UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics

Department of Arctic and Marine Biology

Microbial eukaryotes and their functional importance in the Arctic

A Svalbardian perspective

Magdalena Wutkowska

A dissertation for the degree of Philosophiae Doctor - August 2020

Microbial eukaryotes and their functional importance in the Arctic.

A Svalbardian perspective

Magdalena Wutkowska

Thesis submitted in partial fulfillment of the requirements
for the degree of Philosophiae Doctor (PhD) in Natural Sciences
AUGUST 2020

Department of Arctic Biology
The University Centre in Svalbard

Department of Arctic and Marine Biology
Faculty of Biosciences, Fisheries and Economics
UiT – The Arctic University of Norway

This PhD thesis was supervised by:

Pernille B. Eidesen, PhD

Department of Arctic Biology at The University Centre in Svalbard, Longyearbyen, Norway

Dorothee Ehrich, PhD

Department of Arctic and Marine Biology at UiT – The Arctic University of Norway, Tromsø, Norway

Tove M. Gabrielsen, PhD

Department of Natural Sciences at University of Agder, Kristiansand, Norway
Department of Arctic Biology at The University Centre in Svalbard, Longyearbyen, Norway

““*Life is weird.”*

“As opposed to what?”

– Found on the bathroom stall door of a truck stop in Breezewood, Pennsylvania”

This quote opened a chapter *What is light, really?* in an awesome book *The optics of life: a biologist’s guide to light in nature* by Sönke Johnsen

Acknowledgements

Although there is only one name on the cover of this booklet, it would never happen without help and support from so many people. I want to thank you all - I believe you know who you are! Here I would like to mention only several names, I hope that the rest of you will forgive me for that.

First and foremost, I want to thank my supervisors for coming up with this project that lured me in a totally different world. Pernille, you've taught me a lot at every stage of this project, and I'll always be grateful for your patience, for saving me from troubles and my overthinking. Tove, thank you for trusting me with this enormous metatranscriptomic endeavour, this part of my PhD taught me the most in all possible dimensions. Dorothee, thank you for stepping out of your comfort zone and joining this project with your expertise, for sharing your knowledge and for making me always feel welcome in Tromsø.

I want to thank two Annas without whom this PhD would be a totally different experience. Anna Vader, thank you for being not only a co-author of all of my papers, a critical researcher and the best companion for a two-person journal club! Anna Ejsmond thank you for being the best office- and photo-project companion! Thank you for all the hours we spent talking about science and for tolerating my temperature preferences in our lovely office!

I want to thank all the people who were involved in the MicroFun project, who collected, filtered/sieved/processed the samples that I could use in my project.

I want to thank many people and institutions that were a part of the scientific environment around me. I want to thank all of my co-authors for your valuable contributions. I want to thank Ramiro Logares for an opportunity to learn from you and all the members of the Loglab at the Institute of Marine Sciences in Barcelona. I want to thank members of NEFOM and ARCTOS networks for their support and ForBio for providing excellent learning opportunities. I want to thank all the people whom I could help with their fieldwork, courses and projects throughout this PhD, I learned a lot from you! I want to thank all the people who had so much patience to answer my questions throughout this project - special thanks to Abel supercomputer helpdesk in Oslo - you are the heroes!

Finally, I would like to thank all the people that were an essential part of my non-work life, during dog/kayak/hiking trips, talks about films, philosophy, physics and space, all the art events in town... Massive thank you to all of the people that were geographically far away, but supported me despite the distance: my family and friends.

Table of Contents

Abstract

Microbial eukaryotes, including protists and fungi, play diverse functions in virtually all ecosystems. In the High Arctic, their high biomass and diversity reflects crucial ecological importance and the performance of key ecological processes. Protists are the main primary producers in arctic seas, whereas fungi are an important group of decomposers and symbiotic partners of plants in terrestrial habitats. During the last decade, along with the development of new high-throughput sequencing methods, our knowledge regarding arctic microbial eukaryotes has expanded. Previous studies have identified the major groups of microbial eukaryotes present in Svalbard and how their richness and abundance may vary along various temporal and spatial scales. Those studies used high-throughput sequencing to reveal the dynamics, biodiversity patterns and community composition of diverse microbial eukaryotes such as marine protists, soil and root-associated fungi. However, altogether these studies have just scratched the surface of disentangling the biodiversity and its drivers. Basic questions regarding taxonomic diversity, community composition and their drivers are addressed in a limited manner, often leaving most of the observed variation unexplained. Regarding functionality of these organisms, even less is known. At the same time, these findings have also increased the amount of questions about microbial eukaryotes, their life histories, strategies, seasonality, sensitivity to changes in environmental conditions, as well as functional importance of these organisms at different scales.

Previously unexplained variation and other emerging knowledge gaps regarding microbial eukaryotes formed a backstage for this thesis. The main focus was to look at these organisms from a functional angle regarding variation related to methodology, seasonality and biotic factors through case studies addressing the following knowledge gaps. Firstly, to understand the functionality of biodiversity in a temporal and spatial context of cold soils, we need to determine if our methods estimate biodiversity of the active community of microbial eukaryotes. In other words, does DNA-based detection of species provide good enough approximation to continue or is a different methodology needed? We found that the choice of marker gene template influenced diversity measures and read numbers in abundant fungal groups such as Helotiales and Agaricales. However, it did not impact the community structure. Secondly, the aim was to understand the role of biodiversity and functionality of plant root-associated fungi in relation to host plant performance. We explored putative effects of fungal diversity on plant morphology and the interplay between functional diversity and abiotic factors in a spatial context. Our results revealed the importance of fungal richness and functional diversity, but no impact of community structure on plant morphometrics. Moreover, we showed that temperature affects fungal richness, below- and aboveground parts of the plant in different

ways, making it difficult to predict its impact on the biological outcomes in natural systems. The third aim was to address the lack of polar night investigations of microbial eukaryotes in general, especially with a strong focus on their functions. Here, we investigated the impact of strong seasonality on functions of microbial eukaryotes in the marine environment. Community-level gene expression was driven primarily by seasonal patterns of light availability. Among the most expressed transcripts, nearly ⅔ transcripts were not functionally annotated, providing further evidence for distinct genetic makeup of the Arctic Ocean.

Through these three case studies, this thesis contributed some building blocks to close important knowledge gaps, but also revealed that there are more unknowns to be addressed. This thesis aimed to increase awareness of the importance of functional understanding of the roles of microbial eukaryotes in the High Arctic ecosystems. Finally, it highlights further possibilities and developments that could improve the understanding of possible future responses of these organisms and processes that they control.

List of papers

The thesis contains the following papers:

I.

Wutkowska M.*, Vader A., Mundra S., Cooper E.J. and Eidesen P.B., (2019) Dead or Alive; or Does It Really Matter? Level of Congruency Between Trophic Modes in Total and Active Fungal Communities in High Arctic Soil. Frontiers in Microbiology 9:3243. DOI: 10.3389/fmicb.2018.03243

II.

Wutkowska M.*, Ehrich D., Mundra S., Vader A., and Eidesen P.B., Can root-associated fungi mediate the impact of abiotic conditions on the growth of a High Arctic herb?

[Manuscript available at biorXiv.org, DOI: 10.1101/2020.06.20.157099]

III.

Wutkowska M.*, Vader A., Logares R., Pelletier E., Gabrielsen T.M., Linking extreme seasonality and gene expression in arctic marine protists.

[Manuscript awaiting publication of TARA Oceans metatranscriptomics datasets before being ready to be published]

* - corresponding author

Introduction

Towards a theory in microbial ecology

Investigations of microbial communities usually begin with asking questions concerning the taxonomic identity and phylogenetic relationship between its members (Little et al., 2008). Microbes are somewhat elusive; it is difficult or impossible to see them directly. Most microbial species are rare (Logares et al., 2014; Nemergut et al., 2011), distributed stochastically (Bahram et al., 2016) and some are difficult to culture (Cuvelier et al., 2010) or even to detect using molecular tools (Schoch et al., 2012). Additionally, some can be too similar to distinguish between species (Balasundaram et al., 2015; Zhao et al., 2018). Above all, they live in complex and dynamic communities (Konopka et al., 2015; Tecon et al., 2019), operating at spatial and temporal scales that are difficult to comprehend from a human perspective (Ladau & Eloe-Fadrosh, 2019).

The amount of genetic information stored in microorganisms is higher than in plants and animals (Landenmark et al., 2015). Due to their unique metabolism, they drive global biogeochemical cycles and are indispensable in many pivotal ecological processes (Field et al., 1998). The array of molecular methods available to study environmental microbiology nowadays is vast (Bouchez et al., 2016). The sequencing revolution brought more insights into microbiology, revealing previously unexpected diversity and complexity of microorganisms and their communities (Clark et al., 2018; Loman & Pallen, 2015). However, despite new tools, it is difficult to uncover their response mechanisms to changes in the environment, sometimes because there is no baseline knowledge to compare to. This includes distinguishing between these responses and natural variability in spatial and temporal context. Many of the most urgent problems faced by the global society today could perhaps be slowed down or tackled once we understood mechanisms and relationships within the microbial communities, between them and the environment that they inhabit (Cavicchioli et al., 2019; de Lorenzo, 2017; Gillings & Paulsen, 2014). This includes evidence-based conservation and management of natural resources (Malik et al., 2013), human and animal health and wellbeing (Clemente et al., 2012), as well as slowing down ongoing climate changes or its consequences (Cavicchioli et al., 2019). Yet, microbial ecology lacks a proper theoretical framework, i.e. an ecological theory, that would assure efficient and systematic gathering of information, its repeated testing, interpretation and verification (Escalas et al., 2019; Inkpen et al., 2017; Prosser et al., 2007). Microorganisms differ from macroorganisms in some fundamental ways, including how species are being defined, recognizing spatiotemporal scales and state of activity, dispersal, generation length etc. (Andrews, 2017; Prosser et al., 2007). The ultimate

gain of developing such a theory would allow making predictions concerning microbial communities and their ecosystems, instead of piling up facts.

Looking beyond 'Who is there?'

Description of taxonomic identity of members of microbial communities usually precedes research revealing what they do, or - in other words - what is their function (Little et al., 2008). Function in ecology is context-specific and the understanding of the term is still subjected to a long-lasting debate (e.g. Graham et al., 2016; Jax, 2005; Loreau, 2001; Nunes-Neto et al., 2014). The advent of molecular methods in ecology changed understanding of how function is perceived, especially in unicellular organisms (Vandenkoornhuyse et al., 2010). The variety of scales ranging from ecosystem, through species and organism to a cell or unicellular organism make it even more difficult to use the term 'function' in a consistent way (Farnsworth et al., 2017). Altogether the use of function spans from ecosystem services to biochemical processes carried out by molecules.

Organisms are dynamic, resource-processing systems that thrive in a certain space of physicochemical conditions (Calow, 1987). All organismal assemblages consist of organisms that are suited for thriving in an environment with certain combinations of conditions and co-habitants. These organisms intake necessary elements from the environment and output metabolites changing the environment around them in a particular way. This way of interacting and changing with the environment could be interpreted as their function in the ecosystem. Therefore, functions can be understood as organismal characteristics important in biogeochemical cycling (such as carbon acquisition) or as the energy source powering a cell (Figure 1). Some of the categories can be further divided, such as heterotrophy in groups as fungi, where organic carbon can come from symbiotic relationships (symbiotrophs), decomposition of organic matter (saprotrophs) or harming living cells (pathotrophs). These broad categories have been ecologically useful for describing general functions of organisms (Nguyen et al., 2016). No matter what definition of function of an organism is taken into consideration, these characteristics are specified as the presence of certain genes or their sets in the genome, therefore ultimately functions are encoded in genes. However, a presence of a gene describes only a functional potential of the organism. It does not imply when and how often it is expressed, therefore with what intensity it contributes to biogeochemical processes.

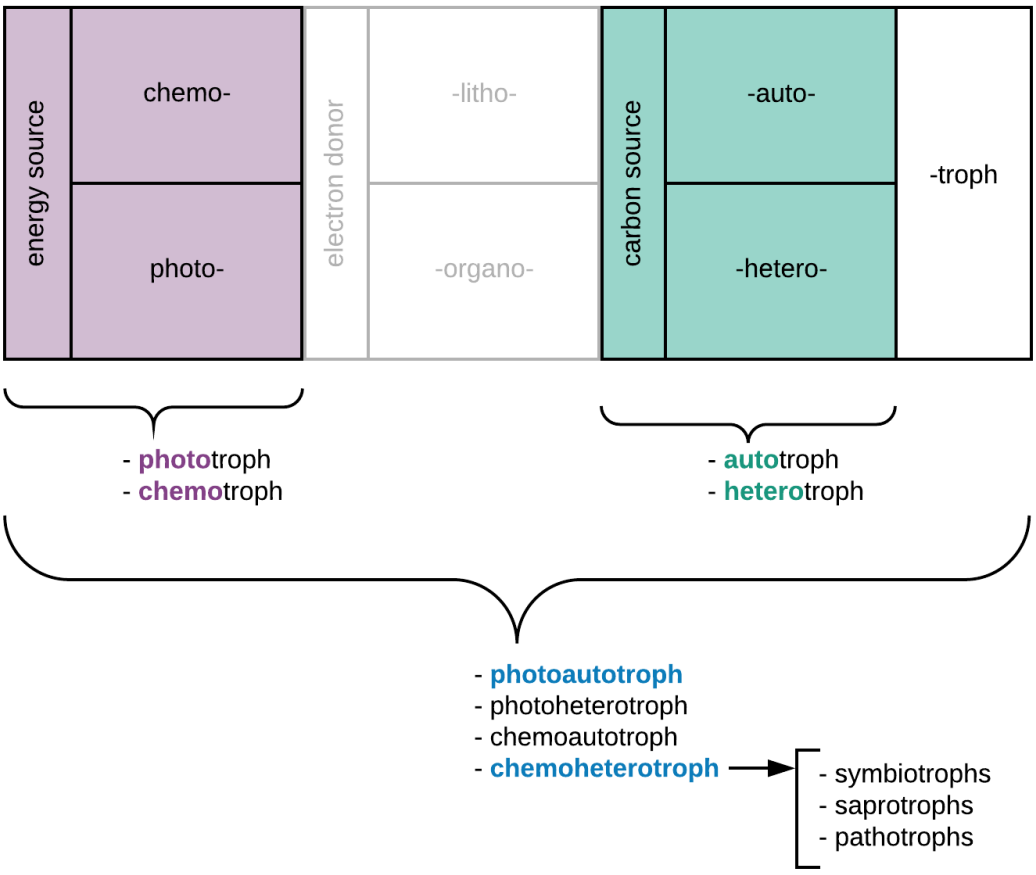


Figure 1 | Primary nutritional groups represent the resource requirements of an organism, therefore they can be understood as a relationship of the organism with its environment and thus its function in this environment. All of this metabolic potential is reserved to bacteria and archaea; whereas eukaryotes including microbial eukaryotes belong to photoautotrophs and chemoheterotrophs (highlighted in blue). Fungi and heterotrophic protists belong to the last category. Traditionally, fungi are further divided into trophic modes based on the origin of organic carbon: symbio- (from other organisms through symbiotic relationship), sapro- (from decomposed organic matter) and pathotrophs (from other living organisms).

Recent advancements in understanding microbial eukaryotes in Svalbard

The umbrella term ‘microbial eukaryotes’ refers to a polyphyletic group of microorganisms containing nuclei in their cells, which includes protists and fungi (Andrews, 2017; Caron et al., 2009; Taylor et al., 2006). The levels of complexity of their cells, genomes, energetics and processes are profoundly different from these in bacteria and archaea (Basile et al., 2019; Lynch, 2006; Lynch & Conery, 2003; Lynch & Marinov, 2017). These differences contribute in many ways to increased efforts and resources required to study them (Keeling & Campo, 2017). Nevertheless, microbial eukaryotes have the second-highest biomass in the biosphere after plants, and the highest biomass in the oceans (Bar-On et al., 2018; Bar-On & Milo, 2019),

which reflects their importance in global ecological processes. They play versatile functions in virtually all environments on Earth, ranging from deep Antarctic seas (López-García et al., 2001), geothermal springs (Oliverio et al., 2018) to Atacama Desert caves (Azúa-Bustos et al., 2009) and the High Arctic. An overwhelming majority of eukaryotic lineages in the tree of life consist only of microbial eukaryotes (Adl et al., 2018; Keeling & Burki, 2019; Patterson, 1999). However, despite their abundance, biodiversity, versatile life histories and contribution to biogeochemical cycles, microbial eukaryotes are often overlooked in microbiology, ecology and medicine (Bik et al., 2012; Keeling & Campo, 2017; Laforest-Lapointe & Arrieta, 2018; Oliverio et al., 2018).

Microbial eukaryotes are major primary producers in the Arctic marine environment, due to low representation of cyanobacteria at high latitudes (Vincent, 2000). In terrestrial habitats they are important plant symbiotrophs and decomposers of organic matter in soils containing large carbon pools (Schuur et al., 2015; Tarnocai et al., 2009). Therefore, microbial eukaryotes are a crucial group of organisms highly engaged in many aspects of carbon cycling. It is not clear how these organisms respond to environmental changes and thus how they will alter carbon cycling. As Svalbard is one of the locations in the Arctic that experiences the most intense repercussions of climate change (Nordli et al., 2020), it is an important spot to research these organisms.

Svalbard, an archipelago located in the European part of the High Arctic (74-81°N, 8-34°E), provides a wide variety of microbial habitats subjected to strong seasonal patterns of physicochemical factors driven primarily by light and nutrient availability, as well as temperature. It is one of the most accessible places in the High Arctic with many research facilities in Ny-Ålesund, Longyearbyen and several remote research stations. However, Svalbard habitats are perhaps not representative for typical Arctic habitats due to many features, such as: patchy landscape, considerable distance from other land masses, geological history, relatively mild climate for such latitude, and the quickest rise of temperatures in the last three decades compared to an Arctic average (Nordli et al., 2020). The remoteness of the archipelago may provide a dispersal barrier for microbes, especially larger size fractions of cells or spores (Wilkinson et al., 2012). These features make Svalbard even more interesting and a valuable location to study microbial life. Molecular tools used in the last decade helped to understand that Svalbard's seemingly barren landscapes teem with microbial life that contribute significantly to biogeochemical cycles. So far, the majority of microbial eukaryotic research in the Arctic, including Svalbard, focused on the community structure, diversity and environmental drivers influencing these communities.

Terrestrial habitats

Fungi are the most researched microbial eukaryotes in terrestrial habitats in the Arctic, including Svalbard. The majority of molecular studies of fungal communities focused on plant root-associated fungi, in particular ectomycorrhizal species (EcM), whereas soil fungi attracted less attention. The first clonal library studies, predating high-throughput metabarcoding era, revealed that despite geographical isolation of the archipelago, EcM communities are more diverse than previously expected (Geml et al., 2011). Moreover, the diversity of root-associated fungi of the common arctic and alpine plant *Dryas octopetala* was shown to be equally high in Svalbard and southern Norway, and did thus not decline with latitude (Bjorbækmo et al., 2010) as previously shown for terrestrial macroorganisms (Hillebrand, 2004). In general, the majority of root-associated fungi belong to EcM, followed by a substantial proportion of saprotrophs (Bjorbækmo et al., 2010; Blaalid et al., 2012; Botnen et al., 2014; Lorberau et al., 2017). EcM fungi tend to be stochastically distributed (Blaalid et al., 2012) and did not show specificity according to host plant species (Botnen et al., 2014). In primary succession gradients, richness of root-associated fungi increased with the distance from the glacier forefront, therefore also with the glacier free-period of the substrate and soil developmental stages (Blaalid et al., 2012; Davey et al., 2015). Communities of soil and root-associated fungi in the Midtre Lovénbreen chronosequence follow distinct development patterns: directional replacement (Dong et al., 2016) and directional-non-replacement (Davey et al., 2015), respectively. Root-associated fungi in Svalbard were studied at different spatial scales starting from centimeters (Mundra, Halvorsen, et al., 2015) to hundreds of kilometers (Blaalid et al., 2014). Root-associated communities show no or little spatial structure at different scales with high levels of heterogeneity (Bjorbækmo et al., 2010; Botnen et al., 2014; Mundra, Halvorsen, et al., 2015). There is a strong need to assess how fungi and other belowground organisms respond to various climate change scenarios such as increased temperature or increased precipitation. However, most of the research presented so far indicates that there are no or little effects of such treatments (Lorberau et al., 2017; Mundra, Halvorsen, et al., 2016).

