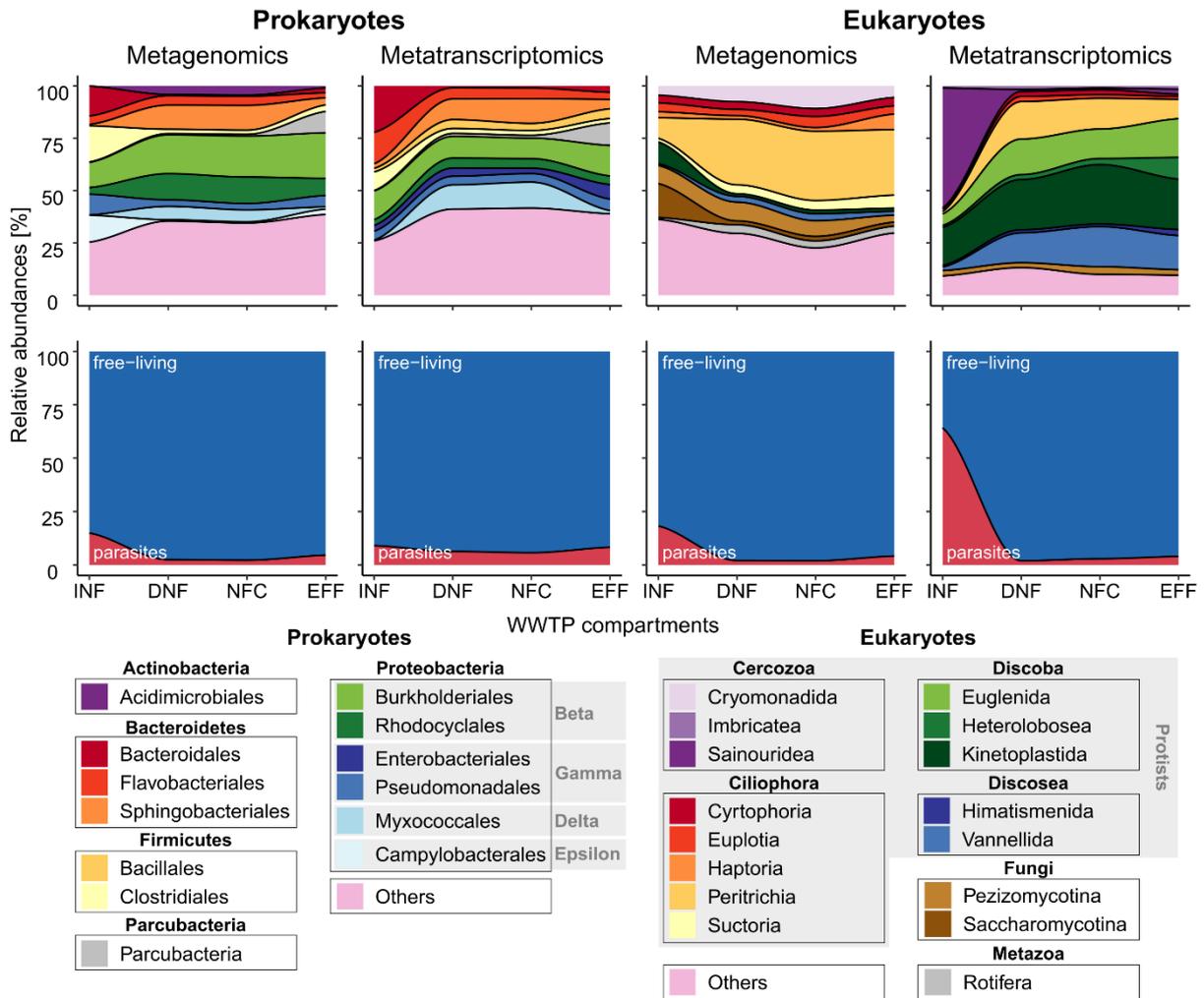
To combine SparCC and SPIEC-EASI methods, networks were first calculated with each approach separately. The same workflow was used for each WWTP compartment type and conducted separately for rDNA and rRNA data. For the Python (v. 2.7.18) based SparCC function non-normalized count data were used (package `sparcc`, v. 0.1.0, Friedman & Alm, 2012). Significant correlations at False Discovery Rate 0.05 were obtained by 100 permutations of randomly shuffled data (function `MakeBootstraps`, package `sparcc`), subjected to network inference. SPIEC-EASI networks (function `spiec.easi`, package `SpiecEasi`), based on non-normalized count data as well, were calculated with the sparse Meinshausen-Buhlmann's neighborhood selection (`mb`) method [32]. The default scaling factor determining the minimum sparsity (`lambda.min.ratio`) was lowered to 0.001 because of the density of the networks. In order to get closer to the target stability threshold (0.05), `nlambda` was set to 50. Finally, only the shared correlations of both network inference methods were retained and visualized in Cytoscape v. 3.8.0 [37] (Supplementary Fig. 7). The complexity of the networks was reduced for visualization [51]. To this end, the nodes were grouped at order level, with the edges indicating the number of genera with shown correlations. To further reduce the complexity of the graphs, only correlations with parasitic taxa are shown in this paper.

### 3 Results

The taxonomic richness of microbial organisms associated with wastewater treatment was found to be high (Supplementary Table 1). We identified a total of 508,250 SSU rDNA sequences via metagenomics and 537,935 SSU rRNA sequences via metatranscriptomics and assigned these to 1947 and 1887 operational taxonomic units (OTUs), respectively. Prokaryotes constituted ~ 94.3 % of the rDNA reads, but only ~ 42.5 % of the rRNA reads. Conversely, protists were less dominant in the rDNA reads (~ 4.6 %) but represented as much as ~ 54.8 % of the rRNA reads. Fungi and microscopic metazoa represented only minor fractions, with slightly higher contributions to the rRNA reads (~ 1.9 % and ~ 0.9 %, respectively) than rDNA reads (~ 0.7 % and ~ 0.4 %, respectively).

In terms of community composition (rDNA, metagenomics), the prokaryotic community was dominated by the orders Burkholderiales, Rhodocyclales, and Sphingobacteriales (bacteria), and the eukaryotic community by the protist orders Peritrichia (Ciliophora) and Cryomonadida (Cercozoa), and the fungal order Pezizomycotina (Fig. 1). The highest activity (rRNA, metatranscriptomics) was found in the prokaryotic orders Burkholderiales, Flavobacteriales, and Sphingobacteriales (bacteria) and the protist orders Kinetoplastida (Discoba), Sainouridea (Cercozoa), and Euglenida (Discoba) (Fig. 1). Taxonomic composition differed significantly between the four WWTP compartments (PERMANOVA, rDNA:  $R^2 = 0.59$ ,  $p = 0.001$ ; rRNA:  $R^2 = 0.35$ ,  $p = 0.001$ ), reflecting community changes during the wastewater treatment process (Supplementary Fig. 5). Interestingly, across all compartments, the

# Chapter 1

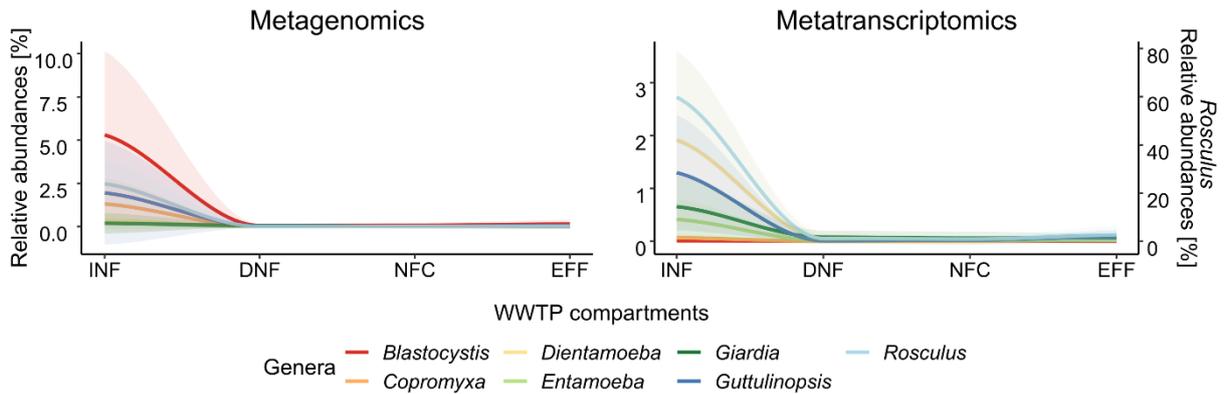


**Fig. 1:** Changes in microbial community composition during wastewater treatment. Area plots showing microbial community composition in four consecutive compartments: INF inflow (sewage), DNF denitrification bioreactor, NFC nitrification bioreactor, and EFF effluent. Numbers in A represent the mean relative abundance (across  $N=10$  WWTP locations) of the most abundant orders and, in B, the mean proportion of potential parasites versus free-living taxa. Numbers are shown separately for prokaryotic taxa (1st and 2nd column) and eukaryotic taxa (3rd and 4th column), comparing relative abundance based on rDNA reads (metagenomics) and rRNA reads (metatranscriptomics).

eukaryotic taxa showed pronounced differences between their relative abundance of rDNA (a measure of biomass) and relative abundance of rRNA (a measure of activity), while these differences were much less pronounced in the prokaryotic taxa (Fig. 1, Supplementary Table 5).

In terms of parasitic taxa, we found that the WWTP microbiome included a large and diverse number of parasitic eukaryotes (up to  $\sim 64\%$  of the eukaryote reads, of which the majority represented parasitic protists and less than 1% represented other parasitic eukaryotes) and a relatively smaller fraction of parasitic bacteria (up to  $\sim 15\%$  of the prokaryote reads). Effective removal of parasites over the course of wastewater treatment was indicated by their pronounced decrease from high abundance in the inflow (mostly raw sewage) to low abundance in the outflow (effluent), both in relative numbers (Fig. 1) and total numbers of ribosomal sequences (Supplementary Fig. 4).

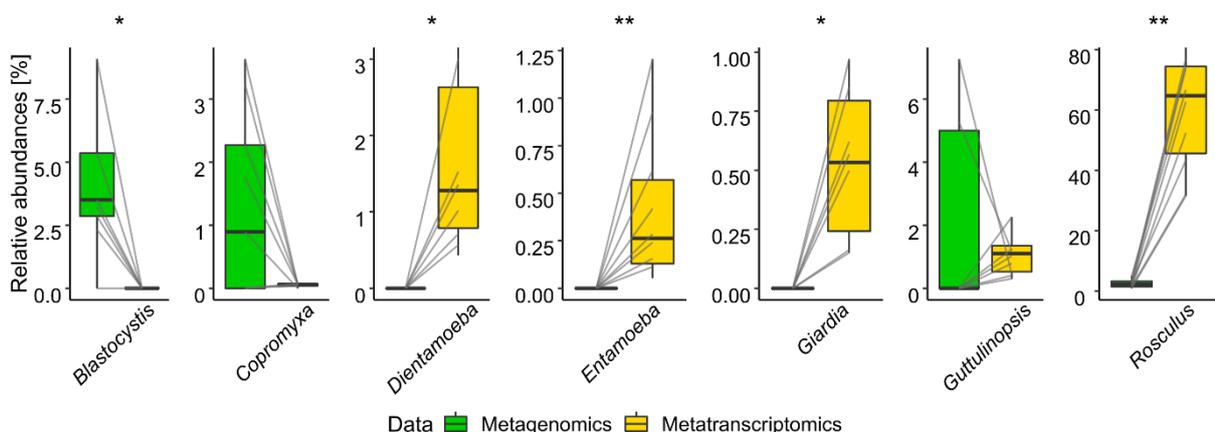
# Chapter 1



**Fig. 2:** Removal of parasitic protists from wastewater. Line plots showing mean relative abundance of selected parasitic protist genera across the four WWTP compartments (INF inflow (sewage), DNF denitrification bioreactor, NFC nitrification bioreactor, EFF effluent). Numbers shown are mean relative protist abundances (across N=10 WWTP locations) of rDNA reads (metagenomics, left-hand side) and rRNA reads (metatranscriptomics, right-hand side). In the latter, abundance of the genus *Rosculus* is shown on a separate Y-axis because of its high number of rRNA reads. Standard deviations are indicated by the transparent areas (colors matching with individual lines).

Based on these observations, we investigated the progress of parasite removal during wastewater treatment in closer detail by comparing changes in relative abundance and activity across WWTP compartments, focusing on selected taxa of parasitic protists (Fig. 2). Here, when comparing the inflow (INF) to denitrification (DNF) compartments, significant decreases were found in the relative abundance of *Blastocystis* and *Rosculus*, and in the activity of all taxa except *Blastocystis* (Sign-Test, Supplementary Table 4). When comparing the nitrification (NFC) bioreactor to the outflow (EFF) compartment, significant increases were found in the activity of *Copromyxa* and *Rosculus* (Sign-Test, Supplementary Table 4).

Given these different changes in abundance versus activity, we contrasted the “detectability” of each of the protist taxa mentioned above, in terms of abundance (rDNA, metagenomics) versus activity



**Fig. 3:** Detection of parasitic protists in wastewater: abundance versus activity. Boxplots showing, for selected taxa, the 25 % and 75 % percentiles and medians of the relative abundance of protist reads in metagenomic data (rDNA, green) and metatranscriptomic data (rRNA, yellow), in samples from the inflow compartment (at N=10 WWTP locations). Significant differences between the rDNA and rRNA abundances are indicated with asterisks (sign test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). The gray lines link the rDNA and rRNA sample pairs from the same location.

## Chapter 1

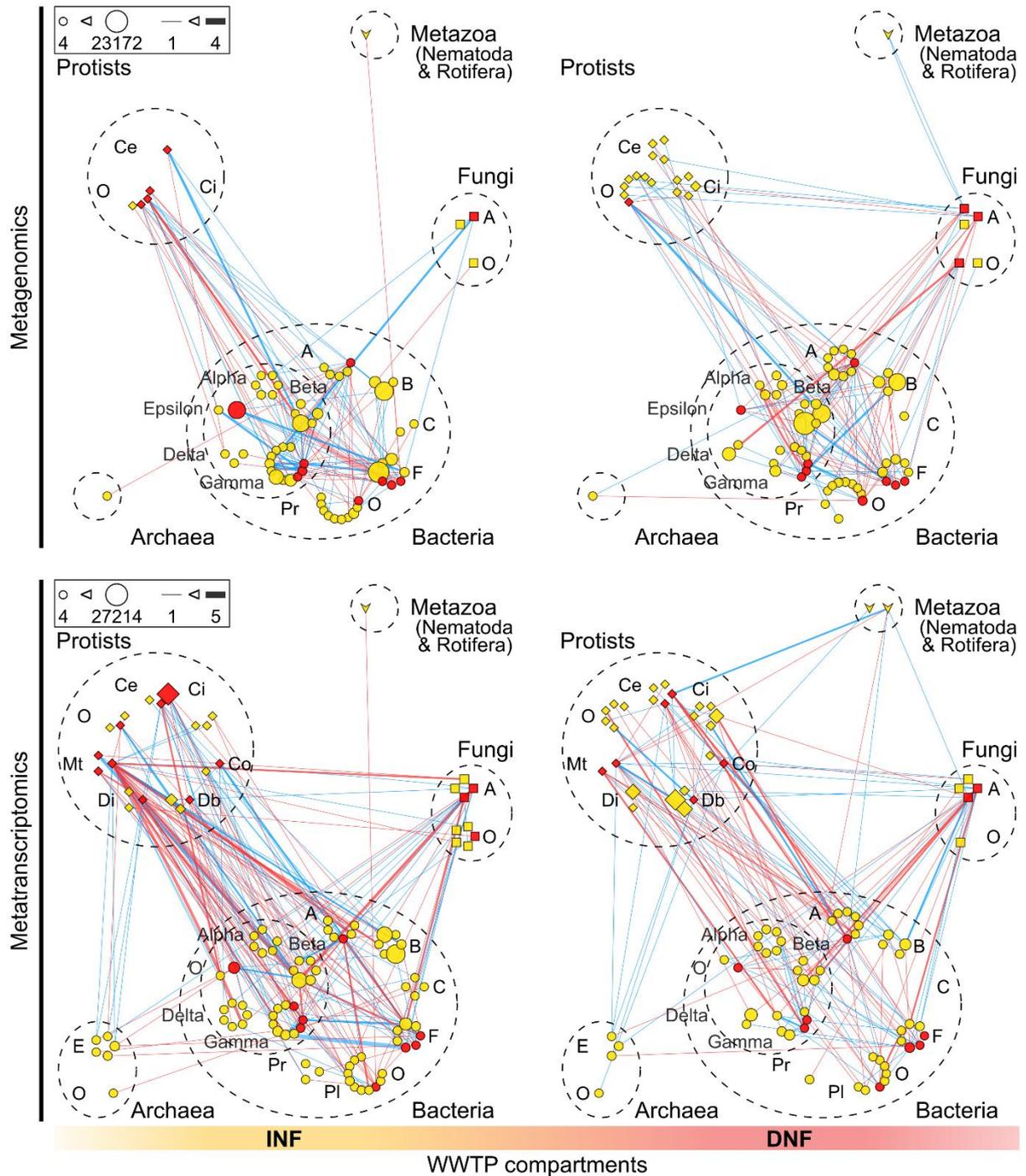
---

(rRNA, metatranscriptomics), focusing on the samples from the inflow (Fig. 3). Intriguingly, this comparison showed that the gut parasites *Dientamoeba*, *Entamoeba*, *Giardia*, and *Rosculus* were hardly detectable in terms of abundance (rDNA) but yielded a high number of reads in terms of activity (rRNA). Conversely, the gut parasites *Blastocystis*, *Copromyxa*, and *Guttulinopsis* were hardly detectable in terms of activity (rRNA) but were clearly present in most rDNA samples. In other words, when present, the latter taxa showed low or no measurable activity.

To investigate whether parasite abundance and activity patterns across wastewater treatment can be explained by microbial community interactions, we conducted network analyses, looking specifically for associations between parasites and potential competitors, predators, and co-associated parasites. When comparing the networks of the communities in the inflow versus denitrification compartments, where most of the parasite removal took place, we observed a surprisingly high number of correlations between bacteria and eukaryotes, in addition to the commonly reported correlations within bacteria (Fig. 4; for other compartments, see Supplementary Fig. 7). Across the networks shown in Fig. 4, correlations within the bacteria accounted for ~ 47 % of all correlations, while correlations between bacteria and eukaryotes accounted for ~ 44 %. Of the latter correlations, 74 % involved protists. Particularly interesting are the correlations found for *Rosculus*, the main genus found in the Cercozoa (“Ce” in Fig. 4). This parasitic and bacterivorous protist was found to be highly active (up to ~ 84 % of the protist rRNA reads) in the inflow, with significantly lower readings in the denitrification bioreactor (only ~ 4 % of the protist reads, Fig. 2, Sign-Test, Supplementary Table 4). Network inference revealed that, in the inflow, *Rosculus* correlated exclusively with bacteria (indicating *Rosculus* feeding on bacteria) while, in the denitrification compartment, it had fewer correlations with bacteria but gained correlations with the rotifers *Adenita* and *Monostyla* (indicating *Rosculus* being preyed upon by rotifers). Other parasitic protist taxa followed the same general pattern of strongly decreasing numbers between the inflow and denitrification compartments (Fig. 2), correlating with bacteria, fungi, and other smaller protists in the inflow, and gaining correlations with predatory ciliates and rotifers in the denitrification bioreactor (Fig. 4).

Interestingly, the correlations discussed above emerged more strongly from the networks derived from activity data (rRNA, metatranscriptomics) than from the networks derived from abundance data (rDNA, metagenomics). As shown in Fig. 4, rDNA and rRNA networks showed distinct differences in density, with respectively 124 versus 192 edges (correlations) and 65 versus 104 nodes (taxa) in the inflow, and 135 versus 151 edges and 78 versus 85 nodes in the denitrification compartment. This greater density of rRNA networks was also observed in the other compartments (Supplementary Fig. 7). Moreover, the rRNA networks revealed associations between parasitic protists and their potential predators (rotifers, in particular) that were not detected in the rDNA networks.

# Chapter 1



**Fig 4: Co-occurrence networks of parasitic orders in the inflow and denitrification bioreactor.** Networks showing correlations derived from co-occurrence network inferences for the inflow (INF) and denitrification bioreactor (DNF) at N=10 WWTP locations, based on metagenomic (first row) and metatranscriptomic (second row) data. Only associations that involve parasites are shown. Nodes represent genera grouped at the order level and trait level (red nodes: parasitic taxa; yellow nodes: free-living taxa), with node size proportional to the total number of reads for each order. Edges represent correlations between taxa (blue lines: positive correlations; red lines: negative correlations), with line thickness proportional to the number of genera per order involved. Abbreviations for archaea: E Euryarchaeota, O others. Abbreviations for bacteria: A Actinobacteria, B Bacteroidetes, C Chloroflexi, F Firmicutes, O Others, PI Planctomycetes, Pr Proteobacteria, T Tenericutes, V Verrucomicrobia. Abbreviations for Proteobacteria: Alpha Alphaproteobacteria, Beta Betaproteobacteria, Gamma Gammaproteobacteria, Delta Deltaproteobacteria, O others. Abbreviations: Fungi: A Ascomycota, B Basidiomycota, O others. Abbreviations for protists: Ce Cercozoa (\*including Rosculus), Ci Ciliophora, Co Conosa, Db Discoba, Di Discosea, Ms Mesomycetozoa, Mt Metamonada, O Others, S Stramenopiles, T Tubulinea.

### 4 Discussion

Our study provides a comprehensive overview of the diversity of microorganisms in wastewater treatment plants (WWTPs), including not only bacteria but also fungi, protists, and microscopic metazoans. Expectedly, bacteria represented the most numerous fraction of the microbial community in terms of metagenomics [6]. However, in terms of metatranscriptomics (activity), the number of eukaryotic reads was higher than the prokaryotic reads. Surprisingly, within the eukaryotic fraction, the abundance and activity of protists were found to be much higher than of fungi. With ~ 55% of all rRNA reads, protists, including many potential parasites, were the most active eukaryotes in the WWTP microbial community. Our results further showed that rDNA (abundance) and rRNA (activity) data showed profoundly different patterns, especially among the eukaryotic taxa.

#### *4.1 Parasite removal – predator facilitated?*

Previous studies have repeatedly shown that wastewater is a hotspot of potential parasites [6, 59, 60]. These studies mostly focused on investigating the bacterial community and potentially parasitic bacteria [6], ignoring the fate of eukaryotic parasites during wastewater treatment. Our study highlights the importance of the latter group, showing a surprising diversity and abundance of gut- and feces-associated parasitic protists, particularly in the inflow (sewage). Our primer-independent findings significantly add to Maritz et al. [23], who detected various parasitic protists in raw sewage using a primer-based approach. While they identified parasitic protists such as *Blastocystis*, *Entamoeba*, and *Trichomonas*, we detected the same taxa plus numerous additional ones, including *Dientamoeba*, *Guttulinopsis*, *Giardia*, and *Rosculus*. Many of these eukaryotic parasites are known to be “long branch organisms,” i.e., organisms with highly divergent marker gene sequences that often cannot be assessed by conventional primer-based sequencing methods [49], which impedes the detection of taxa such as *Giardia* [61]. In contrast, the primer-independent shotgun data used in our study allow to assess the full spectrum of taxa. For example, Wylezich et al. [24, 25] demonstrated the use of this approach to assess the full range of eukaryotic parasites present in swine feces. It can be concluded that primer-based approaches have only limited use for monitoring eukaryotic parasites in wastewater.

An array of studies revealed that potentially harmful (parasitic) bacteria were strongly reduced during the initial phases of wastewater treatment [42, 62, 63]. Our study also shows a pronounced decrease in parasitic taxa between the inflow (sewage) and denitrification bioreactor, but, moreover, shows that this decrease also applies to eukaryotic parasites, protists in particular. This decrease can partly be explained by the transition in chemical conditions of the environment, flocculation, and sedimentation [1,2,3, 42]. However, our network analyses suggest that predation may also play a role, as we found

## Chapter 1

---

parasites to co-occur with a number of other taxa within the microbial WWTP community, indicating that trophic interactions, as suggested by laboratory experiments, could be taking place [13, 14, 60, 64]. In the context of parasite removal from wastewater, the position of gut-parasitic protists is particularly interesting since many of these taxa are both predators (bacterivores) and prey. In our study, the networks of the inflow and denitrification compartments showed a high percentage of correlations (~44 %), thus putative ecological interactions, between bacteria and eukaryotes. Among these correlations, ~ 74 % involved protists, indicating their central role as bacterivorous regulators of bacterial community composition (including potentially preying on gut-parasitic bacteria). Compared to the inflow (sewage), correlations between protists and bacteria decreased in the denitrification compartment, whereas new correlations emerged between protists and their potential predators—i.e., ciliates and rotifers. Previous studies have identified ciliates (protists) and rotifers (metazoans) as potentially the most crucial predators in WWTPs [65, 66]. Our network analyses support this idea, providing evidence of trophic interactions between these predators and their protist prey, *in situ*.

### *4.2 Contrasting abundance versus activity patterns of eukaryotic parasites: consequences for WWTP biomonitoring*

Taxonomic composition of the eukaryotic fraction strongly differed between rDNA data (metagenomics) and rRNA data (metatranscriptomics) (Fig. 1). In extreme cases, we found some of the parasitic protists to be abundantly present in rDNA data but below detection level in rRNA data, or completely the other way around (Fig. 3). Such differences are generally not found in prokaryotes, where it is possible to assess “normalized activity” (RNA/DNA quotient as a measure of activity per individual) based on metagenome-assembled genomes (MAG) (see for example Herold et al. [67] and Arbas et al. [68], reporting on WWTP bacteria). Our findings show that it is not feasible to calculate this quotient for eukaryotes (especially when either rDNA or rRNA is zero); in addition, current technology does not yet allow to assess eukaryote MAGs, since eukaryotes have much larger genomes and higher variation in ribosomal gene duplication than bacteria [69].

As said, the difference in abundance (rDNA reads) versus activity (rRNA reads) was particularly strong for some of the parasitic protists (Fig. 3). In the absence of their natural hosts, we expected these parasites to become dormant, i.e., low in activity and potentially forming resting stages [70]. This was indeed found for the taxa *Blastocystis*, *Copromyxa*, and *Guttulinopsis*, whose presence could be detected via rDNA but whose activity was so low that it mostly fell below the sensitivity threshold of our rRNA sequencing (Fig. 3). In contrast, the taxa *Dientamoeba*, *Entamoeba*, *Giardia*, and *Rosculus* were hardly present in the rDNA data but showed a high expression of ribosomal genes, indicating high activity and even potential reproduction [69]. Outstanding was the high proportion of *Rosculus* in rRNA

# Chapter 1

---

data, making up to ~ 84 % of the eukaryotic fraction. *Rosculus* is known to be highly abundant and active in feces [49], and, as this study indicates, also in sewage within WWTPs.

The importance and ecological meaning of these differences in abundance versus activity data were further revealed in our network analysis (Fig. 4). As expected, the activity-based rRNA networks showed a higher number of edges, i.e., putative interactions, than the rDNA-based networks, because rRNA data reflect the active part of the community. More importantly, the rRNA-based networks revealed associations between parasitic protists and their potential predators (ciliates and rotifers) that were not detected in the rDNA networks. Thus, the very low abundance (rDNA reads) of the parasitic protists *Dientamoeba*, *Entamoeba*, *Giardia*, and *Rosculus* may be explained by predation. At the same time, their observed high activity (rRNA reads) and strong network correlations with bacteria suggest that these protists, while being preyed upon, themselves were actively feeding on bacteria (Fig. 4).

## 5 Conclusions

Our results are of particular interest for biomonitoring to evaluate wastewater treatment efficiency [22, 63, 71]. Especially in developing countries, the treatment of wastewater may be insufficient before re-introduction into the water system or re-use for agricultural purposes [6, 17, 59, 72]. According to Cai et al. [42], this applies, for instance, to around 80 % of sewage in India. Subsequently, potentially infectious and harmful parasitic microorganisms become widespread and form a threat to public health when present in drinking water, water recreation areas, and aquatic food production systems [42, 73,74,75]. Our results clearly show that biomonitoring of wastewater treatment efficiency via molecular methods (“-omics”) can be greatly improved by using primer-independent shotgun approaches to ensure adequate detection of parasitic protists. Combining shotgun metagenomics with shotgun metatranscriptomics allows to monitor both abundance and activity of this important group of microeukaryotic parasites. This improvement is crucial for reducing the public health risks associated with insufficiently treated wastewater.

## Abbreviations

WWTPs: Wastewater treatment plants; rDNA: Ribosomal DNA; rRNA: Ribosomal RNA; INF: Sewage-inflow; DNF: Denitrification bioreactor; NFC: Nitrification bioreactor; EFF: Effluent; OTUs: Operational taxonomic units; MAG: Metagenome-assembled genomes.

## Chapter 1

---

### Author contributions

K.D. conceived the study. J.F. and K.D. analyzed the data and prepared all illustrations. K.D. and J.F. wrote the manuscript with input from all co-authors. The authors read and approved the final manuscript.

### Availability of data and materials

The analyzed data was submitted to the European Nucleotide Archive under the accession numbers PRJEB28815, PRJEB28830, PRJEB30264, PRJEB28728, PRJEB28831, and PRJEB28737. Tables with metadata are included in the publication Ju et al. [28]. The code for the presented analyses are available over GitHub under the following link: <https://github.com/JFreude/MicroeukaryoticGutParasitesInWastewaterTreatmentPlants>.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Funding

Open Access funding enabled and organized by Projekt DEAL. The study was supported by the NRP72 project 407240\_167116 “Swiss River Resistome”.

### Acknowledgments

We thank Christien Ettema for her thorough editing of the manuscript.

## References

1. Rodríguez E, García-Encina PA, Stams AJM, Maphosa F, Sousa DZ. Meta-omics approaches to understand and improve wastewater treatment systems. *Rev Environ Sci Biotechnol*. 2015;14:385–406 Kluwer Academic Publishers.
2. Arregui L, Pérez-Uz B, Salvadó H, Serrano S. Progresses on the knowledge about the ecological function and structure of the protists community in activated sludge wastewater treatment plants. *Curr Res Technol Edu Top Appl Microbiol Microb Biotechnol*. 2010;2:972–9.
3. Walczyńska A, Sobczyk M, Fiałkowska E, Pajdak-Stós A, Fyda J, Wiąckowski K. Interaction between a bacterivorous ciliate *Aspidisca cicada* and a rotifer *Lecane inermis*: doozers and fraggles in aquatic flocs. *Microb Ecol*. 2018;75:569–81.
4. Ferrera I, Sánchez O. Insights into microbial diversity in wastewater treatment systems: how far have we come? *Biotechnol Adv*. 2016;34:790–802.
5. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat Microbiol*. 2019;4:1183–95.
6. Aghalari Z, Dahms H-U, Sillanpää M, Sosa-Hernandez JE, Parra-Saldívar R. Effectiveness of wastewater treatment systems in removing microbial agents: a systematic review. *Glob Health*. 2020;16:13.
7. Bass D, del Campo J. Micro-eukaryotes in animal and plant microbiomes: ecologies of disease? *Eur J Protistol*. 2020;76:125719 Elsevier GmbH.
8. del Campo J, Bass D, Keeling PJ. The eukaryome: diversity and role of microeukaryotic organisms associated with animal hosts. *Funct Ecol*. 2020;34:2045–54.
9. Foissner W. Protists as bioindicators in activated sludge: identification, ecology and future needs. *Eur J Protistol*. 2016;55:75–94 Elsevier GmbH.
10. Flues S, Bass D, Bonkowski M. Grazing of leaf-associated Cercomonads (Protists: Rhizaria: Cercozoa) structures bacterial community composition and function. *Environ Microbiol*. 2017;19:3297–309.
11. Glücksman E, Bell T, Griffiths RI, Bass D. Closely related protist strains have different grazing impacts on natural bacterial communities. *Environ Microbiol*. 2010;12:3105–13.
12. Chabé M, Lokmer A, Ségurel L. Gut protozoa: friends or foes of the human gut microbiota? *Trends Parasitol*. 2017;33:925–34 Elsevier Ltd.
13. Lee NM, Welander T. Use of protozoa and metazoa for decreasing sludge production in aerobic wastewater treatment. *Biotechnol Lett*. 1996;18:429–34.
14. Kim T-D, Unno H. The roles of microbes in the removal and inactivation of viruses in a biological wastewater treatment system. *Water Sci Technol*. 1996;33:243–50.
15. Walochnik J. Parasitic protozoa of farm animals and pets. In: Florin-Christensen M, Schnittger L, editors. *Parasit protozoa farm Anim*; 2018. p. 1–438.
16. Razzolini MTP, Breternitz BS, Kuchkarian B, Bastos VK. *Cryptosporidium* and *giardia* in urban wastewater: a challenge to overcome. *Environ Pollut*. 2020;257:113545 Elsevier Ltd.
17. Siwila J, Mwaba F, Chidumayo N, Mubanga C. Food and waterborne protozoan parasites: the African perspective. *Food Waterborne Parasitol*. 2020;20:e00088.
18. Efstratiou A, Ongert JE, Karanis P. Waterborne transmission of protozoan parasites: review of worldwide outbreaks - an update 2011–2016. *Water Res*. 2017;114:14–22.
19. Angelici MC, Walochnik J, Calderaro A, Saxinger L, Dacks JB. Free-living amoebae and other neglected protistan pathogens: health emergency signals? *Eur J Protistol*. 2021;77:125760.
20. Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Mol Ecol*. 2014;23:3341–55.
21. Willis AD, Minot SS. Strategies to facilitate translational advances from microbiome surveys. *Trends Microbiol*. 2020;28:329–30 Elsevier Ltd.
22. Hirakata Y, Hatamoto M, Oshiki M, Watari T, Kuroda K, Araki N, et al. Temporal variation of eukaryotic community structures in UASB reactor treating domestic sewage as revealed by 18S rRNA gene sequencing. *Sci Rep*. 2019;9:1–11 Nature Publishing Group.
23. Maritz JM, Ten Eyck TA, Elizabeth Alter S, Carlton JM. Patterns of protist diversity associated with raw sewage in New York City. *ISME J*. 2019;13:2750–63 Nature Publishing Group.
24. Wylezich C, Belka A, Hanke D, Beer M, Blome S, Höper D. Metagenomics for broad and improved parasite detection: a proof-of-concept study using swine faecal samples. *Int J Parasitol*. 2019;49:769–77 Elsevier Ltd.

# Chapter 1

---

25. Wylezich C, Caccio SM, Walochnik J, Beer M, Höper D. Untargeted metagenomics shows a reliable performance for synchronous detection of parasites. *Parasitol Res.* 2020;119:2623–9 Springer.
26. Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, et al. Metatranscriptomic census of active protists in soils. *ISME J.* 2015;9:2178–90 Nature Publishing Group.
27. Inkinen J, Jayaprakash B, Siponen S, Hokajärvi AM, Pursiainen A, Ikonen J, et al. Active eukaryotes in drinking water distribution systems of ground and surface waterworks. *Microbiome.* 2019;7:99 BioMed Central Ltd.
28. Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, Singer HP, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* 2019;13:346–60 Nature Publishing Group.
29. Meyer F, Paarmann D, D’Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:386 BioMed Central.
30. Kassambara A. Ggpubr: “ggplot2” based publication ready plots. R package version 040. Comprehensive R archive network (CRAN); 2020.
31. Kassambara A. Rstatix: pipe-friendly framework for basic statistical tests. R package version 070. Comprehensive R archive network (CRAN); 2021.
32. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol.* 2015;11:e1004226 Public Library of Science.
33. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, et al. Vegan: community ecology package. R package version 25-7. Comprehensive R archive network (CRAN); 2020.
34. R Core Team. R: a language and environment for statistical computing. *R Found Stat Comput.* 2019.
35. Arnold JB. Ggthemes: extra themes, scales and geoms for “ggplot2.” R package version 424. Comprehensive R Archive Network (CRAN); 2021.
36. Wickham H. ggplot2. *Wiley Interdiscip Rev Comput Stat.* 2011;3:180–5 John Wiley & Sons, Ltd.
37. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504 Cold Spring Harbor Lab.
38. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 2007;35:7188–96 Oxford Academic.
39. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 2013;41:597–604 Oxford Academic.
40. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421 BioMed Central.
41. Choma M, Bárta J, Šantrůčková H, Urich T. Low abundance of Archaeorhizomycetes among fungi in soil metatranscriptomes. *Sci Rep.* 2016;6:38455.
42. Cai L, Ju F, Zhang T. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl Microbiol Biotechnol.* 2014;98:3317–26 Springer Verlag.
43. Varela AR, Manaia CM. Human health implications of clinically relevant bacteria in wastewater habitats. *Environ Sci Pollut Res.* 2013;20:3550–69 Springer.
44. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol.* 2019;66:4–119 Blackwell Publishing Inc.
45. Assress HA, Selvarajan R, Nyoni H, Ntushelo K, Mamba BB, Msagati TAM. Diversity, co-occurrence and implications of fungal communities in wastewater treatment plants. *Sci Rep.* 2019;9:14056.
46. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Sci Transl Med.* 2012;4:165rv13.
47. Mishra PK, Palma M, Bleich D, Loke P, Gause WC. Systemic impact of intestinal helminth infections. *Mucosal Immunol.* 2014;7:753–62.
48. Savioli L, Gabrielli AF, Montresor A. Helminthic diseases: intestinal nematode infections. In: Quah SR, editor. *Int Encycl public heal.* 2nd ed. Oxford: Academic; 2017. p. 568–75.
49. Schuler GA, Tice AK, Pearce RA, Foreman E, Stone J, Gammill S, et al. Phylogeny and classification of novel diversity in Sainouroidea (Cercozoa, Rhizaria) sheds light on a highly diverse and divergent clade. *Protist.* 2018;169:853–74 Elsevier GmbH.

# Chapter 1

---

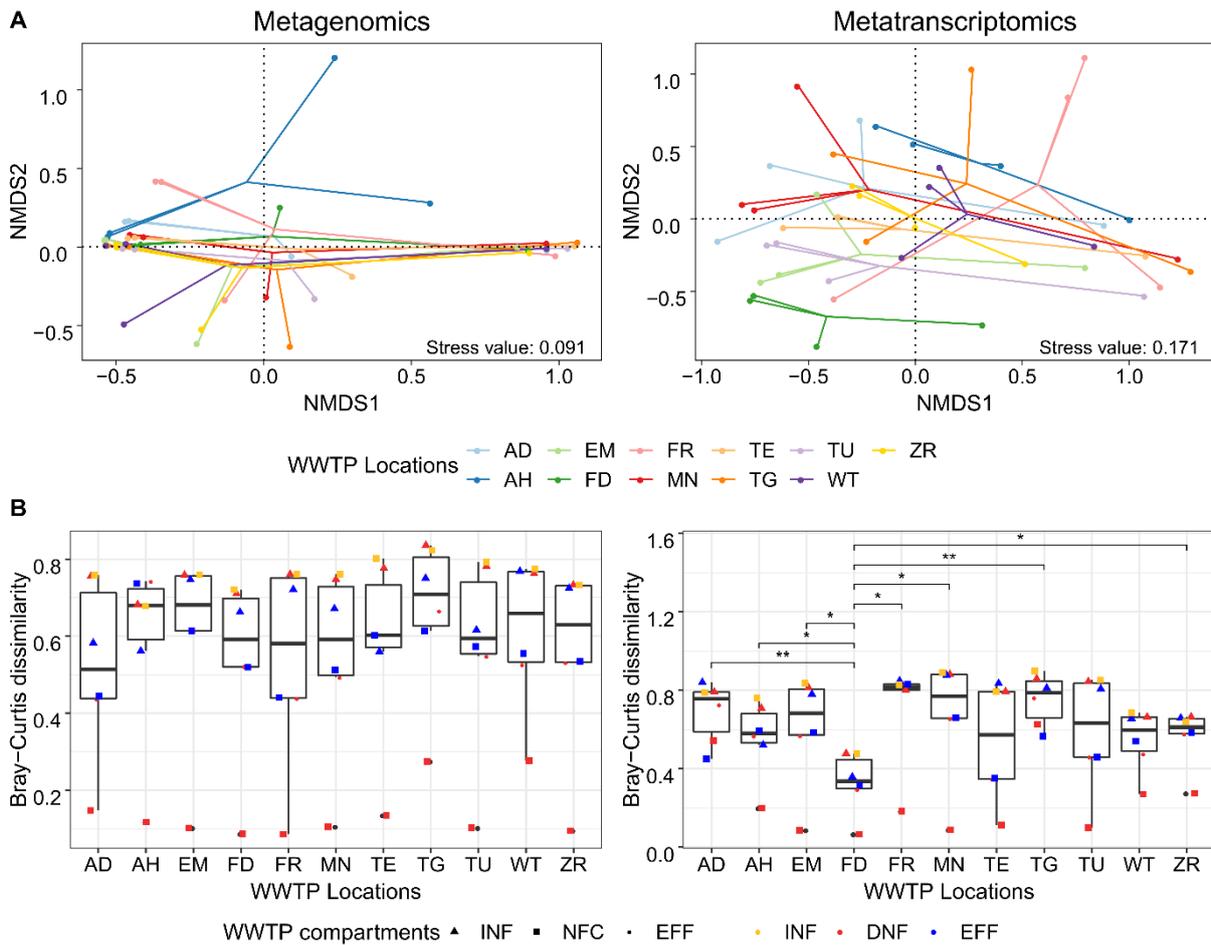
50. Röttgers L, Faust K. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol Rev.* 2018;42:761–80.
51. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol.* 2012;10:538–50 Nature Publishing Group.
52. Yang Y, Chen N, Chen T. Inference of environmental factor-microbe and microbe-microbe associations from metagenomic data using a hierarchical Bayesian statistical model. *Cell Syst.* 2017;4:129–137.e5.
53. Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol.* 2009;75:5111–20.
54. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995;57:289–300 Royal Statistical Society, Wiley.
55. Chen L, Collij V, Jaeger M, van den Munckhof ICL, Vich Vila A, Kurilshikov A, et al. Gut microbial co-abundance networks show specificity in inflammatory bowel disease and obesity. *Nat Commun.* 2020;11:1–12 Nature Research.
56. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol.* 2012;8:e1002687 Public Library of Science.
57. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. *Nat Methods.* 2012;9:796–804.
58. Wang H, Wei Z, Mei L, Gu J, Yin S, Faust K, et al. Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biol Biochem.* 2017;105:227–35.
59. Al-Gheethi AA, Efaq AN, Bala JD, Norli I, Abdel-Monem MO, Ab. Kadir MO. Removal of pathogenic bacteria from sewage-treated effluent and biosolids for agricultural purposes. *Appl Water Sci.* 2018;8:74.
60. Wen Q, Tutuka C, Keegan A, Jin B. Fate of pathogenic microorganisms and indicators in secondary activated sludge wastewater treatment plants. *J Environ Manag.* 2009;90:1442–7.
61. Zahedi A, Greay TL, Papparini A, Linge KL, Joll CA, Ryan UM. Identification of eukaryotic microorganisms with 18S rRNA next-generation sequencing in wastewater treatment plants, with a more targeted NGS approach required for *cryptosporidium* detection. *Water Res.* 2019;158:301–12 Elsevier Ltd.
62. Huang K, Mao Y, Zhao F, Zhang XX, Ju F, Ye L, et al. Free-living bacteria and potential bacterial pathogens in sewage treatment plants. *Appl Microbiol Biotechnol.* 2018;102:2455–64 Springer Verlag.
63. Xue J, Schmitz BW, Caton K, Zhang B, Zabaleta J, Taylor CM, et al. Assessing the spatial and temporal variability of bacterial communities in two Bardenpho wastewater treatment systems via Illumina MiSeq sequencing. *Sci Total Environ.* 2019;657:1543–52 Elsevier B.V.
64. Lee NM, Welander T. Reducing sludge production in aerobic wastewater treatment through manipulation of the ecosystem. *Water Res.* 1996;30:1781–90.
65. Pauli W, Jax K, Berger S. Protozoa in wastewater treatment: function and importance. In: Beek B, editor. *Biodegrad persistence.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2001. p. 203–52.
66. Lapinski J, Tunnacliffe A. Reduction of suspended biomass in municipal wastewater using bdelloid rotifers. *Water Res.* 2003;37:2027–34.
67. Herold M, Martínez Arbas S, Narayanasamy S, Sheik AR, Kleine-Borgmann LAK, Lebrun LA, et al. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. *Nat Commun.* 2020;11:5281.
68. Arbas SM, Narayanasamy S, Herold M, Lebrun LA, Hoopmann MR, Li S, et al. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. *Nat Microbiol.* 2020:1–13.
69. Weber AA-T, Pawlowski J. Can abundance of protists be inferred from sequence data: a case study of foraminifera. *PLoS one.* Public library of. Science. 2013;8:e56739.
70. Weedall GD, Hall N. Sexual reproduction and genetic exchange in parasitic protists. *Parasitology.* 2015;142:S120–7 2014/12/22. Cambridge University Press.
71. Aw TG, Rose JB. Detection of pathogens in water: from phylochips to qPCR to pyrosequencing. *Curr Opin Biotechnol.* 2012;23:422–30.
72. Mahmoud MA, Abdelsalam M, Mahdy OA, El Miniawy HMF, Ahmed ZAM, Osman AH, et al. Infectious bacterial pathogens, parasites and pathological correlations of sewage pollution as an important threat to farmed fishes in Egypt. *Environ Pollut.* 2016;219:939–48.
73. Ferguson AS, Layton AC, Mailloux BJ, Culligan PJ, Williams DE, Smartt AE, et al. Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. *Sci Total Environ.* 2012;431:314–22 Elsevier.

## Chapter 1

---

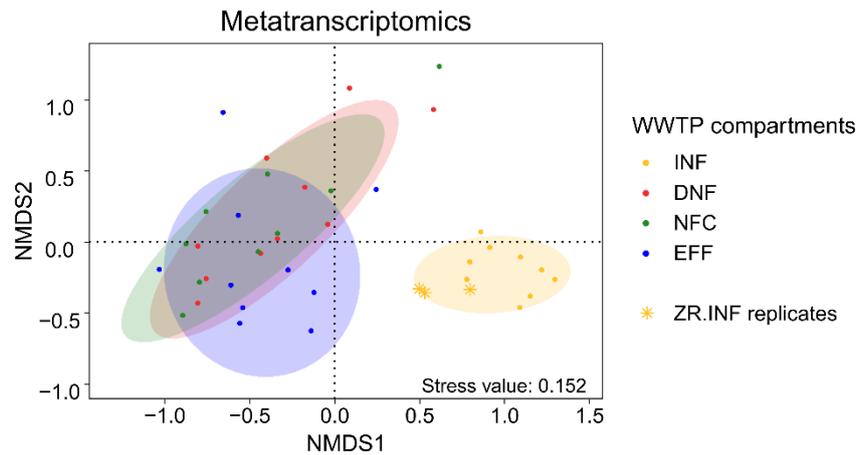
74. Pohl N, Solbach MD, Dumack K. The wastewater protist *Rhogostoma minus* (Thecofilosea, Rhizaria) is abundant, widespread, and hosts Legionellales. *Water Res.* 2021;203:117566 Elsevier Ltd.
75. Yuan L, Wang Y, Zhang L, Palomo A, Zhou J, Smets BF, et al. Pathogenic and indigenous denitrifying bacteria are transcriptionally active and key multi-antibiotic-resistant players in wastewater treatment plants. *Environ Sci Technol.* 2021;55:10862–74.

Supplementary material

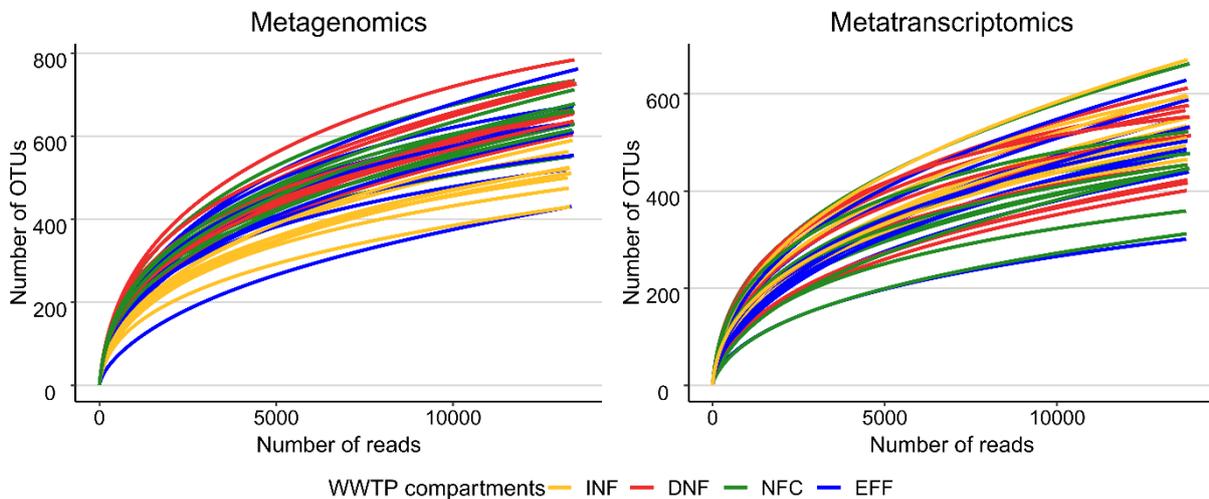


**Supplementary Figure 1:** Comparing microbial communities between WWTP locations to identify outliers. Graphs showing multivariate dispersion **(A)** and beta diversity **(B)** of microbial community composition at N=11 WWTP locations, based on metagenomic data (left-hand side) and metatranscriptomic data (right-hand side). For multivariate dispersion **(A)**, NMDS plots were calculated based on Bray-Curtis dissimilarities. Lines (color-coded by location) link samples at each location to their centromere. For beta diversity **(B)**, boxplots show the 25 % and 75 % percentiles and medians of Bray-Curtis dissimilarities. Points are color and symbol-coded by WWTP compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water). In beta diversity based on metatranscriptomic data (lower row, right-hand side), significant differences between location “FD” and the other locations are indicated with asterisks (unpaired two-sample Wilcoxon test, \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ). **Based on these results, location FD was excluded from further analysis.**

# Chapter 1

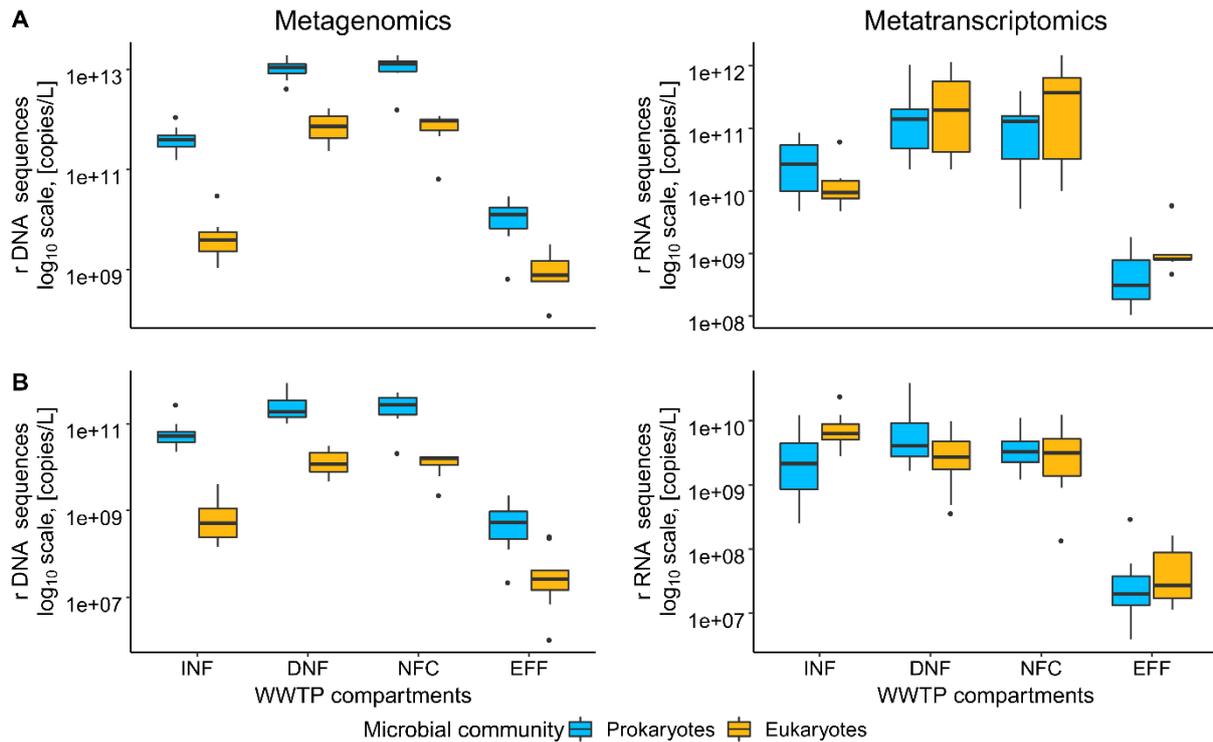


**Supplementary Figure 2:** Assessment of the variation caused by sampling processing (sequencing). NMDS plot based on Bray-Curtis dissimilarities derived from metatranscriptomic data, comparing microbial community composition across WWTP compartments and locations. The variation caused by sample processing is shown for one location, “ZR”, showing three sequencing replicates from the inflow (INF) (replicates indicated by yellow asterisks). **Based on this comparison, we concluded that variation caused by sequencing was low.** Compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water).

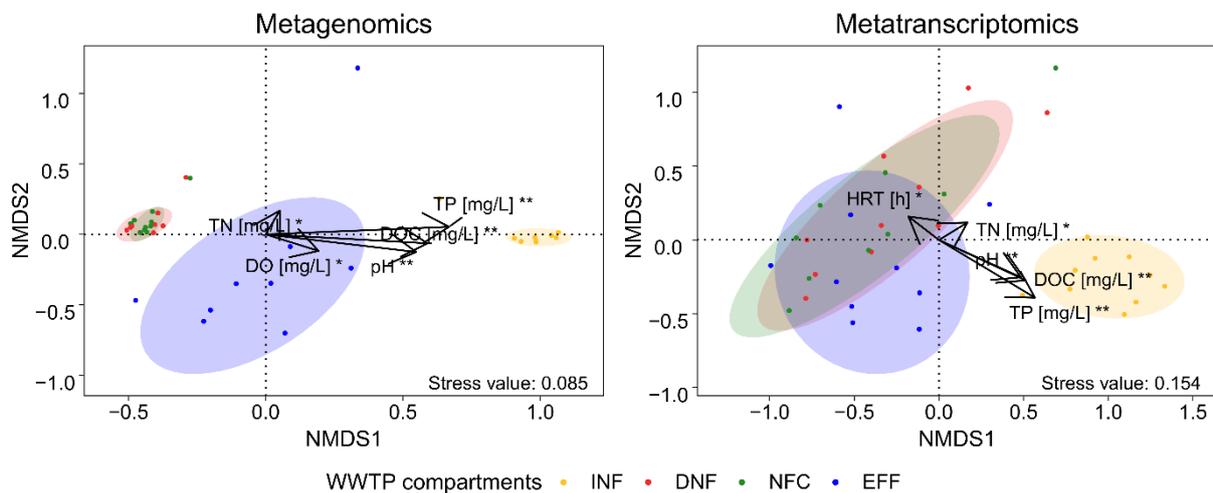


**Supplementary Figure 3:** Rarefaction curves for metagenomic (rDNA) and metatranscriptomic (rRNA) data. Curves showing the number of reads as a function of the number of OTUs identified (N=37 samples, i.e. one sample from each WWTP compartment (4) at each WWTP location (10), excluding 3 samples because of exceptionally low sequencing-depth). Samples are color-coded by compartment: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water).

# Chapter 1

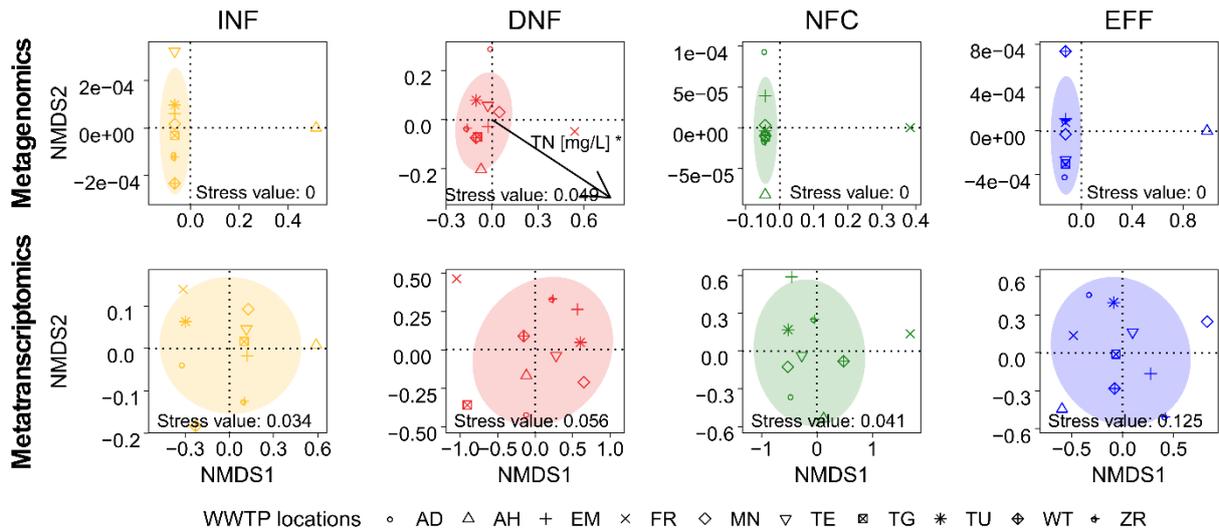


**Supplementary Figure 4:** Total number of rDNA and rRNA sequences. Boxplots showing the 25 % and 75 % percentiles and medians of the total number of rDNA (metagenomics) and rRNA (metatranscriptomics) sequences for (A) the total community and (B) the parasitic community, comparing prokaryotes (blue) and eukaryotes (yellow). Compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water).



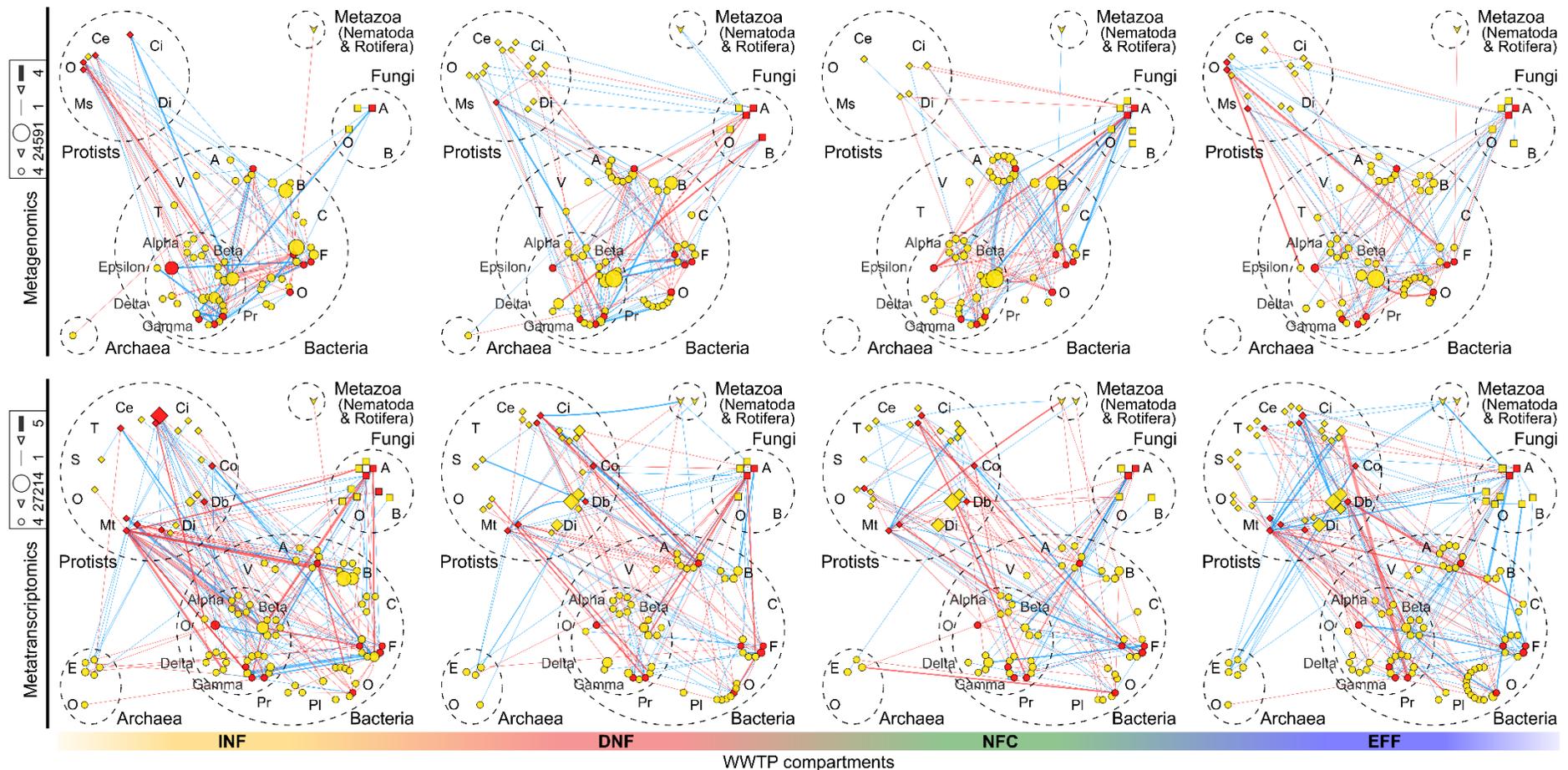
**Supplementary Figure 5:** Microbial community structure and environmental factors across WWTPs. NMDS biplots based on Bray-Curtis dissimilarities showing microbial community composition across WWTP compartments and locations, in association with environmental data. Metagenomic and metatranscriptomic data are shown separately. Samples are color-coded and grouped (ellipses) by compartment. Significant environmental vectors are shown as arrows (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ). Compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water). Environmental vectors: DO = dissolved oxygen, DOC = dissolved organic carbon, HRT = Hydraulic retention time, TN = total nitrogen, TP = total phosphorus.

# Chapter 1



**Supplementary Figure 6:** Microbial community structure and environmental factors in the separate WWTP compartments. NMDS biplots based on Bray-Curtis dissimilarities, showing microbial community composition in association with environmental data for each WWTP compartment. Metagenomic and metatranscriptomic data are shown separately. The distribution of the samples (symbol-coded by WWTP location) is visualized by the ellipses. Significant environmental vectors are shown as arrows. Compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent (treated water). Environmental vectors: TN = total nitrogen.

# Chapter 1



**Supplementary Figure 7:** Co-occurrence networks of parasitic orders in the four WWTP compartments. Networks showing correlations derived from co-occurrence network inferences for each WWTP compartment, based on metagenomic (first row) and metatranscriptomic (second row) data. Only associations that involve parasites are shown. Nodes represent genera grouped at the order level and trait level (red nodes: parasitic taxa; yellow nodes: free-living taxa), with node size proportional to the total number of reads for each order. Edges represent correlations between taxa (blue lines: positive correlations; red lines: negative correlations), with line thickness proportional to the number of genera per order involved. Compartments: INF = inflow (sewage), DNF = denitrification bioreactor, NFC = nitrification bioreactor, EFF = effluent. Abbreviations for Archaea: E = Euryarchaeota, O = Others. Abbreviations for bacteria: A = Actinobacteria, B = Bacteroidetes, C = Chloroflexi, F = Firmicutes, O = Others, PI = Planctomycetes, Pr = Proteobacteria, T = Tenericutes, V = Verrucomicrobia. Abbreviations for Proteobacteria: Alpha = Alphaproteobacteria, Beta = Betaproteobacteria, Gamma = Gammaproteobacteria, Delta = Deltaproteobacteria, O = Others. Abbreviations Fungi: Ascomycota = A, Basidiomycota = B, Others = O. Abbreviations for protists: Ce = Cercozoa (\*including Rosculus), Ci = Ciliophora, Co = Conosa, Db = Discoba, Di = Discosea, Ms = Mesomycetozoa, Mt = Metamonada, O = Others, S = Stramenopiles, T = Tubulinea.

## Chapter 1

**Supplementary Table 1:** Microbial community composition after quality filtering. Total number and relative number (%) of ribosomal reads and OTUs in the metagenomic (rDNA) and metatranscriptomic (rRNA) data of 10 WWTP locations, for prokaryotes (bacteria and Archaea) and eukaryotes (protists, fungi and microscopic

		Metagenomics				Metatranscriptomics			
		Number of reads	[%]	Number of OTUs	[%]	Number of reads	[%]	Number of OTUs	[%]
Eukaryotes	Prokaryotes	479487	~ 94.34	1366	~ 70.16	228569	~ 42.49	1142	~ 60.52
	Protists	23198	~ 4.56	388	~ 19.93	294510	~ 54.75	480	~ 25.44
	Fungi	3763	~ 0.74	126	~ 6.47	9980	~ 1.86	196	~ 10.39
	Metazoa	1802	~ 0.35	67	~ 3.44	4876	~ 0.91	69	~ 3.66

## Chapter 1

**Supplementary Table 2:** Parasitic genera in WWTPs based on both metagenomic and metatranscriptomic data. Overview of all parasitic genera identified in the WWTP samples.

Genera	Microbial community	Genera	Microbial community
<i>Arcobacter</i>	Prokaryotes	<i>Crithidia</i>	Protists
<i>Bacillus</i>	Prokaryotes	<i>Cryptosporidium</i>	Protists
<i>Campylobacter</i>	Prokaryotes	<i>Dientamoeba</i>	Protists
<i>Clostridium</i>	Prokaryotes	<i>Entamoeba</i>	Protists
<i>Corynebacterium</i>	Prokaryotes	<i>Enterobryus</i>	Protists
<i>Enterococcus</i>	Prokaryotes	<i>Enteromonas</i>	Protists
<i>Helicobacter</i>	Prokaryotes	<i>Giardia</i>	Protists
<i>Klebsiella</i>	Prokaryotes	Gregarines_XX	Protists
<i>Legionella</i>	Prokaryotes	<i>Guttulinopsis</i>	Protists
<i>Leptospira</i>	Prokaryotes	<i>Helkesimastix</i>	Protists
<i>Listeria</i>	Prokaryotes	<i>Herpetomonas</i>	Protists
<i>Mycobacterium</i>	Prokaryotes	<i>Hexamita</i>	Protists
<i>Pseudomonas</i>	Prokaryotes	Hexamitinae-Enteromonadida_X	Protists
<i>Salmonella</i>	Prokaryotes	<i>Ichthyophonus</i>	Protists
<i>Vibrio</i>	Prokaryotes	<i>Lacustera</i>	Protists
<i>Yersinia</i>	Prokaryotes	<i>Leishmania</i>	Protists
<i>Acremonium</i>	Fungi	<i>Leptomonas</i>	Protists
<i>Aspergillus</i>	Fungi	<i>Monocystis</i>	Protists
<i>Candida</i>	Fungi	<i>Paratrypanosoma</i>	Protists
<i>Cladophialophora</i>	Fungi	Perkinsida_XXX	Protists
<i>Cryptococcus</i>	Fungi	<i>Phytomonas</i>	Protists
<i>Fusarium</i>	Fungi	<i>Pseudotrichomonas</i>	Protists
<i>Mucor</i>	Fungi	<i>Rhinosporidium</i>	Protists
<i>Ochroconis</i>	Fungi	Rhynosporidae_X	Protists
<i>Penicillium</i>	Fungi	<i>Rosculus</i>	Protists
<i>Rhodotorula</i>	Fungi	<i>Sainouron</i>	Protists
<i>Trichosporon</i>	Fungi	<i>Sappinia</i>	Protists
<i>Ascaris</i>	Metazoa	<i>Sphaerothecum</i>	Protists
<i>Acanthamoeba</i>	Protists	<i>Tetratrichomonas</i>	Protists
<i>Anurofeca</i>	Protists	<i>Trepomonas</i>	Protists
<i>Blastocystis</i>	Protists	<i>Trichomitus</i>	Protists
<i>Blastodinium</i>	Protists	Trichomonadidae_X	Protists
<i>Blechromonas</i>	Protists	<i>Trichomonas</i>	Protists
<i>Copromyxa</i>	Protists	<i>Trimitus</i>	Protists
<i>Creolimax</i>	Protists	Trypanosomatidae_X	Protists

## Chapter 1

**Supplementary Table 3:** Comparing the total number of rDNA and rRNA sequences. Pair-wise comparison of the total number of eukaryotic and prokaryotic sequences for the total community and the parasitic community, contrasting the inflow (INF) with the denitrification bioreactor (DNF), the denitrification bioreactor (DNF) with the nitrification bioreactor (NFC), and the nitrification bioreactor (NFC) with the effluent (EFF). Sign test (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

Data	Microbial community	Data subset	Group 1	Group 2	p-value	
Metagenomics	Eukaryotes	All	INF	DNF	0.00195	**
Metagenomics	Eukaryotes	All	DNF	NFC	1.00000	
Metagenomics	Eukaryotes	All	NFC	EFF	0.00195	**
Metagenomics	Eukaryotes	Parasites	INF	DNF	0.00195	**
Metagenomics	Eukaryotes	Parasites	DNF	NFC	0.75400	
Metagenomics	Eukaryotes	Parasites	NFC	EFF	0.00195	**
Metagenomics	Prokaryotes	All	INF	DNF	0.00195	**
Metagenomics	Prokaryotes	All	DNF	NFC	1.00000	
Metagenomics	Prokaryotes	All	NFC	EFF	0.00195	**
Metagenomics	Prokaryotes	Parasites	INF	DNF	0.00195	**
Metagenomics	Prokaryotes	Parasites	DNF	NFC	1.00000	
Metagenomics	Prokaryotes	Parasites	NFC	EFF	0.00195	**
Metatranscriptomics	Eukaryotes	All	INF	DNF	0.00195	**
Metatranscriptomics	Eukaryotes	All	DNF	NFC	1.00000	
Metatranscriptomics	Eukaryotes	All	NFC	EFF	0.00195	**
Metatranscriptomics	Eukaryotes	Parasites	INF	DNF	0.10900	
Metatranscriptomics	Eukaryotes	Parasites	DNF	NFC	0.75400	
Metatranscriptomics	Eukaryotes	Parasites	NFC	EFF	0.00195	**
Metatranscriptomics	Prokaryotes	All	INF	DNF	0.02150	*
Metatranscriptomics	Prokaryotes	All	DNF	NFC	1.00000	
Metatranscriptomics	Prokaryotes	All	NFC	EFF	0.00195	**
Metatranscriptomics	Prokaryotes	Parasites	INF	DNF	0.10900	
Metatranscriptomics	Prokaryotes	Parasites	DNF	NFC	1.00000	
Metatranscriptomics	Prokaryotes	Parasites	NFC	EFF	0.00195	**

## Chapter 1

**Supplementary Table 4:** Comparing the abundance of parasitic protists between WWTP compartments. Pair-wise comparison of relative abundances of parasitic protist taxa in metagenomic and metatranscriptomic data, contrasting the inflow (INF) with the denitrification bioreactor (DNF), and the nitrification bioreactor (NFC) with the effluent (EFF). Sign test (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

Data	Genera	Group 1	Group 2	p-value	
Metagenomics	<i>Blastocystis</i>	DNF	INF	0.00781	**
Metagenomics	<i>Copromyxa</i>	DNF	INF	0.0625	
Metagenomics	<i>Dientamoeba</i>	DNF	INF	0.5	
Metagenomics	<i>Entamoeba</i>	DNF	INF	1	
Metagenomics	<i>Giardia</i>	DNF	INF	0.625	
Metagenomics	<i>Guttulinopsis</i>	DNF	INF	0.25	
Metagenomics	<i>Rosculus</i>	DNF	INF	0.00391	**
<hr style="border-top: 1px dashed black;"/>					
Metatranscriptomics	<i>Blastocystis</i>	DNF	INF	0.625	
Metatranscriptomics	<i>Copromyxa</i>	DNF	INF	0.00781	**
Metatranscriptomics	<i>Dientamoeba</i>	DNF	INF	0.00195	**
Metatranscriptomics	<i>Entamoeba</i>	DNF	INF	0.00195	**
Metatranscriptomics	<i>Giardia</i>	DNF	INF	0.0215	*
Metatranscriptomics	<i>Guttulinopsis</i>	DNF	INF	0.00195	**
Metatranscriptomics	<i>Rosculus</i>	DNF	INF	0.00195	**
<hr style="border-top: 1px dashed black;"/>					
Metagenomics	<i>Blastocystis</i>	EFF	NFC	0.687	
Metagenomics	<i>Copromyxa</i>	EFF	NFC	1	
Metagenomics	<i>Dientamoeba</i>	EFF	NFC	1	
Metagenomics	<i>Entamoeba</i>	EFF	NFC	1	
Metagenomics	<i>Giardia</i>	EFF	NFC	1	
Metagenomics	<i>Guttulinopsis</i>	EFF	NFC	0.5	
Metagenomics	<i>Rosculus</i>	EFF	NFC	0.25	
<hr style="border-top: 1px dashed black;"/>					
Metatranscriptomics	<i>Blastocystis</i>	EFF	NFC	1	
Metatranscriptomics	<i>Copromyxa</i>	EFF	NFC	0.0313	*
Metatranscriptomics	<i>Dientamoeba</i>	EFF	NFC	1	
Metatranscriptomics	<i>Entamoeba</i>	EFF	NFC	1	
Metatranscriptomics	<i>Giardia</i>	EFF	NFC	0.18	
Metatranscriptomics	<i>Guttulinopsis</i>	EFF	NFC	0.508	
Metatranscriptomics	<i>Rosculus</i>	EFF	NFC	0.0391	*

## Chapter 1

**Supplementary Table 5:** Overview of the most numerous orders in WWTPs. Overview of the most numerous orders shown in Fig. 1. Numbers show their mean relative abundances across all compartments and locations (total N=40 samples) for both rDNA and rRNA data, as well as the absolute difference between these relative DNA and RNA abundances, per order.

Microbial community	Order	DNA [%]	RNA [%]	Absolute difference [%]	
Prokaryotes	Acidimicrobiales	2.36			
Prokaryotes	Bacteroidales	4.46	6.78	2.32	
Prokaryotes	Flavobacteriales	3.82	7.19	3.37	
Prokaryotes	Sphingobacteriales	6.82	7.08	0.26	
Prokaryotes	Bacillales		3.51		
Prokaryotes	Clostridiales	6.14	3.91	2.23	
Prokaryotes	Parcubacteria_XX	2.96	3.42	0.46	
Prokaryotes	Burkholderiales	17.98	12.11	5.87	
Prokaryotes	Proteobacteria	Rhodocyclales	9.1	4.06	5.05
Prokaryotes		Enterobacteriales		4.05	
Prokaryotes		Pseudomonadales	5.4	4.48	0.92
Prokaryotes		Myxococcales	3.3	6.45	3.15
Prokaryotes		Campylobacteriales	4.17		
Prokaryotes	Others	33.49	36.96	3.47	
Eukaryotes	Protists	Cryomonadida	7.08		
	Protists	Imbricatea_X		1.29	
	Protists	Sainouridea		15.43	
	Protists	Cyrtophoria	3.77	1.57	2.2
	Protists	Euplotia	4.12	1.68	2.44
	Protists	Haptoria	3.43		
	Protists	Peritrichia	26.41	10.73	15.68
	Protists	Suctoria	4.16		
	Protists	Euglenida		13.77	
	Protists	Heterolobosea_X		4.13	
	Protists	Kinetoplastida	3.75	23.67	19.92
	Protists	Himatismenida		1.57	
	Protists	Vannellida	2.18	12.86	10.68
	Fungi	Pezizomycotina	7.11	2.79	4.32
	Fungi	Saccharomycotina	5.53		
	Metazoa	Rotifera_X	2.91		
	Eukaryotes	Others	29.56	10.52	19.04

---

## CHAPTER 2

### Root cap is an important determinant of rhizosphere microbiome assembly

Lioba Ruger<sup>1</sup>, Minh Ganther<sup>2</sup>, Jule Freudenthal<sup>1</sup>, Jan Jansa<sup>3</sup>, Anna Heintz-Buschart<sup>4</sup>, Mika Tapio Tarkka<sup>2</sup> and Michael Bonkowski<sup>1</sup>

<sup>1</sup>Terrestrial Ecology, Institute of Zoology, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zulpicher Str 47b, 50674 Koln, Germany

<sup>2</sup>Helmholtz-Centre for Environmental Research, UFZ, Theodor-Lieser-Str 4, 06120 Halle (Saale), Germany

<sup>3</sup>Institute of Microbiology, Academy of Sciences of the Czech Republic, Vıdensk 1083, 14220 Praha 4 - Kr, Czech Republic

<sup>4</sup>Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands

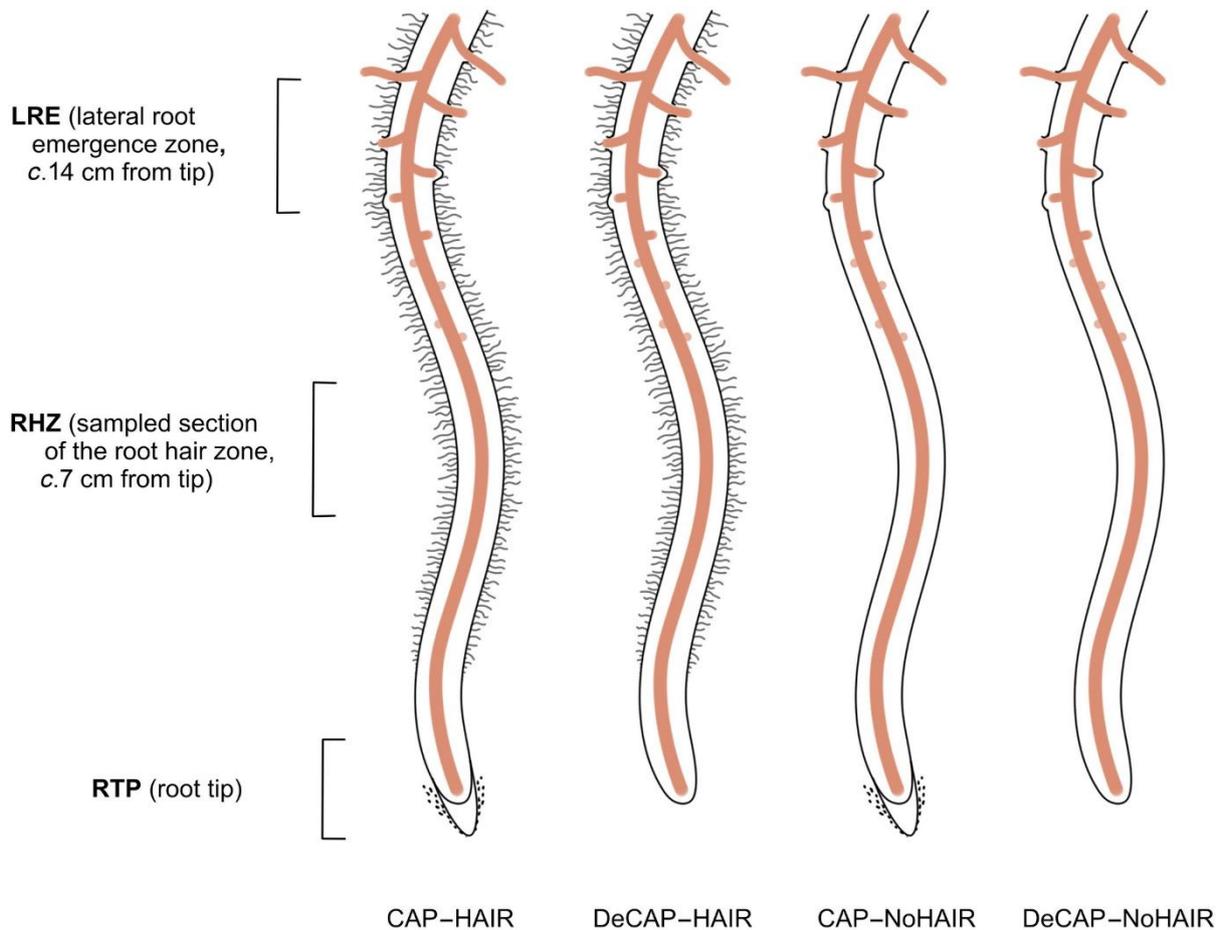
*Keywords: bacteria, maize, molecular control points, protists, rhizosphere microbiome, root cap, root exudates, root hairs*

#### Summary

- Plants impact the development of their rhizosphere microbial communities. It is yet unclear to what extent the root cap and specific root zones contribute to microbial community assembly.
- To test the roles of root caps and root hairs in the establishment of microbiomes along maize roots (*Zea mays*), we compared the composition of prokaryote (archaea and bacteria) and protist (Cercozoa and Endomyxa) microbiomes of intact or decapped primary roots of maize inbred line B73 with its isogenic root hairless (*rth3*) mutant. In addition, we tracked gene expression along the root axis to identify molecular control points for an active microbiome assembly by roots.
- Absence of root caps had stronger effects on microbiome composition than the absence of root hairs and affected microbial community composition also at older root zones and at higher trophic levels (protists). Specific bacterial and cercozoan taxa correlated with root genes involved in immune response.
- Our results indicate a central role of root caps in microbiome assembly with ripple-on effects affecting higher trophic levels and microbiome composition on older root zones.

### 1 Introduction

Root microbiomes are critical to plant health and productivity (Giri *et al.*, 2018; Cantó *et al.*, 2020; Trivedi *et al.*, 2020). As a root grows into the bulk soil, the assembly of its rhizosphere microbiome begins at the root tip and proceeds toward the older root zones (Zelenev *et al.*, 2005; Dupuy & Silk, 2016; R ger *et al.*, 2021). Differences in rhizodeposition at both the tip (Humphris *et al.*, 2005; Benizri *et al.*, 2007) and the root hair zone appear crucial in the microbial assembly process (Robertson-Albertyn *et al.*, 2017; Kelly *et al.*, 2018; Korenblum *et al.*, 2020). The root tip sheds root border cells and actively secretes mucilage, a water-soluble, high-molecular-weight polysaccharide and protein matrix to lubricate root movement and to protect the apical root meristem (Iijima *et al.*, 2000; Nguyen, 2003). Microbial selection occurs mainly through feeding on mucilage (Iijima *et al.*, 2000) and the antimicrobial and signaling molecules it contains (Gu *et al.*, 2016; Hu *et al.*, 2018; Driouich *et al.*, 2021). For example, antimicrobial compounds act against pathogens (Driouich *et al.*, 2013), while microorganisms with glycosyl hydrolases to degrade mucilage are attracted (Amicucci *et al.*, 2019). Also, ethylene (Hahn *et al.*, 2008) and other secondary metabolites likely play active roles in the microbial community assembly process (Haichar *et al.*, 2014; Hawes *et al.*, 2016; Ravanbakhsh *et al.*, 2018; Zhou *et al.*, 2019; Chen *et al.*, 2020). However, the vast majority of primary metabolites, especially sugars, are thought to be passively released directly behind the root cap, where phloem and xylem vessels of the root central cylinder are not yet closed (Farrar *et al.*, 2003; Dennis *et al.*, 2010). This may lead to a massive microbial proliferation in the root hair zone (R ger *et al.*, 2021) and contribute to its role as a hotspot of microbial activity (Blagodatskaya & Kuzyakov, 2015; Zhang *et al.*, 2020). Root hairs play important roles for the uptake of poorly accessible nutrients such as phosphorus and iron. Facilitated by exudation of organic acids and metal-chelating compounds (Yan *et al.*, 2004; Marschner *et al.*, 2011), this activity further modifies the composition of the rhizosphere microbial community. The root hair zone is a region of intensive crosstalk with beneficial and pathogenic microorganisms (Peleg-Grossman *et al.*, 2009; Libault *et al.*, 2010; Poitout *et al.*, 2017). Therefore, it is not surprising that a great number of specific molecules, including flavonoids, coumarins, phenolics, indoles, amino acids, and proteins are released in this area (Bertin *et al.*, 2003; Badri & Vivanco, 2009; Hassan & Mathesius, 2012; Stringlis *et al.*, 2018) and were found to feedback on microbiome assembly and function (Gochnauer *et al.*, 1989; Korenblum *et al.*, 2020; Gebauer *et al.*, 2021). In maize, root development is characterized by gradual merging of the root hair zone with the zone of lateral root emergence (Fig. 1). Cracks around the breakage sites of emerging lateral roots release metabolites that might fuel microbial growth (Jaeger III *et al.*, 1999) and may be vulnerable to pathogen invasion and host infection (Gopalaswamy *et al.*, 2000; Lagopodi *et al.*, 2002; Sprague *et al.*, 2007). Here, lectins and benzoxazinoids around freshly emerged lateral roots were shown to



**Fig. 1:** Four tested treatment combinations and three sampled zones of the root (*Zea mays*). The treatment combinations included roots with intact (CAP) or removed root caps (DeCAP) and roots with typical root hairs (HAIR) or without root hairs (NoHAIR) in the root hair-deficient mutant *rth3*. The three sampled zones included root tip (RTP) comprising the first cm from the root tip, the root hair zone (RHZ) 2 cm of the region below the emergence of the first lateral root, and the region of lateral root emergence (LRE), that is, 2 cm of the region above the first visible lateral root. Root zones are indicated by brackets.

counteract infections and to modify microbiome composition (Sicker *et al.*, 2000; Park *et al.*, 2004; Péret *et al.*, 2009; Cotton *et al.*, 2019).

Beyond root-driven processes, bacterivore protists control the community composition of prokaryotic microbiomes in the rhizosphere (Gao *et al.*, 2019; Dumack *et al.*, 2022). For example, certain phyla of bacterivorous protists are specifically enriched in the maize rhizosphere (Rüger *et al.*, 2021; Taerum *et al.*, 2022), and exert significant top-down control on microbiome assembly and function. Selective feeding by protists strongly shapes rhizosphere bacterial community composition (Jousset *et al.*, 2008; Rosenberg *et al.*, 2009; Jousset & Bonkowski, 2010) and imposes a selection pressure leading to the alteration of microbial functional traits (Jousset *et al.*, 2006; Flues *et al.*, 2017; Xiong *et al.*, 2018; Gao *et al.*, 2019; Amacker *et al.*, 2020; Bahroun *et al.*, 2021). This can reduce deleterious effects of phytopathogens on plant growth (Weidner *et al.*, 2017; Amacker *et al.*, 2020), while the function of mutualists, such as mycorrhizal fungi, can be significantly enhanced (Herdler *et al.*, 2008; Koller *et al.*, 2013; Rozmoš *et al.*, 2021). Overall, protistan feedbacks were shown to alter root metabolite

## Chapter 2

---

profiles and plant stress responses (Kuppardt *et al.*, 2018). They may even change root architecture (Kreuzer *et al.*, 2006), not only by altering bacterial communities, but also by remobilization of nutrients or direct plant-protist interactions (see Bonkowski, 2004).

The assumption that plants actively orchestrate the microbial assembly process (Haichar *et al.*, 2014) implies specific upstream control points of plant gene regulation underlying the crosstalk between plants and microbes (Phillips *et al.*, 2003). Gene expression patterns differ strongly between root zones along the root axis. Generally, transcripts related to growth are enriched in the maize root tips and decline toward the zone of lateral root emergence, where defense-related genes are upregulated (Cesco *et al.*, 2010; Stelpflug *et al.*, 2016; Yu *et al.*, 2021). In particular, the plant immune system is highly regulated and is assumed to play a crucial role as a molecular control point for microbiome assembly (reviewed in Segonzac & Zipfel, 2011; Hacquard *et al.*, 2015). As a first line of defense, immunoreceptors targeting microbial-associated molecular patterns (MAMP) on the root cell surface trigger defenses against specific classes of microorganisms (MAMP-Triggered Immunity; Bittel & Robatzek, 2007). A second line of defense targets microbial effector molecules (Ceulemans *et al.*, 2021) via pathogen recognition (PR) receptor proteins (effector-triggered immunity (ETI), Hacquard *et al.*, 2017). It is in this context that the phytohormones salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) play critical roles in regulating microbial root colonization (Van Loon *et al.*, 2006; Hause & Schaarschmidt, 2009; Lebeis *et al.*, 2015; Chen *et al.*, 2020). Rhizosphere microorganisms can activate plant defense pathways both locally (Hartmann & Schikora, 2012; Brotman *et al.*, 2013) and systemically (Schuhegger *et al.*, 2006; Van Loon *et al.*, 2006; Henkes *et al.*, 2011; Kurth *et al.*, 2014; Pieterse *et al.*, 2014; Verbon *et al.*, 2017) with direct and indirect feedbacks on microbiome assembly (Pieterse & Ton, 2009; Lebeis *et al.*, 2015; Tanaka *et al.*, 2015; Ravanbakhsh *et al.*, 2018; Jones *et al.*, 2019; Chen *et al.*, 2020). Direct effects involve the activation of reactive oxygen species (ROS) and biosynthesis of secondary metabolites (Jacoby *et al.*, 2020; Schütz *et al.*, 2021), such as phenylpropanoids and their derivatives like flavonoids and antimicrobial phytoalexins (Shaw *et al.*, 2006; Steinkellner *et al.*, 2007). Indirectly, systemic responses affect source–sink relationships in plants (Liu *et al.*, 2010; Schultz *et al.*, 2013), with immediate consequences for root C allocation and microbiome assembly (Henkes *et al.*, 2008, 2018).

In view of the complexity of microbiome assembly, our first aim was to determine which of the specific structures (root cap or root hairs) exerts a stronger effect on microbiome assembly along the longitudinal root axis, and whether the absence of root cap or root hair region mutually influences each other's effect on the microbiome. We hypothesized that decapping of the mucilage-releasing root tips and the absence of root hairs in root hair-deficient mutant maize impact prokaryote communities, which in turn affect their protist consumers through feedback mechanisms. After characterizing the main contributors to microbiome assembly along the root axis of maize, we hypothesized that if plants

## Chapter 2

---

actively orchestrated microbiome assembly, it should be possible to identify potential molecular rhizosphere control points by correlations of microbial taxa with the expression patterns of plant (defense) genes. Finally, we hypothesized that specific trophic relationships between prokaryotes and protists are especially relevant for the rhizosphere microbiome structure, and that co-occurrence networks between prokaryotes and protists indicate such relationships.

## 2 Materials and Methods

### 2.1 Experimental setup

The experiment was carried out in a two-factorial design. Wild-type *Zea mays* L. (inbred line B73) plants with intact root caps or with manipulated root caps (CAP vs DeCAP) were compared with root hair-deficient mutants (*root hairless 3*, *rth3*; HAIR vs NoHAIR) with root caps intact or removed, resulting in four (2 × 2) treatment combinations (Fig. 1). The *rth3* mutant is a highly homozygous line (Hochholdinger *et al.*, 2008), exhibiting impaired root hair elongation. Root caps were removed under a dissection microscope with a sterile scalpel as in Humphris *et al.* (2005). Each cap junction was checked and only roots whose caps came off cleanly at the first attempt were used for further experimentation. Prokaryote (bacteria and archaea) and protist (Rhizaria: Cercozoa and Endomyxa) community composition and plant gene expression were analyzed in different root zones. The experiment was set up in rhizoboxes with 36 replicates per treatment combination. Six additional replicates were set up per treatment combination for measurements of root length and diameter to characterize the effect of decapping on root growth. Maize plants were planted in rhizoboxes filled with an agricultural loam soil with a sand : silt : clay, 33 : 48 : 19, harboring its original microbial community (Supporting information Fig. S1; Methods S1). Plants were grown for 6 d in a climate chamber at 12 h : 12 h, day : night (350  $\mu\text{mol m}^{-2} \text{s}^{-1}$  photosynthetically active radiation) at 24°C : 18°C and 65% humidity, until the first roots reached the bottom of rhizoboxes.

### 2.2 Sampling

Three zones along the primary root of each plant were sampled: the root tip (RTP), the root hair zone (RHZ), and the region of lateral root emergence (LRE; Fig. 1). For microbial DNA extraction and subsequent amplicon sequencing of 16S and 18S rRNA gene fragments and quantification of prokaryotes, 0.25 g soil was collected with a sterile spatula from the direct vicinity of each of the three root zones of each plant. Rhizosphere samples were pooled from two plants to ensure better comparability with the pooled root transcriptome data, resulting in 18 replicates for each treatment combination. For root RNA extractions, corresponding root zones were cut out with a sterile scalpel, vortexed in 0.3% NaCl to remove adhering soil, frozen in liquid nitrogen, and stored at -80°C. Roots

## Chapter 2

---

were pooled per rhizobox, resulting in three replicates per treatment before RNA extraction. Two DNA samples from the prokaryote dataset and one RNA sample were excluded from further analysis due to quality issues. Total root systems were scanned (Epson Perfection V700) and primary root length, the length of the primary root including laterals, and average root diameter was analyzed, by WinRHIZO (v.5.0; Regent Instruments, Quebec City, QC, Canada).

### *2.3 Microbial quantification, amplicon sequencing, plant transcriptome sequencing, and data processing*

For soil DNA extraction and purification, the FastDNA SPIN Kit for soil and the GeneClean Spin Kit (MP Biomedicals, Santa Ana, CA, USA) were used, following the manufacturer's instructions. Prokaryote community abundance was determined in extracted DNA by quantitative polymerase chain reaction (qPCR) using the forward primer Eub 338 (5'-ACTCCTACGGGAGGCAGCAG-3') and the reverse primer Eub518 (5'-ATTACCGCGGCTGCTGG-3'; Methods S2).

For sequencing, an *c.* 250-bp long fragment of the prokaryotic V4 region of the 16S rRNA gene was amplified with the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3'; Caporaso *et al.*, 2011) and the reverse primer 806R (5'-GGACTACNVGGGTWTCTAAT-3'; Apprill *et al.*, 2015). The amplicons were double-indexed with Nextera XT indexes to provide unique index combinations per sample. Concentrations of amplicons were measured using Picogreen fluorescence assay before an equimolar mixture was sequenced on a 2x300 MiSeq Illumina platform (Illumina Inc., San Diego, CA, USA) at the Joint Microbiome Facility, Vienna University, Austria. For sequence processing forward and reverse sequence, reads were first paired, quality checked, and filtered. Sequences were then clustered into OTUs at 97% similarity level, identified, resampled to 26 000 sequences per sample, and clustered again. OTUs represented by < 55 reads were discarded.

To amplify a circa 350-bp long fragment of the cercozoan and endomyxan V4 region of the SSU/18S rRNA, a two-step PCR was conducted (Fiore-Donno *et al.*, 2020). In a first PCR, the forward primers S615F\_Cerco (5'-GTAAAAAGCTCGTAGTTG-3') and S615F\_Phyt (5'-GTAAAAARGCTCGTAGTCG-3') and the reverse primer S963R\_Phyt (5'-CAACTTTCGTTCTTGATTTAA-3') were used. In a subsequent semi-nested PCR, using the forward primer S615F\_Cer (5'-GTAAAAARGCTCGTAGTYG-3') and the reverse primer S947R\_Cer (5'-AAGARGAYATCCTTGGTG-3'), both barcoded, samples were indexed (Table S1). The primers specifically target cercozoan and endomyxan DNA as these groups make up a major part of bacterivorous protists in soils (Bates *et al.*, 2012; Burki & Keeling, 2014; Geisen *et al.*, 2015). General Eukaryote primers were not suitable for this study as they have been shown to exclude a significant part of diversity, are highly biased, and amplify a substantial proportion of multicellular organisms (Aslani *et al.*, 2022; Vaulot *et al.*, 2022), and see Lentendu *et al.* (2014) and discussion in (Fiore-Donno *et al.*, 2018). The PCR products were purified and normalized using SequalPrep Normalization Plate Kit

## Chapter 2

---

(Invitrogen) and sequenced on a 2x300 Illumina MiSeq platform at the Cologne Center for Genomics (Cologne, Germany). As for prokaryotes, forward and reverse reads were paired, quality checked, filtered, and clustered at 97% similarity into OTUs. Those represented by < 1000 reads were removed, and remaining OTUs were assigned to taxa. Chimeras were removed, and samples were resampled to 5290 sequences.

For plant transcriptome sequencing, frozen root samples were homogenized in liquid nitrogen using mortar and pestle. Total RNA was isolated using the NucleoSpin RNA Plant kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions and treated with RNase-free DNase (Qiagen). RNA quantity and integrity were determined spectrophotometrically using NanoDrop ND-1000 (Thermo Fisher Scientific, Waltham, MA, USA) and Bioanalyzer (Agilent RNA 6000 Nano kit, Santa Clara, CA, USA). All samples passed quality control with RIN (RNA integrity number) values > 8. Library preparation and sequencing of three independent biological replicates were conducted by Genewiz (Leipzig, Germany) with a strand-specific paired-end 2 × 150-bp design on the Illumina NovaSeq platform. After adapter removal and quality trimming, reads were aligned to maize B73\_RefGen\_v4 genome and assigned to genes.

From prokaryote and Cercozoa sequencing data, rarefaction curves were calculated, to confirm sufficient sequencing depth. OUT richness, Pielou evenness, Shannon diversity, and prokaryote abundance were compared by analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD). Permutational multivariate analysis of variance (PERMANOVA), variance partitioning, nonmetric multidimensional scaling (NMDS), and comparison of group dispersion were used to analyze beta diversity. In order to assess how much variance in beta diversity of prokaryotes could be explained by diversity of their protistan predators, the first two axes of a principal coordinates analysis (PCoA) of cercozoan Bray–Curtis dissimilarities were extracted and used as explanatory variables for Bray–Curtis dissimilarities of prokaryote data in a db-RDA. Plant gene expression data were checked for differentially expressed genes; a gene set enrichment analysis was conducted, as well as a variance partitioning analysis (see Methods S3, S4 for detailed descriptions of sequencing data processing and statistical analyses).

### *2.4 Network analysis*

Co-occurrence network analysis was performed with FlashWeave (v.0.18.0, Tackmann *et al.*, 2019) implemented in Julia (v.1.5.3, Bezanson *et al.*, 2012) to identify patterns in the associations between prokaryotes, protists, and expression levels of 55 significantly differentially expressed plant genes. Gene expression data, root zones, root cap status, and root hair status were integrated as metadata, with root zones included as ordinal scaled factors (increasing from root tip to lateral roots). To reduce

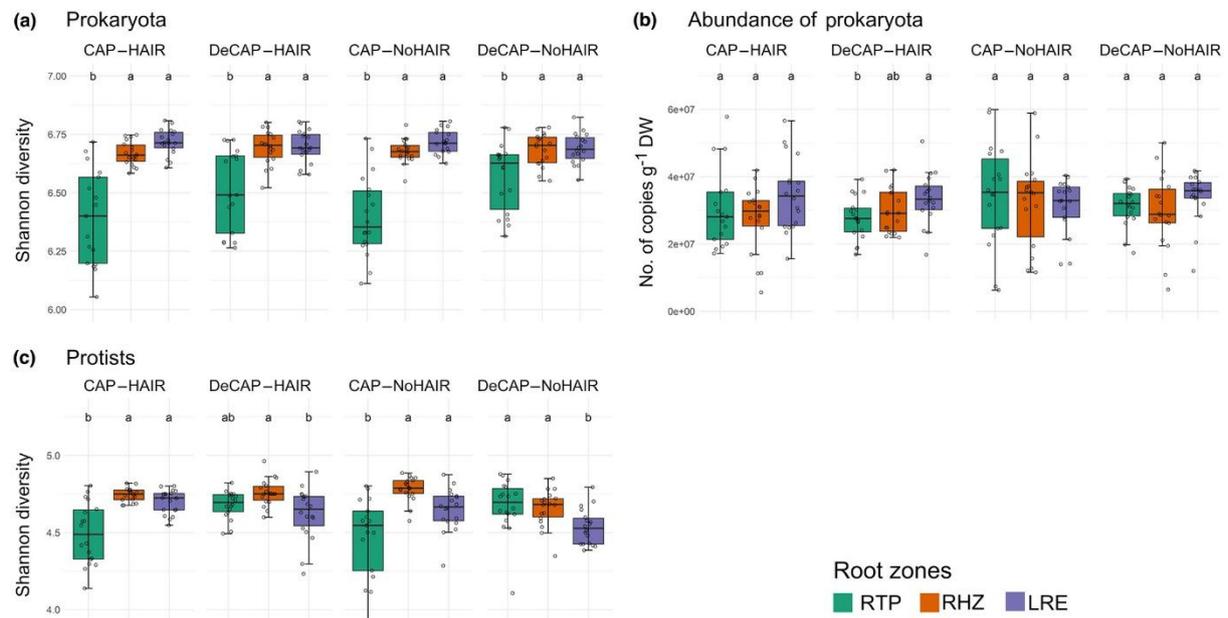
## Chapter 2

spurious edges, rare taxa that only occurred in  $< 1/3$  of all samples were combined into one pseudo taxon (Röttgers & Faust, 2018; Faust, 2021). To account for compositionality of individual datasets (e.g. prokaryotes and protists), datasets were individually normalized by centered log-ratio transformation before networks were calculated. The networks were visualized in Cytoscape (v.3.8.0, Shannon *et al.*, 2003). To cross-compare the abundances of prokaryote and protist OTUs with the gene expression levels in maize roots, the data were combined into one co-occurrence network, which was split into sub-networks including only microbe-microbe interactions or microbe-metadata and metadata-metadata interactions.

### 3 Results

#### 3.1 Influence of decapping on plant transcriptome and root architecture

Low numbers of differentially expressed genes (DEG) and the induction of only few stress response genes in the DEG palette when comparing DeCAP and CAP treatments at RTP indicated successful decapping without harming the root tip (Table S2). Furthermore, decapping caused no obvious differences in root development and had neither a significant effect on primary root length, nor on the total root length including lateral roots. Only the average root diameter increased in DeCAP-HAIR compared with other treatment combinations (Fig. S2; Table S3).



**Fig. 2:** Boxplots illustrating Shannon diversity of (a) prokaryote (bacteria and archaea) and (c) protist (Cercozoa and Endomyxa) communities and (b) prokaryote abundance measured by qPCR (no. of copies  $g^{-1}$  soil dry weight) in the rhizosphere of *Zea mays* in the four treatment combinations with and without root caps (CAP vs DeCAP) and with and without root hairs (HAIR vs NoHAIR) at three root zones: root tip (RTP, green), root hair zone (RHZ, red), and lateral root emergence (LRE, blue). The horizontal line within each box represents the median. Whiskers extend to the lowest and highest scores within 1.5 times the interquartile range from the box. Points represent individual datapoints. Letters indicate significant differences between means (Tukey's HSD).

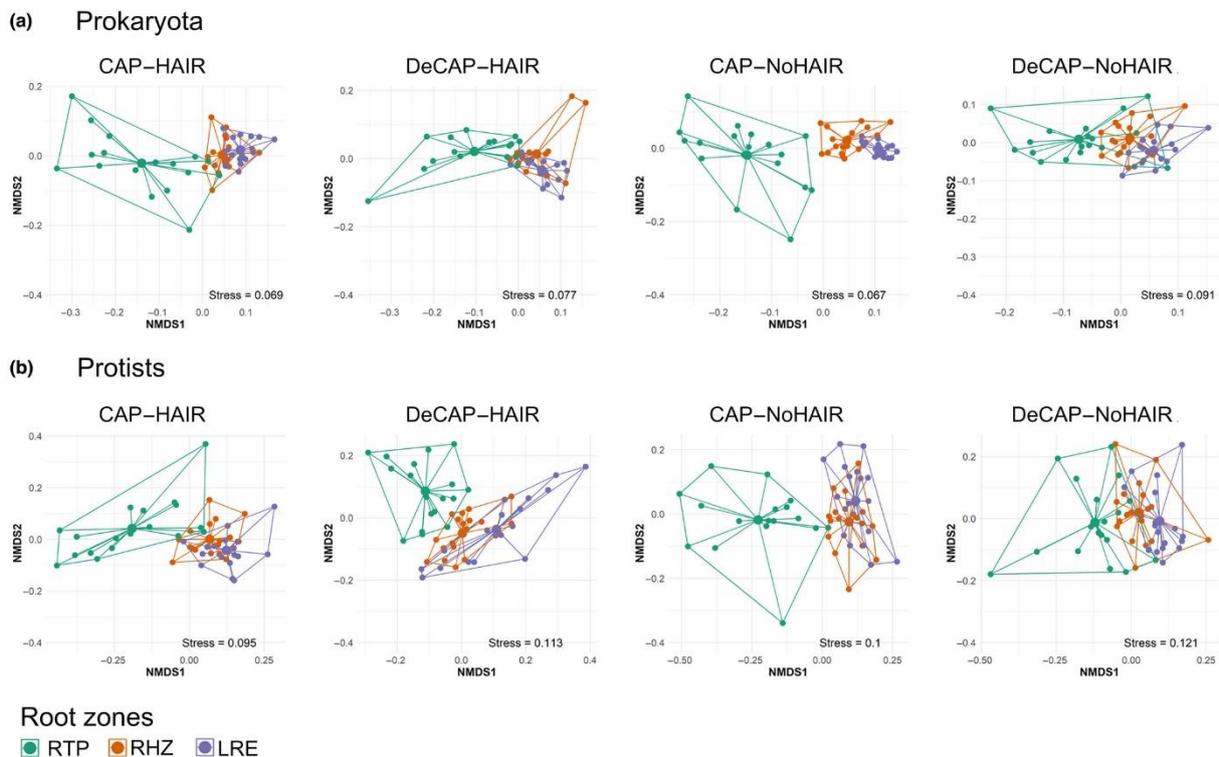
## Chapter 2

### 3.2 Microbial diversity

Prokaryote and protist OTU richness reached saturation and showed excellent coverage of samples (Fig. S3). Their respective local (alpha) diversities were best explained by root zonation, the root cap status only affected prokaryotes, but the interaction of root zonation and the root cap status affected both (Table S4). Prokaryote OTU richness, evenness, and Shannon diversity were lower but highly variable at RTP, and higher but less variable at RHZ and LRE zones (Fig. 2a). Removal of the root cap changed prokaryote Shannon diversity along the root axis, especially in NoHAIR (Table S5). The impact of the tested variables on prokaryote abundance was negligible (Fig. 2b; Tables S6, S7).

Alpha diversity of protists on roots with intact root caps corresponded to patterns in prokaryotes, with the highest variability at RTP (Fig. 2c). Decapping strongly reduced variability of Shannon diversity at RTP and especially NoHAIR had a higher Shannon diversity at RTP compared with LRE (Fig. 2c), because protist evenness decreased at LRE (Fig. S4b).

Irrespective of experimental treatments, prokaryote and protist beta diversity clearly shifted from the root tip toward older root zones (Fig. 3a,b), and variance partitioning explained 7.49% and 8.36% of variation in prokaryote and protist beta diversity between root zones, respectively. The cap explained by far more variance of beta diversity compared with the hair status with 0.76% vs 0.22% in



**Fig. 3:** Nonmetric multidimensional scaling (NMDS) of Bray–Curtis dissimilarities of (a) prokaryote (bacteria and archaea) and (b) protist (Cercozoa and Endomyxa) communities in the rhizosphere of *Zea mays* in the four treatment combinations with and without root caps (CAP vs DeCAP) and with and without root hairs (HAIR vs NoHAIR) at three root zones: root tip (RTP, green), root hair zone (RHZ, red), and lateral root emergence (LRE, blue). Each point represents a microbial community sample, and the color of the point indicates the sample source. Group centroids are marked with a larger dot. The two axes (NMDS1 and NMDS2) represent the two most significant dimensions of variation among the samples.

## Chapter 2

---

prokaryotes and 1.84% vs 0.39% in protists, respectively (Table S8). In general, the communities of both prokaryotes and protists of root tips (RTP) were clearly distinct from RHZ and LRE (Fig. S5), but the effect size was dependent on root cap status (Table S8). Beta diversity dispersal, comparing the differences of community composition among individual root zones, was always higher for protists than prokaryotes and generally highest on root tips, exception the DeCAP–HAIR treatment for prokaryotes (Fig. S5a,b; Table S8). Beta diversity of Cercozoa (first two axes of PCoA) explained 5.8% of prokaryote beta diversity in db-RDA, indicating a significant influence of predation on community structure of bacteria, because when all treatment factors (Root zones, CAP, HAIR) were further included in db-RDA, in total, 10.3% variance of prokaryote beta diversity could be explained.

### *3.3 Plant gene expression patterns*

The three different root zones along the primary root showed clearly distinct patterns of gene expression as shown by PCA, variance partitioning (Fig. S6a,b), and PERMANOVA ( $R^2 = 0.765$ ,  $P < 0.001$ ). The presence or absence of root cap or root hairs had a more subtle influence on gene expression patterns, which were not significant on a global level (PERMANOVA,  $P = 0.071$  and  $P = 0.126$ , respectively).

A highly dynamic transition of gene expression related to growth and development, stress response, metabolism, signaling, and transport characterized the three root zones (Fig. S7; Tables S9, S10). The largest number of differentially expressed genes (DEG) were found between RTP and RHZ (9237,  $P < 0.01$ ,  $|LFC| > 1$ ), followed by RTP vs LRE (6784) and RHZ vs LRE (3173; Table S10). Genes with a specific upregulation at the RTP reflected the functional organization of root tip growth with an enrichment of Gene Ontology (GO) terms ‘cell tip growth’ and ‘plant-type cell wall organization’ (Table S9). However, genes indicative of stress responses were more highly expressed in RTP than in other zones, reflected by enriched GO terms ‘response to karrikin’ and ‘response to water deprivation’. The RHZ was characterized by higher expression of genes related to exudation, nitrate, and water transport (‘nicotianamine biosynthetic process’, ‘nitrate transport’, and ‘water transport’), as well as a range of genes involved in defense response and signaling (‘response to jasmonic acid’, ‘phenylpropanoid biosynthetic process’, and ‘response to wounding’). In the LRE zone beta-glucosidases, expansins, and cellulases were upregulated, indicative of cell wall-structuring functions, as well as several *NRT1/PTR* transporters and nicotianamine transporters *YSL2*, pointing toward transport and exudation.



## Chapter 2

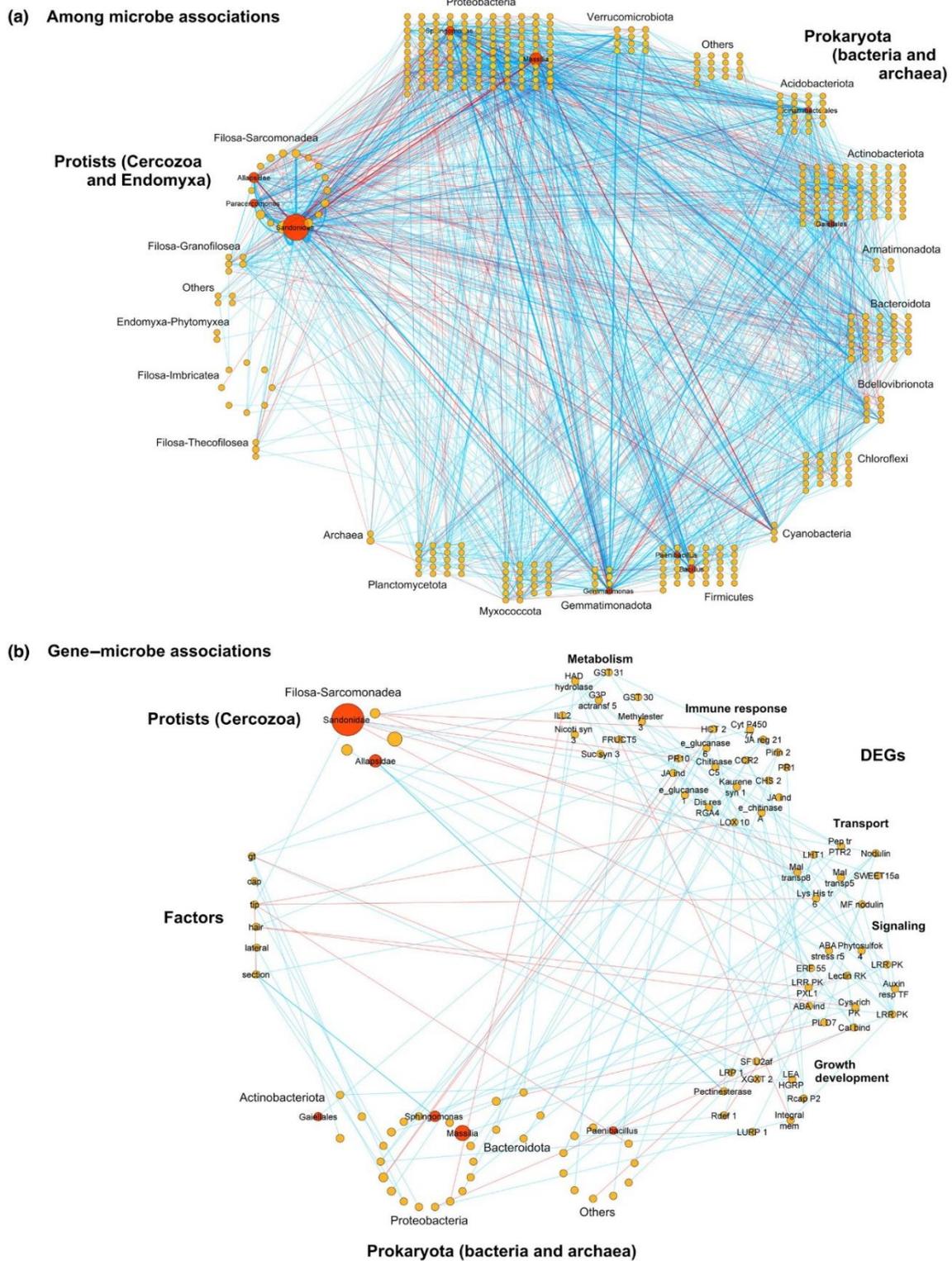
---

### *3.4 Co-occurrence of prokaryotes and protists and their associations with DEGs*

To cross-compare the abundances of prokaryote and protist OTUs with the gene expression levels in maize roots, the data were combined into a co-occurrence network (1910 edges). For better visualization, subnetworks including only microbe–microbe interactions (Fig. 5a) and correlations of microbes, metadata, and root gene expression (Fig. 5b) were separated. The microbe–microbe association network (Fig. 5a) was characterized by a particularly high number of edges among prokaryotes (1294) and between prokaryotes and protists (397), indicating potential interactions due to competition and predation. Especially, bacterivore cercozoan taxa in the class Filosa-Sarcomonadea showed negative associations with various prokaryote taxa. Among these protists, the Sandonidae (Cercozoa) showed an explicitly high proportion of negative associations with different bacterial genera (positive: 49, negative: 31), especially with *Massilia* (edge weight = 3). By contrast, among prokaryotes, the number of positive edges exceeded negative ones almost fivefold (positive: 1061, negative: 233). Links between the treatment factors and plant gene expression levels or microbes were scarce (18).

Among prokaryotes, 32 genera were associated with the expression level of specific genes. For example, the highly abundant genus *Massilia* (Oxalobacteraceae, Burkholderiales, and Betaproteobacteria) with a degree centrality (DC) of 88 was positively linked with the expression of *Cys-rich PK* (Cysteine-rich protein kinase, that is, MAMP-Triggered Immunity signaling). Another genus of the Burkholderiales was positively associated with the expression of *JA ind* (encoding jasmonate-induced protein, immune response). The genus *Lysobacter* (Xanthomonadales, Gammaproteobacteria) with a DC of 44 showed negative associations with the expression of *HCT2* and *JA ind* (encoding hydroxycinnamoyltransferase 2 and jasmonate-induced protein, immune response), and a genus of the Gaiellales (Actinobacteria) with a high DC (52) was positively linked with the expression of *JA reg 21* (jasmonate-regulated 21, immune response). Within Cercozoa, genera of five families were linked to the expression levels of specific genes. The Sandonidae (DC of 112), marked by the highest number of reads, were associated with the expression of various plant genes. For example, a negative association was found with the expression of *ERF 55* (Ethylene-responsive transcription factor 55, signaling), and positive associations with expression levels of genes encoding *JA ind* and *e\_glucanase 1* (jasmonate-induced protein and endoglucanase 1, immune response). The Allapsidae (DC of 54) showed positive associations to plant genes involved in growth and development, such as *LRP 1* (lateral root primordium 1), and especially strong (edge weight = 2) with *Pectinesterase*.

## Chapter 2



**Fig. 5:** Co-occurrence networks illustrating (a) associations among and between prokaryotes (bacteria and archaea) and protists (Cercozoa and Endomyxa), and (b) between the microorganisms, 55 selected DEGs of *Zea mays* and the factors genotype (gt; i.e. HAIR vs NoHAIR), and cap (i.e. CAP vs DeCAP) and root zones (section) with root tip (tip), root hair zone (hair), and lateral root zone (lateral). Nodes represent taxa summed at genus level and grouped at class (protists) or phylum level (prokaryota) for better visualization. Nodes representing genes are grouped into five categories regarding their function. Nodes with a degree centrality  $\geq 50$  are highlighted (orange). The size of nodes indicates the number of reads (normalized). The color of edges indicates whether an association is positive (blue) or negative (red). The edge width reflects the edge weight, that is, the numbers of associations between taxa or metadata.

### 4 Discussion

We explored the roles of root border cells and mucilage secretion (CAP vs DeCAP) and root hairs (HAIR vs NoHAIR) on the assembly of rhizosphere microbial communities along the primary root axis of maize. At the same time, we examined associations of microbial community composition with root gene expression, indicative of rhizosphere control points (*sensu* Phillips & Strong, 2003). Finally, we investigated the effects of a dominant group of protistan predators (Cercozoa and Endomyxa) on the community structure of the prokaryote microbiomes and the effect of prokaryotes on protists. As hypothesized, root cap removal affected microbiome assembly patterns, particularly at root tips. Despite CAP and HAIR treatments, rhizosphere microbiome composition was still mainly driven by the microbial succession from root tips to older root zones. In accordance with R ger *et al.* (2021), variability of local communities (alpha diversity) and communities between individual roots (beta diversity) of prokaryotes and of their protistan predators was significantly higher at root tips compared with older root zones (Figs 2, S5), assuming randomness through priority effects of early colonizers that initially leads to dominance of different taxa on different root tips (Chase, 2003; Fukami, 2015; Attia *et al.*, 2022). The reduced variability of all components of alpha diversity (i.e. OTU richness, evenness, and Shannon index) at RHZ and LRE is a strong indication of the fast formation of a distinct microbiome along the root axis and is further corroborated by reduced dispersal of beta diversity from RTP to RHZ and LRE (Fig. S5). The increased Shannon diversity at RHZ and LRE was due to both, increased numbers of different taxa (i.e. OTU richness), but with reduced dominance of single taxa (i.e. enhanced evenness; Fig. S4). At first sight, this pattern appears counterintuitive, as one would expect fewer, specialized taxa gaining dominance during the assembly process if plants favor certain taxa over others. Instead, increased evenness and taxon richness resemble typical outcomes of predation, where the fastest growing, dominant taxa are preferentially consumed and greater numbers of competitive subordinate species can coexist (Rosenberg *et al.*, 2009; Karako  *et al.*, 2020). The structuring impact of predation on the prokaryote microbiome was supported by a high proportion of bacterial beta diversity explained by cercozoa in a constrained (db-RDA) analysis.

CAP and HAIR treatments modified microbiome assembly in distinct ways. The absence of root caps had far stronger effects than the absence of root hairs, which would imply that root caps are more important than root hairs for microbiome assembly. Microbial communities of DeCAP–HAIR differed from other treatments first of all by reduced abundance of prokaryotes at root tips, likely a result of decreased availability of mucilage for microbial growth (Benizri *et al.*, 2007). Taxon richness, evenness, and Shannon diversity of prokaryotes, however, appeared rather unaffected. Instead, the DeCAP–HAIR treatment affected more strongly the diversity of protists. Apparently, the reduced availability of prey via decapping reduced the variability of protist alpha diversity at RTP. It furthermore caused ripple effects that were still noticeable at LRE: Reduced evenness and decreased Shannon diversity, which

## Chapter 2

---

resulted in far higher variability of beta diversity at LRE than on roots with caps (Fig. S5). The absence of root hairs led to less distinct differences in beta diversity between root tips and older root zones in communities of both prokaryotes and protists, and only when the root caps had been removed (DeCAP–NoHAIR, Fig. 3). This suggests that the minor differences in bacterial community composition that have been associated with NoHAIR maize (Gebauer *et al.*, 2021) or barley (Robertson-Albertyn *et al.*, 2017) are amplified by the absence of root cap, supporting the root cap's importance as the first selector and amplifier of rhizosphere compatible taxa. Rhizosphere community assembly processes at the tip feedback on root hair zones.

For root gene expression profiles, root cap removal also attenuated patterning of root zonation, evident by smaller numbers of differentially expressed genes between root zones. In accordance with the patterns of microbiome assembly, the changes in root gene expression after cap removal were most prominent in the root tip. Genes relating to pathways of water/exudation (e.g. *ABA stress ripening 5*, *nicotianamine synthases*) and defense response, especially of the phenylpropanoid-flavonoid biosynthesis pathway, were upregulated after decapping. Plant-derived flavonoids are known to mediate interactions between plant host and both symbiotic as well as pathogenic microorganisms (Treutter, 2005; Mierziak *et al.*, 2014; Block *et al.*, 2018). For example, Yu *et al.* (2021) showed how flavone production promotes the enrichment of bacteria from Oxalobacteraceae, like *Massilia* in the maize rhizosphere, and their data suggest that this led to stimulation of maize growth and enhanced nitrogen acquisition.

Similarly, like flavonoids, coumarins (phenylpropanoid pathway) were found to alter microbiome composition through strain-specific antimicrobial effects (Stringlis *et al.*, 2018; Voges *et al.*, 2019). Specifically, benzoxazinoids play an acknowledged role in fine-tuning microbial communities of maize roots (Cotton *et al.*, 2019; Kudjordjie *et al.*, 2019; Schütz *et al.*, 2021). Other phenylpropanoids such as chlorogenic acid were shown to improve plant resistance against herbivores and pathogens (Leiss *et al.*, 2009). Furthermore, gene functions relating to plant hormone signaling via ABA, SA, and JA, which also play a role in microbe assembly (Jacobsen *et al.*, 2021), were enriched after decapping. The upregulation of such genes – related to defense and immunity – has been associated with basal root zones (Stelpflug *et al.*, 2016). At the root tip, this could assist to reduce the initial random assembly by more deterministic processes, similar to those known for older root zones, thus strengthening microbiome assembly at subsequent root zones. In addition to the observed expression patterns of defense and immunity-related genes, enhanced gene expression of nicotianamine synthases and mineral nutrient transporters indicates a change in root exudation. Nicotianamine synthases are involved in the biosynthesis of phytosiderophores (Mizuno *et al.*, 2003) to facilitate iron and zinc mobilization and plant uptake (Wirén *et al.*, 1996). The iron nutritional status of the plant is connected to root colonization of beneficial rhizobacteria, such as *Paenibacillus polyxyma* and *Bacillus subtilis*

## Chapter 2

---

(Zhang *et al.*, 2009; Zhou *et al.*, 2016), which enhance plant resistance to microbial pathogens by competing for iron (Verbon *et al.*, 2017). Consequently, subtle defense responses at the root tip in concert with changes in exudate supply and composition appear to have a considerable influence on the assembly of the rhizosphere microbiome, even at more distant root zones. Changes in gene expression were not indicative of heightened stress or accompanied by a stark reduction in growth; therefore, we expect that the influence of root cap removal has a direct effect on the microbial community assembly.

Presence of root hairs had a surprisingly marginal effect on the root zonal distribution of the rhizosphere microbiome as well as on root gene expression. The absence of associations between prokaryotes or protists with the factor 'hair' in the network (Fig. 5b) together with the marginal effects of the NoHAIR treatment on microbiome assembly sustain the assumption of R ger *et al.* (2021) that the root hair zone might play a rather small regulatory role in the assembly of the microbiome in maize. This is surprising, as root hairs were shown to significantly enhance the carbon input into the rhizosphere (Holz *et al.*, 2018), but the lack of root hairs did not reduce prokaryote abundance in our experiment (Fig. 2b). Potentially, the carbon input by root hairs is mainly used as substrate to fuel enzyme production for microbial nutrient mining as proposed by Zhang *et al.* (2020) and has little selective effect because all fast-growing copiotrophic rhizobacteria are stimulated simultaneously (R ger *et al.*, 2021). At the level of root gene expression, the minor impact of *rth3* mutants – exerted mostly at the level of cell wall biosynthesis and organization-related gene expression – was confirmed for maize root systems at four-leaf stadium (Ganther *et al.*, 2021). The significant upregulation of genes involved in nitrate transport and signaling in RHZ and LRE, such as *NRT1* together with *sugar transporter 1*, might indicate a stronger role of these root zones in guiding sink–source allocation patterns (Remans *et al.*, 2006; Krouk *et al.*, 2010; Schultz *et al.*, 2013; Wang *et al.*, 2019). This sink–source allocation may be linked to trophic relationships between protists and prokaryotes (Kuikman & Van Veen, 1989; Kuikman *et al.*, 1991), considering the constant release of nitrogen by protists from consumed microbial biomass (Clarholm, 1985), their stimulation of bacterial turnover, and nitrification in the rhizosphere (Bonkowski *et al.*, 2000; Bonkowski, 2004), as well as on auxin balance in roots (Krome *et al.*, 2010).

Only in the DeCAP–NoHAIR treatment, Shannon diversity of protists decreased significantly from root tip to base, yet in the DeCAP–HAIR treatment, almost similar patterns were found. The low alpha diversity at older root zones resulted from a decrease in evenness (Fig. S4), indicating enhanced dominance of specific cercozoan taxa. Apparently, the structuring effect that the removal of the root cap had on protist communities, was enhanced by the lack of root hairs, but primarily exerted at the region of LRE.

## Chapter 2

---

Such an interaction between root cap and root hairs was not found for plant gene expression. The NoHAIR treatment did not show any globally enriched GO terms, but induced rootzone-specific effects at the RTP, which was unexpected considering that root hair formation occurs from the root elongation zone upward (Bibikova & Gilroy, 2003). Root hairs increase the depletion zone for immobile nutrients such as P and K (Jungk, 2001) and facilitate the spatial diffusion of exudates such as organic acids, siderophores, or exoenzymes (Jungk, 2001; Marschner *et al.*, 2011; Bilyera *et al.*, 2022). In line with this, genes related to nutrient uptake, predominantly iron, and nitrate were upregulated in NoHAIR-RTP, suggesting nutrient depletion at root tips. How processes in the root hair zone are expressed in the root tip remains unclear at present. But, reactive oxygen species (ROS)-associated signaling has been implicated in regulation between the developmental zones of the root (Yamada *et al.*, 2020), and interestingly, NoHAIR caused in RTP a limited but significant induction of genes related to hypersensitive response, which is associated with ROS formation (Hacquard *et al.*, 2015). Taken together, the initial microbiome assembly at the root tip is highly variable and prone to priority effects (see R ger *et al.*, 2021). Therefore, it would be most advantageous if the plants would gain better control over the community assembly process already at root tips. Prime candidates of molecular rhizosphere control points are genes involved in plant defense and stress responses. Plant gene expression indicates roles of jasmonate and ethylene signals triggered by immune response and suggests pleiotropic effects on microbiome assembly through the expression of genes related to ROS signaling and flavonoid production. As the absence of root hairs caused an extremely low impact on root gene expression, in this experiment as well as in earlier studies (Ganther *et al.*, 2021, 2022), and the plants were not grown under nutrient deficiency that could potentially enhance their dependency on root hairs (Bienert *et al.*, 2021), pleiotropic effects by the *rth3* mutation appear unlikely.

Network analysis confirmed significant correlations of root gene expression with microbial taxa that could be indicative of molecular control points in plant roots. Feedbacks between bacterial and plant signaling through systemic changes in root defense and exudation may significantly contribute to microbiome assembly in the rhizosphere. For example, bacterial quorum-sensing molecules were shown to activate systemic induced resistance in tomato through the induction of the SA- and ET-dependent defense genes, PR1 and chitinases (Schuhegger *et al.*, 2006).

*Lysobacter*, a ubiquitous bacterial genus, of which several species carry potential plant protective biocontrol traits (Hayward *et al.*, 2010), was negatively linked with the gene expression of *jasmonate-induced protein and hydroxycinnamoyl transferase 2* (Sullivan & Zarnowski, 2011). Both plant genes are associated with immune responses. A high abundance of *Lysobacter* might lead to a reduced expression of specific genes involved in defense (negative correlation) possibly through antimicrobial, plant-beneficial effects or, on the contrary, the abundance of *Lysobacter* might be controlled by the plant immune response.

## Chapter 2

---

A bacterial taxon of the common plant-associated Burkholderiales (Estrada-De Los Santos *et al.*, 2001), that contain beneficial and plant-pathogenic species (Compant *et al.*, 2008; Suarez-Moreno *et al.*, 2012), showed a positive association with gene expression of *jasmonate-induced protein*, potentially inducing a plant immune response.

Further bacterial and also cercozoan taxa correlated to genes with functions associated with transport, signaling, metabolism, development, and growth. Remarkably, there was a negative association between a cercozoan taxon in Sandonidae and the expression level of *ethylene-responsive transcription factor 55*. ET response factors play regulatory roles in stress signaling with likely feedbacks on microbiome assembly (Müller & Munné-Bosch, 2015; Ravanbakhsh *et al.*, 2018; Chen *et al.*, 2020). Another negative association occurred in DeCAP treatments between a protist in the genus *Cercomonas* and *pathogenesis-related protein 10* (PR10) that plays an important role in ETI host resistance (Chen *et al.*, 2010). These results corroborate findings of Kuppardt *et al.* (2018) who demonstrated that protists decreased typical plant stress responses, such as metabolites connected to the phenolic metabolism in the maize rhizosphere.

A protist taxon in the Allapsidae was positively linked to the expression levels of *Pectinesterase* and *lateral root primordium 1*, both genes involved in plant growth and development. Beneficial effects of protists on root growth have been repeatedly reported and are thought to work directly through remobilization of nutrients, as well as indirectly by changing the bacterial community composition (Bonkowski, 2004; Krome *et al.*, 2009; Weidner *et al.*, 2017). As the respective Allapsid was also highly associated with various prokaryote genera, direct effects on root gene expressions are as likely as indirect effects via changes in bacterial microbiome composition. Overall, our data indicate cross-communication between plants and microbes at different trophic levels in the rhizosphere.

Besides plant–microbe interactions, interactions among microbes significantly drive the self-organization of the rhizosphere microbiome (Bonkowski *et al.*, 2021) – this was indicated by comparable assembly patterns of prokaryotes and their protistan predators along the longitudinal root axis and the particularly high number of associations among microorganisms in co-occurrence networks. Such interactions can be facilitative or mutualistic, especially in the presence of predators at high resource supply from rhizodeposition (Nakajima & Kurihara, 1994; Leibold, 1996), because predator preferences are directed toward less defended bacterial taxa and benefit their grazing-resistant competitors (Jousset *et al.*, 2010; Flues *et al.*, 2017). The large proportion of negative edges between Cercozoa, especially Sandonidae, and specific prokaryotes are likely indicative of consumer relationships, and the negative edges among prokaryotic taxa emphasized the role of competition in community assembly.

### 5 Concluding remarks

Overall, the absence of root caps had by far stronger effects on microbiome composition than the absence of root hairs. Decapping affected microbial community composition at older root zones (bacteria and archaea) as well as higher trophic levels (protists). The absence of root hairs had surprisingly small effects on microbiome assembly, which corresponded well with the low levels of differentially expressed genes in NoHAIR treatments. The bacterial genus *Massilia* (Oxalobacteraceae, Burkholderiales) held a central position in the microbiota network, and Burkholderiales together with other bacterial genera like *Lysobacter* showed associations with plant immune response genes. Community structure of cercozoan protists explained almost as much variation of prokaryote community turnover as the influence of different root zones and manipulations of the root cap and the root hair region. However, the role of heterotrophic protists in shaping the plant microbiome may not only be limited to predator–prey interactions, as specific cercozoan taxa were directly associated with plant immune responses. Overall, our results indicate a central role of root tips in microbiome assembly with ripple-on effects affecting higher trophic levels as well as microbial succession on older root zones.

### Author contributions

MB, MTT, MG and LR conceived the study and planned the experiment. MG and LR performed the experiments and analyzed the data together with JF and AH-B. Analysis of prokaryota was conducted by JJ. All authors wrote the manuscript. MG and LR contributed equally to this work.

### Data availability

Raw RNA-Seq reads and protist and prokaryotic sequences were deposited at the NCBI Sequence Read Archive (SRA) under the BioProject PRJNA77586.

### Competing interests

None declared.

### Acknowledgments

We acknowledge Caroline Marcon and Frank Hochholdinger (University of Bonn) for providing the maize seeds used in this experiment. We thank two unknown referees for their constructive comments. This project was carried out in the framework of the priority program 2089 ‘Rhizosphere spatiotemporal organization – a key to rhizosphere functions’ funded by the DFG (project nos. 403635931 and 403641192). Further support for prokaryote sequencing was granted by Czech Science Foundation (project no. 21-07275S). Open Access funding enabled and organized by Projekt DEAL.

### References

- Amacker N, Gao Z, Agaras BC, Latz E, Kowalchuk GA, Valverde CF, Jousset A, Weidner S. 2020.** Biocontrol traits correlate with resistance to predation by protists in soil Pseudomonads. *Frontiers in Microbiology* 11: 614194.
- Amicucci MJ, Galermo AG, Guerrero A, Treves G, Nandita E, Kailemia MJ, Higdon SM, Pozzo T, Labavitch JM, Bennett AB *et al.* 2019.** Strategy for structural elucidation of polysaccharides: elucidation of a maize mucilage that harbors diazotrophic bacteria. *Analytical Chemistry* 91: 7254–7265.
- Aprill A, McNally S, Parsons R, Weber L. 2015.** Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* 75: 129–137.
- Aslani F, Geisen S, Ning D, Tedersoo L, Bahram M. 2022.** Towards revealing the global diversity and community assembly of soil eukaryotes. *Ecology Letters* 25: 65–76.
- Attia S, Russel J, Mortensen MS, Madsen JS, Sørensen SJ. 2022.** Unexpected diversity among small-scale sample replicates of defined plant root compartments. *The ISME Journal* 16: 997–1003.
- Badri DV, Vivanco JM. 2009.** Regulation and function of root exudates. *Plant, Cell & Environment* 32: 666–681.
- Bahrour A, Jousset A, Mrabet M, Mhamdi R, Mhadhbi H. 2021.** Protists modulate *Fusarium* root rot suppression by beneficial bacteria. *Applied Soil Ecology* 168: 104158.
- Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N. 2012.** Global biogeography of highly diverse protistan communities in soil. *The ISME Journal* 7: 652–659.
- Benizri E, Nguyen C, Piutti S, Slezack-Deschaumes S, Philippot L. 2007.** Additions of maize root mucilage to soil changed the structure of the bacterial community. *Soil Biology and Biochemistry* 39: 1230–1233.
- Bertin C, Yang X, Weston LA. 2003.** The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* 256: 67–83.
- Bezanson J, Karpinski S, Shah VB, Edelman A. 2012.** Julia: a fast dynamic language for technical computing. *arXiv*. doi: 10.48550/arXiv.1209.5145.
- Bibikova T, Gilroy S. 2003.** Root hair development. *Journal of Plant Growth Regulation* 21: 383–415.
- Bienert MD, Werner LM, Wimmer MA, Bienert GP. 2021.** Root hairs: the villi of plants. *Biochemical Society Transactions* 49: 1133–1146.
- Bilyera N, Hummel C, Daudin G, Santangeli M, Zhang X, Santner J, Lippold E, Schlüter S, Bertrand I, Wenzel W *et al.* 2022.** Co-localised phosphorus mobilization processes in the rhizosphere of field-grown maize jointly contribute to plant nutrition. *Soil Biology and Biochemistry* 165: 108497.
- Bittel P, Robatzek S. 2007.** Microbe-associated molecular patterns (MAMPs) probe plant immunity. *Current Opinion in Plant Biology* 10: 335–341.
- Blagodatskaya E, Kuzyakov Y. 2015.** Microbial hotspots and hot moments in soil: concept & review. *Soil Biology and Biochemistry* 83: 184–199.
- Block AK, Vaughan MM, Schmelz EA, Christensen SA. 2018.** Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). *Planta* 249: 21–30.
- Bonkowski M. 2004.** Protozoa and plant growth: the microbial loop in soil revisited. *New Phytologist* 162: 617–631.
- Bonkowski M, Griffiths B, Scrimgeour C. 2000.** Substrate heterogeneity and microfauna in soil organic ‘hotspots’ as determinants of nitrogen capture and growth of ryegrass. *Applied Soil Ecology* 14: 37–53.
- Bonkowski M, Tarkka M, Razavi BS, Schmidt H, Blagodatskaya E, Koller R, Yu P, Knief C, Hochholdinger F, Vetterlein D. 2021.** Spatiotemporal dynamics of maize (*Zea mays* L.) root growth and its potential consequences for the assembly of the rhizosphere microbiota. *Frontiers in Microbiology* 12: 619499.
- Brotman Y, Landau U, Takayuki T, Fernie AR, Chet I, Viterbo A, Willmitzer L. 2013.** Trichoderma-plant root colonization: escaping early plant defense responses and activation of the antioxidant machinery for saline stress tolerance. *PLoS Pathogens* 9: e1003221.
- Burki F, Keeling PJ. 2014.** Rhizaria. *Current Biology* 24: R103–R107.
- Cantó CF, Simonin M, King E, Moulin L, Bennett MJ, Castrillo G, Laplaze L. 2020.** An extended root phenotype: the rhizosphere, its formation and impacts on plant fitness. *The Plant Journal* 103: 951–964.
- Caporaso JG, Lauber CL, Walters WA, Berg-lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011.** Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA* 108: 4516–4522.
- Cesco S, Neumann G, Tomasi N, Pinton R, Weisskopf L. 2010.** Release of plant-borne flavonoids into the rhizosphere and their role in plant nutrition. *Plant and Soil* 329: 1–25.
- Ceulemans E, Ibrahim HMM, De Coninck B, Goossens A. 2021.** Pathogen effectors: exploiting the promiscuity of plant signaling hubs. *Trends in Plant Science* 26: 780–795.
- Chase JM. 2003.** Community assembly: when should history matter? *Oecologia* 136: 489–498.

## Chapter 2

---

- Chen Y, Bonkowski M, Shen Y, Griffiths BS, Jiang Y, Wang X, Sun B. 2020. Root ethylene mediates rhizosphere microbial community reconstruction when chemically detecting cyanide produced by neighbouring plants. *Microbiome* 8: 4.
- Chen Z-Y, Brown RL, Damann KE, Cleveland TE. 2010. PR10 expression in maize and its effect on host resistance against *Aspergillus flavus* infection and aflatoxin production. *Molecular Plant Pathology* 11: 69–81.
- Clarholm M. 1985. Interactions of bacteria, protozoa and plants leading to mineralization of soil nitrogen. *Soil Biology and Biochemistry* 17: 181–187.
- Compant S, Nowak J, Coenye T, Clement C, Barka EA. 2008. Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiology Reviews* 32: 607–626.
- Cotton TEA, Pétriacq P, Cameron DD, Meselmani MA, Schwarzenbacher R, Rolfe SA, Ton J. 2019. Metabolic regulation of the maize rhizobiome by benzoxazinoids. *The ISME Journal* 13: 1647–1658.
- Dennis PG, Miller AJ, Hirsch PR. 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology* 72: 313–327.
- Driouich A, Follet-Gueye M-L, Vicré-Gibouin M, Hawes M. 2013. Root border cells and secretions as critical elements in plant host defense. *Plant Biology* 16: 489–495.
- Driouich A, Gaudry A, Pawlak B, Moore JP. 2021. Root cap-derived cells and mucilage: a protective network at the root tip. *Protoplasma* 258: 1179–1185.
- Dumack K, Feng K, Flues S, Sapp M, Schreiter S, Grosch R, Rose LE, Deng Y, Smalla K, Bonkowski M. 2022. What drives the assembly of plant-associated protist microbiomes? Investigating the effects of crop species, soil type and bacterial microbiomes. *Protist* 173: 125913.
- Dupuy LX, Silk WK. 2016. Mechanisms of early microbial establishment on growing root surfaces. *Vadose Zone Journal* 15: 1–13.
- Estrada-De Los Santos P, Bustillos-Cristales R, Caballero-Mellado J. 2001. *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Applied and Environmental Microbiology* 67: 2790–2798.
- Farrar J, Hawes M, Jones D, Lindow S. 2003. How roots control the flux of carbon to the rhizosphere. *Ecological Society of America* 84: 827–837.
- Faust K. 2021. Open challenges for microbial network construction and analysis. *The ISME Journal* 15: 3111–3118.
- Fiore-donno AM, Richter-Heitmann T, Bonkowski M. 2020. Contrasting responses of protistan plant parasites and phagotrophs to ecosystems, land management and soil properties. *Frontiers in Microbiology* 11: 1823.
- Fiore-Donno AM, Rixen C, Rippin M, Glaser K, Samolov E, Karsten U, Becker B, Bonkowski M. 2018. New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Molecular Ecology Resources* 18: 229–239.
- Flues S, Bass D, Bonkowski M. 2017. Grazing of leaf-associated *Cercomonads* (Protists: Rhizaria: Cercozoa) structures bacterial community composition and function. *Environmental Microbiology* 19: 3297–3309.
- Fukami T. 2015. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics* 46: 1–23.
- Ganther M, Lippold E, Bienert MD, Bouffaud M-L, Bauer M, Baumann L, Bienert GP, Vetterlein D, Heintz-Buschart A, Tarkka MT. 2022. Plant age and soil texture rather than the presence of root hairs cause differences in maize resource allocation and root gene expression in the field. *Plants* 11: 2883.
- Ganther M, Vetterlein D, Heintz-Buschart A, Tarkka MT. 2021. Transcriptome sequencing analysis of maize roots reveals the effects of substrate and root hair formation in a spatial context. *Plant and Soil* 478: 211–228.
- Gao Z, Karlsson I, Geisen S, Kowalchuk G, Jousset A. 2019. Protists: puppet masters of the rhizosphere microbiome. *Trends in Plant Science* 24: 165–176.
- Gebauer L, Bouffaud M, Ganther M, Yim B, Vetterlein D, Smalla K, Buscot F, Heintz-buschart A, Tarkka MT. 2021. Soil texture, sampling depth and root hairs shape the structure of ACC deaminase bacterial community composition in maize rhizosphere. *Frontiers in Microbiology* 12: 616828.
- Geisen S, Tveit AT, Clark IM, Richter A, Svenning MM, Bonkowski M, Ulrich T. 2015. Metatranscriptomic census of active protists in soils. *The ISME Journal* 9: 2178–2190.
- Giri B, Prasad R, Varma A. 2018. *Root biology*. Cham, Switzerland: Springer.
- Gochnauer MB, Mccully ME, Labbé H. 1989. Different populations of bacteria associated with sheathed of roots of field-grown maize. *Plant and Soil* 114: 107–120.
- Gopalaswamy G, Kannaiyan S, O'Callaghan KJ, Davey MR, Cocking EC. 2000. The xylem of rice (*Oryza sativa*) is colonized by *Azorhizobium caulinodans*. *The Royal Society* 267: 103–107.

## Chapter 2

---

- Gu Y, Wei Z, Wang X, Friman V, Huang J, Wang X, Mei X, Xu Y, Shen Q, Jousset A. 2016.** Pathogen invasion indirectly changes the composition of soil microbiome via shifts in root exudation profile. *Biology and Fertility of Soils* 52: 997–1005.
- Hacquard S, Garrido-Oter R, Gonza A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R et al. 2015.** Review microbiota and host nutrition across plant and animal kingdoms. *Cell Host & Microbe* 17: 603–616.
- Hacquard S, Spaepen S, Garrido-Oter R, Schulze-Lefert P. 2017.** Interplay between innate immunity and the plant microbiota. *Annual Review of Phytopathology* 55: 565–589.
- Hahn A, Zimmermann R, Wanke D, Harter K, Edelmann HG. 2008.** The root cap determines ethylene-dependent growth and development in maize roots. *Molecular Plant* 1: 359–367.
- Haichar Z, Santaella C, Heulin T, Achouak W. 2014.** Root exudates mediated interactions belowground. *Soil Biology and Biochemistry* 77: 69–80.
- Hartmann A, Schikora A. 2012.** Quorum sensing of bacteria and trans-kingdom interactions of *N*-acyl homoserine lactones with eukaryotes. *Journal of Chemical Ecology* 38: 704–713.
- Hassan S, Mathesius U. 2012.** The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany* 63: 3429–3444.
- Hause B, Schaarschmidt S. 2009.** The role of jasmonates in mutualistic symbioses between plants and soil-born microorganisms. *Phytochemistry* 70: 1589–1599.
- Hawes M, Allen C, Turgeon BG, Curlango-rivera G, Tran TM, Huskey DA, Xiong Z. 2016.** Root border cells and their role in plant defense. *Annual Review of Phytopathology* 54: 143–161.
- Hayward A, Fegan N, Fegan M, Stirling G. 2010.** *Stenotrophomonas* and *Lysobacter*: ubiquitous plant-associated gamma-proteobacteria of developing significance in applied microbiology. *Journal of Applied Microbiology* 108: 756–770.
- Henkes GJ, Jousset A, Bonkowski M, Thorpe MR, Scheu S, Lanoue A, Schurr U, Röse USR. 2011.** *Pseudomonas fluorescens* CHA0 maintains carbon delivery to *Fusarium graminearum*-infected roots and prevents reduction in biomass of barley shoots through systemic interactions. *Journal of Experimental Botany* 62: 4337–4344.
- Henkes GJ, Kandeler E, Marhan S, Scheu S, Bonkowski M. 2018.** Interactions of mycorrhiza and protists in the rhizosphere systemically alter microbial community composition, plant shoot-to-root ratio and within-root system nitrogen allocation. *Frontiers in Environmental Science* 6: 117.
- Henkes GJ, Thorpe MR, Minchin PEH, Schurr U, Röse USR. 2008.** Jasmonic acid treatment to part of the root system is consistent with simulated leaf herbivory, diverting recently assimilated carbon towards untreated roots within an hour. *Plant, Cell & Environment* 31: 1229–1236.
- Herdler S, Kreuzer K, Scheu S, Bonkowski M. 2008.** Interactions between arbuscular mycorrhizal fungi (*Glomus intraradices*, Glomeromycota) and amoebae (*Acanthamoeba castellanii*, Protozoa) in the rhizosphere of rice (*Oryza sativa*). *Soil Biology and Biochemistry* 40: 660–668.
- Hochholdinger F, Wen T, Zimmermann R, Chimot-marolle P, Bruce W, Lamkey KR, Wienand U, Schnable PS. 2008.** The maize (*Zea mays* L.) roothairless3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *The Plant Journal* 54: 888–898.
- Holz M, Zarebanadkouki M, Kuzyakov Y, Pausch J, Carminati A. 2018.** Root hairs increase rhizosphere extension and carbon input to soil. *Annals of Botany* 121: 61–69.
- Hu L, Robert CAM, Cadot S, Zhang X, Ye M, Li B, Manzo D, Chervet N, Steinger T, Heijden MGA et al. 2018.** Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications* 9: 2738.
- Humphris SN, Bengough AG, Griffiths BS, Kilham K, Rodger S, Stubbs V, Valentine TA, Young IM. 2005.** Root cap influences root colonisation by *Pseudomonas fluorescens* SBW25 on maize. *FEMS Microbiology Ecology* 54: 123–130.
- Iijima M, Griffiths B, Bengough AG. 2000.** Sloughing of cap cells and carbon exudation from maize seedling roots in compacted sand. *New Phytologist* 145: 477–482.
- Jacobsen AGR, Jervis G, Xu J, Topping JF, Lindsey K, Lindsey K. 2021.** Root growth responses to mechanical impedance are regulated by a network of ROS, ethylene and auxin signalling in Arabidopsis. *New Phytologist* 231: 225–242.
- Jacoby RP, Chen L, Schwier M, Koprivova A, Kopriva S. 2020.** Recent advances in the role of plant metabolites in shaping the root microbiome. *F1000Research* 9: 151.
- Jaeger CH III, Lindow SE, Miller W, Clark E. 1999.** Mapping of sugar and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan. *Applied and Environmental Microbiology* 65: 2685–2690.

## Chapter 2

---

- Jones P, Garcia BJ, Furches A, Tuskan GA, Jacobson D. 2019. Plant host-associated mechanisms for microbial selection. *Frontiers in Plant Science* 10: 862.
- Jousset A, Bonkowski M. 2010. The model predator *Acanthamoeba castellanii* induces the production of 2,4-DAPG by the biocontrol strain *Pseudomonas fluorescens* Q2-87. *Soil Biology and Biochemistry* 42: 1647–1649.
- Jousset A, Lara E, Wall LG, Valverde C. 2006. Secondary metabolites help biocontrol strain *Pseudomonas fluorescens* CHA0 to escape protozoan grazing. *Applied and Environmental Microbiology* 72: 7083–7090.
- Jousset A, Rochat L, Scheu S, Bonkowski M, Keel C. 2010. Predator-prey chemical warfare determines the expression of biocontrol genes by rhizosphere-associated *Pseudomonas fluorescens*. *Applied and Environmental Microbiology* 76: 5263–5268.
- Jousset A, Scheu S, Bonkowski M. 2008. Secondary metabolite production facilitates establishment of rhizobacteria by reducing both protozoan predation and the competitive effects of indigenous bacteria. *Functional Ecology* 22: 714–719.
- Jungk A. 2001. Root hairs and the acquisition of plant nutrients from soil. *Journal of Plant Nutrition and Soil Science* 164: 121–129.
- Karakoç C, Clark T, Chatzinotas A. 2020. Diversity and coexistence are influenced by time-dependent species interactions in a predator–prey system. *Ecology Letters* 23: 983–993.
- Kelly S, Mun T, Stougaard J, Ben C, Andersen SU, Murray JD. 2018. Distinct *Lotus japonicus* transcriptomic responses to a spectrum of bacteria ranging from symbiotic to pathogenic. *Frontiers in Plant Science* 9: 1218.
- Koller R, Scheu S, Bonkowski M, Robin C. 2013. Protozoa stimulate N uptake and growth of arbuscular mycorrhizal plants. *Soil Biology and Biochemistry* 65: 204–210.
- Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, Meir S, Rogachev I, Aharoni A. 2020. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proceedings of the National Academy of Sciences, USA* 117: 3874–3883.
- Kreuzer K, Adamczyk J, Iijima M, Wagner M, Scheu S, Bonkowski M. 2006. Grazing of a common species of soil protozoa (*Acanthamoeba castellanii*) affects rhizosphere bacterial community composition and root architecture of rice (*Oryza sativa* L.). *Soil Biology and Biochemistry* 38: 1665–1672.
- Krome K, Rosenberg K, Bonkowski M, Scheu S. 2009. Grazing of protozoa on rhizosphere bacteria alters growth and reproduction of *Arabidopsis thaliana*. *Soil Biology and Biochemistry* 41: 1866–1873.
- Krome K, Rosenberg K, Dickler C, Kreuzer K, Ludwig-Müller J, Ullrich-Eberius C, Scheu S, Bonkowski M. 2010. Soil bacteria and protozoa affect root branching via effects on the auxin and cytokinin balance in plants. *Plant and Soil* 328: 191–201.
- Krouk G, Lacombe B, Bielach A, Perrine-Walker F, Malinska K, Mounier E, Hoyerova K, Tillard P, Leon S, Ljung K et al. 2010. Nitrate-regulated auxin transport by NRT1.1 defines a mechanism for nutrient sensing in plants. *Developmental Cell* 18: 927–937.
- Kudjordjie EN, Sapkota R, Steffensen SK, Fomsgaard IS, Nicolaisen M. 2019. Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome* 7: 59.
- Kuikman PJ, Jansen AG, Van Veen JA. 1991. 15N-nitrogen mineralization from bacteria by protozoan grazing at different soil moisture regimes. *Soil Biology and Biochemistry* 23: 193–200.
- Kuikman PJ, Van Veen JA. 1989. The impact of protozoa on the availability of bacterial nitrogen to plants. *Biology and Fertility of Soils* 8: 13–18.
- Kuppardt A, Fester T, Härtig C, Chatzinotas A. 2018. Rhizosphere protists change metabolite profiles in *Zea mays*. *Frontiers in Microbiology* 9: 857.
- Kurth F, Mailänder S, Bönn M, Feldhahn L, Herrmann S, Große I, Buscot F, Schrey SD, Tarkka MT. 2014. Streptomyces-induced resistance against oak powdery mildew involves host plant responses in defense, photosynthesis, and secondary metabolism pathways. *The American Phytopathological Society* 27: 891–900.
- Lagopodi AL, Ram AFJ, Lamers GEM, Punt PJ, Hondel CAMJJVD, Lugtenberg BJJ, Bloemberg GV. 2002. Novel aspects of tomato root colonization and infection by *Fusarium oxysporum* f. sp. *radicis-lycopersici* revealed by confocal laser scanning microscopic analysis using the green fluorescent protein as a marker. *Molecular Plant–Microbe Interactions* 15: 172–179.
- Lebeis SL, Paredes SH, Lundberg DS, Glavina T, Jones CD, Tringe SG, Dangl JL. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349: 860–864.
- Leibold MA. 1996. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. *The American Naturalist* 147: 784–812.
- Leiss KA, Maltese F, Choi YH, Verpoorte R, Klinkhamer PGL. 2009. Identification of chlorogenic acid as a resistance factor for thrips in chrysanthemum. *Plant Physiology* 150: 1567–1575.

## Chapter 2

---

- Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M. 2014.** Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Molecular Ecology* 23: 3341–3355.
- Libault M, Brechenmacher L, Cheng J, Xu D, Stacey G. 2010.** Root hair systems biology. *Trends in Plant Science* 15: 641–650.
- Liu G, Ji Y, Bhuiyan NH, Pilot G, Selvaraj G, Zou J, Wei Y. 2010.** Amino acid homeostasis modulates salicylic acid-associated redox status and defense responses in Arabidopsis. *Plant Cell* 22: 3845–3863.
- Marschner P, Crowley D, Rengel Z. 2011.** Rhizosphere interactions between microorganisms and plants govern iron and phosphorus acquisition along the root axis – model and research methods. *Soil Biology and Biochemistry* 43: 883–894.
- Mierziak J, Kostyn K, Kulma A. 2014.** Flavonoids as important molecules of plant interactions with the environment. *Molecules* 19: 16240–16265.
- Mizuno D, Higuchi K, Sakamoto T, Nakanishi H, Mori S, Nishizawa NK. 2003.** Three nicotianamine synthase genes isolated from maize are differentially regulated by iron nutritional status. *Plant Physiology* 132: 1989–1997.
- Müller M, Munné-Bosch S. 2015.** Ethylene response factors: a key regulatory hub in hormone and stress signaling. *Plant Physiology* 169: 32–41.
- Nakajima T, Kurihara Y. 1994.** Evolutionary changes of ecological traits of bacterial populations through predator-mediated competition 2. Theoretical considerations. *Oikos* 71: 35.
- Nguyen C. 2003.** Rhizodeposition of organic C by plants: mechanisms and controls. *Agronomie* 23: 375–396.
- Park WJ, Hochholdinger F, Gierl A. 2004.** Release of the benzoxazinoids defense molecules during lateral- and crown root emergence in *Zea mays*. *Journal of Plant Physiology* 161: 981–985.
- Peleg-Grossman S, Golani Y, Kaye Y, Melamed-Book N, Levine A. 2009.** NPR1 protein regulates pathogenic and symbiotic interactions between rhizobium and legumes and non-legumes. *PLoS ONE* 4: e8399.
- Péret B, Larrieu A, Bennett MJ. 2009.** Lateral root emergence: a difficult birth. *Journal of Experimental Botany* 60: 3637–3643.
- Phillips DA, Ferris H, Cook DR, Strong DR. 2003.** Molecular control points in rhizosphere food webs. *Ecology* 84: 816–826.
- Phillips DA, Strong DR. 2003.** Rhizosphere control points: molecules to food webs. *Ecology* 84: 815.
- Pieterse CMJ, Ton J. 2009.** Priming of plant innate immunity by rhizobacteria and  $\beta$ -aminobutyric acid: differences and similarities in regulation. *New Phytologist* 183: 419–431.
- Pieterse MJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014.** Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology* 52: 347–375.
- Poitout A, Martinière A, Kucharczyk B, Queruel N, Silva-Adina J, Mashkooor S, Gamet L, Varoquaux F, Paris N, Sentenac H et al. 2017.** Local signalling pathways regulate the Arabidopsis root developmental response to *Mesorhizobium loti* inoculation. *Journal of Experimental Botany* 68: 1199–1211.
- Ravanbakhsh M, Sasidharan R, Voesenek LACJ, Kowalchuk GA, Jousset A. 2018.** Microbial modulation of plant ethylene signaling: ecological and evolutionary consequences. *Microbiome* 6: 52.
- Remans T, Nacry P, Pervent M, Filleur S, Diatloff E, Mounier E, Tillard P, Forde BG, Gojon A. 2006.** The Arabidopsis NRT1.1 transporter participates in the signaling pathway triggering root colonization of nitrate-rich patches. *Proceedings of the National Academy of Sciences, USA* 103: 19206–19211.
- Robertson-Albertyn S, Terrazas RA, Balbirnie K, Blank M, Janiak A, Szarejko I, Chmielewska B, Karcz J, Morris J, Hedley PE et al. 2017.** Root hair mutations displace the barley rhizosphere microbiota. *Frontiers in Plant Science* 8: 1094.
- Rosenberg K, Bertaux J, Krome K, Hartmann A, Scheu S, Bonkowski M. 2009.** Soil amoebae rapidly change bacterial community composition in the rhizosphere of *Arabidopsis thaliana*. *The ISME Journal* 3: 675–684.
- Röttgers L, Faust K. 2018.** From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiology Reviews* 42: 761–780.
- Rozmoš M, Bukovská P, Hršelová H, Kotianov M, Dudáš M, Gančarčíková K, Jansa J. 2021.** Organic nitrogen utilisation by an arbuscular mycorrhizal fungus is mediated by specific soil bacteria and a protist. *The ISME Journal* 16: 676–685.
- Rüger L, Feng K, Dumack K, Freudenthal J, Chen Y, Sun R, Wilson M, Yu P, Sun B, Deng Y et al. 2021.** Assembly patterns of the rhizosphere microbiome along the longitudinal root axis of maize (*Zea mays* L.). *Frontiers in Microbiology* 12: 614501.
- Schuhegger R, Ihring A, Gantner S, Bahnweg G, Knappe C, Vogg G, Hutzler P, Schmid M, van Breusegem F, Eberl L et al. 2006.** Induction of systemic resistance in tomato by *N*-acyl-L-homoserine lactone-producing rhizosphere bacteria. *Plant, Cell & Environment* 29: 909–918.

## Chapter 2

---

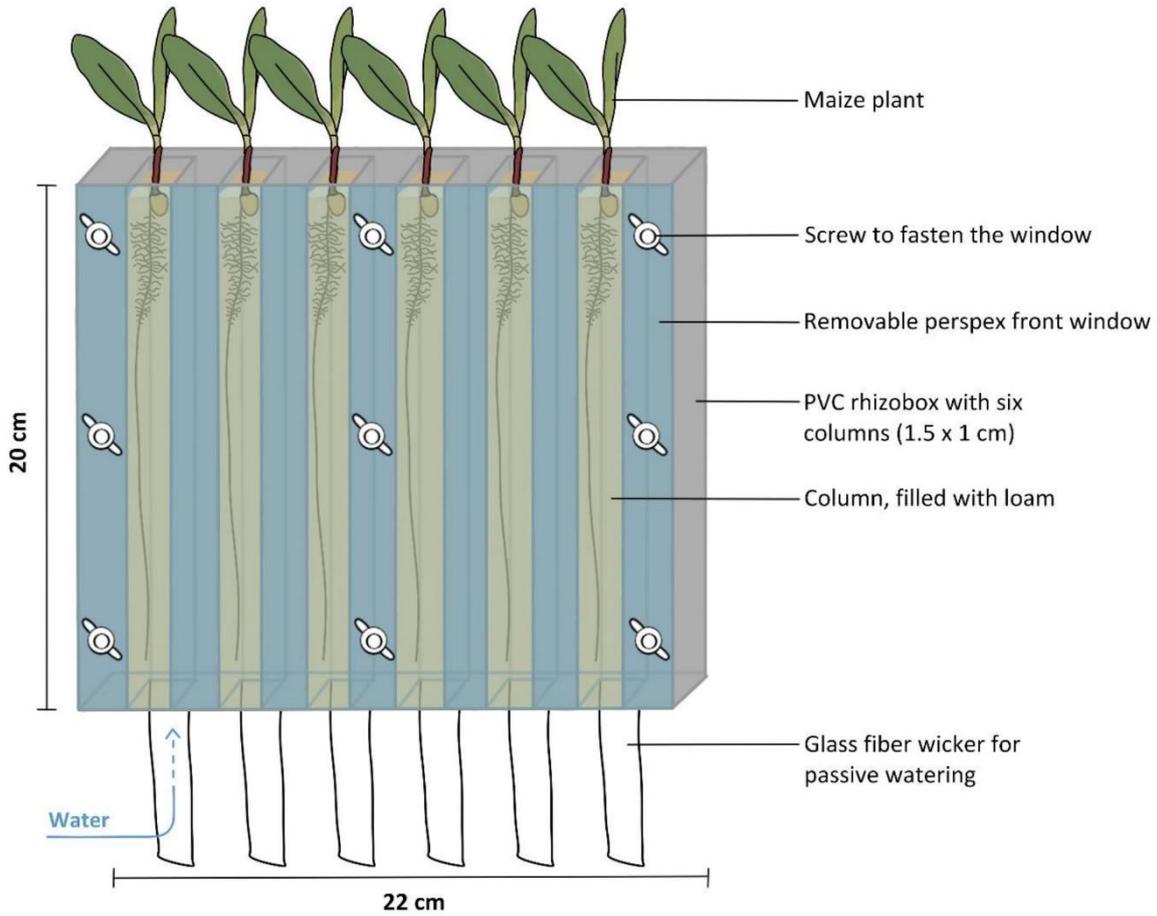
- Schultz JC, Appel HM, Ferrieri AP, Arnold TM. 2013.** Flexible resource allocation during plant defense responses. *Frontiers in Plant Science* 4: 324.
- Schütz V, Frindte K, Cui J, Zhang P, Hacquard S, Schulze-Lefert P, Knief C, Schulz M, Dörmann P. 2021.** Differential impact of plant secondary metabolites on the soil microbiota. *Frontiers in Microbiology* 12: 666010.
- Segonzac C, Zipfel C. 2011.** Activation of plant pattern-recognition receptors by bacteria. *Current Opinion in Microbiology* 14: 54–61.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003.** Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13: 2498–2504.
- Shaw LJ, Morris P, Hooker JE. 2006.** Perception and modification of plant flavonoid signals by rhizosphere microorganisms. *Environmental Microbiology* 8: 1867–1880.
- Sicker D, Freyt M, Schulz M, Gierlt A. 2000.** Role of natural benzoxazinones in the survival strategy of plants. *International Review of Cytology* 193: 319–346.
- Sprague SJ, Watt M, Kirkegaard JA, Howlett BJ. 2007.** Pathways of infection of *Brassica napus* roots by *Leptosphaeria maculans*. *New Phytologist* 176: 211–222.
- Steinkellner S, Lenzemo V, Langer I, Schweiger P, Khaosaad T, Toussaint J-P, Vierheilig H. 2007.** Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules* 12: 1290–1306.
- Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, De LN, Kaeppler SM. 2016.** An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *The Plant Genome* 9: 1–16.
- Stringlis IA, Yu K, Feussner K, De Jonge R, Van BS, Van VMC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018.** MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proceedings of the National Academy of Sciences, USA* 115: E5213–E5222.
- Suarez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonca-Previato L, James EK, Venturi V. 2012.** Common features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microbial Ecology* 63: 249–266.
- Sullivan ML, Zarnowski R. 2011.** Red clover HCT2, a hydroxycinnamoyl-coenzyme A:malate hydroxycinnamoyl transferase, plays a crucial role in biosynthesis of phelic acid and other hydroxycinnamoyl-malate esters *in vivo*. *Plant Physiology* 155: 1060–1067.
- Tackmann J, Rodrigues JFM, Mering C v. 2019.** Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Systems* 9: 286–296.
- Taerum SJ, Micciulla J, Corso G, Steven B, Gage DJ, Triplett LR. 2022.** 18S rRNA gene amplicon sequencing combined with culture-based surveys of maize rhizosphere protists reveal dominant, plant-enriched and culturable community members. *Environmental Microbiology Reports* 14: 110–118.
- Tanaka S, Han X, Kahmann R. 2015.** Microbial effectors target multiple steps in the salicylic acid production and signaling pathway. *Frontiers in Plant Science* 6: 349.
- Treutter D. 2005.** Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biology* 7: 581–591.
- Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020.** Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology* 18: 607–621.
- Van Loon LC, Rep M, Pieterse CMJ. 2006.** Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* 44: 135–162.
- Vaulot D, Geisen S, Mahe F, Bass D. 2022.** pr2-primers: an 18S rRNA primer database for protists. *Molecular Ecology Resources* 22: 168–179.
- Verbon EH, Trapet PL, Stringlis IA, Kruijs S, Bakker PAHM, Pieterse MJ. 2017.** Iron and immunity. *Annual Review of Phytopathology* 55: 355–375.
- Voges MJEEE, Bai Y, Schulze-Lefert P, Sattely ES. 2019.** Plant-derived coumarins shape the composition of an Arabidopsis synthetic root microbiome. *Proceedings of the National Academy of Sciences, USA* 116: 12558–12565.
- Wang W, Hu B, Li A, Chu C. 2019.** NRT1.1s in plants: functions beyond nitrate transport. *Journal of Experimental Botany* 71: 4373–4379.
- Weidner S, Latz E, Agaras B, Valverde C, Jousset A. 2017.** Protozoa stimulate the plant beneficial activity of rhizospheric pseudomonads. *Plant and Soil* 410: 509–515.
- Wirén NV, Marschner H, Römheld V. 1996.** Roots of iron-efficient maize also absorb phyto siderophore-chelated zinc. *Plant Physiology* 111: 1119–1125.

## Chapter 2

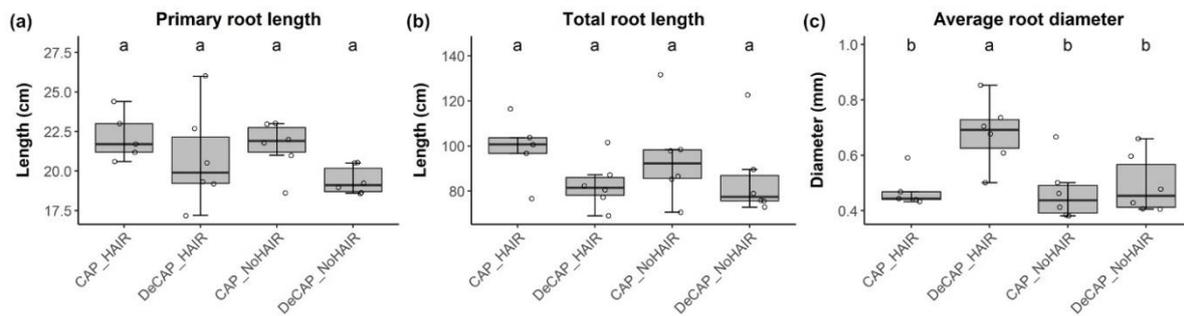
---

- Xiong W, Jousset A, Guo S, Karlsson I, Zhao Q, Wu H, Kowalchuk GA, Shen Q, Li R, Geisen S. 2018.** Soil protist communities form a dynamic hub in the soil microbiome. *The ISME Journal* 12: 634–638.
- Yamada M, Han X, Benfey PN. 2020.** RGF1 controls root meristem size through ROS signalling. *Nature* 577: 85–88.
- Yan X, Liao H, Beebe SE, Blair MW, Lynch JP. 2004.** QTL mapping of root hair and acid exudation traits and their relationship to phosphorus uptake in common bean. *Plant and Soil* 265: 17–29.
- Yu P, He X, Baer M, Beirinckx S, Tian T, Moya YAT, Zhang X, Deichmann M, Frey FP, Bresgen V et al. 2021.** Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nature Plants* 7: 481–499.
- Zelenev VV, Bruggen AHCV, Semenov AM. 2005.** Modeling wave-like dynamics of oligotrophic and copiotrophic bacteria along wheat roots in response to nutrient input from a growing root tip. *Ecological Modelling* 188: 404–417.
- Zhang H, Sun Y, Xie X, Kim M, Dowd SE, Paré PW. 2009.** A soil bacterium regulates plant acquisition of iron via deficiency-inducible mechanisms. *The Plant Journal* 58: 568–577.
- Zhang X, Kuzyakov Y, Zang H, Dippold MA, Shi L, Spielvogel S, Razavi BS. 2020.** Rhizosphere hotspots: root hairs and warming control microbial efficiency, carbon utilization and energy production. *Soil Biology and Biochemistry* 148: 107872.
- Zhou C, Guo J, Zhu L, Xiao X, Xie Y, Zhu J, Ma Z, Wang J. 2016.** *Paenibacillus polymyxa* BFKC01 enhances plant iron absorption via improved root systems and activated iron acquisition mechanisms. *Plant Physiology and Biochemistry* 105: 162–173.
- Zhou S, Zhang YK, Kremling KA, Ding Y, Bennett JS, Bae JS, Kim DK, Ackerman HH, Kolomiets MV, Schmelz EA et al. 2019.** Ethylene signaling regulates natural variation in the abundance of antifungal acetylated diferuloylsucroses and *Fusarium graminearum* resistance in maize seedling roots. *New Phytologist* 221: 2096–2111.

Supplementary material

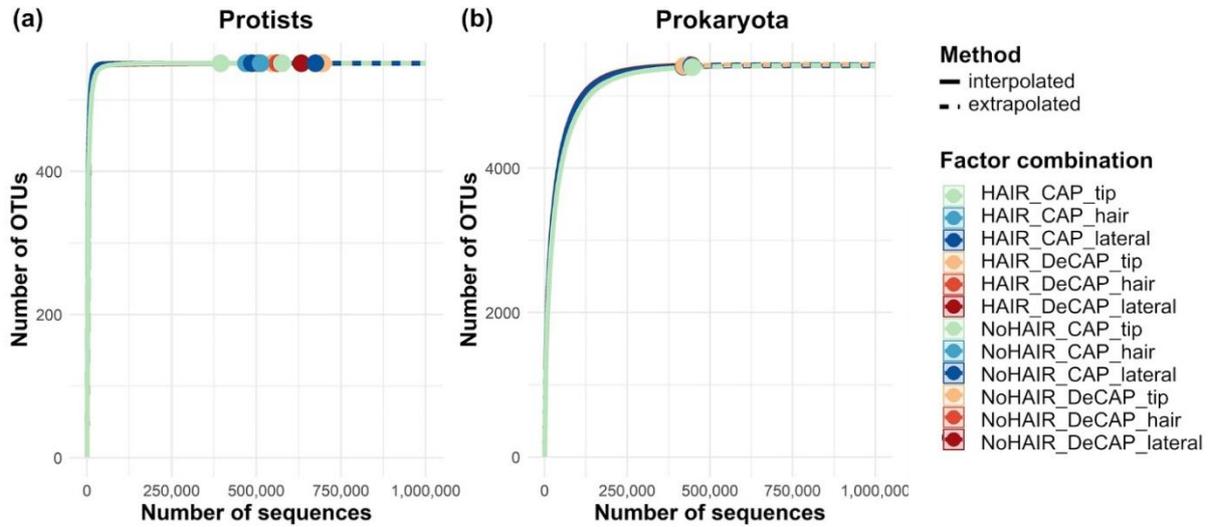


**Fig S1:** Assembled rhizobox with six columns, filled with loam and planted with individual maize plants (*Zea mays* L.). To enable root zone-specific sampling, rhizoboxes were constructed with a removable perspex front window. The bottom openings of the columns were closed, leaving only a gap for a protruding strip of a glass fiber wicker for passive watering by immersing it in a container with water.

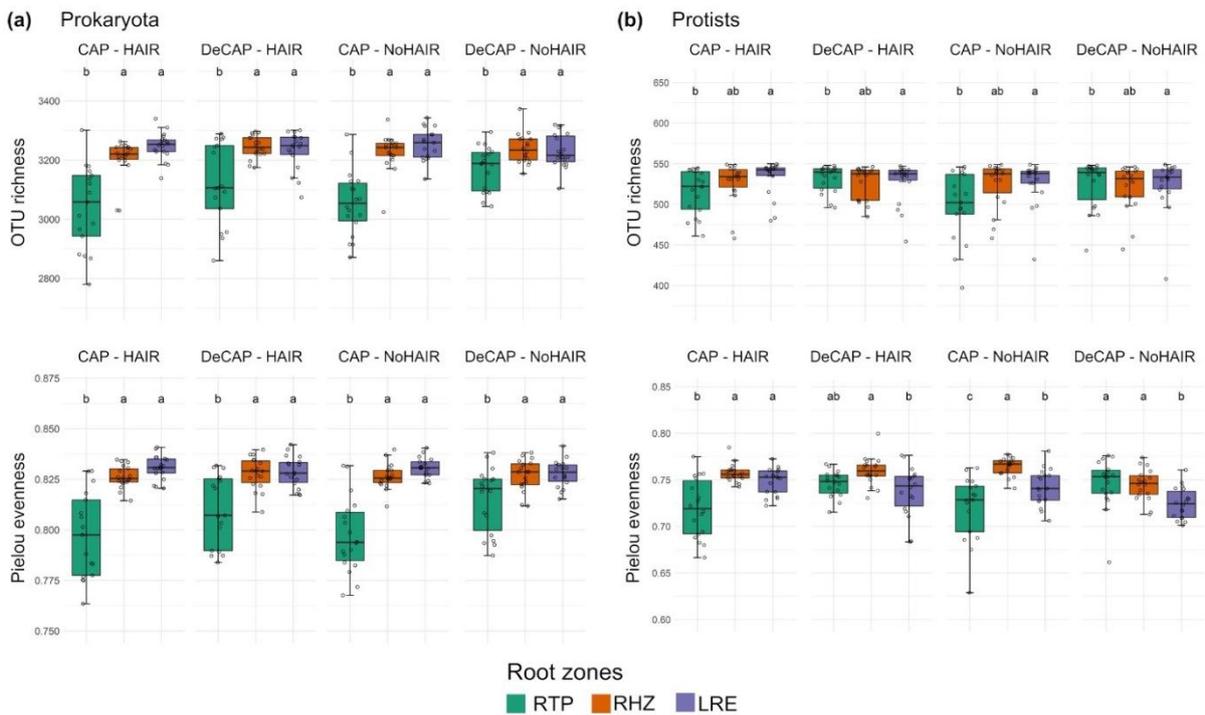


**Fig. S2:** Boxplots of (a) the total root length, (b) length of the primary root and (c) the average root diameter of plants (*Zea mays* L.) with (HAIR) and without root hairs (NoHAIR) and with (DeCAP) and without applied decapping treatment (CAP). The horizontal line within each box represents the median. Whiskers extend to the lowest and highest scores within 1.5 times the interquartile range from the box. Points represent individual datapoints. Lower case letters above boxplots indicate differences between means (Tukey's HSD), n=6, except in the CAP-HAIR treatment n=5.

## Chapter 2

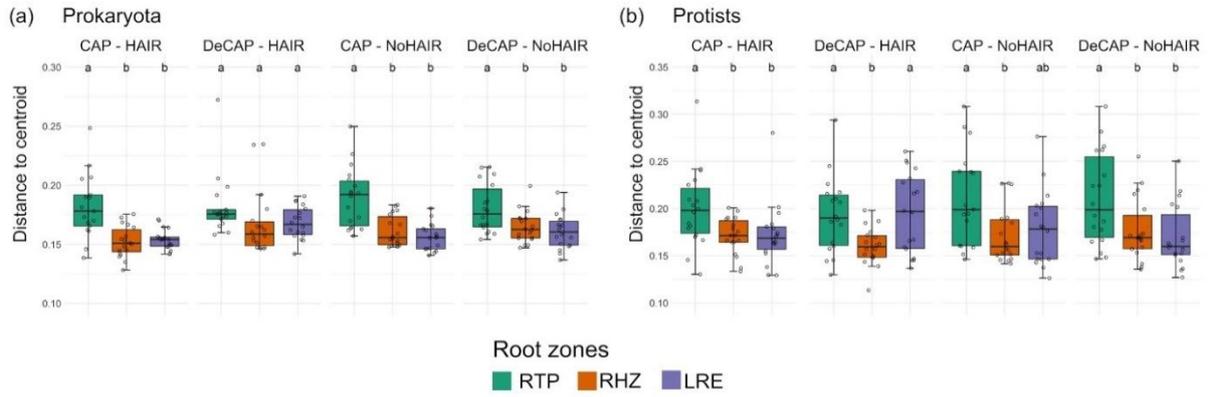


**Fig. S3:** Rarefaction curves for (a) protists (Cercozoa and Endomyxa) and (b) prokaryotes (bacteria and archaea) with interpolated number of OTUs for subsampled sequences (solid line) and extrapolated number of OTUs with increasing number of sequences (dashed line).

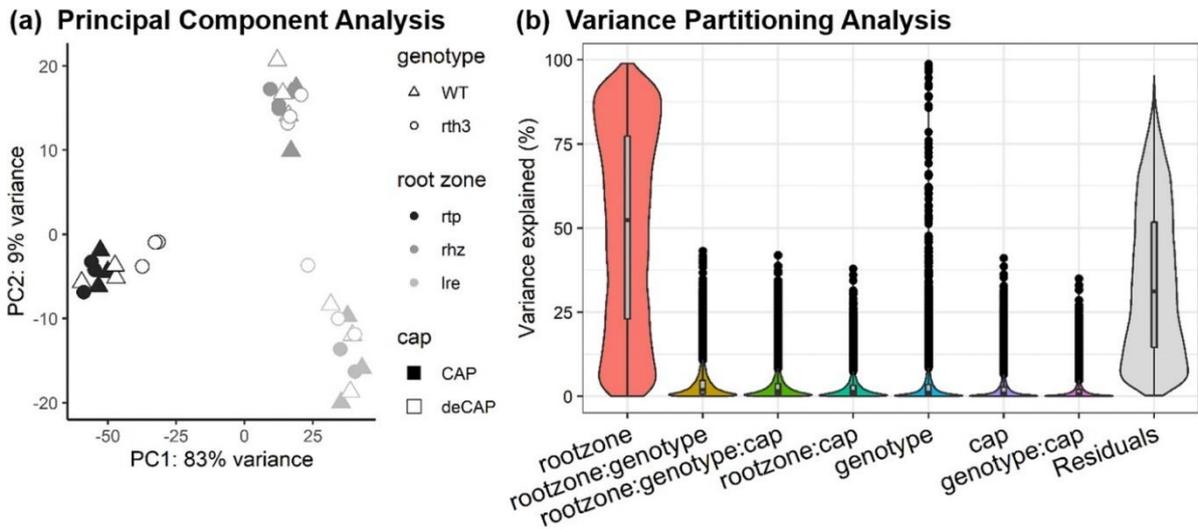


**Fig. S4:** OTU richness and Pielou evenness of (a) prokaryote (bacteria and archaea) and (b) protist (Cercozoa and Endomyxa) communities at the three root zones: root tip, root hair zone and lateral root emergence zone of *Zea mays* L., with HAIR or NoHAIR, and CAP or DeCAP treatment. The horizontal line within each box represents the median. Whiskers extend to the lowest and highest scores within 1.5 times the interquartile range from the box. Points represent individual datapoints. Letters indicate significant differences between root zones (Tukey's HSD).

## Chapter 2

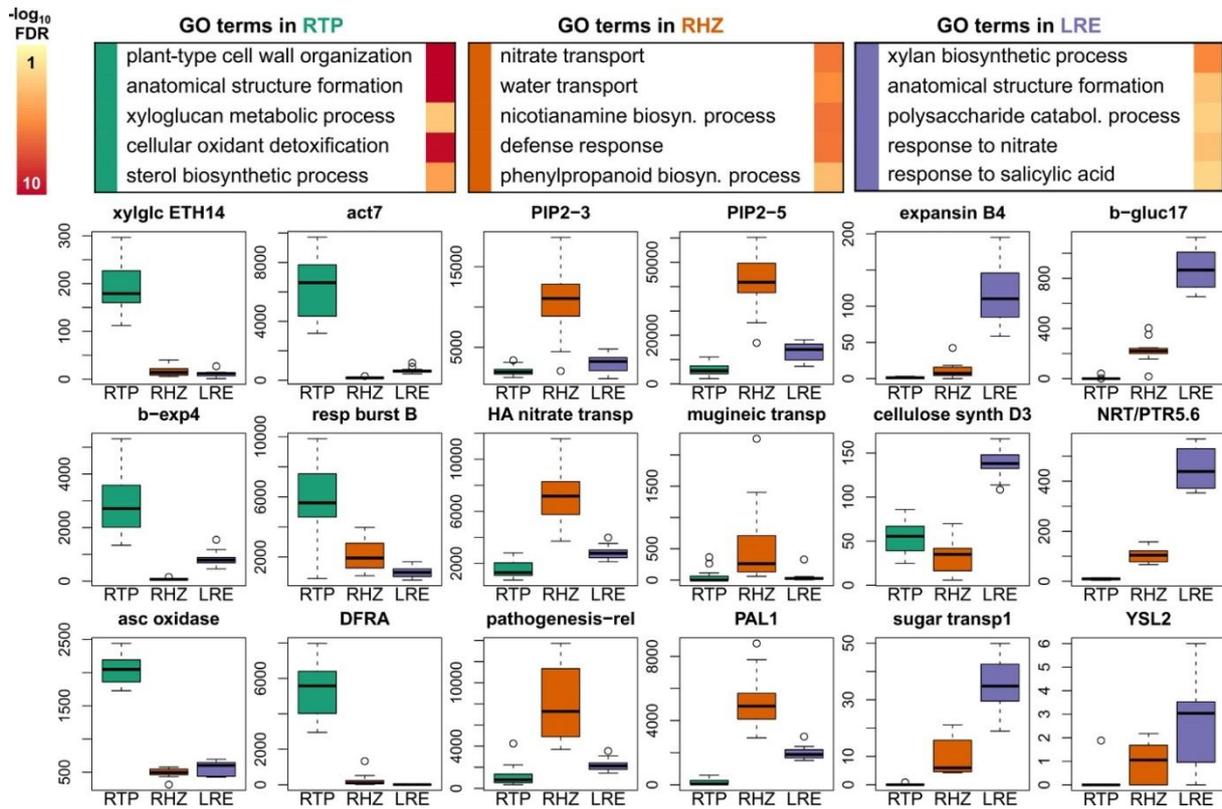


**Fig. S5:** Group dispersion of beta diversity in (a) prokaryote (bacteria and archaea) and (b) protist (Cercozoa and Endomyxa) communities at the three root zones: root tip, root hair zone and lateral root emergence zone of *Zea mays* L., with HAIR or NoHAIR, and CAP or DeCAP treatment. The horizontal line within each box represents the median. Whiskers extend to the lowest and highest scores within 1.5 times the interquartile range from the box. Points represent individual datapoints. Letters indicate significant differences between root zones (Tukey's HSD).