Marine habitats

Historically, the most researched marine microbial eukaryotes in waters around Svalbard were some of the bloom forming nano- and micro planktonic plastid-bearing species taxonomically identified with microscopy. However, molecular tools revealed that there is a tremendous diversity among smaller cells: pico- (0.2-2µm) and nanoplanktonic (2-20µm) microbial eukaryotes (Marquardt et al., 2016; Sørensen et al., 2012). As a result of climate changes the Arctic Ocean becomes warmer and less saline enhancing stratification of the water column (Wassmann et al., 2011). These conditions may favour organisms with a higher surface-area-

to-volume ratio, which are more efficient in absorbing nutrients, such as picoplankton (Li et al., 2009). Despite small sizes (<10µm), microbial pico- and small nanoeukaryotes contribute to 50% of the primary production in the Barents Sea (Hodal & Kristiansen, 2008). Key phototrophs belonging to these groups were detected as active during prolonged period of darkness (at 78°N ~ 4 months) during the polar night (Marquardt et al., 2016; Vader et al., 2014). Establishing the world's northernmost time series station in Adventfjorden helped to answer some of the fundamental questions on the seasonality of the microbial eukaryotic community in relation to abiotic factors. These communities exhibit distinct phases throughout the year (Kubiszyn et al., 2017; Marquardt et al., 2016). The well-mixed water column containing the low biomass winter community was the most diverse throughout the year and primarily heterotrophic (Kubiszyn et al., 2017; Marquardt et al., 2016). The return of light prompted winter to spring transition with a rapid increase in biomass of photosynthetic species within the community and is characterised by low diversity (Marquardt et al., 2016; Iversen & Seuthe, 2011; Vaqué et al., 2008). These trends continued through the spring bloom, however, with a changed species composition compared to the early phase after the light returned. Post-bloom stage encompasses summer and fall, when the diversity increases, the community becomes more heterotrophic and the overall biomass decreases before the start of the polar night.

Together these studies characterizing the diversity of microbial eukaryotes in and around Svalbard in relation to environmental gradients create an excellent basis to look beyond community composition and move towards better ecological understanding.

Objectives

Molecular tools like high-throughput sequencing have rapidly advanced our knowledge of diversity and community structure in arctic microbial eukaryotes, but at the same time revealed that our commonly measured environmental variables are far from sufficient to explain the spatial and temporal variation revealed in these systems. The overarching objective of this study was to understand more of this spatial and temporal variation by exploring different methodological approaches and focus on the functional importance of microbial eukaryotes in Svalbard. This thesis attempted to explore this broad objective by looking into specific research questions that were:

- Does the type of template matter when describing microbial eukaryote communities? A comparison of results based on rDNA and rRNA templates of the same marker gene (**PAPER I**).
- Do functional groups matter? Will the diversity and community structure within fungal functional groups respond in concert with or independently from the environmental variables? (**PAPER I, PAPER II**)
- Is the interplay between root-associated fungi, environmental factors, and host-plant performance influenced by fungal diversity, community structure and/or functional diversity? **(PAPER II)**
- What are the functions of marine pelagic microbial eukaryotes throughout the year and which environmental factors influence these functions? **(PAPER III)**
- How different are the functional profiles of microbial eukaryotes during the polar night? Are there similarities between them in two consecutive polar nights? **(PAPER III)**

Approach

Samples

All samples used in this PhD were collected in 2011-2013, mainly as a part of the MicroFun Project led at UNIS in 2012-2016. The project aimed at describing the identity and diversity of microbial eukaryotes in Svalbard using high-throughput sequencing (mostly DNA metabarcoding). The marine side of the project focused on protists (Marquardt et al., 2019; 2016; Meshram et al., 2017; Vader et al., 2014, 2018; Wiedmann et al., 2016), whereas the terrestrial part focused on soil and root-associated fungi (Lorberau et al., 2017; Mundra, Bahram, et al., 2015, 2016; Mundra, Halvorsen, et al., 2015, 2016). These studies described the taxonomic identity of major players, spatiotemporal patterns in communities of microbial eukaryotes and some of the abiotic variables driving these patterns. All of these investigations revealed only the tip of the iceberg when it comes to understanding these communities, leaving a substantial proportion of observed variation unexplained. They also identified many knowledge gaps. Thus, these results suggested a change of approach in future research, perhaps looking at the data again from a functional perspective in order to explain more of the immense diversity these studies revealed. Therefore, the present project was developed in a way that addresses some of the key questions regarding functional aspects of microbial eukaryotes in arctic habitats in Svalbard and showcases how this topic could be approached and perhaps further developed. Besides looking at microbial eukaryotes through the lens of their functions, this project tested the usefulness of expressed marker genes (rRNA) in comparison to rDNA in evaluating fungal diversity and community structure in soil. For functional investigation of microbial eukaryotes in the temporal marine study we used metatranscriptomics which enables investigation of an entire pool of polyadenylated genes expressed by all the cells in the sample (Figure 2).

Uncovering functions using high-throughput sequencing

Culture-independent high-throughput methods have revolutionised microbial ecology (Su et al., 2012). Especially, the next generation sequencing-based approaches became increasingly available due to many technological developments. Currently, there is an array of diverse high-throughput sequencing methods to assess different types of molecules in an environmental sample (Figure 2), such as genes (metabarcoding, metagenomics), expressed genes (metatranscriptomics) etc. In order to showcase possible approaches to describe communities of microbial eukaryotes we used two approaches (Figure 2, outlined in green).

In PAPERS I & II we used organisms' identities obtained from metabarcoding datasets to infer functions for each taxonomic annotation using a comprehensive curated database; whereas in PAPER III identities of expressed genes were used to assess their functions. In PAPERS I and II functions were assigned by querying fungal taxonomic identities against the FUNGuild database gathering literature references of fungal trophic modes and guilds (Nguyen et al., 2016). This powerful tool combines functional information from literature for over 13000 fungal taxa. In PAPER III we assigned functions directly to expressed genes using a unified database of gene functions across all species, namely The Gene Ontology (Ashburner et al., 2000).

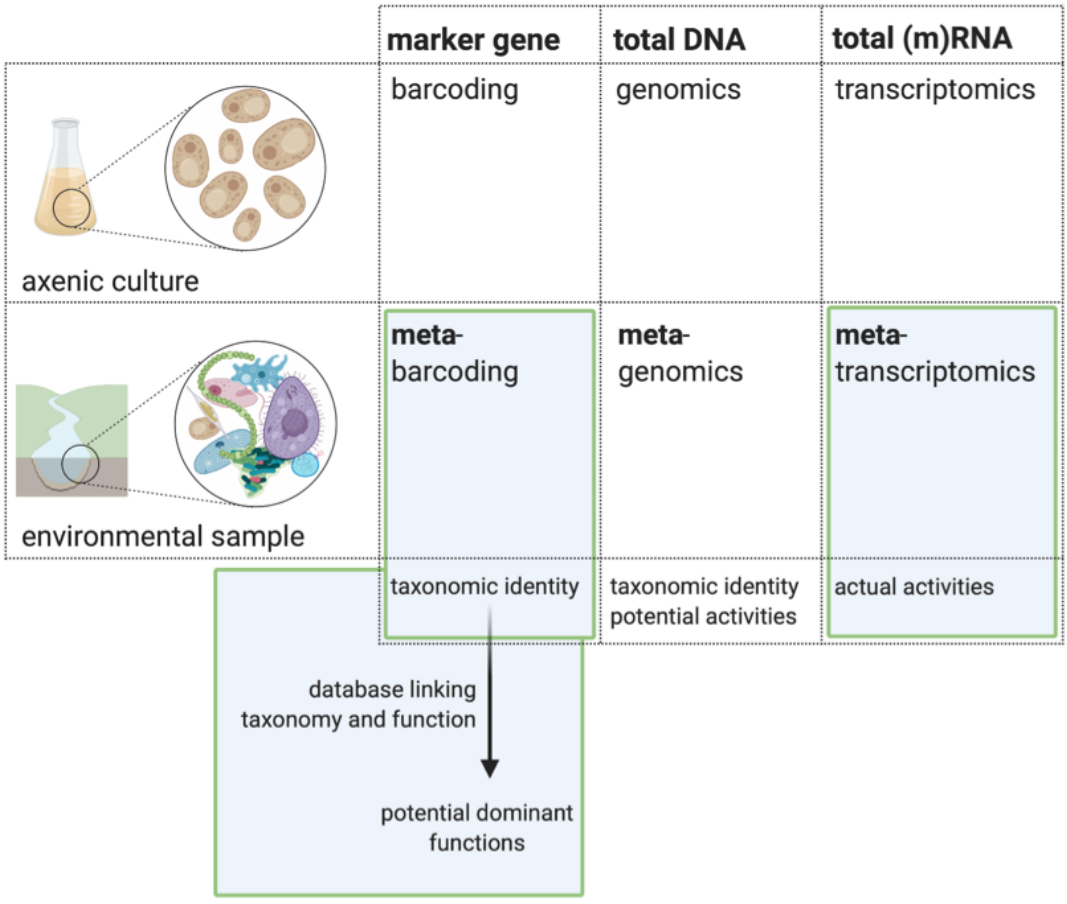


Figure 2 | Summary of some high-throughput approaches that use sequencing to directly or indirectly address questions related to the function of microbial eukaryotes. Approaches highlighted in green were used in this thesis.

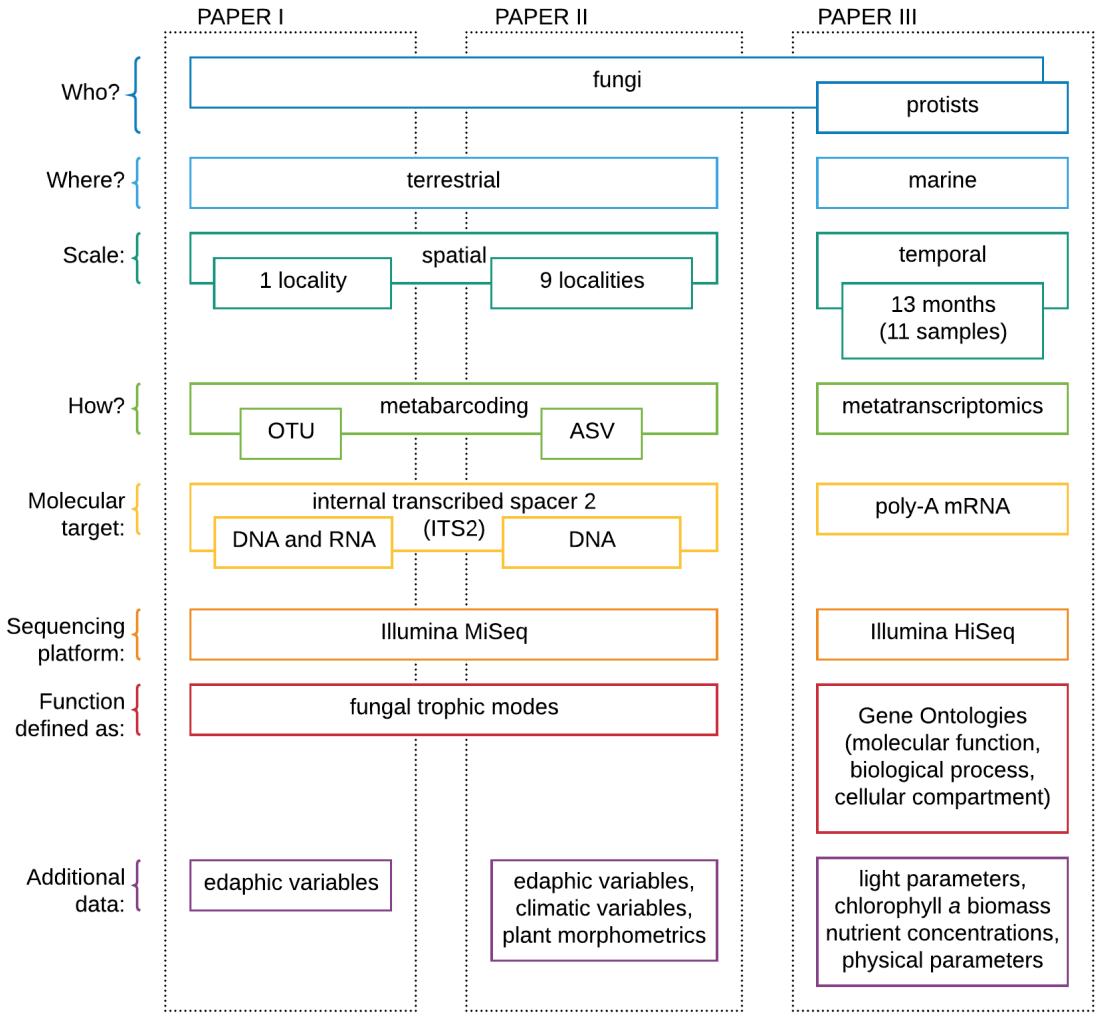


Figure 3 | An outline of approaches used in each of the case studies.

Sequencing data analyses

Although the characteristics of sequencing datasets differed and the data analysis was distinct for each of the studies (Figure 3 and 4), there were two fundamental types of steps similar in all three approaches. These are processing steps and queries against databases (Figure 4).

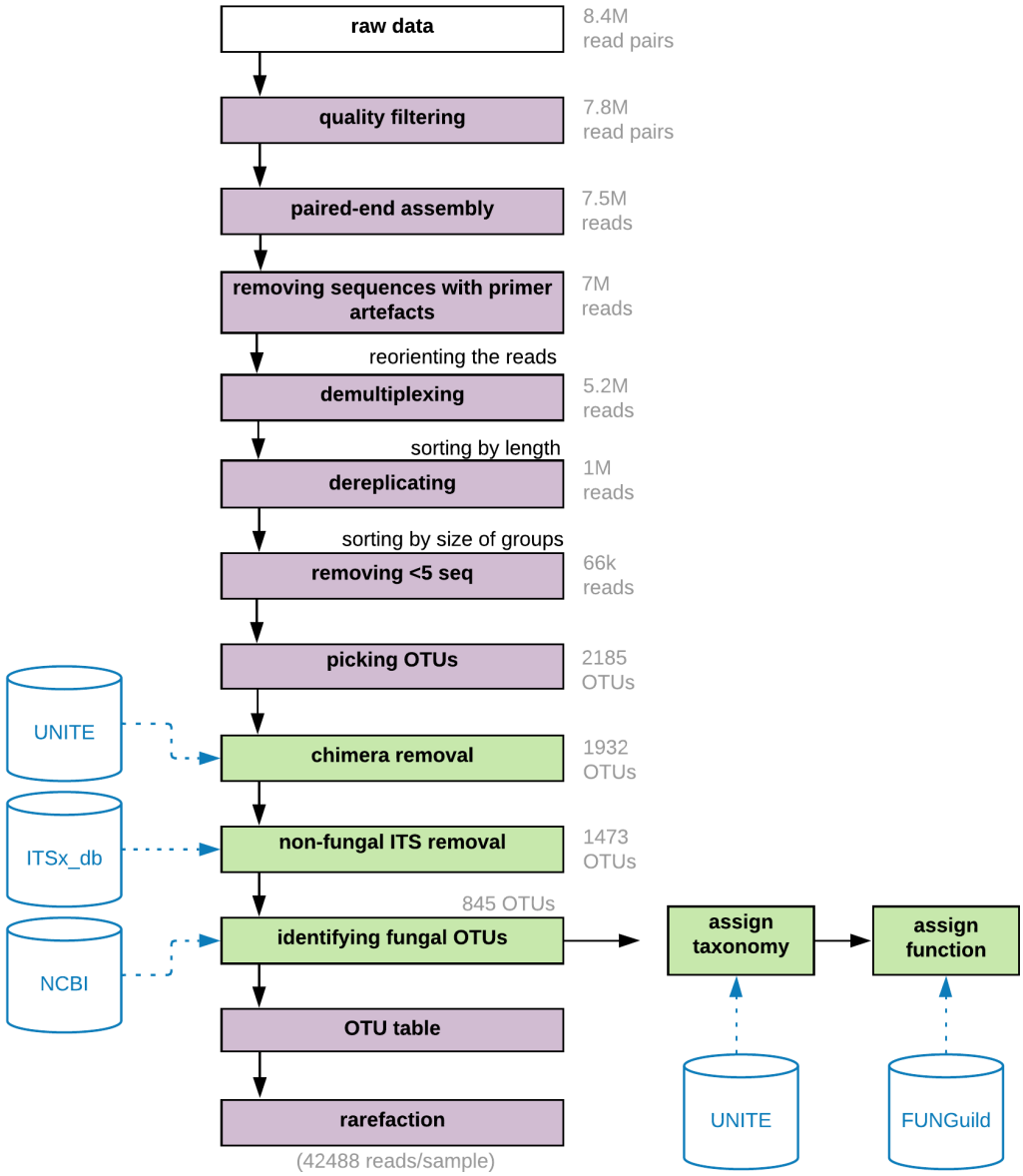

raw data
8.4M read pairs
quality filtering
7.8M read pairs
paired-end assembly
7.5M reads
removing sequences with primer artefacts
7M reads
reorienting the reads
demultiplexing
5.2M reads
sorting by length
dereplicating
1M reads
sorting by size of groups
removing <5 seq
66k reads
picking OTUs
2185 OTUs
UNITE
chimera removal
1932 OTUs
ITSx_db
non-fungal ITS removal
1473 OTUs
845 OTUs
NCBI
identifying fungal OTUs
assign taxonomy
assign function
OTU table
UNITE
FUNGuild
rarefaction
(42488 reads/sample)