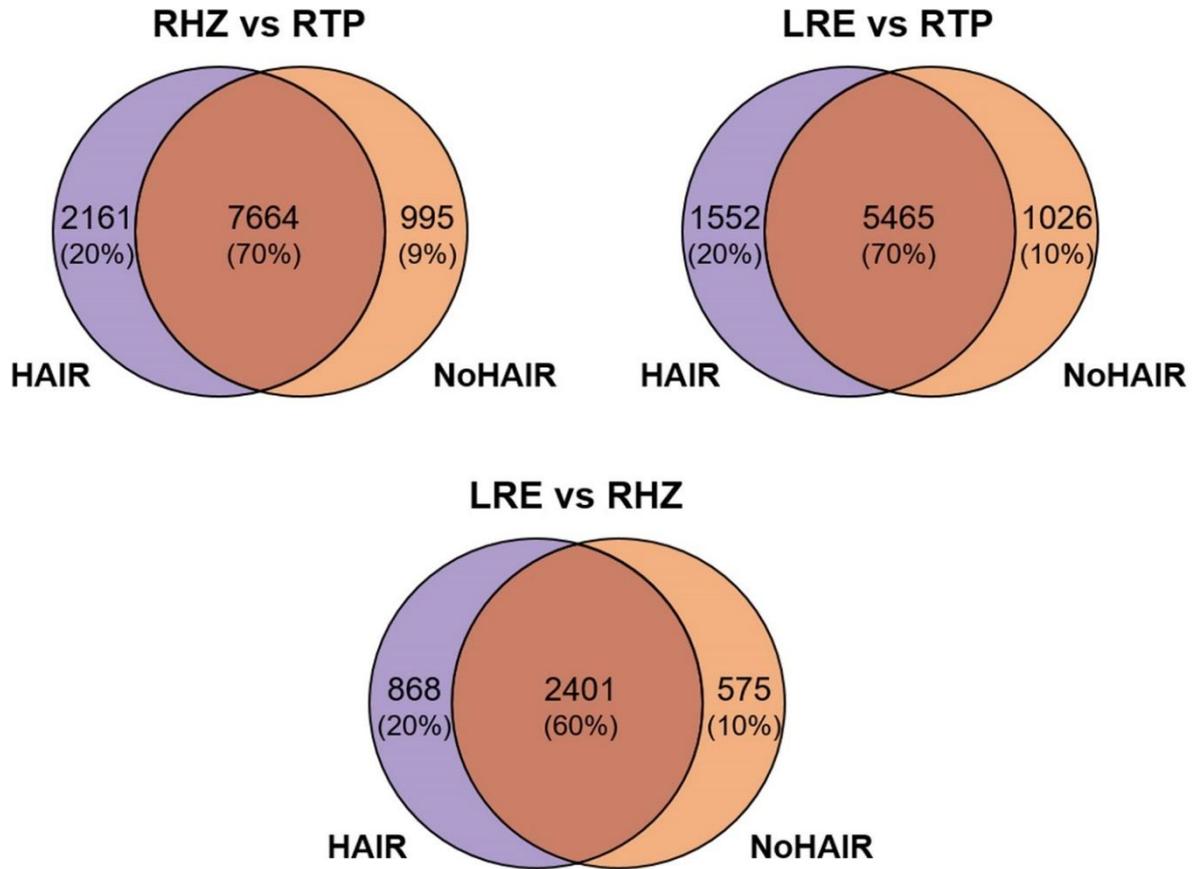


**Fig. S6:** (a) Principal Component Analysis and (b) Variance Partitioning Analysis of RNAseq root gene expression data. Explained variance is plotted for each gene along the y-axis boxplot, as well as the gene frequency (vertical histogram/violin shape). Factors are sorted by median explained variance, from highest to lowest.

## Chapter 2



**Fig. S7:** Root zone-specific genes and enriched GO terms ( $p_{adj} < 0.01$ ,  $|LFC| > 1$ ). Gradient-colored boxes to the right of each GO term indicates its significance. Below are shown expression count boxplots (normalized counts in DESeq2) for selected genes related to the enriched GO term categories. The horizontal line within each box represents the median. Whiskers extend to the lowest and highest scores within 1.5 times the interquartile range from the box. Data beyond this range are represented by individual points. Colors (green, orange, blue) in the boxplots refer to the root zone (RTP, RHZ, LRE).



**Fig. S8:** Venn diagrams of root zone differences in either HAIR or NoHAIR samples. Shown are the number and the overlap of differentially expressed genes for each pairwise root zone contrast (RHZ vs RTP, LRE vs RTP, LRE vs RHZ; Condition vs baseline-Condition) in HAIR or NoHAIR samples.

## Chapter 2

**Table S1:** Combinations of primer tag sequences used for barcoding of cercozoan and endomyxan sequences during the second PCR amplification.

A separate file has been provided for Tables S1, please refer to the original publication available at <https://doi.org/10.1111/nph.19002>.

**Table S2:** Differentially expressed genes ( $p < 0.01$ ) by root zonation, decapping, and root hair treatment. Contrasts are denoted as Condition vs Baseline-Condition. Differential expression by root hair genotype and decap treatment was also compartmentalized for each root zone (denoted as ROOTZONE\_Condition vs Baseline-Condition (e.g. RTP\_DeCAP vs CAP).

A separate file has been provided for Tables S2, please refer to the original publication available at <https://doi.org/10.1111/nph.19002>.

**Table S3:** Analysis of Variance (ANOVA) conducted on primary root length, total root length and average root diameter of *Zea mays* L. plants with HAIR, and CAP treatment. Nominator, denominator degrees of freedom (df), F-values and p-values of one-way ANOVA.

<b>ANOVA</b>									
	<b>Primary root length</b>			<b>Total root length</b>			<b>Average root diameter</b>		
	<b>df</b>	<b>F</b>	<b>P</b>	<b>df</b>	<b>F</b>	<b>P</b>	<b>df</b>	<b>F</b>	<b>P</b>
CAP	1, 19	4.38	0.0501	1, 19	3.11	0.094	1, 19	7.27	0.014
HAIR	1, 19	1.55	0.229	1, 19	0.001	0.97	1, 19	5.34	0.032
CAP : HAIR	1, 19	0.23	0.635	1, 19	0.227	0.639	1, 19	4.15	0.056

## Chapter 2

**Table S4:** Analysis of Variance (ANOVA) conducted on protist (Cercozoa and Endomyxa) and prokaryote (archaea and bacteria) OTU-richness, Shannon diversity ( $H'$ ) and Pielou evenness ( $J'$ ) in the rhizosphere of *Zea mays* L. The factors root zone, CAP, and HAIR were included in the analysis. Nominator, denominator degrees of freedom (df), F-values and p-values.

		ANOVA								
		OTU-richness			$H'$			$J'$		
		df	F	P	df	F	P	df	F	P
<b>Prokaryota</b>	root zone	2, 201	65.89	< 0.001***	2, 201	92.924	< 0.001***	2, 201	91.56	< 0.001***
	cap status	1, 201	9.82	0.002**	1, 201	8.76	0.003**	1, 201	7.82	0.006**
	hair status	1, 201	1.6	0.207	1, 201	0.32	0.571	1, 201	0.15	0.7
	root zone : cap status	2, 201	9.33	< 0.001***	2, 201	11.37	< 0.001***	2, 201	10.97	< 0.001***
	root zone : hair status	2, 201	0.67	0.513	2, 201	0.46	0.634	2, 201	0.38	0.686
	cap status : hair status	1, 201	0	0.99	1, 201	0.05	0.816	1, 201	0.08	0.783
	root zone : cap status: hair status	2, 201	0.77	0.466	2, 201	0.47	0.627	2, 201	0.38	0.685
<b>Protists</b>	root zone	2, 203	2.42	0.092	2, 203	22.71	< 0.001***	2, 203	23.85	< 0.001***
	cap status	1, 203	1.9	0.17	1, 203	0.49	0.484	1, 203	0.12	0.733
	Hair status	1, 203	2.94	0.088	1, 203	3.73	0.055	1, 203	2.37	0.125
	root zone : cap status	2, 203	4.72	0.0099**	2, 203	21.76	< 0.001***	2, 203	18.77	< 0.001***
	root zone : hair status	2, 203	0.45	0.641	2, 203	0.65	0.523	2, 203	1.03	0.359
	cap status: hair status	1, 203	0.22	0.64	1, 203	2.01	0.158	1, 203	2.68	0.103
	root zone : cap status : hair status	2, 203	0.31	0.736	2, 203	1.32	0.27	2, 203	1.13	0.324

## Chapter 2

**Table S5:** Analysis of Variance (ANOVA) conducted on prokaryote (archaea and bacteria) abundance, protist (Cercozoa and Endomyxa) and prokaryote (archaea and bacteria) OTU-richness, Shannon diversity ( $H'$ ), Pielou evenness ( $J'$ ) and beta dispersal in the rhizosphere of *Zea mays* L. The root regions root tip (RTP), root hair zone (RHZ) and lateral root emergence zone (LRE) were compared in treatments with (CAP) and without (DeCAP) root caps and root hairs (HAIR and NoHAIR) individually. The table contains F-values with nominator, denominator degrees of freedom (df) and p-values.

		ANOVA treatment combinations														
		Abundance (qPCR)			OUT-richness			$H'$			$J'$			Beta dispersal		
Treatment		df	F	P	df	F	P	df	F	P	df	F	P	df	F	P
Prokaryota	CAP HAIR	2, 50	2.03	0.143	2, 50	23.8	< 0.001 ***	2, 50	34.6	< 0.001 ***	2, 50	35.0	< 0.001 ***	2, 50	14.97	< 0.001 ***
	DeCAP HAIR	2, 51	3.18	0.050	2, 50	9.9	< 0.001 ***	2, 50	13.8	< 0.001 ***	2, 50	13.9	< 0.001 ***	2, 50	2.2	0.12
	CAP NoHAIR	2, 50	0.32	0.729	2, 50	32.5	< 0.001 ***	2, 50	45.2	< 0.001 ***	2, 50	42.6	< 0.001 ***	2, 50	17.9	< 0.001 ***
	DeCAP NoHAIR	2, 51	1.09	0.344	2, 51	6.1	0.004 **	2, 51	8.0	< 0.001 ***	2, 51	7.8	0.001 **	2, 51	7.1	0.002 **
Protists	CAP HAIR				2, 51	2.4	0.105	2, 51	20.4	< 0.001 ***	2, 51	16.1	< 0.001 ***	2, 51	4.9	0.011 *
	DeCAP HAIR				2, 51	0.3	0.72	2, 51	5.8	0.005 **	2, 51	5.8	0.005 **	2, 51	5.1	0.0095 **
	CAP NoHAIR				2, 50	4.0	0.025 *	2, 50	14.0	< 0.001 ***	2, 50	14.9	< 0.001 ***	2, 50	4.2	0.021 *
	DeCAP NoHAIR				2, 51	0.1	0.891	2, 51	5.1	0.0098 **	2, 51	6.1	0.004 **	2, 51	5.7	0.006 **

## Chapter 2

**Table S6:** Analysis of Variance (ANOVA) conducted on prokaryote (archaea and bacteria) abundance in the rhizosphere of *Zea mays* L. The factors root zone, CAP, and HAIR were included in the analysis. The table contains nominator, denominator degrees of freedom (df), F-values and p-values.

		<b>ANOVA</b>		
		<b>Abundance (qPCR)</b>		
		<i>df</i>	<i>F</i>	<i>P</i>
<b>Prokaryota</b>	root zone	2, 202	2.44	0.099
	cap status	1, 202	0.31	0.576
	hair status	1, 202	1.52	0.219
	root zone : cap status	2, 202	0.73	0.481
	root zone : hair status	2, 202	1.3	0.275
	cap status : hair status	1, 202	0.05	0.825
	root zone : cap status : hair status	2, 202	0.84	0.434

**Table S7:** Prokaryote abundances measured by qPCR, given in number of copies per gram soil dry weight.

A separate file has been provided for Tables S7, please refer to the original publication available at <https://doi.org/10.1111/nph.19002>.

## Chapter 2

**Table S8:** Permutational Multivariate Analysis of Variance (PERMANOVA) using Bray-Curtis dissimilarities, permuted 999 times, and variance partitioning to test for differences in beta diversity of prokaryota (bacteria and archaea) and protists (Cercozoa and Endomyxa) in the rhizosphere of *Zea mays* L. The factors root zone, CAP and HAIR and their interactions were included in the analysis. The table contains F-values with nominator, denominator degrees of freedom (df) and R<sup>2</sup> (PERMANOVA) or variance explained (variance partitioning) and p-values.

		PERMANOVA			
		<i>df</i>	<i>F</i>	<i>R</i> <sup>2</sup>	<i>P</i>
<b>Prokaryota</b>	root zone	2, 201	18.5	0.15	0.001 ***
	cap status	1, 201	4.0	0.02	0.001 ***
	hair status	1, 201	2.2	0.01	0.016 *
	root zone : cap status	2, 201	2.1	0.02	0.002 **
	root zone : hair status	2, 201	1.0	0.01	0.361
	cap status : hair status	1, 201	1.7	0.01	0.040 *
	root zone : cap status : hair status	2, 201	1.0	0.01	0.306
<b>Protists</b>	root zone	2, 203	17.1	0.13	0.001 ***
	cap status	1, 203	6.3	0.02	0.001 ***
	hair status	1, 203	2.2	0.01	0.007 **
	root zone : cap status	2, 203	3.8	0.03	0.001 ***
	root zone : hair status	2, 203	1.1	0.01	0.255
	cap treatment : hair status	1, 203	2.1	0.01	0.008 **
	root zone : cap status : hair status	2, 203	1.6	0.01	0.014 *
PERMANOVA treatment combinations					
		<i>df</i>	<i>F</i>	<i>R</i> <sup>2</sup>	<i>P</i>
<b>Prokaryota</b>	CAP - HAIR	2, 50	6.6	0.21	0.001 ***
	DeCAP - HAIR	2, 50	3.7	0.13	0.001 ***
	CAP - NoHAIR	2, 50	7.0	0.22	0.001 ***
	DeCAP -NoHAIR	2, 51	2.9	0.10	0.001 ***
<b>Protists</b>	CAP - HAIR	2, 51	7.7	0.23	0.001 ***
	DeCAP - HAIR	2, 51	4.6	0.15	0.001 ***
	CAP - NoHAIR	2, 50	8.1	0.25	0.001 ***
	DeCAP -NoHAIR	2, 51	3.5	0.12	0.001 ***
Variance partitioning					
		Variance expl. (%)	<i>df</i>	<i>F</i>	<i>P</i>
<b>Prokaryota</b>	root section	7.49	1, 211	18.16	<0.001
	cap treatment	0.76	1, 211	2.36	<0.001
	genotype	0.22	1, 211	1.47	0.029
	total	8.55	1, 211	7.61	<0.001
<b>Protists</b>	root section	8.36	1, 213	20.22	<0.001
	cap treatment	1.84	1, 213	4.80	<0.001
	genotype	0.39	1, 213	1.75	0.019
	total	10.47	1, 213	9.34	<0.001

## Chapter 2

---

**Table S9:** Enriched Gene Ontology terms for root zone comparisons and root zone-specific gene expression. Genes included in the analysis were differentially expressed between root zone contrasts ( $p_{adj} < 0.05$ ,  $|LFC| > 1$ ). Contrasts are denoted as Condition vs Baseline-Condition. Root zone-specific genes were considered genes that were significantly higher expressed in one root zone compared to the remaining root zones.

A separate file has been provided for Tables S9, please refer to the original publication available at <https://doi.org/10.1111/nph.19002>.

**Table S10:** Number of DEG between root zones, root hair genotypes and decap treatments. Genes included in the analysis were differentially expressed between root zone contrasts ( $p_{adj} < 0.05$ ,  $|LFC| > 1$ ). Contrasts are denoted as Condition vs Baseline-Condition.

A separate file has been provided for Tables S10, please refer to the original publication available at <https://doi.org/10.1111/nph.19002>.

## Chapter 2

---

### **Methods S1** Rhizobox construction and preparation of maize seedlings.

To enable root zone-specific sampling, rhizoboxes made of PVC were constructed with a removable perspex front window. The bottom openings of the columns were closed by tape, leaving only a gap for a protruding strip of a glass fiber wicker for passive watering by immersing it in a container with water (Supporting information Fig. 1). Ethanol-sterilized rhizobox columns were filled with sieved loam (1 mm) from a Haplic Phaeozem close to Schladebach in Saxony-Anhalt with a sand/silt/clay ratio of 33/48/19 (for soil parameters see Vetterlein *et al.*, 2021), harboring its original microbial community (the same soil was used in Ruger *et al.*, 2021). Maize seeds were surface-sterilized with 10% H<sub>2</sub>O<sub>2</sub> under vacuum for 10 min, rinsed with sterile water and subsequently placed on sterile wet filter paper in Petri dishes. Sealed Petri dishes were placed in a 30° vertical position in the dark at 20 °C for germination. Next, root caps of half of the seedlings with straight vertical primary roots were removed as in Humphris *et al.* (2005). Each cap junction was checked and only roots whose caps came off cleanly at the first attempt were used for further experimentation. Special care was taken that the root tip was not harmed by the dissection, and successful dissection of the cap was reflected by root gene expression with low numbers of DEG and the induction of only few stress response genes in the DEG palette when comparing DeCAP and CAP treatments at RTP. Roots of seedlings were then placed into holes in the center of each soil column, which were prepared using a sterile needle. To ensure root growth along the rhizobox windows, boxes were placed on racks at an angle of 30° to the vertical.

### **Methods S2** Quantification of prokaryota by qPCR.

Quantitative polymerase chain reaction (qPCR) to quantify prokaryote abundance was performed as described in Bukovská *et al.* (2021). Briefly, calibration was conducted with amplicons generated from soil DNA extracts with the same primers which were then used for the qPCR analyses (Eub338F/Eub518R), using the *TP HS DNA-free 2x Master Mix* (TopBio, Vestec, Czech Republic). Amplicons were purified from the PCR mixture by *QIAquick PCR purification kit* (Qiagen, Holden, Germany), quantified by Picogreen fluorescence, and serially diluted to reach a broad range of fragment copy numbers (from dozens to hundreds of millions per microliter). The qPCR was performed in a total volume of 20 µl containing 10 µl *Luna Universal qPCR 2x Master Mix* (New England Biolabs, M3003), 2 µl template DNA, 0.5 µl of 10 µM forward primer Eub 338 (5'-ACTCCTACGGGAGGCAGCAG-3') and 0.5 µl of 10 µM reverse primer Eub518 (5'-ATTACCGCGGCTGCTGG-3'), using a LightCycler 480 II Instrument (Roche Molecular Systems, Rotkreuz, CHE). The following cycling steps were repeated for 55 cycles, after initial denaturation at 95°C for 5 min: denaturation at 95 °C for 10 s (ramp 4.4 °C s<sup>-1</sup>), annealing at 55 °C for 20 s (ramp 2.0 °C s<sup>-1</sup>), and amplification at 72 °C for 25 s (ramp 4.4°C s<sup>-1</sup>). The qPCR results were recorded as the second derivation maximum of the amplification curves. Melting curve profiles were checked to eliminate false positives showing unusually low melting temperatures (T<sub>m</sub>) (e.g., due to non-specific primer dimer formation).

### **Methods S3** Processing of sequencing data

#### **Prokaryota**

Based on the index combinations, sequences were demultiplexed to individual samples. Thereafter, sequencing adapters were removed using cutadapt (Martin, 2011). All subsequent steps were carried out in SEED2 software (Větrovský *et al.*, 2018). First, forward and reverse sequence reads were paired when an overlap of 20 bp was detected, with a maximum allowed mismatch set at 15%. Sequences

## Chapter 2

---

with an average quality score of less than 30, a per-base quality score of less than 7, and a length of less than 200 bp were removed. Further, potentially chimeric sequences were removed, and remaining sequences clustered at 97 % similarity level (using usearch v. 8.1.1861 embedded in SEED2 software). All resulting OTUs were identified by comparing (using Blastn) with the SILVA SSU database. All sequences belonging to Eukaryota (mitochondria, chloroplasts) were then removed and remaining data resampled (rarefied) to 26,000 sequences per sample (except for three individual samples which were slightly below this threshold). The resulting sequences were clustered again at 97% similarity level, and the most abundant sequences from each OTU were used for identification via RDP classifier with the SILVA database as reference. All OTUs represented by less than 55 reads (accounting for < 0.001% of all reads) were discarded.

### **Protists**

Reads were processed using the customized MOTHUR pipeline version 1.45.3 (Schloss *et al.*, 2009). First paired-end reads were merged with a minimum overlap of 200 bp and only if there were no mismatches in primer and barcode sequences and maximum two mismatches and one ambiguity in the target sequence. All sequences which did not meet the default parameters of the quality control, or which were smaller than 300 bp were removed. Sequences were clustered into operational taxonomic units (OTUs) using VSEARCH (Rognes *et al.*, 2016) according to the abundance-based greedy algorithm (agc) with a similarity threshold of 97%. OTUs represented by less than 1000 reads were removed as likely to represent amplification or sequencing noise (Fiore-Donno *et al.*, 2018). OTUs were assigned to taxa using BLAST+ (Camacho *et al.*, 2009) with an e-value of 1-50 and the PR2 database (Guillou *et al.*, 2013), keeping only the best hit. Cercozoan and endomyxan sequences were aligned with the template provided by Fiore-Donno *et al.* (2018) allowing gaps of maximum 5 nucleotides. Chimeras were identified using UCHIME (Edgar *et al.*, 2011) and removed. Samples were re-sampled (rarefied) to 5290 sequences per sample, with no further OUT exclusion.

### **Root gene expression**

Data processing and analysis were conducted as described before (Ganther *et al.*, 2020). Briefly, adapter removal and quality trimming were performed using the default parameters of 'Trimmomatic' (Bolger *et al.*, 2014). Cleaned reads were aligned onto the maize B73\_RefGen\_v4 genome with 'HISAT2' (Kim *et al.*, 2019). Reads were assigned to genes using 'featurecounts' of the 'subread' package (Liao *et al.*, 2019). Gene descriptions from the Plants Ensembl database (plants.ensembl.org) were assigned to the gene identifiers.

### **Methods S4** Statistical analysis.

#### **Microbial abundance and diversity**

All analyses, data handling and visualization were performed in R version 4.0.3 (R Core Team, 2020) using the packages: 'dplyr' (Wickham *et al.*, 2018), 'agricolae' (de Mendiburu, 2021), 'vegan' (Oksanen *et al.*, 2017), 'car' (Fox and Weisberg, 2019), 'gridExtra' (Aguie, 2017) 'RColorBrewer' (Neuwirth, 2014) and 'ggplot2' (Wickham, 2016). Rarefaction curves were calculated with the package 'iNEXT' (Hsieh *et al.*, 2020), to confirm sufficient sequencing depth of microbial diversity. One sample each of the protist and of the prokaryote dataset were identified as outliers and removed. Calculations of alpha and beta diversity were based on relative abundances of OTUs per sample. OTU richness, Pielou evenness, Shannon diversity and prokaryote abundance were compared by analysis of variance (ANOVA) with Tukey's Honestly Significant Difference (agricolae: HSD.test). Permutational Multivariate Analysis of Variance (PERMANOVA) using Bray-Curtis dissimilarities, permuted 999 times, tested for

## Chapter 2

---

differences in beta diversity (vegan: adonis (Anderson, 2001)), visualized by Non-Metric Multidimensional Scaling (NMDS, vegan: metaMDS). In order to analyze if root cap or hair status or differences between root zones had stronger effects on microbial community composition, variance partitioning of beta diversities (vegan: varpart) was applied. Variation of microbial beta diversity on respective root zones was calculated as multivariate homogeneity of group dispersions and compared by PERMDISP2 (vegan: betadisper (Anderson *et al.*, 2006)) and Tukey's HSD. To assess how much variance in beta diversity of prokaryotes could be explained by diversity of their protistan predators, the first two axes of a principal coordinates analysis (PCoA) of cercozoan Bray Curtis dissimilarities were extracted and used as explanatory variables for Bray Curtis dissimilarities of prokaryote data in a db-RDA (vegan: capscale).

### Root gene expression

Differential gene expression was determined with the R package 'DESeq2' (Love *et al.*, 2014). After filtering genes with low-expression counts (< 10 counts in all samples combined), 29,981 genes remained for differential gene expression analysis. Genes with a corrected p-value (Benjamini-Hochberg adjustment, padj) <0.01 were considered as differentially expressed. Gene set enrichment analysis was conducted using the R package 'Goseq', utilizing the maize-GAMER gene ontology (GO) annotation by Wimalanathan *et al.*, (2018). For this, differentially expressed genes with an absolute log<sub>2</sub> fold change (LFC) > 1 were included. Variance partitioning analysis was conducted with the R package 'VariancePartitioning' (Hoffman & Schadt, 2016).

### Supplementary references

- Anderson MJ. 2001.** A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Anderson MJ, Ellingsen KE, McArdle BH. 2006.** Multivariate dispersion as a measure of beta diversity. *Ecology Letters*. doi: 10.1111/j.1461-0248.2006.00926.x
- Auguie B. 2015.** *gridExtra: Miscellaneous functions for "Grid" graphics*. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra> [accessed 1 June 2022]
- Bukovská P, Rozmoš M, Kotianová M, Gančarčíková K, Dudáš M, Hršelová H, Jansa J. 2021.** Arbuscular mycorrhiza mediates efficient recycling from soil to plants of nitrogen bound in chitin. *Frontiers in Microbiology*. doi: 10.3389/fmicb.2021.574060
- Bolger AM, Lohse M, Usadel B. 2014.** Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. doi: 10.1093/bioinformatics/btu170
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics*. doi: 10.1186/1471-2105-10-421
- de Mendiburu F. 2021.** *Agricolae: statistical procedures for agricultural research*. R package version 1.3-3. <https://CRAN.R-project.org/package=agricolae> [accessed 1 June 2022]
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011.** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. doi: 10.1093/bioinformatics/btr381
- Fiore-Donno AM, Rixen C, Rippin M, Glaser K, Samolov E, Karsten U, Becker B, Bonkowski M. 2018.** New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12729
- Fox J, Weisberg S. 2019.** *An R companion to applied regression, third edition. USA: Sage publications*. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/> [accessed 1 June 2022]
- Ganther M, Yim B, Ibrahim Z, Bienert MD, Lippold E, Maccario L, Sørensen SJ, Bienert GP, Vetterlein D, Heintzbuschart A, et al. 2020.** Compatibility of X-ray computed tomography with plant gene expression, rhizosphere bacterial communities and enzyme activities. *Journal of Experimental Botany*. doi: 10.1093/jxb/eraa262
- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boute C, Burgaud G, Vargas C de, Decelle J, et al. 2013.** The protist ribosomal reference database (PR 2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*. doi: 10.1093/nar/gks1160

## Chapter 2

---

- Hoffman GE, Schadt EE. 2016.** variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. doi: 10.1186/s12859-016-1323-z
- Hsieh TC, Ma KH, Chao A. 2020.** *iNEXT: interpolation and extrapolation for species diversity*. R package version 2.0.20. [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/) [accessed 1 June 2022]
- Humphris SN, Bengough AG, Griffiths BS, Kilham K, Rodger S, Stubbs V, Valentine TA, Young IM. 2005.** Root cap influences root colonisation by *Pseudomonas fluorescens* SBW25 on maize. *FEMS Microbiology Ecology*. doi: 10.1016/j.femsec.2005.03.005
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019.** Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*. doi: 10.1038/s41587-019-0201-4
- Liao Y, Smyth GK, Shi W. 2019.** The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*. doi: 10.1093/nar/gkz114
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. doi: 10.1186/s13059-014-0550-8
- Martin M. 2011.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. doi: 10.14806/ej.17.1.200
- Neuwirth E. 2014.** *RColorBrewer: ColorBrewer palettes*. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer> [accessed 1 June 2022]
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D. 2017.** *Vegan: community ecology package*. R package version 2.5-7. <https://CRAN.R-project.org/package=vegan> [accessed 1 June 2022]
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016.** VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. doi: 10.7717/peerj.2584
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009.** Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. doi: 10.1128/AEM.01541-09
- Rüger L, Feng K, Dumack K, Freudenthal J, Chen Y, Sun R, Wilson M, Yu P, Sun B, Deng Y, et al. 2021.** Assembly patterns of the rhizosphere microbiome along the longitudinal root axis of maize (*Zea mays* L.). *Frontiers in Microbiology*. doi: 10.3389/fmicb.2021.614501
- Vetterlein D, Lippold E, Schreiter S, Phalempin M, Fahrenkamp T, Hochholdinger F, Marcon C, Tarkka M, Oburger E, Ahmed M, et al. 2021.** Experimental platforms for the investigation of spatiotemporal patterns in the rhizosphere—Laboratory and field scale. *Journal of Plant Nutrition and Soil Science*. doi: 10.1002/jpln.202000079
- Větrovský T, Baldrian P, Morais D. 2018.** SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics*. doi: 10.1093/bioinformatics/bty071
- Wickham H, François R, Henry L, Müller K. 2018.** *dplyr: a grammar of data manipulation*. R Package Version 1.0.6. <https://CRAN.R-project.org/package=dplyr>. [accessed 1 June 2022]
- Wickham H. 2016.** *ggplot2: elegant graphic for data analysis*. New York, USA: Springer-Verlag.
- Wimalanathan K, Friedberg I, Andorf CM, Lawrence-dill CJ. 2018.** Maize GO annotation—methods, evaluation ,and review (maize-GAMER). *Plant Direct*. doi: 10.1002/pld3.52

## CHAPTER 3

### Microeukaryotic predators shape the wastewater microbiome

Nils Heck<sup>1</sup>, Jule Freudenthal<sup>1</sup>, Kenneth Dumack<sup>1</sup>

<sup>1</sup>Terrestrial Ecology, Institute of Zoology, University of Cologne, Zùlpicher Str. 47b, Köln 50674, Germany

*Keywords: protists, wastewater treatment, food web, predation, biotic interactions, metatranscriptomics*

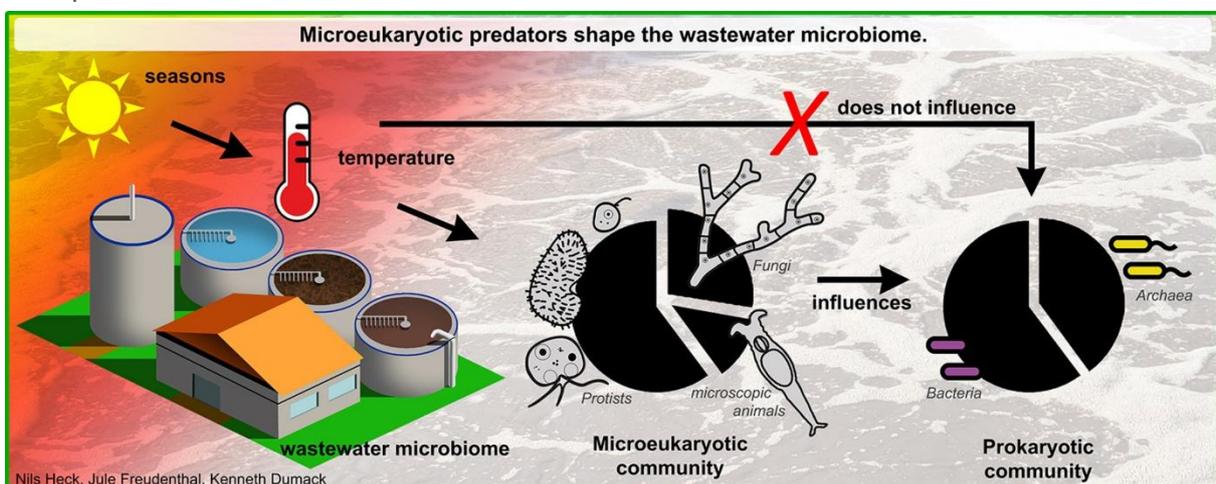
#### Highlights

- Wastewater prokaryotes undergo seasonal changes.
- Wastewater prokaryotes are unaffected by seasonal changes in water temperature.
- Wastewater microeukaryotes are affected by seasonal changes in water temperature.
- Seasonal effect shaping prokaryotes is thus indirectly exerted by predatory protists.

#### Abstract

The physicochemical parameters that shape the prokaryotic community composition in wastewater have been extensively studied. In contrast, it is poorly understood whether and how biotic interactions affect the prokaryotic community composition in wastewater. We used metatranscriptomics data from a bioreactor sampled weekly over 14 months to investigate the wastewater microbiome, including often neglected microeukaryotes. Our analysis revealed that while prokaryotes are unaffected by seasonal changes in water temperature, they are impacted by a seasonal, temperature-induced change in the microeukaryotic community. Our findings suggest that selective predation pressure exerted by microeukaryotes is a significant factor shaping the prokaryotic community in wastewater. This study underscores the importance of investigating the entire wastewater microbiome to develop a comprehensive understanding of wastewater treatment.

#### Graphical Abstract



### 1 Introduction

Prokaryotes are the most abundant microbial entities in wastewater treatment plants (WWTPs), playing a crucial role in the removal of excess nitrogen, phosphorus, and particulate matter (Arregui et al., 2010; Ferrera and Sánchez, 2016; Pan et al., 2018; Wu et al., 2019; Aragaw, 2021). Consequently, extensive research has been conducted to understand the physicochemical parameters that influence the community composition and functioning of prokaryotes in wastewater. Even before the microbial community composition of wastewater was extensively determined, it was recognized that physicochemical parameters affect functioning, and thus wastewater treatment is achieved by altering physicochemical conditions to meet specific processes. For example, denitrification is an anaerobic process, so it functions effectively under anoxic conditions (Lu et al., 2014). In contrast, nitrification is an aerobic process that requires an environment rich in oxygen (Okabe et al., 2011; Ge et al., 2015). Wastewater treatment relies on the coupling of both processes, and thus bioreactors are often periodically aerated. With the advent of molecular tools, it was discovered that physicochemical parameters also significantly influence the community composition of prokaryotes. For instance, pH is known to correlate with the relative abundance of certain prokaryotic taxa, and seasonal effects are frequently reported (Gao et al., 2016; Herold et al., 2020; LaMartina et al., 2021). Season, however, is a sum of numerous environmental changes, for instance, in the light regime and temperature. Accordingly, seasonality affects community composition and subsequently function and performance in wastewater treatment through accompanying environmental changes whose interplay is often poorly understood (Kim et al., 2013; Liu et al., 2016; González-Camejo et al., 2018; Schages et al., 2020).

Apart from prokaryotes, the microbiome in WWTPs also includes microscopic animals and microeukaryotes such as fungi and protists. Although microscopic animals, fungi, and phototrophic protists (i.e., algae) are occasionally included in surveys of wastewater, predatory protists were long marginalized (Cuellar-Bermudez et al., 2017; Freudenthal et al., 2022). This knowledge gap is mainly due to the challenges in taxonomic identification and enumeration. While primer-based metabarcoding is currently the most commonly used method to assess microbiomes, protists are genetically diverse and paraphyletic, leading to contradictory results in protist assessments (Lentendu et al., 2014; Sibbald and Archibald, 2017; Hirakata et al., 2019; Maritz et al., 2019; Burki et al., 2020). However, with shotgun meta-omics, it is now possible to assess the protistan community without any primer bias (Freudenthal et al., 2022). Furthermore, these techniques enable the assessment of the entire microbiome, from prokaryotes to microbial eukaryotes. Nevertheless, these promising methods have rarely been used to investigate microbial communities in wastewater, and if so, the data were mostly not screened for protists.

## Chapter 3

---

Despite past methodological hurdles, our knowledge of how microbial eukaryotes, particularly predators, affect wastewater treatment still relies largely on decades-old surveys or experiments (Foissner, 2016). While these findings can describe certain outcomes, they are often insufficient to link these changes to respective species in the microbial community. Predatory protists were found to be involved in floc formation, clarification, and the removal of parasites, but whether their specific predation pressure significantly shapes the prokaryotic community composition in wastewater is unknown (Lee and Welander, 1996a, 1996b; Pauli et al., 2001; Arregui et al., 2010; Freudenthal et al., 2022).

This study aims to address this knowledge gap by analyzing a publicly available metatranscriptomic dataset provided by Herold et al. (2020), who sampled the anoxic tank of a WWTP every week for about 14 months. The dataset allowed us to identify the abiotic and biotic factors that change the community composition of the entire wastewater microbiome over time. Our specific objective was to assess the wastewater microbiome in its entirety, identify which environmental factors shape its composition, and particularly investigate whether biotic interactions shape the community.

## 2 Material and methods

### *2.1 Data access, quality filtering, and assembly*

We utilized a publicly available dataset provided by Herold et al. (2020) for this study. Briefly, the samples were collected on a weekly basis over a period of 14 months from the anoxic (denitrification) zone of the aeration tank of a municipal biological wastewater treatment plant located in Schifflange, Luxembourg (49°30'48.29"N; 6°1'4.53"E; Herold et al., 2020). The treatment plant in Schifflange, after primary clarification, uses an activated sludge process. This takes place in two aeration basins, which have an outer aerobic nitrification zone and an inner anaerobic denitrification zone each, with a total volume of 17,400 m<sup>3</sup> (SIVEC, 2023). Samples were taken from floating sludge islets in the denitrification zone. After the samples had been collected, RNA was extracted as part of a sequential co-isolation procedure. For RNA extraction and cDNA library preparation, please refer to Herold et al. (2020) and Roume et al. (2013). The code for the presented analyses including set parameters is available on GitHub under the following link: <https://github.com/N-Heck-1/Microeukaryotic-predators-shape-the-wastewater-microbiome>. The quality of the raw data received was assessed using fastqc v. 0.11.9 (Andrews, 2010). We used TrimGalore v. 0.6.7 (Krueger et al., 2021) to detect and remove adapters and perform quality trimming, which removed bases with a quality score of <30 and the last ten bases of the 3' end. Mothur v. 1.45.3 (Schloss et al., 2009) was used to assemble the paired-end reads into contigs. The contigs were screened for a

## Chapter 3

---

minimum length of 90 bp and a maximum of 2 bp in ambiguities and mismatches after determining appropriate screening parameters.

### *2.2 Taxonomic assignment*

After quality filtering, we screened the sequences using the BLASTN algorithm (Camacho et al., 2009) against the respective databases. To identify eukaryotic taxa, we used the PR<sup>2</sup> v. 4.11.1 database (Guillou et al., 2013), and for prokaryotic taxa, we used the SILVA 138 SSU Ref Nr. 99 database (Pruesse et al., 2007). We kept only the best hit for each result, and filtered the results using an e-value threshold of  $1e^{-35}$  for PR2 and  $1e^{-30}$  for SILVA results, a bit-score threshold of 140, and an identity threshold of >92% across the data set. As our investigation focused on the microbial community, we removed sequences derived from chloroplasts, macroscopic Metazoa, and Embryophyta. Taxon counts were binned at the genus level to form the OTUs for all further analyses. We performed data filtering and analysis in R v. 4.0.5 (R Core Team, 2021), using the packages lattice v. 0.20–41 (Deepayan, 2008), tidyr v. 1.1.3 (Wickham and Girlich, 2022), dplyr v. 1.0.7 (Wickham et al., 2022), ape v. 5.6–2 (Paradis, 2006), and vegan v. 2.5–7 (Dixon, 2003; Oksanen et al., 2020). Additionally, we used ggplot2 v. 3.3.4 for visualization (Wickham, 2016).

### *2.3 Removal of outliers*

To ensure the uniformity of sequencing results, we initially screened the data and removed two samples with much lower sequencing depth and which were sampled much earlier than the others (i.e., 2010–10–04, 2011–01–25). We also removed singleton to quintuplet OTUs before proceeding to further analysis. To identify potential outliers among samples, we utilized multivariate dispersion based on normalized and Bray-Curtis transformed data (Bray and Curtis, 1957) and generated an NMDS plot using the metaMDS function from the vegan package. Based on the plot, we excluded three additional samples (i.e., 2011–04–05, 2011–03–29, 2011–03–21; see Supplementary Figure 1). We then produced rarefaction curves to assess sequencing depth across the remaining samples and excluded those with a depth below  $1e^6$  ribosomal reads (i.e., 2011–11–23, 2011–11–16, 2011–11–29; see Supplementary Figure 2). To facilitate comparability between samples, we rarefied the data to a depth of 1,066,681 ribosomal reads per sample, resulting in the removal of 43 rare OTUs out of a total of 4386.

### *2.4 Statistical analyses*

To visualize temporal changes in community composition, stacked bar plots were generated. An NMDS plot was also computed to visualize the impact of environmental factors on the prokaryotic and

## Chapter 3

---

eukaryotic communities. The environmental variables considered were season, water temperature, conductivity, and oxygen saturation, which were measured as described in Herold et al. (2020). The potential effect of these parameters on community composition was explored using the `envfit` and `orditorp` functions from the `vegan` package. To investigate the influence of biotic factors on prokaryotic and eukaryotic community structure, principal coordinate analysis (PCoA) was employed to compute a measure of community structure using the `pcoa` function from the `ape` package (Gower, 1966). The PCoA revealed that the first two axes explained 67.7% and 12.9% of the variation in the prokaryotic community, respectively, and 41.5% and 17.8% of the variation in the microbial eukaryotes. The coordinates of the axes with the highest explanatory power were used for further analyses. To investigate the impact of environmental variables and biotic interactions on the eukaryotic and prokaryotic communities, PERMANOVA was performed using the `adonis` function from the `vegan` package (Anderson, 2001). To investigate the factors and interactions that contribute significantly to the variation in prokaryotic and eukaryotic communities, we employed variance partitioning analysis. To address the issue of numerous zero values in the abundance data, a Hellinger transformation was applied prior to analysis using the `varpart` function from the `vegan` package (Legendre, 2008). To identify the variables that significantly explained the variation in community composition, we conducted forward selection using the `ordistep` function from the `vegan` package.

### *2.5 Network analyses*

To visualize and locate biotic interactions, networks were calculated. Sample heterogeneity caused by rare taxa was reduced to lessen their influence on network precision through co-absence (Röttjers and Faust, 2018). A prevalence filter was applied, excluding all OTUs present in less than 50% of the samples, resulting in 1559 OTUs being included in the network analyses. FlashWeave v. 0.19.0 (Tackmann et al., 2019) was chosen to investigate the multitude of complex associations within the WWTP microbiome. As a cross-sectional tool for network inference, it is well-suited for data with varying time intervals (Gerber, 2014). FlashWeave provides good scalability for large data sets, performs well on data with many heterogeneous samples, and enables the inclusion of environmental data (Matchado et al., 2021). The tool was run in sensitive mode with default settings in the Julia v. 1.7.3 environment (Bezanson et al., 2017). The resulting network was simplified for visualization in Cytoscape v. 3.9.1 (Shannon et al., 2003), by aggregating calculated nodes at the order level. This step included the influences of unique OTUs as edges in the network. If edges with conflicting signs were introduced by this step, both were kept, reflecting evenly contrasting interactions (Röttjers and Faust, 2018).

## Chapter 3

---

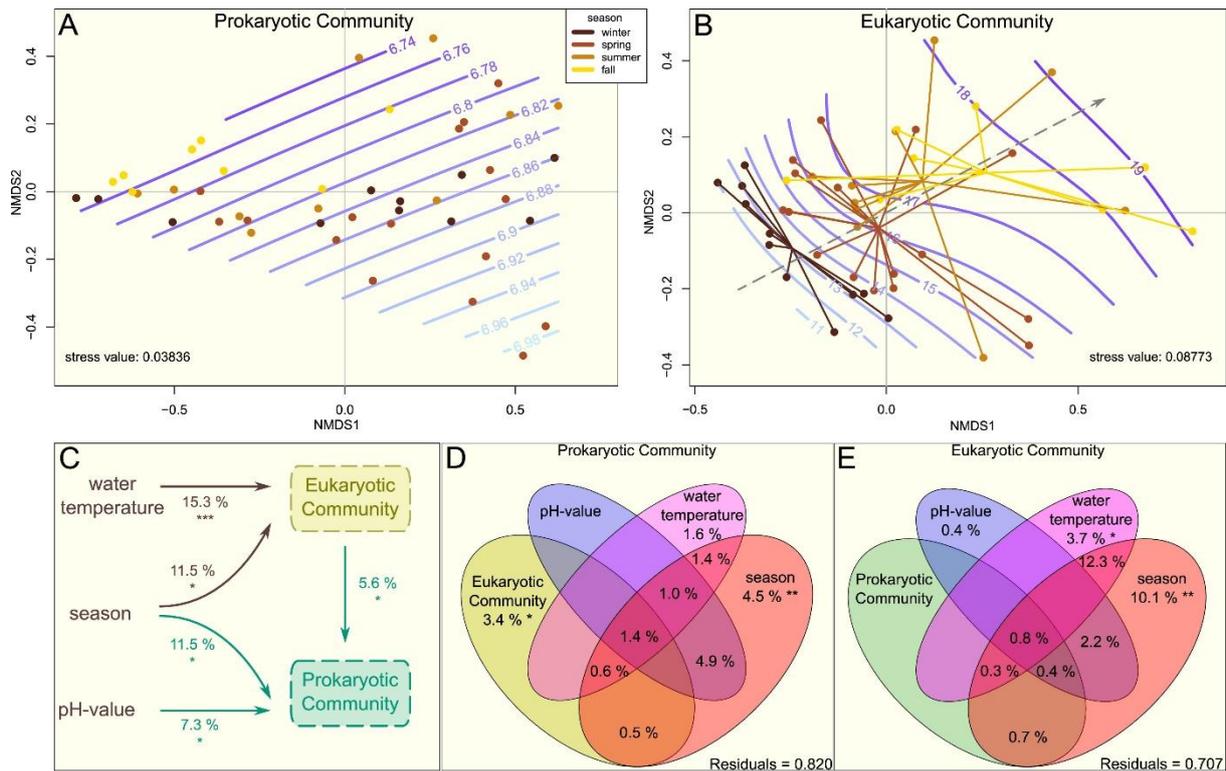
On the OTU level, 2673 edges were found between 1530 nodes and aggregated to 1767 edges between 327 nodes at the order level. Associations within functional groups (here defined as Metazoa, fungi, bacteria, Archaea, and protists) accounted for 883 edges. Among these, 610 edges represented associations among protists, 228 edges among bacteria, and 51 among fungi. Node size was chosen to denote the total number of ribosomal reads across all samples for that order, and node color indicated the number of OTUs that were aggregated. Edge thickness was related to the number of individual edges between the two orders, while edge color denoted the sign. To further ease the visualization of the network, only edges between functional groups and their respective nodes were shown, and nodes were grouped by higher taxonomic levels.

### 3 Results

The microbial community in the wastewater treatment plant underwent seasonal changes, with the eukaryotic community experiencing more pronounced changes compared to the prokaryotic community (Fig. 1). Prokaryotic seasonal clusters were not clearly separated (Fig. 1A), while a separation of gradually changing winter and summer clusters was observed for eukaryotes (Fig. 1B). Seasonal changes shaping the microeukaryotic community can be traced back to a response of the microeukaryotes to changes in season-dependent water temperature (Fig. 1C). This observation was supported by variance partitioning, which revealed that both season and temperature accounted for significant proportions of the variation in the microeukaryotic data, with a substantial overlap of 12.3% (Fig. 1E). In contrast, we did not observe a clear seasonal response of the prokaryotic community to changes in water temperature. To determine which seasonal changes, if not water temperature, influenced the changes in the prokaryotic community, it was tested whether the seasonal variation of the prokaryotes could be explained by the changes found in the eukaryotic community composition. The analysis revealed that the variation in the prokaryotic community composition (Fig. 1C) was indeed explained by the variation in the eukaryotic community composition (PCoA1). Supporting these findings, approximately 14% of the variation in the prokaryotic dataset was attributed to seasonal changes, while a smaller but still significant proportion (about 6%) was explained by the eukaryotic community composition, as determined by variance partitioning (Fig. 1D). In addition to the described seasonal and biotic effects, the pH value was found to explain variation in prokaryotes, too (Fig. 1A and C). The other tested chemical parameters showed no effects.

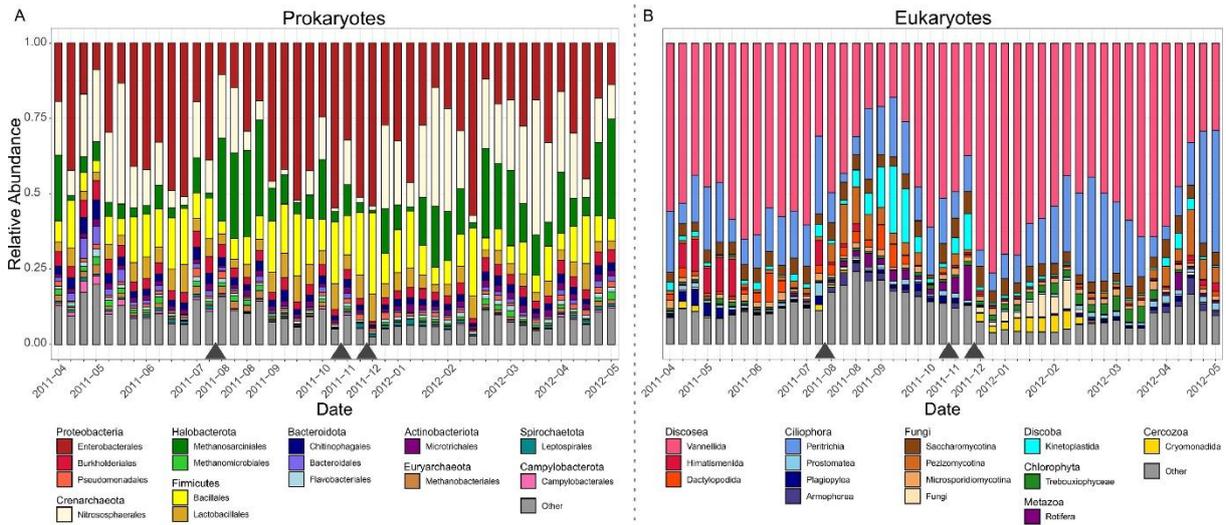
Since we found a shaping influence of the eukaryotic community composition on the prokaryotic community composition, we investigated which pro- and eukaryotic taxa changed over time. We found a total of 4343 operational taxonomic units (OTUs) composed of 47,995,763 SSU rRNA sequences in the bioreactor. In general, the prokaryotic community was dominated by the orders Enterobacterales

## Chapter 3



**Fig. 1:** Influences on the prokaryotic and eukaryotic community composition in wastewater. **(A)** The prokaryotic wastewater community is depicted, with the strongest abiotic shaping force, pH, highlighted – seasonal effects are not clearly visible. In contrast, the eukaryotic community composition is strongly shaped by season and related water temperature **(B)**. The color of each point indicates the season **(A + B)**. Smooth surfaces (blue lines) represent the influence of the abiotic parameter with the highest explanatory power on the respective community compositions, i.e., pH for prokaryotes and water temperature for eukaryotes, respectively. The central arrow in B indicates the orientation of seasonal cluster centromeres along the temperature gradient. **(C)** Predictors for prokaryotic and eukaryotic community composition are shown. Arrows indicate significant influences on the microeukaryotic and prokaryotic communities, respectively. The structure (PCoA1) of the eukaryotic community was used as a predictor for prokaryotic data, and the structure (PCoA1) of the prokaryotic community was used as a predictor for eukaryotic data. The  $R^2$ -values indicate the percentage of the respective factor explained. Asterisks indicate the significance codes: \* for  $0.05 \geq p > 0.01$ , \*\* for  $0.01 \geq p > 0.001$ , and \*\*\* for  $p \leq 0.001$ . Only significant factors are shown. **(D + E)** Venn diagrams show the partitioned variance explaining the composition of the prokaryotic community **(D)** and the eukaryotic community **(E)**. Significant effects according to forward selection are marked by stars:  $0.05 \geq * > 0.01 \geq **$ . Note that here too a significant effect of the eukaryotic community composition on the prokaryotic community composition was found **(D)** and that a large percentage of the seasonal effect explaining the variance of the eukaryotic community composition overlaps with a temperature effect **(E)**. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

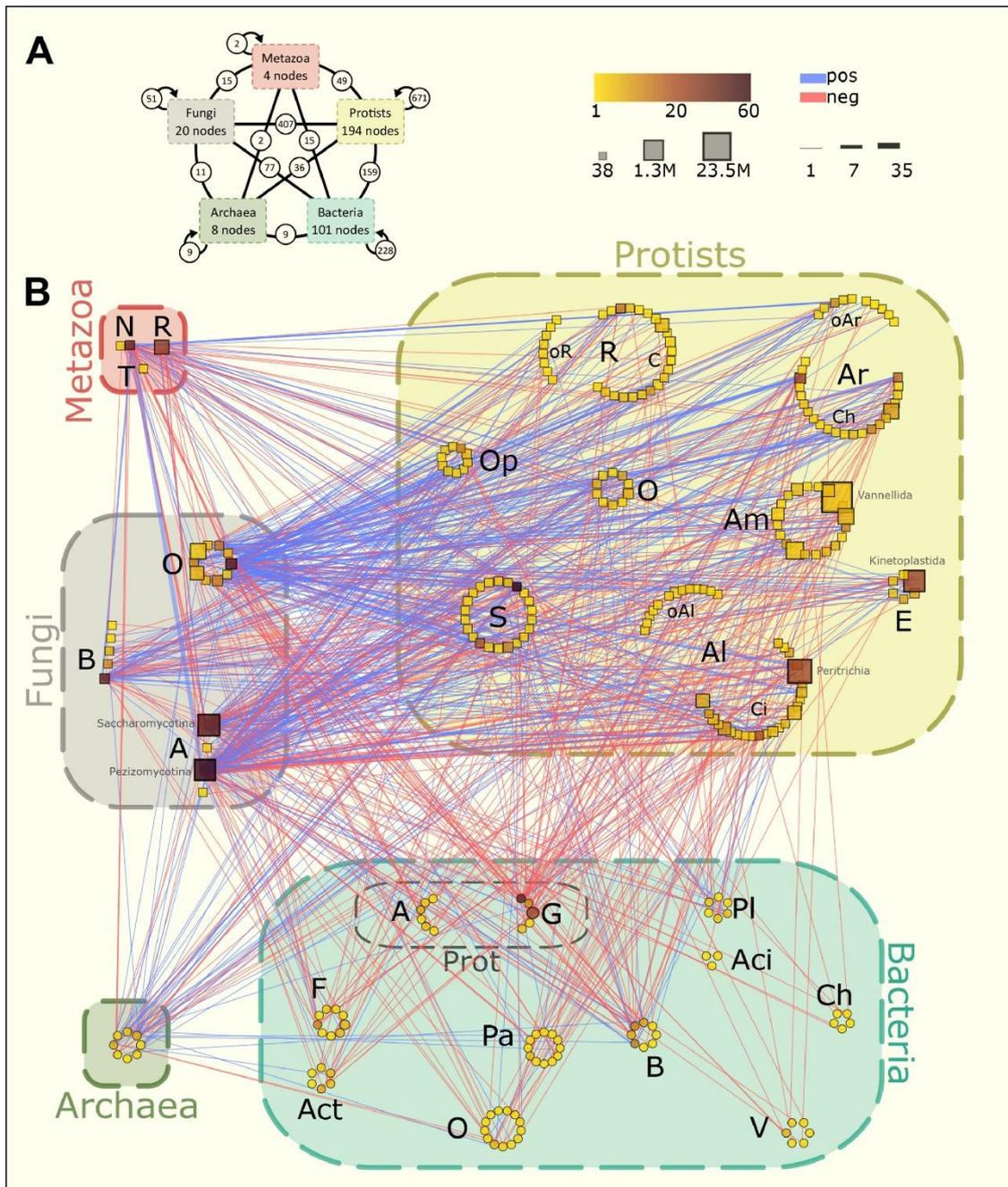
(Gammaproteobacteria), followed by Nitrososphaerales (Crenarchaeota), Methanosarcinales (Halobacterota), and Bacillales and Lactobacillales (Firmicutes; Fig. 2A). The eukaryotic community was dominated by Vannellida (Amoebozoa), making up almost half of all eukaryotic rRNA reads, followed by the second most abundant order Peritrichia (Ciliophora). Among the other highly abundant groups were Prostomatea (Ciliophora), Kinetoplastida (Discoba), Cryomonadida (Cercozoa), Trebouxiophyceae (Chlorophyta, green algae), and the fungal taxa Saccharomycotina and Pezizomycotina (Fig. 2B).



**Fig. 2:** Microbial community composition in the anoxic compartment of a wastewater treatment plant over 13 months. Stacked bar plots display the relative abundance of the 15 most abundant prokaryotic (A) and eukaryotic (B) orders. Each bar represents the community in one sample, and gray triangles on the x-axis indicate gaps in the dataset due to the removal of low-quality samples.

Many of the dominant eukaryotic orders displayed a higher relative abundance in samples taken at lower water temperatures in winter and early spring, such as Vannelliida (Amoebozoa), Cryomonadida (Cercozoa), and the green algae Trebouxiophyceae (Chlorophyta; Supplementary Fig. 3A–C). In contrast, Himatistenida and Dactylopodida (both Amoebozoa; Supplementary Figure 3D), as well as Pezizomycotina (fungi) showed an increase in their relative abundance with higher water temperature, mostly from April to September (Fig. 2B). Kinetoplastida (Discoba) and Rotifera also showed higher abundances from late summer to early winter, while both were hardly detectable from December to April.

To investigate which biotic interactions in the microbial community may have led to the observed changes, we employed network analysis. The most pronounced result was the numerous negative associations between protists and bacteria (~80.1%), potentially indicating predator-prey relationships. A total of 785 edges indicated associations between functional groups, with 407 (~51.8%) between fungi and protists, 159 (~20.3%) between bacteria and protists, and 77 (~9.8%) between bacteria and fungi (Fig. 3). Most associations between protists and fungi were positive (~66.9%).



**Fig. 3:** Co-occurrence network of microbial orders in the anoxic compartment of a biological wastewater treatment plant. Nodes were aggregated at order level. **(A)** shows a summary of the network presented in **(B)**, highlighting the number of nodes in each functional group and the number of edges within and between each group. In **(B)**, the node color indicates the number of genera included and node size the total number of reads over all samples, as shown in the legend. Blue edges show positive association, red edges show negative association, and the thickness of the edges shows the number of genera involved. Only edges between the functional groups are displayed in the network itself **(B)**, and only nodes with at least one edge are displayed. Abbreviations for Metazoa: N=Nematoda, R=Rotifera, T=Tardigrada. Abbreviations for Fungi: A=Ascomycota, B=Basidiomycota, O=Other. Abbreviations for Protists: Al=Alveolata, Am=Amoebozoa, Ar=Archaeplastida, C=Cercozoa, Ci=Ciliophora, Ch=Chlorophyta, E=Excavata, O=Other, oAl=other Alveolates, oAr=other Archaeplastida, Op=protistan Opisthokonta, oR=other Rhizaria, R=Rhizaria, S=Stramenopiles. Abbreviations for Bacteria: A=Alphaproteobacteria, Aci=Acidobacteriota, Act=Actinobacteriota, B=Bacteroidota, Ch=Chloroflexi, F=Firmicutes, G=Gammaproteobacteria, O=Other, Pa=Patescibacteria, PI=Planctomycetota, Prot=Proteobacteria, V=Verrucomicrobiota. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 4 Discussion

The performance of a wastewater treatment plant is highly dependent on the composition and activity of its microbial community (Arregui et al., 2010; Liu et al., 2016; Aghalari et al., 2020). Therefore, it is crucial to have a detailed understanding of the factors influencing the microbial community to ensure consistent and efficient WWTP performance. This study reveals that the seasonally changing microeukaryotic community composition is a significant factor shaping the prokaryotic community.

Researchers have been investigating the environmental parameters that shape the prokaryotic community composition in WWTPs for decades (Kim et al., 2013; Liu et al., 2016; Herold et al., 2020). Although the potential importance of eukaryotes for wastewater treatment has been described, it has not received much attention due to a lack of suitable methodology (Pauli et al., 2001; Arregui et al., 2010; Foissner, 2016). Accordingly, a previous study reported no clear seasonality for protists (Utz and Bohrer-Morel, 2008), while another study found that seasonally changing water temperature influenced protistan growth (Hirakata et al., 2019). In this study, we found clear seasonal variations in both prokaryotes and eukaryotes. Although the temperature in the anoxic tank fluctuated seasonally with the ambient temperature, the seasonal effect on the prokaryotic community composition could not be explained by a direct effect of water temperature. Instead, we conclude that the eukaryotic community composition was impacted by seasonally changing water temperatures, which, in turn, shaped the composition of the prokaryotic community. This finding underscores the importance of focusing on the whole microbial community and its interactions to facilitate our understanding of wastewater treatment processes.

For the first time in a WWTP, we were able to disentangle a shaping force exerted by the microeukaryotic community on the prokaryotic community composition. Our use of network analysis revealed numerous negative associations between eukaryotic and prokaryotic groups, further visualizing the shaping impact that had been found statistically. Such negative edges may have diverse ecological explanations, including competition and predation (Faust and Raes, 2012). We argue that this shaping force is largely caused by protistan predation. All of the most numerous protists in this study are known bacterivores (Böhme et al., 2009; Vaerewijck et al., 2011; Samba-Louaka et al., 2019), including sessile peritrich ciliates like *Vorticella*, which are considered important for treatment efficiency, effluent clarity, and nitrification, as well as Kinetoplastida and Cryomonadida, both of which are known to feed on wastewater bacteria (Pauli et al., 2001; Öztoprak et al., 2020; Arregui et al., 2012; Foissner, 2016; Liu et al., 2016; Pohl et al., 2021; Freudenthal et al., 2022). The most abundant protistan taxon in this study, bacterivorous Vannelida, is often absent from wastewater surveys that rely on primer-based methods, but studies based on culturing or metatranscriptomic methods have repeatedly reported their presence in WWTP bioreactors (Ramirez et al., 2015; Freudenthal et al.,

## Chapter 3

---

2022). We thus consider this to be a primer bias against the amoebozoan Vannellida (Urich et al., 2008; Geisen et al., 2015; Fiore-Donno et al., 2016).

It is still uncertain to what degree changes in prokaryotic community composition caused by protistan predation affect their functioning. It is known that there is a certain redundancy in prokaryotic functioning in wastewater, where multiple different prokaryotes are involved in the same nitrogen removal processes (Pan et al., 2018). Given that protists feed selectively, it is reasonable to assume that not all of these taxa are preyed upon at the same rate. Such redundancy suggests that changes in the prokaryotic community due to protistan predation may only have a limited effect on wastewater treatment functions, but this requires further investigation (Ju et al., 2014). Our analysis revealed a high abundance of Nitrosphaerales, which are ammonia-oxidizing Archaea involved in wastewater nitrification with increased abundance in hypoxic conditions (Park et al., 2006; Limpiyakorn et al., 2013; Rodríguez et al., 2015; Ferrera and Sánchez, 2016). However, it is not well understood whether, and at what rate, protists prey on Archaea.

In addition to predation, protists also compete with prokaryotes. For example, phototrophic protists (i.e., algae) are known to compete with prokaryotes for nutrients such as nitrogen and phosphorus (Cuellar-Bermudez et al., 2017; González-Camejo et al., 2018). Nevertheless, we argue that competition for nutrients is likely less significant than predation throughout the year. Phototrophic protists were generally less abundant than predatory ones, but they peaked in their abundance during winter. Therefore, eukaryote-prokaryote competition may be more pronounced in winter.

## 5 Conclusion

The results presented in this study, along with several other recent studies, emphasize the importance of giving more attention to microeukaryotes, particularly protists, in the investigation and improvement of wastewater treatment (Foissner, 2016; Assress et al., 2019; Freudenthal et al., 2022). Additionally, we demonstrate the significant benefits of primer-independent metatranscriptomics in generating comprehensive datasets that encompass the entire microbial diversity.

## Data availability

The code for the presented analyses is available over GitHub under the following link: <https://github.com/N-Heck-1/Microeukaryotic-predators-shape-the-wastewater-microbiome>.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

We thank Marcel Dominik Solbach for providing the drawing of the wastewater treatment plant that we included in the graphical abstract.

### References

- Aghalari, Z., Dahms, H.U., Sillanpää, M., Sosa-Hernandez, J.E., Parra-Saldívar, R., 2020. Effectiveness of wastewater treatment systems in removing microbial agents: a systematic review. *Global Health* 16, 13. <https://doi.org/10.1186/s12992-0200546-y>.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26, 32–46. <https://doi.org/10.1111/j.14429993.2001.01070.pp.x>.
- Andrews S., 2010. FastQC.
- Aragaw, T.A., 2021. Functions of various bacteria for specific pollutants degradation and their application in wastewater treatment: a review. *Int. J. Environ. Sci. Technol.* 18, 2063–2076. <https://doi.org/10.1007/s13762-020-03022-2>.
- Arregui, L., Liébana, R., Rodríguez, E., Murciano, A., Conejero, F., Pérez-Uz, B., Serrano, S., 2012. Analysis of the usefulness of biological parameters for the control of activated sludge wastewater treatment plants in an interlaboratory study context. *J. Environ. Monit.* 14, 1444. <https://doi.org/10.1039/c2em10861e>.
- Arregui, L., Pérez-Uz, B., Salvadó, H., Serrano, S., 2010. Progresses on the knowledge about the ecological function and structure of the protists community in activated sludge wastewater treatment plants. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, 2, pp. 972–979.
- Assress, H.A., Selvarajan, R., Nyoni, H., Ntushelo, K., Mamba, B.B., Msagati, T.A.M., 2019. Diversity, Co-occurrence and implications of fungal communities in wastewater treatment plants. *Sci. Rep.* 9, 14056. <https://doi.org/10.1038/s41598019-50624-z>.
- Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: a fresh approach to numerical computing. *SIAM Rev.* 59, 65–98. <https://doi.org/10.1137/141000671>.
- Böhme, A., Risse-Buhl, U., Küsel, K., 2009. Protists with different feeding modes change biofilm morphology. *FEMS Microbiol. Ecol.* 69, 158–169. <https://doi.org/10.1111/j.1574-6941.2009.00710.x>.
- Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of southern wisconsin. *Ecol. Monogr.* 27, 325–349.
- Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The new tree of eukaryotes. *Trends Ecol. Evol.* 1, 43–55. <https://doi.org/10.1016/j.tree.2019.08.008>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Cuellar-Bermudez, S.P., Aleman-Nava, G.S., Chandra, R., Garcia-Perez, J.S., Contreras-Angulo, J.R., Markou, G., Muylaert, K., Rittmann, B.E., Parra-Saldívar, R., 2017. Nutrients utilization and contaminants removal. a review of two approaches of algae and cyanobacteria in wastewater. *Algal Res.* 24, 438–449. <https://doi.org/10.1016/j.algal.2016.08.018>.
- Deepayan, S., 2008. *Lattice: Multivariate Data Visualization with R*. Springer.
- Dixon, P., 2003. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
- Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. <https://doi.org/10.1038/nrmicro2832>.
- Ferrera, I., Sánchez, O., 2016. Insights into microbial diversity in wastewater treatment systems: how far have we come? *Biotechnol. Adv.* 34, 790–802. <https://doi.org/10.1016/j.biotechadv.2016.04.003>.
- Fiore-Donno, A.M., Weinert, J., Wubet, T., Bonkowski, M., 2016. Metacommunity analysis of amoeboid protists in grassland soils. *Sci. Rep.* 6, 19068. <https://doi.org/10.1038/srep19068>.
- Foissner, W., 2016. Protists as bioindicators in activated sludge: identification, ecology and future needs. *Eur. J. Protistol.* 55, 75–94. <https://doi.org/10.1016/j.ejop.2016.02.004>.
- Freudenthal, J., Ju, F., Bürgmann, H., Dumack, K., 2022. Microeukaryotic gut parasites in wastewater treatment plants: diversity, activity, and removal. *Microbiome* 10, 27. <https://doi.org/10.1186/s40168-022-01225-y>.
- Gao, P., Xu, W., Sontag, P., Li, X., Xue, G., Liu, T., Sun, W., 2016. Correlating microbial community compositions with environmental factors in activated sludge from four full-scale municipal wastewater treatment plants in Shanghai, China. *Appl. Microbiol. Biotechnol.* 100, 4663–4673. <https://doi.org/10.1007/s00253-016-73070>.
- Ge, S., Wang, S., Yang, X., Qiu, S., Li, B., Peng, Y., 2015. Detection of nitrifiers and evaluation of partial nitrification for wastewater treatment: a review. *Chemosphere* 140, 85–98. <https://doi.org/10.1016/j.chemosphere.2015.02.004>.
- Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., de Groot, G.A., 2015. Not all are freeliving: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Mol. Ecol.* 24, 4556–4569. <https://doi.org/10.1111/mec.13238>.

## Chapter 3

- Gerber, G.K., 2014. The dynamic microbiome. *FEBS Lett.* 588, 4131–4139. <https://doi.org/10.1016/j.febslet.2014.02.037>.
- González-Camejo, J., Barat, R., Pachés, M., Murgui, M., Seco, A., Ferrer, J., 2018. Wastewater nutrient removal in a mixed microalgae–bacteria culture: effect of light and temperature on the microalgae–bacteria competition. *Environ. Technol.* 39, 503–515. <https://doi.org/10.1080/09593330.2017.1305001>.
- Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–338. <https://doi.org/10.1093/biomet/53.3-4.325>.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaultot, D., Zimmermann, P., Christen, R., 2013. The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucl. Acids Res.* 41, 597–604. <https://doi.org/10.1093/nar/gks1160>.
- Herold, M., Martínez Arbas, S., Narayanasamy, S., Sheik, A.R., Kleine-Borgmann, L.A.K., Lebrun, L.A., Kunath, B.J., Roume, H., Bessarab, I., Williams, R.B.H., Gillece, J.D., Schupp, J.M., Keim, P.S., Jäger, C., Hoopmann, M.R., Moritz, R.L., Ye, Y., Li, S., Tang, H., Heintz-Buschart, A., May, P., Muller, E.E.L., Laczny, C.C., Wilmes, P., 2020. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. *Nat. Commun.* 11, 5281. <https://doi.org/10.1038/s41467020-19006-2>.
- Hirakata, Y., Hatamoto, M., Oshiki, M., Watari, T., Kuroda, K., Araki, N., Yamaguchi, T., 2019. Temporal variation of eukaryotic community structures in UASB reactor treating domestic sewage as revealed by 18S rRNA gene sequencing. *Sci. Rep.* 9, 12783. <https://doi.org/10.1038/s41598-019-49290-y>.
- Ju, F., Guo, F., Ye, L., Xia, Y., Zhang, T., 2014. Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ. Microbiol. Rep.* 6, 80–89. <https://doi.org/10.1111/17582229.12110>.
- Kim, T.S., Jeong, J.Y., Wells, G.F., Park, H.D., 2013. General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl. Microbiol. Biotechnol.* 97, 1755–1765. <https://doi.org/10.1007/s00253-012-4002-7>.
- Krueger F., James F., Ewels P., Afyounian E., Schuster-Boeckler B., 2021. TrimGalore. 10.5281/zenodo.5127899.
- LaMartina, E.L., Mohaimani, A.A., Newton, R.J., 2021. Urban wastewater bacterial communities assemble into seasonal steady states. *Microbiome* 9, 116. <https://doi.org/10.1186/s40168-021-01038-5>.
- Lee, N.M., Welander, T., 1996a. Reducing sludge production in aerobic wastewater treatment through manipulation of the ecosystem. *Water Res.* 30, 1781–1790. [https://doi.org/10.1016/0043-1354\(96\)00059-0](https://doi.org/10.1016/0043-1354(96)00059-0).
- Lee, N.M., Welander, T., 1996b. Use of protozoa and metazoa for decreasing sludge production in aerobic wastewater treatment. *Biotechnol. Lett.* 18, 429–434. <https://doi.org/10.1007/BF00143465>.
- Legendre, P., 2008. Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *J. Plant Ecol.* 1, 3–8. <https://doi.org/10.1093/jpe/rtm001>.
- Lentendu, G., Wubet, T., Chatzinotas, A., Wilhem, C., Buscot, F., Schlegel, M., 2014. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil : a multiple barcoding approach. *Mol. Ecol.* 23, 3341–3355. <https://doi.org/10.1111/mec.12819>.
- Limpiyakorn, T., Fürhacker, M., Haberl, R., Chodanon, T., Srithep, P., Sonthiphand, P., 2013. amoA-encoding archaea in wastewater treatment plants: a review. *Appl. Microbiol. Biotechnol.* 97, 1425–1439. <https://doi.org/10.1007/s00253-012-46507>.
- Liu, T., Liu, S., Zheng, M., Chen, Q., Ni, J., 2016. Performance assessment of full-scale wastewater treatment plants based on seasonal variability of microbial communities via high-throughput sequencing. *PLoS One* 11, e0152998. <https://doi.org/10.1371/journal.pone.0152998>.
- Lu, H., Chandran, K., Stensel, D., 2014. Microbial ecology of denitrification in biological wastewater treatment. *Water Res.* 64, 237–254. <https://doi.org/10.1016/j.watres.2014.06.042>.
- Maritz, J.M., Ten Eyck, T.A., Elizabeth Alter, S., Carlton, J.M., 2019. Patterns of protist diversity associated with raw sewage in New York City. *ISME J.* 13, 2750–2763. <https://doi.org/10.1038/s41396-019-0467-z>.
- Matchado, M.S., Lauber, M., Reitmeier, S., Kacprowski, T., Baumbach, J., Haller, D., List, M., 2021. Network analysis methods for studying microbial communities: a mini review. *Comput. Struct. Biotechnol. J.* 19, 2687–2698. <https://doi.org/10.1016/j.csbj.2021.05.001>.
- Okabe, S., Aoi, Y., Satoh, H., Suwa, Y., 2011. Nitrification in Wastewater Treatment. In: Ward, B.B., Arp, D.J., Klotz, M.G. (Eds.), *Nitrification*. John Wiley & Sons Ltd, pp. 405–433. <https://doi.org/10.1128/9781555817145.ch16>.
- Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlenn D., Minchin P.R., O’Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., Wagner H., 2020. Vegan community ecology package.

## Chapter 3

---

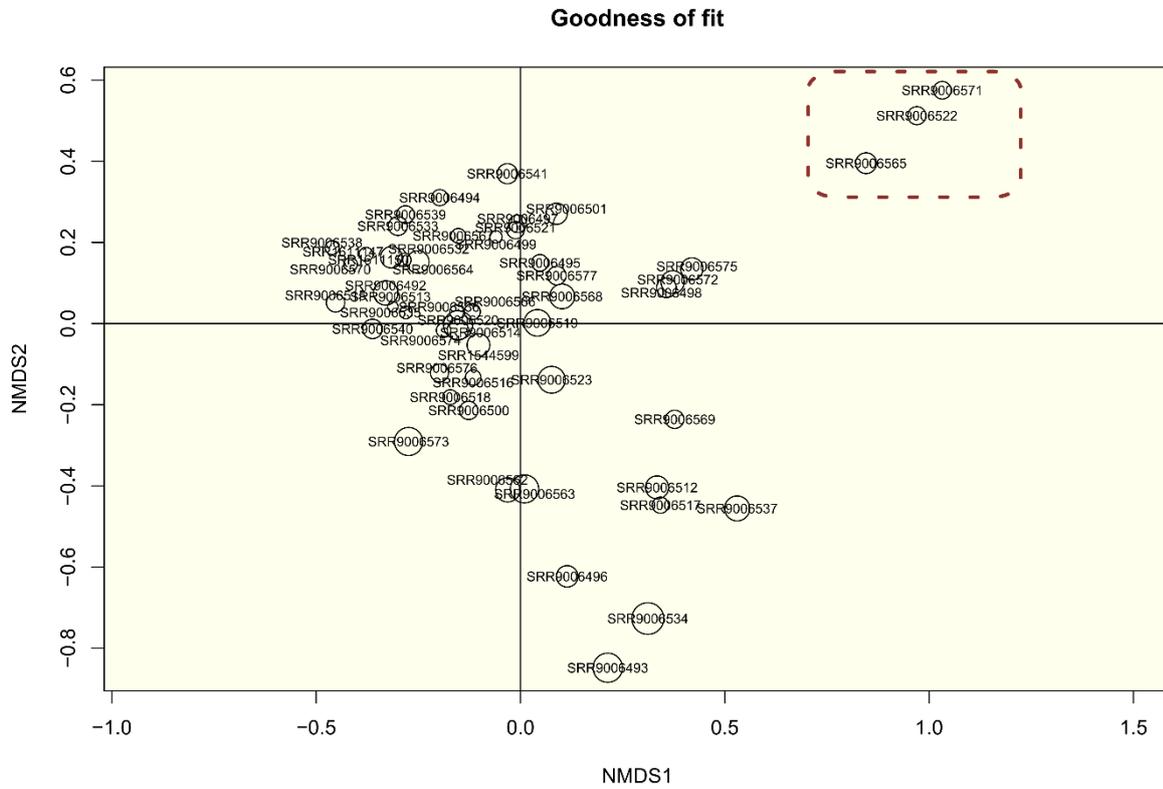
- Öztoprak, H., Walden, S., Heger, T., Bonkowski, M., Dumack, K., 2020. What drives the diversity of the most abundant terrestrial cercozoan family (Rhogostomidae, cercozoa, rhizaria)? *Microorganisms* 8, 1123. <https://doi.org/10.3390/microorganisms8081123>.
- Pan, K.L., Gao, J.F., Fan, X.Y., Li, D.C., Dai, H.H., 2018. The more important role of archaea than bacteria in nitrification of wastewater treatment plants in cold season despite their numerical relationships. *Water Res.* 145, 552–561.
- Paradis, E., 2006. *Analysis of Phylogenetics and Evolution with R*. Springer, New York, NY. <https://doi.org/10.1007/978-0-387-35100-1>.
- Park, H.D., Wells, G.F., Bae, H., Criddle, C.S., Francis, C.A., 2006. Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors. *Appl. Environ. Microbiol.* 72, 5643–5647. <https://doi.org/10.1128/AEM.00402-06>.
- Pauli, W., Jax, K., Berger, S., Beek, B., 2001. Protozoa in Wastewater Treatment: function and Importance. Biodegradation and Persistence, *The Handbook of Environmental Chemistry*. Springer-Verlag, Berlin/Heidelberg, pp. 203–252. [https://doi.org/10.1007/10508767\\_3](https://doi.org/10.1007/10508767_3).
- Pohl, N., Solbach, M.D., Dumack, K., 2021. The wastewater protist *Rhogostoma minus* (Thecofilosea, Rhizaria) is abundant, widespread, and hosts Legionellales. *Water Res.* 203, 117566 <https://doi.org/10.1016/j.watres.2021.117566>.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.* 35, 7188–7196. <https://doi.org/10.1093/nar/gkm864>.
- R Core Team, 2021. R: a language and environment for statistical computing.
- Ramirez, E., Martinez, B., Gonzalez, M.E., Robles, E., Choncohua, E., Galan, C., 2015. Microbiological and physicochemical characteristics of a pilot plant of activated sludge. *Int. J. Eng. Res. Technol.* 4, 1236–1240.
- Rodríguez, E., García-Encina, P.A., Stams, A.J.M., Maphosa, F., Sousa, D.Z., 2015. Metaomics approaches to understand and improve wastewater treatment systems. *Rev. Environ. Sci. Bio Technol.* 14, 385–406. <https://doi.org/10.1007/s11157-015-9370x>.
- Röttjers, L., Faust, K., 2018. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol. Rev.* 42, 761–780. <https://doi.org/10.1093/femsre/fuy030>.
- Roume, H., EL Muller, E., Cordes, T., Renaut, J., Hiller, K., Wilmes, P., 2013. A biomolecular isolation framework for eco-systems biology. *ISME J.* 7, 110–121. <https://doi.org/10.1038/ismej.2012.72>.
- Samba-Louaka, A., Delafont, V., Rodier, M.H., Cateau, E., Héchard, Y., 2019. Free-living amoebae and squatters in the wild: ecological and molecular features. *FEMS Microbiol. Rev.* 43, 415–434. <https://doi.org/10.1093/femsre/fuz011>.
- Schages, L., Wichern, F., Kalscheuer, R., Bockmühl, D., 2020. Winter is coming – Impact of temperature on the variation of beta-lactamase and mcr genes in a wastewater treatment plant. *Sci. Total Environ.* 712, 136499 <https://doi.org/10.1016/j.scitotenv.2020.136499>.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: opensource, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- Sibbald, S.J., Archibald, J.M., 2017. More protist genomes needed. *Nat. Ecol. Evol.* 1, 0145 <https://doi.org/10.1038/s41559-017-0145>.
- SIVÉC, 2023 Station D’Epuration [WWW Document]. SIVÉC – syndicat intercommunal à vocation ECologique. URL <https://sivec.lu/installation/station-depuration/> (accessed 6.21.23).
- Tackmann, J., Matias Rodrigues, J.F., von Mering, C., 2019. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Syst.* 9 <https://doi.org/10.1016/j.cels.2019.08.002>, 286–296. e8.
- Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 3, e2527. <https://doi.org/10.1371/journal.pone.0002527>.
- Utz, L.R.P., Bohrer-Morel, M.B.C., 2008. Characterization of the zooplankton community of the secondary wastewater treatment system of an oil refinery in Southern Brazil. *BIOCIÊNCIAS* 16, 1–14.
- Vaerewijck, M.J.M., Sabbe, K., Baré, J., Houf, K., 2011. Occurrence and diversity of freeliving protozoa on butterhead lettuce. *Int. J. Food Microbiol.* 147, 105–111. <https://doi.org/10.1016/j.ijfoodmicro.2011.03.015>.

## Chapter 3

---

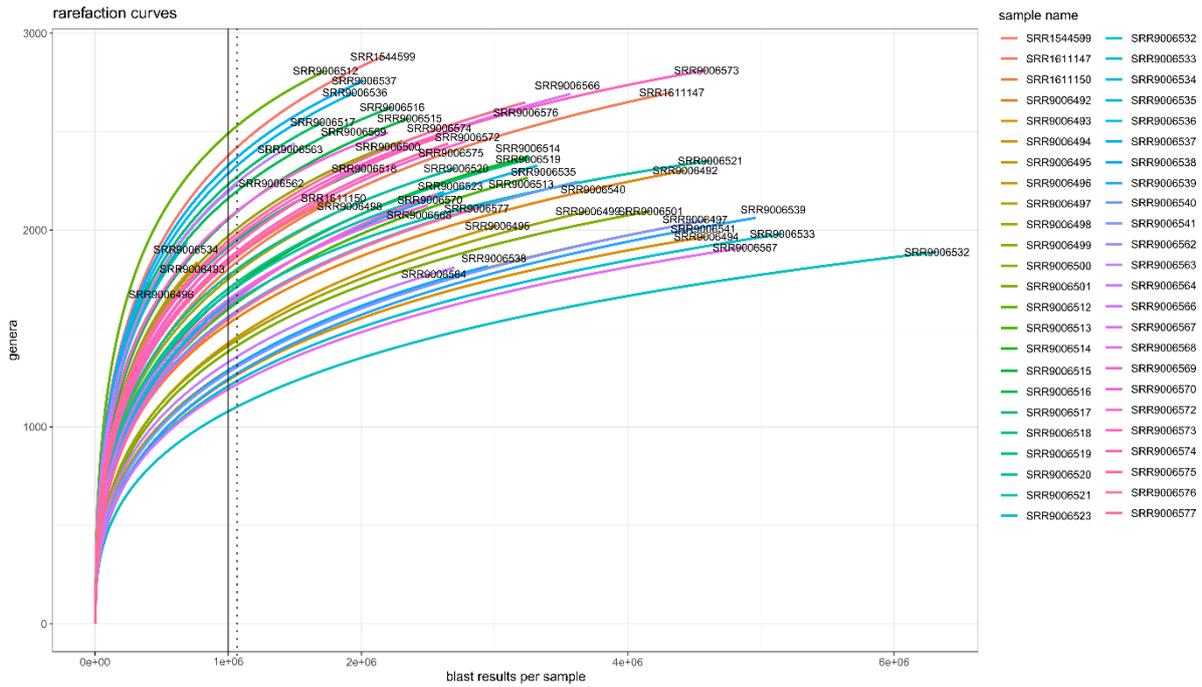
- Wickham, H., 2016. *ggplot2: Elegant Graphics For Data Analysis*, 2nd ed. Springer, New York, NY. 2016 Edition. ed.
- Wickham H., François R., Henry L., Müller K., 2022. *dplyr: a grammar of data manipulation*. 1.0.7.
- Wickham H., Girlich M., 2022. *Tidyr: tidy messy data*.
- Wu, Linwei, Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Qiuting, Brown, M.R., Li, Z., Van Nostrand, J.D., Ling, F., Xiao, N., Zhang, Ya, Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Acevedo, D., Agullo-Barcelo, M., Alvarez, P.J.J., Alvarez-Cohen, L., Andersen, G.L., de Araujo, J.C., Boehnke, K.F., Bond, P., Bott, C. B., Bovio, P., Brewster, R.K., Bux, F., Cabezas, A., Cabrol, L., Chen, S., Criddle, C.S., Deng, Y., Etchebehere, C., Ford, A., Frigon, D., Sanabria, J., Griffin, J.S., Gu, A.Z., Habagil, M., Hale, L., Hardeman, S.D., Harmon, M., Horn, H., Hu, Z., Jauffur, S., Johnson, D.R., Keller, J., Keucken, A., Kumari, S., Leal, C.D., Lebrun, L.A., Lee, J., Lee, M., Lee, Z.M.P., Li, Y., Li, Z., Li, M., Li, X., Ling, F., Liu, Y., Luthy, R.G., Mendonça-Hagler, L.C., de Menezes, F.G.R., Meyers, A.J., Mohebbi, A., Nielsen, P.H., Ning, D., Oehmen, A., Palmer, A., Parameswaran, P., Park, J., Patsch, D., Reginatto, V., de los Reyes, F.L., Rittmann, B.E., Noyola, A., Rossetti, S., Shan, X., Sidhu, J., Sloan, W.T., Smith, K., de Sousa, O.V., Stahl, D.A., Stephens, K., Tian, R., Tiedje, J.M., Tooker, N.B., Tu, Q., Van Nostrand, J.D., De los Cobos Vasconcelos, D., Vierheilig, J., Wagner, M., Wakelin, S., Wang, A., Wang, B., Weaver, J.E., Wells, G.F., West, S., Wilmes, P., Woo, S.G., Wu, Linwei, Wu, J.H., Wu, Liyou, Xi, C., Xiao, N., Xu, M., Yan, T., Yang, Y., Yang, M., Young, M., Yue, H., Zhang, B., Zhang, P., Zhang, Qiuting, Zhang, Ya, Zhang, T., Zhang, Qian, Zhang, W., Zhang, Yu, Zhou, H., Zhou, J., Wen, X., Curtis, T.P., He, Q., He, Z., Brown, M.R., Zhang, T., He, Z., Keller, J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis, T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., Consortium, G.W.M., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* 4, 1183–1195. <https://doi.org/10.1038/s41564-019-0426-5>.

Supplementary material



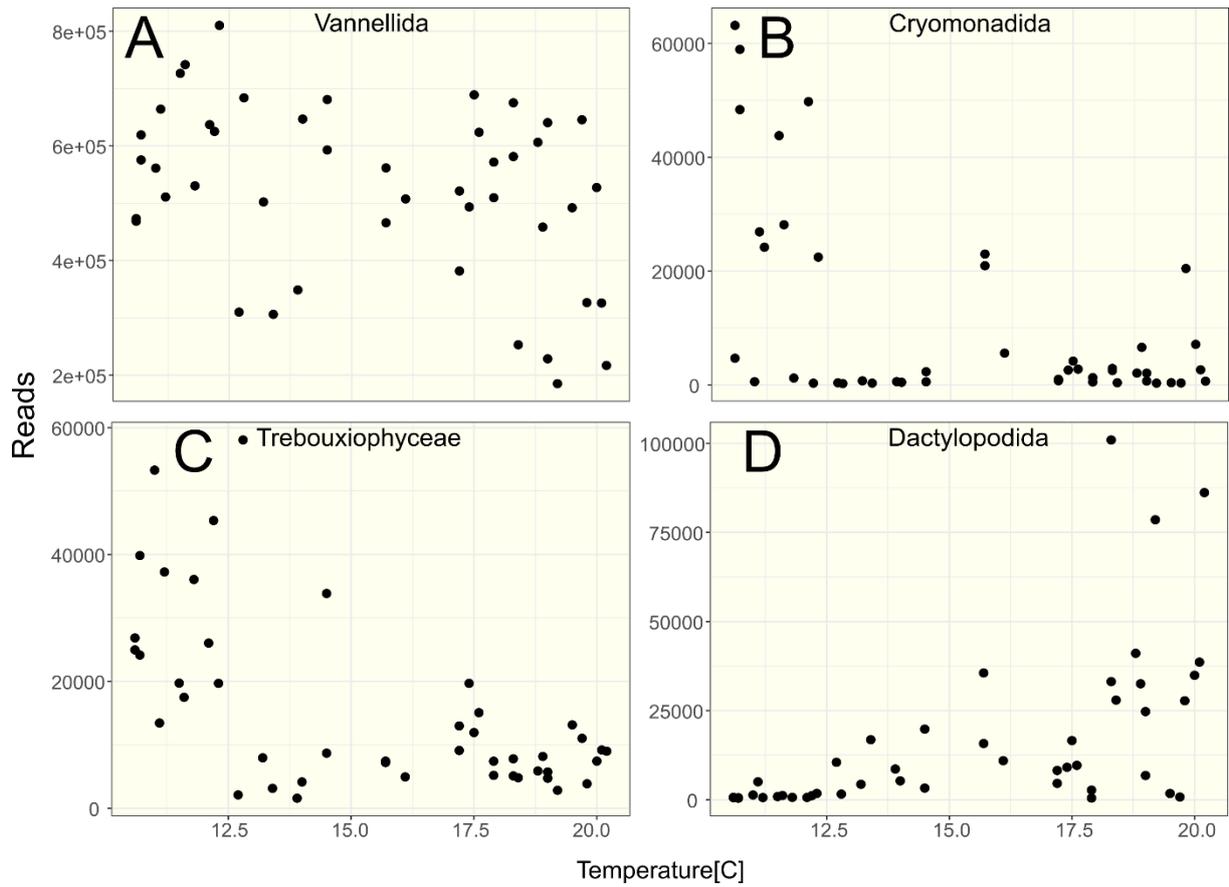
**Supplementary Figure 1: Nonmetric multidimensional scaling plot used to determine outliers.** Each site represents one sample. Circle size represents the goodness of fit. Outlier samples to be removed are marked by a red frame.

# Chapter 3



**Supplementary Figure 2: Rarefaction curves.** The number of OTUs at the genus level is plotted against the number of reads for each sample. The black line denotes  $1e^6$  rRNA reads, which was chosen as the cutoff for sequencing depth. The dotted line is set at the number of reads to which all samples were rarefied – 1,066,81.

## Chapter 3



**Supplementary Figure 3: Influence of temperature on selected prominent eukaryotic orders.** The abundance of selected eukaryotic orders plotted against water temperature. Note that Vannellida (A, Amoebozoa), Cryomonadida (B, Cercozoa), and not further determined Trebouxiophyceae (C) decreased with increasing temperature, while Dactylopodida (D, Amoebozoa) increased.

---

## CHAPTER 4

### Biotic interactions explain seasonal dynamics of the alpine soil microbiome

Anna Maria Fiore-Donno<sup>1</sup>, Jule Freudenthal<sup>1</sup>, Mathilde Borg Dahl<sup>2</sup>, Christian Rixen<sup>3,4</sup>, Tim Urich<sup>2</sup>, Michael Bonkowski<sup>1</sup>

<sup>1</sup>Institute of Zoology, University of Cologne, Zuelpicher Str. 47b, 50674 Cologne, Germany

<sup>2</sup>Institute of Microbiology, University of Greifswald, 17489 Greifswald, Germany

<sup>3</sup>WSL Institute for Snow and Avalanche Research SLF, 7260 Davos Dorf, Switzerland

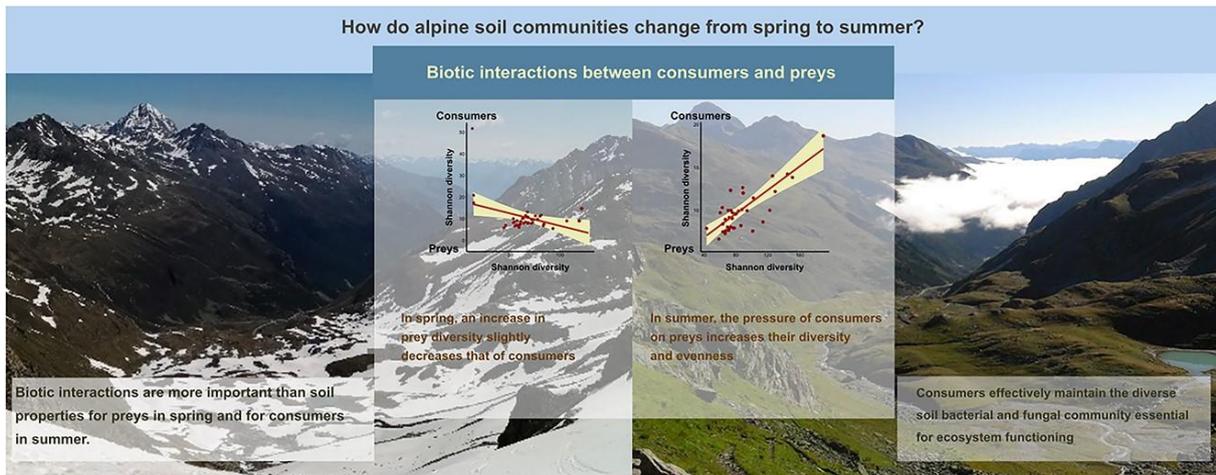
<sup>4</sup>Climate Change, Extremes and Natural Hazards in Alpine Regions Research Centre CERC, 7260 Davos Dorf, Switzerland

*Keywords: biotic interactions, alpine ecology, protists, metatranscriptomics, soil food web, soil ecology, community ecology*

#### Abstract

While it is acknowledged that alpine soil bacterial communities are primarily driven by season and elevation, there is no consensus on the factors influencing fungi and protists. Here we used a holistic approach of the microbiome to investigate the seasonal dynamics in alpine grasslands, focusing on soil food web interactions. We collected 158 soil samples along elevation transects from three mountains in the Alps, in spring during snowmelt and in the following summer. Using metatranscriptomics, we simultaneously assessed prokaryotic and eukaryotic communities, further classified into trophic guilds. Our findings reveal that the consumers' pressure increases from spring to summer, leading to more diverse and evenly distributed prey communities. Consequently, consumers effectively maintain the diverse soil bacterial and fungal communities essential for ecosystem functioning. Our research highlights the significance of biotic interactions in understanding the distribution and dynamics of alpine microbial communities.

## Graphical Abstract



We investigated the seasonal dynamics and interactions of soil food webs in the Alps. Our metatranscriptomic results from 158 samples reveal (i) that biotic interactions can explain more variation in the microbial communities than topographic and edaphic variables; (ii) a seasonal dynamic in biotic interactions: consumers pressure on preys increases from spring to summer; (iii) that, contrary to current models, there is no shift in communities during thaw.

## 1 Introduction

Understanding the seasonal dynamics of the soil microbiome is an important step towards modelling the effects of climate warming, i.e. whether and how the balance between the carbon stored in the soil and the CO<sub>2</sub> released to the atmosphere will be altered [1]. It is not yet clear how the main components of the soil food web—bacterial and fungal primary decomposers and their main consumers (predatory bacteria, heterotrophic protists and bacterivorous and fungivorous nematodes)—are structured in alpine regions, and in particular which of the abiotic drivers, elevation or season, has a preponderant influence [2, 3]. Along altitudinal gradients, bacterial and fungal diversity generally decreases with elevation [4], while protistan diversity increases [3]. While most studies agree on the importance of pH and elevation in shaping bacterial communities in general, the unexplained variance is usually greater for fungi and protists [3-5].

Recently, models that include biotic interactions in addition to abiotic environmental variables have been shown to explain significantly more variation in microbial metacommunity assembly [6-8]. Biotic interactions, such as competition and predation, can promote the coexistence or exclusion of species [9]. Thus, the inclusion of biotic interactions is of great importance for process-based understanding and prediction of ecological responses [10]. For instance, ecosystem services, such as C and N cycling, in which soil microbes play an important role, can be altered by consumer–prey interactions [11]: It has been suggested that in soil microbiomes with higher productivity there is increased predation on

lower trophic levels, influencing carbon flow through the belowground food web [12]. In addition, predation contributes to the formation of the soil microbial necromass, the importance of which, accounting for up to 50% of soil organic matter, has only recently been recognized [13]. It is therefore crucial to emphasize the role of dynamic trophic relationships, also as a cause of microbial mortality [14].

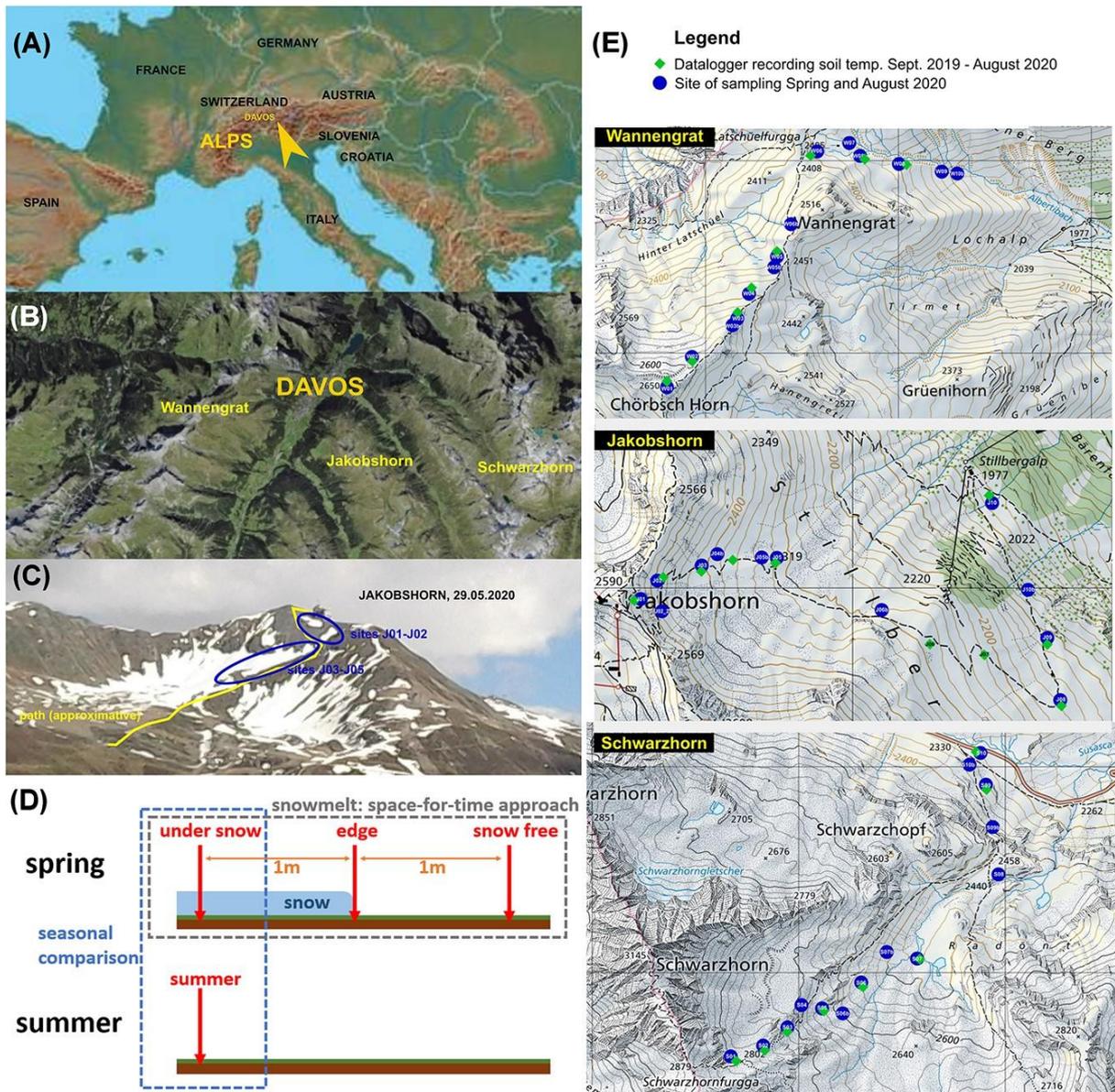
Our study aimed at disentangling the interactions of seasonal, topographic, edaphic and biotic factors and how they shape the soil microbiome. To this end, we used metatranscriptomics to simultaneously assess the prokaryotic and eukaryotic microbial diversity in alpine grasslands in Switzerland, near Davos (Fig. 1A and B), at a regional scale (158 samples on three different mountains), taking into account elevation (altitudinal transects from c. 1900 to 2800 m a.s.l) and season (spring and summer; Fig. 1C-E). Small subunit ribosomal RNAs (16S and 18S) were used to identify the bacterial, archaeal, fungal, metazoan, and protistan communities to the genus level. To investigate biotic interactions, identified genera were further classified, where possible, according to nutrition and lifestyle.

We addressed the following key questions: (i) which are the dominant components of the soil microbiome during snowmelt and what is the fate of these communities in summer? (ii) Do consumers contribute to the seasonal turnover of the bacterial and fungal communities? (iii) Which are the main drivers of alpine soil communities, topography, soil properties, seasonality or biotic interactions? We hypothesized that consumers would play a major role in shaping prey communities, with an increase in consumers and a decrease in preys during the season, with season playing a more important role than elevation. Finally, we hypothesized that each trophic guild would differ in its response to environmental and seasonal changes.

## 2 Materials and methods

### *2.1 Sampling design*

The study area is located around the town of Davos in the eastern central Alps in Switzerland (Fig. 1). The climate is characterized by an average annual temperature of 3.5 °C, 193 frost days, a total annual precipitation of 1022 mm and fresh snow on 69.1 days per year [15]. Sampling was designed to capture the biogeochemical changes that occur during snowmelt in alpine grasslands. For this, we selected 15–16 sites per mountain in spring 2020, where snow patches were still present. At each site, we collected three soil samples constituting a time series: the first sample under the snow, the second at the edge of the patch, i.e. exactly during snowmelt, and the third in soil recently cleared of snow (Fig. 1D). To assess a medium-scale reproducibility of our data, we repeated the same scheme in three mountains (Fig. 1E), while the altitudinal gradient from 1972 to 2816 m would allow comparing the



**Figure 1:** Maps of the collection sites and sampling scheme. (A) The Alps and Davos in Europe. (B) The three mountains where the study took place. (C) View of the sites at the top of the Jakobshorn during the spring sampling. (D) A scheme of the sampling design per site. (E) Sites and dataloggers position along the altitudinal gradient in each mountain.

biogeochemical data between early and later snowmelt periods. To establish a seasonal cycle, we sampled again in August 2020 the same sites that were under snow in spring.

## 2.2 One year of soil temperature records

In each mountain, 10 dataloggers (ibutton DS1922L, Maxim Integrated Products, San Jose, CA, US) set up to record the temperature every 7200 s to an accuracy of 0.0625 °C were placed in early September 2019 c. every 50 m along the elevation gradient, and recovered in August 2020, as previously described (Rindt et al. 2023) (Fig. S1). Coordinates and heights were determined using a GPS (Trimble Geo XH 6000, Trimble Inc. Sunnyvale, CA, US) with an accuracy of a few cm (Table S1).

## Chapter 4

---

### *2.3 Soil sampling*

Spring sampling took place between 19.5 and 24.6.2020; the three highest sites on the Schwarzhorn could only be sampled on 6 July. The summer sampling took place from 20 to 25 August. Soil temperature at the time of collection was recorded using the described dataloggers, inserted at a depth of c. 5–8 cm, and left for c. 10 min. to record the temperature to an accuracy of 0.0625 °C (Table S1). Two to 4 g of wet soil for RNA extraction were collected with a clean plastic spoon and immediately placed in a sterile, RNase-free 15 ml plastic tube containing 6.5 ml of Life Guard soil RNA (Qiagen GmbH, Hilden, Germany) and stored in an insulated box with cooling packs. The samples were stored at –20 °C as soon as we returned to the WSL Institute in Davos. They were not allowed to thaw until extraction. Soil (c. 200 g) for determining edaphic properties was collected with a clean metal spoon and stored at 4 °C in a polyethylene bag.

### *2.4 RNA extraction, reverse-transcription, library preparation and sequencing*

Prior starting the following steps, great care was taken to work in an RNase-free environment, notably by treating all objects that would come into contact with the samples with RNaseZap, an RNase decontamination solution (Sigma-Aldrich, MO, USA). The tubes containing the soil samples were thawed, centrifuged, and the buffer removed. Circa 1 g of wet soil was removed with a spatula and transferred to the RNeasy PowerSoil Total RNA kit vials (Qiagen GmbH, Hilden, Germany). The manufacturer's protocol was strictly followed, except for the disruption step, which was carried on an MP Biomedicals FastPrep-24 homogenizer for 30 s at 5 m/sec. The RNA was eluted in 50 µl of SR7 buffer, with the addition of 1 µl of recombinant RNasin ribonuclease inhibitor (Promega, Madison, WI, USA). DNAs were digested with DNase I (New England BioLabs, MA, USA) and proteins and small RNAs were removed using the Megaclear kit (Invitrogen, CA, USA), following the manufacturer's protocol. Samples were eluted with 50 µl of preheated elution buffer, and quantified using a Qubit 30 Fluorometer (Invitrogen, CA, USA) using 2 µl of the RNA in the high sensitivity buffer. Quality was estimated with a 2100 Bioanalyzer (Agilent, CA, USA) using the Prokaryote Total RNA Nano assay. Samples with a concentration < 11 ng/µl were precipitated with 1:10 volume of 5 M ammonium acetate and washed with ethanol, according to the protocol of the Megaclear kit, to reach an RNA concentration > 10 ng/µl. Libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, Ipswich, MS, USA) without rRNA removal or mRNA selection. The incubation time of the first strand cDNA synthesis at 42 °C was increased from 15 to 50 min. To select cDNA fragments of 370–600 bp after the second strand synthesis, the fragmentation time was reduced to 10 min. The library size option "400 bp" was selected and the final libraries were amplified with 12 polymerase chain reaction (PCR) cycles. The libraries were sequenced in a single complete run of

## Chapter 4

---

NovaSeq SP FC (Illumina Inc., San Diego, CA, US), length of paired sequences of 250 bp, at the Cologne Genomic Centre, University of Cologne, Germany.

### *2.5 Sequence analyses - filtering and identification*

We obtained  $1.02 \times 10^9$  raw sequences, which were submitted to the PhyloFLASH processing pipeline using default settings [16]. In brief, the pipeline identifies SSU rRNA sequences by aligning unpaired sequences to a filtered SILVA database (v.138, NR99), from which LSU, low-complexity, and cloning vector fragments have been removed. SSU sequences were identified using a short sequence aligner for DNA and RNA-seq data [17], with the default setting of a minimum identity of 70%. Bacteria and Archaea were taxonomically assigned by taking the last common ancestor of the taxonomy strings of all the hits. Eukaryotic forward and reverse SSU sequences were assembled using FLASH [18] and low quality sequences were filtered out with default settings. Eukaryotic sequences were identified to the genus level using a slightly modified PR2 database [19], using Blast + [20] with an e-value of  $1e-10$  and keeping only the best hit. Unicellular eukaryotes were classified as protists, and additional information on lifestyle (free-living, plant, or animal parasite) and nutrition (heterotroph, autotroph, or mixotroph) was added whenever possible, according mostly to [21]. Nematodes were classified as eukaryvore, bacterivore, plant-feeding, fungivore, animal parasite, or omnivore [22, 23]. To assess the biotic interactions, we considered broadly defined groups of consumers, i.e. bacterial predators—the phyla Myxococcota and Bdellovibrionota (94.7% of the assemblage), heterotrophs and free-living protists (4.3%), and free-living nematodes (1%), and preys, i.e. all other bacteria (99.2%), fungi (0.77%), and autotrophic protists (unicellular algae, 0.03%) (Table S2).

### *2.6 Edaphic properties and vegetation survey*

Soil water content, pH, soil organic C and total N, soil microbial biomass C and N, dissolved organic C, and total dissolved N were measured as previously described [24]. Vegetation was recorded from June to August in 2020 and 2021, within a 40 cm diameter circle around the spot where the soil sample was taken. All vascular plants rooted within the surface were identified according to [25], and the percentage of the surface they covered was estimated (Table S1).

### *2.7 Statistical analyses*

All statistical analyses were carried out within the R environment (R v. 4.1.3) [26] on the taxonomic abundance/sample (Table S2) and on the sample topographic, edaphic, and biotic characteristics (Table S1). Unless otherwise specified, community analyses were performed with the vegan package 2.5–7 [27]. To assess whether more sampling and sequencing effort would have revealed more

## Chapter 4

---

richness, we performed an analysis based on accumulation curves (function *specaccum*) and rarefaction curves (package and function iNEXT 3.0.0), using the abundance table, with a 97% confidence interval, 50 bootstraps and 50 knots; the latter function also calculates species richness (observed and estimated) and sample coverage. Exponential Shannon indices were calculated with the function *renyi* (with the hill parameter, on sample-standardized data, with function *decostand*, method “total”). Significant differences in sample-standardized sequence counts, alpha diversity and evenness between seasons were determined by analysis of variance and Tukey tests (package agricolae 1.3-5, function *aov* and *HSD.test*) with a  $P \leq 0.05$ , while correlations between the same data were determined with the function *lm*.

Beta diversity between mountains, altitude and snow coverage in spring was inferred by Principal Coordinate Analysis (function *cmdscale*), using Bray–Curtis dissimilarities (function *vegdist*, method “bray”) on the sample-standardised taxa of interest (Bacteria, Archaea, protists, Fungi and the functional groups consumers and preys), then plotted with the package ggplot2 3.3.5. Principal component analysis revealed an influence of altitude and mountain on bacterial beta diversity, but this effect was mainly driven by the three highest sites in the Schwarzhorn (Fig. S2). When these outliers were removed, a decrease of the variation explained by the first axis was observed (40.1 to 31.3%), and a trend for altitude was visible. Fungal, metazoan, protistan and consumer communities displayed no clear trend with increasing elevation (Fig. S2). Variation partitioning (function *varpart* applied to the Hellinger-transformed taxa dataset and using RDA, function *rda*) was used to assess the proportion of beta diversity explained by each of the factors mountain, altitude and spring snow coverage.

Differential abundances of the most abundant taxa across seasons were calculated with the package DESeq2 1.30.1 [28]; DESeq objects were created with function *DESeqDataSetFromMatrix* and normalised (function *estimateSizeFactors*), then the differential expression was calculated using *DESeq* with the parameters `minReplicatesForReplace = Inf`, `sftype = “poscounts”`. Results with an adjusted p-value  $< 0.01$  and an absolute  $\log_2$ fold  $> 0.5$  were considered as significant and plotted with ggplot2 (*geom\_segment*).

Distance-based redundancy analysis (dbRDA, function *dbRDA*) was conducted to describe the influence of environmental factors (scaled with function *scale*) on the distribution of the abovementioned taxa of interest (standardized by samples as above). The most influential variables were identified with the function *ordistep* based on the Akaike Information Criterion, and the resulting model tested with *anova*.

To estimate the proportion of variance accounted for by topographic (mountain and altitude), biotic (Shannon vascular plants, Shannon preys or consumers), and edaphic (soil temperature, water content, pH and organic C) factors in the diversity of consumers and preys by season, variance partitioning analyses were performed. The variance attributed to each category of factors

(topographic, biotic, and edaphic) and their intersections was estimated using the  $R^2$  of the linear models (function *lm*) of log-transformed Shannon indices of preys or consumers versus the other categories of factors.

To investigate associations between consumers and preys and among consumers, network analyses were performed for each season on the most abundant taxa (Table S3), further filtered by prevalence, i.e. selecting taxa that were present in more than one third of the sites. The filtered out taxa were binned into a “pseudo-taxon”, which was used for inferences to avoid altering the ratio between taxa, as recommended by [29], but was not shown in the final results. The co-occurrence network was calculated at the genus level using FlashWeave [30], a package implemented in julia v.1.7.1 [31] with parameters for homogeneous data (sensitive = true and heterogeneous = false). Co-occurrences were estimated in several steps: first, for each genus, all directly associated neighbours were computed; then, individual neighbours were connected through a combinatorial strategy to form a global association graph, with an optimized sequence of statistical tests for conditional independence. The final set of directly associated neighbours contained only genera that were conditionally dependent on each other [30]. The network was then summarized (using R) to higher taxa and functions and visualized using Cytoscape v.3.9.1 [32], which was also used to analyse original networks (not summarized). We excluded self-loops, nodes with only one association and associations within preys.

### 3 Results

#### *3.1 Metatranscriptomics*

From the 158 collected soil samples, we obtained more than one billion RNA sequences (on average c. 2.5 million/sample), resulting in c. 8000 identified taxa at the genus level (on average 2349 per sample) (Table S4). Rarefaction curves describing the observed number of genera as a function of the number of sequences (Fig. S3) suggested that our sequencing effort was sufficient, confirmed by the estimated sample coverage (between 99.95 and 100%, Table S2). Therefore, and because all samples were sequenced in a single run, we considered that differences in the number of sequences between, e.g. taxonomic or trophic groups, was informative per se. Accumulation curves describing the number of genera as a function of the number of samples, binned by altitude, mountain, and snow coverage (Fig. S4) did not reach a plateau and showed slight differences between environments, thus indicating differences in beta diversity.

#### *3.2 Taxonomic and functional diversity*

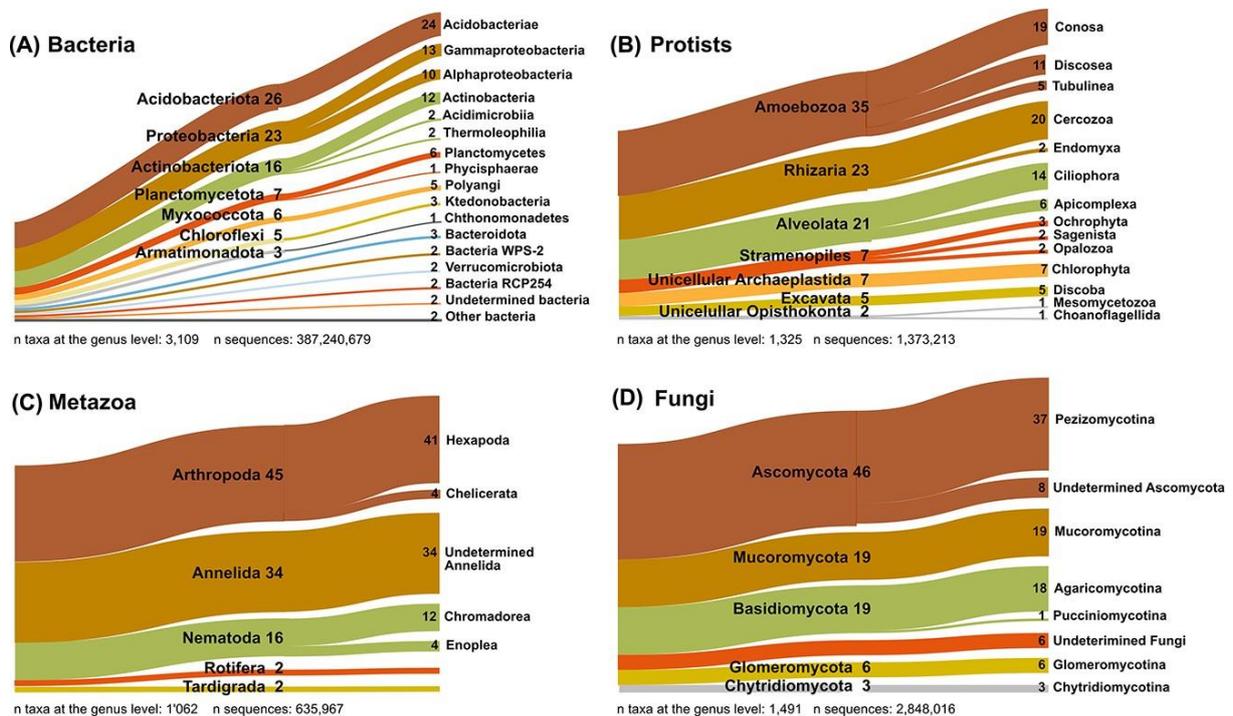
Prokaryotes dominated the total assemblage with 98.02% of the identified taxa, of which only 0.06% were Archaea (Table S2). Among bacteria, it is noteworthy that the predatory Myxococcota

## Chapter 4

represented almost 6% of the bacterial genera (Fig. 2a and Note S1), making it the main consumer in soil, surpassing the heterotrophic protists in abundance. Among eukaryotes, fungi and metazoa accounted each for 36% of the SSU rRNAs, protists for 17% and multicellular plants for 11% (Table S2). More than a third of the protistan sequences were assigned to the phylum Amoebozoa (35%; Fig. 2B), including the very elusive slime mould Myxogastria. The phylum Rhizaria (23%) was in majority composed of Cercozoa (20%). Our functional classification by nutrition showed that 88% of the protists were heterotrophs, 8% autotrophs, and 2% mixotrophs. In terms of lifestyle, 90% were free-living, 7% were animal parasites, and 2% were plant parasites (Table S2).

Among the animals (Metazoa), insects (41%), and ringed worms (34%) were most abundant (Fig. 2C), followed by nematodes (16%). Nematodes were classified as bacterivores (29%), plant-feeding (25%), fungivores (16%), eukaryvores (10%), omnivores (3%), and animal parasites (1%; Table S2).

The fungi were dominated by Ascomycota (46%; Fig. 2D), mainly Pezizomycotina (37%), which also makes up most of Ascomycota in terms of described species. Of the 53% of the fungi that could be functionally identified, saprotrophs were the largest group (35%), followed by plant parasites (7%), endomycorrhizals (6%), ectomycorrhizals (5%), lichenized (5%), ericoid-mycorrhizals (2%), and other parasites and endosymbionts each representing <1% (Table S2).



**Figure 2:** Sankey diagrams showing the relative proportion of high-rank taxonomic groups, based on percentage of sequences identified to the genus level. Taxa <1% are not shown. (A) Bacteria. (B) Protists. (C) Metazoa. (D) Fungi.

## Chapter 4

---

### *3.3 Soil temperatures and edaphic parameters*

During our study, the snow cover lasted on average 228 days (max. = 286, min. = 186, sd = 21.2), 22 days longer on the high than on the low altitudes (Table S1). Soil temperatures were on average 2 °C lower on the high than on the low altitudes (Table S1) and showed a day/night variation before and after the snow covered the soil; during winter under the snow cover, temperatures were stable and remained constantly >0 °C (Fig. S1). Temperatures recorded during the sampling increased during snowmelt and further increased in summer, mirrored by an opposite trend in soil water content. The pH did not significantly vary during snowmelt. Microbial biomass and dissolved C increased during and immediately after snowmelt, as well as from spring to summer (Table S1) [24].

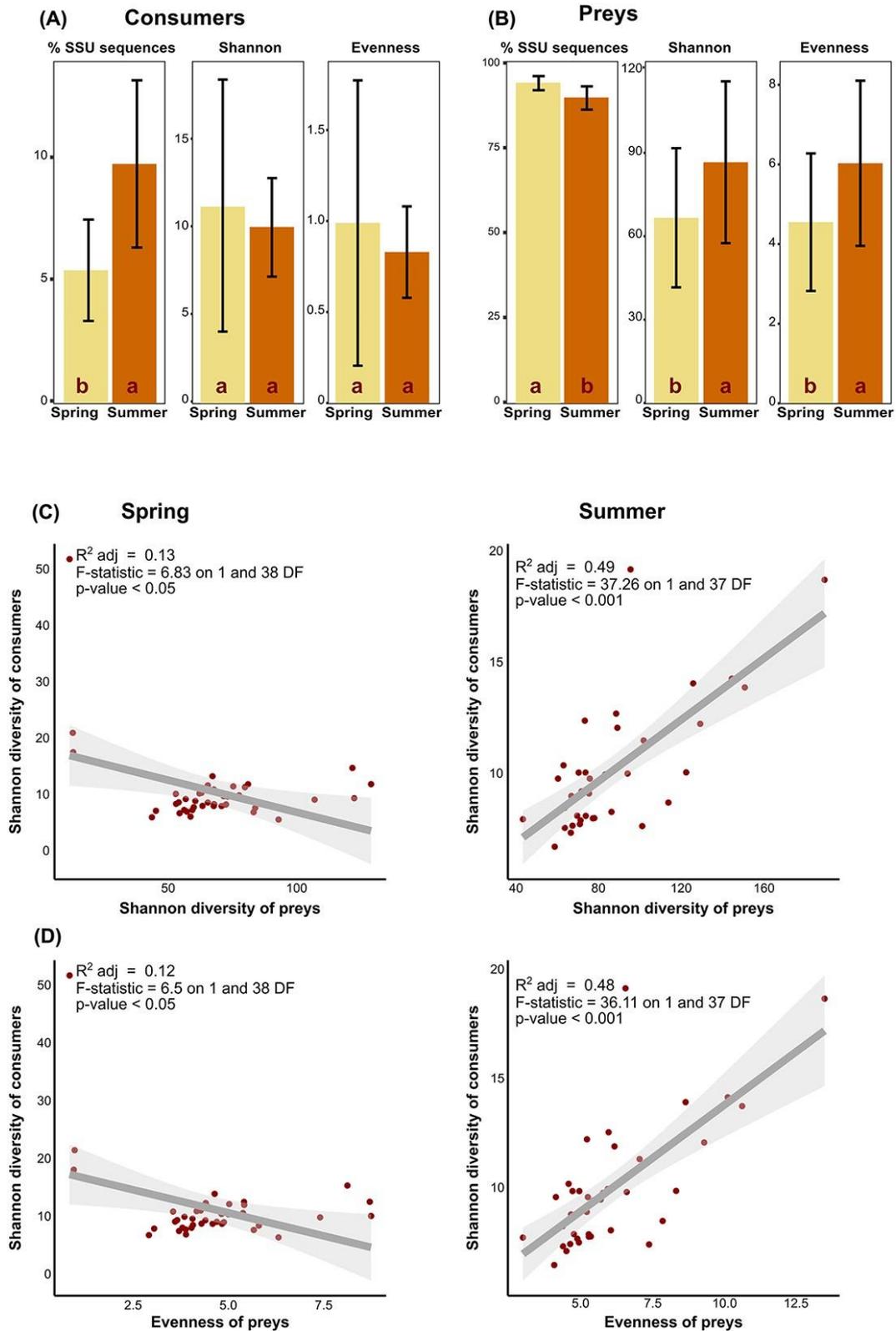
### *3.4 Dynamics during snowmelt and from spring to summer*

In spring, despite sharp environmental changes across the snowmelt gradient, no significant differences in the abundance, diversity or evenness of the communities of consumers and preys were observed (Fig. S5). In contrast, abundance, diversity, and evenness varied between the sample collected under the snow in spring and the sample collected at the same spot in summer, with differences between functional guilds. For instance, the standardized counts of consumer SSU rRNAs sequences was significantly more abundant in summer than in spring, but no significant differences in diversity and evenness were found (Fig. 3A—expressed as percentage of total sequences). In contrast, the prey SSU rRNAs were more abundant in spring, but preys were more diverse and even in summer (Fig. 3B). Changes in diversity and evenness of consumers and their potential prey were correlated. There were negative correlations between the diversity and the evenness of consumers and those of preys in spring (Fig. 3C and D). On the contrary, the same negative correlations were strong and positive in summer (Fig. 3C and D). We questioned whether the increase of consumers in spring was related to the main abiotic changes occurring from spring to summer, i.e. a warmer and drier soil; we found positive linear correlations between soil temperature and water content and the richness of Myxococcota ( $F_{[1, 156]} = 7.6$ ,  $P = .007$ ) and Rhizaria ( $F_{[1, 156]} = 14.27$ ,  $P < .001$ ), but not for that of Amoebozoa.

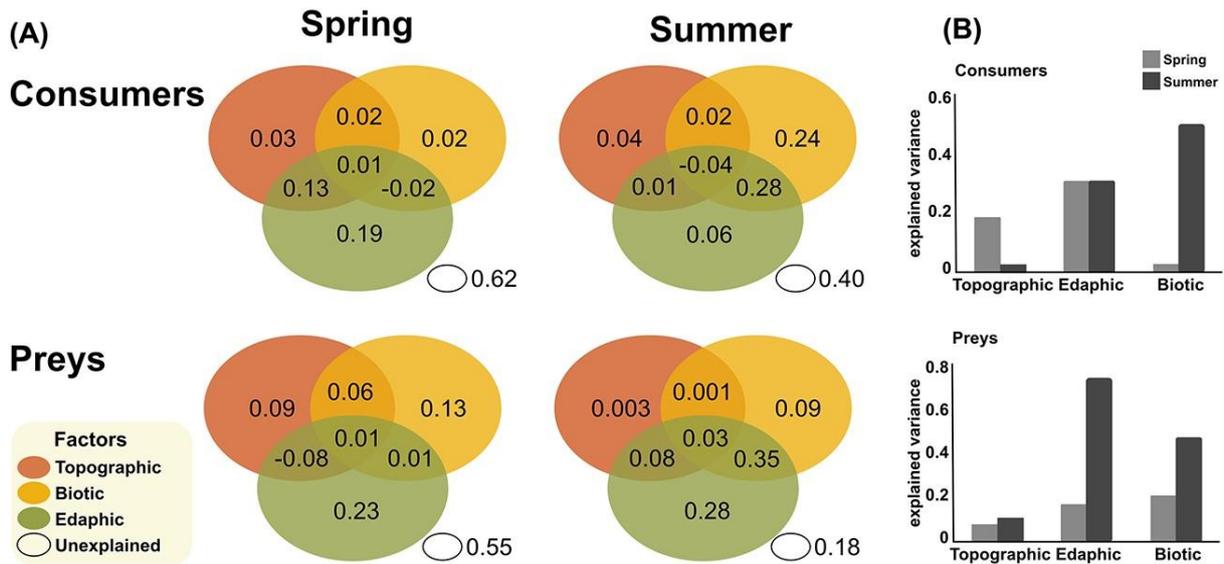
### *3.5 Most influential environmental parameters*

We tested how consumers and preys were responding to different factors, categorized as topographic (altitude, mountain), edaphic (water content, pH, soil temperature, and organic C), and biotic (coverage and diversity of the vascular plants, diversity of bacteria, fungi, protists, and consumers). Models (dbRDA) were estimated during spring snowmelt and from spring to summer (Table S5). The models retained pH and prey diversity as the main drivers for consumers in summer; remarkably, the

## Chapter 4



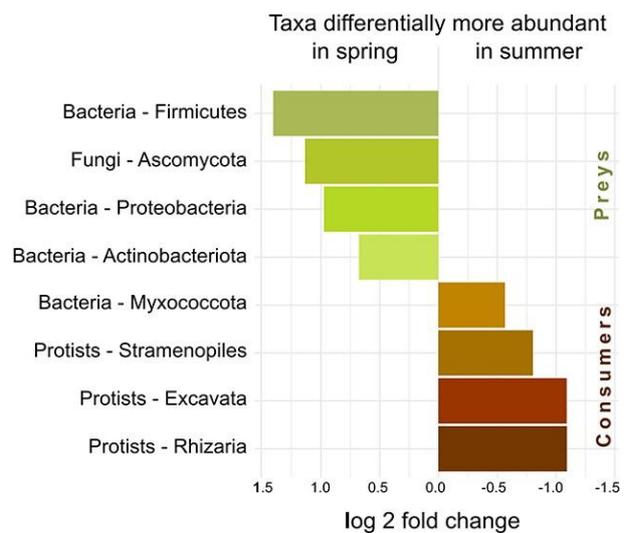
**Figure 3:** (A, B) seasonal variation in the standardized SSU rRNA sequences, diversity (Shannon index) and evenness, from the sample under the snow in spring to summer. Significant changes (analysis of variance,  $P$ -value  $\leq 0.05$ ) are indicated by “a” (higher) and “b” (lower). Standard errors bars are shown. (A) Consumers (predatory bacteria, heterotrophic and free-living protists, selected nematodes). (B) Preys (non-predatory bacteria, fungi and autotrophic protists). Consumers increase from spring to summer (without significant changes in diversity or evenness); preys decrease, while their diversity and evenness increase. (C, D) linear correlations between consumers and preys (y and x axis, respectively), in spring and summer. (C) Shannon indices. (D) Shannon index of consumers versus evenness of preys. Dots = samples. Grey surface = 95% confidence interval.



**Figure 4:** Relative influence of topographic, biotic and edaphic factors on the diversity of consumers and preys, by season. (A) The proportion of variance explained by topographic (upper left ellipse), biotic (upper right) and edaphic (bottom) factors, and the unexplained variance. (B) Comparison of explained total variance by group of factors (the sum of each ellipse), in spring and summer.

prey diversity was more important than pH in summer (Table S5). Preys were also more influenced by the diversity of consumers under the snow than by topographic and edaphic factors. Bacteria were influenced by altitude and edaphic parameters in different combinations, depending on the sample; biotic factors had little influence. Less variation was explained in models estimated for protists, with low F values and inconsistent results between samples. Fungi did not respond to topographic nor edaphic factors, but sporadically to bacterial and protistan diversity. The mountains had no effect in this analysis (Table S5).

For a reliable interpretation of our results, it was important to test the influence of our sampling design, i.e. the four samples per site (Fig. 1D), the altitudinal gradient (Fig. 1E), and the three mountains (Fig. 1B) on the microbial communities. The first axis of principal component analysis only explained 31.3% of the variation of bacteria, with a slight effect of altitude. As in the dbRDA models, the three mountains had no influence, and the fungal, protistan, and consumer communities displayed no clear clustering trend with respect to the parameters tested (Fig. S2).



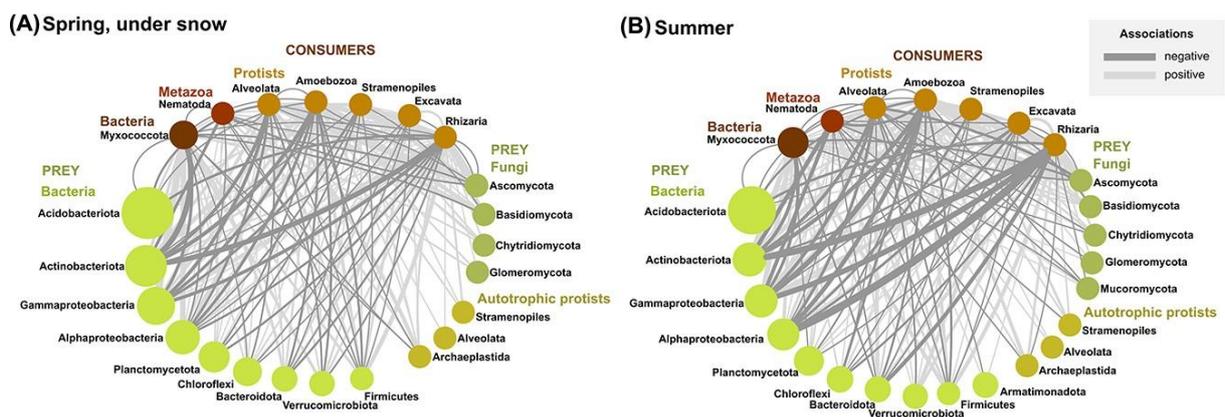
**Figure 5:** Differential abundances. Taxa were filtered according to abundance and presence in one-third of the samples. All taxa differentially more abundant in spring and summer are preys and consumers, respectively.

### 3.6 Biotic interactions

Variation partitioning, conducted to disentangle the relative influence of topographic, edaphic, and biotic factors, indicated a clear seasonal trend—all factors together explained more variance in summer than in spring, for both consumers and preys (Fig. 4A). A seasonal increase in the relative influence of biotic factors and biotic + edaphic factors on community variance was observed, higher for consumers (c. 17 times) than for preys (c. 2 times; Fig. 4B).

Differential expression analysis revealed which taxa were significantly more abundant in spring or in summer. Strikingly, all groups selected by this analysis as more abundant in spring were preys (Bacteria: Firmicutes, Actinobacteria and Proteobacteria, Fungi: Ascomycota), whereas those more abundant in summer were consumers (Bacteria: Myxococcota, protists: Rhizaria, Stramenopiles, and Excavata; Fig. 5).

The co-occurrence network confirmed the dominant role of biotic interactions between consumers and preys in summer, showing more associations—and more negative ones—in summer than in spring (Fig. 6A and B, Table S6). In summer, Rhizaria stood out among the consumers with strong negative associations with several major bacterial phyla, i.e. Actinobacteria and Alpha- and Gammaproteobacteria. No seasonal differences were found in the networks of associations between preys, which reflect competitive or facilitative interactions or the sharing of ecological niches (Fig. S6). However, there were no noticeable differences in the structure of the networks, which were only affected by shifts in the relative proportions of the associations.



**Figure 6:** Co-occurrence networks of abundant phyla of consumers and preys. (A) Spring, under the snow. (B) Summer. The size of the nodes (dots) is proportional to the number of sequences. Edges (connecting lines) represent positive (light grey) or negative (dark grey) correlations, with line width proportional to the number of correlations. Self-loops, taxa with a single edge and connections between preys are not shown.

### 4 Discussion

#### *4.1 How biotic interactions shape the soil communities*

The opposite and correlated changes in SSU rRNA relative abundance, diversity and evenness of consumers and preys from spring to summer strongly suggest an effect of predation: the increase of grazers' abundance reduces the prevalence of preys, whilst increasing their diversity and evenness—predators shape the preys' communities (Figs 3, 5 & 6). It has been repeatedly demonstrated, particularly in aquatic environments, that predation prevents competitive exclusion—i.e. the dominance of few better adapted species profiting from the resources of a given habitat [33-35]. Thus, as observed here, the increase in consumers' abundance leads to an increase in preys' evenness and diversity, while without predators, competition for resources may result in the domination of fewer species [36]. The significance for ecosystem functioning is unmistakable: highly uneven communities, with an extreme dominance by few species, are less resistant to environmental stress [37]. Thus, both the number and relative abundances of species must be sustained to achieve a vigorous ecosystem functioning [38]. In this study, we show that in alpine grasslands with a constant snow cover in winter, the increase in consumers from spring to summer effectively contributes in maintaining a diverse bacterial and fungal community.

The communities of consumers and preys were organized in highly interconnected networks, with stronger negative interactions occurring in summer (Fig. 6). This, in addition to the previous results, likely reflects predator–prey interactions. Among protistan consumers, Cercozoa and Amoebozoa display the strongest negative associations with the major bacterial phyla. Supporting a consumer effect—but not proving it—the dominant cercozoan taxon is bacterivorous (Glissomonadida, 43% of all cercozoan sequences; Table S2). Negative associations between Cercozoa and Actinobacteria (important polysaccharide decomposers) [39] and Alphaproteobacteria were also observed in the rhizosphere [40, 41] where the glissomonads were most abundant (compared to bulk soil and litter) [42]. A link between the decrease of Actinobacteria and the grazing of heterotrophic protists was also suggested [43]. However, negative associations may be solely abiotic, e.g. due to opposite sensitivity to environmental conditions, or biotic but driven by prey defences, such as Actinobacteria secreting secondary metabolites to evade predation [44]. In addition to predation, competitive interactions have been shown to be a major driver in bacterial community composition [45]. For instance, fungal–bacterial competition explained 32% of the variance within planktonic bacterial communities [6]. Accordingly, we observed intricate networks between preys, not or only slightly affected by season (Fig. S6). As our aim was to observe if and how biotic interactions changed between spring and summer, we performed our analyses using a software [30] that attempts to remove associations driven

## Chapter 4

---

by environmental data, e.g. shared niches. Accordingly, the networks calculated with (Note S1) and without (Fig. 6) environmental data were quite similar.

Our findings are in line with multiple studies which demonstrate that bacterial communities are consistently driven by edaphic or topographic parameters (pH, altitude, water content, and organic carbon), whereas eukaryotic communities display weaker trends in response to environmental gradients in comparison to bacteria (Fig. S2 and Table S5) [4, 46]. Subsequently, while it is generally agreed that bacterial communities are predominantly driven by pH, there is no consensus as to a single main driver for fungal and protistan communities [5, 46, 47]. Indeed, distinct taxa or functional groups of fungi and protists have different optima along environmental gradients and therefore are differentially affected by seasonal and/or altitudinal changes [4, 5, 47, 48]. Similarly, protistan trophic guilds, e.g. consumers, parasites, and phototrophs, differentially respond to altitude and edaphic factors [49]; it is noteworthy that only 40% of the species turnover in these communities could be explained by abiotic factors [47].

It logically follows that the drivers of the fungal and protistan biogeographies must be sought elsewhere. Our models indicated that the most influential response was between functional guilds—in particular, consumers displayed a strong and consistent response to prey diversity, and more so in summer (Table S5). Additionally, variation partitioning (Fig. 4) demonstrated that the combined influence of biotic and edaphic factors explained the largest share of variance in consumer and prey communities. Thus, the intricate interplay between the environment and competitive and/or predatory interactions is best observed when functional guilds are taken into account.

### *4.2 Changes during snowmelt*

Our findings reveal that alpine soil microbial communities, including protists, undergo gradual changes from spring to summer, without any sudden shift during snowmelt (Fig. S5), despite the drastic environmental changes occurring during thaw [24]. The soil microbial biomass in alpine grassland typically attains its annual peak during winter, just before snowmelt [50]. In soils frozen during winter, the soil microbial biomass suddenly declined at snowmelt [46, 47], and cold-adapted winter soil microbial communities died and were swiftly replaced by summer ones [48, 49]. Since we did not observe a sudden collapse of the microbial biomass at snowmelt [24], we consider that under a previous winter snow cover with stable soil temperatures  $>0^{\circ}\text{C}$  (Fig. S1), a specific winter-adapted microbial community does not develop nor dies at thaw. The dynamics and composition of the alpine microbiome is thus dependent on climatic conditions. Current global warming is already reducing the alpine winter snow cover [51], resulting in colder, often frozen soils. This may challenge the stability we observed and result in spring shifts of the microbial communities and their functions [52].

### *4.3 Taxonomic and functional diversity*

Our study challenges two commonly held assumptions concerning the diversity and the functional groups of the soil microbiome. The biases induced by “universal eukaryotic primers” during PCR overestimate the “SAR” (or Harosa) clade, particularly the ciliates, while strongly underestimating Amoebozoa (especially Conosa and the slime moulds), as previously signalled [53, 54]. RNA-based studies unequivocally agree on revealing Amoebozoa as an important (when not dominant) protistan lineage, alongside Rhizaria, in soil and litter [55-57]. In our study, amoebozoans not only dominate the protistan assemblage, but also played a major role as consumers (Fig. 6)—it follows that neglecting them will result in an incomplete view of the soil food web. Our results indubitably shows that the bacterial predators (Myxococcota essentially) outnumber the protistan predators. They play an essential, and mostly unrecognized, role in shaping the microbial communities (Fig. 6), as already noticed [58].

The low proportion of protistan parasites (9%) is consistent with a previous study in the Swiss Alps, where their relative abundance decreased at the altitudes at which our survey was conducted [47]. Consumers were also the most abundant in temperate regions [59]. This is in contrast to tropical soils where parasites dominated [60, 61], probably related to the high abundance and diversity of insects [60].

### *4.4 Methodological discussion*

It increasingly appears that metatranscriptomics might become the preferred method for molecular monitoring of complex environmental microbiomes. Using mock communities, it has been shown that rRNA-based methods outperform metagenomics in identifying taxa [62]. DNA-based surveys are inappropriate for monitoring short- to middle-term shifts, since they include a fair portion of dead organisms [63]. Possible biases are discussed in Note S1.

## 5 Concluding remarks

The soil microbial food web in alpine environments shows striking seasonal dynamics, with biotic interactions explaining a higher proportion in variability of consumer turnover than soil properties (e.g. pH and carbon content) or topography (elevation, spatial difference between mountains). Our study stands out by applying metatranscriptomics to a large-scale ecological appraisal of the entire soil microbiota. We achieved a sequencing depth that surpasses the descriptive limits of classical amplicon-based approaches enabling inter-domain and inter-sample comparisons. We complement our protocol with trait-based approaches to enhance basic knowledge of the soil food web functioning. Thus, our

## Chapter 4

---

study contributes to the understanding of the alpine ecosystem by showing the importance of the biotic interactions in shaping the seasonal dynamics of the soil microbiome.

### Author contributions

AMFD and CR collected samples. CR conducted the vegetation survey. AMFD and TU performed molecular biological analyses. AMFD and MBD conducted bioinformatics. AMFD, JF, MB and TU and performed statistics. AMFD wrote the manuscript. All authors contributed to the interpretation of the data and the writing/editing of the manuscript. AMFD devised and coordinated the project.

### Data availability

Sequencing data and raw sequences are available under NCBI BioProject PRJNA850398.

### Competing interests

The authors declare that they do not have any competing interests in relation to the work described.

### Funding

This work was funded by the German Research Foundation (DFG), grant number FI 1686/2-1.

### Acknowledgments

For help in collecting samples and doing the vegetation survey in Davos, we are grateful to Sven Buchmann, Fiona Schwaller, Sabrina Keller, and Francesca Jaroszynska. We thank Jana Täumer and Verena Gross at the laboratory at the University of Greifswald, for helping with RNA extraction. At the University of Cologne, we are grateful to Irene Brockhaus, Anna Herzog, Andrea Meis, and Oscar Rindt for the measurement of edaphic properties, under the supervision of Christoph Rosinger. We are indebted to Caterina Penone for conducting the variance partitioning analysis. We thank Alexandra Elbakyan for facilitating access to scientific articles.

### References

1. Cheng L, Zhang N, Yuan M *et al.* Warming enhances old organic carbon decomposition through altering functional microbial communities. *ISME J.* 2017;**11**:1825–35. <https://doi.org/10.1038/ismej.2017.48>.
2. Zhu B, Li C, Wang J *et al.* Elevation rather than season determines the assembly and co-occurrence patterns of soil bacterial communities in forest ecosystems of mount Gongga. *Appl Microbiol Biotechnol* 2020;**104**:7589–602. <https://doi.org/10.1007/s00253-020-10783-w>.
3. Shen C, He J-Z, Ge Y. Seasonal dynamics of soil microbial diversity and functions along elevations across the treeline. *Sci Total Environ* 2021;**794**:148644. <https://doi.org/10.1016/j.scitotenv.2021.148644>.
4. Adamczyk M, Hagedorn F, Wipf S *et al.* The soil microbiome of GLORIA mountain summits in the Swiss Alps. *Front Microbiol* 2019;**10**:1080. <https://doi.org/10.3389/fmicb.2019.01080>.
5. Shen C, Gunina A, Luo Y *et al.* Contrasting patterns and drivers of soil bacterial and fungal diversity across a mountain gradient. *Environ Microbiol Rep* 2020;**22**:3287–301. <https://doi.org/10.1111/1462-2920.15090>.
6. Li G, Wang Y, Lib H *et al.* Quantifying relative contributions of biotic interactions to bacterial diversity and community assembly by using community characteristics of microbial eukaryotes. *Ecol Indic* 2023;**146**:109841. <https://doi.org/10.1016/j.ecolind.2022.109841>.
7. Lee CK, Laughlin DC, Bottos EM *et al.* Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. *Comm Biol* 2019;**2**:62. <https://doi.org/10.1038/s42003-018-0274-5>.
8. García-Girón J, Heino J, García-Criado F *et al.* Biotic interactions hold the key to understanding metacommunity organisation. *Ecography* 2020;**43**:1180–90. <https://doi.org/10.1111/ecog.05032>.
9. Chesson P, Kuang JJ. The interaction between predation and competition. *Nature* 2008;**456**:235–8. <https://doi.org/10.1038/nature07248>.
10. Dormann CF, Bobrowski M, Dehling DM *et al.* Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Glob Ecol Biogeogr* 2018;**27**: 1004–16. <https://doi.org/10.1111/geb.12759>.
11. de Vries FT, Thébaud E, Liiri M *et al.* Soil food web properties explain ecosystem services across European land use systems. *Proc Natl Acad Sci USA* 2014;**110**:14296–301. <https://doi.org/10.1073/pnas.1305198110>.
12. Sokol NW, Slessarev E, Marschmann GL *et al.* Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* 2022;**20**:415–30. <https://doi.org/10.1038/s41579-022-00695-z>.
13. Ludwig M, Achtenhagen J, Miltner A *et al.* Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. *Soil Biol Biochem* 2015;**81**:311–22. <https://doi.org/10.1016/j.soilbio.2014.12.002>.
14. Camenzind T, Mason-Jones K, Mansour I *et al.* Formation of necromass-derived soil organic carbon determined by microbial death pathways. *Nat Geosci* 2023;**16**:115–22. <https://doi.org/10.1038/s41561-022-01100-3>.
15. *Federal Office for Meteorology and Climatology.* [www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail.html](http://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail.html) Last accessed: March 2023.
16. Gruber-Vodicka HR, Seah BK, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. *mSystems* 2020;**5**:e00920. <https://doi.org/10.1128/mSystems.00920-20>.
17. Bushnell, B. BBmap. <https://sourceforge.net/projects/bbmap/>. <https://sourceforge.net/projects/bbmap/> Last accessed: August 2023.
18. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;**27**: 2957–63. <https://doi.org/10.1093/bioinformatics/btr507>.
19. Guillou L, Bachar D, Audic S *et al.* The protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 2013;**41**:D597–604. <https://doi.org/10.1093/nar/gks1160>.
20. Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinf* 2008;**10**:421. <https://doi.org/10.1186/1471-2105-10-421>.
21. Adl SM, Simpson AGB, Lane CE *et al.* The revised classification of eukaryotes. *J Euk Microbiol* 2012;**59**:429–514. <https://doi.org/10.1111/j.1550-7408.2012.00644.x>.
22. Yeates G, Bongers T, De Goede R *et al.* Feeding habits in soil nematode families and genera - an outline for soil ecologists. *J Nematol* 1993;**25**:315–31.
23. Ferris, H. *Nemaplex.* <http://nemaplex.ucdavis.edu/IndexFiles/etol.html> Last accessed: August 2023.

## Chapter 4

---

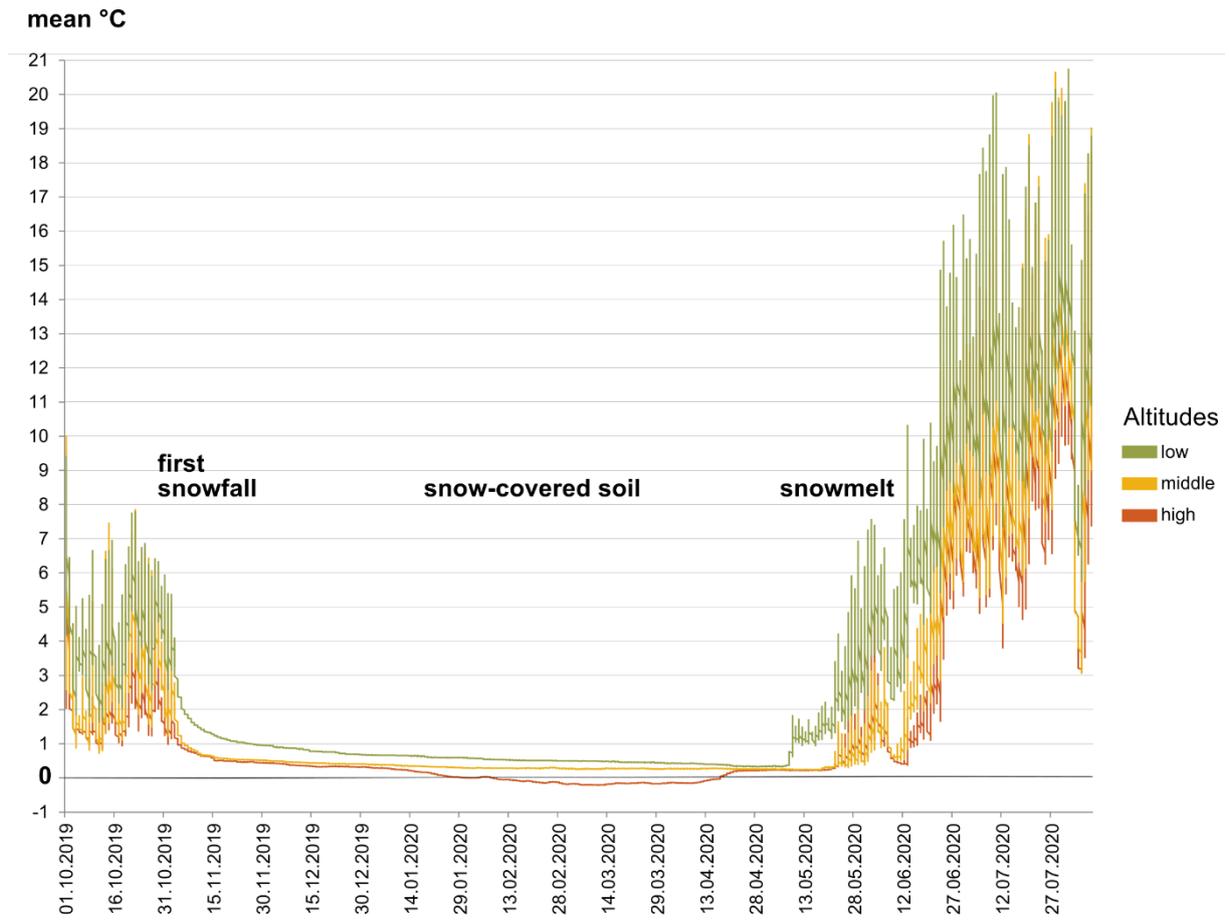
24. Rindt O, Rosinger C, Bonkowski M *et al.* Biogeochemical dynamics during snowmelt and in summer in the Alps. *Biogeochemistry* 2023;**162**:257–66. <https://doi.org/10.1007/s10533-02201005-8>.
25. Juillerat, P., Bäumler, B., Bornand, C., *et al.* *Flora Helvetica checklist 2017: der Gefäßpflanzen der Schweiz*. 385 pp. (InfoFlora, Geneva, 2017). <https://doi.org/10.5167/uzh-165877>.
26. R Development Core Team. R: *A Language and Environment for Statistical Computing*. <http://www.R-project.org/>. Last accessed: Access. Date.
27. Oksanen, J., Blanchet, F. G., Kindt, R., *et al.* Vegan: community ecology package. R package version 2.0-10. <http://CRAN.Rproject.org/package=vegan>. (2013).
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. <https://doi.org/10.1186/s13059-014-0550-8>.
29. Faust K. Open challenges for microbial network construction and analysis. *ISME J.* 2021;**15**:3111–8. <https://doi.org/10.1038/s41396-021-01027-4>.
30. Tackmann J, Rodrigues JFM, von Mering C. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Systems* 2019;**9**: 286–296.e8. <https://doi.org/10.1016/j.cels.2019.08.002>.
31. Bezanson J, Karpinski S, Shah VB *et al.* Julia: a fast dynamic language for technical computing. *arXiv* 2012;**1209**:5145. <https://doi.org/10.48550/arXiv.1209.5145>.
32. Shannon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504. <https://doi.org/10.1101/gr.1239303>.
33. Gauzens B, Legendre S, Lazzaro X *et al.* Intermediate predation pressure leads to maximal complexity in food webs. *Oikos* 2015;**125**:595–603. <https://doi.org/10.1111/oik.02627>.
34. Saleem M, Fetzer I, Dormann CF *et al.* Predator richness increases the effect of prey diversity on prey yield. *Nat Commun* 2012;**3**:1305. <https://doi.org/10.1038/ncomms2287>.
35. Burian A, Pinn D, Peralta-Maraver I *et al.* Predation increases multiple components of microbial diversity in activated sludge communities. *ISME J* 2022;**16**:1086–94. <https://doi.org/10.1038/s41396-021-01145-z>.
36. Corno G, Caravati E, Callieri C *et al.* Effects of predation pressure on bacterial abundance, diversity, and size-structure distribution in an oligotrophic system. *J Limnol* 2008;**67**:107–19. <https://doi.org/10.4081/jlimnol.2008.107>.
37. Wittebolle L, Marzorati M, Clement L *et al.* Initial community evenness favours functionality under selective stress. *Nature* 2009;**458**:623–6. <https://doi.org/10.1038/nature07840>.
38. Crowder DW, Northfield TD, Strand MR *et al.* Organic agriculture promotes evenness and natural pest control. *Nature* 2010;**466**: 109–12. <https://doi.org/10.1038/nature09183>.
39. Tveit A, Urich T, Svenning MM. Metatranscriptomic analysis of Arctic peat soil microbiota. *Appl Environ Microbiol* 2014;**80**: 5761–72. <https://doi.org/10.1128/AEM.01030-14>.
40. Rossmann M, Pérez-Jaramillo J, Kavamura V *et al.* Multitrophic interactions in the rhizosphere microbiome of wheat: from bacteria and fungi to protists. *FEMS Microbiol Ecol* 2020;**96**:fiae032. <https://doi.org/10.1093/femsec/fiae032>.
41. Rürger L, Feng K, Dumack K *et al.* Assembly patterns of the rhizosphere microbiome along the longitudinal root axis of maize (*Zea mays* L.). *Front Microbiol* 2021;**12**:614501. <https://doi.org/10.3389/fmicb.2021.614501>.
42. Fiore-Donno AM, Human ZR, Štursová M *et al.* Soil compartments (bulk soil, litter, root and rhizosphere) as main drivers of soil protistan communities distribution in forests with different nitrogen deposition. *Soil Biol Biochem* 2022;**168**:108628. <https://doi.org/10.1016/j.soilbio.2022.108628>.
43. Schostag M, Priemé A, Jacquiod S *et al.* Bacterial and protozoan dynamics upon thawing and freezing of an active layer permafrost soil. *ISME J.* 2019;**13**:1345–59. <https://doi.org/10.1038/s41396-019-0351-x>.
44. Eckert EM, Baumgartner M, Huber IM *et al.* Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environ Microbiol Rep* 2013;**15**:2019–30. <https://doi.org/10.1111/1462-2920.12083>.
45. Romdhane S, Spor A, Aubert J *et al.* Unraveling negative biotic interactions determining soil microbial community assembly and functioning. *ISME J.* 2022;**16**:296–306. <https://doi.org/10.1038/s41396-021-01076-9>.
46. Li J, Li C, Kou Y *et al.* Distinct mechanisms shape soil bacterial and fungal co-occurrence networks in a mountain ecosystem. *FEMS Microbiol Ecol* 2020;**96**:fiae030. <https://doi.org/10.1093/femsec/fiae030>.
47. Mazel F, Malard L, Niculita-Hirzel H *et al.* Soil protist function varies with elevation in the Swiss Alps. *Environ Microbiol Rep* 2021;**24**:1689–702. <https://doi.org/10.1111/1462-2920.15686>.

## Chapter 4

---

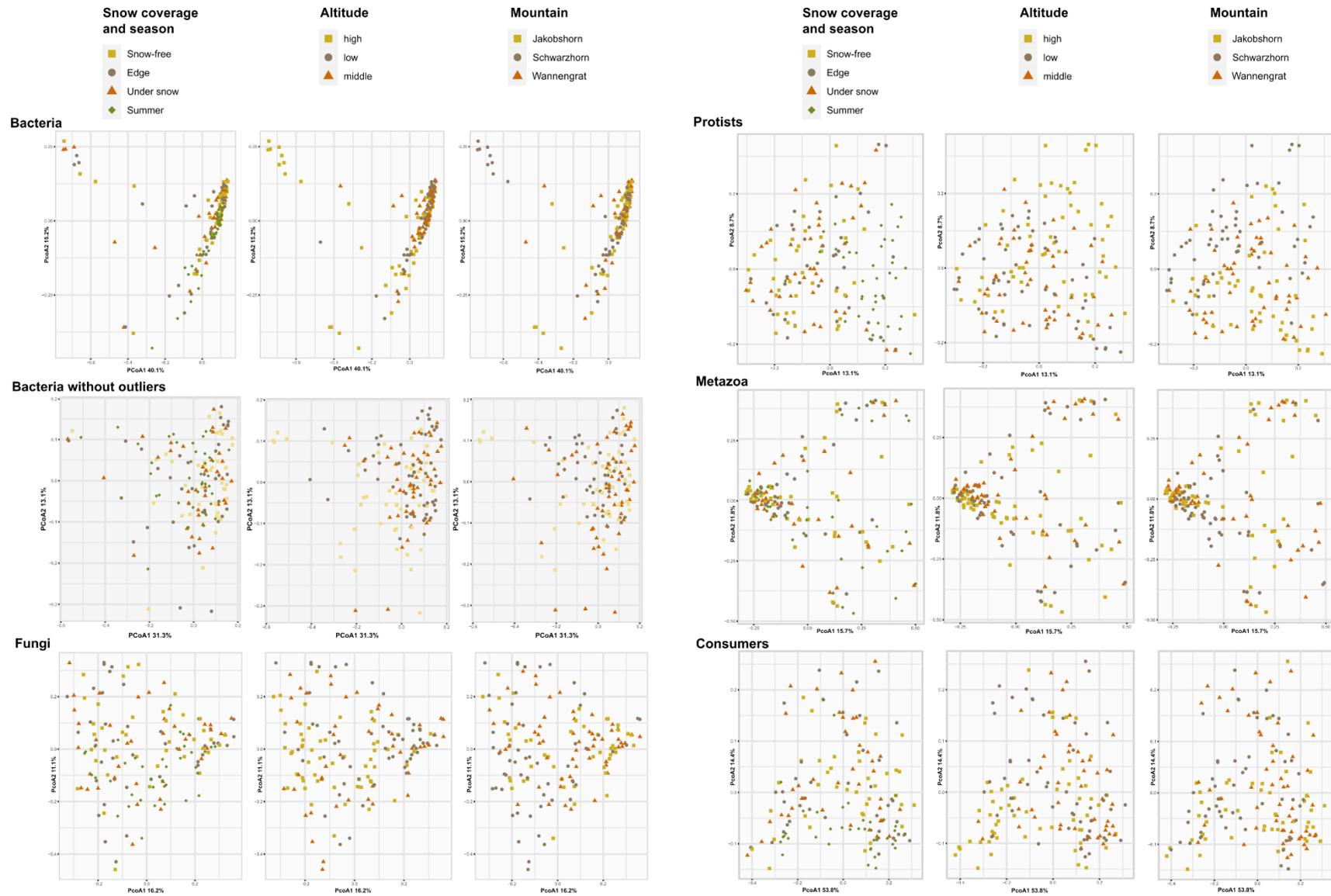
48. Yashiro E, Pinto-Figueroa E, Buri A *et al.* Local environmental factors drive divergent grassland soil bacterial communities in the western Swiss Alps. *Appl Environ Microbiol* 2016;**82**:6303–16. <https://doi.org/10.1128/AEM.01170-16>.
49. Nguyen B-AT, Chen Q-L, Yan Z-Z *et al.* Distinct factors drive the diversity and composition of protistan consumers and phototrophs in natural soil ecosystems. *Soil Biol Biochem* 2021;**160**:108317. <https://doi.org/10.1016/j.soilbio.2021.108317>.
50. Schmidt SK, Lipson DA. Microbial growth under the snow: implications for nutrient and allelochemical availability in temperate soils. *Plant Soil* 2004;**259**:1–7. <https://doi.org/10.1023/B:PLSO.0000020933.32473.7e>.
51. Rumpf SB, Gravey M, Brönnimann O *et al.* From white to green: snow cover loss and increased vegetation productivity in the European Alps. *Science* 2022;**376**:1119–22. <https://doi.org/10.1126/science.abn6697>.
52. Broadbent AA, Snell HS, Michas A *et al.* Climate change alters temporal dynamics of alpine soil microbial functioning and biogeochemical cycling via earlier snowmelt. *ISME J.* 2021;**15**: 2264–75. <https://doi.org/10.1038/s41396-021-00922-0>.
53. Geisen S, Tveit A, Clark IM *et al.* Metatranscriptomic census of active protists in soils. *ISME J.* 2015;**9**:2178–90. <https://doi.org/10.1038/ismej.2015.30>.
54. Fiore-Donno AM, Weinert J, Wubet T *et al.* Metacommunity analysis of amoeboid protists in grassland soils. *Sci Rep* 2016;**6**:19068. <https://doi.org/10.1038/srep19068>.
55. Urich T, Lanzén A, Qi J *et al.* Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 2008;**3**:e2527. <https://doi.org/10.1371/journal.pone.0002527>.
56. Voss C, Fiore-Donno AM, Guerreiro MA *et al.* Metatranscriptomics reveal unsuspected protistan diversity in leaf litter across temperate beech forests, with Amoebozoa the dominating lineage. *FEMS Microbiol Ecol* 2019;**95**:fiz142. <https://doi.org/10.1093/femsec/fiz142>.
57. Harkes P, Suleiman AKA, van den Elsen SJJ *et al.* Conventional and organic soil management as divergent drivers of resident and active fractions of major soil food web constituents. *Sci Rep* 2019;**9**:13521. <https://doi.org/10.1038/s41598-019-49854-y>.
58. Petters S, Groß V, Söllinger A *et al.* The soil microbial food web revisited: predatory myxobacteria as keystone taxa? *ISME J* 2021;**15**:2665–75. <https://doi.org/10.1038/s41396-021-00958-2>.
59. Seppéy CVW, Broennimann O, Buri A *et al.* Soil protist diversity in the Swiss western Alps is better predicted by topo-climatic than by edaphic variables. *JBiogeogr* 2020;**47**:866–78. <https://doi.org/10.1111/jbi.13755>.
60. Mahé F, De Vargas C, Bass D *et al.* Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol* 2017;**1**:0091. <https://doi.org/10.1038/s41559-017-0091>.
61. Schulz G, Schneider D, Brinkmann N *et al.* Changes in trophic groups of protists with conversion of rainforest into rubber and oil palm plantations. *Front Microbiol* 2019;**10**:240. <https://doi.org/10.3389/fmicb.2019.00240>.
62. Hempel CA, Wright N, Harvie J *et al.* Metagenomics versus total RNA sequencing: most accurate data-processing tools, microbial identification accuracy and perspectives for ecological assessments. *Nucleic Acids Res* 2022;**50**:9279–93. <https://doi.org/10.1093/nar/gkac689>.
63. Carini P, Marsden PJ, Leff J *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 2016;**2**:16242. <https://doi.org/10.1038/nmicrobiol.2016.242>.

Supplementary material



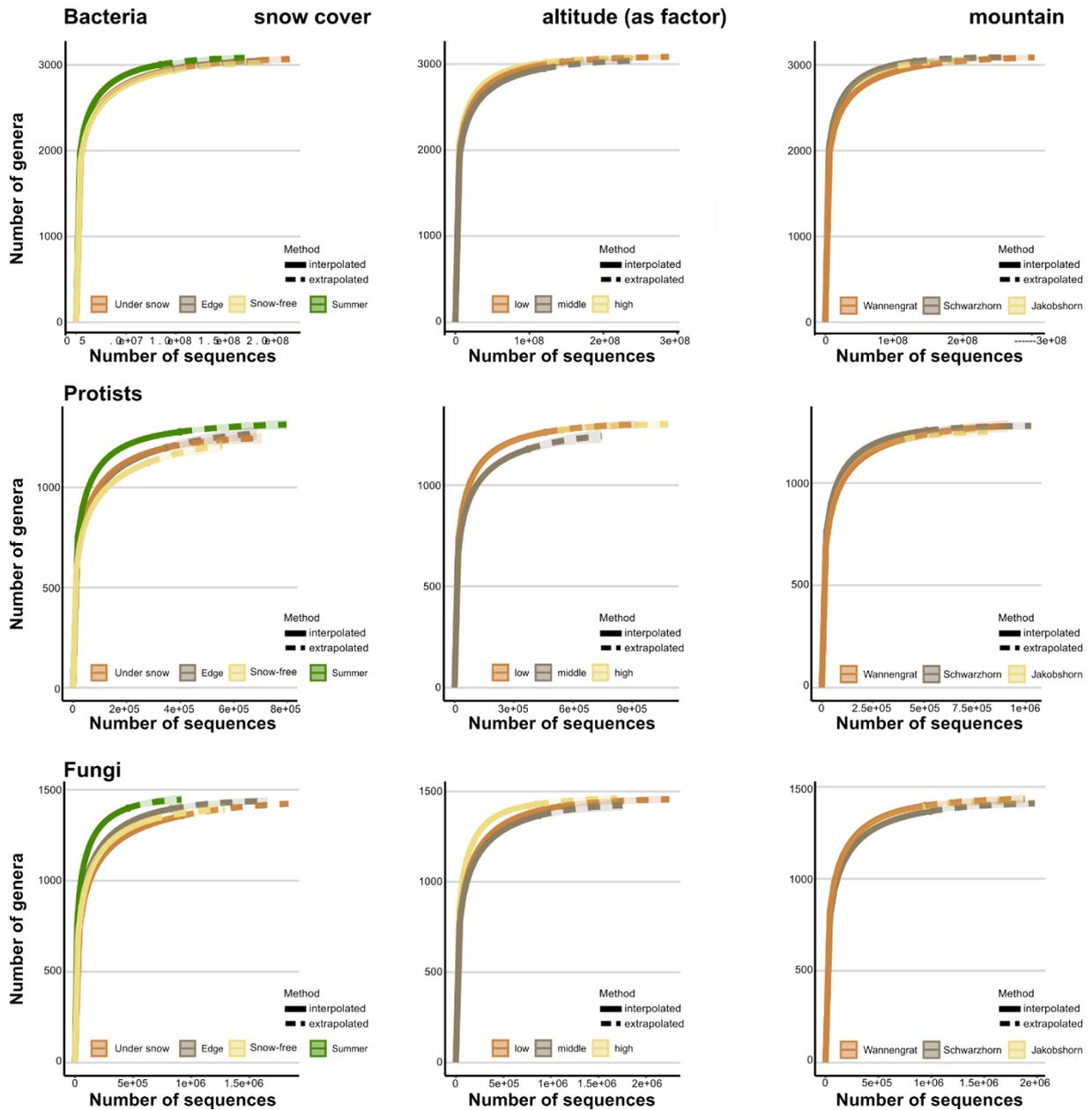
**Figure S1:** Soil temperature recorded every two hours from autumn 2019 to summer 2020, averaged by altitudinal categories. In winter the soil remained at a constant temperature just above zero degrees, except in the highest sites, where temperatures dropped slightly below 0°C but stayed above -1°C.

# Chapter 4



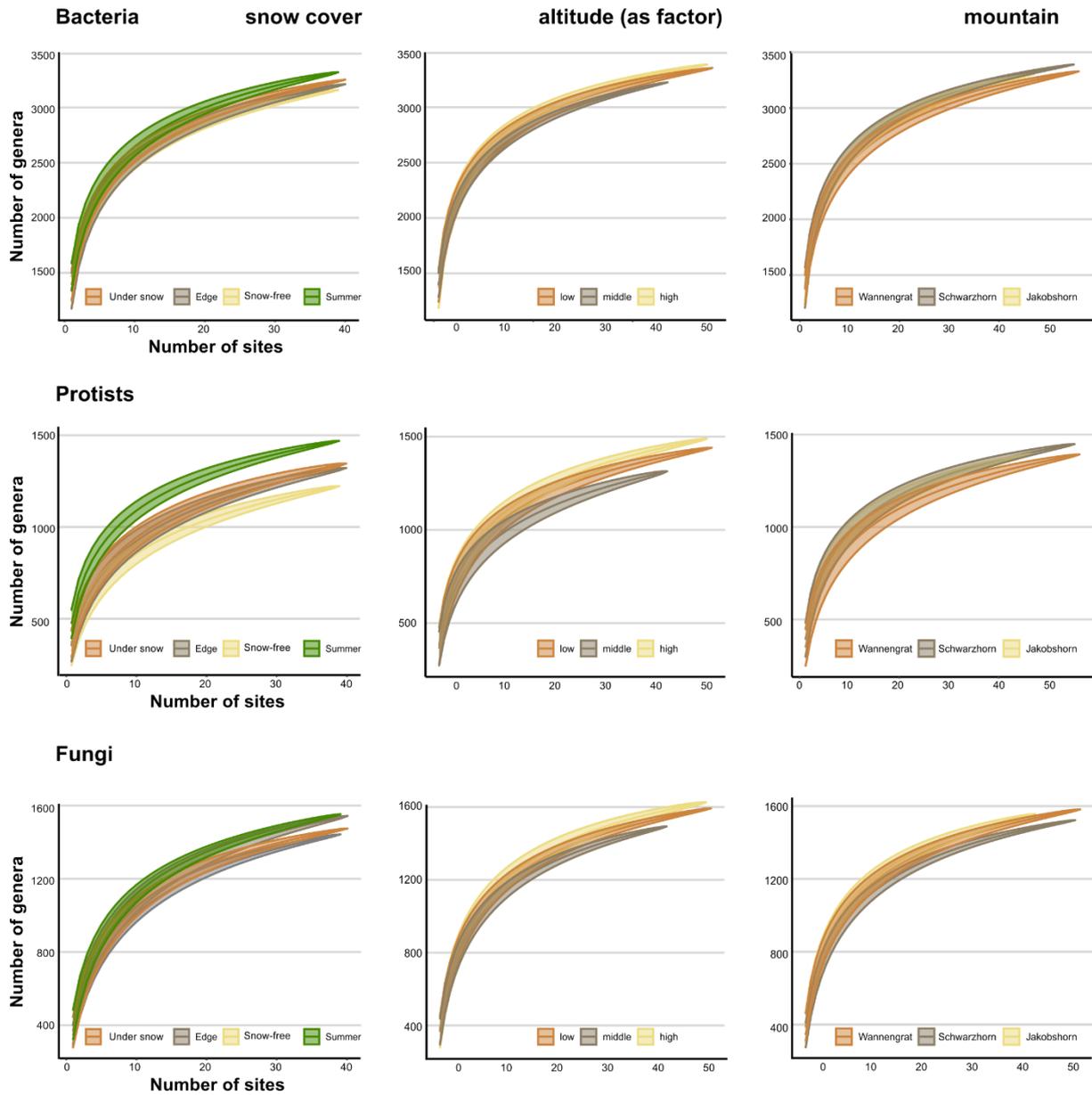
**Figure S2:** Principal Component Analysis of the Bray-Curtis dissimilarity indices of the main taxa and functions, showing that snow coverage, season, altitude and mountain have little influence in shaping the communities. The functional group "preys" was nearly identical to bacteria, and thus not shown.

## Chapter 4



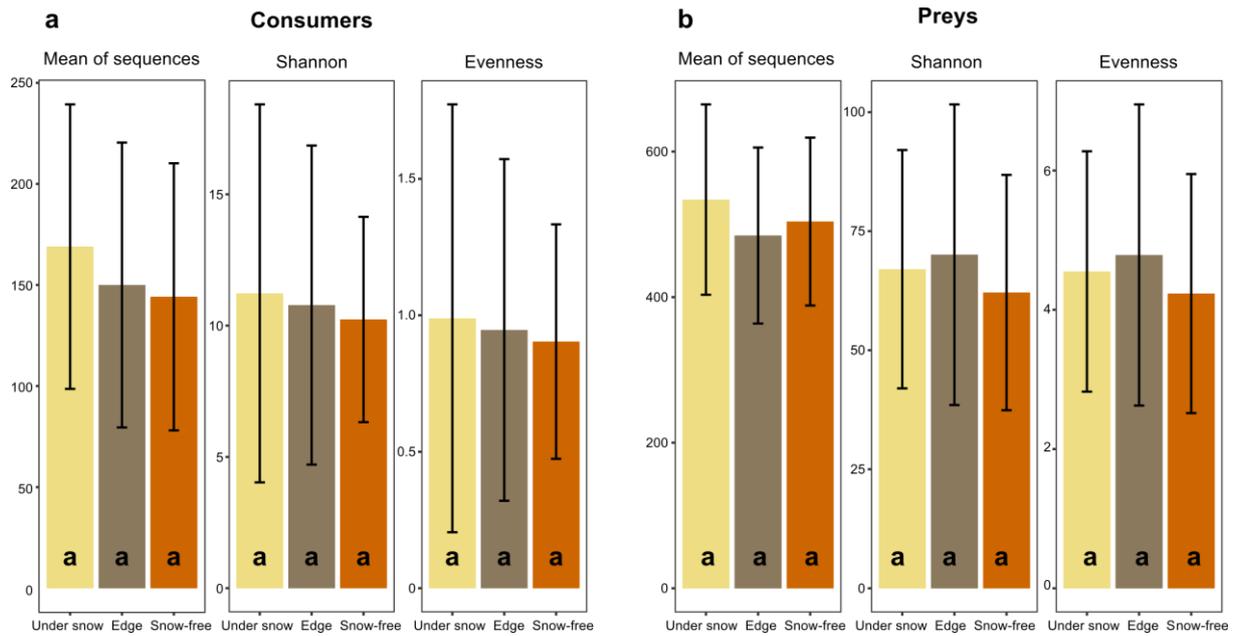
**Figure S3:** Rarefaction curves for bacteria, protists and fungi, by snow cover, altitude and mountain. They were calculated using the iNEXT package, on raw abundances, with a 97% confidence interval, 50 bootstraps and 50 knots.

## Chapter 4

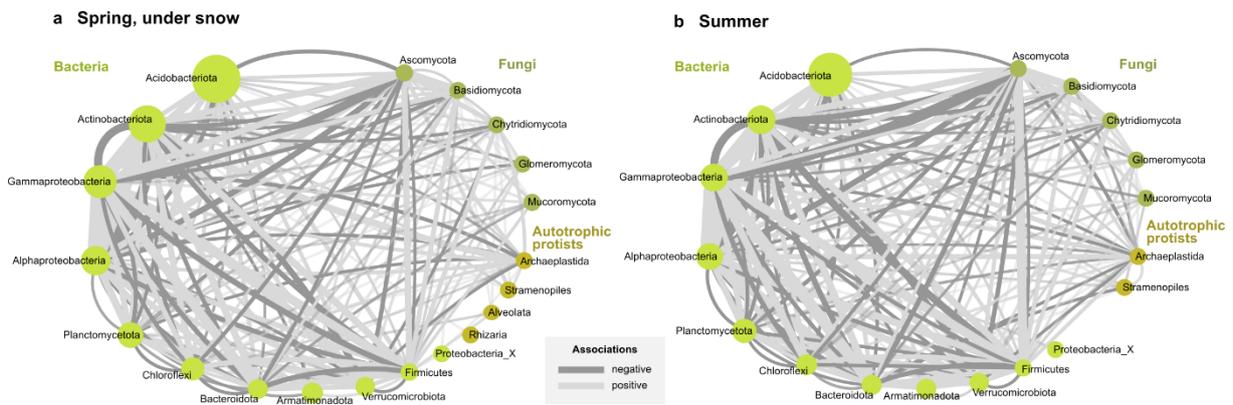


**Figure S4:** Accumulation curves for bacteria, protists and fungi, by snow cover, altitude and mountain. They were calculated using the *specaccum* function, on raw abundances.

## Chapter 4



**Figure S5:** Variation in the mean of SSU sequences, diversity (Shannon index) and evenness, from the samples under the snow, at the edge of the snow patch and snow-free. There were no significant changes (Tukey-test,  $p$ -value 0.05), they are indicated by "a". Standard errors bars are shown. **a**, Consumers (predatory bacteria, heterotrophic and free-living protists, selected nematoda). **b**, Preys (non-predatory bacteria, fungi and autotrophic protists).



**Figure S6:** Co-occurrence networks of abundant phyla of preys. **a**, Spring, under the snow. **b**, Summer. The size of the nodes (dots) are proportional to the number reads. Edges (connecting lines) represent positive (light grey) or negative (dark grey) correlations, with line width proportional to the number of correlations. Self-loops and taxa with a single edge are not shown.

## Chapter 4

**Table S1:** Database of the topographic, edaphic and biotic parameters of the samples.

A separate file has been provided for Tables S1, please refer to the original publication available at <https://doi.org/10.1093/ismeco/ycae028>.

**Table S2:** Database of the abundance of each identified taxon per sample (n=158), with their taxonomic and functional assignment. Last line = total sequences per sample.

A separate file has been provided for Tables S2, please refer to the original publication available at <https://doi.org/10.1093/ismeco/ycae028>.

**Table S3.** Thresholds applied to select the most abundant taxa at different taxonomic levels, and the number of taxa and percentage of sequences deleted.

<b>Taxonomic rank</b>	<b>Threshold</b>	<b>total taxa</b>	<b>Nb taxa deleted</b>	<b>Taxa remaining</b>	<b>% sequences deleted</b>	
Bacteria; Phylum	0.5	85	72	13	2.05	Bacteria_X deleted
Bacteria; Class	0.4	111	87	24	2.88	
Bacteria; Order	0.3	201	153	48	4.08	
Bacteria; Family	0.2	266	206	60	3.97	
Fungi; Class	0.8	12	7	5	1.71	
Fungi; Order	0.5	14	7	7	0.60	
Fungi; Family	0.5	37	23	14	2.05	
Protists; Domain	2	10	4	6	3.20	
Protists; Phylum	1	25	13	12	1.86	
Protists; Class	0.5	97	73	24	2.94	
Protists; Order	0.2	117	71	46	3.07	
Protists; Family	0.1	169	86	83	1.60	
Metazoa; Class	0.5	30	25	5	0.54	
Metazoa; Order	0.5	11	4	7	0.19	
Metazoa; Family	0.5	14	6	8	0.53	
Archaea; Phylum	0.1	8	3	5	0.03	
Archaea; Class	0.1	12	6	6	0.09	
Archaea; Order	0.1	12	3	9	0.10	
Archaea; Family	0.1	13	4	9	0.09	

## Chapter 4

**Table S4:** Sequences obtained from the 158 soil samples and results at each step of their processing.

Raw sequences	Quality filtered	Small-subunit rRNA gene	Chloroplast seq. excluded	Binned to the genus level <sup>1</sup>		"Genera" per sample	Sequences per sample
1.02 x 10 <sup>9</sup>	91 ± 2.6%	394,527,938 (39%)	572,167 (0.14%)	7,694	average = min. = max. = stdev =	2,349 1,363 2,975 295	2,502,341 916,180 3'930'068 635183
<sup>1</sup> Genus level ≠ genus, some taxa could only be identified to higher levels. "Genera" represented by only one to three sequences were excluded							

**Table S5:** Models of the most influential topographic, edaphic and biotic variables, selected by distance-based redundancy and variable selection (functions dbrda and ordistep), for consumers and preys and also bacteria, protists and fungi. Models evaluated by anova, F-values and model p-values are shown.

A separate file has been provided for Tables S5 please refer to the original publication available at <https://doi.org/10.1093/ismeco/ycae028>.

**Table S6:** Network metrics. a Number of associations found in spring, under the snow and in summer. Note the increase in negative associations in summer compared to spring. b Network characteristics, note that there are more connections in summer than in spring, i.e. more nodes and edges and especially the doubling of connected components.

<b>a</b>	<b>Spring, under snow</b>				<b>Summer</b>			
	Total	Negative	Positive	% of negative	Total	Negative	Positive	% of negative
Total	2820	580	2240	21	2972	668	2304	22
Prey	2213	429	1784	19	2255	464	1791	21
Consumers	178	8	170	4	234	10	224	4
Between consumers & prey	429	143	286	<b>33</b>	483	194	289	<b>40</b>

<b>b</b>	<b>Spring</b>	<b>Summer</b>
	Number of Nodes	1892
Number of Edges	2820	2972
Avg. Number of neighbors	3	3
Network diameter	18	22
Network radius	11	12
Characteristic path length	7.74	7.76
Clustering coefficient	0.002	0.002
Network density	0.002	0.002
Network heterogeneity	0.472	0.482
Network centralization	0.004	0.003
Connected components	<b>8</b>	<b>16</b>

## Chapter 4

---

**Supplementary Note 1.** Additional information on taxonomic diversity and methodological discussion.

### **Taxonomic diversity**

Prokaryotes totalized 98.02% of the genera, with Bacteria and Archaea accounting for 99.94% and 0.06% of their sequences, respectively (Supplementary Table 2). Bacteria were dominated by three phyla, Acidobacteriota (26%), Proteobacteria (23%), and Actinobacteriota (16%). The Proteobacteria were mostly Gammaproteobacteria (~55%) and Alphaproteobacteria (~45%). The predatory Myxococcota represented 5.72% (Fig. 2A). In Archaea the phylum Crenarchaeota (95%) and the class Nitrososphaeria (94%) dominated (Supplementary Table 2).

Among eukaryotes, Fungi and Metazoa accounted for 36% of the sequences each, protists for 17%, and multicellular plants for 11% (Supplementary Table 2). More than one third of the protistan sequences were assigned to the phylum Amoebozoa (35%) (Fig. 2B). Among it, most of the sequences were identified as Conosa (19%), then Discosea (11%, mostly represented here by *Acanthamoeba*) and Tubulinea (5%). Conosa mainly was composed of Variosea and the slime-moulds Myxogastria (83% and 16% of the conosan sequences, respectively). In Tubulinea, the amoebae with shell (Arcellinida) and Euamoebida (mainly the genus *Copromyxa*, a dung-inhabiting amoeba) accounted for 37% and 25% of the tubulinean sequences, respectively. Rhizaria (23%) was in majority composed of Cercozoa (20%), and only of 2% of the mostly parasitic Endomyxa. Alveolata accounted for 21%, with 14% of Ciliophora and 6% of Apicomplexa. The Stramenopiles (also named Heterokonta, 7%), with 3% of microscopic brown algae (Ochrophyta), 2% of Sagenista (labyrinthulids) and 2% of Opalozoa (mostly parasitic, including the Oomycetes).

Among the animals (Metazoa), insects (41%) and ringed worms (34%) were most abundant (Fig. 2C), followed by nematodes (16%). Arthropoda, composed mostly of insects (41%) and Chelicerata (4%), among them mostly arachnids, accounted for 45% of the sequences. The microscopic rotiferans and tardigrads accounted for 2% each.

The fungal diversity was definitely dominated by Ascomycota (46%) (Fig. 2D), of which mostly Pezizomycotina (37%), which also makes up most of Ascomycota in terms of described species. The Mucoromycota (19%) was composed in majority of undetermined taxa and the genus *Mortierella*, a soil saprotroph. Basidiomycota (19%) was dominated by the Agaricomycotina (18%) and only 1% of the plant parasites Pucciniomycotina. The mostly arbuscular mycorrhizal Glomeromycota accounted for 6%, and the mostly saprophytic chytrids for 3%. Interestingly, one of the most abundant genera of fungi (6.6%) belonged to the recently described class Archaeorhizomycetes (*incertae sedis* in Ascomycota), a very elusive group that doesn't form fruiting bodies. Archaeorhizomyces were widespread and abundant in the Alps, since it was the most abundant genus (6.6% of fungal sequences), as in Pinto-Figueroa et al. (2019), where it represented 13.3.% of all fungal ITS sequences.