Figure 4A

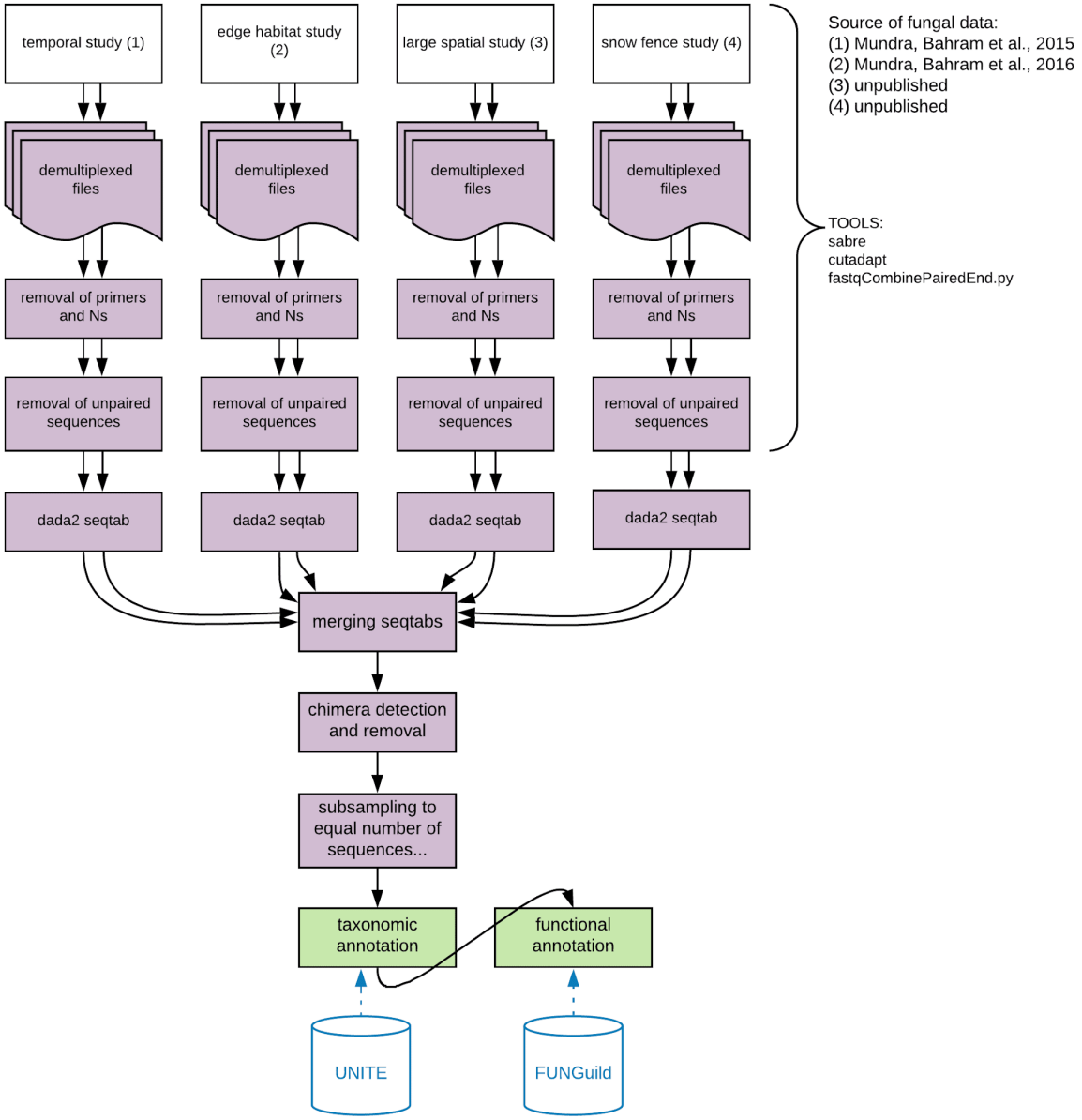

temporal study (1)
edge habitat study (2)
large spatial study (3)
snow fence study (4)
demultiplexed files
removal of primers and Ns
removal of unpaired sequences
dada2 seqtab
merging seqtabs
chimera detection and removal
subsampling to equal number of sequences...
taxonomic annotation
functional annotation
UNITE
FUNGuild
Source of fungal data:
(1) Mundra, Bahram et al., 2015
(2) Mundra, Bahram et al., 2016
(3) unpublished
(4) unpublished
TOOLS:
sabre
cutadapt
fastqCombinePairedEnd.py

Figure 4B

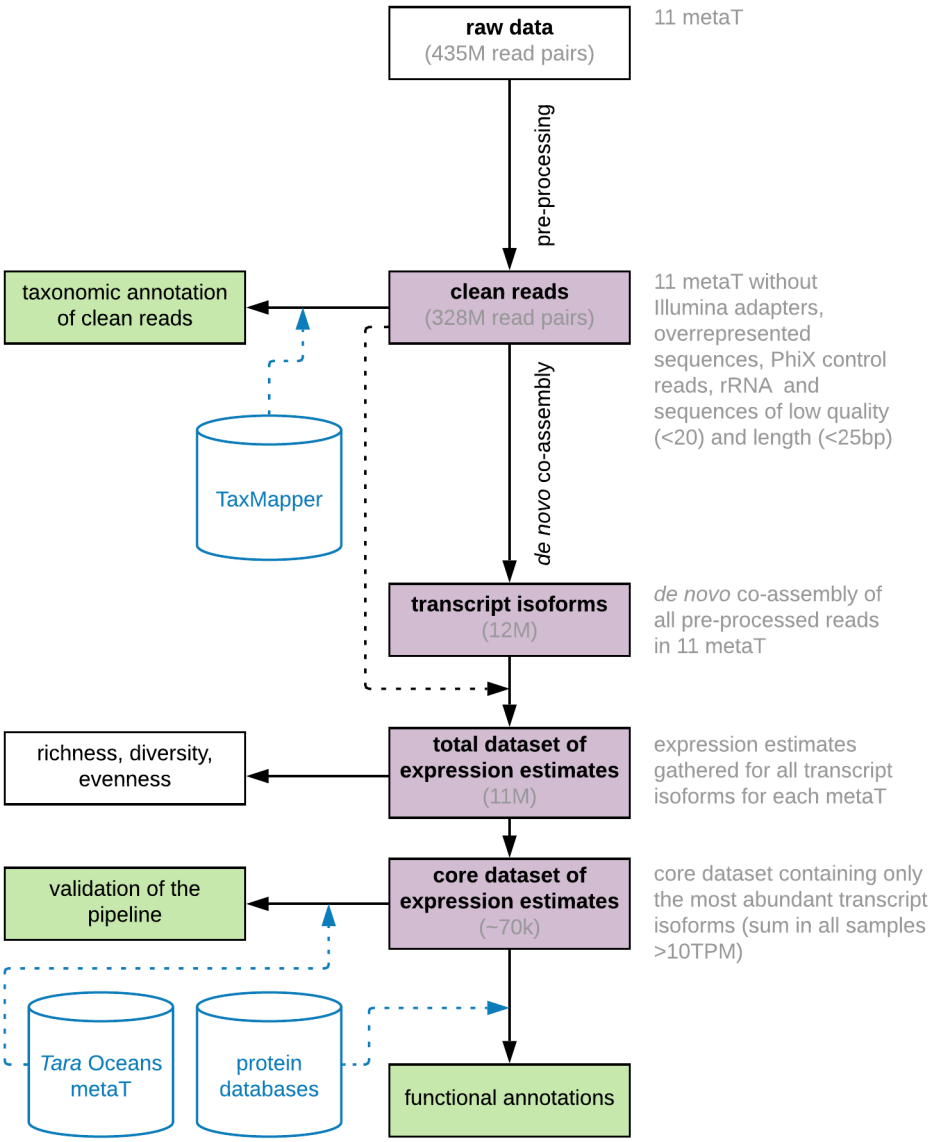


Figure 4C

Figure 4 | Summary of bioinformatics pipelines from each study indicating similarities between steps. Processing steps (purple) and queries against databases (green) some of the most crucial steps in handling various high-throughput sequencing data. Figure 4A and 4B depict two distinct workflows in metabarcoding, operational taxonomic unit and amplicon sequence variant approaches, respectively. Figure 4C shows an example of handling metatranscriptomic datasets.

Processing steps

The use of processing steps, and sometimes their order, was enforced by the methods used for generating data and a specific analytical approach. For example, samples for PAPERS I

and II were multiplexed by attaching variable length barcodes to amplified marker genes, therefore during early steps of sequencing data analyses these datasets had to be demultiplexed to decipher the sequence provenance. Demultiplexing was the first analytical step in an amplicon sequence variant pipeline (ASV, PAPER II, Figure 4B), because all other analyses required a per sample approach (Callahan et al., 2016); whereas in an operational taxonomic unit pipeline (OTU, PAPER I, Figure 4A) it could be done in later stages, so that the sample non-specific bulk removal could be handled first saving computational resources (Bálint et al., 2014). Most of the early processing steps removed reads or their parts that were non-informative for inferring their biological meaning (Table 1). Other processing steps ensured better alignment opportunities due to sequence pairing, therefore providing an increased length that was not possible to capture within the 300 bp limit of Illumina MiSeq sequencing technology. All of these were ultimately used to enable the analysis and decrease the unnecessary resources needed for next steps of the analyses.

Table 1 | Overview of data that were removed from sequencing datasets in processing steps of bioinformatics pipelines in PAPER I, II & III.

Type of removed data	P.I	P.II	P. III
primers, barcodes and adapters	x	x	x
reads with ambiguous bases	x	x	x
reads with inappropriate length	x	x	x
unpaired reads	x	x	x
chimeric reads	x	x	
reads/parts of reads with low-quality	x	x	
reads that were ‘too’ rare	x	x	
non-fungal ITS	x		
reads that were ‘too’ numerous			x
Phi X control			x
overrepresented reads			x
rRNA sequences			x

Queries against databases

A nucleic acid sequence gains its human-interpretable biological meaning when it is identified and labelled. To identify the sequence, it needs to be compared with already known and annotated sequences to assess a level of similarity between them (databases used in this

thesis were gathered together in Table 2). There are many methods for sequence comparisons that differ fundamentally in their analytical approaches and implementation, however, discussing this topic goes far beyond the scope of this thesis.

Table 2 | Overview of databases/datasets used in the bioinformatics pipelines in PAPERS I, II & III.

Database name	Annotation type	P.I	P.II	P.III	Reference
ITSx_db	taxonomic	x			Bengtsson-Palme et al., 2013
NCBI	taxonomic	x			NCBI Resource Coord., 2016
UNITE	taxonomic	x	x		UNITE Community, 2019
SortMeRNA	taxonomic			x	Kopylova et al., 2012
Silva					Pruesse et al., 2007
Rfem					Burge et al., 2012
TaxMapper	taxonomic			x	Beisser et al., 2017
FUNguild	functional	x	x		Nguyen et al., 2016
UniProt	functional			x	The UniProt Consortium, 2017
Pfam	functional			x	Finn et al., 2016
eggNOG 3.0	functional			x	Powell et al., 2012
The Gene Ontology	functional			x	Ashburner et al., 2000
KEGG	functional			x	Kanehisa, 2000; 2016
***Tara* Oceans metaT**	validation			x	unpublished

Statistical analysis

Two types of samples were collected in each study: for nucleic acid extraction and for associated environmental parameters. Therefore, each study consisted of at least two types of data: nucleic acid sequences with quality scores and environmental parameters (Figure 4). Most of the environmental measurements were measured *in situ* or in the laboratory, but some were inferred from meteorological models, such as precipitation and temperature in PAPER II (Schuler & Østby, 2020). The nature of the input data in PAPERS I, II & III required the use of both univariate and multivariate statistics to explore the data and to test hypotheses. To infer causal relationships between edaphic and climatic variables, fungal diversity and plant morphometrics in PAPER II we used structural equation modelling. All the statistical methods have been executed in R (R Core Team, 2018) and are described in detail in each of the papers.

Summary of main findings

The broad study objectives were explored by establishing three case studies (Figure 3) addressing some of the most crucial knowledge gaps regarding the functional importance of microbial eukaryotes in the High Arctic.

In **PAPER I**, we looked at the possible differences inferring on ecological roles of soil fungi when using rDNA and rRNA of the same marker gene. Arctic soils characterized by limited decomposition capacity in low temperatures, are thought to be prone to prevent dead cells from decomposing and their genetic material from decay. HTS studies that use rDNA as a template do not discriminate between living and dead cells in the samples. Therefore, there is a concern that DNA-based results could disturb the current ecological interpretation of the functional identity of fungi in the soil. The samples for the study were taken from a field experimental site testing one of the predicted climate change scenarios which assumes increased snow precipitation in the Arctic. The setup of snow fences spans over two distinct vegetation types: heath and meadow. We looked at the possible differences in functional roles of fungi between the two templates at three different levels: community composition, OTU richness and read abundances. At the community composition level, we found that the vegetation type influenced soil fungal community composition more than the choice of metabarcoding template. Therefore, the functions of fungi were shaped by abiotic and probably also biotic factors developed and evolving over a long period of time in the particular location. At the read abundance level, the functional role of fungi inferred based on both templates showed similar trends irrespective of vegetation type. Symbiotrophic reads were the most abundant in each combination of template and vegetation type. However, rRNA revealed twice as many saprotrophic and functionally unassigned reads as rDNA, also regardless of the vegetation type. Although the templates differed in read abundances between trophic modes, the overall picture was very similar. More pronounced differences between the templates were revealed at taxonomic and biodiversity levels within the functional groups. At the OTU richness level, symbiotrophs showed higher mean OTU richness in rRNA, compared to rDNA. Richness, unlike other levels, was influenced by the choice of metabarcoding templates. We found no evidence of fungal community composition or richness being affected by the deep snow regime in the field experiment.

In **PAPER II**, we looked at the relationship between root-associated fungal communities of an arctic herb (n=214) facing different levels of environmental stressors in each of nine distinct localities in Spitsbergen. We were specifically interested to find out if fungal parameters mediate the influence of abiotic stressors on *Bistorta vivipara* performance and growth. Fungal

parameters were represented by diversity, ratio of symbio-to saprotrophs and community composition, separately using presence-absence and abundance ASV table. Plant morphological measurements were used as proxies for storage (rhizome volume), photosynthetic (longest leaf length) and reproductive capabilities (ratio of inflorescence to the total stem length). Three fungal parameters were used in abundance and presence-absence models in relation to a host-plant and edaphic and climatic variables: diversity, functional diversity (ratio of symbio- to saprotrophs) and a proxy for community composition. We tested seven biological hypotheses regarding relationships between these plant morphometrics, fungal parameters and abiotic factors using structural equation modelling. Models using presence-absence and abundance fungal parameters showed a distinct picture. The best-fitting presence-absence model supported our hypothesis that the fungal community composition did not impact plant parameters and additionally the ratio of inflorescence to stem length was not affected by any fungal parameters. It showed that fungal diversity (number of ASV) as well as the functional diversity (ratio of symbio- to saprotrophic ASVs) did influence plant morphometrics (rhizome volume and leaf length, respectively). Community structure was not important for *B. vivipara* measurements. The best-fitting abundance model did not find any significant relationship between fungi and plant variables. Both models showed an important contribution of temperature to fungal and plant variables that differs in direction and its magnitude. Variance in plant response to abiotic factors was on average better explained than fungal variance. Both measures of variance increased when locality was considered as a random factor in our equations.

In **PAPER III**, we looked at the gene expression patterns of marine microbial eukaryotes in a temporal perspective. Here, we tried to address a long-standing question of what happens in the sea during the polar night, therefore the particular focus of the study was on polar night. Samples were collected at the northernmost marine time series station (IsA) at 11-time points, at local noon, from 25m depth and captured plankton in the 0.45-10 µm size fraction. The study spanned over 13 months, from December 2011 to January 2013; it included two consecutive polar nights with two and three samples respectively. mRNA was extracted, reverse transcribed, amplified and sequenced. *De novo* assembled transcripts were taxonomically and functionally annotated. The functions were defined as molecular functions, biological activities and cellular compartment standardized as the Gene Ontology (GO). Environmental parameters fluctuated throughout the year. For instance, at 25m depth photosynthetically active radiation (PAR) was detectable between April and September, whereas the nutrients were depleted from May (the onset of the spring bloom) to August. Diversity and evenness of transcripts were higher during polar night than polar day; a September sample with mixed light regime had a similar number of transcripts as the polar

day average. For annotations the total dataset was subsampled to a core dataset of nearly 70 000 most abundant transcript isoforms (with a sum of transcripts per million across all samples >10). The level of taxonomic annotations of transcripts was similar throughout the study (33-42%), that left the majority of transcripts taxonomically unannotated. Alveolates dominated taxonomic annotations throughout the year. Dinophyceae transcripts dominated polar night and September samples, whereas Ciliophora transcripts were more abundant during polar day. The number of functional annotations was low. Environmental variables fitted into dissimilarity matrices of biological processes and molecular functions revealed the structural importance of light parameters (day length, declination and PAR), but not water masses or temperature. The most abundant biological processes were connected to housekeeping functions, and the majority of them were represented during polar day. Only very few the most abundant GO terms were overrepresented during polar night, such as one-carbon metabolic processes, response to stress and phototransduction. All light-dependent processes were overrepresented during polar night, except for phototransduction. Most light-dependent processes were present during polar night, beyond PAR availability period at 25m depth. Among most abundant molecular functions four categories were overrepresented during polar night: DNA binding, adenosylhomocysteine activity, photoreceptor and light-activated channel activity. Despite low levels of annotations, *de novo* assembled transcript isoforms in the core datasets mapped to *Tara Oceans* datasets, especially the Arctic samples (up to 75% of our transcript isoforms mapped to surface samples, up to 78% mapped to the deep chlorophyll maximum layer and up to 74% to the mesopelagic zone. Overall, we found that the two polar nights were similar to each other based on all the characteristics that we have looked at.

Discussion

The three case studies presented in this thesis explored microbial eukaryotes in methodological, spatial or temporal contexts increasing our knowledge of their functional importance in Svalbard. Furthermore, this thesis shed light on how the immense but unexplained variation may be further explained by looking through the lens of functionality. All the metabarcoding studies from Svalbard that were focusing on microbial eukaryotes, in particular on root-associated fungi, reported a high proportion of unexplained variation in these microbial communities. Since understanding sources of variation in the environment seems to be crucial to evaluate important factors influencing communities and underlying ecological processes, it means that these communities are not fully understood. It is necessary to identify and distinguish between possible sources of this unexplained variation. Is it stochastic or could it have been explained by using other approaches? Or perhaps there are some more crucial parameters in the environment that might play a role in explaining these unknown sources? By using high-throughput sequencing we attempted to showcase some of the possible approaches to study functions of microbial eukaryotes in the Arctic. The three studies presented here shed light on the nature of the knowledge gaps. The results from each case study were discussed in detail in corresponding papers. Here, however, I would like to emphasize the implications and context of the results in a broader perspective.

Insights from the thesis in a broader context

The comparison between rDNA and rRNA metabarcoding templates (PAPER I) was to our knowledge the first attempt to distinguish total from active fungal communities in the High Arctic soils. Previous metabarcoding attempts to distinguish between the total and active microbial communities in the High Arctic focused on soil bacteria (Schostag et al., 2015) and marine protists (Marquardt et al., 2016; Onda et al., 2017; Vader et al., 2014). Distinguishing active and total community is a particularly important issue because the vast majority of fungal metabarcoding studies in the Arctic rely only on the use of rDNA markers (e.g. Blaalid et al., 2014; Botnen et al., 2014; Davey et al., 2015; Lorberau et al., 2017; Mundra, Bahram, et al., 2015, 2016; Mundra, Halvorsen, et al., 2015, 2016). Determination of viability of microorganisms has been a non-trivial task since they were discovered (Emerson et al., 2017). Culture-dependent methods, such as growing microorganisms on agar (Postgate, 1969), prove unequivocally that they are alive when forming colonies. However, soil is a complex environment containing many species that are difficult or yet impossible to culture, as well as microorganisms in various seemingly low metabolic or resting stages that will not grow (Davey, 2011). In the High Arctic the combination of environmental conditions in soils favours

preservation of dead organic matter, including parts of fungal cells containing DNA, therefore possibly biasing the results. The low levels of nutrients in some High Arctic soils along with low temperatures may also favour low levels of metabolism, keeping cells in resting stages or spores (Robinson, 2001). The high congruence in community composition of total and active communities of symbio-, sapro and pathotrophs in our results may indicate that it does not matter which metabarcoding template should be used (PAPER I). However, there are several problems with this assumption. First of all, the edaphic drivers of community composition differed for total and active communities. This is probably also related to a response to snow regime only by rRNA-based community composition, although the overall effects were very small; similar trends however, could not be detected using rDNA template. Secondly, our results explore community composition at one time point and it is not clear how the total and active soil fungal communities fluctuate throughout the year, especially in the onset of edaphic changes introduced by snow fences (Cooper, 2014; Mörsdorf et al., 2019). Recent evidence suggests that at low soil temperatures bacterial rRNA can have a very slow turnover, i.e. 16 days when kept at 4°C and even 215 days at -4°C (Schostag et al., 2020). At the time of sampling, the daily mean temperature in the topsoil was higher than 4°C (Mundra, Halvorsen, et al., 2016). If the same turnover rate applied to fungal rRNA, then the rRNA-based results would provide a snapshot of fungi active also for up to two weeks prior to actual sampling. Different time of sampling, especially during the major edaphic changes introduced by the snow fence setup (Mörsdorf et al., 2019), would most likely also blur the picture of the active community composition, and affect our ecological conclusions on fungal trophic modes. The slow rRNA turnover probably blurs some microbial richness responses, especially at times of relatively fast fungal shifts due to sudden changes in the environmental parameters. Besides temperature, there are also other environmental parameters that could further alter the rRNA turnover, such as grazing by insects etc. Therefore, the use of either metabarcoding template carries some drawbacks. The choice of a template and study design should ideally be fine-tuned depending on the aims of a study. In either case, analyses grouping taxons (OTUs) in functional groups proved to be a valuable tool to address ecological issues.