### **Methodological discussion**

#### **a. rRNA biases, copy numbers and microbial abundance**

In our analyses, we broadly assume that the rRNA copy numbers, binned to the genus level, represent the abundance of these genera, which is tentative. There are indeed recognized biases, the first being sequencing errors overestimating the diversity, which we mitigated by deleting genera represented by one to three reads, and conducted, when necessary, analyses only on the most abundant taxa. Secondly, an undetermined proportion of wrong assignments is due to the inaccuracy of the reference databases, still containing mislabelled sequences. Since the higher the taxonomic level, the less likely are those errors, we conducted our analyses at the orders, classes and phyla levels. Thirdly, the

## Chapter 4

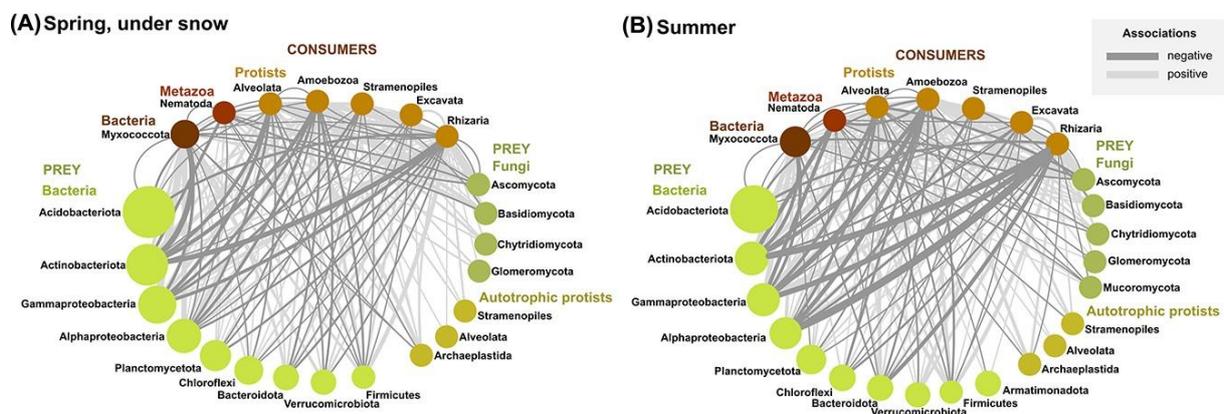
amplification step during library preparation may be less efficient for extremely A-T or G-C rich regions (Shi et al. 2021), which we could not test or evaluate in our study.

It is often assumed that metatranscriptomics only target active soil microbes, e.g. (Geisen et al. 2015; Harkes et al. 2019). While this is certainly true for messenger RNAs, it has been shown that dormant cells can also contain high numbers of ribosomes ((Blazewicz et al. 2013 and references therein). Therefore, rRNA-based studies of the soil microbiome, which is likely to include a high proportion of dormant cells, cannot provide information on microbial activity.

It is still currently unclear how to relate rRNA transcript numbers with microbial cell numbers for inferring abundances (Geisen et al. 2015; Petters et al. 2021; Söllinger et al. 2022). Biases are probably introduced by varying ribosomal RNAs copy numbers per cell, between prokaryotes and eukaryotes, between taxa and even inside taxa (Harkes et al. 2019). Recently, a down-regulation of the bacterial protein biosynthesis machinery has been observed in response to warming (Söllinger et al. 2022). At present, these biases are unquantifiable, but some hints suggest that they may only be cavils. A study showed that the cell abundance of small planktonic eukaryotes was correlated with the rDNA abundance (Gong et al. 2020) but for two species of ciliates, it varied with cell-size, so that the rRNA concentration was higher in smaller than larger cells (Fu et al. 2017). Accordingly, the ratio of the phospholipids' fatty acid-derived biomass of fungi to bacteria is 8.6 (6.7~11.0) in cold climates (He et al. 2020), while in metatranscriptomics bacteria dominated by large the counts of SSUs (Urich et al. 2008) and the present study. Thus, the higher biomass of soil fungi, which have larger cells than bacteria, is not mirrored by rRNA-based approach, and the results from both methods cannot be compared. General quantification methods are still wanting and will represent a huge step forward when available.

### b. Network analyses

In addition to the analyses conducted as described in the main text and showed in Fig. 6 - associations between consumers and consumers and preys in spring and summer (and Fig. S6, associations between preys only), we also tested the effect of selected environmental parameters on these networks, i.e. mountain, altitude, soil temperature, days under snow, soil water content, pH, microbial C, microbial N, microbial CN, Organic C, Dissolved N. The biotic associations of the obtained networks with and without environmental data were quite similar. Thus, FlashWeave's predictions appear to be robust



**Co-occurrence networks of abundant phyla of consumers and preys, with selected environmental factors. a** Spring, under the snow. **b** Summer. The size of the nodes (dots) is proportional to the number of reads. Edges (connecting lines) represent positive (light grey) or negative (dark grey) correlations, with line width proportional to the number of correlations. Self-loops, taxa with a single edge and connections between preys are not shown. Note the similarity with networks shown in Fig. 6.

## Chapter 4

against shared-niche biases, i.e. inflation of associations between taxa driven by environmental factors, even with missing environmental data.

### Network metrics

	<b>Snow</b>	<b>Summer</b>
<b>Number of Nodes</b>	1869	2017
<b>Number of Edges</b>	2779	3003
<b>Avg. Number of neighbors</b>	2.998	3010
<b>Network diameter</b>	19	22
<b>Network radius</b>	12	12
<b>Characteristic path length</b>	7.723	7.694
<b>Clustering coefficient</b>	0.002	0.002
<b>Network density</b>	0.002	0.002
<b>Network heterogeneity</b>	0.467	0.483
<b>Network centralization</b>	0.004	0.003
<b>Connected components</b>	12	16

### Supplementary references

- Blazewicz, S. J., R. L. Barnard, R. A. Daly & M. K. Firestone (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* 7: 2061-2068.
- Fu, R. & J. Gong (2017). Single cell analysis linking ribosomal (r)DNA and rRNA copy numbers to cell size and growth rate provided insights into molecular protistan ecology. *J. Euk. Microbiol.* 64: 885–896.
- Geisen, S., A. Tveit, I. M. Clark, A. Richter, M. M. Svenning, M. Bonkowski, et al. (2015). Metatranscriptomic census of active protists in soils. *ISME J.* 9: 2178–2190.
- Gong, F., G. Li, Y. Wang, Q. Liu, F. Huang, K. Yin, et al. (2020). Spatial shifts in size structure, phylogenetic diversity, community composition and abundance of small eukaryotic plankton in a coastal upwelling area of the northern South China Sea. *J. Plankton Res.* 6: 650-667.
- Harkes, P., A. K. A. Suleiman, S. J. J. van den Elsen, J. J. de Haan, M. Holterman, E. E. Kuramae, et al. (2019). Conventional and organic soil management as divergent drivers of resident and active fractions of major soil food web constituents. *Sci. Rep.* 9: 13521.
- He, L., J. L. Mazza Rodrigues, N. A. Soudzilovskaia, M. Barceló, P. Axel Olsson, C. Song, et al. (2020). Global biogeography of fungal and bacterial biomass carbon in topsoil. *Soil Biol. Biochem.*
- Petters, S., V. Groß, A. Söllinger, M. Pichler, A. Reinhard, M. M. Bengtsson, et al. (2021). The soil microbial food web revisited: Predatory myxobacteria as keystone taxa? *ISME J.* 15: 2665-2675.
- Pinto-Figueroa, E. A., E. Seddon, E. Yashiro, A. Buri, H. Niculita-Hirzel, J. Roelof van der Meer, et al. (2019). Archaeorhizomycetes spatial distribution in soils along wide elevational and environmental gradients reveal co-abundance patterns with other fungal saprobes and potential weathering capacities. *Front. Microbiol.* 10: 656.
- Shi, H., Y. Zhou, E. Jia, M. Pan, Y. Bai & Q. Ge (2021). Bias in RNA-seq library preparation: Current challenges and solutions. *BioMed Res. Int.* 2021: 6647597.
- Söllinger, A., J. Séneca, M. Borg Dahl, L. L. Motleleng, J. Prommer, E. Verbruggen, et al. (2022). Down-regulation of the bacterial protein biosynthesis machinery in response to weeks, years, and decades of soil warming. *Sci. Adv.* 8: eabm3230.
- Urlich, T., A. Lanzén, J. Qi, D. H. Huson, C. Schleper & S. C. Schuster (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* 3: e2527.

---

## CHAPTER 5

# Algae-fungi symbioses and bacteria-fungi co-exclusion drive tree species-specific differences in canopy bark microbiomes

Jule Freudenthal<sup>a</sup>, Kenneth Dumack<sup>a</sup>, Stefan Schaffer<sup>b</sup>, Martin Schlegel<sup>c,d</sup> and Michael Bonkowski<sup>a</sup>

<sup>a</sup>Terrestrial Ecology, Institute of Zoology, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zùlpicher Str. 47b, 50674 Köln, Germany

<sup>b</sup>Molecular Evolution and Systematics of Animals, Institute of Biology, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany

<sup>c</sup>Biodiversity and Evolution, Institute of Biology, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany

<sup>d</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle Jena Leipzig, Puschstraße 4, 04103 Leipzig, Germany

*Keywords: forest ecosystems, microbial diversity, host specificity, eukaryotes, prokaryotes, lichen, plant holobiont*

### Abstract

With over 3 trillion trees, forest ecosystems comprise nearly one-third of the terrestrial surface of the Earth. Very little attention has been given to the exploration of the above-ground plant microbiome of trees, its complex trophic interactions, and variations among tree species. To address this knowledge gap, we applied a primer-independent shotgun metatranscriptomic approach to assess the entire living canopy bark microbiome comprising prokaryotic and eukaryotic primary producers, decomposers, and various groups of consumers. With almost 1500 genera, we found a high microbial diversity on three tree species with distinct bark textures: oak (*Quercus robur*), linden (*Tilia cordata*), both with rough bark, and maple (*Acer pseudoplatanus*) with smooth bark. Core co-occurrence network analysis revealed a rich food web dominated by algal primary producers, and bacterial and fungal decomposers, sustaining a diverse community of consumers, including protists, microscopic metazoans, and predatory bacteria. Whereas maple accommodated a depauperate microbiome, oak and linden accommodated a richer microbiome mainly differing in their relative community composition: Bacteria exhibited an increased dominance on linden, whereas co-occurring algae and fungi dominated on oak, highlighting the importance of algal-fungal lichen symbioses even at the microscopic scale. Further, due to bacteria-fungi co-exclusion, bacteria on bark are not the main beneficiaries of algae-derived carbon compounds as it is known from aquatic systems.

### 1 Introduction

Forests cover almost one-third of the terrestrial surface of the Earth, with more than 3 trillion trees worldwide [1, 2]. They form the primary interface between terrestrial biomes and the atmosphere, of which bark surfaces, the rhytidome, constitute a huge and important part [3, 4]. However, microbial pathogens, such as *Oomycota* (e.g. Sudden Oak Death, Lime Disease) or fungi (e.g. Sooty Bark Disease of maple), are increasingly threatening forest ecosystems [5, 6]. Thus, a deeper knowledge of the diversity and composition of what might be considered a “healthy microbiome” of forest trees is urgently needed.

Tree bark surfaces provide long-lasting habitats colonized by algae and cyanobacteria as primary producers, prokaryotes and fungi as decomposers, and predatory bacteria, heterotrophic protists, and microscopic metazoans as consumers. They collectively form a complex microbial food web—the bark microbiome. Bark microbiomes may differ between tree species, due to substantial differences in the physical surface structures and chemical characteristics of tree barks [7, 8]. For example, rough bark may retain humidity and nutrients better than smooth bark and may provide increased protection against environmental stressors such as UV radiation and wash-off by rainfall [9, 10]. Primer-based DNA 16S, ITS, and 18S rRNA gene amplicon sequencing studies of selected microbial taxa indicated tree species-specific differences for bacteria and fungi on bark [9, 11, 12], but not for protists [13]. There are also indications that the bark structure influences the composition of algae and microscopic metazoan communities [14, 15]. However, no study has yet comprehensively analyzed the entire canopy bark microbiome and its potential trophic interactions [7].

So how diverse are the entire bark microbiomes among tree species and are they shaped by differences in bark surface texture? Using a primer-independent shotgun metatranscriptomic approach, the present study analyzes the diversity and composition of the bark microbiomes of three tree species in a floodplain forest. The strength of such an approach lies in the simultaneous assessment of the entire diversity of prokaryotes and eukaryotes [16] and avoiding the selectivity and biases inherent to primer-based methods, in particular the strong biases associated to universal eukaryotic primers [17, 18]. Additionally, metatranscriptomics have shown to be more accurate than metagenomics for the taxonomic identification of microbial communities at equal sequencing depths [19]. Further, this RNA-based approach mainly targets living microorganisms and is therefore rather insensitive to dead microorganisms and environmental DNA [20, 21] which can be highly enriched on tree bark [22]. This approach allowed, to (i) comprehensively investigate the entire bark microbial and microfauna food web of canopies using one single methodology and dataset, (ii) explore potential relationships between primary producers, decomposers and consumers in these food webs, and (iii) analyze differences between microbial communities of different tree species. We hypothesized that (i) bark microbiomes of different tree species would differ and that (ii) microbial communities on the two tree

## Chapter 5

---

species with rough bark surfaces pedunculate oak (*Quercus robur*) and small-leaved linden (*Tilia cordata*) would be more similar than those on sycamore maple (*Acer pseudoplatanus*) with a distinctly smoother bark texture.

## 2 Materials and Methods

### 2.1 Sampling, RNA Extraction and Sequencing

Bark samples in tree canopies were collected on May 23, 2022, in the Leipzig floodplain forest (51.3657 N, 12.3094 E) in Germany in cooperation with the Leipzig Canopy Crane Facility. Five different tree canopies were sampled, for each of three tree species respectively: *Quercus robur*, *Tilia cordata*, and *Acer pseudoplatanus*. Samples were taken at the average mid-canopy height of  $23 \pm 3.5$  m. To further reduce the influence of spatial variation in canopies, accessible branches at all four celestial directions were sampled by scraping off the biocrust on the bark using a sterile scalpel blade. Special care was taken to avoid any macroscopic organisms, including any visible lichen or mosses. The material was collected in sterile tubes (SARSTEDT AG & Co. KG, Nümbrecht, Germany), immediately placed on dry ice, and subsequently stored at  $-80^{\circ}\text{C}$  until RNA extraction.

RNA was extracted from 0.2 g of the collected material with the RNeasy PowerSoil Total RNA Kit (Qiagen GmbH, Hilden, Germany) according to the instructions of the manufacturer, except for the disruption step, which was carried out on a FastPrep-24 homogenizer (MP Biomedicals, Eschwege, Germany) for 30 sec at  $5.5 \text{ m sec}^{-1}$ . Subsequently, the RNA was eluted in 90  $\mu\text{l}$  buffer SR7, directly followed by DNA digestion using Ambion Dnase I (Thermo Fisher Scientific Inc, Darmstadt, Germany) and RNA purification using the MEGAClear kit (Thermo Fisher Scientific Inc). RNA concentrations were quantified by Qubit RNA High sensitivity Assay-Kit (Thermo Fisher Scientific Inc) and Qubit 4 Fluorometer (Thermo Fisher Scientific Inc).

For library preparation, the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MS, USA) was used with 100 ng total RNA and without the depletion of ribosomal RNAs or selection of mRNAs. Sequencing was performed on a NovaSeq Sequencing System (Illumina Inc., San Diego, CA, USA) with a paired-end sequence length of 150 bp at the Cologne Center for Genomics (Köln, Germany).

### 2.2 Sequence Processing

After quality assessment of the raw data using FastQC v. 0.11.9 [23], quality filtering was performed using FastP v 0.23.2 [24]. All read pairs that contained any low-quality base  $< 10\%$  or had more than  $10\%$  bases with a quality score of  $< 25$  were excluded. Furthermore, adapter sequences and read pairs with low complexity or ambiguous bases were removed (Supplementary Table 1). Paired-end reads

## Chapter 5

---

were assembled into contigs using Mothur v. 1.48.0 [25]. Subsequently, all contigs were filtered for a minimum contig length of 100 bp without ambiguities or mismatches and a minimum overlap of 10 bases (Supplementary Table 1). Additionally, the proportion of messenger RNA (mRNA) in the samples was estimated using SortMeRNA v. 4.3.4 [26], which, on average, accounted for only 4.2% of the total RNA.

For the taxonomic assignment BLASTN v. 2.10.0 [27] was employed to compare the sequences against the SILVA 138 SSU Ref Nr. 99 database for the prokaryotes and the PR<sup>2</sup> database v. 4.14.0 [28] for the eukaryotes. The sequences were filtered by an *E* value threshold of  $1e^{-50}$  keeping only the best match (Supplementary Table 1). However, considering the limitations of the sequencing method, i.e. sequencing of random fragments with a length of 150 bp, a very conservative taxonomic assignment at genus level was performed (similarity threshold of  $\geq 85\%$ ) to avoid an overestimation of the microbial diversity [29, 30]. Further, genera with proportional reads below 0.001% across all samples as well as putative contaminants were excluded, such as macroscopic animals, land plants (*Embryophyceae*), and chloroplasts. Fungi were classified based on the taxonomic assignment into functional groups as lichen-forming fungi, (facultative) yeasts, and plant parasites [31], and lichen-forming algae were identified according to a previously published database [32]. Because the vast majority of organisms were microbial, we will refer to the assessed communities as “microbial communities”, although they also included microscopic metazoans, i.e. rotifers, tardigrades, and nematodes. Further, we will refer to the prokaryotic community as “bacterial community”, as *Archaea* represented on average only  $0.002 \pm 0.004\%$  of the prokaryotic community.

### 2.3 Statistical Analyses

All statistical analyses and data visualizations were carried out in R v. 4.3.1 [33] with the packages ggalt v. 0.4.0, ggplot2 v. 3.4.2, ggpubr v. 0.6.0, ggrepel v. 0.9.3, ggsignif v. 0.6.4, ggthemes v. 4.2.4 [34–39]. Rarefaction curves were calculated from count data (vegan v. 2.6-4::rarecurve [40]) to check whether the sampling effort covered the taxonomic diversity (Supplementary Fig. 1). As all rarefaction curves reached complete saturation, data analysis was proceeded without normalizing for sequencing depth (Supplementary Fig. 1).

The relative proportion of the 10 most abundant classes of bacteria, algae, fungi, heterotrophic protists, and microscopic metazoans that constituted more than 1% of the respective community as well as the 10 most abundant genera classified to genus level was visualized by Sankey diagrams (riverplot v. 0.10::makeRiver [41]). Core co-occurrence network analyses were conducted using Sparse and Compositionally Robust Inference of Microbial Ecological Networks (SPIEC-EASI [42]). This method considers the compositionality of the data and reduces indirect edges by using sparse neighborhood

or inverse covariance selection. To improve the robustness of the analysis, only genera accounting for more than 0.01% across all samples and present in all samples were included, as co-absence may yield spurious edges [43]. The remaining genera were combined into a pseudo-taxon. The sparse Meinshausen-Buhlmann's neighborhood selection method, with a  $\lambda_{\min}$  ratio=0.001 and  $n_{\lambda}=100$  (package SpiecEasi v. 1.1.2) was used for network inference. The number of associations between the genera was summarized for functional trophic and taxonomic groups, i.e algae and cyanobacteria (primary producers), fungi and other bacteria (decomposers), and predatory bacteria, protists and microscopic metazoans (consumers), showing the ratio of positive to negative associations. The core co-occurrence network was visualized using Cytoscape v. 3.9.0 [44].

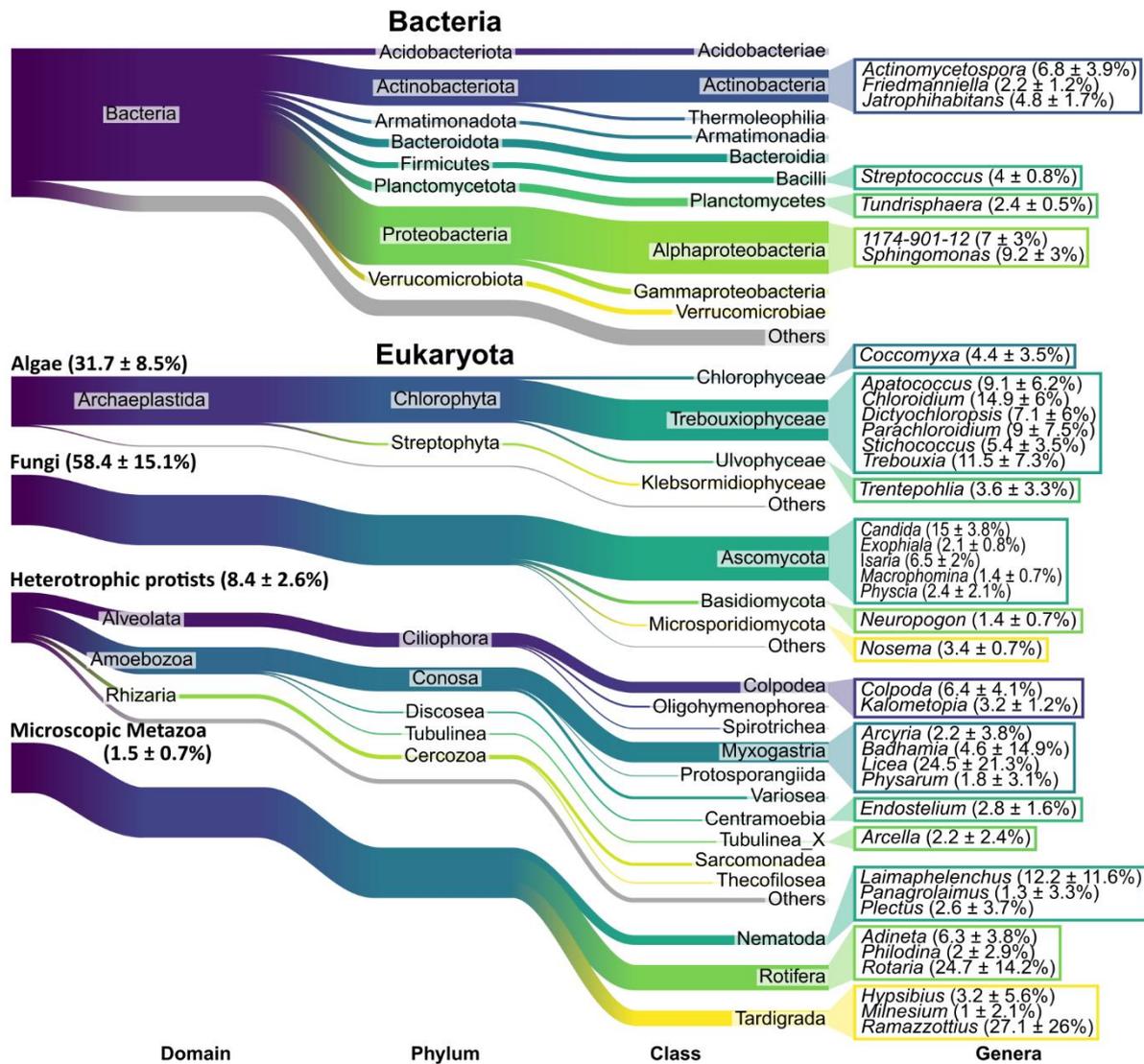
To compare alpha and beta diversities of microbiomes between tree species, we considered the whole community as well as the 10 most abundant classes (fungi and microscopic metazoans) or domains (protists and algae) of the eukaryotes and the 10 most abundant phyla of the bacteria represented by at least 10 genera (referred to as microbial groups). Alpha diversity was assessed based on genus richness and Pielou's Evenness index using the functions `vegan v. 2.6-4::diversity` and `vegan v. 2.6-4::specnumber` [40]. Alpha diversity indices among tree species were compared by using a Kruskal-Wallis test and Wilcoxon post-hoc test (`rstatix v. 0.7.2::kruskal_test` and `rstatix v. 0.7.2::wilcox_test` [45]), corrected for multiple comparisons [46]. To visualize shifts in microbial community compositions in relation to tree species, non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity of relative abundances was conducted (`vegan v. 2.6-4::metaMDS` [40]). Differences in community composition between tree species were tested by a permutational multivariate analysis of variance (perMANOVA, `vegan v. 2.6-4::adonis2` [40]) and the function `pairwise.adonis2` for pairwise comparisons (`pairwiseAdonis v. 0.4.1` [47]). In addition, we performed NMDS for algae, bacteria, fungi, and protists, respectively, and correlated the relative abundance of the microbial groups as well as of the respective 10 most abundant genera that could be classified to the genus level onto the ordination (`vegan v. 2.6-4::envfit` [40]), only significant ( $P$  value < 0.05) correlations are shown.

### 3 Results

#### *3.1 Overall microbial diversity of bark surfaces*

We obtained  $16 \pm 1.5$  million ribosomal RNA sequences per sample, revealing altogether a microbiome diversity of almost 1500 genera from 28 algal, 53 bacterial, nine fungal, 33 protistan, and three microscopic metazoan classes on the tree bark surfaces (Fig. 1, Supplementary Fig. 2). On average,  $60.3 \pm 9.4\%$  of these sequences could be assigned to in total 645 bacterial genera and  $39.7 \pm 9.4\%$  to 113 algal, 558 fungal, 155 protistan, and 16 microscopic metazoan genera. Algae as primary producers

## Chapter 5



**Figure 1: Microbial community composition in tree canopies.** Sankey diagrams with the mean ( $\pm$  1SD) relative abundance (%) across all tree species (N=15) showing the 10 most abundant classes of bacteria, algae, fungi, heterotrophic protists, and microscopic metazoa that accounted for more than 1% of the respective community. Additionally, the 10 most abundant genera that were classified to genus level are shown for each community, respectively.

accounted for  $31.7 \pm 8.5\%$  of the eukaryotic reads, whereas cyanobacteria represented only a minor fraction ( $0.2 \pm 0.1\%$  of bacterial reads). Accordingly, primary producers were dominated by green algae in *Trebouxiophyceae* ( $83.2 \pm 30.8\%$  of algal reads), with dominant genera such as *Chloroidium*, *Trebouxia*, and *Apatococcus* ( $14.9 \pm 6$ ,  $11.5 \pm 7.3$ ,  $9.1 \pm 6.2\%$  of algal reads, respectively) as well as *Coccomyxa* (*Chlorophyceae*,  $4.4 \pm 3.5\%$  of algal reads) and *Trentepohlia* (*Ulvophyceae*,  $3.6 \pm 3.3\%$  of algal reads). Among decomposers, by far the most abundant bacteria were *Alphaproteobacteria* ( $34.9 \pm 5.2\%$  of bacterial reads), dominated by *Sphingomonas* accounting for nearly 10% of the bacterial reads, followed by *Actinobacteria*, *Planctomycetes*, *Bacteroidia*, and *Acidobacteria* ( $22 \pm 7$ ,  $6 \pm 3.6$ ,  $5.4 \pm 1.7$ , and  $4.4 \pm 3.9\%$  of bacterial reads, respectively). Fungal decomposers accounted for more than half of the eukaryotic reads ( $58.4 \pm 15.1\%$ ). The great majority belonged to *Ascomycota*

## Chapter 5

---

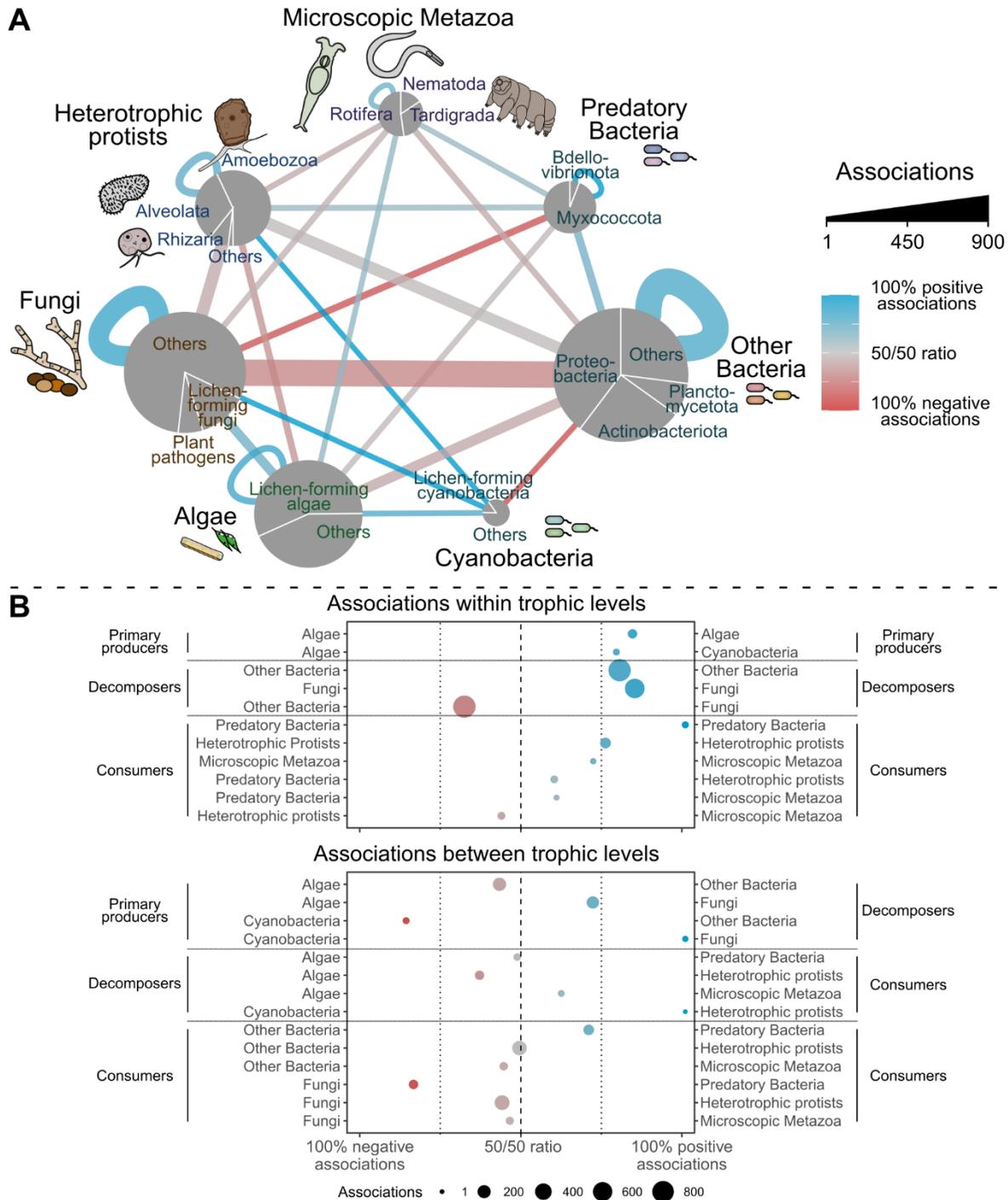
(88.1 ± 24% of fungal reads), followed by *Basidiomycota* (6.8 ± 2.1% of fungal reads). However, 18 ± 1% of the fungal reads were identified as yeast-forming taxa like *Candida* (15 ± 3.8% of fungal reads). Only a minor proportion of potential plant pathogens (7%) were detected among fungi. Consumers were dominated by protists, which accounted for 8.4 ± 2.6% of eukaryotic reads, predominantly the amoebozoan classes *Myxogastria* and *Variosea* (42.5 ± 24.6 and 5.5 ± 1.6% of protistan reads, respectively), and ciliates in *Colpodea* and cercozoans in *Sarcomonadea* (22.2 ± 9.5 and 5.3 ± 2.4% of protistan reads, respectively). The genera *Licea* and *Colpoda* together accounted for almost half of the community of protists (24.5 ± 21.3 and 22.2 ± 9.5% of protistan reads, respectively). Microscopic metazoans accounted for 1.5 ± 0.7% of eukaryotic reads and were dominated by *Rotifera* and *Tardigrada* (50.6 ± 27 and 32.9 ± 27% of metazoan reads, respectively), with the tardigrade genus *Ramazzottius* accounting for more than one-quarter of the total metazoan community. The remaining microscopic metazoans were nematodes with 16.5 ± 13.3% reads at low frequency. Predatory bacteria occurred only in low abundance (2.3 ± 1.1% of bacterial reads) and were dominated by *Myxococcota* (93.6 ± 3.2% of predatory bacteria reads).

### *3.2 Core co-occurrence network analyses of the microbial food web*

Network analysis across all tree species revealed a complex core microbial food web composed of 579 genera with 3,936 associations (Fig. 2, Supplementary Fig. 3). The food web was dominated by bacterial and fungal decomposers (247 and 198 genera, respectively). Additionally, we found a high proportion of lichen-forming genera (57% of algal, 15% of cyanobacterial, and 13% of fungal reads) such as *Chloroidium*, *Trebouxia*, and *Apatococcus* (the three dominant algal genera), *Gloeocapsa* (cyanobacteria), as well as *Physcia*, *Lecanora*, and *Lecidea* (fungi). Moreover, also a few associations between previously described lichen symbionts, such as between *Chlorella* and *Trapeliopsis*, were detected.

Most associations (67.2%) were found within trophic levels, and the majority (64.1%) of these associations were positive. Outstanding, however, was the high proportion of negative associations (68.6%) between bacterial and fungal decomposers. Between trophic levels, fungi showed primarily positive associations with cyanobacteria and algae (100% and 71.3% of associations, respectively), indicating potential symbiotic relationships. In contrast to fungi, bacteria showed highly negative associations with cyanobacteria and algae (86.7% and 57.7% of associations, respectively). Moreover, 68.8% of all associations between trophic levels occurred between consumers, i.e. protists, microscopic metazoans and predatory bacteria, and their potential prey. Predatory bacteria (*Myxobacteria*, *Bdellovibrio*) exhibited mainly positive associations with bacterial decomposers (70% of associations) but mainly negative associations with fungal decomposers (84.4% of associations),

whereas protists and microscopic metazoans did not show clear patterns in the ratio of negative to positive associations.

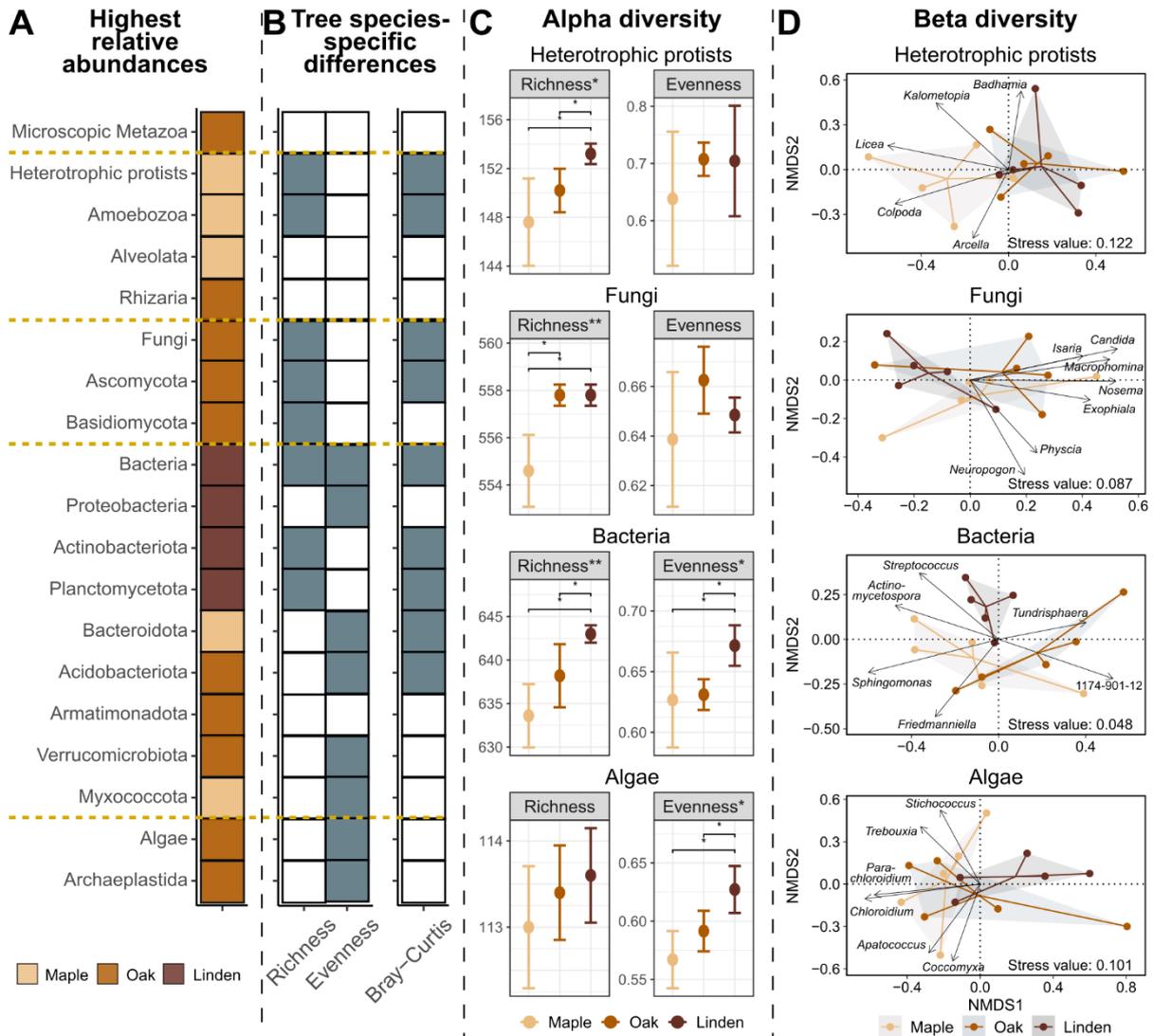


**Figure 2: Summarized core co-occurrence networks of the microbial community in tree canopies.** The core co-occurrence network (A) shows associations between genera on the bark surfaces in tree canopies (N=15). The associations are summarized for different taxonomic groups, i.e. for algae and cyanobacteria (primary producers), fungi and other bacteria (decomposers), and predatory bacteria, heterotrophic protists and microscopic metazoans (consumers). The node size is proportional to the log-transformed relative number of reads for bacteria and eukaryotes respectively. The thickness of the edges indicates the number of associations between taxonomic groups and the color code for the ratio of negative to positive associations. A summary of the network as a dot plot (B) shows the ratio of negative to positive associations with the size of the points corresponding to the number of associations between and within trophic levels.

3.3 Differences of bark microbiomes between the tree species

A comparison of the relative abundance, alpha, and beta diversity of the bark microbiomes revealed significant differences between maple, oak, and linden (Fig. 3, Supplementary Fig. 4-6). However, these differences were not uniform but strongly differed between microbial groups.

Maple accommodated a depauperate microbiome, with a significantly lower genus richness of bacteria (*Actinobacteriota*, *Planctomycota*), fungi and protists (*Amoebozoa*). Still, maple showed the highest



**Figure 3: Tree species-specific differences.** Heat map for selected microbial groups showing (A) highest relative abundance of specific microbial taxa on the respective tree species and (B) differences in genus richness and Pielou's evenness (alpha diversity) as well as in Bray-Curtis dissimilarities (beta diversity) between the tree species (N=5). Grey-colored areas indicate significant differences (alpha diversity= Kruskal-Wallis test, beta diversity=PERMANOVA,  $P$  value < 0.05). Dot plots (C) show the alpha diversity indices per tree species (mean  $\pm$  1 SD). Significant differences across all tree species (Kruskal-Wallis test) are marked in the graph title, significant differences for pairwise comparisons of tree species (Wilcoxon test) are marked in the graph with stars (\*  $P$  value < 0.05; \*\*  $P$  value < 0.01; \*\*\*  $P$  value < 0.001). Non-metric multidimensional scalings of Bray-Curtis dissimilarity (D) show the multivariate dispersion of the samples. Lines are color-coded by tree species and link samples of each tree species to their centromere. The relative abundances of the ten most abundant genera classified to genus level were correlated onto the respective ordinations, significant correlations ( $P$  value < 0.05) are shown as arrows.

## Chapter 5

---

relative abundance of the bacterial phyla *Bacteroidota*, *Myxococcota* (Fig. 3A), and the genera *Sphingomonas* and *Actinomyces* (Fig. 3D). Likewise, among eukaryotes, the algal genera *Coccomyxa*, *Chloridium*, *Parachloroidium*, and *Stichococcus* (Fig. 3D), as well as the protistan genera *Licea* (*Amoebozoa*) and *Colpoda* (*Alveolata*) (Fig. 3A, 3D) dominated on maple. This resulted in a low evenness of algae and bacteria on maple (Fig. 3B & C) and a distinct community composition of protists compared to the other tree species (Fig. 3B & D, Supplementary Fig. 4).

The oak microbiome showed the highest relative abundance of algae and fungi, *Rhizaria* (protists), and microscopic metazoans (Fig. 3A). Many lichen-forming algae such as *Apatococcus*, *Dictyochloropsis*, *Trebouxia*, and *Trentepohlia* together with lichen-forming fungi such as *Physcia* and yeast cells (*Candida*) showed their highest relative abundance on oak (Fig. 3D), as well as microscopic metazoans in *Adineta*, *Rotaria* (*Rotifera*), *Ramazzottius* (*Tardigrada*), and fungivorous *Laimaphelenchus* (*Nematoda*). In terms of genus richness, oak exhibited an intermediate genus richness between maple and linden (Fig. 3B & C). Similarly, the genus evenness of algae and bacteria was higher on oak than on maple but lower than on linden (Fig. 3B & C). However, the oak microbiome exhibited no distinct community composition in terms of beta diversity (Fig. 3B & D, Supplementary Fig. 4).

The linden microbiome showed the highest relative abundance of bacteria, with the dominant phyla *Proteobacteria*, *Actinobacteriota*, and *Planctomycetota* (Fig. 3A), along with the highest genus richness of bacteria and protists (Fig. 3C). Additionally, with the highest evenness of algae and bacteria, linden bark hosted quite diverse communities of primary producers and decomposers (Fig. 3C). The composition of bacterial and fungal communities on linden differed significantly from maple and oak (Fig. 3D, Supplementary Fig. 4).

## 4 Discussion

### 4.1 Bark microbiome

Our comprehensive analysis revealed a diverse bark microbiome in tree canopies, comprising nearly 1500 genera. Quantitatively, our method revealed its strengths in uncovering a substantially higher diversity of eukaryotes compared to earlier amplicon sequencing studies on bark microbiomes [e.g. 9, 48–50], in particular of algae and protists. Qualitatively, our findings correspond well with data obtained by light microscopy and culture-based studies [e.g. 51–54].

Tree bark surfaces typically constitute sunlight-exposed but water and nutrient-limited habitats for microorganisms. Primary producers form the basis of the microbial food web. Their ability to utilize sunlight for carbon fixation forms an important nutrient and energy source for heterotrophic organisms [3, 7, 55]. Algae were by far the most diverse and dominant primary producers on bark with 113 genera in 28 classes, whereas cyanobacteria contributed only a small proportion. We confirmed

## Chapter 5

---

the prevalence of *Chlorophyta* [48–50], especially the *Trebouxiophyceae*, which comprise spherical or ellipsoid solitary algae, such as *Chloroidium* and *Trebouxia*, and sarcinoid colony-forming algae, like *Apatococcus* [14, 52]. In addition, we found a high abundance of uniseriate filament-forming algae such as *Trentepohlia* (*Trentepohliales*), one of the most widespread terrestrial algae [56].

The decomposer communities on bark were dominated by bacteria with 644 genera in 52 classes, whereas *Archaea* were rather scarce. The general prevalence of *Alphaproteobacteria*, *Actinobacteria*, *Bacteroidia*, and *Acidobacteria* corresponds to previous amplicon sequencing and metagenomics studies [9, 12, 48, 49, 57]. With almost 10%, the alphaproteobacterium *Sphingomonas* was especially abundant in our bacterial reads. *Sphingomonas* is a well-known inhabitant of bark surfaces [e.g. 9, 48] and may promote plant growth and resistance by producing various phytohormones and antifungal compounds [58, 59]. Fungi were dominated by *Ascomycota* comprising 88% of fungal reads. Among yeasts *Candida* (15% of fungal reads) reached an exceptionally high relative abundance, a result also sustained by culture-based studies [51, 60]. The ability of *Candida* yeasts to degrade cellulose and hemicelluloses may favor its presence on bark [61].

Among consumers, protists were by far the most diverse group with 155 genera in 33 classes. A particular strength of our metatranscriptomic study is that it allows a comprehensive comparison of the dominance of protistan taxa and their life strategies in tree canopies. The protists were dominated by *Amoebozoa*, which are rarely detected and grossly underestimated by DNA-based methods due to their inherent high sequence variability, and variable sequence length of the V4 barcoding region in the 18S ribosomal RNA gene [16, 62]. Our RNA-based approach instead reliably identified a high richness of the amoebozoan class *Myxogastria* with various corticolous (bark-loving) genera such as *Licea*, *Arcyria*, *Physarum*, and *Echinostelium* [54, 63], which confirms the findings of microscopic/culture-based studies [54, 63]. The corticolous *Variosea* and *Myxogastria* are particularly well adapted to the variable life conditions on bark surfaces, as they can form resistant cysts, sporocarps, and spores within hours at declining moisture conditions [64, 65]. Other protists appear particularly predisposed for life on bark surfaces as they can complete their whole life cycle in just a few hours. This is true for flagellates and amoeboflagellates in *Sarcomonadea* (*Cercozoa*) [66] and in particular for *Colpoda* (*Colpodea*) species among ciliates. Early light-microscopic investigations of protists on plant surfaces revealed a prevalence of *Colpoda cucullus*, a ciliate with an impressively short life cycle due to cell division within drought-resistant resting cysts [67, 68]. In our study, *Colpodea* clearly dominated the ciliate community, however, aside from the dominant *Colpodea*, a considerable number of other tree-adapted species was assessed by previous studies [53, 69]. Another adaptation to life on bark was detected by comparing communities of canopy protists with those in the litter layer and soil on the ground [70]. Apart from the dominance of small r-strategists among protists discussed above, a striking prevalence of testate amoebae was described [70], whose shells may prolong their

## Chapter 5

---

foraging times due to enhanced protection against drought. Molecular studies of cercozoan diversity in tree canopies of the Leipzig floodplain forest [13] confirmed the prevalence of drought-resistant testate amoeba (e.g. *Thecofilosea*) as well as of *Sarcomonadea* with rapid life cycles; a pattern well-reflected in our data (Fig. 1, Supplementary Fig. 2). However, protists are not the only consumers in microbial food webs. Recent studies highlighted predatory bacteria as one of the key consumers of bacteria in soil [71, 72]. According to the low proportion of predatory bacteria in our data, they appear to be of much lower importance on bark. Nematodes were quite rare, but dominated by the genus *Laimaphelenchus*, common colonizers of tree bark, known feed on algae, mosses and lichen [73]. Instead, the microfauna on trees was dominated by *Ramazzottius*, a widespread tardigrade genus [74], accounting for more than a quarter of the metazoan community.

The soundness of the metatranscriptomics data to results obtained by direct microscopic estimations of protistan diversity, the reliable detection of groups commonly isolated from bark that evade detection by general eukaryotic primers [17, 18], and coherence to the diversity obtained at much greater sequencing depth with taxon-specific primers [e.g. 13], provide strong evidence for the reliability of this method to reflect the true diversity of the bark microbiome.

### 4.2 Core microbial food web

The ubiquitous taxa across the three tree species comprised the core microbial food web. It was composed of 579 distinct genera with a great richness of bacterial (247) and fungal (198) genera among the decomposers, 51 genera of primary producers and 83 genera of consumers.

The network analysis indicates that lichen symbioses are a characteristic feature of the core bark microbiome. Among the primary producers, almost 60% of algal reads and 15% of cyanobacterial reads could be assigned to potential lichen symbionts (photobionts). The dominant *Trebouxia* (11.7% of all algal reads) is known as the most common algal lichen symbiont [32, 75]. Correspondingly, 13% of the fungal reads were assigned to typical lichen-forming taxa (mycobionts) such as *Physcia*, *Lecanora*, and *Lecidea* [76, 77]. In addition, 71.3% of algae-fungi associations were positive. These even include some associations between previously described lichen partners such as *Chlorella* and *Trapeliopsis* [32] although identifying lichen partners is challenging, as more and more sequencing studies reveal unsuspected promiscuity among fungi and algal partners in lichen symbioses [78–80]. The strong positive associations between algae and fungi and the high proportion of lichen symbionts, despite the strict avoidance of sampling any visible lichens, indicate widespread early stages of symbioses, e.g. loosely associated cells, hyphal webs or propagules [32, 81, 82].

The high proportion of negative associations of bacteria with algae and fungi indicates an antagonistic relationship. In contrast to aquatic systems, where bacteria are the main beneficiaries of phytoplankton

## Chapter 5

---

exudates and where symbiotic interactions between algae and bacteria are widespread [83, 84], bacteria compete for algal exudates with fungi on bark. Fungal symbioses is likely favored by the dry conditions on bark [85], and may be further enhanced by antibacterial metabolites of algae and fungi [82, 86]. *Vice versa*, *Actinobacteria* and *Sphingomonas* which dominated the bacterial bark community, are known for producing antifungal metabolites [59, 87, 88]. This again may feed back on potential plant pathogenic fungi or oomycetes [5, 89], which were only detected at low frequency (fungi) or absent (oomycetes).

The bark food web contained a substantial proportion of consumers at higher trophic levels. Assuming a gross growth efficiency of 30-50% [90], the higher trophic levels on the tree bark must be sustained by a substantially larger proportion of primary, bacterial, and fungal production. Protists are the primary consumers in our food web (8.4% of eukaryotic reads), followed by predatory bacteria (2.3% of bacterial reads) and microscopic metazoans (1.5% of eukaryotic reads). Many protists and metazoans of the phyllo- and rhizosphere are omnivores and consume a broad range of algae, bacteria, and fungi [91–93], which is reflected by the numerous associations between these groups in the network analysis. Most of the protists and microscopic metazoans did not show any clear directional ratio of negative to positive associations, as can be expected from omnivores with frequent prey shifts. An exception was *Colpodea* (*Colpoda* and *Exocolpoda*) exhibiting numerous associations with different bacterial taxa, including strongly positive associations with cyanobacteria. In agreement with previous studies [71, 72], predatory bacteria exhibited strong positive co-occurrence with bacterial decomposers in our network analyses. Predatory bacteria also exhibited strong co-exclusion with fungi which might be partly explained by their ability to inhibit fungal growth [72], but more likely reflects an indirect association driven by the co-exclusion of bacterial and fungal decomposers on bark.

### *4.3 Tree species-specific differences of their bark microbiomes*

We hypothesized that trees with rough bark (linden and oak) would host more similar microbiomes as compared to trees with smooth bark (maple). Tree species-specific differences have so far been shown for specific microbial groups on bark surfaces [9, 11, 12, 14, 15, 65, 94], but not for the entire microbiome. In this study, we confirmed tree species-specific differences among the bark microbiomes, driven by different microbial groups.

Maple with smooth bark accommodated a poorer microbiome, characterized by low genus richness and evenness, compared to oak and linden with rough bark. Beta diversity patterns showed an increased relative abundance of (a) desiccation-tolerant genera such as the bacterial *Sphingomonas* and the algae *Chloroidium*, *Trebouxia*, *Stichococcus*, and *Coccomyxa* [14, 52, 95, 96]; and (b) UV protected taxa such as the pigmented bacteria in *Chitinophagaceae* [97] and eukaryotes in

## Chapter 5

---

*Myxogastrina* [98]. The smooth maple bark may provide less protection against environmental stressors like UV radiation or desiccation, thus supporting a sparser microbiome. However, tree species-specific differences in the composition of canopy microbiomes, could only partly be explained by bark topology. Whereas differences between smooth and rough bark communities were mainly driven by genus richness and evenness, microbial groups on oak and linden with rough bark differed mostly in their relative proportions. Oak was characterized by a higher relative abundance of algae and fungi, including many lichen-forming genera. Linden was characterized by a higher relative abundance of bacteria. Furthermore, microbial communities on linden exhibited the highest evenness and a distinct community composition in terms of beta diversity. These differences in evenness and beta diversity were mainly driven by primary producers and decomposers suggesting that they are stronger shaped by selective pressures than higher trophic levels.

### 5 Concluding remarks

The shotgun metatranscriptomic approach allowed the simultaneous assessment and thus comparison of the entire living prokaryotic and eukaryotic microbial canopy bark communities among oak, linden and maple. We identified highly diverse and tree species-specific differences in the composition of canopy microbiomes, only partly explained by bark topology. Still, all tree species also harbored a joint and taxonomic diverse core microbiome. Strong algal-fungal co-occurrence indicated microscopic lichen symbioses. Also, the life-strategies of the dominant protistan taxa reflected a variety of specific adaptations to the harsh environmental conditions on bark. We detected strongly negative associations of algae and fungi with bacteria. Their consumers were myxobacteria, microscopic metazoans, and especially protists. The many omnivores among protists, also consuming algae, yeasts, and other protists, reflect more complex food-web interactions. Although potential plant pathogens were rare, the method allowed their assessment relative to the total microbial diversity, and network analysis showed great potential to identify likely microbial interdependencies. Understanding the microbiome dynamics is essential, as they impact host plant fitness, function, and productivity, thereby influencing tree health and ecosystem productivity.

### Author contributions

KD, MS and MB designed the study. JF and SS conceived and conducted the sampling with help from all authors. JF and SS conducted the RNA extraction, MS and KD helped with the laboratory work. JF performed the bioinformatic and statistical analyses and wrote the first draft of the manuscript. All co-authors commented on the manuscript and approved the submitted version.

### Data Availability

The raw data are available under NCBI BioProject PRJNA1105877. The code for the presented analyses is available at GitHub: <https://github.com/JFreude/BarkMicrobiome>.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgments

We would like to thank Ronny Richter for his support by operating the canopy crane. We thank the Leipzig Canopy Crane Platform of the German Center for Integrative Biodiversity Research (iDiv) for allowing us to sample the trees from their field trial. This study was supported by grants from the German Research Foundation (DFG) in the framework of the priority program SPP 1991: TAXON-OMICS (Project IDs 447013012 and 221301018). We acknowledge the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded High Performance Computing (HPC) system CHEOPS (Funding number: INST 216/512/1FUGG) as support.

### References

1. Aerts R, Honnay O. Forest restoration, biodiversity and ecosystem functioning. *BMC Ecol* 2011;**11**:29. <https://doi.org/10.1186/1472-6785-11-29>
2. Crowther TW, Glick HB, Covey KR *et al.* Mapping tree density at a global scale. *Nature* 2015;**525**:201–5. <https://doi.org/10.1038/nature14967>
3. Bringel F, Couée I. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Front Microbiol* 2015;**06**:486. <https://doi.org/10.3389/fmicb.2015.00486>
4. Ozanne CMP, Anhuf D, Boulter SL *et al.* Biodiversity meets the atmosphere: a global view of forest canopies. *Science* 2003;**301**:183–6. <https://doi.org/10.1126/science.1084507>
5. Jung T, Vannini A, Brasier CM. Progress in understanding *Phytophthora* diseases of trees in Europe 2004–2007. In: Goheen E.M., Frankel S.J. (eds.), *Phytophthoras in Forests and Natural Ecosystems: Proceedings of the Fourth Meeting of the International Union of Forest Research Organizations (IUFRO) Working Party S07.02.09*. United States: Department of Agriculture, 2009, 3–24.
6. Kespohl S, Riebesehl J, Grüner J *et al.* Impact of climate change on wood and woodworkers-*Cryptostroma corticale* (sooty bark disease): a risk factor for trees and exposed employees. *Front Public Health* 2022;**10**:973686. <https://doi.org/10.3389/fpubh.2022.973686>
7. Baldrian P. Forest microbiome: diversity, complexity and dynamics. *FEMS Microbiol Rev* 2017;**41**:109–30.
8. Rosell JA. Bark in woody plants: understanding the diversity of a multifunctional structure. *Integr Comp Biol* 2019;**59**:535–47. <https://doi.org/10.1093/icb/icz057>
9. Arrigoni E, Antonielli L, Pindo M *et al.* Tissue age and plant genotype affect the microbiota of apple and pear bark. *Microbiol Res* 2018;**211**:57–68. <https://doi.org/10.1016/j.micres.2018.04.002>
10. Buck JW, Lachance MA, Traquair JA. Mycoflora of peach bark: population dynamics and composition. *Canad J Bot* 1998;**76**:345–54. <https://doi.org/10.1139/cjb-76-2-345>
11. Hofmann B, Dreyling L, Dal Grande F *et al.* Habitat and tree species identity shape aboveground and belowground fungal communities in central European forests. *Front Microbiol* 2023;**14**:1067906. <https://doi.org/10.3389/fmicb.2023.1067906>
12. Lambais MR, Lucheta AR, Crowley DE. Bacterial community assemblages associated with the phyllosphere, dermosphere, and rhizosphere of tree species of the Atlantic forest are host taxon dependent. *Microb Ecol* 2014;**68**:567–74. <https://doi.org/10.1007/s00248-014-0433-2>
13. Jauss R-T, Walden S, Fiore-Donno AM *et al.* From forest soil to the canopy: increased habitat diversity does not increase species richness of Cercozoa and Oomycota in tree canopies. *Front Microbiol* 2020;**11**:592189. <https://doi.org/10.3389/fmicb.2020.592189>
14. Lüttge U, Büdel B. Resurrection kinetics of photosynthesis in desiccation-tolerant terrestrial green algae (Chlorophyta) on tree bark. *Plant Biol* 2010;**12**:437–44. <https://doi.org/10.1111/j.1438-8677.2009.00249.x>
15. Ptatscheck C, Milne PC, Traunspurger W. Is stemflow a vector for the transport of small metazoans from tree surfaces down to soil? *BMC Ecol* 2018;**18**:43. <https://doi.org/10.1186/s12898-018-0198-4>
16. Urich T, Lanzén A, Qi J *et al.* Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 2008;**3**:e2527. <https://doi.org/10.1371/journal.pone.0002527>
17. Freudenthal J, Ju F, Bürgmann H *et al.* Microeukaryotic gut parasites in wastewater treatment plants: diversity, activity, and removal. *Microbiome* 2022;**10**:27. <https://doi.org/10.1186/s40168-022-01225-y>
18. Lentendu G, Wubet T, Chatzinotas A *et al.* Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Mol Ecol* 2014;**23**:3341–55. <https://doi.org/10.1111/mec.12819>
19. Hempel CA, Wright N, Harvie J *et al.* Metagenomics versus total RNA sequencing: most accurate data-processing tools, microbial identification accuracy and perspectives for ecological assessments. *Nucleic Acids Res* 2022;**50**:9279–93. <https://doi.org/10.1093/nar/gkac689>
20. Carini P, Marsden PJ, Leff JW *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 2016;**2**:16242. <https://doi.org/10.1038/nmicrobiol.2016.242>
21. Geisen S, Tveit AT, Clark IM *et al.* Metatranscriptomic census of active protists in soils. *ISME J* 2015;**9**:2178–90. <https://doi.org/10.1038/ismej.2015.30>
22. Magyar D. The tree bark: a natural spore trap. *Asp Appl Biol* 2008;**89**:7–16.

## Chapter 5

---

23. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. 2010. <https://www.bioinformatics.babraham.ac.uk/index.html>. Babraham Institute, Cambridge, United Kingdom.
24. Chen S, Zhou Y, Chen Y *et al*. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinform* 2018;**34**:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>
25. Schloss PD, Westcott SL, Ryabin T *et al*. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41. <https://doi.org/10.1128/AEM.01541-09>
26. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinform* 2012;**28**:3211–7. <https://doi.org/10.1093/bioinformatics/bts611>
27. Camacho C, Coulouris G, Avagyan V *et al*. BLAST+: architecture and applications. *BMC Bioinforma* 2009;**10**:421. <https://doi.org/10.1186/1471-2105-10-421>
28. Guillou L, Bachar D, Audic S *et al*. The Protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 2013;**41**:D597–604. <https://doi.org/10.1093/nar/gks1160>
29. Cai L, Ju F, Zhang T. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl Microbiol Biotechnol* 2014;**98**:3317–26. <https://doi.org/10.1007/s00253-013-5402-z>
30. Choma M, Bárta J, Šantrůčková H *et al*. Low abundance of Archaeorhizomycetes among fungi in soil metatranscriptomes. *Sci Rep* 2016;**6**:38455. <https://doi.org/10.1038/srep38455>
31. Nguyen NH, Song Z, Bates ST *et al*. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol* 2016;**20**:241–8. <https://doi.org/10.1016/j.funeco.2015.06.006>
32. Sanders WB, Masumoto H. Lichen algae: the photosynthetic partners in lichen symbioses. *Lichenologist* 2021;**53**:347–93. <https://doi.org/10.1017/S0024282921000335>
33. R Core Team. *R: A Language and Environment for Statistical Computing*. 2023. <https://www.r-project.org/>. R Foundation for Statistical Computing, Vienna, Austria.
34. Rudis B, Bolker B, Marwick B, Schulz J, Matev R. Ggalt: extra coordinate systems, ‘geoms’, statistical transformations, scales and fonts for ‘ggplot2’. 2022. <https://CRAN.R-project.org/package=ggalt>.
35. Wickham H. *ggplot2*. *Wiley Interdiscip Rev Comput Stat* 2011;**3**:180–5. <https://doi.org/10.1002/wics.147>
36. Kassambara A. *Ggpubr: ‘ggplot2’ Based Publication Ready Plots*. 2023. <https://cran.r-project.org/package=ggpubr>.
37. Slowikowski K, Schep A, Hughes S *et al*. *Ggrepel: Automatically Position Nonoverlapping Text Labels with ‘ggplot2’*. 2023. <https://CRAN.R-project.org/package=ggrepel>.
38. Ahlmann-Eltze C. *Ggsignif: Significance Brackets for ‘ggplot2’*. 2022. <https://CRAN.R-project.org/package=ggsignif>.
39. Arnold JB. *Ggthemes: Extra Themes, Scales and Geoms for ‘ggplot2’*. 2021. <https://cran.r-project.org/package=ggthemes>.
40. Oksanen J, Blanchet FG, Friendly M *et al*. *Vegan: Community Ecology Package*. 2022. <https://cran.r-project.org/package=vegan>.
41. Weiner J. *Riverplot: Sankey or Ribbon Plots*. 2022. <https://CRAN.R-project.org/package=riverplot>.
42. Kurtz ZD, Müller CL, Miraldi ER *et al*. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 2015;**11**:e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>
43. Röttjers L, Faust K. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol Rev* 2018;**42**:761–80. <https://doi.org/10.1093/femsre/fuy030>
44. Shannon P, Markiel A, Ozier O *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504. <https://doi.org/10.1101/gr.1239303>
45. Kassambara A. *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. 2023. <https://cran.r-project.org/package=rstatix>.
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;**57**:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
47. Martinez Arbizu P. *pairwiseAdonis: Pairwise Multilevel Comparison Using Adonis*. 2020. <https://github.com/pmartinezarbizu/pairwiseAdonis>.
48. Aguirre-von-Wobeser E, Alonso-Sánchez A, Méndez-Bravo A *et al*. Barks from avocado trees of different geographic locations have consistent microbial communities. *Arch Microbiol* 2021;**203**:4593–607. <https://doi.org/10.1007/s00203-021-02449-6>

## Chapter 5

---

49. Dreyling L, Schmitt I, Dal Grande F. Tree size drives diversity and community structure of microbial communities on the bark of beech (*Fagus sylvatica*). *Front For Glob Change* 2022;**5**:858382. <https://doi.org/10.3389/ffgc.2022.858382>
50. Dreyling L, Penone C, Schenk NV *et al.* Biotic interactions outweigh abiotic factors as drivers of bark microbial communities in central European forests. *ISME Commun* 2024;**4**:ycae012. <https://doi.org/10.1093/ismeco/ycae012>
51. Dabassa Koricha A, Han D-Y, Bacha K *et al.* Occurrence and molecular identification of wild yeasts from Jimma zone, south West Ethiopia. *Microorganisms* 2019;**7**:633. <https://doi.org/10.3390/microorganisms7120633>
52. Darienko T, Gustavs L, Mudimu O *et al.* *Chloroidium*, a common terrestrial coccoid green alga previously assigned to *chlorella* (Trebouxiophyceae, Chlorophyta). *Eur J Phycol* 2010;**45**:79–95. <https://doi.org/10.1080/09670260903362820>
53. Foissner W. *Terrestrial and Semiterrestrial Ciliates*. (Ciliophora) from Venezuela and Galápagos: Protozoa, 2015.
54. Stephenson SL. Distribution and ecology of myxomycetes in temperate forests. II. Patterns of occurrence on bark surface of living trees, leaf litter, and dung. *Mycologia* 1989;**81**:608–21. <https://doi.org/10.1080/00275514.1989.12025792>
55. Thapa S, Prasanna R. Prospecting the characteristics and significance of the phyllosphere microbiome. *Ann Microbiol* 2018;**68**:229–45. <https://doi.org/10.1007/s13213-018-1331-5>
56. Brooks F, Rindi F, Suto Y *et al.* The Trentepohliales (Ulvophyceae, Chlorophyta): an unusual algal order and its novel plant pathogen. *Cephaleuros Plant Dis* 2015;**99**:740–53. <https://doi.org/10.1094/PDIS-01-15-0029-FE>
57. Cregger MA, Veach AM, Yang ZK *et al.* The Populus holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* 2018;**6**:31. <https://doi.org/10.1186/s40168-018-0413-8>
58. Asaf S, Numan M, Khan AL *et al.* *Sphingomonas*: from diversity and genomics to functional role in environmental remediation and plant growth. *Crit Rev Biotechnol* 2020;**40**:138–52. <https://doi.org/10.1080/07388551.2019.1709793>
59. Wachowska U, Kucharska K, Pluskota W *et al.* Bacteria associated with winter wheat degrade *Fusarium* mycotoxins and triazole fungicide residues. *Agron* 2020;**10**:1673. <https://doi.org/10.3390/agronomy10111673>
60. Yalçın HT, Findik B, Terzi Y *et al.* Isolation and molecular identification of industrially important enzyme producer yeasts from tree barks and fruits. *Arch Microbiol* 2021;**203**:1079–88. <https://doi.org/10.1007/s00203-020-02104-6>
61. Šuchová K, Fehér C, Ravn JL *et al.* Cellulose- and xylan-degrading yeasts: enzymes, applications and biotechnological potential. *Biotechnol Adv* 2022;**59**:107981. <https://doi.org/10.1016/j.biotechadv.2022.107981>
62. Voss C, Fiore-Donno AM, Guerreiro MA *et al.* Metatranscriptomics reveals unsuspected protistan diversity in leaf litter across temperate beech forests, with Amoebozoa the dominating lineage. *FEMS Microbiol Ecol* 2019;**95**:fiz142. <https://doi.org/10.1093/femsec/fiz142>
63. Everhart SE, Keller HW, Ely JS. Influence of bark pH on the occurrence and distribution of tree canopy myxomycete species. *Mycologia* 2008;**100**:191–204. <https://doi.org/10.1080/15572536.2008.11832476>
64. Ing B. Tansley review no. 62 the phytosociology of myxomycetes. *New Phytol* 1994;**126**:175–201. <https://doi.org/10.1111/j.1469-8137.1994.tb03937.x>
65. Shadwick JDL, Stephenson SL, Spiegel FW. Distribution and ecology of protostelids in great Smoky Mountains National Park. *Mycologia* 2009;**101**:320–8. <https://doi.org/10.3852/08-167>
66. Flues S, Bass D, Bonkowski M. Grazing of leaf-associated Cercomonads (Protists: Rhizaria: Cercozoa) structures bacterial community composition and function. *Environ Microbiol* 2017;**19**:3297–309. <https://doi.org/10.1111/1462-2920.13824>
67. Bamforth SS. Population dynamics of soil and vegetation protozoa. *Am Zool* 1973;**13**:171–6. <https://doi.org/10.1093/icb/13.1.171>
68. Mueller JA, Mueller WP. *Colpoda cucullus*: a terrestrial aquatic. *Am Midl Nat* 1970;**84**:1–12. <https://doi.org/10.2307/2423721>
69. Foissner W, Agatha S, Berger H. *Soil Ciliates (Protozoa, Ciliophora) from Namibia (Southwest Africa), with Emphasis on Two Contrasting Environments, the Etosha Region and the Namib Desert. Part I: Text and Line Drawings. Part II: Photographs*. Linz, Austria: Biologiezentrum des Oberösterreichischen Landesmuseums, 2002.

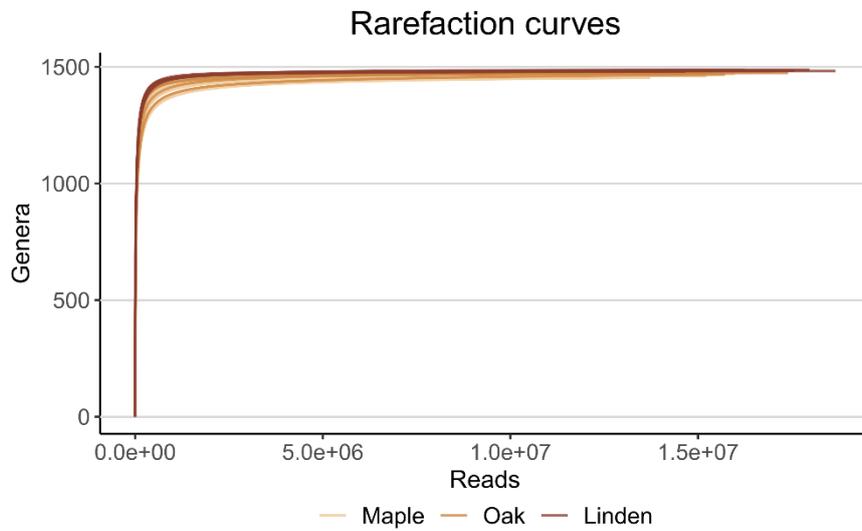
70. Bamforth SS. Protozoa from aboveground and ground soils of a tropical rain forest in Puerto Rico. *Pedobiologia* 2007;**50**:515–25. <https://doi.org/10.1016/j.pedobi.2006.10.009>
71. Gros V, Reinhard A, Petters S *et al.* Adding complexity to soil food webs: Myxobacteria have broad predation spectra with bacteria, yeasts and filamentous fungi *in vitro*. *Eur J Soil Biol* 2023;**117**:103508. <https://doi.org/10.1016/j.ejsobi.2023.103508>
72. Petters S, Gros V, Sollinger A *et al.* The soil microbial food web revisited: predatory myxobacteria as keystone taxa? *ISME J* 2021;**15**:2665–75. <https://doi.org/10.1038/s41396-021-00958-2>
73. Yeates GW, Bongers T, Degoede RGM *et al.* Feeding habits in soil nematode families and genera - an outline for soil ecologists. *J Nematology* 1993;**25**: 315-331. <https://pubmed.ncbi.nlm.nih.gov/19279775/>
74. Guidetti R, Cesari M, Giovannini I *et al.* Morphology and taxonomy of the genus *Ramazzottius* (Eutardigrada; Ramazzottiidae) with the integrative description of *Ramazzottius kretschmanni* sp. nov. *Eur Zool J* 2022;**89**:346–70. <https://doi.org/10.1080/24750263.2022.2043468>
75. Muggia L, Leavitt S, Barreno E. The hidden diversity of lichenised Trebouxiophyceae (Chlorophyta). *Phycologia* 2018;**57**:503–24. <https://doi.org/10.2216/17-134.1>
76. Aptroot A. Changes in the epiphytic flora on four *tilia* trees in Belgium over 59 years. *Herz* 2010;**25**:39–45. <https://doi.org/10.13158/heia.25.1.2010.39>
77. Jürriado I, Liira J, Paal J *et al.* Tree and stand level variables influencing diversity of lichens on temperate broad-leaved trees in boreo-nemoral floodplain forests. *Biodivers Conserv* 2009;**18**:105–25. <https://doi.org/10.1007/s10531-008-9460-y>
78. Dal Grande F, Rolshausen G, Divakar PK *et al.* Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol* 2018;**217**:277–89. <https://doi.org/10.1111/nph.14770>
79. Moya P, Molins A, Martinez-Alberola F *et al.* Unexpected associated microalgal diversity in the lichen *Ramalina farinacea* is uncovered by pyrosequencing analyses. *PLoS One* 2017;**12**:e0175091. <https://doi.org/10.1371/journal.pone.0175091>
80. Yahr R, Vilgalys R, DePriest PT. Geographic variation in algal partners of *Cladonia subtenuis* (Cladoniaceae) highlights the dynamic nature of a lichen symbiosis. *New Phytol* 2006;**171**:847–60. <https://doi.org/10.1111/j.1469-8137.2006.01792.x>
81. Hom EFY, Murray AW. Niche engineering demonstrates a latent capacity for fungal-algal mutualism. *Science* 2014;**345**:94–8. <https://doi.org/10.1126/science.1253320>
82. Krespach MKC, Garcia-Altare M, Flak M *et al.* Lichen-like association of *Chlamydomonas reinhardtii* and *aspergillus nidulans* protects algal cells from bacteria. *ISME J* 2020;**14**:2794–805. <https://doi.org/10.1038/s41396-020-0731-2>
83. Amin SA, Hmelo LR, van Tol HM *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015;**522**:98–101. <https://doi.org/10.1038/nature14488>
84. Grossart H-P, Simon M. Interactions of planktonic algae and bacteria: effects on algal growth and organic matter dynamics. *Aquat Microb Ecol* 2007;**47**:163–76. <https://doi.org/10.3354/ame047163>
85. Honegger R. The lichen symbiosis—what is so spectacular about it? *Lichenologist* 1998;**30**:193–212. <https://doi.org/10.1006/lich.1998.0140>
86. Uma R, Sivasubramanian V, Niranjali DS. Preliminary phycochemical analysis and *in vitro* antibacterial screening of green micro algae, *Desmococcus olivaceus*, *Chlorococcum humicola* and *Chlorella vulgaris*. *J Algal Biomass Utiln* 2011;**2**:74–81.
87. Boukhatem ZF, Merabet C, Tsaki H. Plant growth promoting Actinobacteria, the most promising candidates as bioinoculants? *Front Agron* 2022;**4**:849911. <https://doi.org/10.3389/fagro.2022.849911>
88. Palaniyandi SA, Yang SH, Zhang L *et al.* Effects of actinobacteria on plant disease suppression and growth promotion. *Appl Microbiol Biotechnol* 2013;**97**:9621–36. <https://doi.org/10.1007/s00253-013-5206-1>
89. Jauss R-T, Walden S, Fiore-Donno AM *et al.* A parasite’s paradise: biotrophic species prevail oomycete community composition in tree canopies. *Front For Glob Change* 2021;**4**:668895. <https://doi.org/10.3389/ffgc.2021.668895>
90. Esteban GF, Fenchel TM, Esteban GF *et al.* Bioenergetics. In: Esteban GF, Fenchel TM (eds). *Ecology of Protozoa: The Biology of Free-Living Phagotrophic Protists*. 2nd Edition. Springer International Publishing, Cham, 2020. pp. 55–62, [https://doi.org/10.1007/978-3-030-59979-9\\_5](https://doi.org/10.1007/978-3-030-59979-9_5).
91. Dumack K, Pundt J, Bonkowski M. Food choice experiments indicate selective fungivorous predation in *Fisculla terrestris* (Thecofilosea, Cercozoa). *J Eukaryot Microbiol* 2019;**66**:525–7. <https://doi.org/10.1111/jeu.12680>

## Chapter 5

---

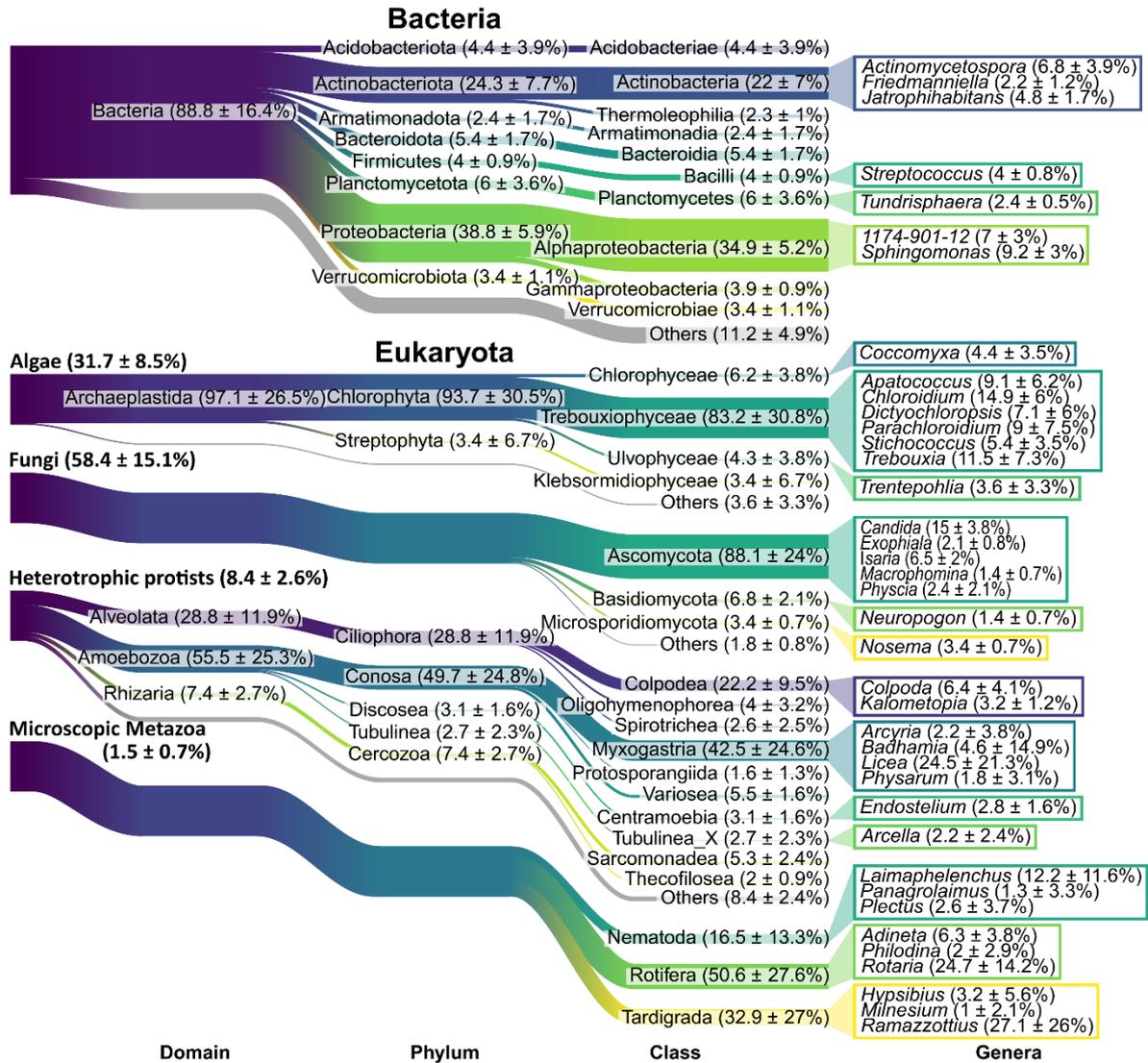
92. Dumack K, Feng K, Flues S *et al.* What drives the assembly of plant-associated protist microbiomes? Investigating the effects of crop species, soil type and bacterial microbiomes. *Protist* 2022;**173**:125913. <https://doi.org/10.1016/j.protis.2022.125913>
93. Estermann AH, Teixeira Pereira Bassiaridis J, Loos A *et al.* Fungivorous protists in the rhizosphere of *Arabidopsis thaliana* – diversity, functions, and publicly available cultures for experimental exploration. *Soil Biol Biochem* 2023;**187**:109206. <https://doi.org/10.1016/j.soilbio.2023.109206>
94. Spier L, van Dobben H, van Dort K. Is bark pH more important than tree species in determining the composition of nitrophytic or acidophytic lichen floras? *Environ Pollut* 2010;**158**:3607–11. <https://doi.org/10.1016/j.envpol.2010.08.008>
95. Romani M, Carrion C, Fernandez F *et al.* High bacterial diversity in pioneer biofilms colonizing ceramic roof tiles. *Int Biodeterior Biodegradation* 2019;**144**:104745. <https://doi.org/10.1016/j.ibiod.2019.104745>
96. Freystein K, Salisch M, Reisser W. Algal biofilms on tree bark to monitor airborne pollutants. *Biol* 2008;**63**:866–72. <https://doi.org/10.2478/s11756-008-0114-z>
97. Krieg NR, Parte A, Ludwig W *et al.* (eds) *Bergey's manual of systematic bacteriology: Volume 4 the Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes.* 2nd edn. New York, NY: Springer New York, 2010.
98. Fiore-Donno AM, Berney C, Pawlowski J *et al.* Higher-order phylogeny of plasmodial slime molds (Myxogastria) based on elongation factor 1-a and small subunit rRNA gene sequences. *J Eukaryot Microbiol* 2005;**52**:201–10. <https://doi.org/10.1111/j.1550-7408.2005.00032.x>

### Supplementary material



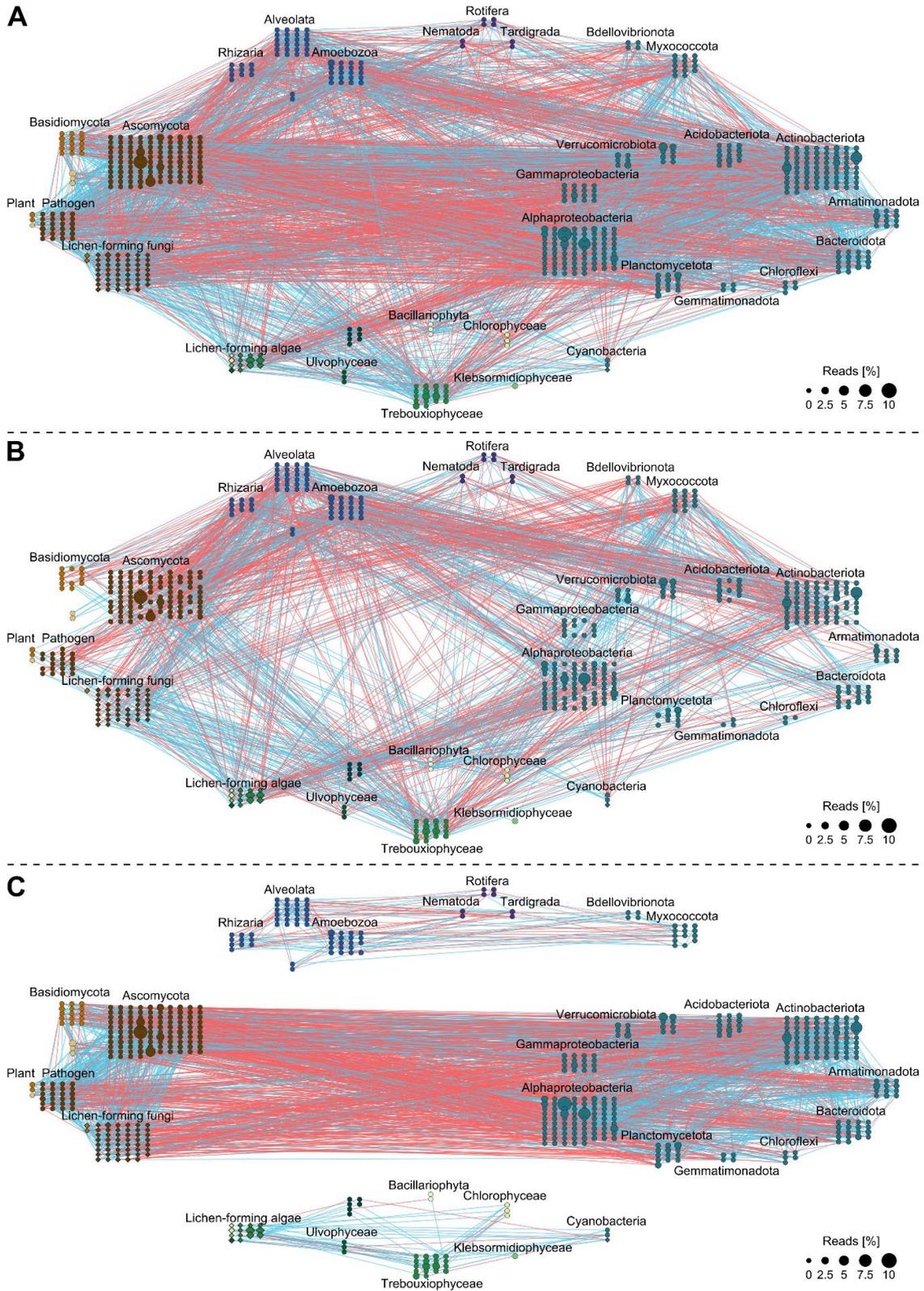
**Supplementary Figure 1: Rarefaction curves.** Rarefaction curves show the number of reads in relation to the number of genera (N=15 samples). All samples reached complete saturation. The samples are color-coded by the tree species.

## Chapter 5



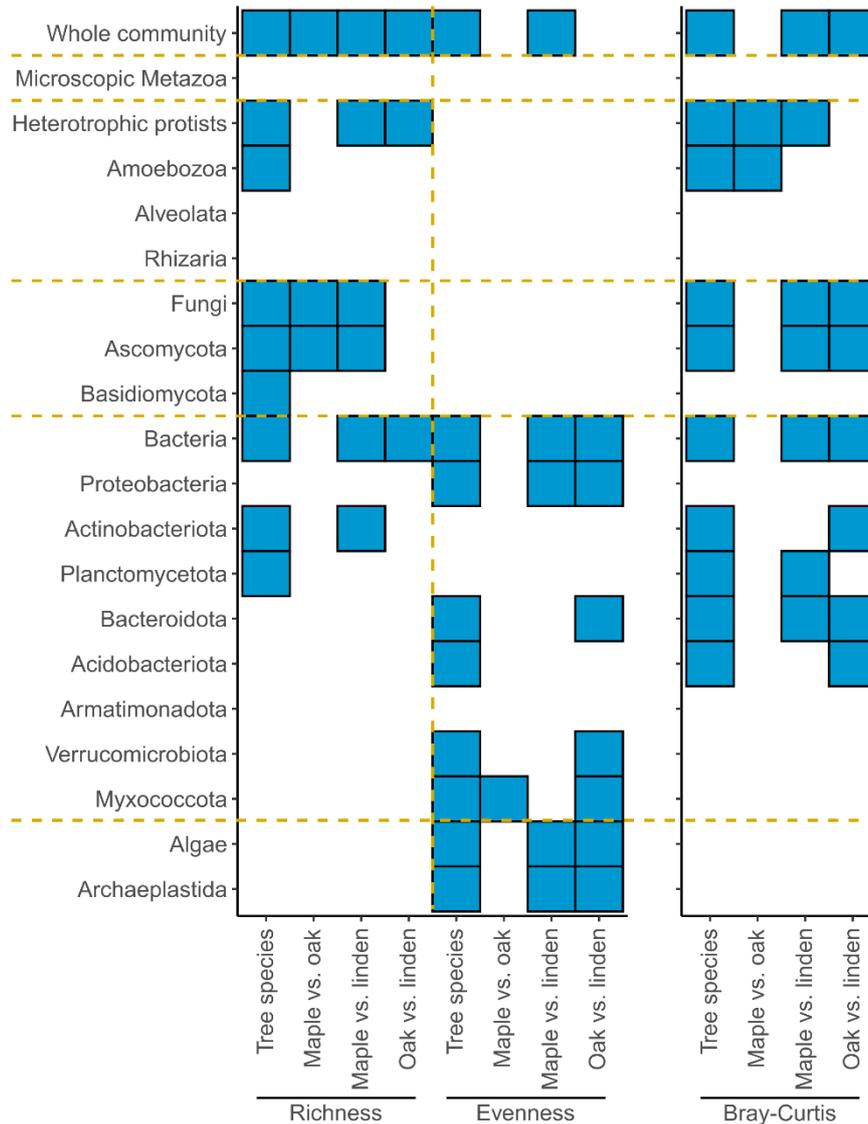
**Supplementary Figure 2: Microbial community composition in tree canopies.** Sankey diagrams show the mean relative abundance (percent) and standard deviation across all tree species (N=15) of the 10 most abundant classes of bacteria, algae, fungi, heterotrophic protists, and microscopic metazoa that accounted for more than 1% of the respective community. Additionally, the 10 most abundant genera that were classified to genus level are shown for each community, respectively.

# Chapter 5



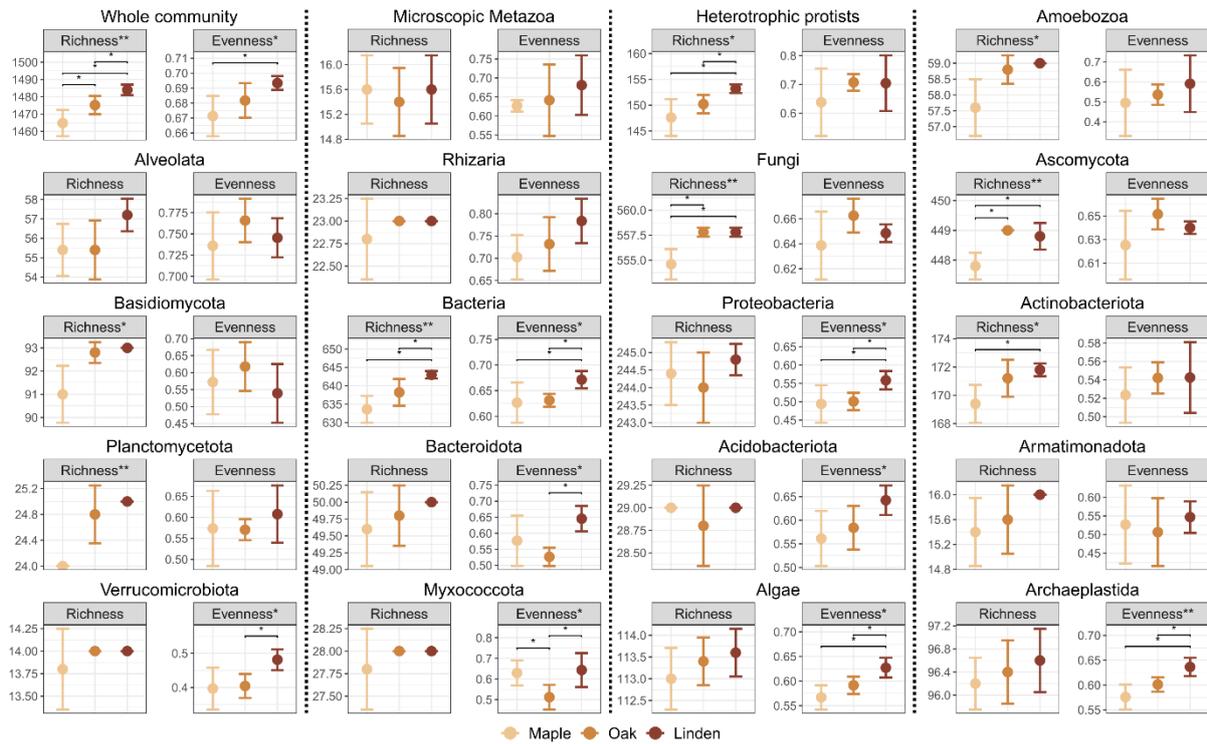
**Supplementary Figure 3: Co-occurrence networks of the microbial community in tree canopies.** The co-occurrence networks show associations between genera found on the bark surfaces in tree canopies (N=15). The networks are arranged by different trophic levels with primary producers at the bottom, decomposers in the middle, and consumers at the top. Associations are shown between all genera (A), between the trophic levels (B), and within the trophic levels (C). The node size is proportional to the relative number of reads for bacteria and eukaryotes respectively. Edges represent negative (red) and positive (blue) associations between the genera.

## Chapter 5



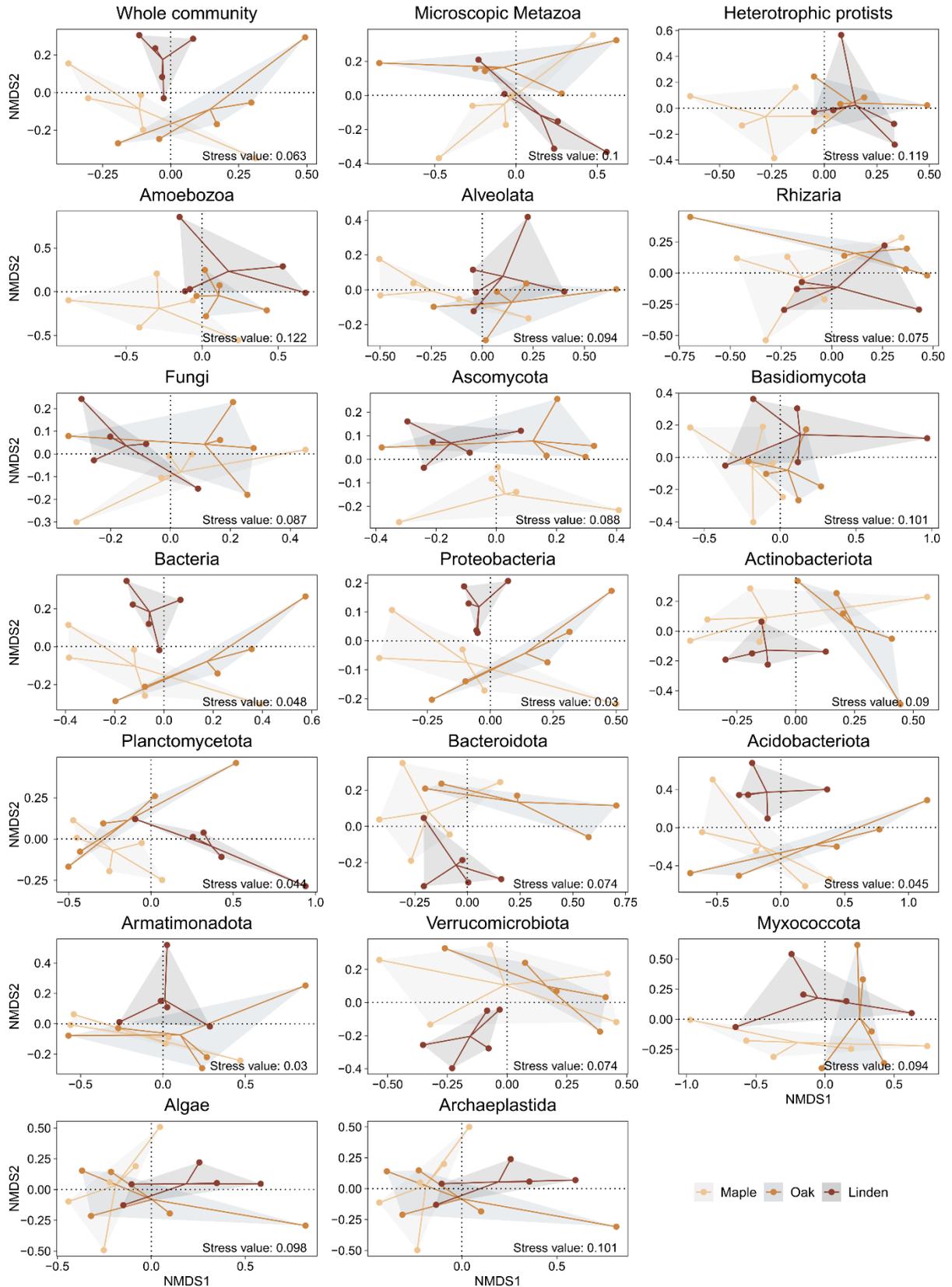
**Supplementary Figure 4: Overview of significant differences of tree species-specific alpha and beta diversity indices.** The heat map shows significant differences in genus richness and Pielou's evenness as well as Bray-Curtis dissimilarities across all tree species (first column respectively, alpha diversity=Kruskal-Wallis test, beta diversity=perMANOVA) as well as differences in pairwise comparisons of tree species (alpha diversity=Wilcoxon test, beta diversity= Pairwise perMANOVA) for selected taxonomic groups. Blue colored areas indicate significant differences ( $P$  value < 0.05).

# Chapter 5



**Supplementary Figure 5: Overview of tree species-specific alpha diversity indices.** Dot plots show the genus richness and Pielou's evenness per tree species for selected taxonomic groups. The dots represent the mean value, the error bars the standard deviation. Significant differences across all tree species (Kruskal-Wallis test) are marked in the graph title significant differences for pairwise comparisons of tree species (Wilcoxon test) are marked in the graph with stars (\*  $P$  value < 0.05; \*\*  $P$  value < 0.01; \*\*\*  $P$  value < 0.001).

# Chapter 5



**Supplementary Figure 6: Overview of tree species-specific beta diversity.** Non-metric multidimensional scalings (NMDS) of Bray-Curtis dissimilarity show the multivariate dispersion of the samples for selected taxonomic groups. Lines are color-coded by tree species and link samples of each tree species to their centromere.

## Chapter 5

**Supplementary Table 1: Overview of the number of reads per processing step.** For each sample, the number of reads after quality filtering, the number of reads successfully assembled into contigs, and the average contig length are shown for each sample. In addition, the number of hits for the PR2 and Silva databases are shown.

Tree	Replicate	Number of reads		Number of contigs	Mean length of contigs	Database matches	
		Total	Filtered			PR <sup>2</sup>	SILVA
Maple	1	62'066'095	54'782'009	34'267'056	218	6'857'979	10'859'015
Maple	2	61'182'230	54'834'717	30'029'067	230	6'617'627	9'902'565
Maple	3	57'642'915	51'623'620	26'500'849	230	7'510'904	7'786'301
Maple	4	68'321'722	60'934'697	31'625'283	232	7'256'503	9'636'821
Maple	5	67'760'636	60'176'238	32'602'613	227	6'160'160	11'345'074
Oak	1	63'551'723	57'268'788	33'384'234	228	9'073'024	9'934'790
Oak	2	65'635'053	58'974'935	32'298'611	229	5'224'063	12'335'655
Oak	3	64'953'704	58'186'565	34'978'135	226	9'861'618	9'500'957
Oak	4	61'439'467	54'765'861	30'547'978	228	8'545'337	8'572'851
Oak	5	66'993'556	60'139'618	27'541'468	236	7'182'410	8'003'008
Linden	1	80'623'991	72'008'768	37'192'288	233	5'520'558	13'818'189
Linden	2	63'944'549	57'229'574	29'761'371	232	6'919'860	9'201'105
Linden	3	70'370'573	63'065'424	37'658'841	225	5'462'431	13'574'846
Linden	4	75'151'907	66'997'536	38'528'179	229	8'390'810	12'738'239
Linden	5	70'830'597	63'392'067	35'355'216	229	6'384'769	12'035'556

---

## CHAPTER 6

### A novel protistan trait database reveals functional redundancy and complementarity in terrestrial protists (Amoebozoa & Rhizaria)

Jule Freudenthal<sup>a</sup>, Martin Schlegel<sup>b,c</sup>, Michael Bonkowski<sup>a</sup> and Kenneth Dumack<sup>a</sup>

<sup>a</sup>Terrestrial Ecology, Institute of Zoology, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zùlpicher Str. 47b, 50674 Köln, Germany

<sup>b</sup>Biodiversity and Evolution, Institute of Biology, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany

<sup>c</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle Jena Leipzig, Puschstraße 4, 04103 Leipzig, Germany

*Keywords: ecological traits, functional trait analysis, microbial ecology, functional trait reference database, soil, predator-prey*

#### Abstract

The inclusion of functional traits of protists in environmental sequencing surveys, in addition to the traditional taxonomic framework, is essential for a better understanding of their roles and impacts on ecosystem processes. We provide a database of functional traits for a widespread and important clade of protists – the Amoebozoa – based on extensive literature research in eight trait categories: Habitat, locomotion, nutrition, morphology, morphotype, size, spore formation, and disease-relatedness. The comparison of community traits of the Amoebozoa with sympatric but highly divergent Cercozoa (Rhizaria) revealed both convergent evolution of morphology or locomotion and distinct differences in habitat preference and feeding selectively. Amoebozoa seem to be rather unselective in their prey choice compared to Cercozoa. Indeed, the feeding preferences of Amoebozoa appeared to be related to cell size, whereas Cercozoa selectively feed on prey. Applications to metatranscriptomic data from soil, litter, and bark surfaces revealed differences in the average community trait compositions and ecosystem functioning, such as an increased proportion of disease-related Amoebozoa in soil or different proportions of nutrition types of Amoebozoa and Cercozoa on bark. This database will facilitate ecological analyses of sequencing data and improve our understanding of the diversity of adaptations of Amoebozoa to the environment and their functional roles in ecosystems.

---

## 1 Introduction

Refined information on the functional diversity of organisms, in addition to the traditional taxonomic framework, may greatly improve our knowledge of their function in ecosystem processes (Bouskill et al., 2012; Krause et al., 2014), but also, for example, how abiotic and biotic drivers shape communities (Briones, 2014; Fiore-Donno et al., 2019). To meet the analytical demands of environmental sequencing projects, trait-based data must be collated and tools developed to easily assign functional traits to existing sequencing databases.

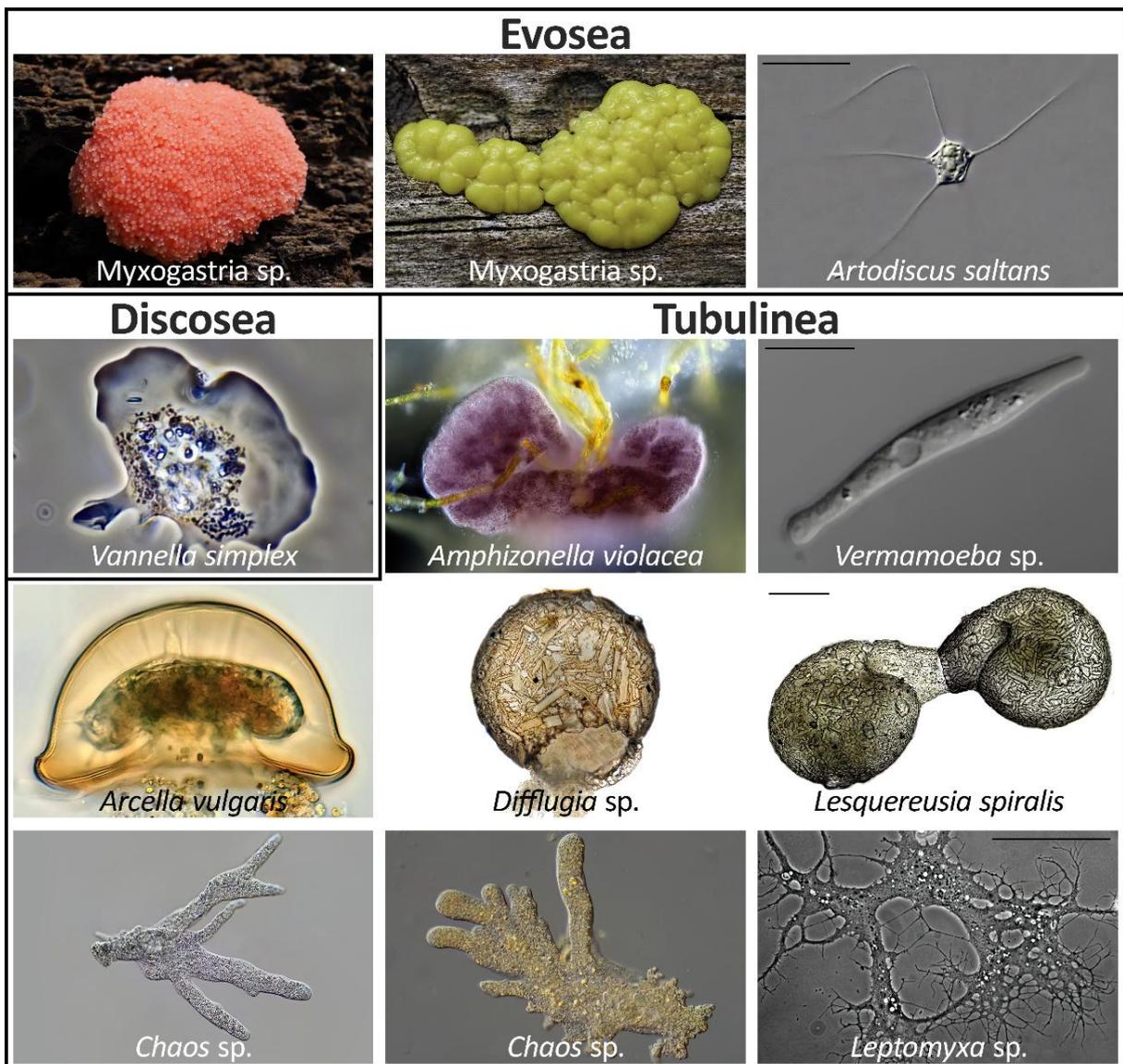
Trait-based community analyses aim to link species diversity to ecosystem functioning (Lavorel and Garnier, 2002; Violle et al., 2007). Traits, on the one hand, determine the performance and fitness of an organism (response traits) by directly reflecting its adaptations to physical, chemical, and biotic environmental drivers. On the other hand, traits such as feeding mode capture their potential impact (effect traits) on the environment and, thus, species' contributions to ecosystem functioning (Krause et al., 2014; Suding et al., 2008). Accordingly, trait-based community analyses may provide detailed information on the niche space occupied by communities (Lennon et al., 2012) or covered by a taxonomic group (Díaz et al., 2016) but also allow for an upscaling of ecosystem processes (Mulder et al., 2013).

Trait-based surveys are widely established for plants, animals, and prokaryotes (Beier et al., 2022; Bouskill et al., 2012; Louca et al., 2016). However, a sound functional understanding of the super-diverse communities of microbial eukaryotes is challenging, as their over 20 phyla comprise a multitude of completely independent evolutionary trajectories (Ruggiero et al., 2015). Furthermore, taxonomic and functional diversity are generally not necessarily coupled (Louca et al., 2016). For example, closely related taxa may exhibit different predatory impacts (Glücksman et al., 2010). Accordingly, studies covering a broad range of the diversity of microbial eukaryotes may so far provide only limited information on their functions (Aslani et al., 2022; Giachello et al., 2023; Köninger et al., 2023).

Over the past decade, there has been an increase in trait-based environmental surveys focusing on the lesser-investigated part of the microbial diversity, the protists (Amacker et al., 2022; Fiore-Donno et al., 2019; Flues et al., 2017; Jauss et al., 2021; Lamentowicz et al., 2020). The vast majority of microbial eukaryotic diversity is represented by protists, a paraphyletic assemblage of mostly unicellular eukaryotes. Protists fulfill numerous functions in terrestrial and aquatic ecosystems, like primary production and the exertion of distinct predation patterns, but some taxa are also important parasites of plants and animals. It is crucial to include functional traits to understand their roles in ecosystem processes. For example, the metabolic basis of protistan functional traits has been used to identify the main drivers of the shift between net heterotrophy and autotrophy in the oceans and to establish models predicting phytoplankton blooms (Alexander et al., 2015). Moreover, the importance of

symbioses among the planktonic eukaryotes was only revealed after compiling the planktonic Protist Interaction DAtabase (PIDA, Bjorbækmo et al., 2020).

In soils, Amoebozoa and Cercozoa (Rhizaria) are the most dominant terrestrial protistan supergroups (Domonell et al., 2013; Dumack et al., 2016; Fiore-Donno et al., 2024; Urich et al., 2008; Voss et al., 2019). A detailed trait database exists for the Cercozoa and Endomyxa (Dumack et al., 2020). This trait database allows an easy assignment of traits to environmental sequences and thus enables functional insights into the structure of microbial food webs. Fiore-Donno et al. (2019) showed a relative increase in the abundance of shell-bearing Cercozoa with drier soils, supporting the long-assumed function of their shells, i.e., increased drought resistance due to reduced evaporation.



**Figure 1: Overview of the morphological diversity of the Amoebozoa.** The graphic shows (1) the Evosea, represented by *Myxomycetes* sp. and *Artodiscus saltans* (Conosea), (2) the Discosea, represented by *Vannella simplex* (Flabellinia), and (3) the Tubulinea, represented by *Amphizonella violacea* (Corycida), *Vermamoeba* sp. (Echinamoebida), *Arcella vulgaris*, *Diffflugia* sp., *Lesquereusia spiralis* (Arcellinida), *Chaos* sp. (Euamoebida), and *Leptomyxa* sp. (Leptomyxida).

---

The supergroup Amoebozoa is equally diverse (Fig. 1) and abundant as Cercozoa and comprises three major lineages (Kang et al., 2017; Tekle et al., 2022): First, the Evosea, some of which with flagellated cells and complex life cycles, some are giant such as the plasmodia of Myxomycetes (clade: Cutosea and Conosea). Second, the Discosea, comprising the Flabellinia with flattened cells with separate hyaloplasm of which (lobose) subpseudopodia protrude (orders: Stygamoebida, Thecamoeboda, Dactylopodida, Vannellida, Dermamoebida), and the Centramoebida, some of which with scales (orders: Acanthamoebida, Pellitida, Himatismenida). Third, the Tubulinea, comprising shell-bearing but mostly naked lobose amoebae, with highly variable cell sizes ranging from 20  $\mu\text{m}$  to several centimeters, taxa with larger cells often with branching (ramose) or network-forming (reticulose) pseudopodia (orders: Leptomyxida, Euamoebida, Arcellinida, Echinamoebida, Corycida). Moreover, the Amoebozoa include important human parasites like *Acanthamoeba* spp., *Balamuthia mandrillaris*, and *Entamoeba histolytica* (Fiore-Donno et al., 2016; Geisen et al., 2014; Tice et al., 2016; Walochnik, 2018). Traditional approaches early on identified the significance of Amoebozoa in soil systems (Azam et al., 1983; Clarholm, 1985; de Ruyter et al., 1995). Unfortunately, the widespread use of “general eukaryotic” primers in metabarcoding studies led to a consistent and dramatic underrepresentation of amoebozoan sequences in surveys of terrestrial eukaryote diversity. Metatranscriptomics does not suffer from these extreme primer biases and led to sequencing results largely concordant with microscopic surveys illustrating the dominance of Amoebozoa (Fiore-Donno et al., 2024; Freudenthal et al., 2022; Heck et al., 2023; Voss et al., 2019). Now, as the molecular methodology to assess the taxonomic richness and diversity of Amoebozoa is established, a trait database is highly needed to understand the diversity of their functional roles in terrestrial and aquatic communities.

Here, we provide a trait database for Amoebozoa to serve as a common reference and to facilitate functional ecological studies. We showcase the usage of amoebozoan traits on recently published metatranscriptomic data of soil, leaf litter, and bark surfaces, and we compare the traits of the two most dominant soil protistan supergroups – Amoebozoa and Cercozoa.

## 2 Materials and Methods

### *2.1 Intentional use of the trait database as a justification of its structure and content*

As a baseline for our literature research, we screened the curated diversity of amoebozoan 18S rDNA sequences in the PR<sup>2</sup> database v. 5.0.0 (Guillou et al., 2012) for all currently included amoebozoan genera. We attributed traits by means of the literature, including original descriptions, or accessing other available meta-analyses and already collated data. All consulted references are provided in the database. We attempted to be as exhaustive as possible in selecting functional traits. However, given the vast divergence within Amoebozoa, the trait database is still a strongly simplified representation

---

of their functional diversity. Even though the database may not fulfill the expectations of taxonomists, it is primarily meant to facilitate functional analyses of sequencing data, especially by (microbial) ecologists who are not necessarily experts in protistan diversity. Accordingly, it was crucial to provide traits in discrete categories that can be easily subjected to statistical analyses, i.e., each trait for any taxon can only be assigned once in each category. This is especially problematic for Amoebozoa with complex life cycles (Keller et al., 2022; Tice et al., 2016). Consequently, this database is a simplified approximation for amoebozoan traits and contains a sum of compromises to increase its practical application.

## *2.2 Justification of genus-level trait assignment*

We consider the genus level to be most suitable for assigning functional traits, as most traits (e.g., nutrition, locomotion, morphotype) in protists are conserved at the genus level (Dumack et al., 2020). In addition, sequences in reference databases are typically not assigned to species, as short reads in environmental sequencing data often do not allow for reliable taxonomic assignment at the species level. However, some traits, particularly cell size, may differ considerably between taxa, even between species within one genus, or between the different stages in the life cycle (Berney et al., 2015; Kylin, 2001). Therefore, instead of recording size as a continuous variable, we assigned a fixed (common) size range to each genus, which, however, needs to be considered with care, as variability can be large. Accordingly, comments and references are given for each genus. Moreover, although amoebozoans are phylogenetically more divergent to Cercozoa than animals to plants, we tried to keep the traits most comparable to the already published Cercozoa database (Dumack et al., 2020) but accounted also for traits intrinsic to Amoebozoa (for instance spore formation which is absent in Cercozoa).

## *2.3 Functional traits*

We considered the organisms' prey range, rough morphology, and morphotype, locomotion, known habitat preference, animal disease-relatedness (whether as vector or immediate parasite), presence/absence of spore formation, and size range. Prey range categories were grouped according to bacterivory, omnivory (feeding on bacteria and eukaryotes), eukaryvory (feeding on fungi, microfauna, algae, or other protists), and saprotrophy. We could not assign more precise categories (e.g., fungivory, algivory) for a lack of information (or contradictory reports, i.e., likely multiple trophic modes) for most taxa. Morphology was mainly specified by the presence/absence of a shell and flagella. As amoebozoan amoebae, although variable, show well-recognizable shapes (Smirnov and Brown, 2004), we further defined simplified morphotypes, i.e., disc, tubule, palm, and reticulate. Two main locomotion modes were recognized: organisms bound to the substrate, i.e., gliding or freely

---

swimming. However, amoebae, amoeboflagellates, and flagellates differ not only in their locomotion but amoeboid cells are surface feeders, whereas prey capture of flagellates likely is much more selective due to their larger handling time. We considered habitat preferences of soil and freshwater taxa and marine taxa because soil-inhabiting and freshwater taxa may easily switch habitats, while marine taxa are rarely found in terrestrial or limnic habitats (Smirnov and Brown, 2004). Genera accommodating species from marine and soil and freshwater habitats were considered to have evolved ubiquitous habitat preferences. Spore formation is an important trait to enhance dispersal. For simplicity, we did not consider different spore formation strategies. Suggestions for updates can be addressed to the corresponding author. An R package for the easy assignment of the traits with updated versions of the database will be available at <https://github.com/JFreude/Functional-Traits-Amoebozoa>.

## *2.4 Statistics*

The statistical data analyses were conducted in R v. 4.3.1 (R Core Team, 2023). The data were visualized with the R packages *ggplot2* v. 3.5.1 and *ggpubr* v. 0.6.0 (Kassambara, 2023a; Wickham, 2011). An overview of the relative genus richness per functional traits within each category of the Amoebozoa database was given by a Sankey diagram. Only taxa for which a trait could be assigned to the respective category were considered for the relative genus richness. To explore potential convergent evolution between terrestrial Amoebozoa and Cercozoa, we also included data for Cercozoa and Endomyxa (Rhizaria) and their respective functional trait database (Dumack et al., 2020). For convenience, we will refer to Cercozoa and Endomyxa as Cercozoa.

For comparing the sizes of bacterivorous, eukaryvorous, and omnivorous taxa, the given size was used, or the mean size was calculated if a size range was given. Specifications such as "up to" or "larger than" were not considered, taxa with a size of "up to macroscopic" were regarded as 1000  $\mu\text{m}$  in size. The sizes across feeding types were compared using a Kruskal-Wallis test and Dunn's post-hoc test (*rstatix* v. 0.7.2::*kruskal\_test* and *rstatix* v. 0.7.2::*dunn\_test* (Kassambara, 2023b). Pairwise comparisons were corrected for multiple testing according to Benjamini & Hochberg (1995).

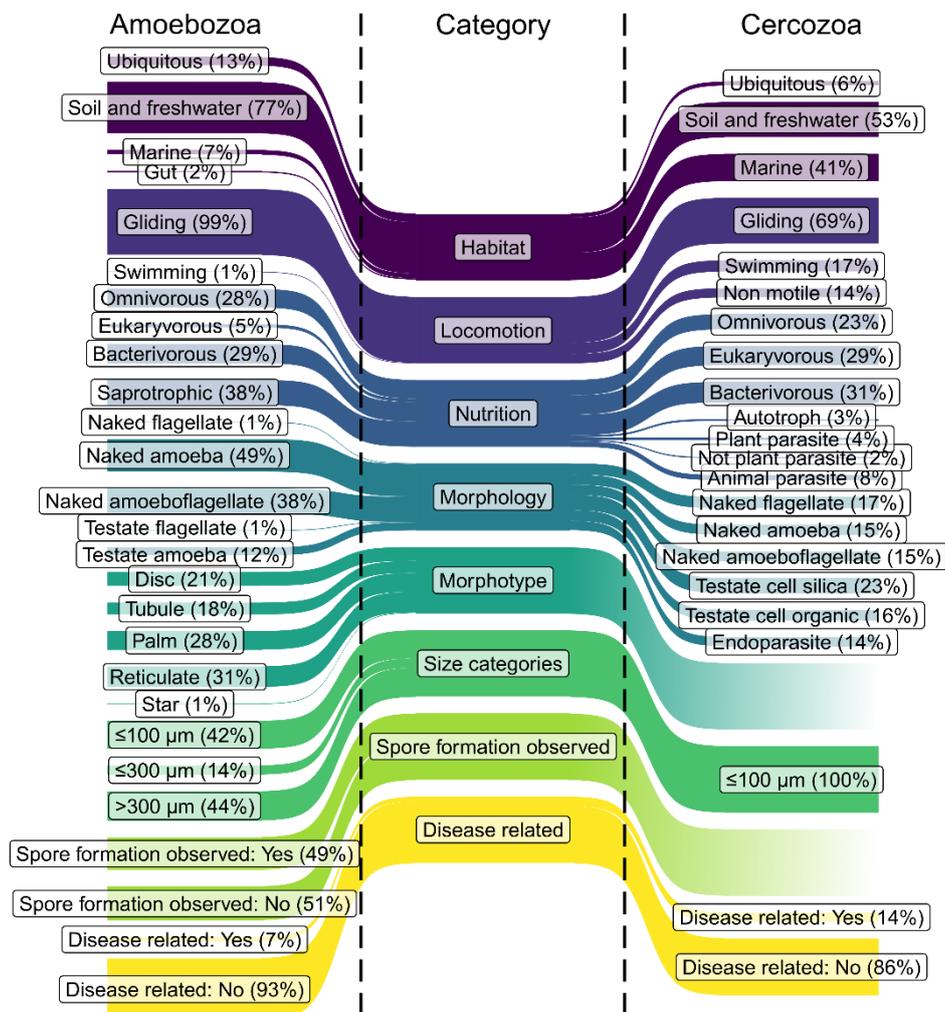
To showcase the usage of the trait database, we visualized the variations in functional traits of Amoebozoa and Cercozoa communities across different habitats. We used publicly available metatranscriptomic datasets from bark (Freudenthal et al., 2024), litter (Voss et al., 2019), and soil (Fiore-Donno et al., 2024). From the latter, we only used samples that were collected in the summer (see Fiore-Donno et al., 2024). After assigning the traits, the mean and standard deviation of the relative community composition for each trait category were calculated and visualized in a point

diagram for each habitat and community (Amoebozoa and Cercozoa), respectively. The proportion of taxa with missing trait information is not shown.

### 3 Results and Discussion

We provide a functional database for Amoebozoa allowing to easily add ecological meaning to molecular studies. The database comprises functional traits, i.e., habitat, locomotion, nutrition morphology, and size (Fig. 2). Additionally, we included information on morphotype, if spore formation was observed, or if they may cause diseases. All sources consulted for the functional trait database are provided in the supplementary Table 1.

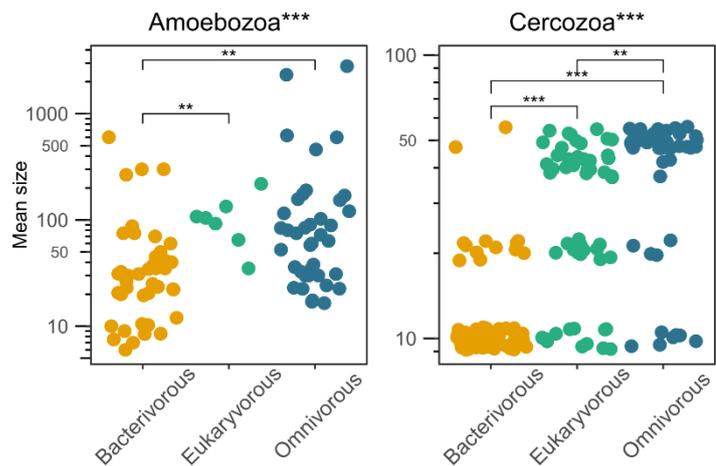
A comparison of the genus richness per trait of Amoebozoa with the Cercozoa revealed striking similarities but also distinct differences (Fig. 2). Both taxa show convergent evolution, i.e., both groups show similar morphological variation as they include shell-bearing amoebae, naked amoebae, and



**Figure 2: Overview of the relative genus richness per functional trait within each category of the Amoebozoa and Cercozoa databases.** The sankey diagrams show the percentual genus richness calculated for the given traits of each category for the Amoebozoa (left) and Cercozoa (right) databases.

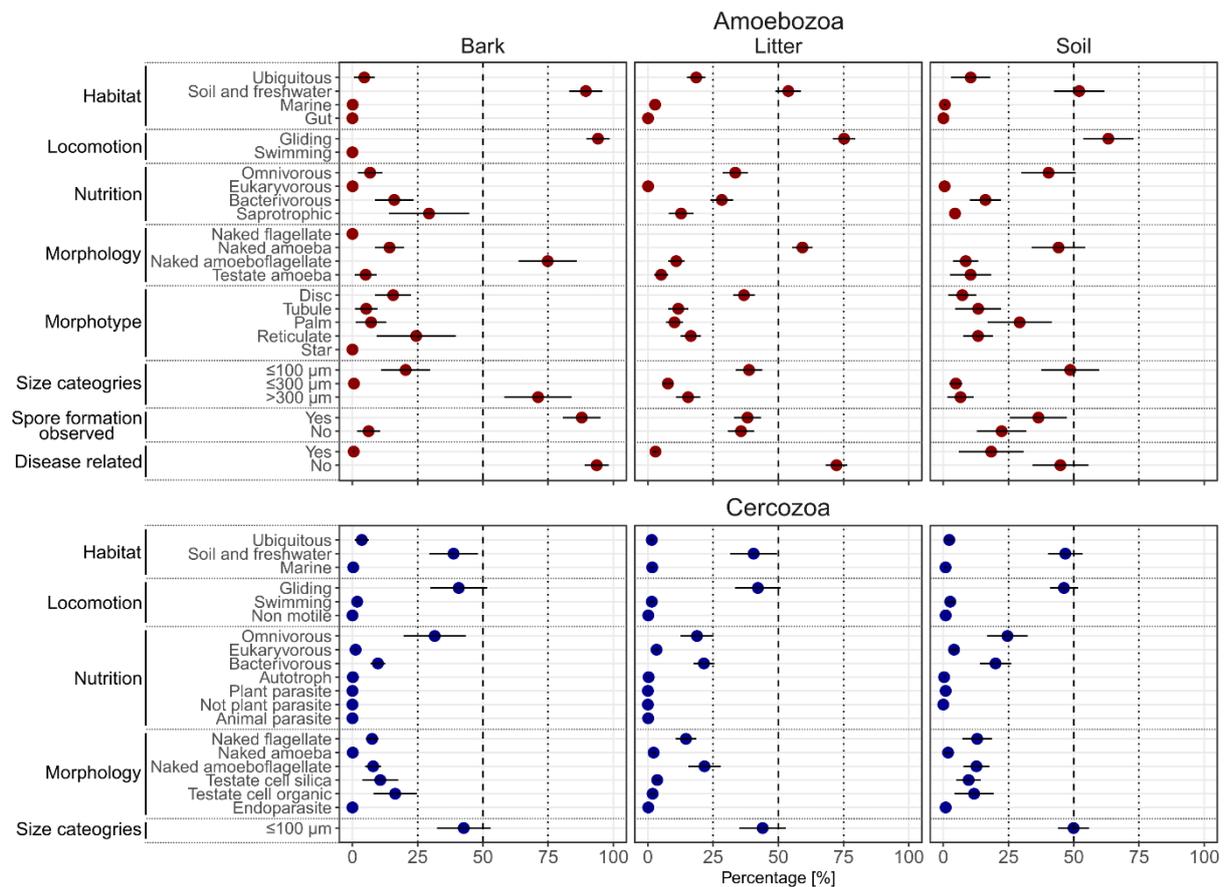
flagellated taxa. Furthermore, the majority of species are gliding, an adaptation to surface feeding in soil habitats. However, next to similarities, our database revealed striking differences. For example, almost 80 % of the known amoebozoan genera occur in soil or freshwater and only a small fraction in marine environments, whereas for Cercozoa, the ratio of soil and freshwater to marine genera is nearly balanced.

The mean size of Amobozoa and Cercozoa was associated with the



**Figure 3: Overview of the association between size and feeding type for Amoebzoa and Cercozoa.** The point diagrams show the mean sizes of bacterivorous (yellow), eukaryvorous (green) and omnivorous (blue) for the Amoebzoa (left) and Cercozoa (right) databases. Significant differences across all feeding types (Kruskal-Wallis test) and of pairwise comparisons of the feeding types (Dunn's test) are indicated with stars in the graph title or the graph, respectively (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

feeding type, i.e., bacterivorous taxa were significantly smaller compared to eukaryvorous and omnivorous taxa (Fig. 3). Traditionally, most protists were considered to be bacterivorous. In recent years, however, it has become increasingly clear that many protists indeed exhibit a broad prey spectrum, including algae, fungi, and other heterotrophic protists (Dumack et al., 2019b; Estermann et al., 2023; Geisen et al., 2016; Seppey et al., 2017). For testate amoebae, it has been shown that feeding type can be determined by shell sizes, although with certain limitations (Dumack et al., 2024; Fournier et al., 2015). We now show that this also applies much more broadly to the entirety of Amoboza and Cercozoa. Moreover, Amoebzoa comprise only a very limited number of eukaryvorous taxa, in particular in contrast to Cercozoa. While small Amoebzoa ingest only bacteria, larger ones consume bacteria and single-celled and multicellular eukaryotes., such as nematodes and fungi (Geisen et al., 2015). In other words, if the prey item can be entirely enclosed by an amoebozoan cell, it is suitable prey. Thus, the Amoebzoa are likely much less selective in their prey choice than Cercozoa, but the size of the amoebozoan cell determines which prey can be ingested (Kulishkin et al., 2023). However, few reliable data on feeding preferences exist for Amoebzoa, and more feeding experiments are urgently needed, predominantly in taxa where individual species may differ by several orders of magnitude in cell size, for example, in Variosea (Berney et al., 2015). Nonetheless, aside this generalized pattern that we found, there are some highly specialized consumers among Amoebzoa (Dumack et al., 2024, 2019a; Estermann et al., 2023; Smirnov et al., 2011), such as *Phryganella paradoxa* (Arcellinida) which feeds on pennate diatoms by bending or breaking their frustules (Dumack et al., 2024). However, these specialized consumers appear to be more exceptions than the rule.



**Figure 4: The relative proportion of functional traits of Amoebzoa and Cercozoa communities of bark, litter and soil.** The point diagrams show the percentages that were assigned to the given traits of each category for Amoebzoa and Cercozoa communities of bark (N=15), litter (N=18) and soil (N=39), respectively. The points represent the mean and are color-coded by Amoebzoa (red) and Cercozoa (blue). The error bars represent the standard deviation.

To illustrate the assignment and analyses of the Amoebzoa and Cercozoa traits, we applied the databases on recently published metatranscriptomic data sets of soil, leaf litter, and bark surfaces (Fig. 4). These datasets originate from different locations and years, providing independent inventories for the comparison of traits among communities. For example, the proportion of disease-related amoebzoa was higher for soil (~20 %) than for bark and litter. Furthermore, spore-forming (~90%) and saprotrophic (~30%) Amoebzoa were exceptionally dominant on bark. Communities of Amoebzoa and terrestrial Cercozoa differed in trait composition: Omnivorous Cercozoa were most dominant on bark, while omnivorous Amoebzoa were the least abundant. In addition, the variation of functional traits between bark, litter, and soil was much lower in Cercozoa communities than within Amoebzoa.

The newly provided trait database for Amoebzoa enables an easy assignment of traits to environment sequencing surveys in order to detect trade-offs and evolutionary trajectories in adaptations among different supergroups in protists and to deepen our knowledge of the functional diversity of Amoebzoa and their impact on ecosystem functioning.

---

## Data Availability

The trait database and an R package for the automatic assignment of the Traits to a taxonomy table are available upon request and will be publicly available by the end of the project at the latest at GitHub: <https://github.com/JFreude/FunctionalTraitsAmoebozoa>.

## Competing interests

The authors declare that they have no conflict of interest.

## Funding

This study was supported by grants from the German Research Foundation (DFG) in the framework of the priority program SPP 1991: TAXON-OMICS (Project IDs 447013012 and 221301018).

## Acknowledgments

We would like to thank Ferry Siemensma for providing the amazing pictures of the Amoebozoa.

---

## References

- Alexander, H., Rouco, M., Haley, S.T., Wilson, S.T., Karl, D.M., Dyhrman, S.T., 2015. Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc Natl Acad Sci* 112, E5972–E5979. <https://doi.org/10.1073/pnas.1518165112>
- Amacker, N., Gao, Z., Hu, J., Jousset, A.L.C., Kowalchuk, G.A., Geisen, S., 2022. Protist feeding patterns and growth rate are related to their predatory impacts on soil bacterial communities. *FEMS Microbiol Ecol* 98, fiac057. <https://doi.org/10.1093/femsec/fiac057>
- Aslani, F., Geisen, S., Ning, D., Tedersoo, L., Bahram, M., 2022. Towards revealing the global diversity and community assembly of soil eukaryotes. *Ecol Lett* 25, 65–76. <https://doi.org/10.1111/ele.13904>
- Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., Thingstad, F., 1983. The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* 10, 257–263. <https://doi.org/10.3354/meps010257>
- Beier, S., Werner, J., Bouvier, T., Mouquet, N., Violle, C., 2022. Trait-trait relationships and tradeoffs vary with genome size in prokaryotes. *Front Microbiol* 13. <https://doi.org/10.3389/fmicb.2022.985216>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berney, C., Geisen, S., Van Wichelen, J., Nitsche, F., Vanormelingen, P., Bonkowski, M., Bass, D., 2015. Expansion of the ‘Reticulosphere’: Diversity of novel branching and network-forming amoebae helps to define Varioseae (Amoebozoa). *Protist* 166, 271–295. <https://doi.org/10.1016/j.protis.2015.04.001>
- Bjorbækmo, M.F.M., Evenstad, A., Røsæg, L.L., Krabberød, A.K., Logares, R., 2020. The planktonic protist interactome: where do we stand after a century of research? *ISME J* 14, 544–559. <https://doi.org/10.1038/s41396-019-0542-5>
- Bouskill, N.J., Eveillard, D., Chien, D., Jayakumar, A., Ward, B.B., 2012. Environmental factors determining ammonia-oxidizing organism distribution and diversity in marine environments. *Environ Microbiol* 14, 714–729. <https://doi.org/10.1111/j.1462-2920.2011.02623.x>
- Briones, M.J.I., 2014. Soil fauna and soil functions: A jigsaw puzzle. *Front Environ Sci* 2. <https://doi.org/10.3389/fenvs.2014.00007>
- Clarholm, M., 1985. Interactions of bacteria, protozoa and plants leading to mineralization of soil nitrogen. *Soil Biol Biochem* 17, 181–187. [https://doi.org/10.1016/0038-0717\(85\)90113-0](https://doi.org/10.1016/0038-0717(85)90113-0)
- de Ruiter, P.C., Neutel, A.-M., Moore, J.C., 1995. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269, 1257–1260. <https://doi.org/10.1126/science.269.5228.1257>
- Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Colin Prentice, I., Garnier, E., Bönsch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Chave, J., Joseph Wright, S., Sheremet’ev, S.N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M.D., Gorné, L.D., 2016. The global spectrum of plant form and function. *Nature* 529, 167–171. <https://doi.org/10.1038/nature16489>
- Domonell, A., Brabender, M., Nitsche, F., Bonkowski, M., Arndt, H., 2013. Community structure of cultivable protists in different grassland and forest soils of Thuringia. *Pedobiologia* 56, 1–7. <https://doi.org/10.1016/j.pedobi.2012.07.001>
- Dumack, K., Fiore-Donno, A.M., Bass, D., Bonkowski, M., 2020. Making sense of environmental sequencing data: Ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Mol Ecol Resour* 20, 398–403. <https://doi.org/10.1111/1755-0998.13112>
- Dumack, K., Kahlich, C., Lahr, D.J.G., Bonkowski, M., 2019a. Reinvestigation of *Phryganella paradoxa* (Arcellinida, Amoebozoa) Penard 1902. *J Eukaryot Microbiol* 66, 232–243. <https://doi.org/10.1111/jeu.12665>
- Dumack, K., Koller, R., Weber, B., Bonkowski, M., 2016. Estimated abundance and diversity of heterotrophic protists in South African biocrusts. *S Afr J Sci* 112, 5–5. <https://doi.org/10.17159/sajs.2016/20150302>
- Dumack, K., Lara, E., Duckert, C., Ermolaeva, E., Siemensma, F., Singer, D., Krashevskaya, V., Lamentowicz, M., Mitchell, E.A.D., 2024. It’s time to consider the Arcellinida shell as a weapon. *Eur J Protistol* 92, 126051. <https://doi.org/10.1016/j.ejop.2024.126051>
- Dumack, K., Pundt, J., Bonkowski, M., 2019b. Food choice experiments indicate selective fungivorous predation in *Fisculla terrestris* (Thecofilosea, Cercozoa). *J Eukaryot Microbiol* 66, 525–527. <https://doi.org/10.1111/jeu.12680>
- Estermann, A.H., Teixeira Pereira Bassiaridis, J., Loos, A., Solbach, M.D., Bonkowski, M., Hess, S., Dumack, K., 2023. Fungivorous protists in the rhizosphere of *Arabidopsis thaliana* – Diversity, functions, and publicly available cultures for experimental exploration. *Soil Biol Biochem* 187, 109206. <https://doi.org/10.1016/j.soilbio.2023.109206>

- 
- Fiore-Donno, A.M., Freudenthal, J., Dahl, M.B., Rixen, C., Urich, T., Bonkowski, M., 2024. Biotic interactions explain seasonal dynamics of the alpine soil microbiome. *ISME Commun* 4, ycae028. <https://doi.org/10.1093/ismeco/ycae028>
- Fiore-Donno, A.M., Richter-Heitmann, T., Degrune, F., Dumack, K., Regan, K.M., Marhan, S., Boeddinghaus, R.S., Rillig, M.C., Friedrich, M.W., Kandeler, E., Bonkowski, M., 2019. Functional traits and spatio-temporal structure of a major group of soil protists (Rhizaria: Cercozoa) in a temperate grassland. *Front Microbiol* 10, 1332. <https://doi.org/10.3389/fmicb.2019.01332>
- Fiore-Donno, A.M., Weinert, J., Wubet, T., Bonkowski, M., 2016. Metacommunity analysis of amoeboid protists in grassland soils. *Sci Rep* 6, 19068. <https://doi.org/10.1038/srep19068>
- Flues, S., Bass, D., Bonkowski, M., 2017. Grazing of leaf-associated Cercomonads (Protists: Rhizaria: Cercozoa) structures bacterial community composition and function. *Environ Microbiol* 19, 3297–3309. <https://doi.org/10.1111/1462-2920.13824>
- Fournier, B., Lara, E., Jassey, V.E., Mitchell, E.A., 2015. Functional traits as a new approach for interpreting testate amoeba palaeo-records in peatlands and assessing the causes and consequences of past changes in species composition. *Holocene* 25, 1375–1383. <https://doi.org/10.1177/0959683615585842>
- Freudenthal, J., Dumack, K., Schaffer, S., Schlegel, M., Bonkowski, M., 2024. Algae-fungi symbioses and bacteria-fungi co-exclusion drive tree species-specific differences in canopy bark microbiomes. *ISME J* wrae206. <https://doi.org/10.1093/ismejo/wrae206>
- Freudenthal, J., Ju, F., Bürgmann, H., Dumack, K., 2022. Microeukaryotic gut parasites in wastewater treatment plants: Diversity, activity, and removal. *Microbiome* 10, 27. <https://doi.org/10.1186/s40168-022-01225-y>
- Geisen, S., Fiore-Donno, A.M., Walochnik, J., Bonkowski, M., 2014. *Acanthamoeba* everywhere: high diversity of *Acanthamoeba* in soils. *Parasitol Res* 113, 3151–3158. <https://doi.org/10.1007/s00436-014-3976-8>
- Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T., Bonkowski, M., 2016. The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biol Biochem* 94, 10–18. <https://doi.org/10.1016/j.soilbio.2015.11.010>
- Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., Bonkowski, M., 2015. Pack hunting by a common soil amoeba on nematodes. *Environ Microbiol* 17, 4538–4546. <https://doi.org/10.1111/1462-2920.12949>
- Giachello, S., Cantera, I., Carteron, A., Marta, S., Cipriano, C., Guerrieri, A., Bonin, A., Thuiller, W., Ficetola, G.F., 2023. Toward a common set of functional traits for soil protists. *Soil Biol Biochem* 187, 109207. <https://doi.org/10.1016/j.soilbio.2023.109207>
- Glücksman, E., Bell, T., Griffiths, R.I., Bass, D., 2010. Closely related protist strains have different grazing impacts on natural bacterial communities. *Environ Microbiol* 12, 3105–3113. <https://doi.org/10.1111/j.1462-2920.2010.02283.x>
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2012. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 41, D597–D604. <https://doi.org/10.1093/nar/gks1160>
- Heck, N., Freudenthal, J., Dumack, K., 2023. Microeukaryotic predators shape the wastewater microbiome. *Water Res* 242, 120293. <https://doi.org/10.1016/j.watres.2023.120293>
- Jauss, R.-T., Walden, S., Fiore-Donno, A.M., Schaffer, S., Wolf, R., Feng, K., Bonkowski, M., Schlegel, M., 2021. A parasite's paradise: Biotrophic species prevail oomycete community composition in tree canopies. *Front For Glob Change* 4, 668895. <https://doi.org/10.3389/ffgc.2021.668895>
- Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcântara, D.M.C., Roger, A.J., Shadwick, L.L., Smirnov, A., Kudryavtsev, A., Lahr, D.J.G., Brown, M.W., 2017. Between a Pod and a Hard Test: The Deep Evolution of Amoebae. *Mol Biol Evol* 34, 2258–2270. <https://doi.org/10.1093/molbev/msx162>
- Kassambara, A., 2023a. ggpubr: “ggplot2” Based Publication Ready Plots. R package version 0.6.0.
- Kassambara, A., 2023b. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2.
- Keller, H.W., Everhart, S.E., Kilgore, C.M., 2022. Chapter 1 - The myxomycetes: introduction, basic biology, life cycles, genetics, and reproduction, in: Rojas, C., Stephenson, S.L. (Eds.), *Myxomycetes: Biology, Systematics, Biogeography and Ecology*. Academic Press, pp. 1–45. <https://doi.org/10.1016/B978-0-12-824281-0.00003-8>
- Köninger, J., Ballabio, C., Panagos, P., Jones, A., Schmid, M.W., Orgiazzi, A., Briones, M.J.I., 2023. Ecosystem type drives soil eukaryotic diversity and composition in Europe. *Glob Change Biol* 29, 5706–5719. <https://doi.org/10.1111/gcb.16871>

- 
- Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S., Grossart, H.-P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* 5. <https://doi.org/10.3389/fmicb.2014.00251>
- Kulishkin, N.S., Surkova, A.A., Mesentsev, Y.S., Smirnov, A.V., 2023. Three ways to eat spaghetti: Amoebae from two Amoebozoa lineages are able to feed on cyanobacteria of the genus *Oscillatoria*. *Protistology* 17, 38–49. <https://doi.org/10.21685/1680-0826-2023-17-1-4>
- Kylin, H., 2001. Biodiversity in a slime mould: arthropods associated with *Brefeldia maxima*. *Mycologist* 15, 70–73. [https://doi.org/10.1016/S0269-915X\(01\)80084-7](https://doi.org/10.1016/S0269-915X(01)80084-7)
- Lamentowicz, M., Kajukało-Drygalska, K., Kołaczek, P., Jassey, V.E.J., Gąbka, M., Karpińska-Kołaczek, M., 2020. Testate amoebae taxonomy and trait diversity are coupled along an openness and wetness gradient in pine-dominated Baltic bogs. *Eur J Protistol* 73, 125674. <https://doi.org/10.1016/j.ejop.2020.125674>
- Lavorel, S., Garnier, E., 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct Ecol* 16, 545–556. <https://doi.org/10.1046/j.1365-2435.2002.00664.x>
- Lennon, J.T., Aanderud, Z.T., Lehmkuhl, B.K., Schoolmaster Jr., D.R., 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* 93, 1867–1879. <https://doi.org/10.1890/11-1745.1>
- Louca, S., Parfrey, L.W., Doebeli, M., 2016. Decoupling function and taxonomy in the global ocean microbiome. *Science* 353, 1272–1277. <https://doi.org/10.1126/science.aaf4507>
- Mulder, C., Ahrestani, F.S., Bahn, M., Bohan, D.A., Bonkowski, M., Griffiths, B.S., Guicharnaud, R.A., Kattge, J., Krogh, P.H., Lavorel, S., Lewis, O.T., Mancinelli, G., Naeem, S., Peñuelas, J., Poorter, H., Reich, P.B., Rossi, L., Rusch, G.M., Sardans, J., Wright, I.J., 2013. Connecting the green and brown worlds, in: Woodward, G., Bohan, D.A.B.T.-A. in E.R. (Eds.), *Advances in Ecological Research*. Academic Press, pp. 69–175. <https://doi.org/10.1016/B978-0-12-420002-9.00002-0>
- R Core Team, 2023. R: A language and environment for statistical computing.
- Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., Cavalier-Smith, T., Guiry, M.D., Kirk, P.M., 2015. A higher level classification of all living organisms. *PLOS ONE* 10, e0119248. <https://doi.org/10.1371/journal.pone.0119248>
- Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., Lara, E., 2017. Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. *Soil Biol Biochem* 112, 68–76. <https://doi.org/10.1016/j.soilbio.2017.05.002>
- Smirnov, A., Brown, S., 2004. Guide to the methods of study and identification of soil gymnamoebae. *Protistology* 3, 148–190.
- Smirnov, A.V., Bedjagina, O.M., Goodkov, A.V., 2011. *Dermamoeba algensis* n. sp. (Amoebozoa, Dermamoebidae) – An algivorous lobose amoeba with complex cell coat and unusual feeding mode. *Eur J Protistol* 47, 67–78. <https://doi.org/10.1016/j.ejop.2010.12.002>
- Suding, K.N., Lavorel, S., Chapin Iii, F.S., Cornelissen, J.H.C., Díaz, S., Garnier, E., Goldberg, D., Hooper, D.U., Jackson, S.T., Navas, M.-L., 2008. Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. *Glob Change Biol* 14, 1125–1140. <https://doi.org/10.1111/j.1365-2486.2008.01557.x>
- Tekle, Y.I., Wang, F., Wood, F.C., Anderson, O.R., Smirnov, A., 2022. New insights on the evolutionary relationships between the major lineages of Amoebozoa. *Sci Rep* 12, 11173. <https://doi.org/10.1038/s41598-022-15372-7>
- Tice, A.K., Shadwick, L.L., Fiore-Donno, A.M., Geisen, S., Kang, S., Schuler, G.A., Spiegel, F.W., Wilkinson, K.A., Bonkowski, M., Dumack, K., Lahr, D.J.G., Voelcker, E., Clauß, S., Zhang, J., Brown, M.W., 2016. Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. *Biol Direct* 11, 69. <https://doi.org/10.1186/s13062-016-0171-0>
- Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLOS ONE* 3, e2527. <https://doi.org/10.1371/journal.pone.0002527>
- Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional! *Oikos* 116, 882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x>
- Voss, C., Fiore-Donno, A.M., Guerreiro, M.A., Peršoh, D., Bonkowski, M., 2019. Metatranscriptomics reveals unsuspected protistan diversity in leaf litter across temperate beech forests, with Amoebozoa the dominating lineage. *FEMS Microbiol Ecol* 95, fiz142. <https://doi.org/10.1093/femsec/fiz142>

- 
- Walochnik, J., 2018. Amoebae, in: Florin-Christensen, M., Schnittger, L. (Eds.), *Parasitic Protozoa of Farm Animals and Pets*. Springer International Publishing, Cham, pp. 389–412. [https://doi.org/10.1007/978-3-319-70132-5\\_15](https://doi.org/10.1007/978-3-319-70132-5_15)
- Wickham, H., 2011. ggplot2. *Wiley Interdiscip Rev Comput Stat* 3, 180–185. <https://doi.org/10.1002/wics.147>

---

## CHAPTER 7

### Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics

Jule Freudenthal<sup>a</sup>, Michael Bonkowski<sup>a</sup>, Nele-Joelle Engelmann<sup>a</sup>, Martin Schlegel<sup>b,c</sup> and Kenneth Dumack<sup>a</sup>

<sup>a</sup>Terrestrial Ecology, Institute of Zoology, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zùlpicher Str. 47b, 50674 Köln, Germany

<sup>b</sup>Biodiversity and Evolution, Institute of Biology, University of Leipzig, Talstraße 33, 04103 Leipzig, Germany

<sup>c</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle Jena Leipzig, Puschstraße 4, 04103 Leipzig, Germany

*Keywords: protists, functional traits, microbial ecology, multi-gene phylogeny, parasites, metabolism*

#### Abstract

Protists are vastly diverse, forming over 20 supergroups of the eukaryotic diversity and fulfilling plentiful functions. Rhizaria is a widespread and highly abundant supergroup comprising important parasites and a huge diversity of free-living heterotrophic predators. Despite the diversity and biogeochemical importance of Rhizaria, our understanding of their physiology and metabolic capabilities remains limited, mainly due to a general lack of data and bioinformatic tools for cross-species comparisons of physiological traits. In this study, we assembled a total of 15 transcriptomes of the parasite-related bacterivorous *Rhogostoma* and their eukaryovorous relatives. By phylogenomic analyses and whole transcriptome comparison, we create an evolutionary framework to which we relate physiological traits. The morphologically highly similar *Rhogostoma* strains branch in two distinct clusters differing in orthogroups and gene expression patterns related to cell adhesion and biofilm formation. Furthermore, we reveal considerable intra-genus variation in amino acid and lipid metabolism, which might be explained by an ancient streamlining through gradual specialization to parasitism, bearing the potential for subsequent metabolic radiation. We conclude that even closely related and morphologically similar species in Rhizaria may differ distinctly in their functional repertoire. With the here established and showcased analyses, we create a basis for future characterization of the physiological traits of microeukaryotes.

### 1 Introduction

The Rhizaria are among the least understood but most diverse microbial eukaryotes. Rhizaria exhibits a high morphological and ecological diversity, including (1) Foraminifera, Radiolaria, and Phaeodaria, which are important oceanic predators; (2) phototrophic taxa (algae), such as *Paulinella chromatophora* and Chlorarachniophytes, which are crucial for our understanding of the evolution of photosynthesis; and (3) numerous parasitic species, such as *Plasmodiophora brassicae*, which causes about 15% of cabbage yields losses worldwide (Cavalier-Smith et al., 2018; Dumack et al., 2020; Nakamura and Suzuki, 2015; Neuhauser et al., 2010; Nowack, 2014). In addition, Rhizaria includes a variety of free-living and heterotrophic species, many of which exhibit an enormous species richness and genetic diversity (Bass et al., 2009; Flues et al., 2018; Howe et al., 2009).

Thecofilosea is one of the most abundant Rhizaria phyla across ecosystems, for example, in soils and the polar oceans, as shown by global surveys (Oliverio et al., 2020; Sommeria-Klein et al., 2021). Thecofilosea comprise a remarkable ecological and morphological diversity, including bacterivores, eukaryvores, and specialized, parasite-like predators of algae (Dumack et al., 2020). Among the Thecofilosea, free-living and heterotrophic taxa of the genus *Rhogostoma* (Rhogostomidae) are most abundant in terrestrial ecosystems (Öztoprak et al., 2020; Walden et al., 2021). *Rhogostoma* viciously preys on bacteria, whereas most of its relatives feed on eukaryotes (Seppey et al., 2017). Furthermore, *Rhogostoma* exhibits remarkable 18S marker gene variation but a highly conserved morphology, indicating ongoing cryptic speciation with a hidden diversity that remains to be fully uncovered (Öztoprak et al., 2020).

Despite the widespread occurrence and high abundance of Rhizaria, and *Rhogostoma* in particular, we still lack a well-supported phylogeny of the Rhogostomidae as single- or few-gene phylogenies, in contrast to many other microbial taxa, do not resolve most interspecific phylogenetic relationships (Cavalier-Smith et al., 2018). Furthermore, the transcriptome representation of Rhizaria is scarce (Sibbald and Archibald, 2017), and their physiology and functional diversity remain to be explored. With modern high-throughput sequencing techniques and bioinformatics tools, it is now feasible to explore the physiological traits of individual species in an evolutionary context, which in turn sheds light on their potential ecological impact (Gerbracht et al., 2022; Ribeiro et al., 2020).

In this study, we explore the phylogenetic relationships and physiological traits of Thecofilosea. We present a compiled data set of 12 novel *Rhogostoma* transcriptomes, two Tectofilosida transcriptomes, and one Ebriida sp. transcriptome. Using a multi-gene phylogenomic approach, we resolve the phylogenetic backbone of Thecofilosea, in particular of *Rhogostoma*. Additionally, we perform whole transcriptome comparisons, along with functional annotation and enrichment analyses, to explore the inter- and intraspecific diversity in the physiological traits of the Thecofilosea in an evolutionary context.

## 2 Materials and Methods

### 2.1 RNA Extraction

We analyzed 11 *Rhogostoma* strains previously described and cultured by Öztoprak et al. (2020), Pohl et al. (2021), and Martínez Rendón et al. (2024), along with *Fisculla terrestris* (Solbach et al., 2021) and *Katarium polorum* (Solbach et al., 2024). Additionally, one *Rhogostoma* strain was isolated and cultured from a sample originating from the Leipzig floodplain forest (51.3657 N, 12.3094 E) in Germany by isolating single cells using sterile glass micropipettes and culturing them in cell culture flasks (SARSTEDT AG & Co. KG T25; Nümbrecht, Germany) with wheat grass (WG)-medium, at temperatures ranging from 4 to 21°C.

For each RNA extraction from cultures, monoclonal cells were detached from the bottles by vigorous shaking and/or thorough scraping with a sterile cell scraper. Subsequently, the medium was filtered with a filter pore size of 3 µm (cellulose nitrate membranes, Whatman™, Buckinghamshire, United Kingdom) until the filter was completely covered in cells. The filter was transferred to a 1.5 ml tube (SARSTEDT AG & Co. KG, Nümbrecht, Germany) containing 1 ml ice-cold Sørensen buffer. The tube was vigorously shaken and centrifuged at 1500 rpm for 5 minutes at 4°C to detach the cells from the filter. The filter was carefully removed without disturbing the pellet, followed by an additional centrifugation step for 2 minutes to firm the pellet. The Sørensen buffer was replaced with 1 ml of clean Sørensen buffer and the tube was centrifuged again for 5 min. Finally, the Sørensen buffer was discarded, and 170 µl of ice-cold RLN buffer was added.

RNA extraction was carried out using the RNeasy® Plant Mini Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's instructions starting from step 2 and using only 300 µl ethanol in step 4. RNA concentrations were quantified using the Qubit RNA High Sensitivity Assay Kit (Thermo Fisher Scientific Inc, Germany) and Qubit 4 Fluorometer (Thermo Fisher Scientific Inc, Germany). Sequencing was performed on an Illumina NovaSeq instrument (Illumina Inc., San Diego, CA, USA) at the Cologne Center for Genomics (Köln, Germany) with 2 × 100 bp paired-end reads, polyA selection, and a sequencing depth of about 50 Mio sequences (see Supplementary Table 1).

For single-cell RNA extraction, we followed the protocol of Hagemann-Jensen et al. (2020). Single cells of *Ebriida* sp. were isolated from samples originating from North Slope, Alaska (71.404558 N, 156.530021 W) and grown in F4 medium with nitschoid diatoms as prey. Before RNA extraction, cells were starved for 24 hours and given in the lysis buffer with a micromanipulator. Three replicates were pooled and sequenced with an Illumina NextSeq sequencer (Illumina Inc., San Diego, CA, USA) at the Cologne Center for Genomics (Köln, Germany) with 2 × 100 bp paired-end reads, polyA selection, and a sequencing depth of about 50 Mio sequences as well (see Supplementary Table 1).

### *2.2 Transcriptome assembly*

The quality of the 15 newly sequenced transcriptomes (*Rhogostoma*, *Fisculla terrestris*, *Katarium polorum*, and *Ebriida* sp.) as well as of the two Protaspidae sp. transcriptomes was assessed using FastQC v. 0.11.9 (Andrews, 2010), followed by RNA-seq error corrections with Rcorrector v 1.0.6 (Song and Florea, 2015) and quality filtering and adapter trimming with FastP v 0.23.2 (Chen et al., 2018). Potential contaminations from prokaryotes, plants, fungi, and humans were excluded using Kraken2 v 2.1.2 (Wood et al., 2019). Ribosomal RNA reads were identified using SortMeRNA v 4.3.4 (Kopylova et al., 2012) and blasted against the PR<sup>2</sup> database v. 4.14.0 (Guillou et al., 2013) using BLASTN v. 2.10.0 (Camacho et al., 2009) to confirm any *Rhogostoma*/Thecofilosea sequences. The messenger RNA reads were assembled using Trinity v 2.14.0 (Grabherr et al., 2011), and candidate coding regions were identified using TransDecoder v 5.5.0 (Haas, 2018). Subsequently, the RNA-Seq read representation of the Trinity transcripts was validated using Bowtie2 v 2.4.1 (Langmead et al., 2019; Langmead and Salzberg, 2012).

We extended our data set by including 28 rhizarian transcriptomes/genomes (Balzano et al., 2015; Burki et al., 2013; Gerbracht et al., 2022; Gomaa et al., 2021; Grant et al., 2012; Keeling et al., 2014; Lhee et al., 2021; Rodríguez-Ezpeleta et al., 2007; Schwelm et al., 2015; Sierra et al., 2016, see Supplementary Table 2). Candidate coding regions of nucleotide assemblies were identified using TransDecoder v 5.5.0 (Haas, 2018). The completeness of all assembled transcriptomes was evaluated with benchmarking universal single-copy orthologs (BUSCO) v 5.2.2 (Manni et al., 2021a, 2021b) and the eukaryote odb10 database.

### *2.3 Phylogenomic analyses*

PhyloFisher v. 1.2.13 (Tice et al., 2021) was employed to identify orthologs for the multi-gene phylogeny of the Rhizaria, based on the provided database comprising 240 genes from 304 taxa across the eukaryotic tree of life. To screen for paralogs and contaminants, single gene trees of all 2490 genes were revised manually and using a customized R script. The cleaned data sets were filtered for Rhizaria transcriptomes, considering only taxa with more than 65 % amino acid coverage across all 240 genes. Exceptions were the single-cell transcriptomes of Protaspidae sp. and *Ebriida* sp., which had overall low gene coverage but were still included in the phylogenomic analysis. Further, the 240 genes were filtered, keeping only genes present in more than 50% of the taxa. A concatenated alignment consisting of 222 genes spanning 40 Rhizaria strains was built using PhyloFisher v. 1.2.13 (Tice et al., 2021).

The maximum likelihood (ML) tree was constructed with IQ-Tree v. 2.1.2 (Minh et al., 2020) using the site-heterogeneous mixture model LG + C60 + F +  $\Gamma$  and 1000 Ultra-Fast Bootstrap replicates (UFB) as well as SH-like approximate likelihood ratio test (SH-aLRT).

### *2.4 Transcriptome annotation and comparison*

OrthoFinder v. 2.5.2 (Emms and Kelly, 2019, 2015) was employed to identify phylogenetic hierarchical orthogroups across all *Rhogostoma* strains, *Fisculla terrestris*, and *Katarium polorum*. Beforehand, the redundancy of the Trinity contigs was reduced by clustering at 95% identity over at least 90% of the shorter contig length using CD-HIT v. 4.8.1 (Fu et al., 2012). A customized R script was employed to filter the TransDecoder protein sequences by the clustered Trinity contigs and to split TransDecoder protein sequences with multiple predicted coding regions. In the case of overlapping coding regions, the longer one was retained. To verify the quality of the clustered contigs, we checked their completeness using BUSCO v 5.2.2 (Manni et al., 2021a, 2021b) and the eukaryotic odb10 database, and the RNA-Seq read mapping using Bowtie2 v 2.4.1 (Langmead et al., 2019; Langmead and Salzberg, 2012). Further, the gene expression was quantified with Salmon v 1.9.0 (Patro et al., 2017) in alignment-based mode. The protein sequences were functionally annotated based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, 2019; Kanehisa et al., 2023; Kanehisa and Goto, 2000) using eggNOG-mapper v. 2.1.9 (Cantalapiedra et al., 2021).

The number of shared and unique orthogroups across all strains was visualized using ggupset v. 0.4.0 (Ahlmann-Eltze, 2024). Only the top 25 combinations with the highest number of orthogroups were considered. In addition, a general overview of the relative proportions of functional annotations from 26 categories was provided by bar charts visualized with ggplot2 (Wickham, 2011). The relative proportions of functional annotations were calculated for all species, the core set of orthogroups (i.e., orthogroups shared by all strains), the two *Rhogostoma* clusters, *Katarium polorum* and *Fisculla terrestris*, respectively.

A detailed overview of the metabolic pathways of the amino acid, nucleotide, carbohydrate, and lipid metabolism was created based on the KEGG module. For this, the presence or absence of each KEGG orthology (KO) term was assessed per KEGG module, the minimum KO terms required for each module were determined, and the percentage of KO terms for each reaction step (the number of preset KO terms divided by the minimum number of KO terms required for the respective reaction) was calculated. Only KEGG modules for which (for at least one strain) 50% or more of the KO terms were present and no more than a total of three KO terms were missing were considered for the calculation of the heatmaps and the summarisation of the amino acid, central carbohydrate, and lipid metabolism modules into a graph. The graph was visualized using Cytoscape v. 3.9.0 (Shannon et al., 2003).

For the gene enrichment analysis of the two phylogenetically distinct *Rhogostoma* clusters multiple steps were conducted: First, the quantified gene expressions from salmon were summarized for each strain and orthogroup using tximport v. 1.32.0 (Soneson et al., 2016), generating matrices containing the weighted mean of the contig length, the effective contig length, the number of reads (counts) and

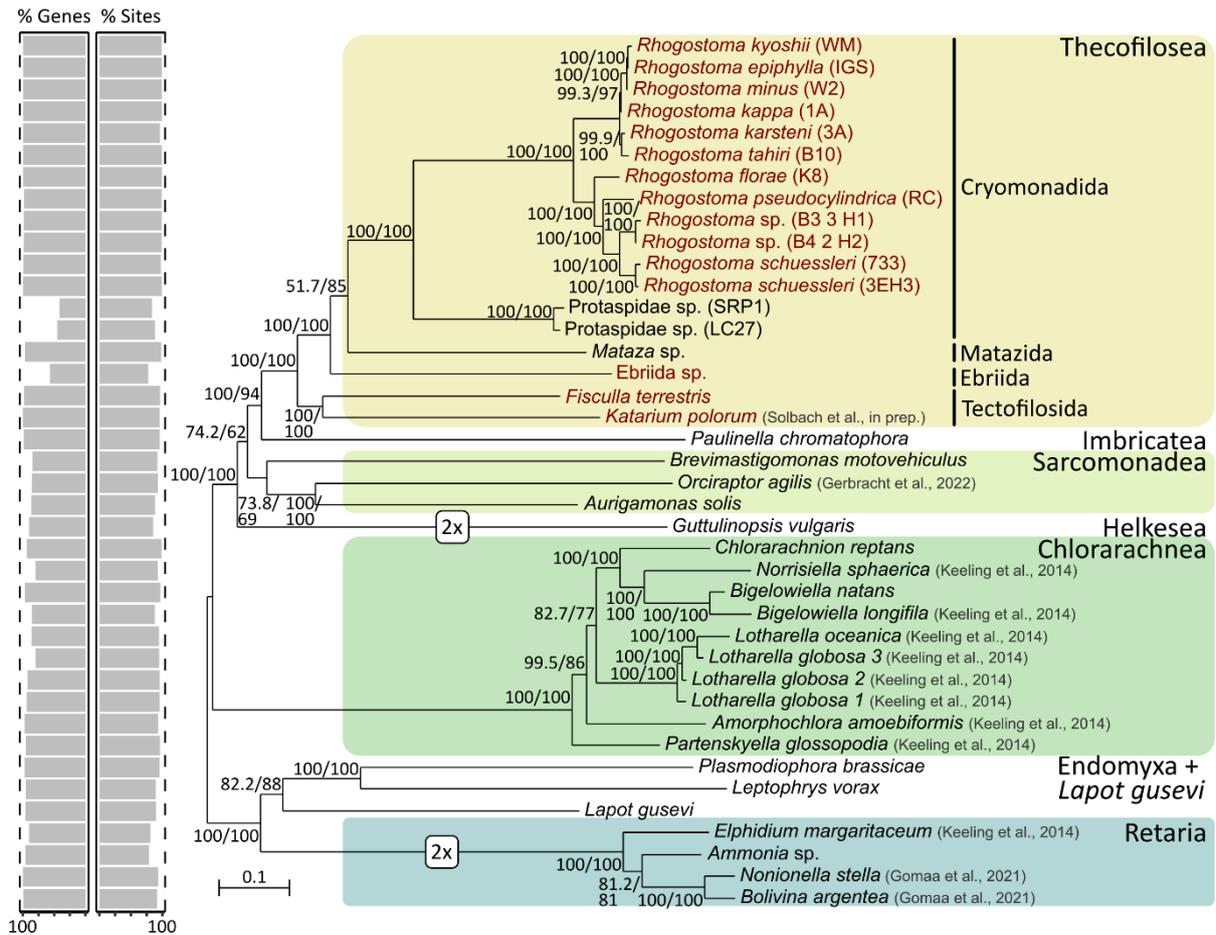
the transcript per million (TPM) for each orthogroup. Second, low count orthogroups were excluded, keeping only orthogroups with counts per million > 1 in six or more strains. Third, principal component analysis (PCA) based on the filtered expression level of the orthogroup was calculated using the functions `vst` and `plotPCA` (DESeq2 v. 1.44.0; (Love et al., 2014)). Fourth, differential expression analysis was conducted with DESeq2 v. 1.44.0 (Love et al., 2014) comparing the two *Rhogostoma* clusters with six strains each. Orthogroups with  $|\log_2 \text{fold change}| \geq 1$  and adjusted p-value < 0.001 were considered as differentially expressed. The enrichment analysis of KO terms was carried out using Goseq v. 1.56.0 (Young et al., 2010). The enrichment analysis was performed for all 12 *Rhogostoma* strains, using the length and the functional assignment of the orthogroups of each strain, respectively. Only KO terms that were significantly enriched in the enrichment analysis of all strains within one *Rhogostoma* cluster were retained.

### 3 Results

#### 3.1 Phylogenomic analysis

To shed light on the evolutionary relationships of Thecofilosea, we conducted a phylogenetic analysis based on a comprehensive multi-gene data set of 222 concatenated genes (71,479 amino acids) derived from, in total, 40 Rhizaria transcriptomes (Fig. 1, Supplementary Table 2, Supplementary Fig. 1 & 2). Among these transcriptomes are a total of 15 newly added Thecofilosea transcriptomes: 14 derived from monoclonal cultures with high gene and site coverage - 12 *Rhogostoma* (Cryomonadida) and two Tectofilosida transcriptomes, as well as one single-cell transcriptome of Ebriida sp. with moderate gene and site coverage, as it is typical for single-cell transcriptomes (Fig. 1, Supplementary Table 1).

The phylogenetic analyses provided a highly supported backbone of the Thecofilosea, Imbricatea, Sarcomonadea and Helkesea (Fig. 1). At the base of the Thecofilosea, *Fisculla terrestris* (Fiscullidae) and *Katarium polorum* (Chlamydropyridae) form a fully supported monophyletic group, the Tectofilosida. The Ebriida sp. branch with high support basal to Matazida. Within Cryomonadida, two distinct, fully supported monophyletic subclades were identified. The first subclade included all Protaspididae strains and clustered at the base of the second subclade, including all *Rhogostoma* strains. The *Rhogostoma* strains were further divided into two fully supported clusters, each comprising six strains: Cluster 1, characterized by short branches, included *R. kyoshii*, *R. epiphylla*, *R. minus*, *R. kappa* and *R. tahiri*, and cluster 2, with longer branches, included *R. florea*, *R. pseudocylindrica* and two strains each of *R. schuessleri* and unidentified *Rhogostoma* species.

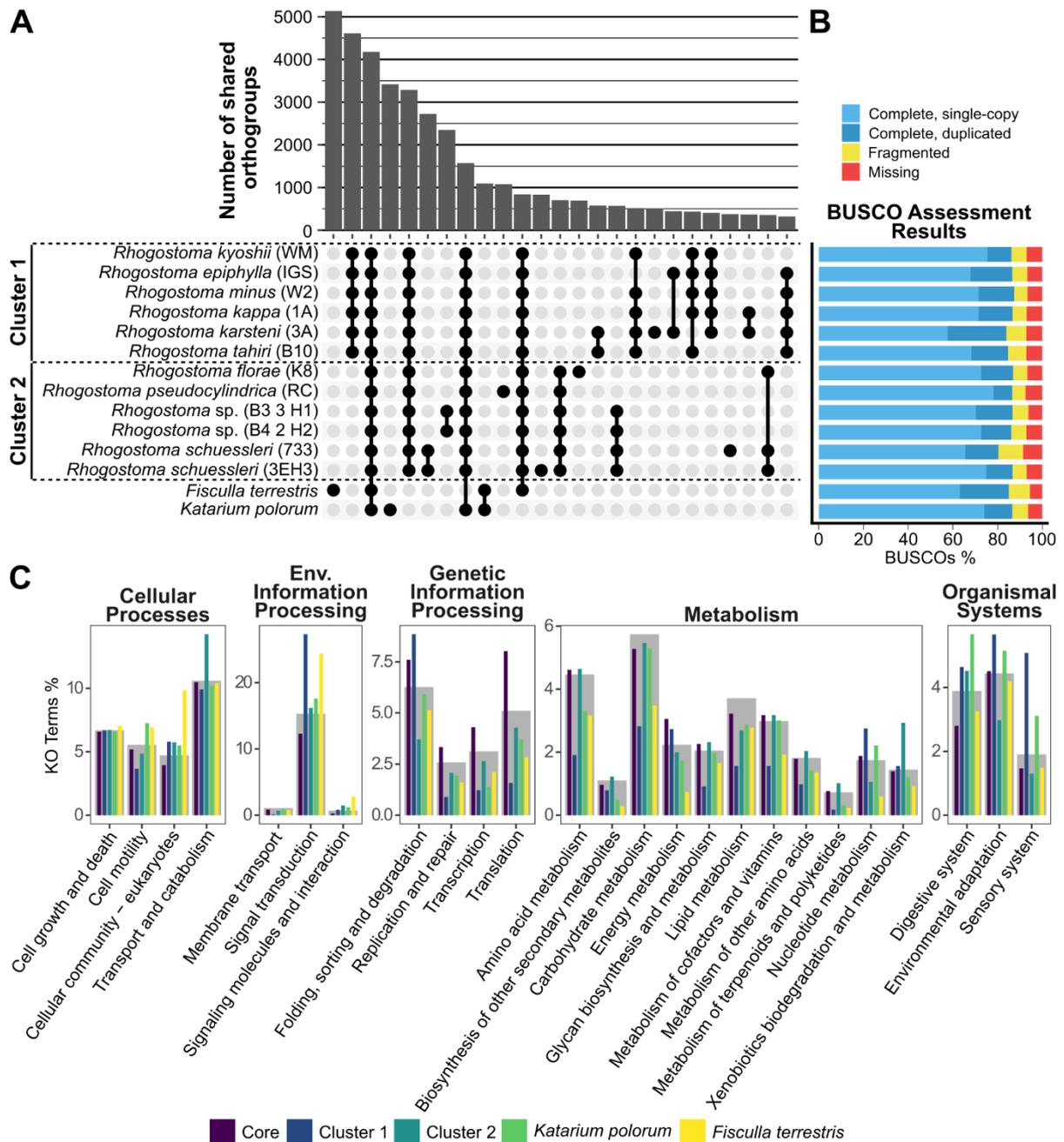


**Figure 1: Multi-gene phylogeny of the Rhizaria.** Maximum likelihood tree (LG+C60+F+G model) based on an alignment comprising 222 concatenated genes (71,479 amino acids) derived from 40 Rhizaria transcriptomes. Support values were obtained from Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) and 1000 ultrafast bootstraps (UFB). The percentage of gene coverage and the percentage of the site coverage of the present genes for each taxon are shown on the left. Newly added transcriptomes are highlighted in red.

### 3.2 Comparative transcriptomics

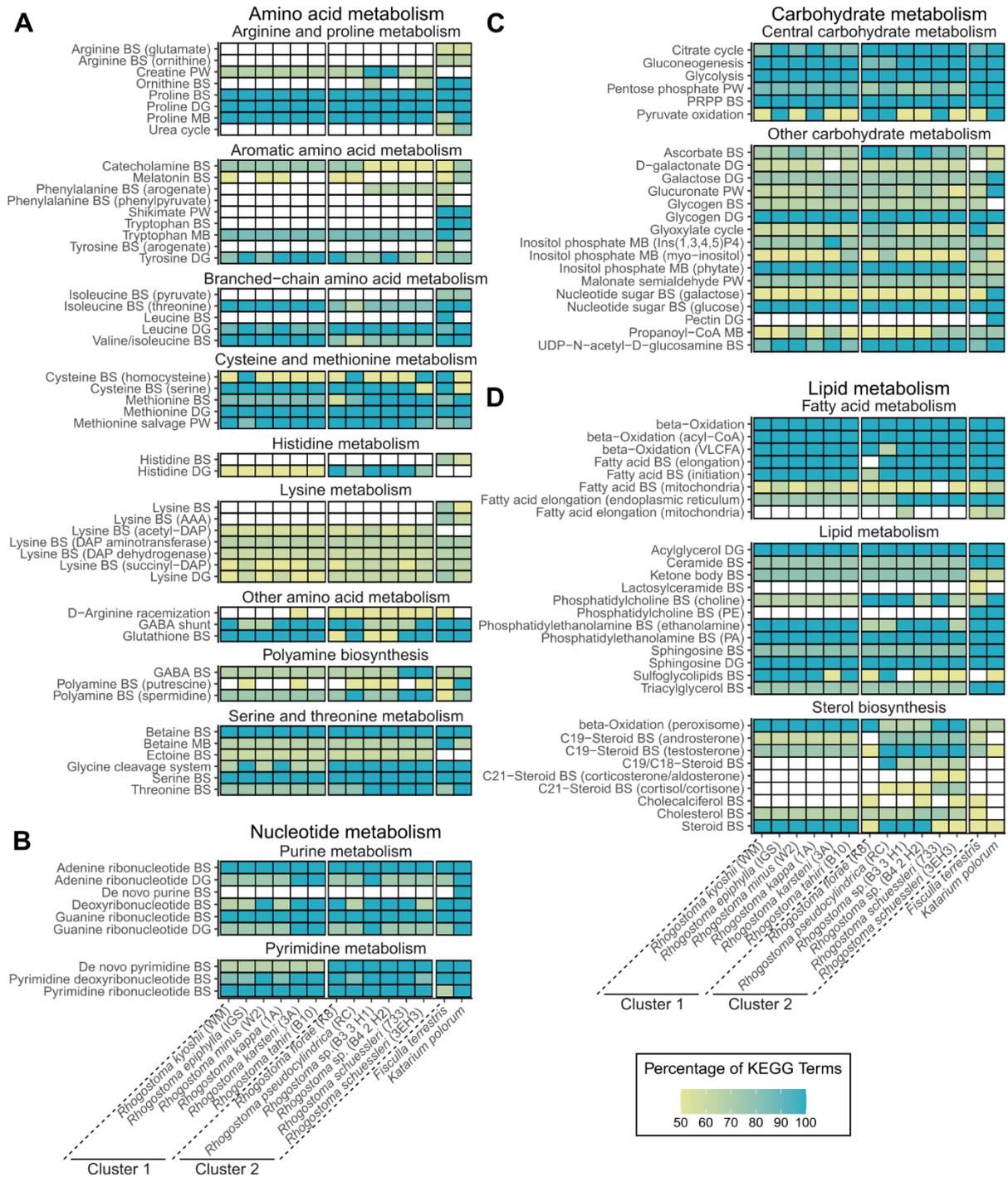
To explore the physiological traits of free-living the heterotrophic Thecofilosea, we compared and functionally annotated the whole transcriptomes of Thecofilosea species based on the KEGG database. Only the culture-based and thus deeply sequenced transcriptomes of *Rhogostoma*, *Fisculla*, and *Katarium* were considered, not the single-cell transcriptome of *Ebriida* sp. due to the comparably lower coverage of single-cell transcriptomes (Supplementary Table 1, Supplementary Figure 3). The transcriptomes comprised an average of 67,766 clustered contigs, with a mean mapping rate of 95% when aligning the reads back to the assembled contigs and a mean completeness of 86% according to benchmarking universal single-copy orthologs (BUSCO) of the Eukaryota dataset (Fig. 2B). Further, these transcriptomes contained an average of 44,972 predicted open reading frames (ORFs, Supplementary Table 1).

Overall, 93% of the ORFs were assigned to 58,676 orthogroups (Fig. 2A). We identified a core set of 4,178 orthogroups that was shared by all investigated Thecofilosea species (Fig. 2A). Compared to the



**Figure 2: Shared orthogroups and functional annotations of the Thecofilosea. (A)** The upper bar chart shows the number of shared orthogroups that are unique to the respective combination of Thecofilosida strains displayed in the matrix below. Only the top 25 strain combinations, based on the number of shared orthogroups, are shown. **(B)** BUSCOs assessment of the assembled Thecofilosida transcriptome, based on the Eukaryota database. **(C)** A selection of the functional annotations (KEGG orthology) of the Thecofilosida transcriptomes. The grey bar charts in the background display the relative proportions of functional annotations of all transcriptomes for the respective category. In addition, the relative proportions of functional annotations of the core set of orthogroups (purple), the *Rhogostoma* clusters 1 (blue) and 2 (dark green), *Katarium polorum* (light green) and *Fisculla terrestris* (yellow) are shown for the respective category.

functional annotations of all orthogroups, the core orthogroups exhibited a higher relative proportion of KO terms related to genetic information processing (Fig. 2C), in particular, related to replication and repair, transcription and translation. The two phylogenetically distinct *Rhogostoma* clusters showed equally distinct sets of orthogroups. Cluster 1 shared a significantly higher number of unique



**Figure 3: Overview of metabolic pathways for the Thecofilosea transcriptomes based on KEGG modules.** The heat maps show the presence and absence of the KEGG modules for amino acid metabolism (A), nucleotide metabolism (B), carbohydrate metabolism (C) and lipid metabolism (D) for each culture-based Thecofilosea transcriptome. The colour indicates the percentage of KO terms present, normalised to the minimum number of KO terms required for each module. White boxes indicate that either more than 50% or more than three of the required KO terms were missing. Modules that were absent in all Thecofilosea transcriptomes are not shown. Abbreviations: BS Biosynthesis, PW Pathway, MB Metabolism, DG Degradation.

orthogroups (4,612) than cluster 2 (704, Fig. 2A). Further, cluster 1 comprised a higher proportion of KO terms related to signal transduction, sensory systems, and environmental adaptation. In contrast, Cluster 2 comprised a higher proportion of KO terms related to transport and catabolism compared to

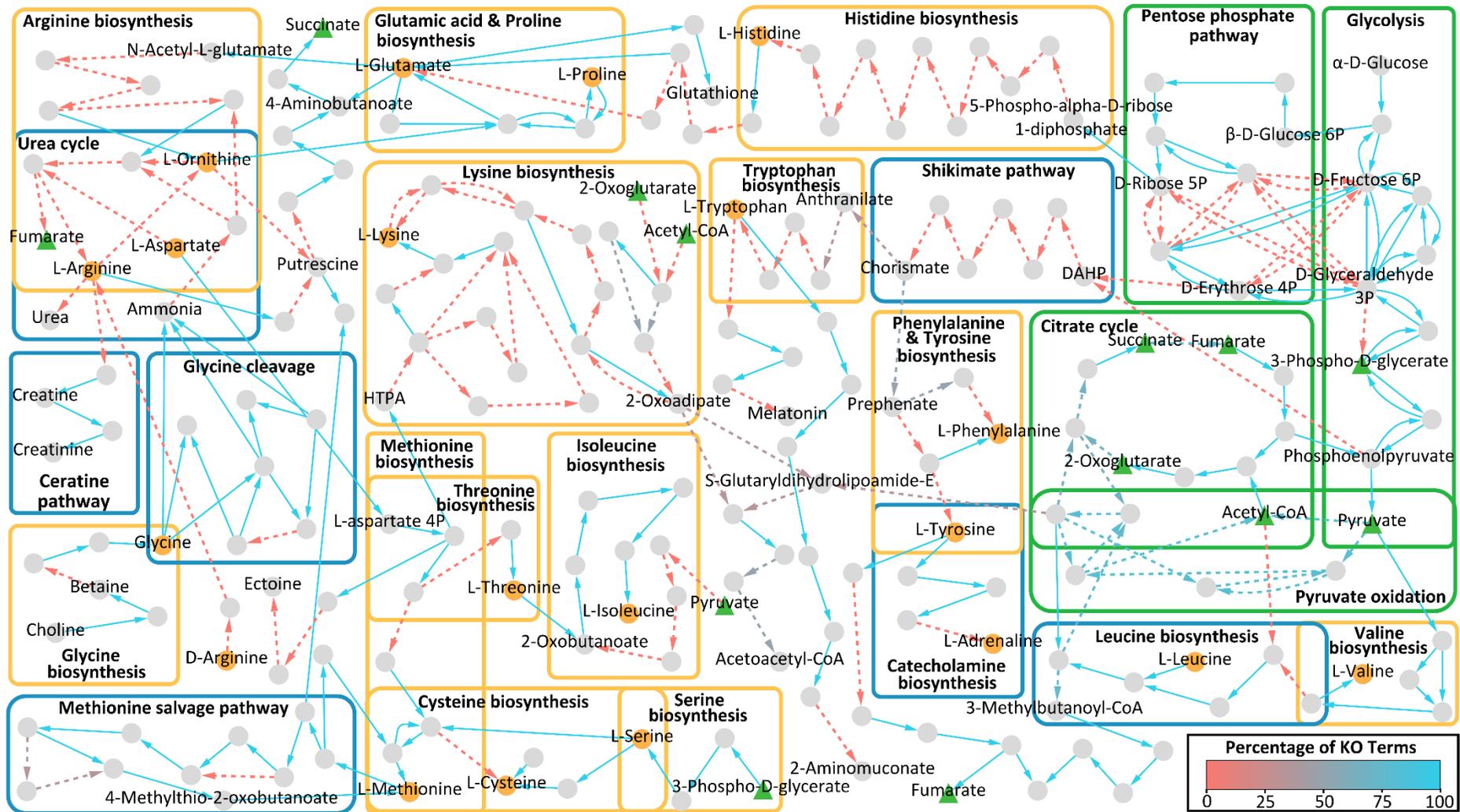
## Chapter 7

---

the overall functional annotation (Fig. 2C). Notably, both Tectofilosida species, *Fisculla terrestris* and *Katarium polorum* exhibited high numbers of unique orthogroups, 5,134 and 3,417, respectively (Fig. 2A). Both species showed a higher relative proportion of KO terms involved in cell motility. *Fisculla terrestris* additionally showed a relative increase of KO terms associated with cellular community processes such as focal adhesion, adherens junctions, tight junctions, and gap junctions compared to the overall functional annotation (Fig. 2C).

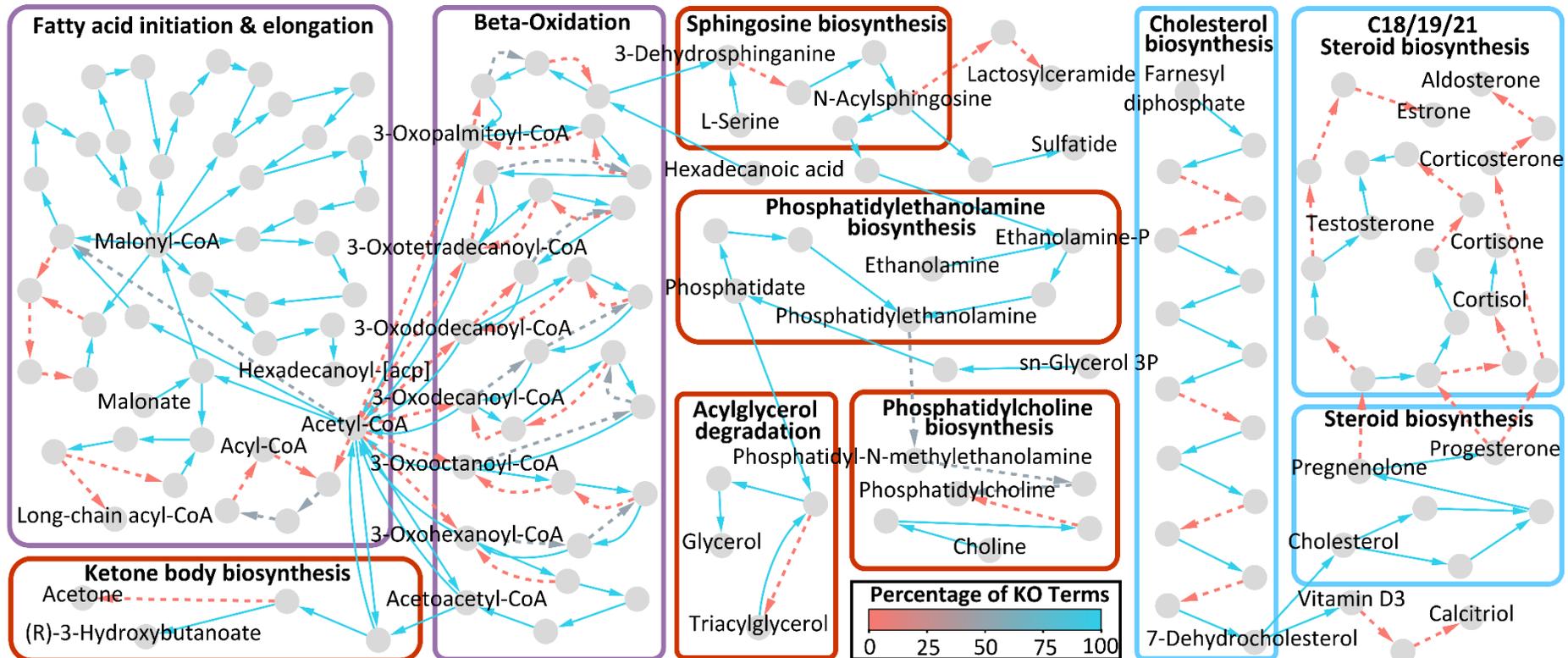
To gain insights into the metabolic adaptations of these free-living, heterotrophic Thecofilosea strains, we analyzed KEGG modules involved in amino acid, carbohydrate, nucleotide, and lipid metabolism (Fig. 3) and reconstructed the pathways for central carbohydrate and amino acid metabolism (Fig. 4 & 5, Supplementary Figures 4-29). The majority of genes involved in the central carbohydrate metabolism were expressed, although the complete pyruvate oxidation pathway could only be reconstructed for 6 out of 14 strains. The *Rhogostoma* cluster 2 lacked few genes for the non-oxidative pentose phosphate pathway. Furthermore, most of the genes involved in nucleotide biosynthesis were present throughout the dataset, except for genes related to the *de novo* purine biosynthesis, which were exclusively found in *Katarium polorum*. In contrast to the relatively conserved central carbohydrate and nucleotide metabolisms, greater variability was observed in amino acid and lipid metabolism. All strains expressed most genes involved in the biosynthesis of proline, valine, isoleucine, cysteine, methionine, serine, threonine, glutamate, and glycine. However, only *Fisculla terrestris* and *Katarium polorum* appeared to synthesize arginine, tryptophan, histidine, and lysine. *Fisculla terrestris* additionally possessed most genes to synthesize leucine, phenylalanine, and tyrosine. Differences in the expression of genes involved in lipid metabolism between the strains were detected, particularly genes involved in sterol biosynthesis. For example, strains belonging to the *Rhogostoma* Cluster 2 expressed a greater number of enzymes involved in the biosynthesis of C18/19/21-steroids.