The role of biodiversity and its effects on ecosystem processes and other organisms still remains an open question (Hooper et al., 2005; Winfree, 2020). Similarly, the role of diversity of microbiota and its impact on the host well-being are still debated (Berg et al., 2017; Valdes et al., 2018). There seems to be a general consensus that the decline of microbial diversity could negatively affect the system, but it is highly context-dependent and difficult to discuss with unknown levels of functional redundancy present in the ecosystem. Symbiotic associations between plants and root-associated fungi are regarded as crucial in Arctic soils and other nutrient-limited environments, mainly because they supply 61-88% of nitrogen found

in plant tissues (Hobbie & Hobbie, 2006). However, the importance of fungal diversity and communities for plant morphometrics in the Arctic has rarely been studied. Therefore, it is not clear if the plant benefits from an increase in the number of symbiotic partners putatively providing more resources but at the same time perhaps increasing the plant's energetic costs of maintaining the symbiotic relationship. Additionally, among some of the most uncertain issues concerning biodiversity is the relationship between community structure, taxonomic and functional diversity (Hooper et al., 2005; Inkpen et al., 2017). We took this a step further and tested these three characteristics of root-associated fungi in relation to a host plant (PAPER II). The relationship between morphometrics of *Bistorta vivipara,* its root-associated fungi and environmental parameters revealed valuable insights on these tripartite dynamics in the High Arctic discussed in detail in PAPER II. An unexpected negative impact of a functional parameter (the ratio belonging to symbio- and saprotrophic ASVs) on *B. vivipara's* leaf length requires more research to understand if there is a mechanistic explanation of this phenomenon. It could be interesting to see if plants that are more competitive than *B. vivipara* would display similar trends in their root-associated fungi. However, it is worth asking whether the outcomes of the study would have remained the same if we had looked at the active fraction of fungal community. Do all active root-associated fungi transport nitrogen to plant roots? Are there other local edaphic factors that could enhance preservation of genetic material in dead organic matter that can be abundant in the soil and vicinity of plant roots?

The establishment of the world's northernmost time series station allowed for following temporal changes of a community composition of marine microbial eukaryotes that are driven by strong seasonal patterns (Marquardt et al., 2016). Similarly, the analysis of a community-level gene expression of small microbial eukaryotes showed that primarily access to light but also nutrients is tightly linked to molecular functions of these communities (PAPER III). Two consecutive polar nights exhibited similar relative abundances of functional annotations. The level of similarity in functions between the two polar nights was striking and to some extent unexpected because of the differences in nutrients' concentrations and temperature between the two polar nights. It seems that the prolonged lack of light (e.g. PAR) is such a fundamental environmental factor, that it triggers a very similar functional response and overrides the response to other environmental variables. Therefore, it acts as a reset for the marine arctic system, before it takes off next season when the light comes back. Perhaps that could be also an important factor controlling which species of protists shifting northwards following climate change, can survive. High level of unannotated transcripts found in our study could encourage further research and bioprospecting efforts in the Arctic, especially focusing on pico- and nanoeukaryotic plankton. This high proportion of the unknown transcripts coincides with a distinct biogeographic hotspot of viromes in the Arctic Ocean (Gregory et al., 2019). So far,

there is no supporting evidence that distinct virome and metatranscriptome of microbial eukaryotes are connected; however, perhaps it could provide a valuable foothold for looking at the genomic relationship between viruses and microbial eukaryotes in the Arctic Ocean. It is necessary to keep in mind that an important difference between eukaryotic organisms, opposed to bacteria and archaea, gene expression control takes place at any moment after it has been produced, including post-transcriptional, translational or even post-translational level (McCarthy, 1998). Eukaryotic cells are not as restricted in energy and resource use as bacteria and archaea, so they produce transcripts that can be destroyed at later stages if necessary (Madigan et al., 2014). As a consequence, there is a substantial probability that an expressed gene will not end up as a functional protein - something that seems to be often forgotten or omitted when discussing eukaryotic metatranscriptomic studies. In the presence of stressful factors in the cell's environment mRNA stability can be altered, either shortened or prolonged, depending on the gene (Fan et al., 2002) that could also impact our results and conclusions, however, it is not clear to what extent. Therefore, it could be interesting to follow up with a similar study looking at the same system from a protein or metabolite perspective. Would we see the similar results? How many transcripts are successfully translated into functional proteins? Additionally, it could also be interesting to look at the influence of temperature on mRNA half-life in marine protists and its determinants.

The importance of the functional approach in face of environmental changes

Among other goals, microbial ecology aims to understand how microbial communities respond to perturbations in their environment (Konopka, 2009; Robinson et al., 2010). These perturbations can be short-term (pulses) or persistent (presses); the consequences of the latter are often more difficult to observe (Bender et al., 1984; Shade et al., 2012). Many environmental shifts taking place in the Arctic, such as rising temperatures, are an example of gradual, large-scale presses that may not disturb the structure of microbial or fungal communities in the first stages of the shift. In fact, many studies report either no or weak responses of microbial communities to experimentally introduced changes in the environment, especially in Svalbard's terrestrial habitats (e.g. Lorberau et al., 2017, Wutkowska et al., *in prep*). Lack of responses could be a sign of community stability (resistance or resilience) or observers' fault. The latter could include the quantification of microbial communities with a delayed timing in respect to the onset of response, inadequate scales or detection tools to confidently detect responses of microbial communities to altered environmental conditions. Moreover, one could reason that the large proportion of unexplained variation masks the existing response. This is an important argument for understanding and structuring the unexplained variation, for instance by moving from taxonomic to functional framework.

Nevertheless, the immediate microbial response, if not lethal, should be visible as altered gene expression or physiological patterns - a typical response of all cells to changes in its environment. Therefore, it seems intuitive to try to detect which molecular functions, especially which expressed genes, fluctuate under the influence of an environmental factor of interest. Natural systems are rich in microbial species intertwined in tightly connected assemblages full of genomic information. By definition “meta-omics” methods look at the preselected information available for the whole community of organisms and might not be suitable for detecting single species responses. Many cells in microbial communities belonging to different taxonomic groups could respond in the same way to the same stressor. Additionally, certain genes can be expressed in a cell as a non-specific cellular response to stress (i.e. expressing genes coding non-specific stress responses such as heat-shock proteins, chaperones etc. which have been described in PAPER III).

Microorganisms, including microbial eukaryotes, differ when it comes to plasticity and stress tolerance (Orosz et al., 2018; Slaveykova et al., 2016), which is of critical importance when the magnitude and temporal scale of an environmental change does not allow for evolving towards coping with the particular stressor. Therefore, these changes in the microbial community might eventually cause extinctions or local dominance of some species. Single molecular “species”, typically one out of thousands OTUs/ASVs in metabarcoding studies, might respond in a weak manner, not really visible for statistical tools. However, when stacking information on the response of many species (ASVs or OTUs) that acquire or use resources in a similar fashion then it could potentially be easier to detect functional shifts in microbial communities under experimental settings. Finding out the best methods and ways to look at microbial eukaryotes and other organisms from the functional perspective could provide a worthwhile foundation for microbial ecology theory in the Arctic and beyond.

Could one method rule them all?

Virtually all the methods used in science are inherently flawed, thus have their limitations. High-throughput sequencing methods used in (microbial) ecology are not an exception (Lemos et al., 2011; Lindahl et al., 2013). The limitations associated with studies based on DNA or complementary DNA sequencing from bulk environmental samples occur at each step, starting from experimental or sampling design to the last step of data analysis and interpretation of the outcomes (Lindahl et al., 2013; Peimbert & Alcaraz, 2016). Specific limitations of the studies were discussed and carefully acknowledged during the interpretation of the results, however, it is important to elaborate on advantages and disadvantages of

methods used in this thesis to further explore functional importance of microbial eukaryotes in Svalbard and in the Arctic.

Metabarcoding

Metabarcoding (PAPERS I & II) is currently a commonly used high-throughput method in microbial ecology for biodiversity studies in virtually all environments (Santoferrara et al., 2020). Its routine use for more than a decade resulted in a plethora of publications describing biases introduced by methodological choices at each step of the metabarcoding study (reviewed in Nilsson et al., 2019). For instance, the choice of primers and marker genes (e.g. their use in fungi was reviewed in Raja et al., 2017) or type of pipeline used for sequencing data analysis (Anslan et al., 2018). Moreover, extensive development of algorithms, bioinformatic tools and even complete pipelines lead to fairly well-established guidelines, however, there are still debates on some issues. Despite all of the methodological considerations and biases, the speed, breadth and depth of information on the identity of organisms in the samples using metabarcoding are incomparable with any previous classic taxonomic methods. However, the usefulness of metabarcoding is strongly linked to the quality of databases used for identification and the level of knowledge connected to each entry, such as functional traits.

Functional annotations of metabarcoding data, in the form of assigning the taxa to broad categories, such as trophic modes, was carried out by matching taxonomic identity of sequences with a database linking taxonomy and functions reported in the literature (Nguyen et al., 2016). However, metabarcoding studies in Svalbard typically detect a high proportion of taxonomically unassigned sequences with no matches to curated comprehensive taxonomic databases. This high proportion decreases the number of functional assignments. Fungal trophic modes or guilds are encoded in multiple genes; therefore, they are not likely to change. Yet, there are fungal species that are difficult to categorize to only one fungal trophic mode, such as *Mycena* (Thoen et al., 2019) or members of Sebacinales (Oberwinkler et al., 2013; Weiss et al., 2004; Weiß et al., 2016), with overall poor resolution of taxonomic assignment in samples from Svalbard. In general, symbiotrophic fungi are able to decompose organic and could thus also be classified as belonging to the saprotrophic trophic mode (Nicolás et al., 2019; Shah et al., 2016). Probably the switching or or exhibiting more than one trophic mode are far more common phenomena that have not yet been resolved due to tedious research capable of answering this question. Additionally, it is not clear how environmental or internal conditions modulate a switch between trophic modes or contribution to organic carbon acquisition in fungi that exhibit more than one trophic mode. Hence, these types of functional

annotations are not free of flaws. Nevertheless, they still provide valuable insights into broad ecological processes in the soil and root-associated fungi that can be used for testing hypotheses or generating new ones that could be then tested using different methods.

The use of similar functional annotations for marine microbial eukaryotes is probably less informative due to the ability of many species to switch between auto- and heterotrophy in distinct ways (Mitra et al., 2016; Stoecker & Lavrentyev, 2018). Instead, a commonly used classification of marine plankton is based on their size or general biogeochemical roles such as photo-, mixo-, heterotrophs and parasitoids (Caron et al., 2017).

Metatranscriptomics

Among high-throughput sequencing of environmental samples, metatranscriptomics is less frequently used, because of the difficulty and complexity of the procedure and data analysis (reviewed in detail in Peimbert & Alcaraz, 2016). Interestingly it provides unprecedented insights into actual activities performed at a given time by all the living organisms in the sample at once. By extraction of all the mRNA present in the sample, the method gives access to information on all the expressed genes. Therefore, it is an excellent tool to capture a snapshot of community-level molecular response patterns that are far more sensitive to changes in the environment (e.g. those connected to climate changes Mackelprang et al., 2016) than monitoring community composition. During post-transcriptional modifications a sequence of nucleotides containing only adenines is added to the 3' end of the majority of eukaryotic transcripts, therefore it is easy to specifically select them after RNA extraction using a poly(A)-tail selection procedure. There have been only a few metatranscriptomics studies in Svalbard with the focus on microbial eukaryotes (e.g. Vader et al., 2018) and there can be several reasons for that. There are indeed a number of challenges and issues to consider before launching a metatranscriptomics study.

Firstly, quick sampling procedures and immediate preservation in liquid nitrogen are required. The average lifespan of mRNA varies between organisms, types of cells and genes, but in unicellular organisms the mRNA half-life oscillates roughly between 3-90 minutes (Bernstein et al., 2002; Wang et al., 2002). These fundamental characteristics of transcripts heavily constrains experimental design and sampling procedures. In ideal conditions samples intended for rRNA or mRNA analysis should be immediately flashfrozen in -80°C which is difficult to assure in remote locations, especially terrestrial ones where all the equipment often needs to be carried. Tanks with liquid nitrogen are commonly used in marine sampling on ships, however that does not always save valuable time for mRNA not to decay. First of all,

because of sampling time, especially when samples are collected from deep parts of the water column. Secondly, it is important to assure comparable volumes of water for metatranscriptomic studies and enough cellular biomass for representative samples. In polar night or deep parts of water column cell counts of microbial eukaryotes are low compared to spring bloom, therefore more sea water needs to be filtered to gather enough biomass which is a time-consuming process (30 liters of sea water were filtered for each date in PAPER III). The majority of the RNA extracted in bulk samples is ribosomal (Kopylova et al., 2012; Kukurba & Montgomery, 2015), therefore a relatively high biomass of cells is usually required to capture their putative molecular functions.

Secondly, some environments are challenging for metatranscriptomics analyses due to their complexity, i.e. high diversity of microorganisms with low relative abundances (Shakya et al., 2019). In this case it is difficult to obtain the optimal or even sufficient depth of sequencing to capture medium or low abundant transcripts from the environment (Peimbert & Alcaraz, 2016; Westreich et al., 2016). Soil is described as a complex environment inhabited by many groups of organisms and many chemical inhibitors for nucleic acid extraction. Compared to sampling cells from sea water it is difficult to pinpoint certain size fractions of organisms in soil samples.

Thirdly, understanding metatranscriptomics data relies heavily on comparison of nucleotide or amino acid sequences to databases. Genomes of very few eukaryotic organisms from polar environments have been sequenced or derived as metagenome-assembled genomes from metagenomic studies. Thus, it is difficult to find appropriate databases that would contain genomic templates to map against their respective functions.

What else is there?

Metabarcoding can help to infer functions of organisms indirectly, whereas thanks to metatranscriptomics we can assign functions to expressed genes. However, there are other methods that could be used to obtain functional information of the microbial eukaryotic communities depending on the specific aim of the study. Metagenomics is the way to infer all the genomic content of the sample and therefore functional potential of the genes present in an environment (Handelsman, 2004). This high-throughput sequencing method will gain much more attention in the coming years in the Arctic research because of its capacity to tackle some major unknowns in microbial ecology (Edwards et al., 2020). Beyond nucleic acid sequencing, the knowledge on functions of microbial communities could be enhanced using methods based on recognizing other types of molecules, such as proteins (metaproteomics; Maron et al., 2007) or small-molecule metabolites (metabolomics; Oliver, 1998). These

methods rely on using physical separation, mass spectroscopy or nuclear magnetic resonance spectroscopy to identify molecules in the samples (Peisl et al., 2018; Yuqiu Wang et al., 2020). Despite decreasing diversity compared to transcripts, environmental samples can be rich in both proteins and metabolites from different organisms making the analysis and the result interpretation tremendously complex (Saito et al., 2019). Studies of functions of microbial communities at the level of molecules are inherently intricate, however at the same time they provide a valuable asset in the microbial ecology toolbox.

The curious case of unexplained variation

The source of unexplained variation in the majority of microbial eukaryotic studies is either stochastic, comes from methodology or limited present understanding of the functioning of these communities. Compared to how ecologists understand assemblages of plants or animals, where ecological requirements and interactions between specific species are studied in length, little is known about most microbial species which are lumped in groups. But one could assume that there are at least as many interactions, dependencies etc. that are modulated by environmental factors for microbes as for macro-organisms. Knowledge of function derived from taxonomic annotations does not seem to decrease the unexplained variability in the ecosystem even with the use of a different template (PAPER I). However, as it has been demonstrated, the functional approach proved to be a valuable way to look for mechanisms underlying ecological processes (PAPER II and PAPER III). Perhaps it is suggesting that the way we measure parameters, for instance in fungal studies, suits vegetation research and are measured at scales inadequate for microbial ones (Madigan et al., 2014). There is a limited number of abiotic factors that are measured in the environment that are intended to explain patterns of microbial communities. It is possible that there are other, perhaps more important factors that would explain the unexplained variation. Or perhaps the available databases of functions do not yet encompass all functions, which was the case in matching functional annotations in PAPERS I, II and III. For instance, in PAPER III, most of the transcripts isoforms were found in *Tara Oceans* dataset but not in available databases used for functional annotations. On the other hand, already identified environmental variables could be measured in an inadequate manner or scale, not considering small scale gradients that might be of crucial importance for microorganisms (Welch et al., 2016). Additionally, the standard approach to study design in high-throughput sequencing methods of microbial communities is to derive a sampling scheme that would be representative for a certain habitat (Zinger et al., 2019). This involves mixing randomly picked volumes of samples. Although it assures adequate representation of species present in a habitat/plot, it does not consider spatial structure of a particular microbial community or

microbial habitat. Despite the above limitations, both functional approaches which use metabarcoding and metatranscriptomics, brought in substantial knowledge on diversity and functioning of arctic microbial eukaryotes and are crucial for building the ecological theory in microbial ecology. Yet, neither metabarcoding nor metatranscriptomics, seem to provide enough information in a stand-alone mode anymore and perhaps neither will suffice independently in future studies of functions of microbial eukaryotes in the Arctic.

Future perspectives – a wish list

There are many future objectives that could follow this thesis in order to enhance understanding of the functional importance of microbial eukaryotes in Svalbard and in other places in the Arctic. As it was outlined in this thesis, looking at functional aspects of microbial eukaryotes in the Arctic involves linking knowledge and skills from many disciplines, therefore the advancements in this field would ideally require enhancements in many separate areas. Some of the suggestions gathered during this project have been grouped into categories and outlined below.

- Let's integrate methods and disciplines

Progress in the environmental microbiology/microbial ecology and ultimately the whole (Arctic) ecology requires integration of methods and separate scientific disciplines. Nucleic acid sequencing accompanied with additional data on biological, physiological, ecological and biogeochemical dynamics in the environment from the studied system would enhance confidence in the data and interpretability of the results. The use of theoretical modelling in combination with observational studies and tests in controlled factorial experiments (Bradley et al., 2016) would increase the predictive capabilities in the system.

- Let's sample beyond summer

The above-ground plant growing season in the High Arctic spans roughly for ¼ of a year. This time coincides with a disproportionate number of sampling for microbial ecology studies compared to other months, and especially polar night. On average, 70% of arctic plant biomass is located below the ground (Poorter et al., 2012) and recent studies on plant phenology revealed that plant "below-ground season" is 50% longer compared to what is observed above the ground (Blume-Werry et al., 2016). Therefore, root-associated microbial eukaryotes are most likely also found active in this extended period and probably even longer - throughout the year. Temporal sampling however, especially in terrestrial habitats, seems to

present many challenges, including accessibility of the sampling sites, lack of relatively non-destructive sampling methods of the soil in experimental plots sampled many times throughout the year etc. There is also a strong need for more arctic marine winter sampling on a regular basis at already established time series stations, such as IsA and beyond, to cover more habitats with distinct influences of water masses.

- Let's learn the scales and represent all the habitats

Many studies of microbial eukaryotes in the Arctic use samples collected in a rather random and unplanned manner that was not designed to encompass spatial or temporal information. There can be many reasons for that including logistic difficulties and high costs of logistics. To understand (eukaryotic) microbial processes, it is necessary to reveal how they change in environmental gradients in a context of space and time, as well as what are the appropriate resolutions to accurately measure both. Some types of habitats in the Arctic have very low or no sampling coverage, which means that they are white spots on the microbial eukaryotic maps.