## Chapter 7



**Figure 4: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma kyoshii* (WM).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma kyoshii* (WM) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7

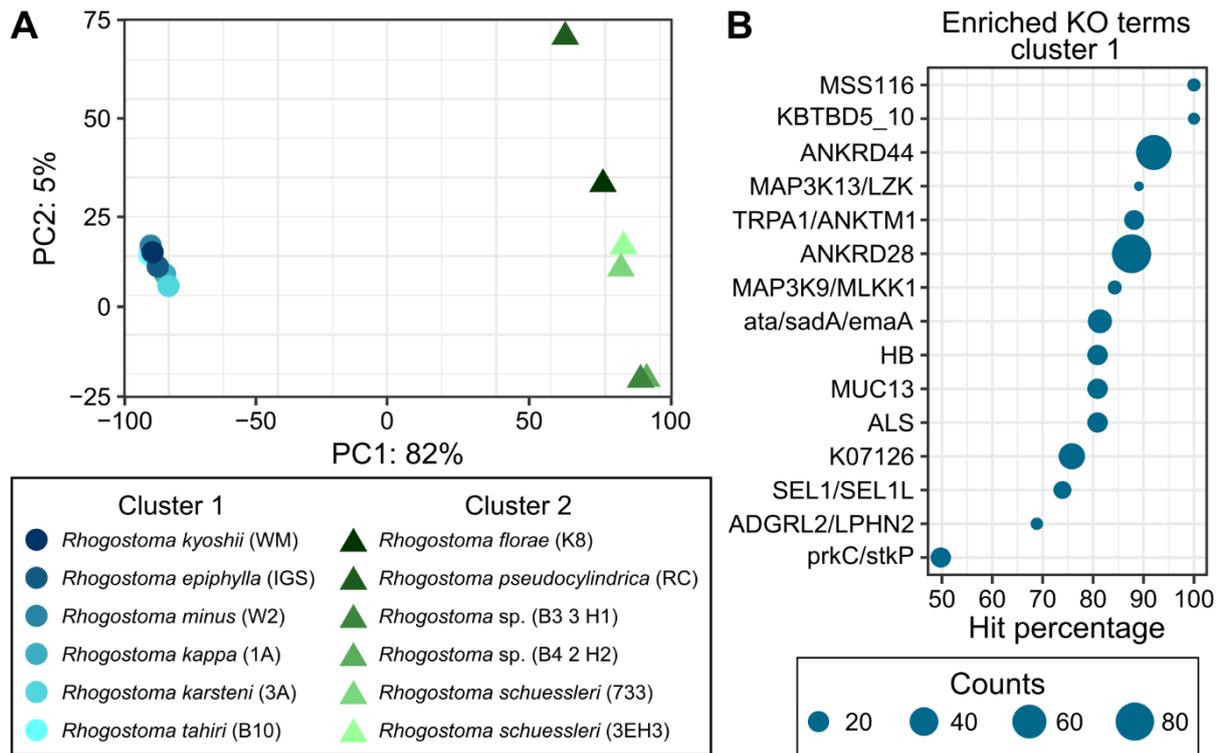


**Figure 5: Overview of the lipid metabolism for *Rhogostoma kyoshii* (WM).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma kyoshii* (WM) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

### *3.3 Enrichment analysis*

In addition to the analyses of presence-absence data, we compared the relative expression patterns of the two *Rhogostoma* clusters based on 20,000 orthogroups. A PCA showed a clear differentiation of *Rhogostoma* clusters 1 and 2, explaining 82% of the total variation (Fig. 6A). In addition, differences in the expression patterns of the long-branched, i.e., evolutionary distant *Rhogostoma* cluster 2 ( Fig. 1) explained 5% of the total variation, with strains of the same species grouping together.

Differential expression analysis revealed 7,502 out of the 20,000 orthogroups to be differentially expressed. About one-third of these orthogroups could be functionally annotated. A subsequent enrichment analysis identified a total of 15 KO terms that were consistently enriched across all six species of *Rhogostoma* cluster 1 (Fig. 6B). In contrast, no KO terms were consistently enriched across all six species of *Rhogostoma* cluster 2. ANKRD28 and ANKRD44 (Ankyrin repeat domains 28 and 44) were the most frequent KO terms of the enriched KO terms of *Rhogostoma* cluster 1, with 84 and 66 counts, respectively. ANKRD28 and ANKRD44 occurred almost exclusively (~90%) in the significantly enriched orthogroups (hit percentage, Fig. 6B). Furthermore, *ata/sadA/emaA* (trimeric autotransporter adhesin), ALS (agglutinin-like protein), MUC13 (mucin 13), and TRPA1/ANKTM1 (transient receptor potential cation channel subfamily A member 1) were identified among the enriched KO terms of *Rhogostoma* cluster 1. They also exhibited a high prevalence, with 26, 18, 18, and 17 counts, respectively, and ~80% hit percentage in the significantly enriched orthogroups (Fig. 6B).



**Figure 6: Comparative transcriptomics of *Rhogostoma*.** (A) Principal component analysis of all *Rhogostoma* strains based on the expression patterns of 20,000 orthogroups. The color and shape of the points encode for *Rhogostoma* clusters 1 (blue shades, round) and 2 (green shades, triangular). (B) Significantly enriched KO terms in all six *Rhogostoma* strains of cluster 1. The enrichment analysis was based on 7,502 out of 20,000 differentially expressed orthogroups (adjusted p-value < 0.001, |log<sub>2</sub> fold change| ≥ 1). The size of the points indicates the frequency of the respective KO terms in the higher expressed orthogroups of *Rhogostoma* cluster 1. The hit percentage describes the ratio of KO terms in higher expressed orthogroups of *Rhogostoma* cluster 1 compared to all orthogroups.

### 4 Discussion

#### 4.1 *The physiology capacity of Thecofilosea (Cecozoa)*

We identified a core set of conserved genes and pathways that were shared across all investigated Thecofilosea species. First, we identified a core set of shared orthogroups that included an exceptionally high proportion of genes associated with genetic information processing, i.e., translation, transcription, and replication. As these genes are essential for all living cells and are known to be highly conserved (Yao and O'Donnell, 2016), it was expected that these genes would be overrepresented in the core set of orthogroups. Second, we show an overall high completeness in the central carbohydrate and nucleotide metabolism, including the presence of glycolysis, the citrate cycle, and the capacity to synthesize all nucleotides. Although few enzymes were missing, this can most likely be explained by differences in the homologs due to rapid evolutionary rates or incompleteness of the transcriptome data. Overall, the robust identification of conserved genes and pathways emphasizes the soundness of our methods and data.

We were able to reconstruct the *de novo* synthesis for at least nine amino acids for all investigated Thecofilosea species. However, we also found variations in the physiological traits of the species. Heterotrophic microorganisms usually lack pathways for synthesizing certain amino acids and consequently depend on salvaging them from their prey. For example, the predatory amoebae *Dictyostelium* and *Arcella* have lost the capability to synthesize 11 and 5 amino acids, respectively (Payne and Loomis, 2006; Ribeiro et al., 2020). Our data indicate the absence of numerous amino acid pathways, for instance, tryptophan and histidine, particularly in *Rhogostoma* – indicating their dependence on the uptake of these compounds from their prey.

#### 4.2 *Physiology in an evolutionary context*

##### 4.2.1 *Multi-gene phylogeny*

By incorporating the new Thecofilosea transcriptomes into existing public Rhizaria data, our phylogenetic analysis expands the latest Cercozoa multi-gene phylogeny of Irwin et al. (2019). In our phylogenetic tree, the superclass Ventrifilosa is highly supported – a hypothesized monophylum of the predominantly shell-bearing and free-living cecozoa groups Thecofilosea and Imbricatea (Cavalier-Smith and Karpov, 2012). We further confirm that the Tectofilosida are indeed monophyletic as suggested by Dumack et al. (2017). Lastly, the Sarcomonadea branch basal to Ventrifilosa, as indicated by numerous single-gene phylogenies (Howe et al., 2011; Scoble and Cavalier-Smith, 2014).

### *4.2.2 Orthogroup similarity reflects the phylogenetic distance*

The high coverage and saturation of our culture-derived transcriptomic data allowed for whole transcriptome comparisons of physiological traits among Thecofilosea. The physiological capabilities based on nearly complete transcriptomes highly reflect evolutionary distance across all studied Thecofilosea strains. Within Tectofilosida, although only represented by two strains, the high evolutionary distance – indicated by long branches in the phylogenetic analyses – was reflected by the high number of unique orthogroups in each species and a moderate amount of shared orthogroups. Notably, *Fisculla terrestris* exhibited a higher proportion of unique orthogroups related to cell adherence, fusion, and cell-to-cell communication, likely reflecting its nature of frequent fusion (Gao et al., 2024).

*Rhogostoma* radiated into two clusters with notable differences in both orthogroups and gene expression patterns. The short-branched, i.e., evolutionary close, *Rhogostoma* cluster 1 exhibited a high number of shared orthogroups and clustered closer together in the PCA, compared to the long-branched, i.e., evolutionary distant, *Rhogostoma* cluster 2. Remarkably, *Rhogostoma* cluster 1 showed a higher proportion of genes related to signal transduction, sensory systems, and environmental adaptation in the orthogroups unique to this cluster. In addition, differences in gene expression patterns revealed enrichment in genes in *Rhogostoma* cluster 1 associated with first, sensory processes, including the reception of heat, pain, or environmental irritants (TRPA1/ANKTM1, Bautista et al., 2006), second protection and lubrication of cell surfaces (Muc13, Williams et al., 2001), third, cell migration and focal adhesion formation (ANKRD28, Tachibana et al., 2009), and fourth, cell adhesion to biotic and abiotic surfaces and biofilm formation (ata/sadA/emaA and ALS, Bentancor et al., 2012; Mintz, 2004; Oh et al., 2019; Raghunathan et al., 2011).

Aside from inter-cluster variation, there is additionally a large inter-specific variation in *Rhogostoma*, showcasing that even closely related and morphologically similar species exhibit distinct physiological traits and thus, potentially distinct ecological influence. It is important to note that although transcriptomic responses exhibit a snapshot of metabolic activity (Martin and Wang, 2011; Raghavan et al., 2022), we analyzed a high density of input cells reflecting the transcriptomic response of thousands of individuals, minimizing temporal and individual variation. To further increase the comparability, we grew all *Rhogostoma* strains in the same medium, strengthening our interpretation of the results.

### 4.2.3 Reduced gene set of *Rhogostoma*

In addition to a core set of orthogroups, each investigated evolutionary clade expressed unique transcripts, providing evidence that protistan functions in ecosystems cannot be easily generalized, even at a low taxonomic level. We show that *Rhogostoma*, a genus containing morphologically highly similar species (Öztoprak et al., 2020) and being the most derived clade in our phylogenetic tree, exhibited an exceptionally high variability in its physiological repertoire. The question arises, which selective forces led to the evolution of different physiological traits in species with highly similar morphologies?

As our phylogenetic tree shows, the Rhogostomidae (Cryomonadida) are closely related to the Protaspididae (Cryomonadida), highly specialized parasites of algae (Drebes et al., 1996; Schnepf and Kühn, 2000). The transition from a free-living to a parasitic lifestyle usually leads to the loss of existing functions and a simplified metabolism, as parasites exploit the resources of their host (Jackson et al., 2016; Poulin and Randhawa, 2015). Thus, the transition to parasitism is often thought to be irreversible, yet several studies, for example on diplomonads, provide evidence for a secondary free-living lifestyle (Wiśniewska et al., 2024; Xu et al., 2016). We hypothesize that the ancestors of the Cryomonadida underwent a loss of functional and genomic diversity with the specialization towards parasitism and that the ancestor of the Rhogostomidae broadened the prior narrow prey spectrum and specialized on feeding on bacteria in addition to eukaryotes.

Horizontal gene transfer (HGT) is, in general, considered to be a minor driver of eukaryote evolution (Keeling, 2024, 2019). Instead, eukaryotes are considered mainly to evolve by gene duplication and the subsequent adaption of homologs to different functions. Nonetheless, there is evidence that HGT in microbial eukaryotes may represent an escape from increasing adaptation to parasitism, which typically involves a loss of genomic and functional diversity (Wiśniewska et al., 2024; Xu et al., 2016), i.e., that HGT may contribute to the reversibility of parasitism and readaptation to a free-living lifestyle through the acquisition of new functions.

Overall, the back-transition from parasitism to free-living could potentially have caused the acquisition of even a single functional gene to cause a new speciation event in *Rhogostoma*, contributing to the current remarkable species diversity of *Rhogostoma*.

### 5 Conclusion

The outstanding high coverage of the transcriptomes and the deeply sampled phylogeny of the free-living and heterotrophic Thecofilosea allowed us to draw conclusions on their physiological capabilities and compare them in an evolutionary context. The Thecofilosea possess a core set of conserved genes involved in genetic information processing, nucleotide metabolism, and central carbohydrate metabolism. However, the high variability in amino acid and lipid metabolism, as well as the differentially expressed and enriched genes, indicate potentially distinct functional roles. These findings emphasize the remarkable physiological diversity even among closely related, morphologically highly similar taxa. Considering that the Thecofilosea represent only a fraction of the tremendous protistan diversity, the need for further research is evident. Our methodological approach paves the way for subsequent studies on a larger scale.

### Data Availability

The original data are available upon request and will be publicly available by the end of the project at the latest.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This study was supported by grants from the German Research Foundation (DFG) in the framework of the priority program SPP 1991: TAXON-OMICS (Project IDs 447013012 and 221301018) and with the grant number 555596351.

### Acknowledgments

We would like to thank Lukas Isenhardt for carrying out most of the RNA extractions and Gordon Lax for providing the single-cell sequences of Protaspida, which will soon be publicly available. We furthermore thank the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded (Funding number: INST 216/512/1FUGG) High Performance Computing (HPC) system CHEOPS as well as support.

### References

- Ahlmann-Eltze, C., 2024. ggupset: Combination Matrix Axis for “ggplot2” to Create “UpSet” Plots. R package version 0.4.0.
- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data.
- Balzano, S., Corre, E., Decelle, J., Sierra, R., Wincker, P., Da Silva, C., Poulain, J., Pawlowski, J., Not, F., 2015. Transcriptome analyses to investigate symbiotic relationships between marine protists. *Front Microbiol* 6.
- Bass, D., Howe, A.T., Mylnikov, A.P., Vickerman, K., Chao, E.E., Edwards Smallbone, J., Snell, J., Cabral, C., Cavalier-Smith, T., 2009. Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas*, *Eocercomonas*, *Paracercomonas*, and *Cavernomonas* gen. nov. *Protist* 160, 483–521. <https://doi.org/10.1016/j.protis.2009.01.004>
- Bautista, D.M., Jordt, S.-E., Nikai, T., Tsuruda, P.R., Read, A.J., Poblete, J., Yamoah, E.N., Basbaum, A.I., Julius, D., 2006. TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. *Cell* 124, 1269–1282. <https://doi.org/10.1016/j.cell.2006.02.023>
- Bentancor, L.V., Camacho-Peiro, A., Bozkurt-Guzel, C., Pier, G.B., Maira-Litrán, T., 2012. Identification of Ata, a multifunctional trimeric autotransporter of *Acinetobacter baumannii*. *J Bacteriol* 194, 3950–3960. <https://doi.org/10.1128/jb.06769-11>
- Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. *Curr Biol* 23, 1541–1547. <https://doi.org/10.1016/j.cub.2013.06.033>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinforma* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38, 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Cavalier-Smith, T., Chao, E.E., Lewis, R., 2018. Multigene phylogeny and cell evolution of chromist infrakingdom Rhizaria: contrasting cell organisation of sister phyla Cercozoa and Retaria. *Protoplasma* 255, 1517–1574. <https://doi.org/10.1007/s00709-018-1241-1>
- Cavalier-Smith, T., Karpov, S.A., 2012. *Paracercomonas* Kinetid Ultrastructure, Origins of the Body Plan of Cercomonadida, and Cytoskeleton Evolution in Cercozoa. *Protist* 163, 47–75. <https://doi.org/10.1016/j.protis.2011.06.004>
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinform* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Drebes, G., Kühn, S.F., Gmelch, A., Schnepf, E., 1996. *Cryothecomonas aestivalis* sp. nov., a colourless nanoflagellate feeding on the marine centric diatom *Guinardia delicatula* (Cleve) Hasle. *Helgolander Meeresunters* 50, 497–515. <https://doi.org/10.1007/BF02367163>
- Dumack, K., Fiore-Donno, A.M., Bass, D., Bonkowski, M., 2020. Making sense of environmental sequencing data: Ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Mol Ecol Resour* 20, 398–403. <https://doi.org/10.1111/1755-0998.13112>
- Dumack, K., Mausbach, P., Hegmann, M., Bonkowski, M., 2017. Polyphyly in the thecate amoeba genus *Lecythium* (Chlamydrophyridae, Tectofilosida, Cercozoa), redescription of its type species *L. hyalinum*, description of *L. jennyae* sp. nov. and the establishment of *Fisculla* gen. nov. and *Fiscullidae* fam. nov. *Protist* 168, 294–310. <https://doi.org/10.1016/j.protis.2017.03.003>
- Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16, 157. <https://doi.org/10.1186/s13059-015-0721-2>
- Flues, S., Blokker, M., Dumack, K., Bonkowski, M., 2018. Diversity of cercomonad species in the phyllosphere and rhizosphere of different plant species with a description of *Neocercomonas epiphylla* (Cercozoa, Rhizaria) a leaf-associated protist. *J Eukaryot Microbiol* 65, 587–599. <https://doi.org/10.1111/jeu.12503>
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinform* 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Gao, S., Solbach, M.D., Bast, J., Dumack, K., 2024. Meiosis-associated expression patterns during starvation-induced cell fusion in the protist *Fisculla terrestris*. *BMC Biol* (In review).
- Gerbracht, J.V., Harding, T., Simpson, A.G.B., Roger, A.J., Hess, S., 2022. Comparative transcriptomics reveals the molecular toolkit used by an algivorous protist for cell wall perforation. *Curr Biol* S0960982222008521. <https://doi.org/10.1016/j.cub.2022.05.049>

## Chapter 7

- Gomaa, F., Utter, D.R., Powers, C., Beaudoin, D.J., Edgcomb, V.P., Filipsson, H.L., Hansel, C.M., Wankel, S.D., Zhang, Y., Bernhard, J.M., 2021. Multiple integrated metabolic strategies allow foraminiferan protists to thrive in anoxic marine sediments. *Sci Adv* 7, eabf1586. <https://doi.org/10.1126/sciadv.abf1586>
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Grant, J.R., Lahr, D.J.G., Rey, F.E., Burleigh, J.G., Gordon, J.I., Knight, R., Molestina, R.E., Katz, L.A., 2012. Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: *Corallomyxa tenera*, *Chilodonella uncinata*, and *Subulatomonas tetraspora*, in: *Protist Genomics*. <https://doi.org/10.2478/prge-2012-0002>
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaultot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* 41, 597–604. <https://doi.org/10.1093/nar/gks1160>
- Haas, B., 2018. TransDecoder [WWW Document]. URL <https://github.com/TransDecoder/TransDecoder/wiki> (accessed 3.12.24).
- Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A.J.M., Faridani, O.R., Sandberg, R., 2020. Smart-seq3 Protocol. <https://doi.org/dx.doi.org/10.17504/protocols.io.bcq4ivyw>
- Howe, A.T., Bass, D., Scoble, J.M., Lewis, R., Vickerman, K., Arndt, H., Cavalier-Smith, T., 2011. Novel cultured protists identify deep-branching Environmental DNA clades of Cercozoa: New genera *Tremula*, *Micrometopion*, *Minimassisteria*, *Nudifila*, *Peregrinia*. *Protist* 162, 332–372. <https://doi.org/10.1016/j.protis.2010.10.002>
- Howe, A.T., Bass, D., Vickerman, K., Chao, E.E., Cavalier-Smith, T., 2009. Phylogeny, taxonomy, and astounding genetic diversity of Glissonadida ord. nov., the dominant gliding zooflagellates in soil (Protozoa: Cercozoa). *Protist* 160, 159–189. <https://doi.org/10.1016/j.protis.2008.11.007>
- Irwin, N.A.T., Tikhonenkov, D.V., Hehenberger, E., Mylnikov, A.P., Burki, F., Keeling, P.J., 2019. Phylogenomics supports the monophyly of the Cercozoa. *Mol Phylogenet and Evol* 130, 416–423. <https://doi.org/10.1016/j.ympev.2018.09.004>
- Jackson, A.P., Otto, T.D., Aslett, M., Armstrong, S.D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J.M., Dacks, J.B., Acosta-Serrano, A., Field, M.C., Ginger, M.L., Berriman, M., 2016. Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr Biol* 26, 161–172. <https://doi.org/10.1016/j.cub.2015.11.055>
- Kanehisa, M., 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 28, 1947–1951. <https://doi.org/10.1002/pro.3715>
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51, D587–D592. <https://doi.org/10.1093/nar/gkac963>
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Keeling, P.J., 2024. Horizontal gene transfer in eukaryotes: aligning theory with data. *Nat Rev Genet* 25, 416–430. <https://doi.org/10.1038/s41576-023-00688-5>
- Keeling, P.J., 2019. Combining morphology, behaviour and genomics to understand the evolution and ecology of microbial eukaryotes. *Philos Trans R Soc B Biol Sci* 374, 20190085. <https://doi.org/10.1098/rstb.2019.0085>
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Ryneerson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaultot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic

## Chapter 7

---

- life in the oceans through transcriptome sequencing. *PLOS Biol* 12, e1001889. <https://doi.org/10.1371/journal.pbio.1001889>
- Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinform* 28, 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Langmead, B., Wilks, C., Antonescu, V., Charles, R., 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinform* 35, 421–432. <https://doi.org/10.1093/bioinformatics/bty648>
- Lhee, D., Lee, J., Ettahi, K., Cho, C.H., Ha, J.-S., Chan, Y.-F., Zelzion, U., Stephens, T.G., Price, D.C., Gabr, A., Nowack, E.C.M., Bhattacharya, D., Yoon, H.S., 2021. Amoeba genome reveals dominant host contribution to plastid endosymbiosis. *Mol Biol Evol* 38, 344–357. <https://doi.org/10.1093/molbev/msaa206>
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021a. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* 38, 4647–4654. <https://doi.org/10.1093/molbev/msab199>
- Manni, M., Berkeley, M.R., Seppey, M., Zdobnov, E.M., 2021b. BUSCO: Assessing genomic data quality and beyond. *Curr Protoc* 1, e323. <https://doi.org/10.1002/cpz1.323>
- Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. *Nat Rev Genet* 12, 671–682. <https://doi.org/10.1038/nrg3068>
- Martínez Rendón, C., Braun, C., Kappelsberger, M., Boy, J., Casanova-Katny, A., Glaser, K., Dumack, K., 2024. Enhancing microbial predator-prey detection with network and trait-based analyses. *Microbiome* (in review).
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Mintz, K.P., 2004. Identification of an extracellular matrix protein adhesin, EmaA, which mediates the adhesion of *Actinobacillus actinomycetemcomitans* to collagen. *Microbiology* 150, 2677–2688. <https://doi.org/10.1099/mic.0.27110-0>
- Nakamura, Y., Suzuki, N., 2015. Phaeodaria: Diverse marine cercozoans of world-wide distribution, in: Ohtsuka, S., Suzuki, T., Horiguchi, T., Suzuki, N., Not, F. (Eds.), *Marine Protists: Diversity and Dynamics*. Springer Japan, Tokyo, pp. 223–249. [https://doi.org/10.1007/978-4-431-55130-0\\_9](https://doi.org/10.1007/978-4-431-55130-0_9)
- Neuhauser, S., Bulman, S., Kirchmair, M., 2010. Plasmodiophorids: The challenge to understand soil-borne, obligate biotrophs with a multiphasic life cycle, in: Gherbawy, Y., Voigt, K. (Eds.), *Molecular Identification of Fungi*. Springer, Berlin, Heidelberg, pp. 51–78. [https://doi.org/10.1007/978-3-642-05042-8\\_3](https://doi.org/10.1007/978-3-642-05042-8_3)
- Nowack, E.C.M., 2014. *Paulinella chromatophora* – rethinking the transition from endosymbiont to organelle. *Acta Soc Bot Pol* 83, 387–397. <https://doi.org/10.5586/asbp.2014.049>
- Oh, S.-H., Smith, B., Miller, A.N., Staker, B., Fields, C., Hernandez, A., Hoyer, L.L., 2019. Agglutinin-like sequence (ALS) genes in the *Candida parapsilosis* species complex: Blurring the boundaries between gene families that encode cell-wall proteins. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.00781>
- Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L., Fierer, N., 2020. The global-scale distributions of soil protists and their contributions to belowground systems. *Sci Adv* 6, eaax8787. <https://doi.org/10.1126/sciadv.aax8787>
- Öztoprak, H., Walden, S., Heger, T., Bonkowski, M., Dumack, K., 2020. What drives the diversity of the most abundant terrestrial cercozoan family (Rhogostomidae, Cercozoa, Rhizaria)? *Microorganisms* 8, 1123. <https://doi.org/10.3390/microorganisms8081123>
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
- Payne, S.H., Loomis, W.F., 2006. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryot Cell* 5, 272–276. <https://doi.org/10.1128/ec.5.2.272-276.2006>
- Pohl, N., Solbach, M.D., Dumack, K., 2021. The wastewater protist *Rhogostoma minus* (Thecofilosea, Rhizaria) is abundant, widespread, and hosts Legionellales. *Water Res* 203, 117566. <https://doi.org/10.1016/j.watres.2021.117566>
- Poulin, R., Randhawa, H.S., 2015. Evolution of parasitism along convergent lines: From ecology to genomics. *Parasitology* 142, S6–S15. <https://doi.org/10.1017/S0031182013001674>
- Raghavan, V., Kraft, L., Mesny, F., Rigerte, L., 2022. A simple guide to *de novo* transcriptome assembly and annotation. *Brief Bioinform* 23, bbab563. <https://doi.org/10.1093/bib/bbab563>

## Chapter 7

- Raghunathan, D., Wells, T.J., Morris, F.C., Shaw, R.K., Bobat, S., Peters, S.E., Paterson, G.K., Jensen, K.T., Leyton, D.L., Blair, J.M.A., Browning, D.F., Pravin, J., Flores-Langarica, A., Hitchcock, J.R., Moraes, C.T.P., Piazza, R.M.F., Maskell, D.J., Webber, M.A., May, R.C., MacLennan, C.A., Piddock, L.J., Cunningham, A.F., Henderson, I.R., 2011. SadA, a trimeric autotransporter from *Salmonella enterica* Serovar Typhimurium, can promote biofilm formation and provides limited protection against infection. *Infect Immun* 79, 4342–4352. <https://doi.org/10.1128/iai.05592-11>
- Ribeiro, G.M., Porfírio-Sousa, A.L., Maurer-Alcalá, X.X., Katz, L.A., Lahr, D.J.G., 2020. De novo sequencing, assembly, and annotation of the transcriptome for the free-living testate amoeba *Arcella intermedia*. *J Eukaryot Microbiol* 67, 383–392. <https://doi.org/10.1111/jeu.12788>
- Rodríguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A.J., Gray, M.W., Philippe, H., Lang, B.F., 2007. Toward Resolving the Eukaryotic Tree: The Phylogenetic Positions of Jakobids and Cercozoans. *Curr Biol* 17, 1420–1425. <https://doi.org/10.1016/j.cub.2007.07.036>
- Schnepf, E., Kühn, S.F., 2000. Food uptake and fine structure of *Cryothecomonas longipes* sp. nov., a marine nanoflagellate incertae sedis feeding phagotrophically on large diatoms. *Helgol Mar Res* 54, 18–32. <https://doi.org/10.1007/s101520050032>
- Schwelm, A., Fogelqvist, J., Knaust, A., Jülke, S., Lilja, T., Bonilla-Rosso, G., Karlsson, M., Shevchenko, A., Dhandapani, V., Choi, S.R., Kim, H.G., Park, J.Y., Lim, Y.P., Ludwig-Müller, J., Dixelius, C., 2015. The Plasmodiophora brassicae genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci Rep* 5, 11153. <https://doi.org/10.1038/srep11153>
- Scoble, J.M., Cavalier-Smith, T., 2014. Scale evolution, sequence phylogeny, and taxonomy of thaumatomonad Cercozoa: 11 new species and new genera *Scutellomonas*, *Cowlomonas*, *Thaumatospina* and *Ovaloplaca*. *Eur J Protistol* 50, 270–313. <https://doi.org/10.1016/j.ejop.2013.12.005>
- Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., Lara, E., 2017. Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. *Soil Biol Biochem* 112, 68–76. <https://doi.org/10.1016/j.soilbio.2017.05.002>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sibbald, S.J., Archibald, J.M., 2017. More protist genomes needed. *Nat Ecol Evol* 1, 0145. <https://doi.org/10.1038/s41559-017-0145>
- Sierra, R., Cañas-Duarte, S.J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., González-García, L.N., Gile, G.H., Slamovits, C.H., Klopp, C., Restrepo, S., Arzul, I., Pawlowski, J., 2016. Evolutionary origins of rhizarian parasites. *Mol Biol Evol* 33, 980–983. <https://doi.org/10.1093/molbev/msv340>
- Solbach, M.D., Bonkowski, M., Dumack, K., 2024. *Katarium polorum* n. sp., n. g., a novel thecofilosean amoeba (Cercozoa, Rhizaria) from the polar oceans. *J Eukaryot Microbiol* (Accepted Manuscript).
- Solbach, M.D., Bonkowski, M., Dumack, K., 2021. Novel endosymbionts in rhizarian amoebae imply universal infection of unrelated free-living amoebae by Legionellales. *Front Cell Infect Microbiol* 11. <https://doi.org/10.3389/fcimb.2021.642216>
- Sommeria-Klein, G., Watteaux, R., Ibarbalz, F.M., Pierella Karlusich, J.J., Iudicone, D., Bowler, C., Morlon, H., 2021. Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science* 374, 594–599. <https://doi.org/10.1126/science.abb3717>
- Soneson, C., Love, M., Robinson, M., 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4. <https://doi.org/10.12688/f1000research.7563.2>
- Song, L., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4, 48. <https://doi.org/10.1186/s13742-015-0089-y>
- Tachibana, M., Kiyokawa, E., Hara, S., Iemura, S., Natsume, T., Manabe, T., Matsuda, M., 2009. Ankyrin repeat domain 28 (ANKRD28), a novel binding partner of DOCK180, promotes cell migration by regulating focal adhesion formation. *Exp Cell Res* 315, 863–876. <https://doi.org/10.1016/j.yexcr.2008.12.005>
- Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., Rokas, A., Shen, X.-X., Strassert, J.F.H., Kolísko, M., Brown, M.W., 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. *PLOS Biol* 19, e3001365. <https://doi.org/10.1371/journal.pbio.3001365>
- Walden, S., Jauss, R.-T., Feng, K., Fiore-Donno, A.M., Dumack, K., Schaffer, S., Wolf, R., Schlegel, M., Bonkowski, M., 2021. On the phenology of protists: recurrent patterns reveal seasonal variation of protistan (Rhizaria: Cercozoa and Endomyxa) communities in tree canopies. *FEMS Microbiol Ecol* 97, 2021.02.15.431229. <https://doi.org/10.1093/femsec/fiab081>
- Wickham, H., 2011. ggplot2. *Wiley Interdiscip Rev Comput Stat* 3, 180–185. <https://doi.org/10.1002/wics.147>

## Chapter 7

---

- Williams, S.J., Wreschner, D.H., Tran, M., Eyre, H.J., Sutherland, G.R., McGuckin, M.A., 2001. MUC13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells. *J Biol Chem* 276, 18327–18336. <https://doi.org/10.1074/jbc.M008850200>
- Wiśniewska, M.M., Salomaki, E.D., Silberman, J.D., Terpis, K.X., Mazancová, E., Táborský, P., Jinatham, V., Gentekaki, E., Čepička, I., Kolisko, M., 2024. Expanded gene and taxon sampling of diplomonads shows multiple switches to parasitic and free-living lifestyle. *BMC Biol* 22, 217. <https://doi.org/10.1186/s12915-024-02013-w>
- Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>
- Xu, F., Jerlström-Hultqvist, J., Kolisko, M., Simpson, A.G.B., Roger, A.J., Svärd, S.G., Andersson, J.O., 2016. On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. *BMC Biol* 14, 62. <https://doi.org/10.1186/s12915-016-0284-z>
- Yao, N.Y., O'Donnell, M.E., 2016. Evolution of replication machines. *Crit Rev Biochem Mol Biol* 51, 135–149. <https://doi.org/10.3109/10409238.2015.1125845>
- Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11, R14. <https://doi.org/10.1186/gb-2010-11-2-r14>

## Chapter 7

### Supplementary material

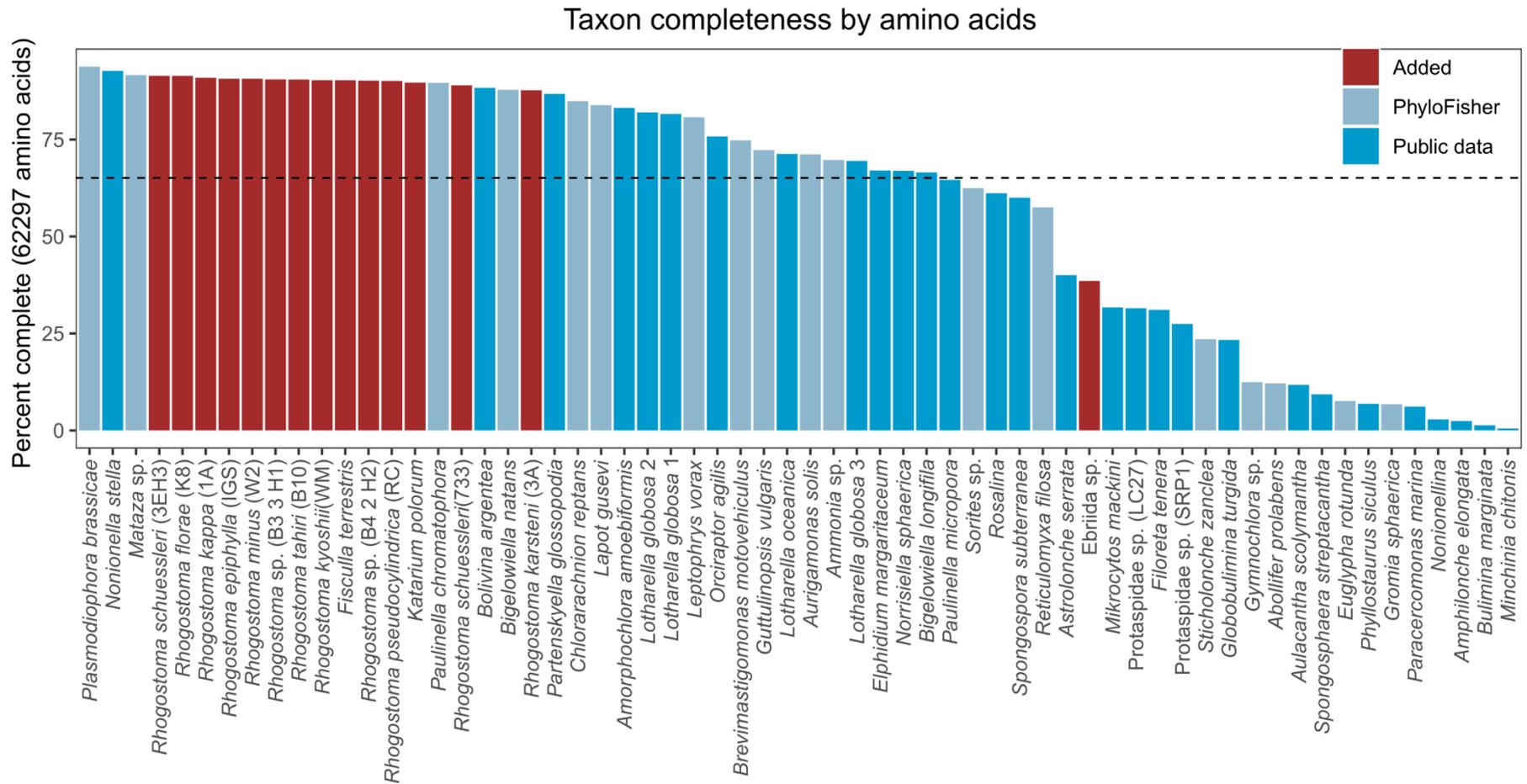
**Supplementary Table 1: Overview of transcriptome assemblies.** The table shows the number of raw, trimmed and filtered reads of all *Rhogostma* strains, *Fisculla terrestris*, *Katarium polorum*, *Protaspida* sp. and *Ebriida* sp. In addition, the number of Trinity contigs and TransDecoder peptides is given, as well as the BUSCO completeness (Eukaryote odb10 database) and the Bowtie RNA-Seq read mapping rate. The latter is given for both the original and clustered Trinity contigs for all *Rhogostma* strains, *Fisculla terrestris* and *Katarium polorum*.

Species	Raw reads	Reads after trimming	Reads after contamination filtering	mRNA reads	Clustered/Not Clustered	Trinity			BUSCO		Mapping rate	TransDecoder		
						Transcripts	N50	Median length	Complete genes	Fragmented genes		Peptides	N50	Median length
<i>Rhogostoma kyoshii</i> (WM)	66533991	53579788	51070053	50742478	Not Clustered	70500	1700	749	86.2	6.7	99.64	54146	501	251
					Clustered	59087	1661	725	86.3	6.7	95.94	43860	488	247
<i>Rhogostoma epiphylla</i> (IGS)	65487075	52520473	49620845	49201702	Not Clustered	87591	1683	692	87.9	6.3	99.63	65641	504	243
					Clustered	71356	1613	655	86.6	6.7	95.68	51349	475	232
<i>Rhogostoma minus</i> (W2)	66314131	58020320	55233041	54923000	Not Clustered	77982	1746	756	87.5	5.9	99.62	60335	512	251
					Clustered	64673	1681	731	87.5	5.9	96.13	48384	488	244
<i>Rhogostoma kappa</i> (1A)	51545724	40019834	37599379	37340949	Not Clustered	78813	1685	697	87.5	5.5	99.5	58373	497	245
					Clustered	64710	1648	692	86.7	6.3	95.46	47114	480	241
<i>Rhogostoma karsteni</i> (3A)	51003791	35252147	33337297	32956763	Not Clustered	112600	1547	656	84.3	8.6	99.29	81278	434	227
					Clustered	91974	1515	638	83.9	9	92.01	64913	414	220
<i>Rhogostoma tahiri</i> (B10)	51142494	36827766	34794910	34503204	Not Clustered	78688	1608	724	85.1	7.8	99.56	59521	467	241
					Clustered	65703	1575	709	84.7	8.2	93.72	48624	452	236
<i>Rhogostoma florum</i> (K8)	62748138	50695419	48711042	48375794	Not Clustered	58503	1876	784	87.4	5.9	99.33	40058	605	312
					Clustered	49440	1817	724	86.6	5.9	96.18	32067	590	307
<i>Rhogostoma pseudocylindrica</i> (RC)	71362011	55869020	54019464	53442745	Not Clustered	70970	1808	721	89.1	5.1	99.4	48490	565	285
					Clustered	56109	1764	729	87	6.3	96.58	37473	549	286
<i>Rhogostoma</i> sp. (B3 3 H1)	57476393	46776628	45348061	44878690	Not Clustered	77985	1910	898	87.8	5.9	99.38	54185	573	296
					Clustered	65178	1858	822	86.7	7.1	95.2	42393	551	285
<i>Rhogostoma</i> sp. (B4 2 H2)	52974991	41960462	40494773	40170318	Not Clustered	65232	1819	881	87.8	5.5	99.51	45318	572	302
					Clustered	54605	1769	804	86.2	6.7	95.13	35310	556	293
<i>Rhogostoma schuessleri</i> (733)	48558834	36144031	34838367	34125325	Not Clustered	74686	1471	706	81.6	9.8	98.53	52855	451	258
					Clustered	61976	1441	637	80.4	11	93.1	41357	444	250
<i>Rhogostoma schuessleri</i> (3EH3)	67459726	53092251	50988304	50202712	Not Clustered	91664	1760	771	87.9	5.1	98.8	62174	548	285
					Clustered	71423	1619	677	86.7	6.3	94.38	44303	503	271
<i>Fisculla terrestris</i>	60284486	54248087.7	50390734.7	50038010.8	Not Clustered	105351	1701	574	86.6	8.2	99.63	66945	534	262
					Clustered	91295	1572	534	85.1	9.4	96.64	54448	494	250
<i>Katarium polorum</i>	64929398	53808760	49336388	48736075	Not Clustered	92384	1500	435	88.2	5.9	98.72	46980	529	266
					Clustered	81202	1436	401	86.6	7.1	92.49	38010	523	261
<i>Ebriida</i> sp.	65095457	55622027	50593796	46007110	Not Clustered	32903	645	435	10.2	19.2	89.83	16924	221	171
<i>Protaspidae</i> sp. (SRP1)	10703734	7585655	6116460	5297090	Not Clustered	14114	911	461	7.4	5.9	98.35	6085	298	197
<i>Protaspidae</i> sp. (LC27)	7428223	5183399	4167035	3694243	Not Clustered	12167	939	479	9	5.9	97.97	5720	290	196

## Chapter 7

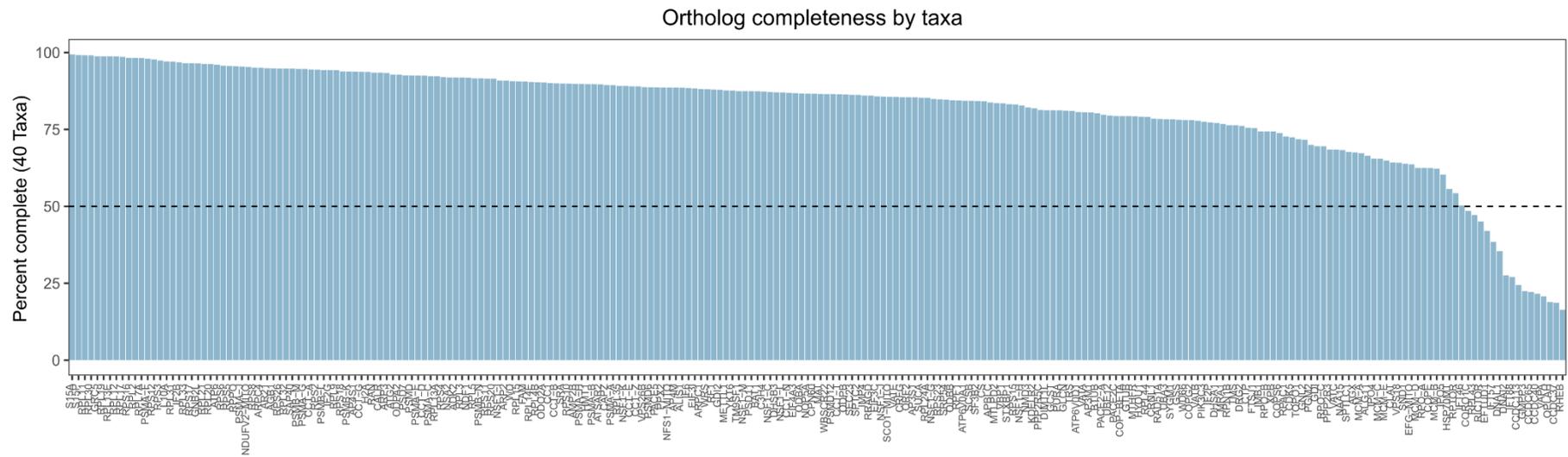
**Supplementary Table 2: Overview of all Rhizaria strains included in the phylogenetic analysis.** The table provides information on the data type and processing status (raw data, assembly or protein sequences), as well as the data origin, i.e., database and ID, and the citation and DOI of the corresponding publication (if available). The Rhizaria species included in the final multi-gene phylogeny are written in bold and marked with an asterisk.

Species	Type	Status	Database	Database ID	Author	DOI
<i>Amorphochlora amoebiformis</i> *	Transcriptome	Assembly	Zenodo	MMETSP0042_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001891">https://doi.org/10.1371/journal.pbio.1001891</a>
<i>Amphilonche elongata</i>	Transcriptome	Assembly	Zenodo	MMETSP1359	Balzano et al., 2015	<a href="https://doi.org/10.3389/fmicb.2015.00098">https://doi.org/10.3389/fmicb.2015.00098</a>
<i>Astrolonche serrata</i>	Transcriptome	Proteins	EukProt v3	EP00494	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
<i>Aulacantha scolymantha</i>	Transcriptome	Assembly	EukProt v3	EP00463	Balzano et al., 2015	<a href="https://doi.org/10.3389/fmicb.2015.00098">https://doi.org/10.3389/fmicb.2015.00098</a>
<i>Bigelowiella longifila</i> *	Transcriptome	Assembly	Zenodo	MMETSP1359	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001898">https://doi.org/10.1371/journal.pbio.1001898</a>
<i>Bolivina argentea</i> *	Transcriptome	Proteins	EukProt v3	EP01084	Gomaa et al., 2021	<a href="https://doi.org/10.1126/sciadv.abf1586">https://doi.org/10.1126/sciadv.abf1586</a>
<i>Brizalina</i> sp.	Transcriptome	Assembly	EukProt v3	EP00481	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
<i>Bulimina marginata</i>	Transcriptome	Assembly	EukProt v3	EP00487	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
Protaspidae sp. (SRP1)*	Transcriptome	Raw	NCBI	SRR31106255	Gordon et al., in prep.	
Protaspidae sp. (LC27)*	Transcriptome	Raw	NCBI	SRR31106254	Gordon et al., in prep.	
<i>Elphidium margaritaceum</i> *	Transcriptome	Assembly	Zenodo	MMETSP1385	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001899">https://doi.org/10.1371/journal.pbio.1001899</a>
<i>Filoreta tenera</i>	Transcriptome	Proteins	EukProt v3	EP00478	Grant et al., 2012	<a href="https://doi.org/10.2478/prge-2012-0002">https://doi.org/10.2478/prge-2012-0002</a>
<i>Globobulimina turgida</i>	Transcriptome	Proteins	EukProt v3	EP00483	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
<i>Lotharella globosa</i> 1*	Transcriptome	Assembly	Zenodo	MMETSP0041_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001890">https://doi.org/10.1371/journal.pbio.1001890</a>
<i>Lotharella globosa</i> 2*	Transcriptome	Assembly	Zenodo	MMETSP0111_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001892">https://doi.org/10.1371/journal.pbio.1001892</a>
<i>Lotharella globosa</i> 3*	Transcriptome	Assembly	Zenodo	MMETSP0112_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001893">https://doi.org/10.1371/journal.pbio.1001893</a>
<i>Lotharella oceanica</i> *	Transcriptome	Assembly	Zenodo	MMETSP0040_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001889">https://doi.org/10.1371/journal.pbio.1001889</a>
<i>Mikrocytos mackini</i>	Transcriptome	Proteins	EukProt v3	EP00477	Burki et al., 2013	<a href="https://doi.org/10.1016/j.cub.2013.06.033">https://doi.org/10.1016/j.cub.2013.06.033</a>
<i>Minchinia chitonis</i>	Transcriptome	Assembly	Zenodo	MMETSP0186	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001895">https://doi.org/10.1371/journal.pbio.1001895</a>
<i>Nonionella stella</i> *	Transcriptome	Proteins	EukProt v3	EP01083	Gomaa et al., 2021	<a href="https://doi.org/10.1126/sciadv.abf1586">https://doi.org/10.1126/sciadv.abf1586</a>
<i>Nonionellina</i> sp.	Transcriptome	Assembly	EukProt v3	EP00488	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
<i>Norrisiella sphaerica</i> *	Transcriptome	Assembly	Zenodo	MMETSP0113_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001894">https://doi.org/10.1371/journal.pbio.1001894</a>
<i>Orciraptor agilis</i> *	Transcriptome	Assembly	ENA	HBWT01000000	Gerbracht et al., 2022	<a href="https://doi.org/10.1016/j.cub.2022.05.049">https://doi.org/10.1016/j.cub.2022.05.049</a>
<i>Paracercomonas marina</i>	EST	Proteins	EukProt v3	EP00460	Rodríguez-Ezpeleta et al., 2007	<a href="https://doi.org/10.1016/j.cub.2007.07.036">https://doi.org/10.1016/j.cub.2007.07.036</a>
<i>Partenskyella glossopodia</i> *	Transcriptome	Assembly	Zenodo	MMETSP1318	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001897">https://doi.org/10.1371/journal.pbio.1001897</a>
<i>Paulinella micropora</i>	Genome	Proteins	EukProt v3	EP00808	Lhee et al., 2021	<a href="https://doi.org/10.1093/molbev/msaa206">https://doi.org/10.1093/molbev/msaa206</a>
<i>Phyllostaurus siculus</i>	Transcriptome	Proteins	EukProt v3	EP00493	Sierra et al., 2016	<a href="https://doi.org/10.1093/molbev/msv340">https://doi.org/10.1093/molbev/msv340</a>
<i>Rosalina</i> sp.	Transcriptome	Assembly	Zenodo	MMETSP0190_2	Keeling et al., 2014	<a href="https://doi.org/10.1371/journal.pbio.1001896">https://doi.org/10.1371/journal.pbio.1001896</a>
<i>Spongosphaera streptacantha</i>	EST	Assembly	EukProt v3	EP00497	Balzano et al., 2015	<a href="https://doi.org/10.3389/fmicb.2015.00098">https://doi.org/10.3389/fmicb.2015.00098</a>
<i>Spongospora subterranea</i>	Transcriptome	Proteins	EukProt v3	EP00474	Schwelm et al., 2015	<a href="https://doi.org/10.1038/srep11153">https://doi.org/10.1038/srep11153</a>

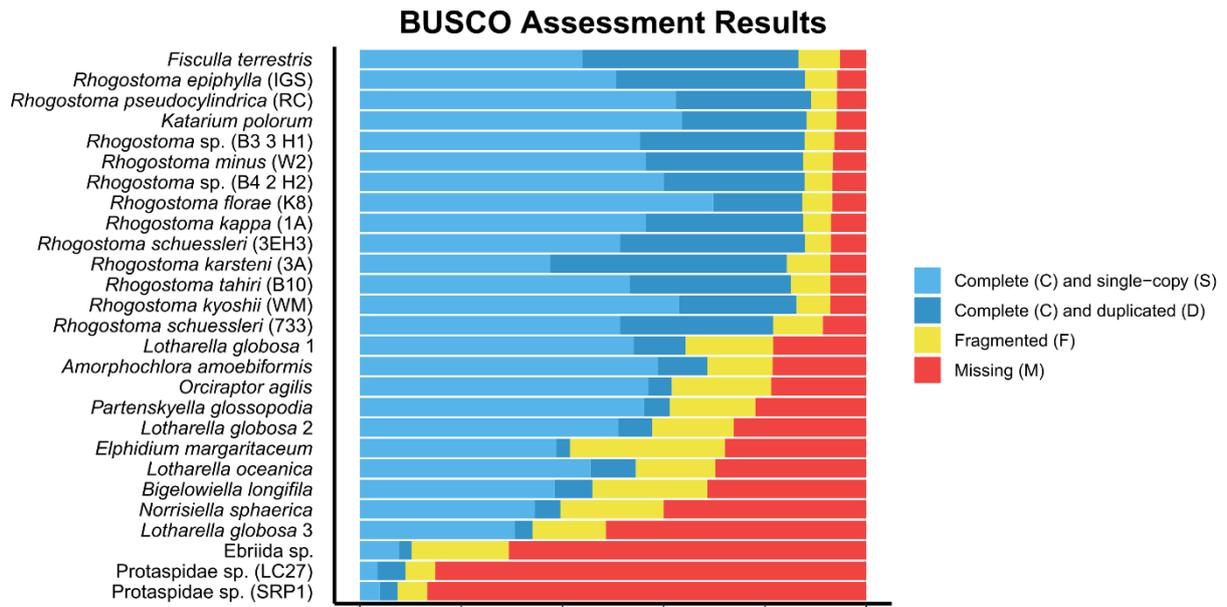


**Supplementary Figure 1: Taxon completeness by amino acids.** The bar charts show the completeness, measured as the percentage of amino acids, of a total of 240 genes selected for phylogenetic analysis, from 62 Rhizaria strains. The colours indicate the origin of the strains, i.e., newly acquired data (red), publicly available data (dark blue) and data provided by PhyloFisher (light blue).

## Chapter 7

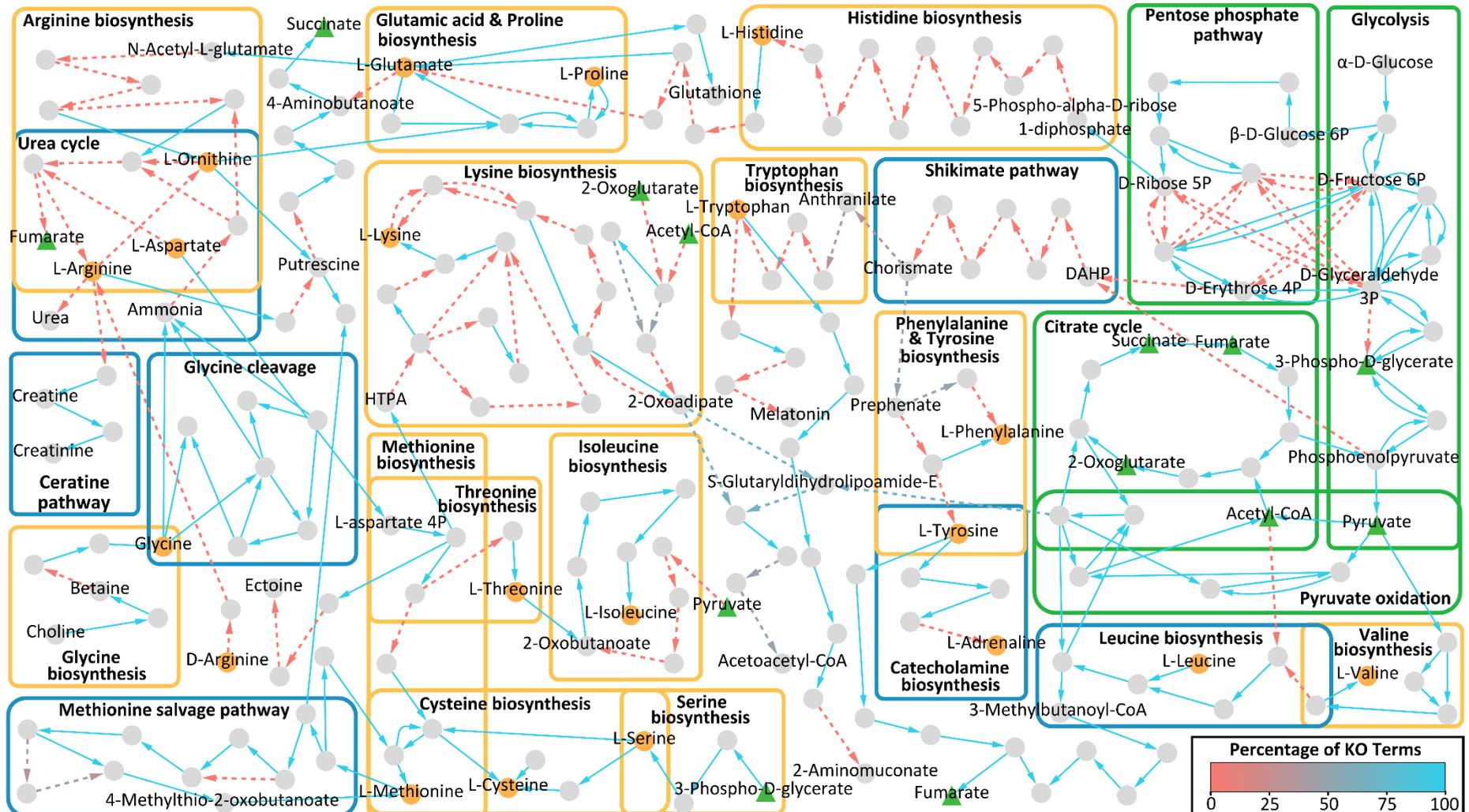


**Supplementary Figure 2: Ortholog completeness by amino acids.** The bar charts show the percentage coverage of 40 Rhizaria strains for 240 genes selected for phylogenetic analysis.



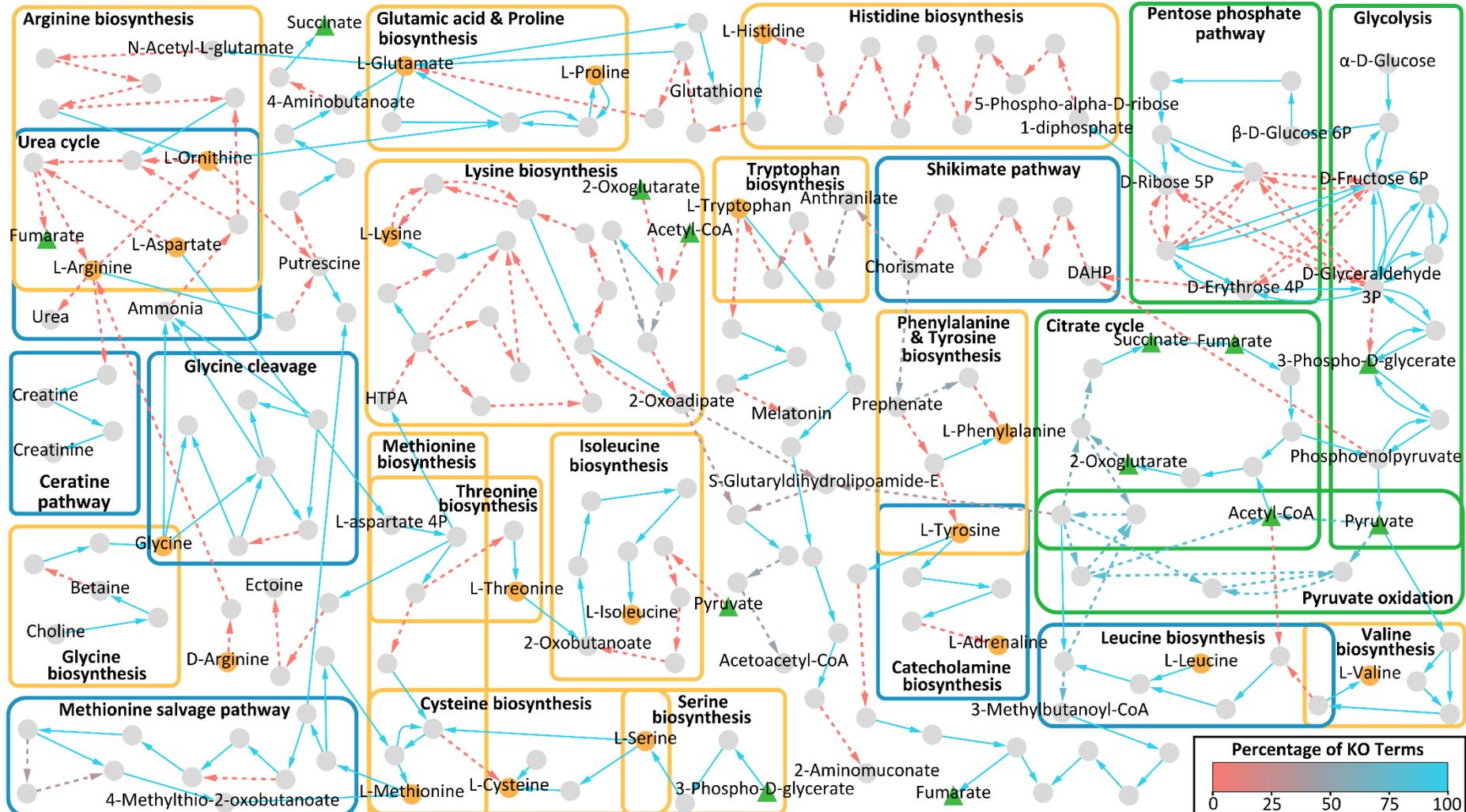
**Supplementary Figure 3: BUSCOs assessment of the Rhizaria transcriptomes.** BUSCO assessment of the Rhizaria transcriptomes based on the Eukaryota database. The colour code for complete single-copy orthologs (light blue), duplicated complete orthologs (dark blue), fragmented orthologs (yellow) and missing orthologs (red).

## Chapter 7



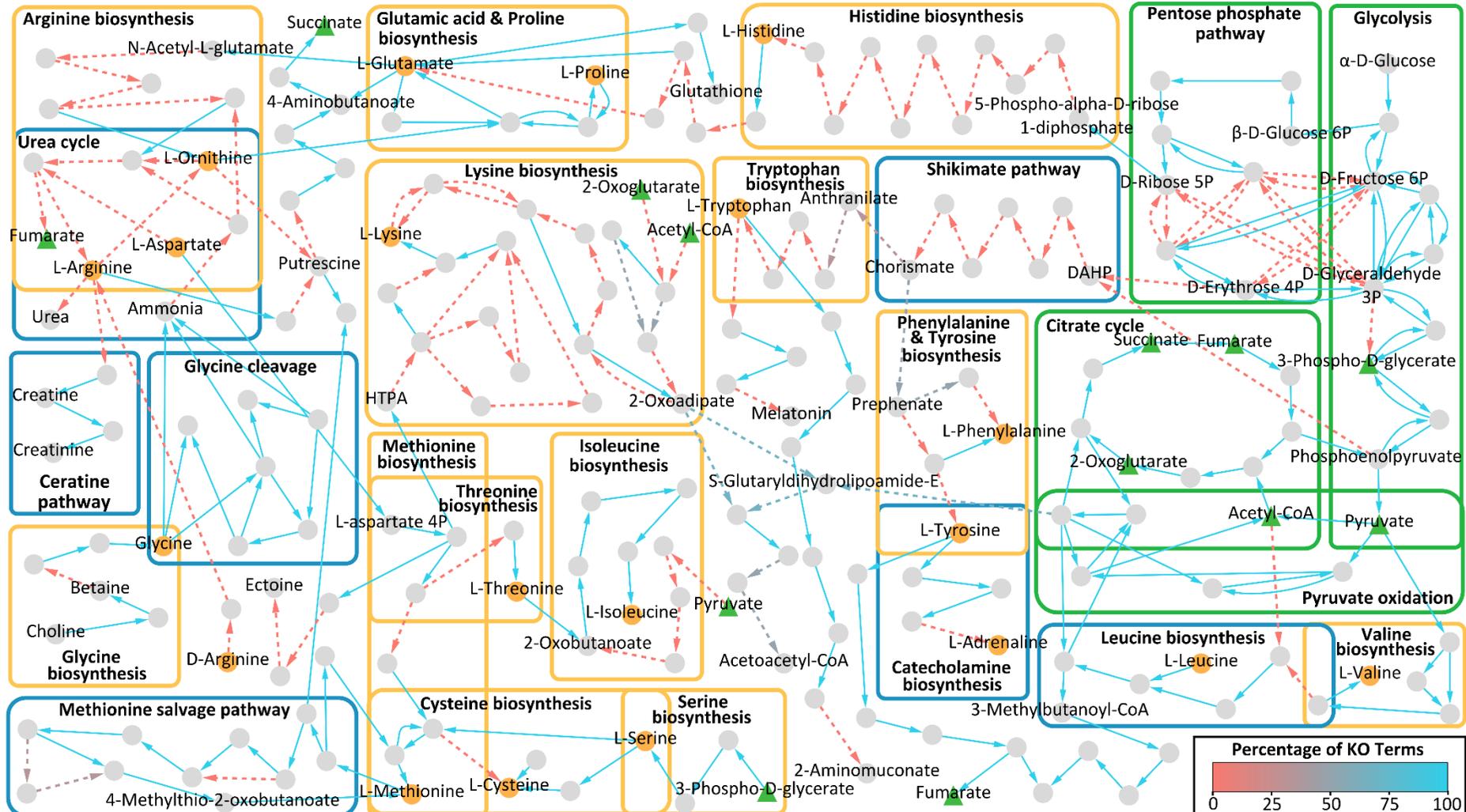
**Supplementary Figure 4: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma epiphylla* (IGS).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma epiphylla* (IGS) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



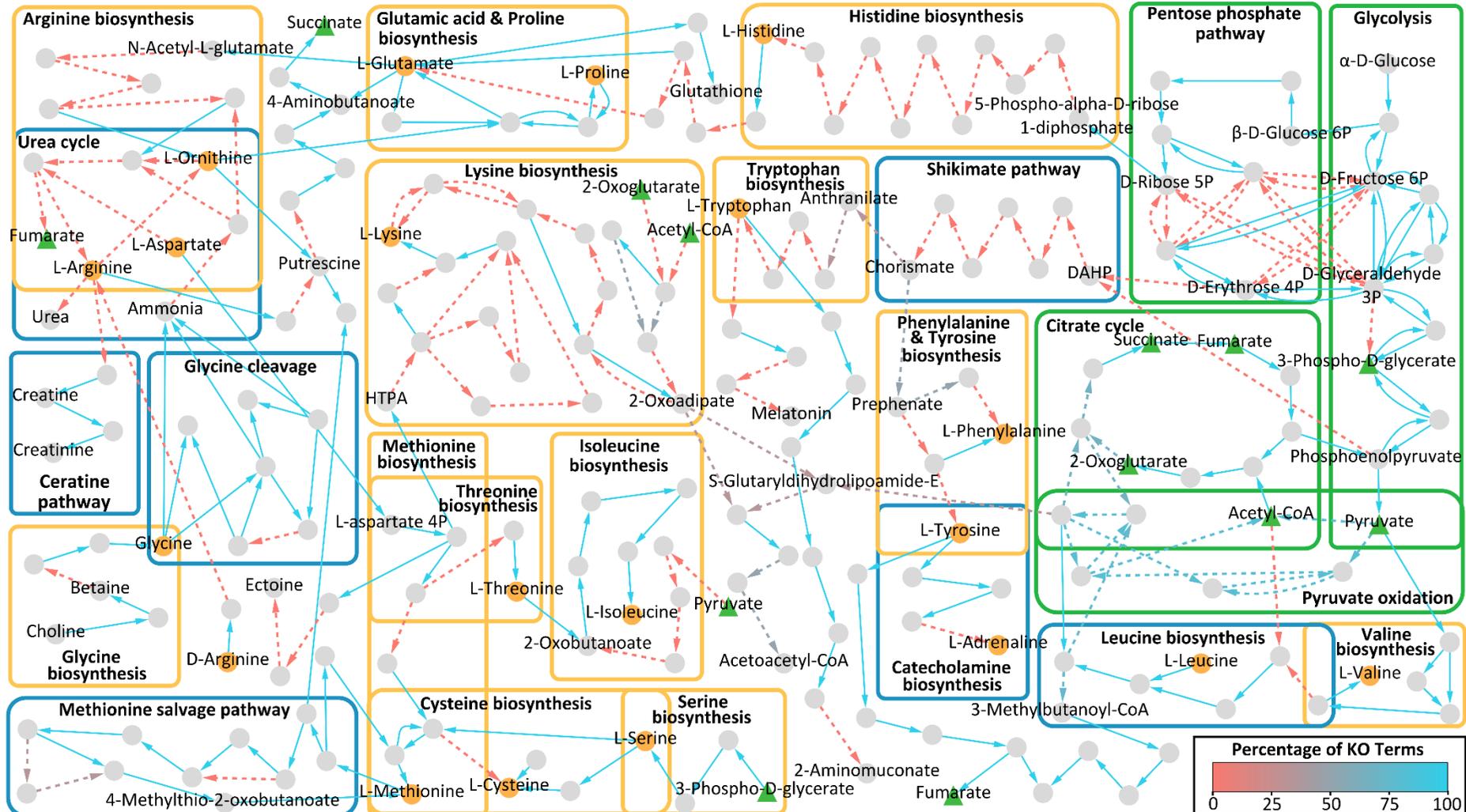
**Supplementary Figure 5: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma minus* (W2).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma minus* (W2) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



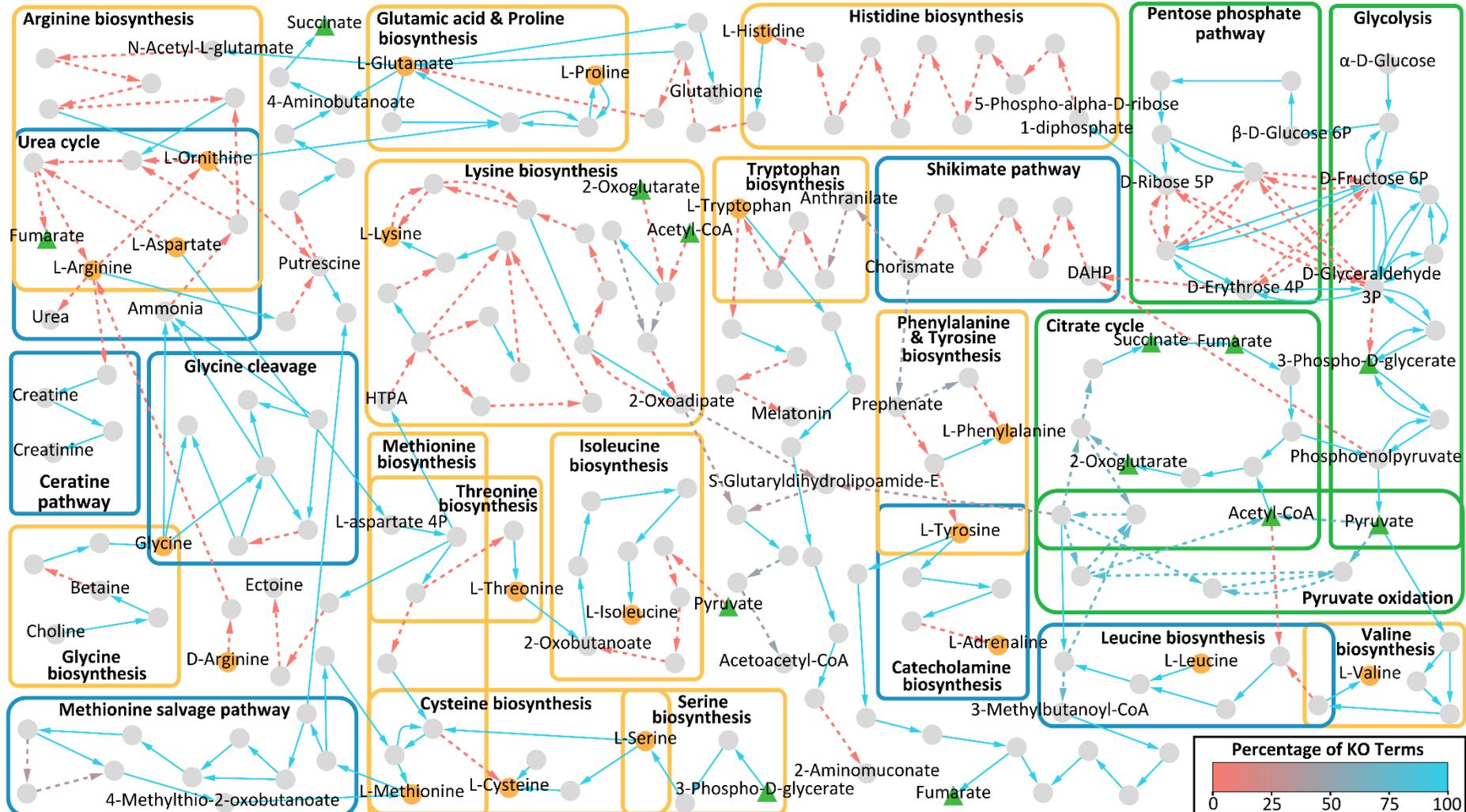
**Supplementary Figure 6: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma kappa* (1A).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma kappa* (1A) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



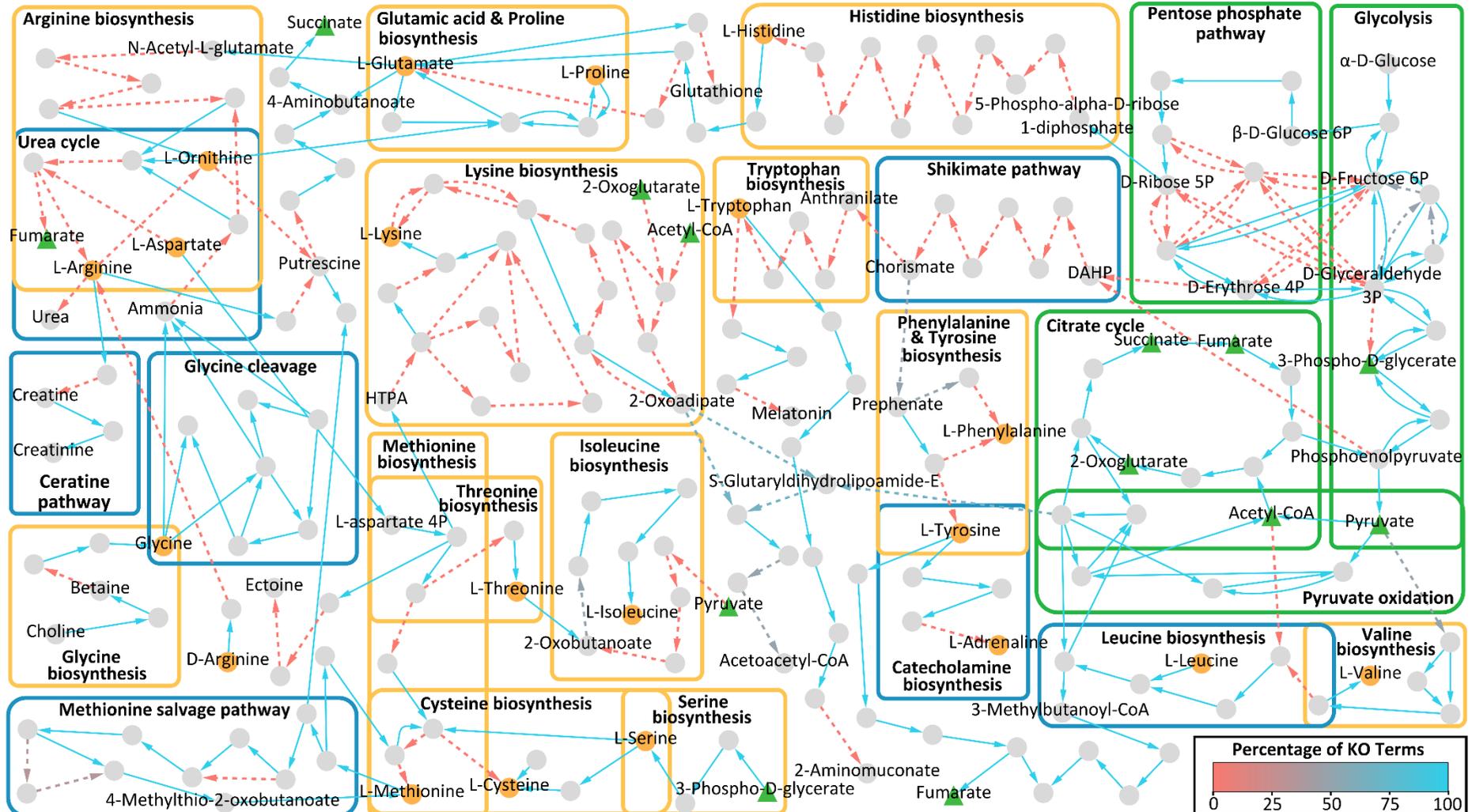
**Supplementary Figure 7: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma karsteni* (3A).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma karsteni* (3A) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



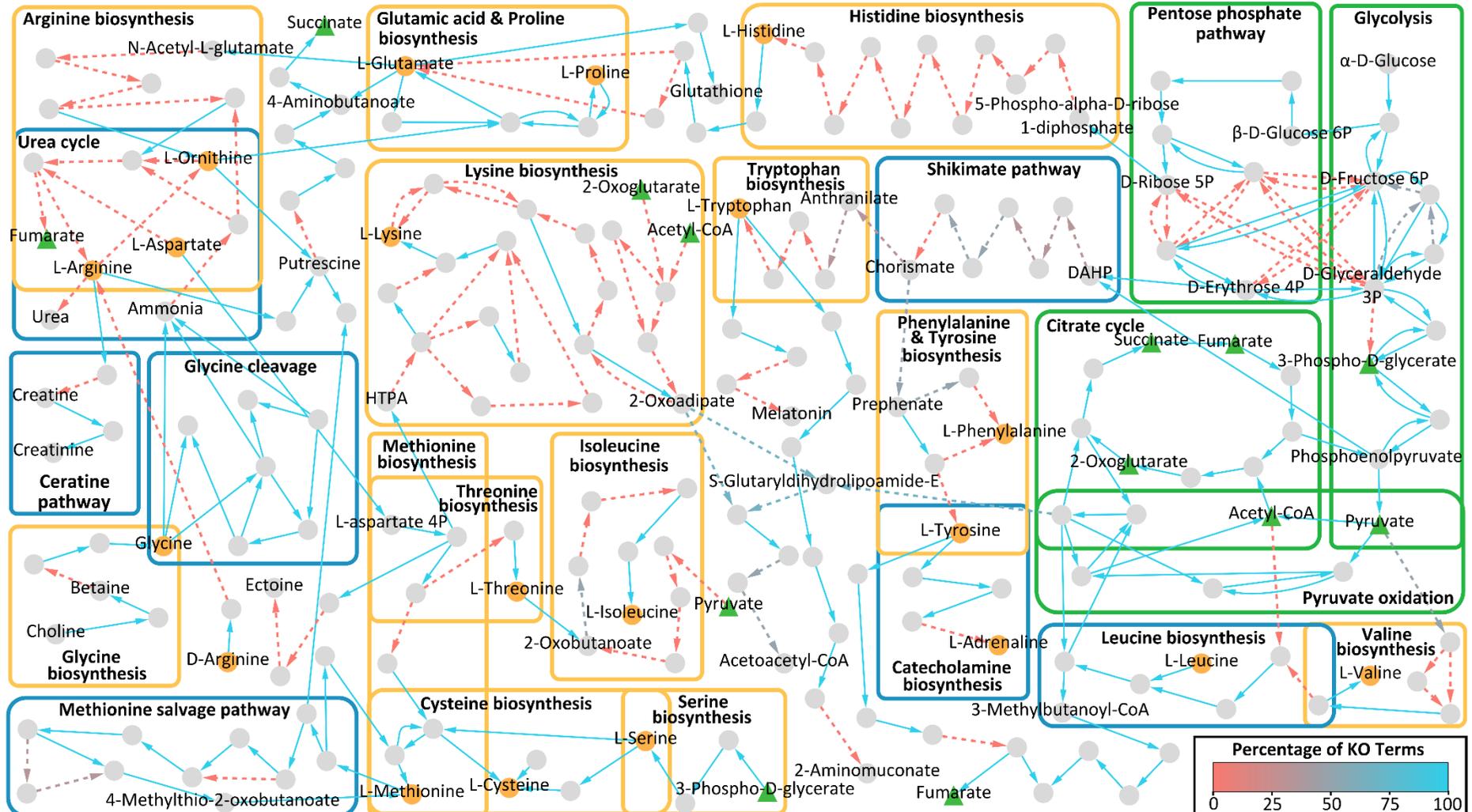
**Supplementary Figure 8: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma tahiri* (B10).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma tahiri* (B10) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



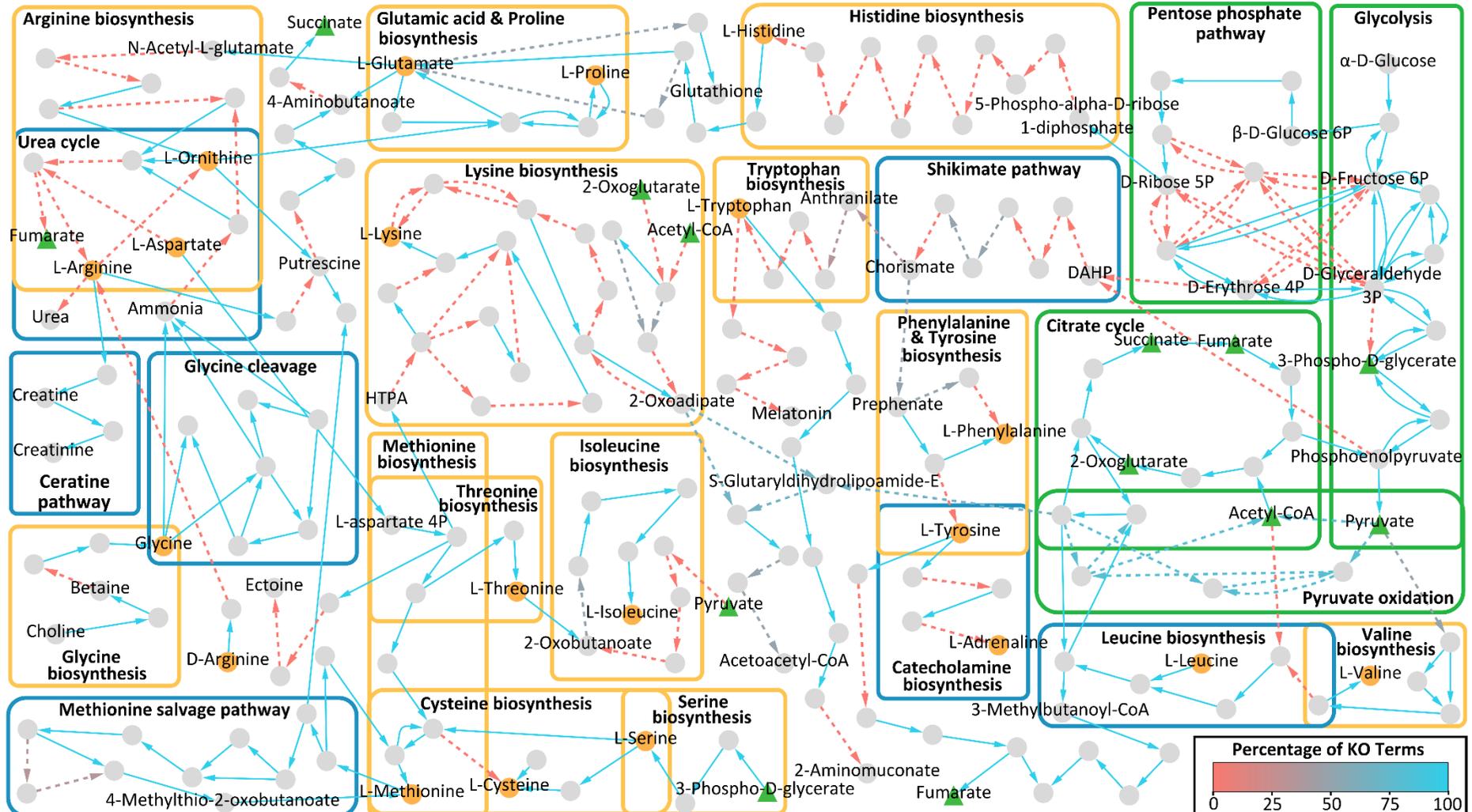
**Supplementary Figure 9: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma floric* (K8).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma floric* (K8) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



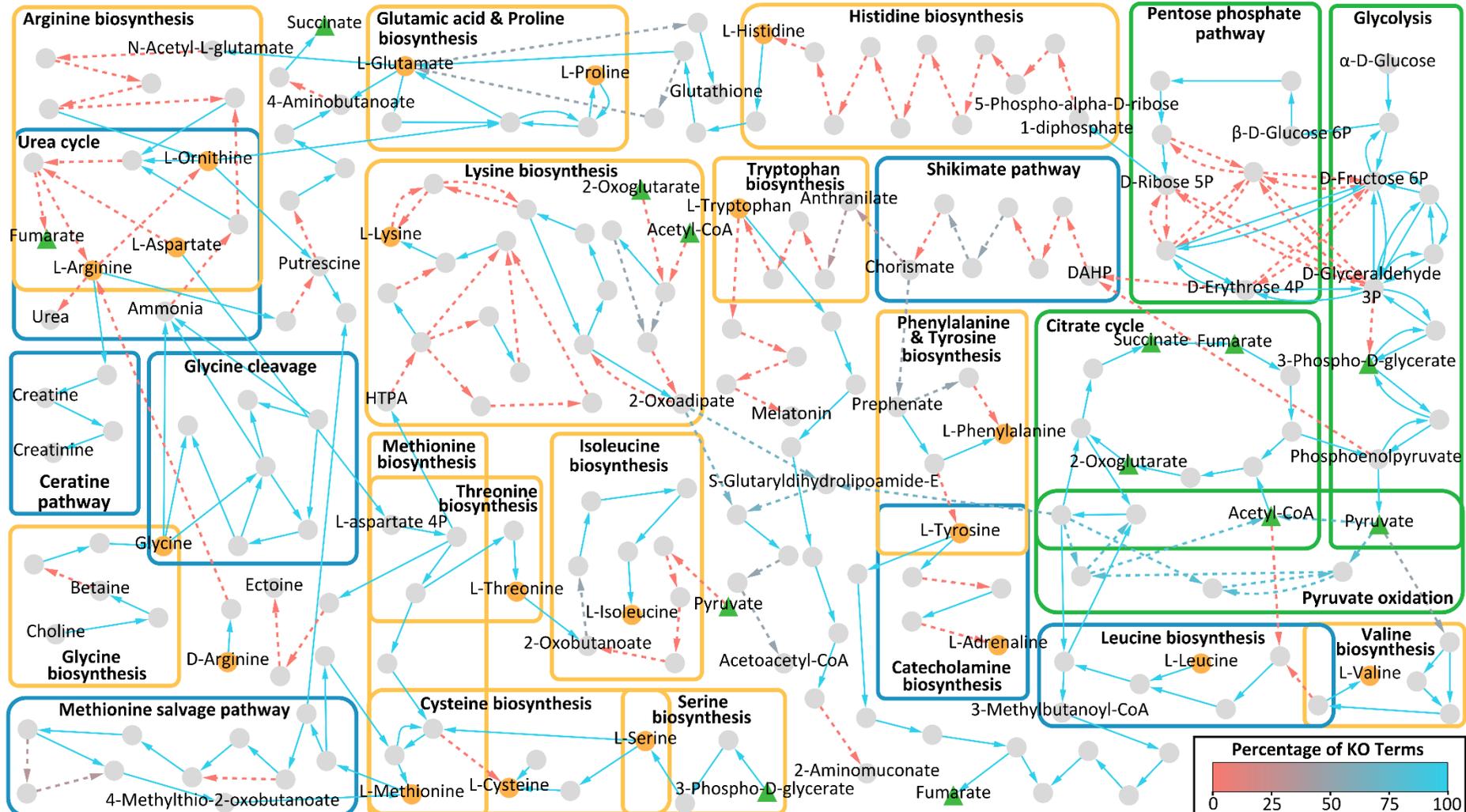
**Supplementary Figure 10: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma pseudocylindrica* (RC).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma pseudocylindrica* (RC) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



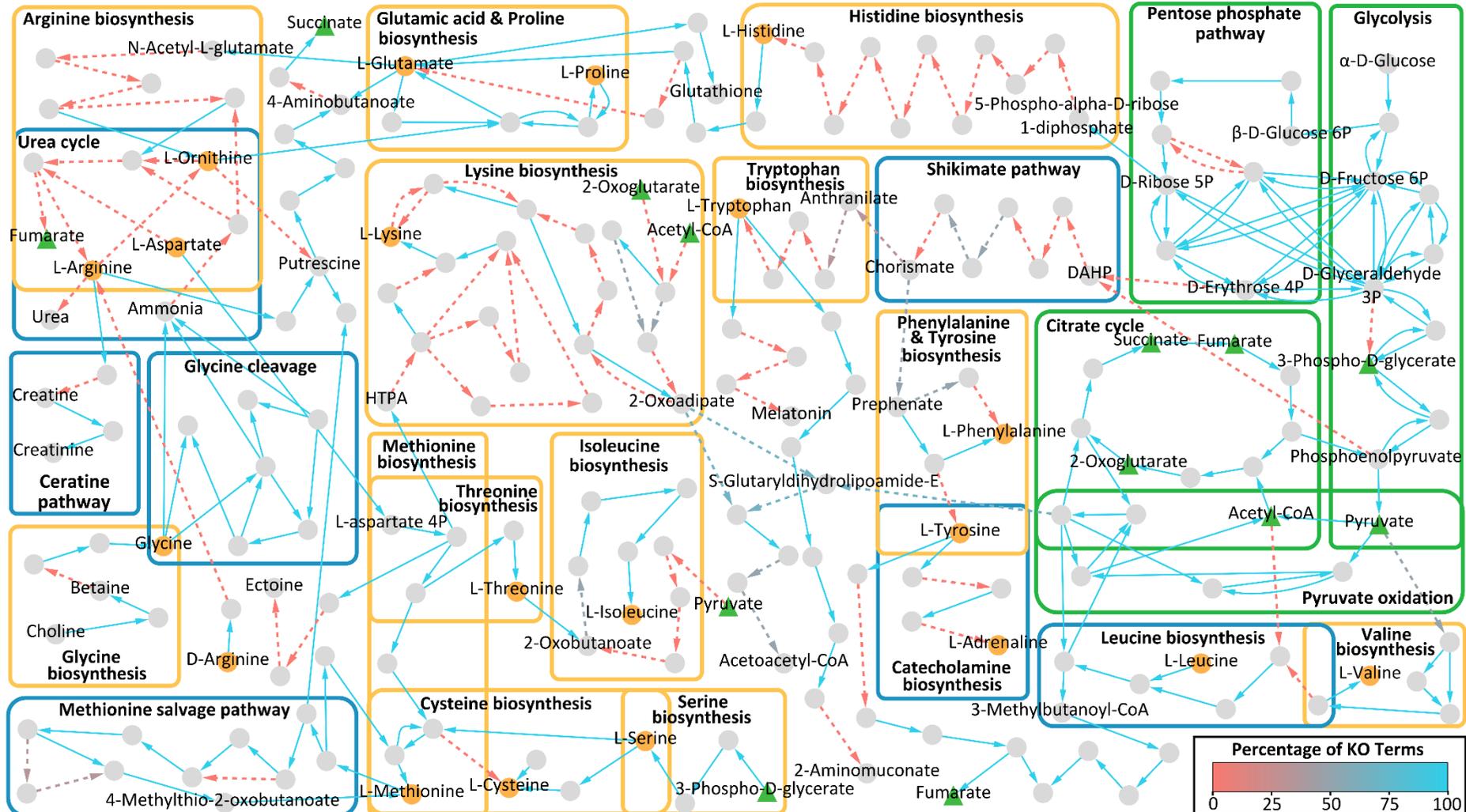
**Supplementary Figure 11: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma sp.* (B3 3 H1).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma sp.* (B3 3 H1) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