- Let's link land and sea

Terrestrial and marine habitats in the Arctic are both affected by rapid climate changes, however, the relationship between the two are biologically seldom studied together in a coherent framework (Webb, 2012). Despite accounting for 1% of total oceans volume, 10% of the global river discharge ends in the Arctic Ocean (Timmermans & Marshall, 2020). Yet, more terrestrial input is transported to the sea with melting glaciers and increased river flows with the warming, as well as more precipitation being observed on land and geomorphology changes. Biogeochemical cycles cross the borders of these habitats all the time. Perhaps a unifying framework of a functional approach combined with measures of process rates could enhance understanding of increased impacts of the climate changes on the whole Arctic biome.

- Let's get to (really!) know who is there and what they can (really!) do

The majority of the microbial eukaryotic species in the Arctic do not have a known genetic makeup, life history or physiology across their lifespan. Many of them are difficult to tell apart from microbial communities that they live in and to be grown in cultures. Therefore, it is difficult to understand their real contribution to ecosystem processes across their lifespan. Getting to know these single species would enhance understanding of their genetic content that impacts their biology, physiology and ecology. It could be also beneficial to look at the population level to address species variability in the natural environment that does not take into consideration variability of the populations and their plasticity in responses to various physicochemical

changes. Culture-dependent or single-cell methods using '-omics', resource requirements in different conditions would build up functional aspects of these microbial eukaryotes. Moreover, it could be a chance to understand how marker gene(s) read abundances correlate with cell volume and biomass. Zooming into the members of microbial eukaryotic communities would provide improved detection of species (van der Linde et al., 2012) and interpretation of high-throughput sequencing results.

Works cited

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., del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., … Zhang, Q. (2018). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*, jeu.12691. https://doi.org/10.1111/jeu.12691

Andrews, J. H. (2017). *Comparative Ecology of Microorganisms and Macroorganisms*. Springer New York. https://doi.org/10.1007/978-1-4939-6897-8

Anslan, S., Nilsson, R. H., Wurzbacher, C., Baldrian, P., Tedersoo, L., & Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, *39*, 29–40. https://doi.org/10.3897/mycokeys.39.28109

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Azúa-Bustos, A., González-Silva, C., Mancilla, R. A., Salas, L., Palma, R. E., Wynne, J. J., McKay, C. P., & Vicuña, R. (2009). Ancient Photosynthetic Eukaryote Biofilms in an Atacama Desert Coastal Cave. *Microbial Ecology*, *58*(3), 485–496. https://doi.org/10.1007/s00248-009-9500-5

Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., & Tedersoo, L. (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *The ISME Journal*, *10*(4), 885–896. https://doi.org/10.1038/ismej.2015.164

Balasundaram, S. V., Engh, I. B., Skrede, I., & Kauserud, H. (2015). How many DNA markers are needed to reveal cryptic fungal species? *Fungal Biology*, *119*(10), 940–945. https://doi.org/10.1016/j.funbio.2015.07.006

Bálint, M., Schmidt, P.-A., Sharma, R., Thines, M., & Schmitt, I. (2014). An Illumina metabarcoding pipeline for fungi. *Ecology and Evolution*, *4*(13), 2642–2653. https://doi.org/10.1002/ece3.1107

Bar-On, Y. M., & Milo, R. (2019). The Biomass Composition of the Oceans: A Blueprint of Our Blue Planet. *Cell*, *179*(7), 1451–1454. https://doi.org/10.1016/j.cell.2019.11.018

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Basile, W., Salvatore, M., Bassot, C., & Elofsson, A. (2019). Why do eukaryotic proteins contain more intrinsically disordered regions? *PLOS Computational Biology*, *15*(7), e1007186. https://doi.org/10.1371/journal.pcbi.1007186

Beisser, D., Graupner, N., Grossmann, L., Timm, H., Boenigk, J., & Rahmann, S. (2017). TaxMapper: An analysis tool, reference database and workflow for metatranscriptome analysis of eukaryotic microorganisms. *BMC Genomics*, *18*(1), 787. https://doi.org/10.1186/s12864-017-4168-6

Bender, E. A., Case, T. J., & Gilpin, M. E. (1984). Perturbation Experiments in Community Ecology: Theory and Practice. *Ecology*, *65*(1), 1–13. https://doi.org/10.2307/1939452

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A. S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, n/a-n/a. https://doi.org/10.1111/2041-210X.12073

Berg, G., Köberl, M., Rybakova, D., Müller, H., Grosch, R., & Smalla, K. (2017). Plant microbial diversity is suggested as the key to future biocontrol and health trends. *FEMS Microbiology Ecology*, *93*(5). https://doi.org/10.1093/femsec/fix050

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, *99*(15), 9697–9702. https://doi.org/10.1073/pnas.112318199

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, *27*(4), 233–243. https://doi.org/10.1016/j.tree.2011.11.010

Bjorbækmo, M., Carlsen, T., Brysting, A., Vrålstad, T., Høiland, K., Ugland, K., Geml, J., Schumacher, T., & Kauserud, H. (2010). High diversity of root associated fungi in both alpine and arctic Dryas octopetala. *BMC Plant Biology*, *10*(1), 244. https://doi.org/10.1186/1471-2229-10-244

Blaalid, R., Carlsen, T., Kumar, S., Halvorsen, R., Ugland, K. I., Fontana, G., & Kauserud, H. (2012). Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing: PRIMARY SUCCESSION OF ROOT-ASSOCIATED FUNGI. *Molecular Ecology*, *21*(8), 1897–1908. https://doi.org/10.1111/j.1365-294X.2011.05214.x

Blaalid, R., Davey, M. L., Kauserud, H., Carlsen, T., Halvorsen, R., Høiland, K., & Eidesen, P. B. (2014). Arctic root-associated fungal community composition reflects environmental filtering. *Molecular Ecology*, *23*(3), 649–659. https://doi.org/10.1111/mec.12622

Blume-Werry, G., Wilson, S. D., Kreyling, J., & Milbau, A. (2016). The hidden season: Growing season is 50% longer below than above ground along an arctic elevation gradient. *New Phytologist*, *209*(3), 978–986. https://doi.org/10.1111/nph.13655

Botnen, S., Vik, U., Carlsen, T., Eidesen, P. B., Davey, M. L., & Kauserud, H. (2014). Low host specificity of root-associated fungi at an Arctic site. *Molecular Ecology*, *23*(4), 975–985. https://doi.org/10.1111/mec.12646

Bouchez, T., Blieux, A. L., Dequiedt, S., Domaizon, I., Dufresne, A., Ferreira, S., Godon, J. J., Hellal, J., Joulian, C., Quaiser, A., Martin-Laurent, F., Mauffret, A., Monier, J. M., Peyret, P., Schmitt-Koplin, P., Sibourg, O., D'oiron, E., Bispo, A., Deportes, I., … Ranjard, L. (2016). Molecular microbiology methods for environmental diagnosis. *Environmental Chemistry Letters*, *14*(4), 423–441. https://doi.org/10.1007/s10311-016-0581-3

Bradley, J. A., Anesio, A. M., & Arndt, S. (2016). Bridging the divide: A model-data approach to Polar and Alpine microbiology. *FEMS Microbiology Ecology*, *92*(3). https://doi.org/10.1093/femsec/fiw015

Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., Eddy, S. R., Gardner, P. P., & Bateman, A. (2012). Rfam 11.0: 10 years of RNA families. *Nucleic*

Acids Research, *41*(D1), D226–D232. https://doi.org/10.1093/nar/gks1005
Calow, P. (1987). Towards a Definition of Functional Ecology. *Functional Ecology*, *1*(1), 57. https://doi.org/10.2307/2389358
Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy, C., Bell, C. J., Bharti, A., Dyhrman, S. T., Guida, S. M., Heidelberg, K. B., Kaye, J. Z., Metzner, J., Smith, S. R., & Worden, A. Z. (2017). Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews Microbiology*, *15*(1), 6–20. https://doi.org/10.1038/nrmicro.2016.160
Caron, D. A., Worden, A. Z., Countway, P. D., Demir, E., & Heidelberg, K. B. (2009). Protists are microbes too: A perspective. *The ISME Journal*, *3*(1), 4–12. https://doi.org/10.1038/ismej.2008.101
Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., Behrenfeld, M. J., Boetius, A., Boyd, P. W., Classen, A. T., Crowther, T. W., Danovaro, R., Foreman, C. M., Huisman, J., Hutchins, D. A., Jansson, J. K., Karl, D. M., Koskella, B., Mark Welch, D. B., … Webster, N. S. (2019). Scientists' warning to humanity: Microorganisms and climate change. *Nature Reviews Microbiology*, *17*(9), 569–586. https://doi.org/10.1038/s41579-019-0222-5
Clark, D. R., Ferguson, R. M. W., Harris, D. N., Matthews Nicholass, K. J., Prentice, H. J., Randall, K. C., Randell, L., Warren, S. L., & Dumbrell, A. J. (2018). Streams of data from drops of water: 21st century molecular microbial ecology. *Wiley Interdisciplinary Reviews: Water*, *5*(4), e1280. https://doi.org/10.1002/wat2.1280
Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell*, *148*(6), 1258–1270. https://doi.org/10.1016/j.cell.2012.01.035
Cooper, E. J. (2014). Warmer Shorter Winters Disrupt Arctic Terrestrial Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 271–295. https://doi.org/10.1146/annurev-ecolsys-120213-091620
Cuvelier, M. L., Allen, A. E., Monier, A., McCrow, J. P., Messie, M., Tringe, S. G., Woyke, T., Welsh, R. M., Ishoey, T., Lee, J.-H., Binder, B. J., DuPont, C. L., Latasa, M., Guigand, C., Buck, K. R., Hilton, J., Thiagarajan, M., Caler, E., Read, B., … Worden, A. Z. (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences*, *107*(33), 14679–14684. https://doi.org/10.1073/pnas.1001665107
Davey, H. M. (2011). Life, Death, and In-Between: Meanings and Methods in Microbiology. *Applied and Environmental Microbiology*, *77*(16), 5571–5576. https://doi.org/10.1128/AEM.00744-11
Davey, M., Blaalid, R., Vik, U., Carlsen, T., Kauserud, H., & Eidesen, P. B. (2015). Primary succession of Bistorta vivipara (L.) Delabre (Polygonaceae) root-associated fungi mirrors plant succession in two glacial chronosequences: Fungi and plants share successional trajectories. *Environmental Microbiology*, *17*(8), 2777–2790. https://doi.org/10.1111/1462-2920.12770
de Lorenzo, V. (2017). Seven microbial bio-processes to help the planet. *Microbial Biotechnology*, *10*(5), 995–998. https://doi.org/10.1111/1751-7915.12816
Dong, K., Tripathi, B., Moroenyane, I., Kim, W., Li, N., Chu, H., & Adams, J. (2016). Soil fungal community development in a high Arctic glacier foreland follows a directional replacement model, with a mid-successional diversity maximum. *Scientific Reports*, *6*(1), 26360. https://doi.org/10.1038/srep26360
Edwards, A., Cameron, K. A., Cook, J. M., Debbonaire, A. R., Furness, E., Hay, M. C., &

Rassner, S. M. E. (2020). Microbial genomics amidst the Arctic crisis. *Microbial Genomics*, *6*(5). https://doi.org/10.1099/mgen.0.000375

Emerson, J. B., Adams, R. I., Román, C. M. B., Brooks, B., Coil, D. A., Dahlhausen, K., Ganz, H. H., Hartmann, E. M., Hsu, T., Justice, N. B., Paulino-Lima, I. G., Luongo, J. C., Lymperopoulou, D. S., Gomez-Silvan, C., Rothschild-Mancinelli, B., Balk, M., Huttenhower, C., Nocker, A., Vaishampayan, P., & Rothschild, L. J. (2017). Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*, *5*(1), 86. https://doi.org/10.1186/s40168-017-0285-3

Escalas, A., Hale, L., Voordeckers, J. W., Yang, Y., Firestone, M. K., Alvarez-Cohen, L., & Zhou, J. (2019). Microbial functional diversity: From concepts to applications. *Ecology and Evolution*, *9*(20), 12000–12016. https://doi.org/10.1002/ece3.5670

Fan, J., Yang, X., Wang, W., Wood, W. H., Becker, K. G., & Gorospe, M. (2002). Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proceedings of the National Academy of Sciences*, *99*(16), 10611–10616. https://doi.org/10.1073/pnas.162212399

Farnsworth, K. D., Albantakis, L., & Caruso, T. (2017). Unifying concepts of biological function from molecules to ecosystems. *Oikos*, *126*(10), 1367–1376. https://doi.org/10.1111/oik.04171

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Geml, J., Timling, I., Robinson, C. H., Lennon, N., Nusbaum, H. C., Brochmann, C., Noordeloos, M. E., & Taylor, D. L. (2011). An arctic community of symbiotic fungi assembled by long-distance dispersers: Phylogenetic diversity of ectomycorrhizal basidiomycetes in Svalbard based on soil and sporocarp DNA: Biodiversity of arctic ectomycorrhizal fungi. *Journal of Biogeography*, *39*(1), 74–88. https://doi.org/10.1111/j.1365-2699.2011.02588.x

Gillings, M. R., & Paulsen, I. T. (2014). Microbiology of the Anthropocene. *Anthropocene*, *5*, 1–8. https://doi.org/10.1016/j.ancene.2014.06.004

Graham, E. B., Knelman, J. E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J. M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J. C., Glanville, H. C., Jones, D. L., Angel, R., Salminen, J., Newton, R. J., Bürgmann, H., Ingram, L. J., … Nemergut, D. R. (2016). Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00214

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., … Wincker, P. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, *177*(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews: MMBR*, *68*(4), 669–685. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Hillebrand, H. (2004). On the Generality of the Latitudinal Diversity Gradient. *The American*

Naturalist, *163*(2), 192–211. https://doi.org/10.1086/381004
Hobbie, J. E., & Hobbie, E. A. (2006). 15N in Symbiotic Fungi and Plants Estimates Nitrogen and Carbon Flux Rates in Arctic Tundra. *Ecology*, *87*(4), 816–822. https://doi.org/10.1890/0012-9658(2006)87[816:NISFAP]2.0.CO;2
Hodal, H., & Kristiansen, S. (2008). The importance of small-celled phytoplankton in spring blooms at the marginal ice zone in the northern Barents Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(20), 2176–2185. https://doi.org/10.1016/j.dsr2.2008.05.012
Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J., & Wardle, D. A. (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, *75*(1), 3–35. https://doi.org/10.1890/04-0922
Inkpen, S. A., Douglas, G. M., Brunet, T. D. P., Leuschen, K., Doolittle, W. F., & Langille, M. G. I. (2017). The coupling of taxonomy and function in microbiomes. *Biology & Philosophy*, *32*(6), 1225–1243. https://doi.org/10.1007/s10539-017-9602-2
Jax, K. (2005). Function and "functioning" in ecology: What does it mean? *Oikos*, *111*(3), 641–648. https://doi.org/10.1111/j.1600-0706.2005.13851.x
Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27
Kanehisa, Minoru, Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070
Keeling, P. J., & Burki, F. (2019). Progress towards the Tree of Eukaryotes. *Current Biology*, *29*(16), R808–R817. https://doi.org/10.1016/j.cub.2019.07.031
Keeling, P. J., & Campo, J. del. (2017). Marine Protists Are Not Just Big Bacteria. *Current Biology*, *27*(11), R541–R549. https://doi.org/10.1016/j.cub.2017.03.075
Konopka, A. (2009). What is microbial community ecology? *The ISME Journal*, *3*(11), 1223–1230. https://doi.org/10.1038/ismej.2009.88
Konopka, A., Lindemann, S., & Fredrickson, J. (2015). Dynamics in microbial communities: Unraveling mechanisms to identify principles. *The ISME Journal*, *9*(7), 1488–1495. https://doi.org/10.1038/ismej.2014.251
Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611
Kubiszyn, A. M., Wiktor, J. M., Wiktor, J. M., Griffiths, C., Kristiansen, S., & Gabrielsen, T. M. (2017). The annual planktonic protist community structure in an ice-free high Arctic fjord (Adventfjorden, West Spitsbergen). *Journal of Marine Systems*, *169*, 61–72. https://doi.org/10.1016/j.jmarsys.2017.01.013
Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, *2015*(11), 951–969. https://doi.org/10.1101/pdb.top084970
Ladau, J., & Eloe-Fadrosh, E. A. (2019). Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends in Microbiology*, *27*(8), 662–669. https://doi.org/10.1016/j.tim.2019.03.003
Laforest-Lapointe, I., & Arrieta, M.-C. (2018). Microbial Eukaryotes: A Missing Link in Gut Microbiome Studies. *MSystems*, *3*(2), e00201-17, /msystems/3/2/msys.00201-17.atom. https://doi.org/10.1128/mSystems.00201-17
Landenmark, H. K. E., Forgan, D. H., & Cockell, C. S. (2015). An Estimate of the Total DNA

in the Biosphere. *PLOS Biology*, *13*(6), e1002168. https://doi.org/10.1371/journal.pbio.1002168

Lemos, L. N., Fulthorpe, R. R., Triplett, E. W., & Roesch, L. F. W. (2011). Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods*, *86*(1), 42–51. https://doi.org/10.1016/j.mimet.2011.03.014

Li, W. K. W., McLaughlin, F. A., Lovejoy, C., & Carmack, E. C. (2009). Smallest algae thrive as the Arctic Ocean freshens. *Science (New York, N.Y.)*, *326*(5952), 539. https://doi.org/10.1126/science.1179798

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., & Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers—A user's guide. *New Phytologist*, *199*(1), 288–299. https://doi.org/10.1111/nph.12243

Little, A. E. F., Robinson, C. J., Peterson, S. B., Raffa, K. F., & Handelsman, J. (2008). Rules of Engagement: Interspecies Interactions that Regulate Microbial Communities. *Annual Review of Microbiology*, *62*(1), 375–401. https://doi.org/10.1146/annurev.micro.030608.101423

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W. H. C. F., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M. C., Romac, S., … Massana, R. (2014). Patterns of Rare and Abundant Marine Microbial Eukaryotes. *Current Biology*, *24*(8), 813–821. https://doi.org/10.1016/j.cub.2014.02.050

Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, *13*(12), 787–794. https://doi.org/10.1038/nrmicro3565

López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., & Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, *409*(6820), 603–607. https://doi.org/10.1038/35054537

Lorberau, K. E., Botnen, S. S., Mundra, S., Aas, A. B., Rozema, J., Eidesen, P. B., & Kauserud, H. (2017). Does warming by open-top chambers induce change in the root-associated fungal community of the arctic dwarf shrub Cassiope tetragona (Ericaceae)? *Mycorrhiza*, *27*(5), 513–524. https://doi.org/10.1007/s00572-017-0767-y

Loreau, M. (2001). Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science*, *294*(5543), 804–808. https://doi.org/10.1126/science.1064088

Lynch, M. (2006). The Origins of Eukaryotic Gene Structure. *Molecular Biology and Evolution*, *23*(2), 450–468. https://doi.org/10.1093/molbev/msj050

Lynch, M., & Conery, J. S. (2003). The Origins of Genome Complexity. *Science*, *302*(5649), 1401–1404. https://doi.org/10.1126/science.1089370

Lynch, M., & Marinov, G. K. (2017). Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *ELife*, *6*, e20437. https://doi.org/10.7554/eLife.20437

Mackelprang, R., Saleska, S. R., Jacobsen, C. S., Jansson, J. K., & Taş, N. (2016). Permafrost Meta-Omics and Climate Change. *Annual Review of Earth and Planetary Sciences*, *44*(1), 439–462. https://doi.org/10.1146/annurev-earth-060614-105126

Madigan, M. T., Martinko, J. M., Bender, K. S., Buckley, D. H., & Stahl, D. A. (2014). *Brock biology of microorganisms.*

Malik, A., Grohmann, E., & Alves, M. (Eds.). (2013). *Management of Microbial Resources in the Environment*. Springer Netherlands. https://doi.org/10.1007/978-94-007-5931-2

Maron, P.-A., Ranjard, L., Mougel, C., & Lemanceau, P. (2007). Metaproteomics: A New Approach for Studying Functional Microbial Ecology. *Microbial Ecology*, *53*(3), 486–493. https://doi.org/10.1007/s00248-006-9196-8

Marquardt, M., Skogseth, R., Wiedmann, I., Vader, A., Reigstad, M., Cottier, F., & Gabrielsen, T. (2019). Vertical export of marine pelagic protists in an ice-free high-Arctic fjord (Adventfjorden, West Spitsbergen) throughout 2011-2012. *Aquatic Microbial Ecology*, *83*(1), 65–82. https://doi.org/10.3354/ame01904

Marquardt, Miriam, Vader, A., Stübner, E. I., Reigstad, M., & Gabrielsen, T. M. (2016). Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). *Applied and Environmental Microbiology*, *82*(6), 1868–1880. https://doi.org/10.1128/AEM.03208-15

McCarthy, J. E. (1998). Posttranscriptional control of gene expression in yeast. *Microbiology and Molecular Biology Reviews: MMBR*, *62*(4), 1492–1553.

Meshram, A. R., Vader, A., Kristiansen, S., & Gabrielsen, T. M. (2017). Microbial Eukaryotes in an Arctic Under-Ice Spring Bloom North of Svalbard. *Frontiers in Microbiology*, *8*, 1099. https://doi.org/10.3389/fmicb.2017.01099

Mitra, A., Flynn, K. J., Tillmann, U., Raven, J. A., Caron, D., Stoecker, D. K., Not, F., Hansen, P. J., Hallegraeff, G., Sanders, R., Wilken, S., McManus, G., Johnson, M., Pitta, P., Våge, S., Berge, T., Calbet, A., Thingstad, F., Jeong, H. J., … Lundgren, V. (2016). Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist*, *167*(2), 106–120. https://doi.org/10.1016/j.protis.2016.01.003

Mörsdorf, M. A., Baggesen, N. S., Yoccoz, N. G., Michelsen, A., Elberling, B., Ambus, P. L., & Cooper, E. J. (2019). Deepened winter snow significantly influences the availability and forms of nitrogen taken up by plants in High Arctic tundra. *Soil Biology and Biochemistry*, *135*, 222–234. https://doi.org/10.1016/j.soilbio.2019.05.009

Mundra, S., Bahram, M., & Eidesen, P. B. (2016). Alpine bistort (Bistorta vivipara) in edge habitat associates with fewer but distinct ectomycorrhizal fungal species: A comparative study of three contrasting soil environments in Svalbard. *Mycorrhiza*, *26*(8), 809–818. https://doi.org/10.1007/s00572-016-0716-1

Mundra, S., Bahram, M., Tedersoo, L., Kauserud, H., Halvorsen, R., & Eidesen, P. B. (2015). Temporal variation of Bistorta vivipara-associated ectomycorrhizal fungal communities in the High Arctic. *Molecular Ecology*, *24*(24), 6289–6302. https://doi.org/10.1111/mec.13458

Mundra, S., Halvorsen, R., Kauserud, H., Bahram, M., Tedersoo, L., Elberling, B., Cooper, E. J., & Eidesen, P. B. (2016). Ectomycorrhizal and saprotrophic fungi respond differently to long-term experimentally increased snow depth in the High Arctic. *MicrobiologyOpen*, *5*(5), 856–869. https://doi.org/10.1002/mbo3.375

Mundra, S., Halvorsen, R., Kauserud, H., Müller, E., Vik, U., & Eidesen, P. B. (2015). Arctic fungal communities associated with roots of Bistorta vivipara do not respond to the same fine-scale edaphic gradients as the aboveground vegetation. *New Phytologist*, *205*(4), 1587–1597. https://doi.org/10.1111/nph.13216

NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *44*(D1), D7–D19. https://doi.org/10.1093/nar/gkv1290

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., Fierer, N., Townsend, A. R., Cleveland, C. C., Stanish, L., & Knight, R. (2011). Global patterns in the biogeography of bacterial taxa: Global bacterial biogeography. *Environmental Microbiology*, *13*(1), 135–144. https://doi.org/10.1111/j.1462-2920.2010.02315.x

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., &

Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Nicolás, C., Martin-Bertelsen, T., Floudas, D., Bentzer, J., Smits, M., Johansson, T., Troein, C., Persson, P., & Tunlid, A. (2019). The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen. *The ISME Journal*, *13*(4), 977–988. https://doi.org/10.1038/s41396-018-0331-6

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, *47*(D1), D259–D264. https://doi.org/10.1093/nar/gky1022

Nordli, Ø., Wyszyński, P., Gjelten, H. M., Isaksen, K., Łupikasza, E., Niedźwiedź, T., & Przybylak, R. (2020). Revisiting the extended Svalbard Airport monthly temperature series, and the compiled corresponding daily series 1898–2018. *Polar Research*. https://doi.org/10.33265/polar.v39.3614

Nunes-Neto, N., Moreno, A., & El-Hani, C. N. (2014). Function in ecology: An organizational approach. *Biology & Philosophy*, *29*(1), 123–141. https://doi.org/10.1007/s10539-013-9398-7

Oberwinkler, F., Riess, K., Bauer, R., Selosse, M.-A., Weiß, M., Garnica, S., & Zuccaro, A. (2013). Enigmatic Sebacinales. *Mycological Progress*, *12*(1), 1–27. https://doi.org/10.1007/s11557-012-0880-4

Oliver, S. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, *16*(9), 373–378. https://doi.org/10.1016/S0167-7799(98)01214-1

Oliverio, A. M., Power, J. F., Washburne, A., Cary, S. C., Stott, M. B., & Fierer, N. (2018). The ecology and diversity of microbial eukaryotes in geothermal springs. *The ISME Journal*, *12*(8), 1918–1928. https://doi.org/10.1038/s41396-018-0104-2

Onda, D. F. L., Medrinal, E., Comeau, A. M., Thaler, M., Babin, M., & Lovejoy, C. (2017). Seasonal and Interannual Changes in Ciliate and Dinoflagellate Species Assemblages in the Arctic Ocean (Amundsen Gulf, Beaufort Sea, Canada). *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00016

Orosz, E., van de Wiele, N., Emri, T., Zhou, M., Robert, V., de Vries, R. P., & Pócsi, I. (2018). Fungal Stress Database (FSD)—a repository of fungal stress physiological data. *Database: The Journal of Biological Databases and Curation*, *2018*. https://doi.org/10.1093/database/bay009

Patterson, D. J. (1999). The Diversity of Eukaryotes. *The American Naturalist*, *154*(S4), S96–S124. https://doi.org/10.1086/303287

Peimbert, M., & Alcaraz, L. (2016). *A Hitchhiker's Guide to Metatranscriptomics* (pp. 313–342). https://doi.org/10.1007/978-3-319-31350-4_13

Peisl, B. Y. L., Schymanski, E. L., & Wilmes, P. (2018). Dark matter in host-microbiome metabolomics: Tackling the unknowns–A review. *Analytica Chimica Acta*, *1037*, 13–27. https://doi.org/10.1016/j.aca.2017.12.034

Poorter, H., Niklas, K. J., Reich, P. B., Oleksyn, J., Poot, P., & Mommer, L. (2012). Biomass allocation to leaves, stems and roots: Meta-analyses of interspecific variation and environmental control: Tansley review. *New Phytologist*, *193*(1), 30–50. https://doi.org/10.1111/j.1469-8137.2011.03952.x

Postgate, J. R. (1969). Chapter XVIII Viable counts and Viability. In J. R. Norris & D. W. Ribbons (Eds.), *Methods in Microbiology* (Vol. 1, pp. 611–628). Academic Press.

https://doi.org/10.1016/S0580-9517(08)70149-1
Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L. J., von Mering, C., & Bork, P. (2012). eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, *40*(D1), D284–D289. https://doi.org/10.1093/nar/gkr1060
Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, A. M., Solan, M., van der Gast, C. J., & Young, J. P. W. (2007). The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, *5*(5), 384–392. https://doi.org/10.1038/nrmicro1643
Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188–7196. https://doi.org/10.1093/nar/gkm864
R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Raja, H. A., Miller, A. N., Pearce, C. J., & Oberlies, N. H. (2017). Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *Journal of Natural Products*, *80*(3), 756–770. https://doi.org/10.1021/acs.jnatprod.6b01085
Robinson, C. H. (2001). Cold adaptation in Arctic and Antarctic fungi. *New Phytologist*, *151*(2), 341–353. https://doi.org/10.1046/j.1469-8137.2001.00177.x
Robinson, C. J., Bohannan, B. J. M., & Young, V. B. (2010). From Structure to Function: The Ecology of Host-Associated Microbial Communities. *Microbiology and Molecular Biology Reviews*, *74*(3), 453–476. https://doi.org/10.1128/MMBR.00014-10
Rokkan Iversen, K., & Seuthe, L. (2011). Seasonal microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): I. Heterotrophic bacteria, picoplankton and nanoflagellates. *Polar Biology*, *34*(5), 731–749. https://doi.org/10.1007/s00300-010-0929-2
Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome Research*, *18*(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761
Santoferrara, L., Burki, F., Filker, S., Logares, R., Dunthorn, M., & McManus, G. B. (2020). Perspectives from Ten Years of Protist Studies by High-Throughput Metabarcoding. *Journal of Eukaryotic Microbiology*, jeu.12813. https://doi.org/10.1111/jeu.12813
Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List, Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K.-D., Bai, F.-Y., Barreto, R. W., Begerow, D., … Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241–6246. https://doi.org/10.1073/pnas.1117018109
Schostag, M. D., Albers, C. N., Jacobsen, C. S., & Priemé, A. (2020). Low Turnover of Soil Bacterial rRNA at Low Temperatures. *Frontiers in Microbiology*, *11*, 962. https://doi.org/10.3389/fmicb.2020.00962
Schostag, M., Stibal, M., Jacobsen, C. S., Bælum, J., Taş, N., Elberling, B., Jansson, J. K.,

Semenchuk, P., & Priemé, A. (2015). Distinct summer and winter bacterial communities in the active layer of Svalbard permafrost revealed by DNA- and RNA-based analyses. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00399

Schuler, T. V., & Østby, T. I. (2020). Sval_Imp: A gridded forcing dataset for climate change impact research on Svalbard. *Earth System Science Data*, *12*(2), 875–885. https://doi.org/10.5194/essd-12-875-2020

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., & Vonk, J. E. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. https://doi.org/10.1038/nature14338

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Bürgmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B. H., Matulich, K. L., Schmidt, T. M., & Handelsman, J. (2012). Fundamentals of Microbial Community Resistance and Resilience. *Frontiers in Microbiology*, *3*. https://doi.org/10.3389/fmicb.2012.00417

Shah, F., Nicolás, C., Bentzer, J., Ellström, M., Smits, M., Rineau, F., Canbäck, B., Floudas, D., Carleer, R., Lackner, G., Braesel, J., Hoffmeister, D., Henrissat, B., Ahrén, D., Johansson, T., Hibbett, D. S., Martin, F., Persson, P., & Tunlid, A. (2016). Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytologist*, *209*(4), 1705–1719. https://doi.org/10.1111/nph.13722

Shakya, M., Lo, C.-C., & Chain, P. S. G. (2019). Advances and Challenges in Metatranscriptomic Analysis. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00904

Slaveykova, V., Sonntag, B., & Gutiérrez, J. C. (2016). Stress and Protists: No life without stress. *European Journal of Protistology*, *55*, 39–49. https://doi.org/10.1016/j.ejop.2016.06.001

Sørensen, N., Daugbjerg, N., & Gabrielsen, T. M. (2012). Molecular diversity and temporal variation of picoeukaryotes in two Arctic fjords, Svalbard. *Polar Biology*, *35*(4), 519–533. https://doi.org/10.1007/s00300-011-1097-8

Stoecker, D. K., & Lavrentyev, P. J. (2018). Mixotrophic Plankton in the Polar Seas: A Pan-Arctic Review. *Frontiers in Marine Science*, *5*, 292. https://doi.org/10.3389/fmars.2018.00292

Su, C., Lei, L., Duan, Y., Zhang, K.-Q., & Yang, J. (2012). Culture-independent methods for studying environmental microorganisms: Methods, application, and perspective. *Applied Microbiology and Biotechnology*, *93*(3), 993–1003. https://doi.org/10.1007/s00253-011-3800-7

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., & Zimov, S. (2009). Soil organic carbon pools in the northern circumpolar permafrost region: SOIL ORGANIC CARBON POOLS. *Global Biogeochemical Cycles*, *23*(2), n/a-n/a. https://doi.org/10.1029/2008GB003327

Taylor, J. W., Turner, E., Townsend, J. P., Dettman, J. R., & Jacobson, D. (2006). Eukaryotic microbes, species recognition and the geographic limits of species: Examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), 1947–1963. https://doi.org/10.1098/rstb.2006.1923

Tecon, R., Mitri, S., Ciccarese, D., Or, D., van der Meer, J. R., & Johnson, D. R. (2019). Bridging the Holistic-Reductionist Divide in Microbial Ecology. *MSystems*, *4*(1),

e00265-18, /msystems/4/1/msys.00265-18.atom. https://doi.org/10.1128/mSystems.00265-18

Thoen, E., Aas, A. B., Vik, U., Brysting, A. K., Skrede, I., Carlsen, T., & Kauserud, H. (2019). A single ectomycorrhizal plant root system includes a diverse and spatially structured fungal community. *Mycorrhiza*, *29*(3), 167–180. https://doi.org/10.1007/s00572-019-00889-z

Timmermans, M., & Marshall, J. (2020). Understanding Arctic Ocean Circulation: A Review of Ocean Dynamics in a Changing Climate. *Journal of Geophysical Research: Oceans*, *125*(4). https://doi.org/10.1029/2018JC014378

The UniProt Consortium. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

UNITE Community. (2019). *Full UNITE+INSD dataset for Fungi*. https://doi.org/10.15156/BIO/786347

Vader, A., Laughinghouse, H. D., Griffiths, C., Jakobsen, K. S., & Gabrielsen, T. M. (2018). Proton-pumping rhodopsins are abundantly expressed by microbial eukaryotes in a high-Arctic fjord. *Environmental Microbiology*, *20*(2), 890–902. https://doi.org/10.1111/1462-2920.14035

Vader, A., Marquardt, M., Meshram, A. R., & Gabrielsen, T. M. (2014). Key Arctic phototrophs are widespread in the polar night. *Polar Biology*, *38*(1), 13–21. https://doi.org/10.1007/s00300-014-1570-2

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ*, k2179. https://doi.org/10.1136/bmj.k2179

van der Linde, S., Holden, E., Parkin, P. I., Alexander, I. J., & Anderson, I. C. (2012). Now you see it, now you don't: The challenge of detecting, monitoring and conserving ectomycorrhizal fungi. *Fungal Ecology*, *5*(5), 633–640. https://doi.org/10.1016/j.funeco.2012.04.002

Vandenkoornhuyse, P., Dufresne, A., Quaiser, A., Gouesbet, G., Binet, F., Francez, A.-J., Mahé, S., Bormans, M., Lagadeuc, Y., & Couée, I. (2010). Integration of molecular functions at the ecosystemic level: Breakthroughs and future goals of environmental genomics and post-genomics: Environmental genomics. *Ecology Letters*, *13*(6), 776–791. https://doi.org/10.1111/j.1461-0248.2010.01464.x

Vaqué, D., Guadayol, Ò., Peters, F., Felipe, J., Angel-Ripoll, L., Terrado, R., Lovejoy, C., & Pedrós-Alioó, C. (2008). Seasonal changes in planktonic bacterivory rates under the ice-covered coastal Arctic Ocean. *Limnology and Oceanography*, *53*(6), 2427–2438. https://doi.org/10.4319/lo.2008.53.6.2427

Vincent, W. F. (2000). Cyanobacterial Dominance in the Polar Regions. In B. A. Whitton & M. Potts (Eds.), *The Ecology of Cyanobacteria: Their Diversity in Time and Space* (pp. 321–340). Springer Netherlands. https://doi.org/10.1007/0-306-46855-7_12

Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., & Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences*, *99*(9), 5860–5865. https://doi.org/10.1073/pnas.092538799

Wang, Yuqiu, Zhou, Y., Xiao, X., Zheng, J., & Zhou, H. (2020). Metaproteomics: A strategy to study the taxonomy and functionality of the gut microbiota. *Journal of Proteomics*, *219*, 103737. https://doi.org/10.1016/j.jprot.2020.103737

Wassmann, P., Duarte, C. M., Agustí, S., & Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem: FOOTPRINTS OF CLIMATE CHANGE. *Global Change Biology*, *17*(2), 1235–1249. https://doi.org/10.1111/j.1365-2486.2010.02311.x

Webb, T. J. (2012). Marine and terrestrial ecology: Unifying concepts, revealing differences. *Trends in Ecology & Evolution*, *27*(10), 535–541. https://doi.org/10.1016/j.tree.2012.06.002

Weiss, M., Selosse, M.-A., Rexer, K.-H., Urban, A., & Oberwinkler, F. (2004). Sebacinales: A hitherto overlooked cosm of heterobasidiomycetes with a broad mycorrhizal potential* *Part 221 of the series Studies in Heterobasidiomycetes from the Botanical Institute, University of Tübingen, Tübingen. *Mycological Research*, *108*(9), 1003–1010. https://doi.org/10.1017/S0953756204000772

Weiß, M., Waller, F., Zuccaro, A., & Selosse, M.-A. (2016). Sebacinales—One thousand and one interactions with land plants. *The New Phytologist*, *211*(1), 20–40. https://doi.org/10.1111/nph.13977

Welch, J. L. M., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., & Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Sciences*, *113*(6), E791–E800. https://doi.org/10.1073/pnas.1522149113

Westreich, S. T., Korf, I., Mills, D. A., & Lemay, D. G. (2016). SAMSA: A comprehensive metatranscriptome analysis pipeline. *BMC Bioinformatics*, *17*(1), 399. https://doi.org/10.1186/s12859-016-1270-8

Wiedmann, I., Reigstad, M., Marquardt, M., Vader, A., & Gabrielsen, T. M. (2016). Seasonality of vertical flux and sinking particle characteristics in an ice-free high arctic fjord—Different from subarctic fjords? *Journal of Marine Systems*, *154*, 192–205. https://doi.org/10.1016/j.jmarsys.2015.10.003

Wilkinson, D. M., Koumoutsaris, S., Mitchell, E. A. D., & Bey, I. (2012). Modelling the effect of size on the aerial dispersal of microorganisms. *Journal of Biogeography*, *39*(1), 89–97. https://doi.org/10.1111/j.1365-2699.2011.02569.x

Winfree, R. (2020). How Does Biodiversity Relate to Ecosystem Functioning in Natural Ecosystems? In A. Dobson, D. Tilman, & R. D. Holt (Eds.), *Unsolved Problems in Ecology* (p. 338).