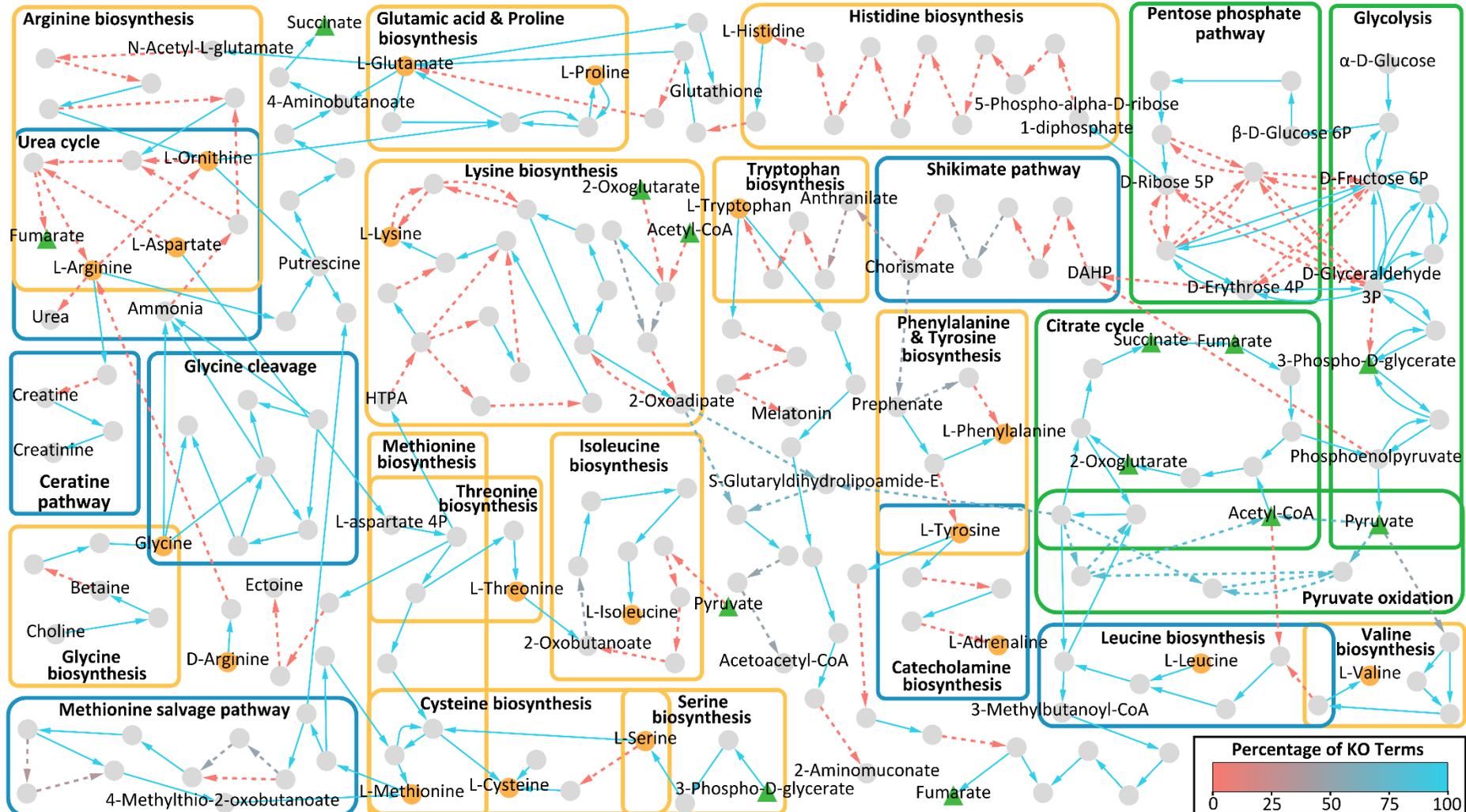
**Supplementary Figure 12: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma sp.* (B4 2 H2).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma sp.* (B4 2 H2) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



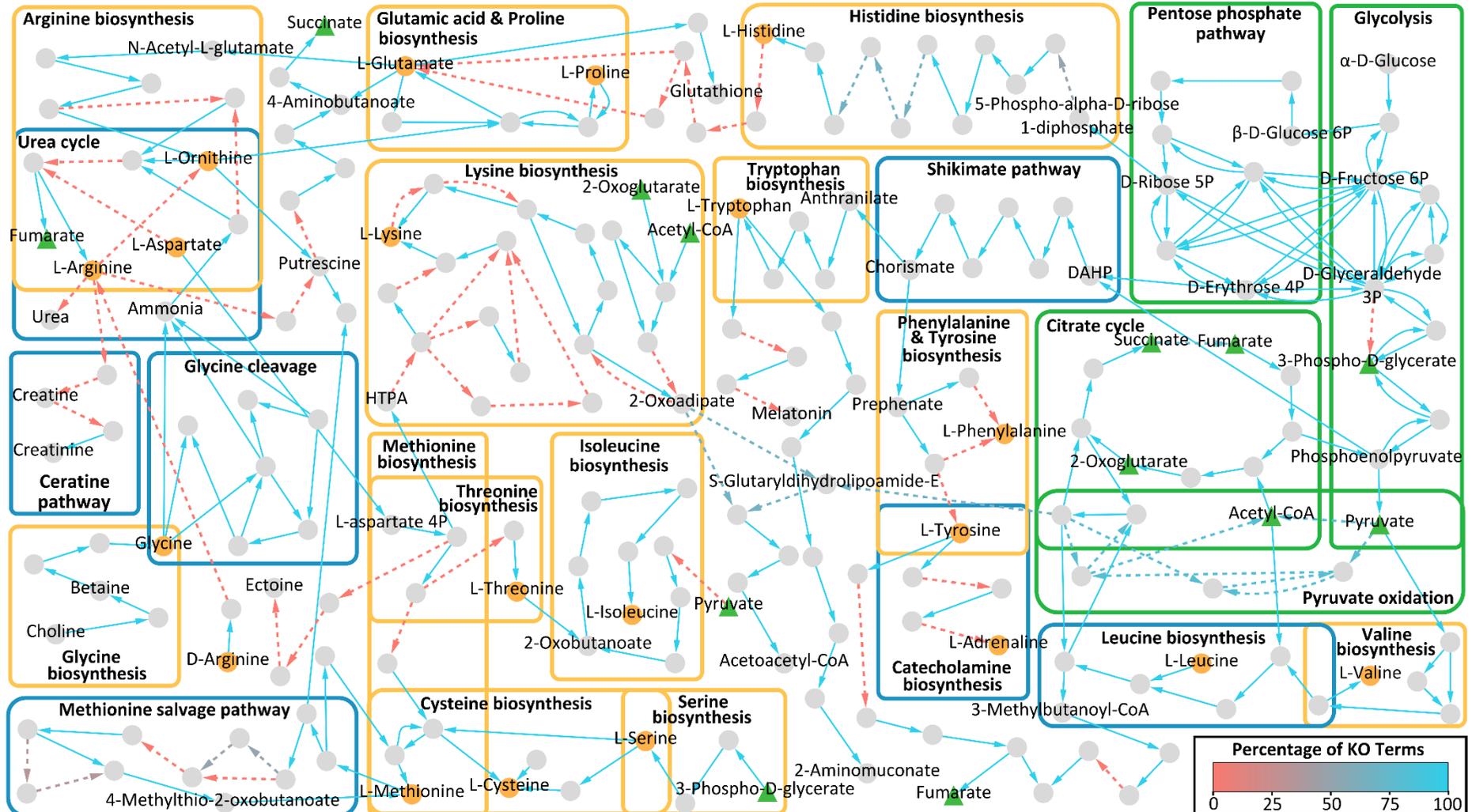
**Supplementary Figure 13: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma schuessleri* (733).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma schuessleri* (733) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



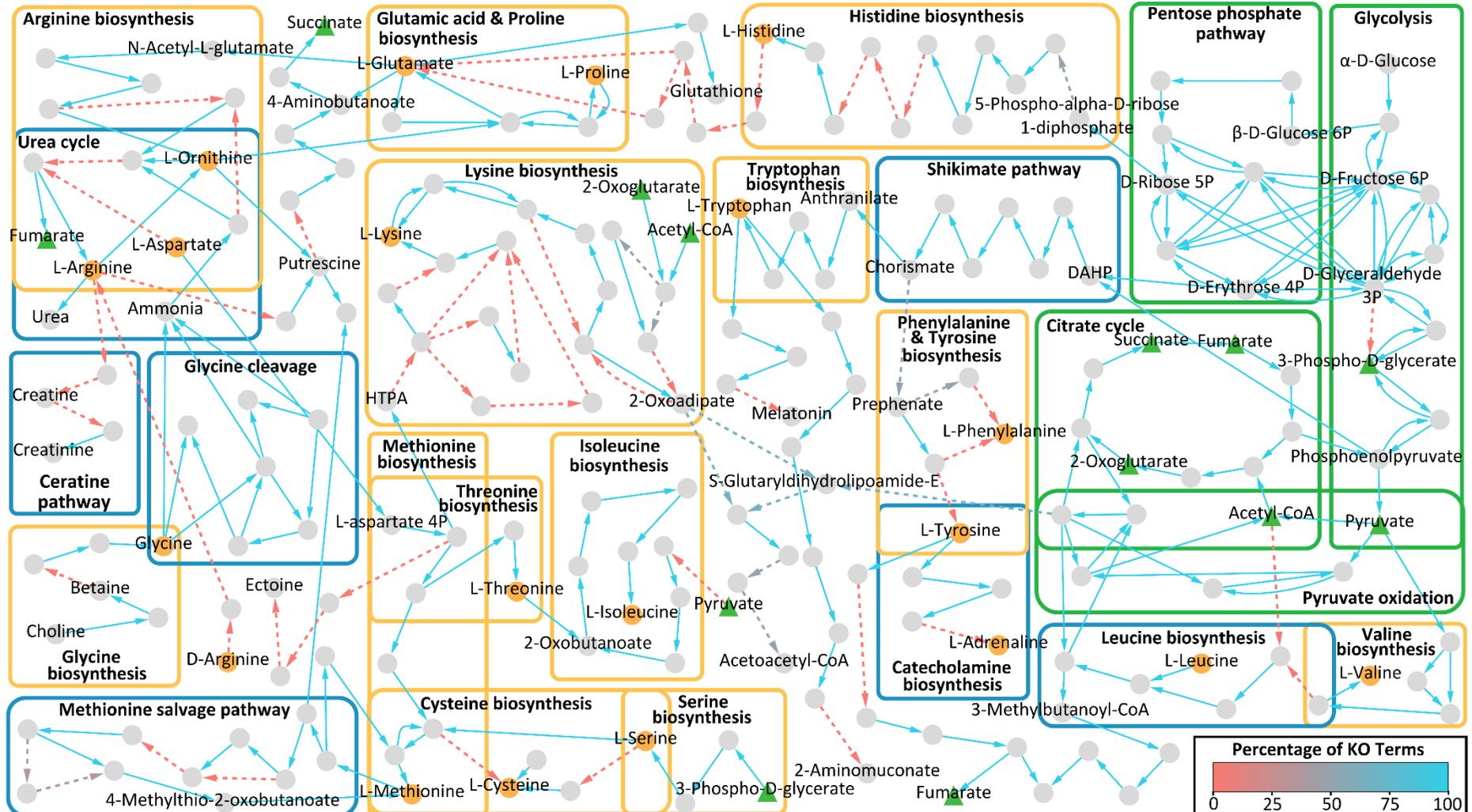
**Supplementary Figure 14: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma schuessleri* (3EH3).** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma schuessleri* (3EH3) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



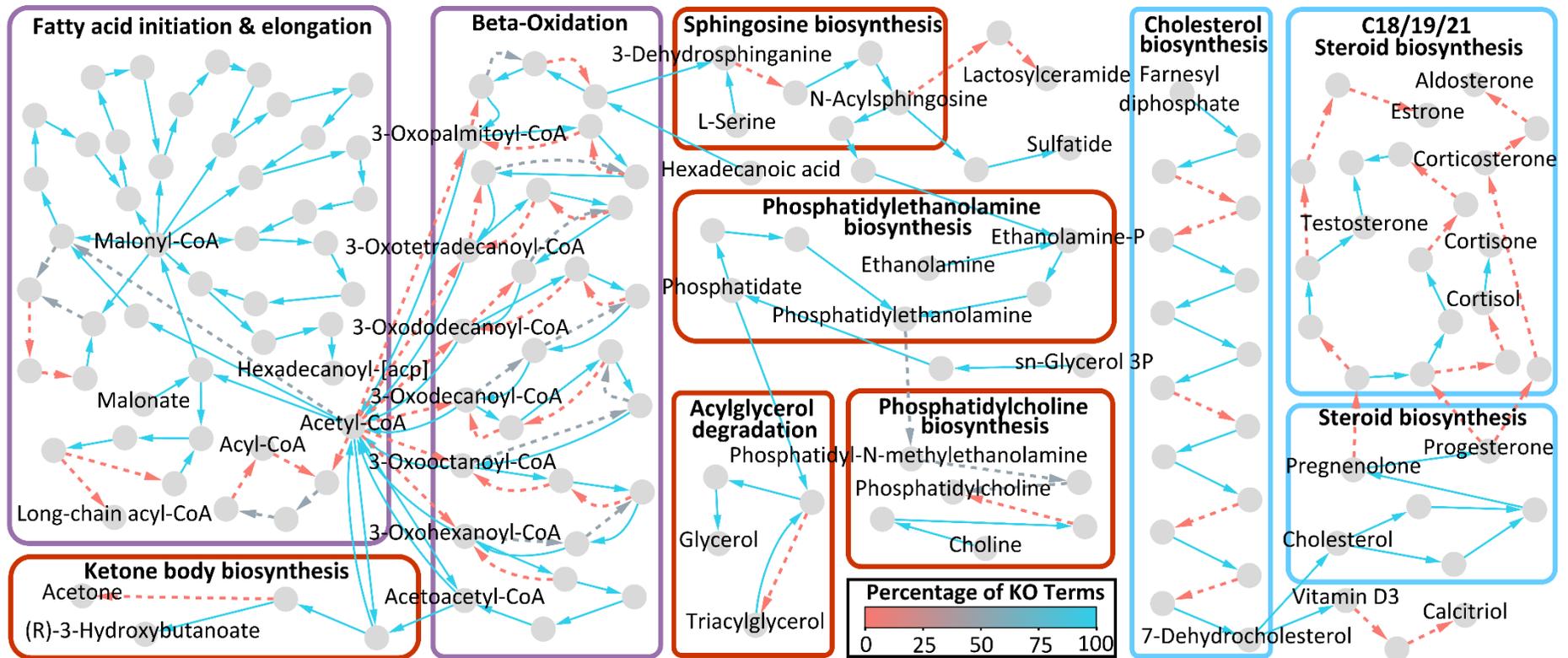
**Supplementary Figure 15: Overview of the central carbohydrate and amino acid metabolism for *Ficculla terrestris*.** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Ficculla terrestris* based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



**Supplementary Figure 16: Overview of the central carbohydrate and amino acid metabolism for *Katarium polorum*.** The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Katarium polorum* based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

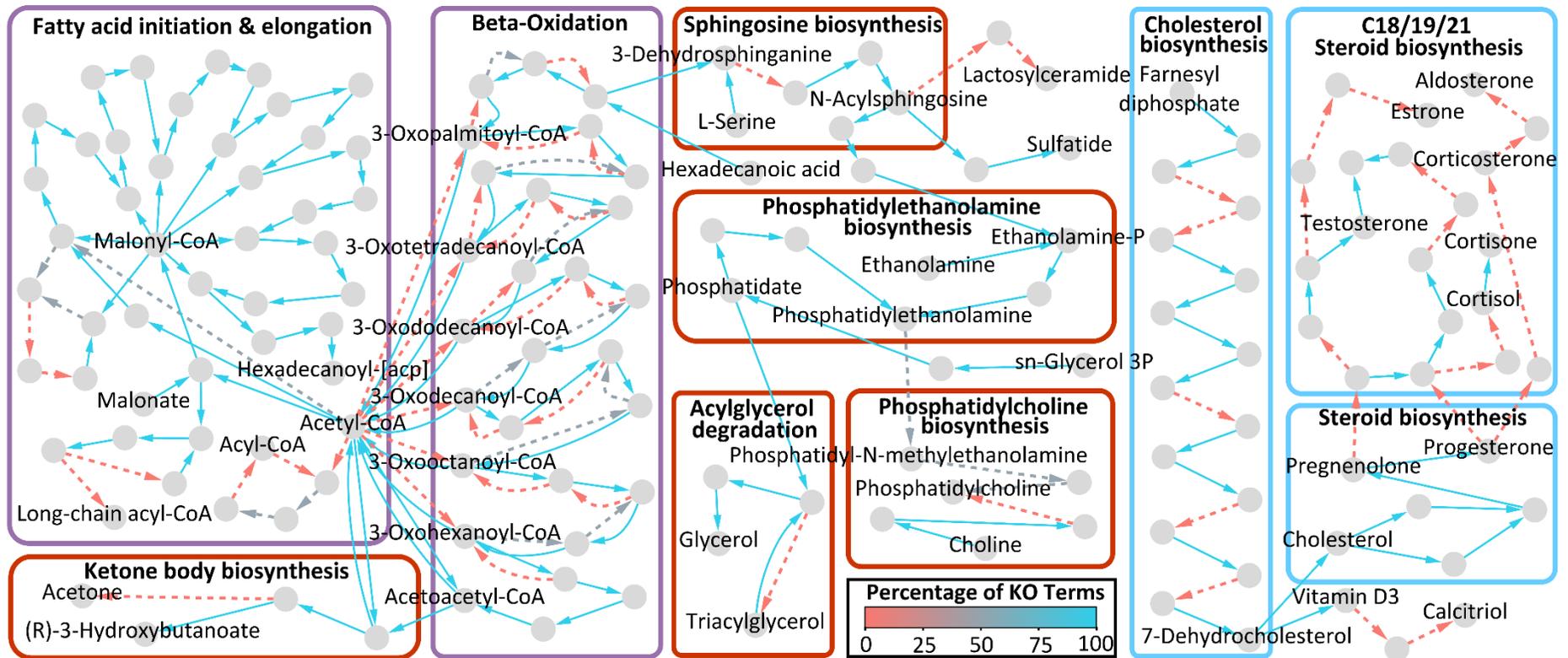
## Chapter 7



**Supplementary Figure 17: Overview of the lipid metabolism for *Rhogostoma epiphylla* (IGS).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma epiphylla* (IGS) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

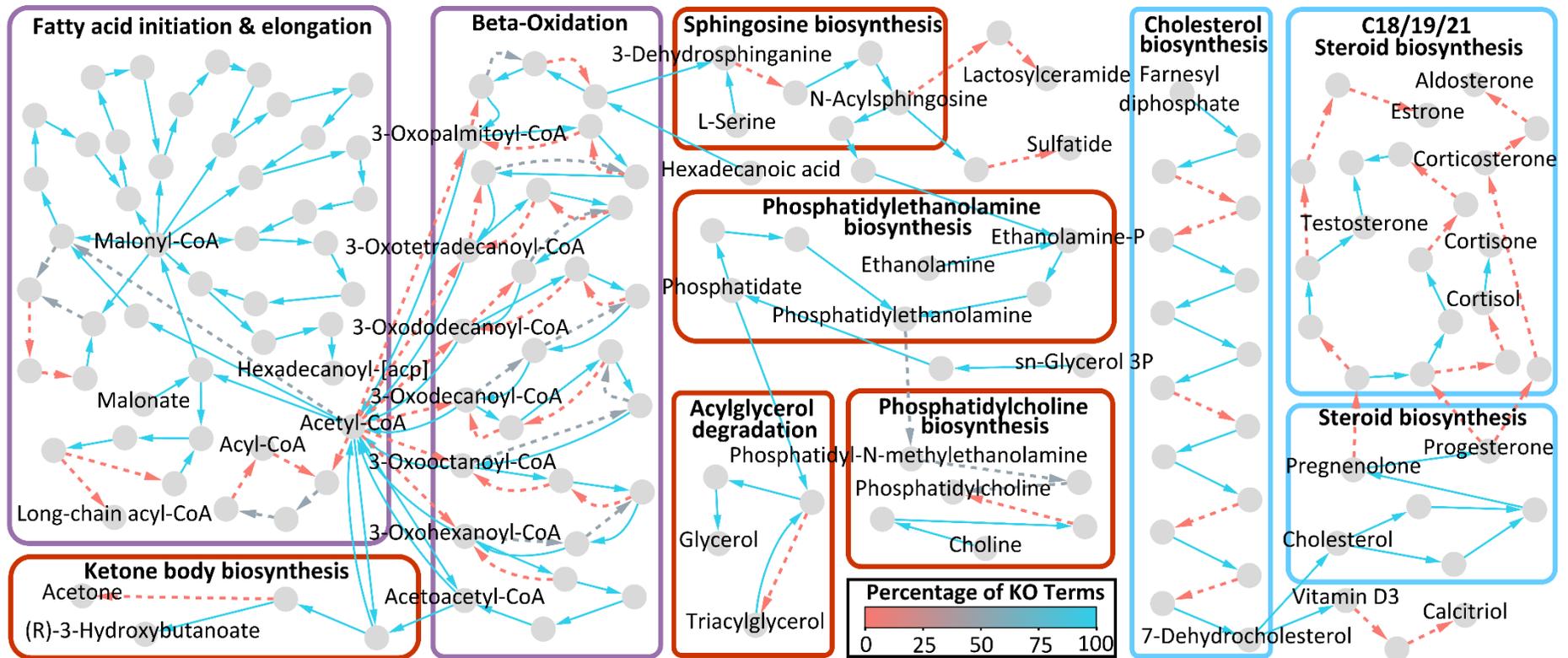


## Chapter 7



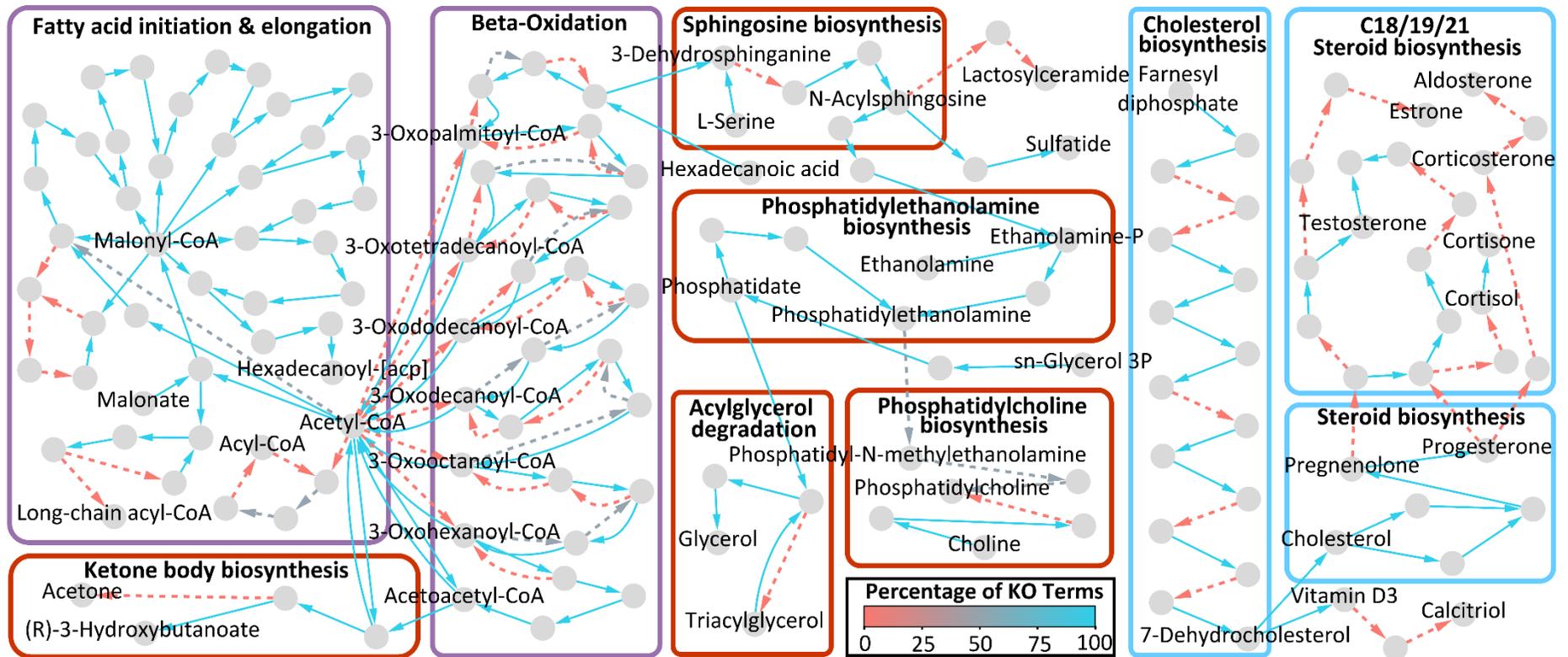
**Supplementary Figure 19: Overview of the lipid metabolism for *Rhogostoma kappa* (1A).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma kappa* (1A) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



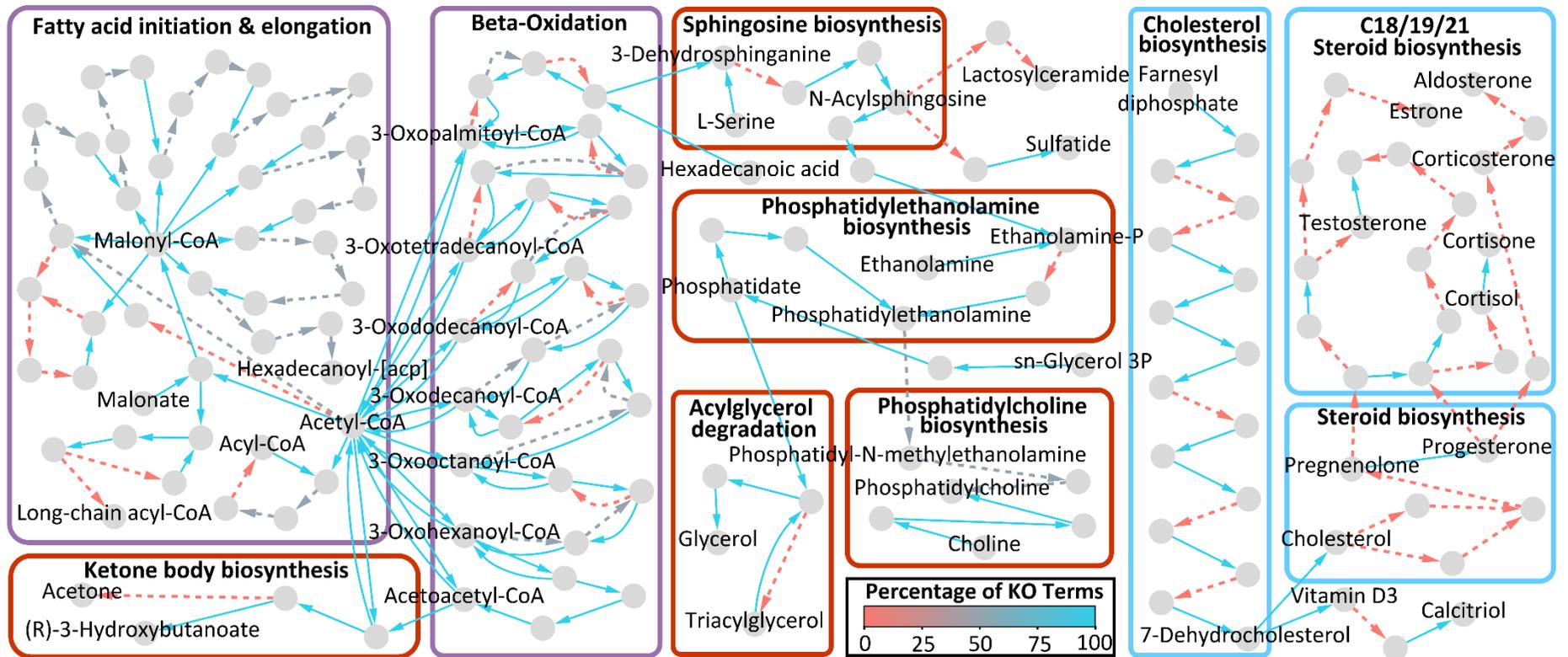
**Supplementary Figure 20: Overview of the lipid metabolism for *Rhogostoma karsteni* (3A).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma karsteni* (3A) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



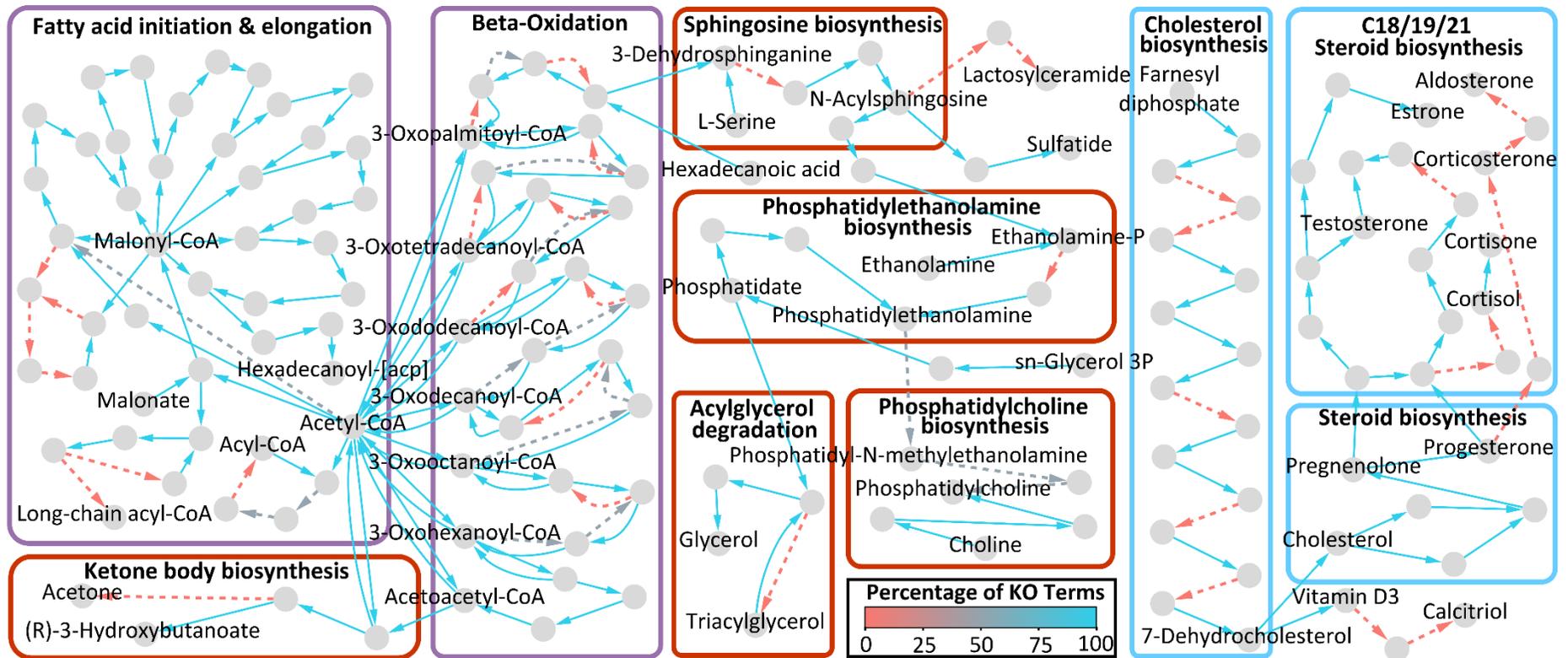
**Supplementary Figure 21: Overview of the lipid metabolism for *Rhogostoma tahiri* (B10).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma tahiri* (B10) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



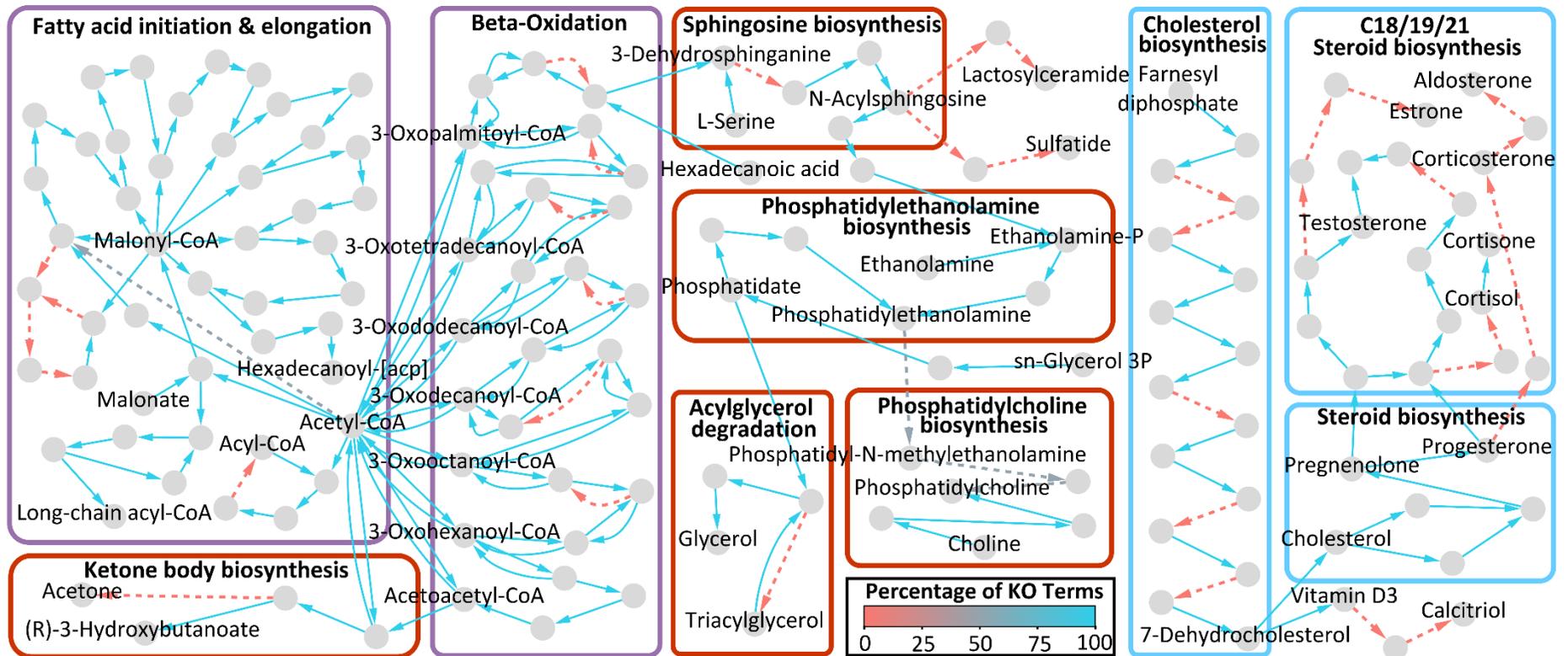
**Supplementary Figure 22: Overview of the lipid metabolism for *Rhogostoma florum* (K8).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma florum* (K8) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



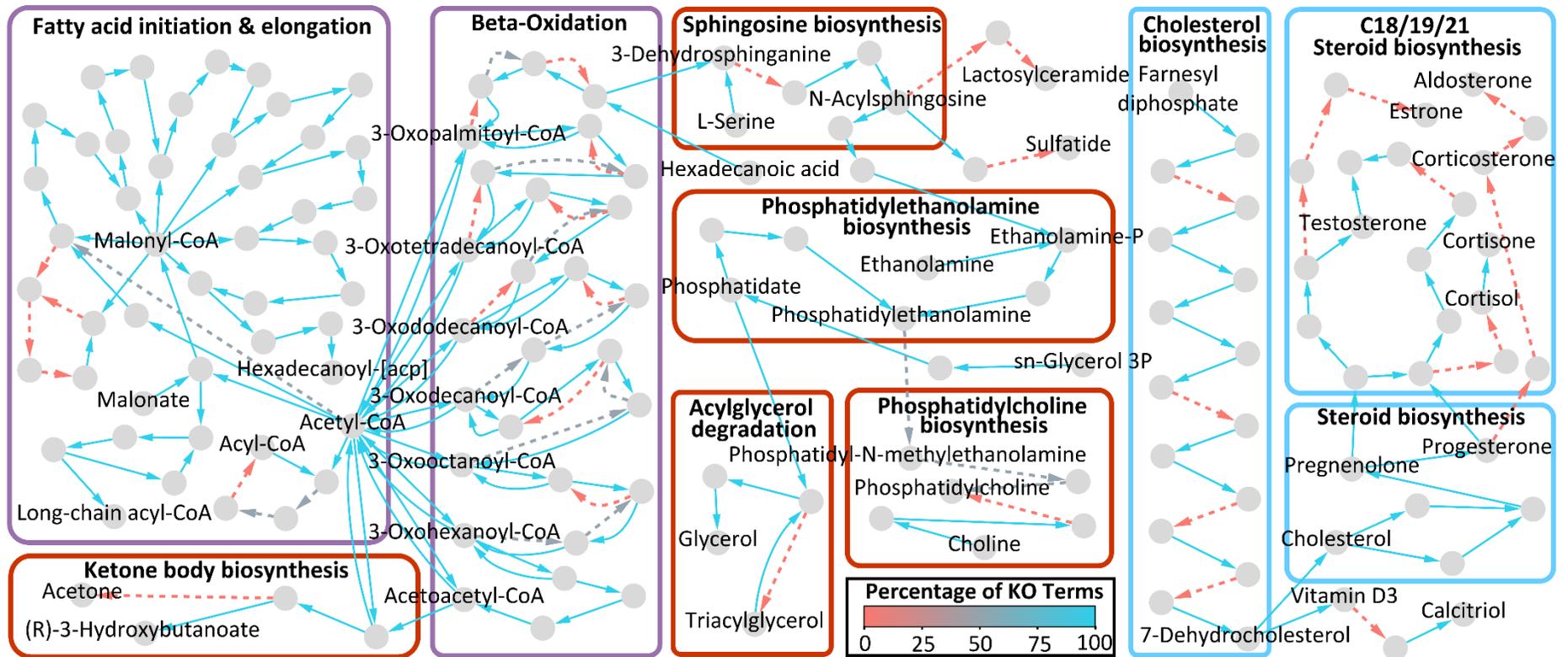
**Supplementary Figure 23: Overview of the lipid metabolism for *Rhogostoma pseudocylindrica* (RC).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma pseudocylindrica* (RC) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



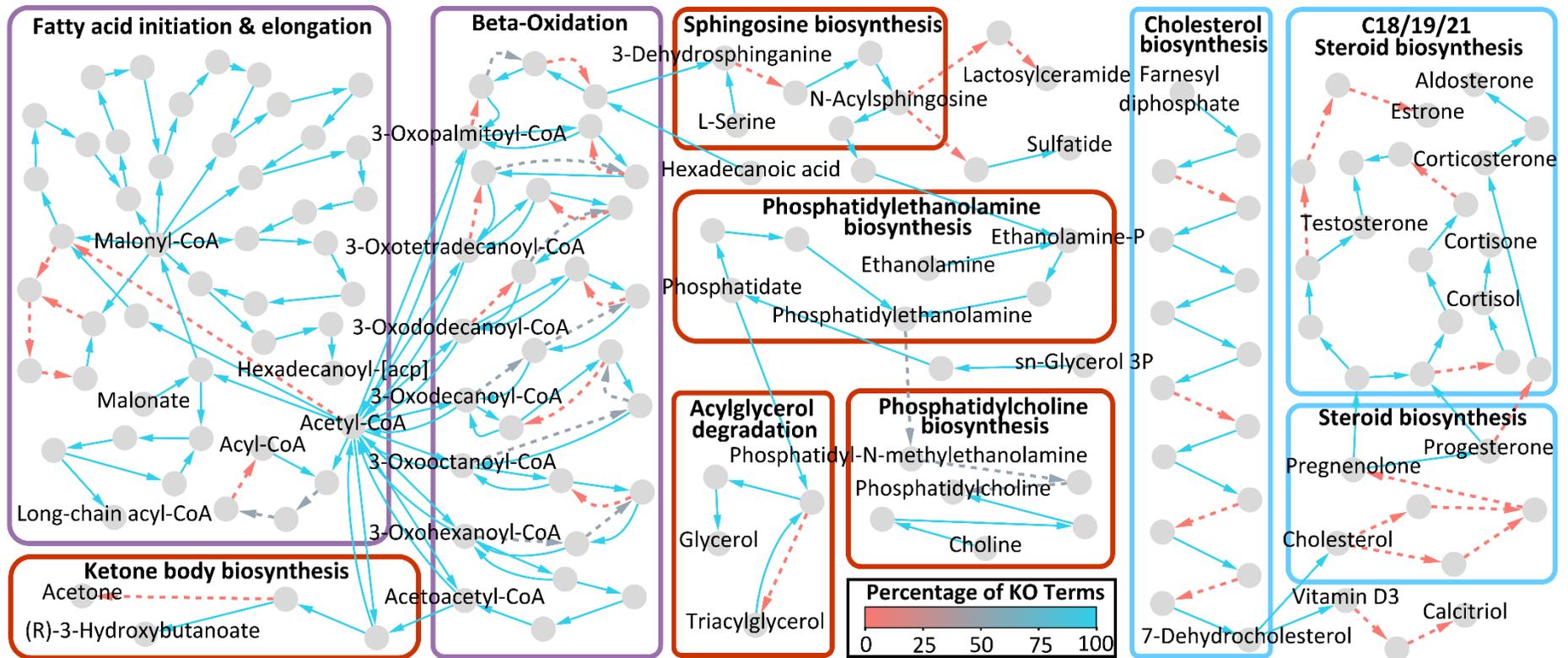
**Supplementary Figure 24: Overview of the lipid metabolism for *Rhogostoma* sp. (B3 3 H1).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma* sp. (B3 3 H1) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



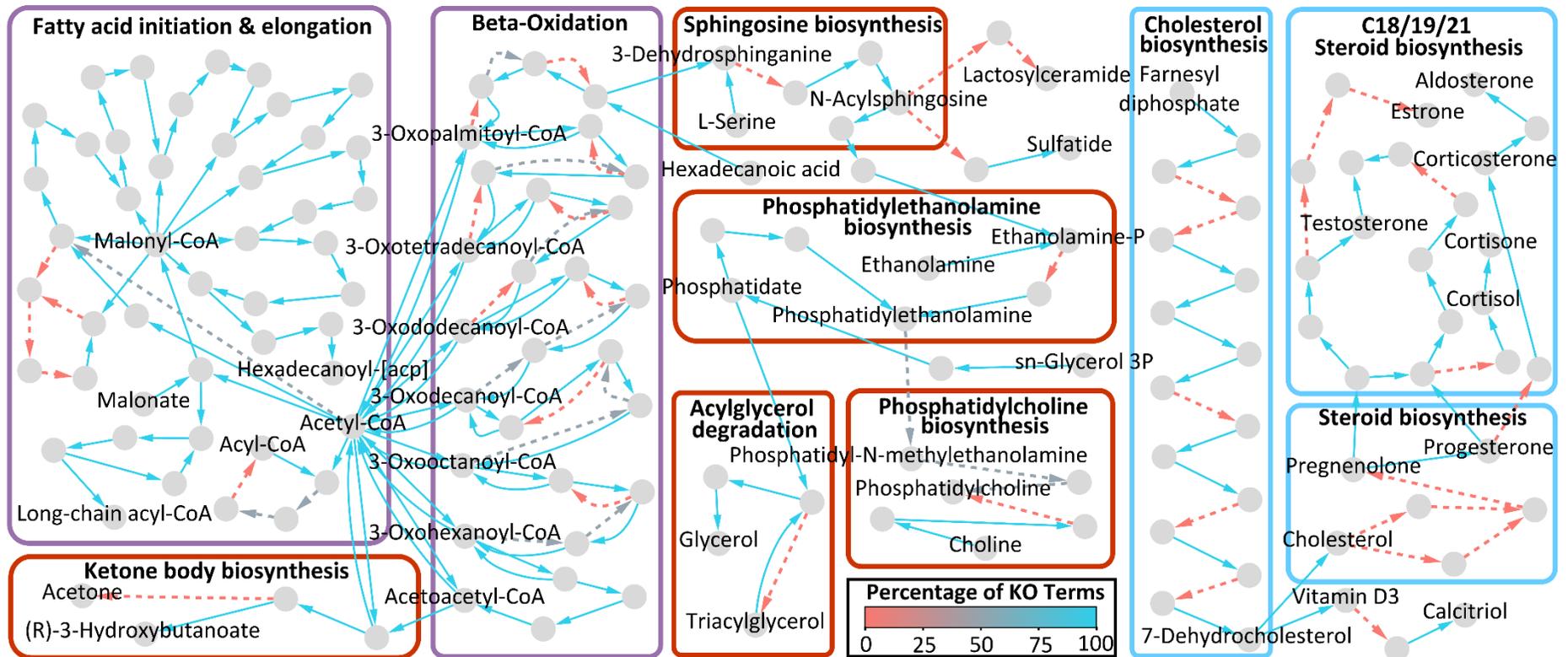
**Supplementary Figure 25: Overview of the lipid metabolism for *Rhogostoma* sp. (B4 2 H2).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma* sp. (B4 2 H2) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



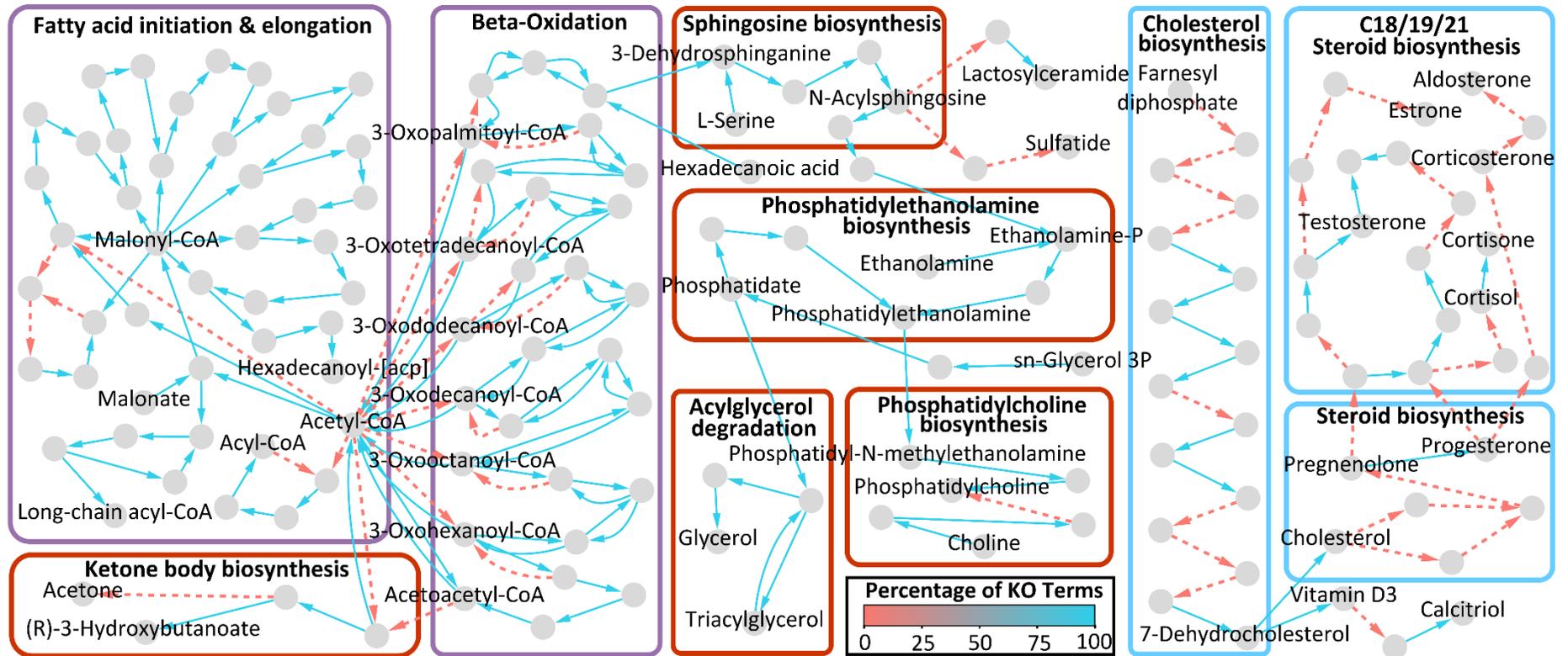
**Supplementary Figure 26: Overview of the lipid metabolism for *Rhogostoma schuessleri* (733).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma schuessleri* (733) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



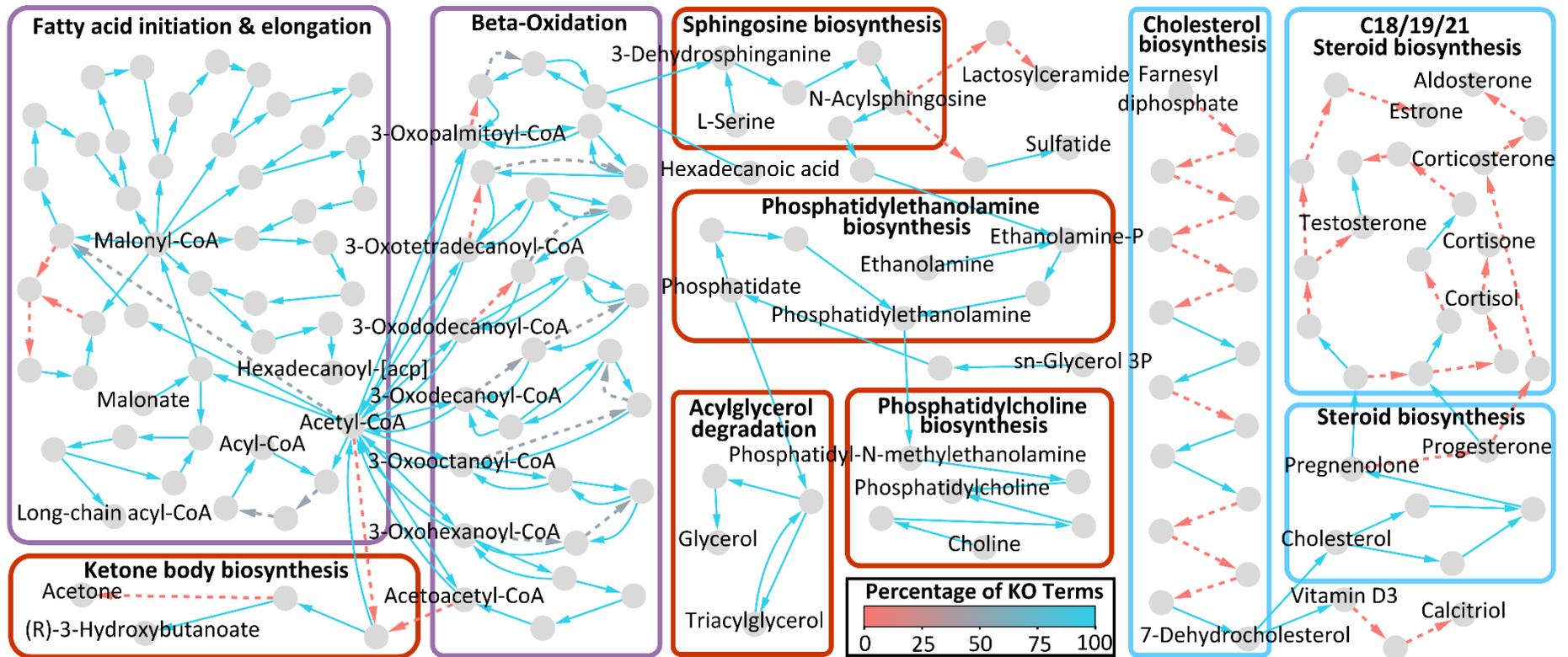
**Supplementary Figure 27: Overview of the lipid metabolism for *Rhogostoma schuessleri* (3EH3).** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma schuessleri* (3EH3) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



**Supplementary Figure 28: Overview of the lipid metabolism for *Ficculla terrestris*.** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Ficculla terrestris* based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

## Chapter 7



**Supplementary Figure 29: Overview of the lipid metabolism for *Katarium polorum*.** The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Katarium polorum* based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

---

## CONCLUSIVE SUMMARY AND GENERAL DISCUSSION

### Biotic and abiotic factors shaping the microbial food web across ecosystems

This thesis provides a comprehensive overview of the microbial food webs of wastewater treatment plants (Chapters 1 and 3), the maize rhizosphere (Chapter 2), alpine soils (Chapter 4), and tree bark in canopies (Chapter 5), exploring and disentangling the biotic and abiotic factors that drive community assembly across distinct ecosystems. We hypothesized that biotic factors, such as predation, competition, and symbiosis, strongly shape microbial food webs and that biotic factors are as important as abiotic factors for microbial community assembly.

We identified patterns indicating that the entire microbial community composition is strongly influenced by predation across all ecosystems using tools such as co-occurrence network analysis (Chapters 1-5). We showed that predatory protists strongly shape the prokaryotic community composition, confirming extensive feeding experiments (Flues et al., 2017; Glücksman et al., 2010; Rosenberg et al., 2009). For example, in the maize rhizosphere (Chapter 2), the community structure of cercozoans (protists) exhibited a strong influence on prokaryotic community turnover, comparable to the influence of the different plant root zones. Further, the expression patterns of root genes involved in immune response were associated with the microbial taxa, indicating molecular control points in the roots. Predation, however, was not restricted to predatory protists but also included predatory bacteria and microscopic metazoans. In the soil microbial food webs, for instance, predatory bacteria were the main consumers (Chapter 4), as implied by recent studies by Petters et al. (2021) and Groß et al. (2023). In contrast, predatory bacteria were less dominant on canopy bark surfaces, indicating their lower importance in this environment (Chapter 5). In wastewater treatment plants, we linked ciliates (protists) and rotifers (microscopic metazoans), known predators in this environment (Lapinski and Tunnacliffe, 2003; Pauli et al., 2001), to parasite removal (Chapter 1). Whereas it is well-recognized that wastewater treatment plants are hotspots for potential parasitic bacteria (Cai et al., 2014; Huang et al., 2018; Xue et al., 2019), we additionally revealed a surprising diversity and relative abundance of potential parasitic protists, which were strongly reduced in the early stage of wastewater treatment. Although the decrease of potential parasitic protists could be partly explained by abiotic factors such as changing chemical conditions, flocculation, or sedimentation (Arregui et al., 2010; Cai et al., 2014; Rodríguez et al., 2015; Walczyńska et al., 2018), we propose that predation greatly contributes to their reduction as well.

## Conclusive Summary and General Discussion

---

On canopy bark surfaces, we disentangled biotic factors, such as symbioses and competition, from abiotic factors, such as UV radiation and desiccation, shaping the microbial food webs (Chapter 5). We propose that abiotic factors were the main drivers for microbial community assembly on smooth bark surfaces. Comparing maple trees with smooth bark surfaces – which provide less protection against environmental stressors such as UV radiation (Arrigoni et al., 2018; Buck et al., 1998) - to acer and linden trees with rough bark surfaces, maple trees exhibited a depauperate microbiome enriched in desiccation-tolerant algal (*Chloroidium*, *Trebouxia*, *Stichococcus*, *Coccomyxa*) and bacterial (*Sphingomonas*) genera (Darienko et al., 2010; Freystein et al., 2008; Lüttge and Büdel, 2010; Romani et al., 2019), as well as UV-protected bacterial (Chitinophagaceae) and protistan (Myxogastria) taxa (Fiore-Donno et al., 2005; Krieg et al., 2010). In contrast, biotic factors were shown to primarily drive microbial community assembly on rough bark surfaces, as indicated by a strong co-exclusion of algae and fungi with bacteria, together with a high proportion of lichen-forming taxa. This contradicts aquatic systems, where bacteria often greatly benefit from algae primary production (Amin et al., 2015; Grossart and Simon, 2007).

Seasonality has been shown to strongly shape microbial food webs across ecosystems (Herold et al., 2020; Shen et al., 2021; Walden et al., 2021). We found that, indeed, a combined effect of biotic and abiotic factors drives seasonal variation in the microbial food webs of wastewater treatment plants and alpine soils (Chapters 3 & 4). In wastewater treatment plants, clear seasonal variations were found for both prokaryotes and eukaryotes (Chapter 3). However, only variations in the eukaryotic, but not in the prokaryotic community, could be directly linked to the temperatures in the treatment tanks, which fluctuated seasonally with the ambient temperature. Instead, we concluded that the treatment tank temperatures affected the eukaryotic community, which, in turn, shaped the prokaryotic community through predation. Similarly, consumer pressure also shaped the prey community in alpine soils (Chapter 4). We discovered that the predation pressure increased from spring to summer, leading to increased diversity and evenness, thus contributing to the resilience of the microbial community (Wittebolle et al., 2009).

This thesis demonstrates that metatranscriptomics is a promising method for investigating complex microbial food webs in various ecosystems (Chapters 1, 3, 4, 5). The strength of metatranscriptomics lies in the simultaneous assessment of the entire diversity of the microbial food web, including both prokaryotes and eukaryotes (Urich et al., 2008). In addition, metatranscriptomics circumvents biases inherent to primer-based approaches using “universal eukaryotic primers” as evident in the reliable detection of, for example, amoebozoans (Fiore-Donno et al., 2016; Geisen et al., 2015a; Lentendu et al., 2014) or “long-branch organisms” such as *Giardia* (Schuler et al., 2018; Zahedi et al., 2019). However, metatranscriptomics also has limitations, particularly the rather short sequence length that usually restricts the taxonomic assignment to the genus level. This limitation can be overcome by the

## Conclusive Summary and General Discussion

---

(additional) use of primer-based approaches with taxon-specific primers (Chapter 2). These sequencing approaches, along with the established methodological workflows, for example, for robust co-occurrence network inference accounting for the nature of environmental sequencing data, allow a comprehensive assessment of the entire complex microbial food web. We highlighted the importance of investigating both the prokaryotic and eukaryotic communities for analyzing microbial food webs by demonstrating, for example, the strong impact of predatory bacteria, protists, and microscopic metazoans on microbial food webs, as well as the complex interplay between biotic and abiotic factors that is only revealed when the entire food web is considered.

### Unveiling the functional diversity of protists

One of the main limitations of high-throughput sequencing studies on protist diversity is the lack of functional trait databases and thus, their convenient integration. However, the assessment of functional traits is essential to fully explore the biotic and abiotic factors shaping microbial food webs (Bouskill et al., 2012; Krause et al., 2014). We hypothesized that protists exhibit high variability in functional traits, even at the strain level, due to their remarkable morphological diversity and evolutionary divergence (Adl et al., 2019; Burki et al., 2020; Geisen et al., 2020). Furthermore, we hypothesized that distinct (functional) groups are differently affected by biotic and abiotic factors.

In this thesis, we addressed the lack of functional trait databases for protists by providing a database for Amoebozoa (Chapter 6), a dominant protist group across all ecosystems (Chapters 1, 3, 4, 5), facilitating future functional studies that will improve our understanding of the diversity of Amoebozoa adaptations to their environment and their role in microbial food webs. The database comprises traits for habitat, locomotion, nutrition, morphology, size, whether spore formation has been observed, and whether they may cause diseases, emphasizing the immense diversity of Amoebozoa. A comparative analysis of the functional traits of Amoebozoa and Cercozoa (Rhizaria) (Dumack et al., 2020) revealed both convergent evolution and striking differences (Chapter 6): Both groups showed similar morphological variation, for instance, but differences in feeding behavior, i.e., the feeding preference in combination with size indicated that amoebozoans taxa are less selective in their prey choice than Cercozoa, but the size of the amoebozoan cell determines the prey size (Kulishkin et al., 2023).

We furthermore showed that biotic and abiotic factors shape the functional community composition of protists. On a large scale, the functional community composition of Amoebozoa and Cercozoa varied between distinct ecosystems (Chapter 6), i.e., alpine soils (Chapter 4), canopy bark surfaces (Chapter 5) leaf litter (Voss et al., 2019). For example, disease-related Amoebozoa were more prevalent in soils than on bark surfaces or litter. Additionally, the overall variability of the functional traits among ecosystems was lower for Cercozoa compared to Amoebozoa communities. However, variations in the

## Conclusive Summary and General Discussion

---

functional community composition of protists were also shaped by biotic and abiotic factors at finer scales. Indeed, the microbial food web, including protists, differed significantly between tree species (Chapter 5), although these differences were not uniform but varied between microbial groups. Only by considering functional traits such as resilience, bark topology in combination with environmental stressors could be identified as potential shaping forces causing heterogeneous differences in the microbial food web of tree species with smooth compared to rough bark surfaces due to a higher proportion of desiccation-tolerant or pigmented (resilient) taxa on smooth bark surfaces. Similarly, in alpine soils (Chapter 4), the interplay between environmental conditions and biotic interactions could only be disentangled by assigning functional guilds.

Beyond community-level analyses, exploring the physiological traits of individual taxa can improve our understanding of their influence on microbial food webs and thus, on ecosystem functioning. To this end, we established a workflow for exploring the physiological traits of individual protist strains, including multi-gene phylogenetic analysis and whole *de novo* transcriptome comparison across species without reference genomes, to place the diversification in an evolutionary context (Chapter 7). By comparing the transcriptomes of Thecofilosea (Rhizaria), we uncovered surprising variations in their physiological traits, even among closely related strains. These variations became even more pronounced with evolutionary distance. Most variation was observed in amino acid and lipid metabolism, indicating that even closely related strains might differ in their functional traits. However, conserved genes and pathways, for example, related to nucleotide metabolism (Yao and O'Donnell, 2016), were robustly identified, emphasizing the soundness of the established methods.

The established methods and tools of this thesis will allow future studies to go beyond taxonomic analyses and additionally address the functional diversity of protists, thus answering questions such as whether changes in protist communities affect only the taxonomic or also the functional composition. Considering the vast, yet largely unexplored, diversity of protists, the here provided insights into the functional traits of the Amoebozoa, as well as the comprehensive analysis of the physiological traits of Thecofilosea, give an idea of the remarkable functional diversity of protists that is to be revealed by future studies.

# GENERAL REFERENCES

- Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V., Zhang, Q., 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol* 66, 4–119. <https://doi.org/10.1111/jeu.12691>
- Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., Morales, R.L., Berthiaume, C.T., Parker, M.S., Djunaedi, B., Ingalls, A.E., Parsek, M.R., Moran, M.A., Armbrust, E.V., 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522, 98–101. <https://doi.org/10.1038/nature14488>
- Arregui, L., Pérez-Uz, B., Salvadó, H., Serrano, S., 2010. Progresses on the knowledge about the ecological function and structure of the protists community in activated sludge wastewater treatment plants. *Curr Res Technol Edu Topics Appl Microbiol Microb Biotechnol* 2, 972–979.
- Arrigoni, E., Antonielli, L., Pindo, M., Pertot, I., Perazzolli, M., 2018. Tissue age and plant genotype affect the microbiota of apple and pear bark. *Microbiol Res* 211, 57–68. <https://doi.org/10.1016/j.micres.2018.04.002>
- Baldrian, P., 2017. Forest microbiome: Diversity, complexity and dynamics. *FEMS Microbiol Rev* 41, 109–130. <https://doi.org/10.1093/femsre/fuw040>
- Baldwin, D.S., Colloff, M.J., Rees, G.N., Chariton, A.A., Watson, G.O., Court, L.N., Hartley, D.M., Morgan, M. J., King, A.J., Wilson, J.S., Hodda, M., Hardy, C.M., 2013. Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Mol Ecol* 22, 1746–1758. <https://doi.org/10.1111/mec.12190>
- Bonkowski, M., Dumack, K., Fiore-Donno, A.M., 2019. The protists in soil - a token of untold eukaryotic diversity, in: Elsas, J.D. van, Trevors, J.T., Rosado, A.S., Nannipieri, P. (Eds.), *Modern Soil Microbiology*. CRC press, Boca Raton, USA, pp. 125–140.
- Bouskill, N.J., Eveillard, D., Chien, D., Jayakumar, A., Ward, B.B., 2012. Environmental factors determining ammonia-oxidizing organism distribution and diversity in marine environments. *Environ Microbiol* 14, 714–729. <https://doi.org/10.1111/j.1462-2920.2011.02623.x>
- Bringel, F., Couée, I., 2015. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Front Microbiol* 6, 486. <https://doi.org/10.3389/fmicb.2015.00486>
- Buck, J.W., Lachance, M.A., Traquair, J.A., 1998. Mycoflora of peach bark: Population dynamics and composition. *Canad J Bot* 76, 345–354.
- Burian, A., Pinn, D., Peralta-Maraver, I., Sweet, M., Mauvisseau, Q., Eyice, O., Bulling, M., Röthig, T., Kratina, P., 2022. Predation increases multiple components of microbial diversity in activated sludge communities. *ISME J* 16, 1086–1094. <https://doi.org/10.1038/s41396-021-01145-z>
- Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The new tree of eukaryotes. *Trends Ecol Evol* 35, 43–55. <https://doi.org/10.1016/j.tree.2019.08.008>
- Cai, L., Ju, F., Zhang, T., 2014. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl Microbiol Biotechnol* 98, 3317–3326. <https://doi.org/10.1007/s00253-013-5402-z>
- Cavalier-Smith, T., Chao, E.E., Lewis, R., 2018. Multigene phylogeny and cell evolution of chromist infrakingdom Rhizaria: contrasting cell organisation of sister phyla Cercozoa and Retaria. *Protoplasma* 255, 1517–1574. <https://doi.org/10.1007/s00709-018-1241-1>
- Cavalier-Smith, T., Chao, E.E.-Y., 2003. Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist* 154, 341–358. <https://doi.org/10.1078/143446103322454112>
- Cavalier-Smith, T., Fiore-Donno, A.M., Chao, E., Kudryavtsev, A., Berney, C., Snell, E.A., Lewis, R., 2015. Multigene phylogeny resolves deep branching of Amoebozoa. *Mol Phylogenet and Evol* 83, 293–304. <https://doi.org/10.1016/j.ympev.2014.08.011>
- Corno, G., Caravati, E., Callieri, C., Bertoni, R., 2008. Effects of predation pressure on bacterial abundance, diversity, and size-structure distribution in an oligotrophic system. *J Limnol* 67, 107–119. <https://doi.org/10.4081/jlimnol.2008.107>
- Darienko, T., Gustavs, L., Mudimu, O., Menendez, C.R., Schumann, R., Karsten, U., Friedl, T., Pröschold, T., 2010. *Chloroidium*, a common terrestrial coccoid green alga previously assigned to *Chlorella* (Trebouxiophyceae, Chlorophyta). *Eur J Phycol* 45, 79–95. <https://doi.org/10.1080/09670260903362820>

## General References

---

- de Ruiter, P.C., Moore, J.C., Zwart, K.B., Bouwman, L.A., Hassink, J., Bloem, J., de Vos, J.A., Marinissen, J.C.Y., Didden, W.A.M., Lebbink, G., Brussaard, L., 1993. Simulation of nitrogen mineralization in the belowground food webs of two winter wheat fields. *J Appl Ecol* 30, 95–106. <https://doi.org/10.2307/2404274>
- de Ruiter, P.C., Neutel, A.-M., Moore, J.C., 1998. Biodiversity in soil ecosystems: The role of energy flow and community stability. *Appl Soil Ecol* 10, 217–228. [https://doi.org/10.1016/S0929-1393\(98\)00121-8](https://doi.org/10.1016/S0929-1393(98)00121-8)
- de Ruiter, P.C., Neutel, A.-M., Moore, J.C., 1995. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* 269, 1257–1260. <https://doi.org/10.1126/science.269.5228.1257>
- Domonell, A., Brabender, M., Nitsche, F., Bonkowski, M., Arndt, H., 2013. Community structure of cultivable protists in different grassland and forest soils of Thuringia. *Pedobiologia* 56, 1–7. <https://doi.org/10.1016/j.pedobi.2012.07.001>
- Dumack, K., Fiore-Donno, A.M., Bass, D., Bonkowski, M., 2020. Making sense of environmental sequencing data: Ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Mol Ecol Resour* 20, 398–403. <https://doi.org/10.1111/1755-0998.13112>
- Dumack, K., Fiore-Donno, A.M., Bass, D., Bonkowski, M., 2019. Making sense of environmental sequencing data: ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Molecular Ecology Resources* 20, 398–403. <https://doi.org/10.1111/1755-0998.13112>
- Dumack, K., Koller, R., Weber, B., Bonkowski, M., 2016. Estimated abundance and diversity of heterotrophic protists in South African biocrusts. *S Afr J Sci* 112, 5–5. <https://doi.org/10.17159/sajs.2016/20150302>
- Dupont, A.Ö.C., Griffiths, R.I., Bell, T., Bass, D., 2016. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environ Microbiol* 18, 2010–2024. <https://doi.org/10.1111/1462-2920.13220>
- Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S., Eisen, J.A., Sundaresan, V., 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci* 112, E911–E920. <https://doi.org/10.1073/pnas.1414592112>
- Estermann, A.H., Teixeira Pereira Bassiaridis, J., Loos, A., Solbach, M.D., Bonkowski, M., Hess, S., Dumack, K., 2023. Fungivorous protists in the rhizosphere of *Arabidopsis thaliana* – Diversity, functions, and publicly available cultures for experimental exploration. *Soil Biol Biochem* 187, 109206. <https://doi.org/10.1016/j.soilbio.2023.109206>
- Falkowski, P.G., 2002. The ocean's invisible forest. *Sci Am* 287, 54–61. <https://doi.org/10.1038/scientificamerican0802-54>
- Faust, K., 2021. Open challenges for microbial network construction and analysis. *ISME J* 15, 3111–3118. <https://doi.org/10.1038/s41396-021-01027-4>
- Faust, K., Raes, J., 2012. Microbial interactions: From networks to models. *Nat Rev Microbiol* 10, 538–550. <https://doi.org/10.1038/nrmicro2832>
- Fiore-Donno, A.M., Berney, C., Pawlowski, J., Baldauf, S.L., 2005. Higher-order phylogeny of plasmodial slime molds (Myxogastria) based on elongation factor 1-A and Small Subunit rRNA gene sequences. *J Eukaryot Microbiol* 52, 201–210. <https://doi.org/10.1111/j.1550-7408.2005.00032.x>
- Fiore-Donno, A.M., Freudenthal, J., Dahl, M.B., Rixen, C., Urich, T., Bonkowski, M., 2024. Biotic interactions explain seasonal dynamics of the alpine soil microbiome. *ISME Commun* 4, ycae028. <https://doi.org/10.1093/ismeco/ycae028>
- Fiore-Donno, A.M., Richter-Heitmann, T., Degrune, F., Dumack, K., Regan, K.M., Marhan, S., Boeddinghaus, R.S., Rillig, M.C., Friedrich, M.W., Kandeler, E., Bonkowski, M., 2019. Functional traits and spatio-temporal structure of a major group of soil protists (Rhizaria: Cercozoa) in a temperate grassland. *Front Microbiol* 10, 1332. <https://doi.org/10.3389/fmicb.2019.01332>
- Fiore-Donno, A.M., Rixen, C., Rippin, M., Glaser, K., Samolov, E., Karsten, U., Becker, B., Bonkowski, M., 2018. New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Mol Ecol Resour* 18, 229–239. <https://doi.org/10.1111/1755-0998.12729>
- Fiore-Donno, A.M., Weinert, J., Wubet, T., Bonkowski, M., 2016. Metacommunity analysis of amoeboid protists in grassland soils. *Sci Rep* 6, 19068. <https://doi.org/10.1038/srep19068>
- Flues, S., Bass, D., Bonkowski, M., 2017. Grazing of leaf-associated Cercomonads (Protists: Rhizaria: Cercozoa) structures bacterial community composition and function. *Environ Microbiol* 19, 3297–3309. <https://doi.org/10.1111/1462-2920.13824>
- Freystein, K., Salisch, M., Reisser, W., 2008. Algal biofilms on tree bark to monitor airborne pollutants. *Biol* 63, 866–872. <https://doi.org/10.2478/s11756-008-0114-z>
- Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., Jousset, A., 2019. Protists: Puppet masters of the rhizosphere microbiome. *Trends Plant Sci* 24, 165–176. <https://doi.org/10.1016/j.tplants.2018.10.011>

## General References

---

- Gauzens, B., Legendre, S., Lazzaro, X., Lacroix, G., 2016. Intermediate predation pressure leads to maximal complexity in food webs. *Oikos* 125, 595–603. <https://doi.org/10.1111/oik.02627>
- Geisen, S., 2016. The bacterial-fungal energy channel concept challenged by enormous functional versatility of soil protists. *Soil Biol Biochem*, Special issue: Food web interactions in the root zone: influences on community and ecosystem dynamics 102, 22–25. <https://doi.org/10.1016/j.soilbio.2016.06.013>
- Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T., Bonkowski, M., 2016. The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biol Biochem* 94, 10–18. <https://doi.org/10.1016/j.soilbio.2015.11.010>
- Geisen, S., Lara, E., Mitchell, E.A.D., Völcker, E., Krashevskaya, V., 2020. Soil protist life matters! *Soil Org* 92, 189–196. <https://doi.org/10.25674/so92iss3pp189>
- Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., de Groot, G.A., 2015a. Not all are free-living: High-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Mol Ecol* 24, 4556–4569. <https://doi.org/10.1111/mec.13238>
- Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., Bonkowski, M., 2015b. Pack hunting by a common soil amoeba on nematodes. *Environ Microbiol* 17, 4538–4546. <https://doi.org/10.1111/1462-2920.12949>
- Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., Urich, T., 2015c. Metatranscriptomic census of active protists in soils. *ISME J* 9, 2178–2190. <https://doi.org/10.1038/ismej.2015.30>
- Gerbracht, J.V., Harding, T., Simpson, A.G.B., Roger, A.J., Hess, S., 2022. Comparative transcriptomics reveals the molecular toolkit used by an algivorous protist for cell wall perforation. *Curr Biol* S0960982222008521. <https://doi.org/10.1016/j.cub.2022.05.049>
- Glücksman, E., Bell, T., Griffiths, R.I., Bass, D., 2010. Closely related protist strains have different grazing impacts on natural bacterial communities. *Environ Microbiol* 12, 3105–3113. <https://doi.org/10.1111/j.1462-2920.2010.02283.x>
- Groß, V., Reinhard, A., Petters, S., Pichler, M., Urich, T., 2023. Adding complexity to soil food webs: Myxobacteria have broad predation spectra with bacteria, yeasts and filamentous fungi *in vitro*. *Eur J Soil Biol* 117, 103508. <https://doi.org/10.1016/j.ejsobi.2023.103508>
- Grossart, H.-P., Simon, M., 2007. Interactions of planktonic algae and bacteria: Effects on algal growth and organic matter dynamics. *Aquat Microb Ecol* 47, 163–176. <https://doi.org/10.3354/ame047163>
- Herold, M., Martínez Arbas, S., Narayanasamy, S., Sheik, A.R., Kleine-Borgmann, L.A.K., Lebrun, L.A., Kunath, B.J., Roume, H., Bessarab, I., Williams, R.B.H., Gillece, J.D., Schupp, J.M., Keim, P.S., Jäger, C., Hoopmann, M.R., Moritz, R.L., Ye, Y., Li, S., Tang, H., Heintz-Buschart, A., May, P., Müller, E.E.L., Laczny, C.C., Wilmes, P., 2020. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. *Nat Commun* 11, 5281. <https://doi.org/10.1038/s41467-020-19006-2>
- Hess, S., Suthaus, A., 2022. The vampyrellid amoebae (Vampyrellida, Rhizaria). *Protist* 173, 125854. <https://doi.org/10.1016/j.protis.2021.125854>
- Huang, K., Mao, Y., Zhao, F., Zhang, X.-X., Ju, F., Ye, L., Wang, Y., Li, B., Ren, H., Zhang, T., 2018. Free-living bacteria and potential bacterial pathogens in sewage treatment plants. *Appl Microbiol Biotechnol* 102, 2455–2464. <https://doi.org/10.1007/s00253-018-8796-9>
- Hunt, H.W., Coleman, D.C., Ingham, E.R., Ingham, R.E., Elliott, E.T., Moore, J.C., Rose, S.L., Reid, C.P.P., Morley, C.R., 1987. The detrital food web in a shortgrass prairie. *Biol Fert Soils* 3, 57–68. <https://doi.org/10.1007/BF00260580>
- Jassey, V.E.J., Walcker, R., Kardol, P., Geisen, S., Heger, T., Lamentowicz, M., Hamard, S., Lara, E., 2022. Contribution of soil algae to the global carbon cycle. *New Phytol* 234, 64–76. <https://doi.org/10.1111/nph.17950>
- Jauss, R.-T., Walden, S., Fiore-Donno, A.M., Dumack, K., Schaffer, S., Wolf, R., Schlegel, M., Bonkowski, M., 2020. From forest soil to the canopy: Increased habitat diversity does not increase species richness of Cercozoa and Oomycota in tree canopies. *Front Microbiol* 11, 592189. <https://doi.org/10.3389/fmicb.2020.592189>
- Johnke, J., Boenigk, J., Harms, H., Chatzinotas, A., 2017. Killing the killer: Predation between protists and predatory bacteria. *FEMS Microbiol Lett* 364, fnx089. <https://doi.org/10.1093/femsle/fnx089>
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K.,

## General References

---

- Romano, G., Rumpho, M.E., Ryneerson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaultot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol* 12, e1001889. <https://doi.org/10.1371/journal.pbio.1001889>
- Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S., Grossart, H.-P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* 5. <https://doi.org/10.3389/fmicb.2014.00251>
- Krieg, N.R., Parte, A., Ludwig, W., Whitman, W.B., Hedlund, B.P., Paster, B.J., Staley, J.T., Ward, N., Brown, D. (Eds.), 2010. *Bergey's manual of systematic bacteriology: Volume 4 the Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*, 2nd Edition. ed. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-68572-4>
- Kulishkin, N.S., Surkova, A.A., Mesentsev, Y.S., Smirnov, A.V., 2023. Three ways to eat spaghetti: Amoebae from two Amoebozoa lineages are able to feed on cyanobacteria of the genus *Oscillatoria*. *Protistology* 17, 38–49. <https://doi.org/10.21685/1680-0826-2023-17-1-4>
- Laforest-Lapointe, I., Messier, C., Kembel, S.W., 2016. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* 4, 27. <https://doi.org/10.1186/s40168-016-0174-1>
- Lapinski, J., Tunnacliffe, A., 2003. Reduction of suspended biomass in municipal wastewater using bdelloid rotifers. *Water Res* 37, 2027–2034. [https://doi.org/10.1016/S0043-1354\(02\)00626-7](https://doi.org/10.1016/S0043-1354(02)00626-7)
- Lentendu, G., Wubet, T., Chatzinotas, A., Wilhelm, C., Buscot, F., Schlegel, M., 2014. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: A multiple barcoding approach. *Mol Ecol* 23, 3341–3355. <https://doi.org/10.1111/mec.12819>
- Litchman, E., De Tezanos Pinto, P., Edwards, K.F., Klausmeier, C.A., Kremer, C.T., Thomas, M.K., 2015. Global biogeochemical impacts of phytoplankton: a trait-based perspective. *J Ecol* 103, 1384–1396. <https://doi.org/10.1111/1365-2745.12438>
- Litchman, E., Edwards, K.F., Boyd, P.W., 2021. Toward trait-based food webs: Universal traits and trait matching in planktonic predator–prey and host–parasite relationships. *Limnol Oceanogr* 66, 3857–3872. <https://doi.org/10.1002/lno.11924>
- Lüttge, U., Büdel, B., 2010. Resurrection kinetics of photosynthesis in desiccation-tolerant terrestrial green algae (Chlorophyta) on tree bark. *Plant Biol* 12, 437–444. <https://doi.org/10.1111/j.1438-8677.2009.00249.x>
- Michel, R., Walochnik, J., Scheid, P., 2014. Article for the “Free-living amoebae Special Issue”: Isolation and characterisation of various amoebophagous fungi and evaluation of their prey spectrum. *Exp Parasitol*, SI: FLAM\_2013 145, S131–S136. <https://doi.org/10.1016/j.exppara.2014.10.005>
- Mizubuti, E.S.G., Fry, W.E., 2006. Potato late blight, in: Cooke, B.M., Jones, D.G., Kaye, B. (Eds.), *The Epidemiology of Plant Diseases*. Springer Netherlands, Dordrecht, pp. 445–471. [https://doi.org/10.1007/1-4020-4581-6\\_17](https://doi.org/10.1007/1-4020-4581-6_17)
- Moore, J.C., Hunt, H.W., 1988. Resource compartmentation and the stability of real ecosystems. *Nature* 333, 261–263. <https://doi.org/10.1038/333261a0>
- Pajdak-Stós, A., Ważny, R., Fiałkowska, E., 2016. Can a predatory fungus (*Zoophagus* sp.) endanger the rotifer populations in activated sludge? *Fungal Ecol* 23, 75–78. <https://doi.org/10.1016/j.funeco.2016.06.005>
- Pauli, W., Jax, K., Berger, S., 2001. Protozoa in wastewater treatment: Function and importance, in: Beek, B. (Ed.), *Biodegradation and Persistence*. Springer-Verlag, Berlin/Heidelberg, pp. 203–252. [https://doi.org/10.1007/10508767\\_3](https://doi.org/10.1007/10508767_3)
- Petters, S., Groß, V., Söllinger, A., Pichler, M., Reinhard, A., Bengtsson, M.M., Urlich, T., 2021. The soil microbial food web revisited: Predatory myxobacteria as keystone taxa? *ISME J* 15, 2665–2675. <https://doi.org/10.1038/s41396-021-00958-2>
- Potapov, A.M., Rozanova, O.L., Semenina, E.E., Leonov, V.D., Belyakova, O.I., Bogatyreva, V.Yu., Degtyarev, M.I., Esaulov, A.S., Korotkevich, A.Yu., Kudrin, A.A., Malysheva, E.A., Mazei, Y.A., Tsurikov, S.M., Zuev, A.G., Tiunov, A.V., 2021. Size compartmentalization of energy channeling in terrestrial belowground food webs. *Ecology* 102, e03421. <https://doi.org/10.1002/ecy.3421>
- Rahman, M.U., Chen, P., Zhang, X., Fan, B., 2023. Predacious strategies of nematophagous fungi as bio-control agents. *Agronomy* 13, 2685. <https://doi.org/10.3390/agronomy13112685>
- Ribeiro, G.M., Porfírio-Sousa, A.L., Maurer-Alcalá, X.X., Katz, L.A., Lahr, D.J.G., 2020. De novo sequencing, assembly, and annotation of the transcriptome for the free-living testate amoeba *Arcella intermedia*. *J Eukaryot Microbiol* 67, 383–392. <https://doi.org/10.1111/jeu.12788>

## General References

---

- Rodríguez, E., García-Encina, P.A., Stams, A.J.M., Maphosa, F., Sousa, D.Z., 2015. Meta-omics approaches to understand and improve wastewater treatment systems. *Rev Environ Sci Biotechnol* 14, 385–406. <https://doi.org/10.1007/s11157-015-9370-x>
- Romani, M., Carrion, C., Fernandez, F., Intertaglia, L., Pecqueur, D., Lebaron, P., Lami, R., 2019. High bacterial diversity in pioneer biofilms colonizing ceramic roof tiles. *Int Biodeterior Biodegradation* 144, 104745. <https://doi.org/10.1016/j.ibiod.2019.104745>
- Romdhane, S., Spor, A., Aubert, J., Bru, D., Breuil, M.-C., Hallin, S., Mounier, A., Ouadah, S., Tsiknia, M., Philippot, L., 2022. Unraveling negative biotic interactions determining soil microbial community assembly and functioning. *ISME J* 16, 296–306. <https://doi.org/10.1038/s41396-021-01076-9>
- Rosenberg, K., Bertaux, J., Krome, K., Hartmann, A., Scheu, S., Bonkowski, M., 2009. Soil amoebae rapidly change bacterial community composition in the rhizosphere of *Arabidopsis thaliana*. *ISME J* 3, 675–684. <https://doi.org/10.1038/ismej.2009.11>
- Röttjers, L., Faust, K., 2018. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol Rev* 42, 761–780. <https://doi.org/10.1093/femsre/fuy030>
- Saleem, M., Fetzer, I., Dormann, C.F., Harms, H., Chatzinotas, A., 2012. Predator richness increases the effect of prey diversity on prey yield. *Nat Commun* 3, 1305. <https://doi.org/10.1038/ncomms2287>
- Sapp, M., Tyborski, N., Linstädter, A., López Sánchez, A., Mansfeldt, T., Waldhoff, G., Bareth, G., Bonkowski, M., Rose, L.E., 2019. Site-specific distribution of oak rhizosphere-associated oomycetes revealed by cytochrome c oxidase subunit II metabarcoding. *Ecol Evol* 9, 10567–10581. <https://doi.org/10.1002/ece3.5577>
- Schuler, G.A., Tice, A.K., Pearce, R.A., Foreman, E., Stone, J., Gammill, S., Willson, J.D., Reading, C., Silberman, J.D., Brown, M.W., 2018. Phylogeny and Classification of Novel Diversity in Sainouroidea (Cercozoa, Rhizaria) Sheds Light on a Highly Diverse and Divergent Clade. *Protist* 169, 853–874. <https://doi.org/10.1016/j.protis.2018.08.002>
- Shen, C., He, J.-Z., Ge, Y., 2021. Seasonal dynamics of soil microbial diversity and functions along elevations across the treeline. *Sci Total Environ* 794, 148644. <https://doi.org/10.1016/j.scitotenv.2021.148644>
- Sibbald, S.J., Archibald, J.M., 2017. More protist genomes needed. *Nature Ecology & Evolution* 1, 0145. <https://doi.org/10.1038/s41559-017-0145>
- Singer, D., Seppey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q., Debroas, D., de Groot, G.A., de Vargas, C., Domaizon, I., Duckert, C., Izaguirre, I., Koenig, I., Mataloni, G., Schiaffino, M.R., Mitchell, E.A.D., Geisen, S., Lara, E., 2021. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ Int* 146, 106262. <https://doi.org/10.1016/j.envint.2020.106262>
- Thapa, S., Prasanna, R., 2018. Prospecting the characteristics and significance of the phyllosphere microbiome. *Ann Microbiol* 68, 229–245. <https://doi.org/10.1007/s13213-018-1331-5>
- Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., Rokas, A., Shen, X.-X., Strasser, J.F.H., Kolísko, M., Brown, M.W., 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. *PLOS Biol* 19, e3001365. <https://doi.org/10.1371/journal.pbio.3001365>
- Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLOS ONE* 3, e2527. <https://doi.org/10.1371/journal.pone.0002527>
- Voss, C., Fiore-Donno, A.M., Guerreiro, M.A., Peršoh, D., Bonkowski, M., 2019. Metatranscriptomics reveals unsuspected protistan diversity in leaf litter across temperate beech forests, with Amoebozoa the dominating lineage. *FEMS Microbiol Ecol* 95, fiz142. <https://doi.org/10.1093/femsec/fiz142>
- Walczyńska, A., Sobczyk, M., Fiałkowska, E., Pajdak-Stós, A., Fyda, J., Wiąckowski, K., 2018. Interaction Between a Bacterivorous Ciliate *Aspidisca cicada* and a Rotifer *Lecane inermis*: Doozers and Fraggles in Aquatic Floccs. *Microb Ecol* 75, 569–581. <https://doi.org/10.1007/s00248-017-1036-5>
- Walden, S., Jauss, R.-T., Feng, K., Fiore-Donno, A.M., Dumack, K., Schaffer, S., Wolf, R., Schlegel, M., Bonkowski, M., 2021. On the phenology of protists: recurrent patterns reveal seasonal variation of protistan (Rhizaria: Cercozoa and Endomyxa) communities in tree canopies. *FEMS Microbiol Ecol* 97, 2021.02.15.431229. <https://doi.org/10.1093/femsec/fiab081>
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* 10, 1669–1681. <https://doi.org/10.1038/ismej.2015.235>
- Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., Boon, N., 2009. Initial community evenness favours functionality under selective stress. *Nature* 458, 623–626. <https://doi.org/10.1038/nature07840>

## General References

---

- Xue, J., Schmitz, B.W., Caton, K., Zhang, B., Zabaleta, J., Garai, J., Taylor, C.M., Romanchishina, T., Gerba, C.P., Pepper, I.L., Sherchan, S.P., 2019. Assessing the spatial and temporal variability of bacterial communities in two Bardenpho wastewater treatment systems via Illumina MiSeq sequencing. *Sci Total Environ* 657, 1543–1552. <https://doi.org/10.1016/j.scitotenv.2018.12.141>
- Yao, N.Y., O'Donnell, M.E., 2016. Evolution of replication machines. *Crit Rev Biochem Mol Biol* 51, 135–149. <https://doi.org/10.3109/10409238.2015.1125845>
- Zahedi, A., Greay, T.L., Paparini, A., Linge, K.L., Joll, C.A., Ryan, U.M., 2019. Identification of eukaryotic microorganisms with 18S rRNA next-generation sequencing in wastewater treatment plants, with a more targeted NGS approach required for *Cryptosporidium* detection. *Water Res* 158, 301–312. <https://doi.org/10.1016/j.watres.2019.04.041>

---

## ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to my supervisor **Prof. Dr. Michael Bonkowski** for offering me the great opportunity to join his research group for my doctoral thesis and for his support, guidance, kindness, and valuable advice. Further, I would like to thank **Prof. Dr. Martin Schlegel** for his support, expertise, fruitful discussions, and constant encouragement. I would like to thank **PD Dr. Kenneth Dumack**. Over the past years, he always fully supported me, offered opportunities and perspectives, gave me insights into the world of academia, accompanied me through challenging times, and celebrated every success with me. Thanks to all of you, for offering me great opportunities to pursue my scientific career, for enabling me to attend numerous conferences and workshops and for creating a working environment in which ideas and open discussions were encouraged. Additionally, I would like to give special thanks to **PD Dr. Kathrin Lampert** for being the second assessor of this thesis. I truly appreciate that you invest the time to read and evaluate this thesis.

I had the pleasure of collaborating with many colleagues. I would like to thank all of them for this opportunity, the valuable insights, your expertise, and the great teamwork, especially **Anna Maria Fiore-Donno, Lioba Rüger, and Nils Heck**. I would also like to thank my entire working group. You all create an amazing working environment with great team spirit, lots of inspiration, and countless moments of shared laughter.

I am deeply grateful to all my friends and family for their support and encouragement throughout this journey. Finally, I would also like to thank those who have not been mentioned by name but who have nevertheless made important contributions.

---

# ERKLÄRUNG ZUR DISSERTATION

gemäß der Promotionsordnung vom 12. März 2020

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.“

## Teilpublikationen:

- Freudenthal, J., Ju, F., Bürgmann, H., Dumack, K., 2022.** Microeukaryotic gut parasites in wastewater treatment plants: Diversity, activity, and removal. *Microbiome* 10, 27. <https://doi.org/10.1186/s40168-022-01225-y>
- Rüger, L., Ganther, M., **Freudenthal, J., Jansa, J., Heintz-Buschart, A., Tarkka, M.T., Bonkowski, M., 2023.** Root cap is an important determinant of rhizosphere microbiome assembly. *New Phytol* 239, 1434–1448. <https://doi.org/10.1111/nph.19002>
- Heck, N., **Freudenthal, J., Dumack, K., 2023.** Microeukaryotic predators shape the wastewater microbiome. *Water Res* 242, 120293. <https://doi.org/10.1016/j.watres.2023.120293>
- Fiore-Donno, A.M., **Freudenthal, J., Dahl, M.B., Rixen, C., Urich, T., Bonkowski, M., 2024.** Biotic interactions explain seasonal dynamics of the alpine soil microbiome. *ISME Commun* 4, ycae028. <https://doi.org/10.1093/ismeco/ycae028>
- Freudenthal, J., Dumack, K., Schaffer, S., Schlegel, M., Bonkowski, M., 2024.** Algae-fungi symbioses and bacteria-fungi co-exclusion drive tree species-specific differences in canopy bark microbiomes. *ISME J*, wrae206. <https://doi.org/10.1093/ismejo/wrae206>. **Accepted manuscript.**
- Freudenthal, J., Schlegel, M., Bonkowski, M., Dumack, K., n.d.** A novel protistan trait database reveals functional redundancy and complementarity in terrestrial protists (Amoebozoa & Rhizaria). **In preparation.**
- Freudenthal, J., Bonkowski, M., Engelmann, N.-J., Schlegel, M., Dumack, K., n.d.** Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics. **In preparation.**