Zhao, Y., Yi, Z., Warren, A., & Song, W. B. (2018). Species delimitation for the molecular taxonomy and ecology of the widely distributed microbial eukaryote genus Euplotes (Alveolata, Ciliophora). *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172159. https://doi.org/10.1098/rspb.2017.2159

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), 1857–1862. https://doi.org/10.1111/mec.15060

PAPER I

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 January 2019
doi: 10.3389/fmicb.2018.03243

Dead or Alive; or Does It Really Matter? Level of Congruency Between Trophic Modes in Total and Active Fungal Communities in High Arctic Soil

Magdalena Wutkowska[1,2], Anna Vader[1], Sunil Mundra[3], Elisabeth J. Cooper[2] and Pernille B. Eidesen[1]*

[1] Department of Arctic Biology, The University Centre in Svalbard (UNIS), Longyearbyen, Norway, [2] Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT - The Arctic University of Norway, Tromsø, Norway, [3] Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, Oslo, Norway

OPEN ACCESS

Edited by:
Samuel Cirés, Universidad Autónoma de Madrid, Spain

Reviewed by:
Clare Helen Robinson, University of Manchester, United Kingdom
Minna Männistö, Natural Resources Institute Finland (Luke), Finland

***Correspondence:**
Magdalena Wutkowska
magdalena.wutkowska@unis.no;
magda.wutkowska@gmail.com

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *16 October 2018*
Accepted: *13 December 2018*
Published: *08 January 2019*

Citation:
Wutkowska M, Vader A, Mundra S, Cooper EJ and Eidesen PB (2019) Dead or Alive; or Does It Really Matter? Level of Congruency Between Trophic Modes in Total and Active Fungal Communities in High Arctic Soil. Front. Microbiol. 9:3243. doi: 10.3389/fmicb.2018.03243

Describing dynamics of belowground organisms, such as fungi, can be challenging. Results of studies based on environmental DNA (eDNA) may be biased as the template does not discriminate between metabolically active cells and dead biomass. We analyzed ribosomal DNA (rDNA) and ribosomal RNA (rRNA) coextracted from 48 soil samples collected from a manipulated snow depth experiment in two distinct vegetation types in Svalbard, in the High Arctic. Our main goal was to compare if the rDNA and rRNA metabarcoding templates produced congruent results that would lead to consistent ecological interpretation. Data derived from both rDNA and rRNA clustered according to vegetation types. Different sets of environmental variables explained the community composition based on the metabarcoding template. rDNA and rRNA-derived community composition of symbiotrophs and saprotrophs, unlike pathotrophs, clustered together in a similar way as when the community composition was analyzed using all OTUs in the study. Mean OTU richness was higher for rRNA, especially in symbiotrophs. The metabarcoding template was more important than vegetation type in explaining differences in richness. The proportion of symbiotrophic, saprotrophic and functionally unassigned reads differed between rDNA and rRNA, but showed similar trends. There was no evidence for increased snow depth influence on fungal community composition or richness. Our findings suggest that template choice may be especially important for estimating biodiversity, such as richness and relative abundances, especially in Helotiales and Agaricales, but not for inferring community composition. Differences in study results originating from rDNA or rRNA may directly impact the ecological conclusions of one's study, which could potentially lead to false conclusions on the dynamics of microbial communities in a rapidly changing Arctic.

Keywords: below-ground processes, fungal trophic mode, fungal functional group, snow regime, arctic vegetation, snow fences

INTRODUCTION

Species loss is a major concern in ecosystem functioning (Cardinale et al., 2012). Amplicon sequencing of DNA extracted from environmental samples has become a common tool for species detection (Bohmann et al., 2014; Bass et al., 2015; Barnes and Turner, 2016). Since only a small fraction of microbes, including fungi, can be cultured in the laboratory, monitoring of these species relies heavily on analysis environmental ribosomal DNA (rDNA) (Creer et al., 2016). Microbes are embedded in multi-species assemblages that closely interact with each other on small spatial scales; genomic methods based on rDNA used to describe their characteristics, such as taxonomic diversity (Konopka, 2009). Despite tremendous advancements in molecular methods, estimating biodiversity and community composition in many groups of organisms, such as fungi, remains challenging (Costello, 2015; Hawksworth and Lücking, 2017).

Critical assessment of results, recommendations and best practices for rDNA metabarcoding is still debated (Goldberg et al., 2016; Shelton et al., 2016). Methodological biases may heavily influence fungal study outcomes; this includes bypassing detection of certain taxonomic groups by choosing particular marker genes (Schoch et al., 2012) or even marker gene regions (Blaalid et al., 2013). In spite of these methodological limitations, a growing body of evidence suggests that the choice between nucleic acid template, namely rDNA, and its transcribed product rRNA, may provide inconsistencies. This is due to the fact that rDNA does not have to correspond with the presence of living cells in the environment (Anderson and Parkin, 2007; Pedersen et al., 2015; Carini et al., 2016). Physicochemical properties of the environment, such as cold temperatures or soil particle adsorption properties, can enhance preservation of DNA from dead organisms (Ogram et al., 1988; Saeki and Kunito, 2010; Saeki et al., 2011). It was recently shown that using rRNA as sequencing template was superior to rDNA in detecting live bacterial cells in water (Li et al., 2017). The turnover rate of DNA is expected to be much slower in soil than in water (Thomsen and Willerslev, 2015). Thus, rDNA metabarcoding of soil samples has a high risk of being biased by dead material.

Risk of bias in biodiversity assessment from dead material is particularly high in samples of soil dwelling organisms from cold climate regions, In the Arctic, lower temperatures slow down the rate of microbial decomposition and cells or extracellular DNA may freeze within permafrost (Gilichinsky et al., 1995; Soina et al., 1995). Old organic material can later intermix through physical processes in the soil, such as cryoturbation, which enables soil from deeper depths to be raised to the top exposing biological material frozen many years ago (Kaiser et al., 2007). To circumvent these problems, an alternative is to use rRNA as a metabarcoding template. RNA degrades rapidly when it is no longer needed in the cell, and therefore gives information about the metabolically active cells that contribute to microbial processes (Blazewicz et al., 2013).

Species can play redundant roles in an ecosystem, therefore recent ecological studies stress targeting functional diversity in ecosystems, as opposed to biodiversity only (Louca et al., 2016; Cernansky, 2017). Many fungal species play redundant roles in ecosystem functioning by exploiting or altering the distribution of the same resources (Moore et al., 2011). In recent years some fungal studies focused on parsing operational taxonomic units (OTUs) into ecologically meaningful groups that play the same function in the ecosystem, such as trophic modes, represented by symbiotrophs, saprotrophs and pathotrophs (Nguyen et al., 2016). All of these trophic modes are important in arctic ecosystems. Saprotrophic fungi acquire their organic carbon through decomposition of dead biomass, and are important for carbon- and nutrient cycling in arctic soils (Buckeridge and Grogan, 2008; Kohler et al., 2015). Symbiotrophic fungi acquire their organic carbon through mutualistic partnerships, especially with plants. This group includes mycorrhizal fungi that play an important ecological role by supporting plant uptake of nutrients and water, notably important in arctic tundra where especially nitrogen may be heavily depleted (Gardes and Dahlberg, 1996; Timling and Taylor, 2012). Pathotrophic fungi that obtain organic carbon by harming host cells play a role in controlling other trophic levels in the ecosystem (Fodor, 2011). Previous studies have suggested that altered climate can change soil carbon balance, affecting vegetation composition through the influence of pathotrophic fungi (Olofsson et al., 2011).

Fungi play important ecological roles in Arctic terrestrial ecosystems and current knowledge on how Arctic fungal biodiversity is shaped by climate changes remains scattered (Timling and Taylor, 2012). Investigating these fungal responses is clearly at risk of being affected by both methodological bias and bias induced by extracellular rDNA, which was estimated to contribute up to 40% of all sequences in soil samples, thus escalating observed richness and misleading conclusions about prokaryotic and fungal communities (Carini et al., 2016). Response of fungal communities to some manifestations of these changes in the Arctic, such as increased winter precipitation, were studied using only rDNA (Morgado et al., 2016; Mundra et al., 2016b; Semenova et al., 2016). Thus, none of these studies discriminated between metabolically active cells, dead matter, spores or relict rDNA.

In this case study we assess how the choice of metabarcoding template (rDNA vs. rRNA) influences the fungal soil community retrieved from soil samples under different environmental conditions. We sampled soil in an experimental setting of snow fences mimicking increased winter precipitation in two distinct vegetation types: heath and meadow (Morgner et al., 2010). Then we sequenced ITS2, analyzing rDNA and rRNA-based metabarcoding data separately. Our main aim was to determine whether results and ecological conclusions based on rDNA and rRNA metabarcoding templates were congruent. We analyzed the data in relation to fungal trophic modes, here defined as symbiotrophs, saprotrophs and pathotrophs (Nguyen et al., 2016). We also compared rDNA and rRNA results in relation to community composition (1) and OTU richness (2). Finally, we looked into how various edaphic variables influenced community composition as well as OTU richness for different fungal trophic modes. Incongruent results between the two metabarcoding templates at any of these levels may potentially point toward types of analyses that can create a misleading picture of the ecosystem.

MATERIALS AND METHODS

Sampling Site, Experimental Setup, and Sample Collection

Snow fences established in Adventdalen, Svalbard (78°10 N, 16°02-16°05 E), altered snow regime since winter 2006/2007, creating approximately 1 m deeper snow in treatment plots compared to controls (Morgner et al., 2010; **Supplements 1** and **1a**). Deep snow regime altered annual patterns of two important physical variables for soil dwelling microorganisms: soil moisture content and temperature (Cooper et al., 2011). Fences were established in blocks of 3 fences with 2 blocks per vegetation type: heath and meadow. Deep snow regime generally had higher soil nutrients (NO_3^-N, NH_4^+N, and K) than ambient (Semenchuk et al., 2015; Mundra et al., 2016b).

Sampling took place on 28 and 30 of August 2012, simultaneously with a study focusing on Bistorta vivipara root associated communities from the same sites (Eidesen, unpublished data). After an individual B. vivipara plant with its whole root system had been excavated using a small shovel, two soil samples were collected, filling 2 ml cryo-tubes, from opposite sides of the resulting hole. The soil samples, procured from 5 to 10 cm depth with a sterile spatula, were immediately frozen in liquid nitrogen. In total 96 samples were collected; 2 holes × 2 soil samples × 2 snow regimes × 3 fences × 2 blocks × 2 vegetation types. Edaphic parameters were measured according to protocols described in Mundra et al. (2015b). To minimize the effect of small-scale spatial variability the two primary samples from the same hole were combined prior to analyses, resulting in 48 true samples (referred to in the remaining text).

Obtaining rDNA and rRNA Sequences

rRNA and rDNA was co-extracted from 1 to 2 g of frozen soil using the PowerSoil Total RNA Isolation Kit (MO BIO Laboratories, United States) and PowerSoil DNA Elution Accessory Kit (MO BIO Laboratories, United States), both according to manufacturer's instructions. Complementary DNA (cDNA) was synthesized using the Maxima First Strand cDNA Synthesis Kit for RT-qPCR, with dsDNase (Thermo Fisher Scientific, United States) following the manufacturers' instructions, except that a 5 min incubation step was used for DNase treatment. After DNase treatment, a 1 μl subsample was used as a no-RT control during subsequent PCR amplification. All no-RT controls were negative, showing that DNase treatment had been successful and that cDNA amplification during RT-PCR was due to rRNA template.

PCRs and library preparations was carried out for rDNA and cDNA as described in Mundra et al. (2016b), using the primers fITS7a (Ihrmark et al., 2012) and ITS4 (White et al., 1990) to amplify the internal transcribed spacer 2 (ITS2) region of the nuclear ribosomal DNA, using 1 μl of DNA/cDNA as templates. The protocol for library preparation is described in Mundra et al. (2016b). The multiplexed samples were paired-end sequenced (2 × 300 bp) on an Illumina MiSeq sequencer at ACGT Inc, Wheeling, United States.

Bioinformatic Analysis of Sequencing Data

The bioinformatic analysis of Illumina sequences followed the pipeline described by Bálint et al. (2014) with minor modifications. A total of 8,413,098 paired-end sequenced reads were filtered using a perl script (supplemented in Bálint et al., 2014). The remaining 7,779,879 high quality paired-end sequenced reads (high quality score > 26) were assembled in PANDAseq 2.6 (Masella et al., 2012). After quality filtering and assembly, 23 rDNA and 19 rRNA samples were retained in the analyses. Sequences with primer artifacts were removed with a python script (supplemented in Bálint et al., 2014), prior to reorientation using fqgrep 0.4.4[1] and the fastx_reverse_compliment function from Fastx Toolkit 0.0.14[2] to reverse sequences identified as oriented in the 3′–5′ direction containing 7,028,992 reads. To demultiplex sequences with variable length barcodes we used the split_library.py script in Qiime 1.9.1 (Caporaso et al., 2010), retaining sequences of 200–500 bp, allowing for 1bp primer mismatch, and maximum length of homopolymer run equal to 8.

5,184,214 demultiplexed sequences were then sorted by length in a range and dereplicated, before sorting groups by size, excluding those containing less than five sequences (Nguyen et al., 2015) in Vsearch 2.7.1 (Rognes et al., 2016). Using 0.97 sequence similarity threshold, 2185 operational taxonomic units (OTUs) were picked by the cluster_otus function (usearch 8.1.1861; Edgar, 2010) and then 232 putative chimera sequences were removed in reference based chimera check with uchime2 (Edgar, 2016) against fungal database UNITE+INSD (Kõljalg et al., 2013; version: UNITE_public_01.12.2017). To retain only ITS2 fragments of fungal origin, sequences were filtered through ITSx v. 1.1b (Bengtsson-Palme et al., 2013), leaving 1473 representative sequences. To further exclude non-fungal sequences we used local blast search (blast+ 2.6.0) against the nucleotide NCBI database (updated 13.12.2017) and parsed these results in MEGAN Community Edition 6.5.10 (Huson et al., 2016) as described in Bálint et al. (2014). Unclustered sequences were mapped against representative sequences identified as fungal in MEGAN to obtain an OTU abundance table, which then was rarefied to 42,488 reads per sample with single_rarefaction.py in Qiime 1.9.1 (Caporaso et al., 2010). The level of rarefaction was set based on output from the demultiplexing step. Several samples with very low read numbers (0–2870 reads), were removed during this step, based on the assumption that these samples had failed during the sequencing reaction. The distribution of failed samples was random, and although leading to a lower number of total samples in the study and hence lower statistical power, should not affect the conclusions of our study.

The final OTU table with rDNA and rRNA samples contained 837 OTUs. Since correct identification of species determines more precise functional assignment, the final taxonomy was assigned by querying representative sequences against the curated fungal database UNITE. In cases where we did not get a blast hit, taxonomy was assigned using the NCBI-NT database. Eight

[1]https://github.com/indraniel/fqgrep

[2]http://hannonlab.cshl.edu/fastx_toolkit/

OTUs were assigned as Rhizaria (all as unidentified class of Cercozoa). We decided to keep these due to the fact that they remained in the dataset through two steps of removing non-fungal OTUs (see above). OTUs were categorized into trophic modes: symbiotrophs, saprotrophs and pathotrophs (**Supplement 1**) using FUNGuild (Nguyen et al., 2016). OTUs assigned to multiple trophic modes, as well as OTUs with taxonomic assignment that precluded accurate assignment to a trophic mode, were marked as "unassigned." The OTU table was divided into separate matrices for rDNA and rRNA, which were analyzed separately for the rest of the study.

Statistical Analysis

Statistical analysis was performed in R v3.4.4 (R Core Team, 2018). All described statistical analyses were performed in parallel for both rDNA and rRNA.

Community Composition

Global non-metric multidimensional scaling (GNMDS; Kruskal, 1964) was used to analyze dissimilarity matrixes within rDNA- and rRNA-based community compositions of the samples containing all OTUs, symbiotrophs, saprotrophs and pathotrophs separately, on presence-absence OTU tables using the Jaccard dissimilarity index. In ordination analyses we used presence-absence data to avoid biases associated with possible differences in RNA copy number. The ordination analyses were performed following Liu et al. (2008). Loss of data during sample preparation and data processing allowed a direct comparison of only nine extracted pairs of rDNA and rRNA samples, which was tested by Mantel's test with 9999 replications (ade4 package 1.7-11, Chessel et al., 2004). Possible relationships among community composition, edaphic variables and experimental factors were investigated. The envfit function in vegan package (v. 2.5-2; Oksanen et al., 2018) was used for multiple regressions of edaphic variables and vegetation type. Permutational multivariate analysis of variance (PERMANOVA) implemented as adonis function in vegan package were used to assess the effect of vegetation type, snow regime, and their possible interaction. In the PERMANOVA, we accounted for spatial variability observed in earlier studies (Mundra et al., 2015a, 2016a,b) by selecting blocks of fences as a random source of variation. Strength of relationships between GNMDS axis, edaphic variables and experimental factors were assessed based on R2 coefficients of determinations and *P*-values.

OTU Richness

OTU richness, as number of OTUs per sample, was calculated using specnumber function in vegan package. We used linear mixed effects models (lmer function in lme4 package; Bates et al., 2015) to test if there were any effects of experimental factors (nucleic acid, snow regime and vegetation type) on richness of all fungi, symbiotrophs and saprotrophs. Random effects reflected the experimental design where blocks of fences and fences are nested in the design. *P*-values were calculated in Anova function from car package (v. 3.0-0; Fox and Weisberg, 2011). In some cases, components of random variance collapsed to 0, meaning that our data were not sufficient to support a model with this level of complexity. A linear model without fitting random factors gave the same estimations, but slightly increased the values of statistical significance of the results.

RESULTS

Assigned OTUs

In our analysis we retained 42 samples (23 rDNA and 19rRNA). The rDNA and rRNA combined OTU table contained 837 OTUs which included 288 symbiotrophs, 105 saprotrophs, 34 pathotrophs, and 410 unassigned OTUs (**Supplement 3**). The number of OTUs assigned to each trophic mode was similar in rDNA and rRNA (**Supplement 3**). However, symbiotrophs, the dominant trophic mode, were relatively less represented in rRNA than rDNA reads. Both saprotrophs and unassigned reads were twice as abundant in rRNA than in rDNA.

Snow regime showed no clear influence on either community composition or richness (**Table 3**, **Supplements 4**, **5**, other data not shown). The "deep snow" and "control" samples within each vegetation type were therefore pooled in the presented analyses.

Community Composition

GNMDS based on the matrix of all OTUs showed a similar overall trend of community composition for rDNA and rRNA (**Figure 1**). Direct comparison of rDNA and rRNA-derived dissimilarity matrices obtained from 9 co-extracted samples showed a strong correlation between the two (Mantel test observed value: 0.73, $p < 0.001$). Fungal community composition based on all OTUs was primarily divided according to vegetation types: heath and meadow, both for rDNA and rRNA (**Figure 1**). Separate GNMDS analyses of rDNA and rRNA for symbiotrophs and saprotrophs showed the same overall trends, with vegetation type as the main driver in shaping their community compositions ($r2 = 0.27–0.66$ with $p = 0.004$ or less).

The two vegetation types, heath and meadow, differed in edaphic parameters (**Supplement 2**). These edaphic variables fitted onto GNMDS (all OTUs) revealed pH as a significant explanatory variable (**Table 1**), but with different explanatory value depending on template (rDNA or rRNA) and trophic mode. While pH had the highest and dominant explanatory value in all analyses based on rDNA (from $r2 = 0.77$ in all OTUs and symbiotrophs; **Table 1**), other variables tended to explain as much variability in the rRNA dataset (especially organic matter content, as well as the connected nitrogen and carbon contents). Carbon/nitrogen ratio was an important edaphic variable for explaining rDNA-derived community composition ($r2 = 0.27–0.39$), but not at all for rRNA ($r2 = 0.03–0.06$).

Community composition of pathotrophs showed distinct trends in regards to clustering in GNMDS and response to edaphic variable, when rDNA and rRNA were compared, while patterns observed in symbiotrophs and saprotrophs were more similar to each other. The 95% confidence intervals on GNMDS plots showed partial (in rDNA) or total (in rRNA) overlap in meadow and heath. Furthermore, no edaphic variables could explain pathotrophic community composition

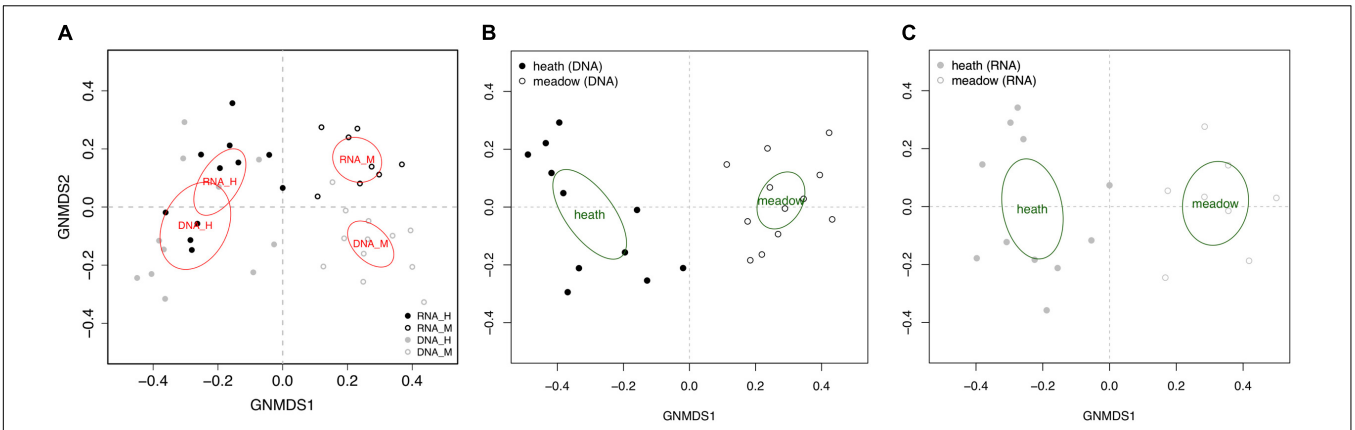


FIGURE 1 | Global non-dimensional scaling of all 42 samples plotted based on presence-absence table that included 837 OTUs **(A)** and according to template (**B**, rDNA – 23 samples; **C**, rRNA – 19 samples) and vegetation type (H, heath; M, meadow).

TABLE 1 | Edaphic variables and vegetation type as a factor fitted into global non-dimensional scaling of all 23 rDNA samples and 19 rRNA samples (plotted based on presence-absence matrixes that included all OTUs, symbiotrophs, saprotrophs, and pathotrophs).

	rDNA		rRNA r2	
	r2	Pr(> r)	r2	Pr(> r)
All_OTUs				
pH	0.77	0.001***	0.67	0.001***
Moisture	0.12	0.245	0.16	0.248
Conductivity	0.22	0.062.	0.36	0.028*
Organic matter	0.12	0.256	0.69	0.001***
Total nitrogen	0.17	0.144	0.61	0.001**
Carbon	0.13	0.239	0.64	0.001**
Carbon/nitrogen ratio	0.39	0.011*	0.03	0.789
Symbiotrophs				
pH	0.77	0.001***	0.63	0.002**
Moisture	0.12	0.277	0.13	0.326
Conductivity	0.22	0.073.	0.36	0.022*
Organic matter	0.12	0.246	0.65	0.001***
Total nitrogen	0.17	0.131	0.50	0.005**
Carbon	0.13	0.214	0.49	0.007**
Carbon/nitrogen ratio	0.39	0.007**	0.06	0.599
Saprotrophs				
pH	0.58	0.001***	0.60	0.002**
Moisture	0.10	0.332	0.17	0.242
Conductivity	0.10	0.340	0.11	0.356
Organic matter	0.10	0.337	0.58	0.001***
Total nitrogen	0.25	0.052.	0.43	0.015*
Carbon	0.19	0.115	0.54	0.002**
Carbon/nitrogen ratio	0.27	0.048*	0.06	0.630
Pathotroph				
pH	0.14	0.224	0.03	0.807
Moisture	0.07	0.457	0.01	0.931
Conductivity	0.01	0.916	0.08	0.500
Organic matter	0.01	0.940	0.04	0.715
Total nitrogen	0.01	0.957	0.04	0.757
Carbon	0.01	0.902	0.03	0.786
Carbon/nitrogen ratio	0.03	0.735	0.01	0.918

*Signif. codes: "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1.*

($r2_{DNA}$ = 0.01–0.14 and $r2_{RNA}$ = 0.01–0.08) with statistical significance ($p > 0.224$).

OTU Richness

Since richness analyses are sensitive to outliers, after initial plotting of these values for all samples, we decided to remove the two highest values (one from each metabarcoding template) that were abnormally high (177 OTUs in rDNA and 159 OTUs in rRNA). Mean richness was higher in rRNA, especially in heath (**Figure 2** and **Table 2**). The same trend was seen in symbiotrophs and unassigned reads, but neither in saprotrophs or pathotrophs (**Figure 2** and **Tables 2**, **4**).

Taking into consideration experimental (metabarcoding template choice and vegetation type) and random factors, the differences in overall OTU richness were driven by the choice of metabarcoding template, rather than by vegetation type (**Table 4**; rDNA < rRNA, model estimation = 16.5, SE = 7.9, p = 0.034 for the template vs. model estimation for vegetation type -2.5, SE = 8.2, p = 0.591). However, based on OTU richness for 9 pairs of co-extracted samples, we saw that the effect of metabarcoding template is important, but not statistically significant (rDNA < rRNA, model estimation = 12.3, SE = 8.1, p = 0.149 for the template).

Overall, out of 827 OTUs, there were 199 OTUs present only in rDNA- and 188 only in rRNA-derived samples. In a subset of 9 co-extracted samples 528 OTUs were detected, from which 135 OTUs were only present in rDNA- and 81 OTUs only in rRNA-based results.

Relative Abundance of Reads

Based on cumulative relative abundances of sequences, symbiotrophs were the dominant group in every combination of factors (metabarcoding template and vegetation type; **Figure 3**). The dominance in relative abundance of symbiotrophic reads was more pronounced in rDNA than rRNA, regardless of vegetation type (by 6.6% of the reads in the meadow and by 15.4% in the heath). Saprotrophs were more abundant in rRNA (by 6% of the reads in heath and 3.3% in meadow). rRNA

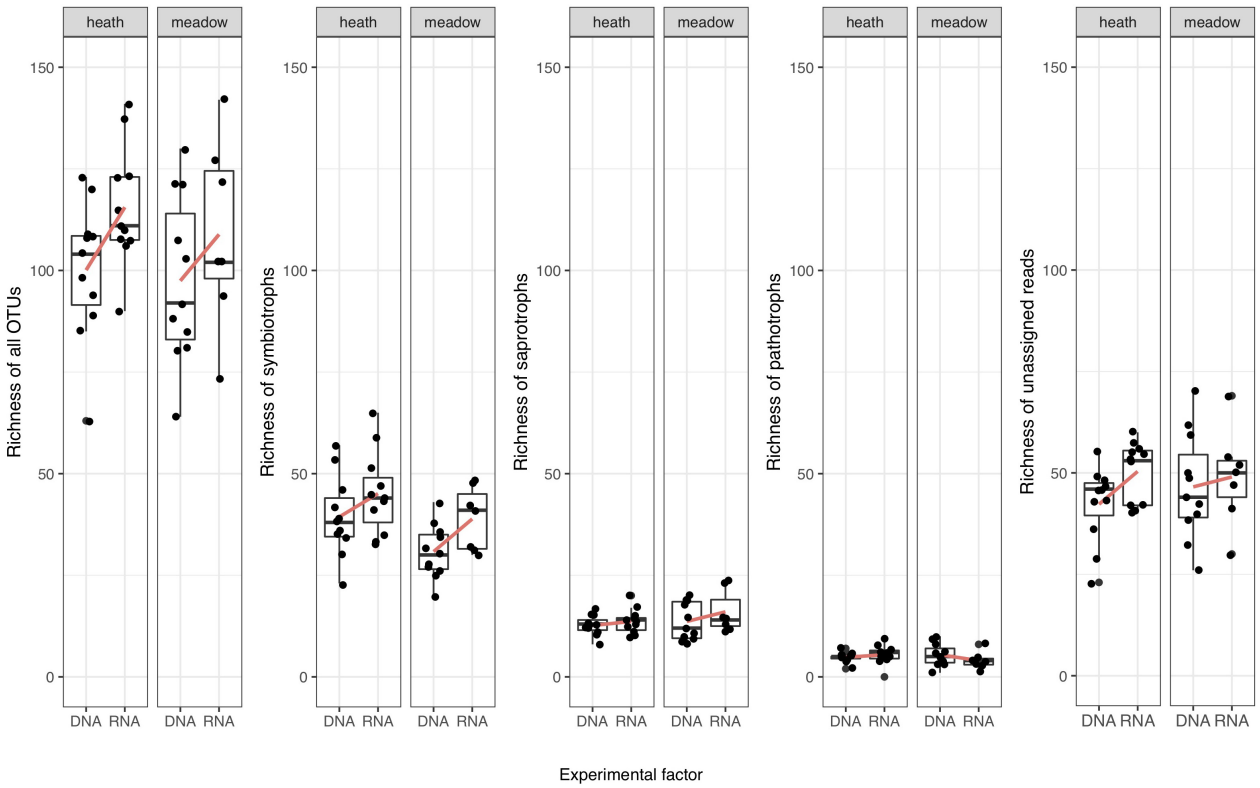


FIGURE 2 | Richness of detected fungal OTUs in meadow and heath (without 2 outliers). Red lines connect mean values.

TABLE 2 | Richness of detected fungal OTUs in a snow fence experimental setup.

	n	μall ± Sd	μ_{Symbio} ± Sd	μ_{sapro} ± Sd	μ_{patho} ± Sd	$\mu_{unassign}$±Sd
DNA_H	11	100.1 ± 17.1	39.4 ± 9.8	12.6 ± 2.5	4.8 ± 1.3	42.3 ± 9.4
RNA_H	11	115.5 ± 14.6	45.1 ± 10.2	13.6 ± 3.0	5.5 ± 2.4	50.4 ± 7.5
DNA_M	12	104.1 ± 30.2	31.4 ± 6.6	15.2 ± 6.9	6.3 ± 3.8	50.3 ± 18.0
RNA_M	8	115.1 ± 27.8	41.2 ± 9.9	16.9 ± 5.5	4.3 ± 2.1	51.8 ± 13.6
DNA_M (no outliers)	11	97.5 ± 20.6	30.8 ± 6.6	13.6 ± 4.6	5.5 ± 2.7	46.5 ± 13.2
RNA_M (no outliers)	7	108.9 ± 23.1	38.6 ± 7.8	16.0 ± 5.3	4.0 ± 2.1	49.0 ± 12.0

n, number of samples; H, heath; M, meadow; μ, mean richness for all (μ_{all}), symbiotrophic (μ_{symbio}), saprotrophic (μ_{sapro}), pathotrophic (μ_{patho}), and unnasigned OTUs ($\mu_{unassign}$); sd, standard deviation.

harbored significantly more functionally unassigned sequences than rDNA, especially in heath where the difference was the highest (10.6% of reads). Similarly to saprotrophs, reads not assigned to any trophic mode, were twice as abundant in rRNA than rDNA-based results. We observed that an increase in relative number of reads from saprotrophic and unassigned trophic modes originated from overall higher richness, as well as highly expressed rRNA in a particular OTU (**Figure 4**).

Taxonomic Groups

Although fungi in each trophic mode are functionally similar in the ecosystem, species can belong to distantly related fungal orders. For combination of trophic modes and vegetation types we detected taxonomic groups that might contribute in varying degree to a bias between rDNA and rRNA-derived results. Within each taxonomic group OTUs responded in different ways: some showed overexpressed rRNA, some were more abundant in rDNA and other OTUs did not change their abundance when rDNA and rRNA-derived results were compared. The most consistent overrepresentation of any taxonomic order in rDNA-results was observed in Agaricales in every combination of trophic mode and vegetation type, except saprotrophs in heath (**Figure 4**). There the numbers of Agaricales reads did not differ between rDNA and rRNA-derived results. Symbiotrophic reads overrepresented in rRNA belonged to Russulales and Thelephorales regardless of vegetation type, whereas overrepresentation of rRNA-derived sequences from Pezizales occurred in the heath. Saprotrophic reads that appeared more often in rRNA in both vegetation types were Helotiales, and additionally – only in the heath: Mortierellales and only

TABLE 3 | Permutational multivariate analysis of variance (PERMANOVA, adonis function in vegan package) based on rDNA and rRNA presence-absence matrixes of all, symbiotrophic, saprotrophic and pathotrophic OTUs.

	Vegetation type		Snow regime		Vegetation X Snow	
	r2	*p*	*r2*	*p*	*r2*	*p*
All OTUs rDNA	0.16	0.133	0.04	0.247	0.04	0.102
All OTUs rRNA	0.14	0.047*	0.06	0.011*	0.05	0.374
Symbiotrophs rDNA	0.15	0.505	0.04	0.617	0.04	0.391
Symbiotrophs rRNA	0.06	0.512	0.05	0.737	0.06	0.247
Saprotrophs rDNA	0.13	0.023*	0.06	0.024*	0.05	0.047*
Saprotrophs rRNA	0.10	0.435	0.07	0.113	0.04	0.897
Pathotrophs rDNA	0.12	0.045*	0.06	0.081.	0.06	0.114
Pathotrophs rRNA	0.08	0.424	0.05	0.402	0.06	0.459

Vegetation type and snow regime factors were first tested in forward selection before testing for interaction. Signif. codes: 0.01 "" 0.05 "." 0.1 " " 1.*

in meadow: Tremellodendropsidales, whereas Sordariales and Hypocreales reads were more numerous in rDNA in the heath than in the meadow. Functionally unassigned reads in rRNA pool were predominantly unassigned taxonomically to order or higher rank in both vegetation types, whereas Sebacinales were found overexpressed only in the heath and Helotiales only in the meadow.

DISCUSSION

Similarities between rDNA and rRNA metabarcoding of microbial or cryptic species has received little attention in cold terrestrial environments. Low temperatures, often below 0°C, can slow microbial processes, including the decomposition of dead biomass. These cells remain in the soil and contribute to a pool of relic rDNA. Our case study contributes to understanding which types of analyses of sequences parsed in ecologically meaningful units may result in most discrepancy between the two metabarcoding templates. Moreover, we made an attempt to link both fungal functions and diversity, to pinpoint possible sources of differences in rDNA and rRNA-derived results from cold environments.

Our comparison between rDNA and rRNA metabarcoding templates unveiled no or little divergence in community composition, also when the sequences were divided into fungal trophic modes. The clustering according to vegetation type agrees with former studies, supporting the importance of a long-lasting interaction between fungal community structure and vegetation type (Chu et al., 2011; Shi et al., 2015). This general trend was also consistent for community composition of symbiotrophs and saprotrophs, demonstrating the fine-tuning of these functional groups with the vegetation type. Ordinations based on pathotrophs, the least represented group, both in number of OTUs and number of reads, did not show clear differences between vegetation types (as other trophic modes); this pattern may be due to their stochastic distribution in the soil (Bahram et al., 2016). We speculate

TABLE 4 | Results of linear mixed models explaining richness of all OTUs, symbiotrophs, saprotrophs and saprotrophs.

Response	Fixed effects					Interaction		Random effects	
Richness	**Intercept ± 1SE**	**Template ± 1SE**	***p***	**Vegetationl ± SE**	***p***	**Template x vegetation type±1SE**	***P***	**Fence:block ± SD**	**Block ± SD**
All OTUs	99.7 ± 5.8	16.5 ± 7.9	0.034	−2.5 ± 8.2	0.591	−4.7 ± 12.0	0.729	13.94 ± 3.7	0
Symbiotrophs	39.0 ± 3.1	6.2 ± 3.7	0.031	−8.2 ± 4.3	0.200	1.8 ± 5.7	0.775	0	5.2 ± 2.3
Saprotrophs (lm)	12.3 ± 1.0	1.0 ± 1.6	0.548	1.0 ± 1.6	0.548	1.4 ± 2.5	0.588		
Pathotrophs (lm)	4.8 ± 0.7	0.6 ± 0.9	0.503	0.6 ± 0.9	0.503	−2.1 ± 1.4	0.015		
Unassigned	41.8 ± 3.4	9.2 ± 4.4	0.080	4.4 ± 4.8	0.916	−6.2 ± 6.7	0.410	11.36 ± 3.4	0

In cases with all OTUs, saprotrophs and functionally unassigned reads variance of random effects, both nested (fence:block) and main (block) could not be estimated (values = 0), so in these cases we used linear model (lm) instead. Table includes all the data, excluding 2 outlier samples.

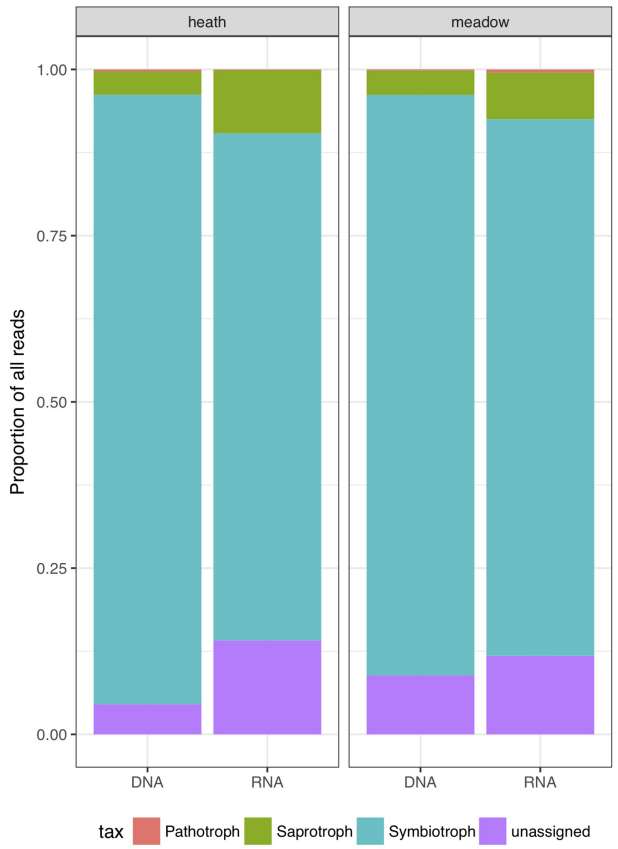


FIGURE 3 | Relative abundances of all reads divided into trophic modes (saprotrophs, symbiotrophs, pathotrophs and functionally unassigned OTUs).

that the strong impact of vegetation type can partly mask effects of other factors, such as metabarcoding template and snow fence treatment (**Supplements 4, 5**). Our findings suggest that a possible bias introduced by rDNA-based metabarcoding does not influence the main conclusions regarding community composition.

Symbiotrophs are usually the dominating fungal functional group in soils, also in the Arctic (Gardes and Dahlberg, 1996; Clemmensen et al., 2006; Mundra et al., 2016b), a trend supported by our study. Both the highest number of OTUs and the largest proportion of sequences belonged to symbiotrophs, especially Agaricales. Although dominating in both templates, a higher proportion of symbiotrophic reads that belong to Agaricales were detected in rDNA than in rRNA regardless of vegetation type. This may suggest that relatively more symbiotrophic rDNA originate from dead cells or extracellular rDNA (Carini et al., 2016). It is plausible that more symbiotrophic rDNA is retained in soil because Agaricales are simply more abundant than other fungi. On the other hand, symbiotrophic Thelephorales, Pezizales or Russulales might be overestimated when rRNA is used as an estimator for abundance. The observed higher proportions of saprotrophic reads in rRNA samples than in rDNA, suggest that saprotrophic OTUs produce relatively more rRNA, especially in Helotiales, hence are more active than the rDNA data would suggest. As we sampled only on one date it is not possible to tell whether data would be similar throughout the year or if there would be taxonomically specific responses to temporal dynamics within the tundra soil.

Fungi with functionally unassigned sequences comprised a substantial proportion of all heath sequences, based on rRNA. Unassigned sequences in this study originated mainly from novel organisms without any database matches but also from unresolved/ambiguous functions that change throughout fungal life cycle or due to changes in the environmental conditions (**Figure 4**). We argue that the taxonomically unresolved component of the fungal community contributes substantially to active fungal community and recommend looking into these unknown OTUs with unknown functions.

Differences in the explanatory power of environmental variables between rDNA and rRNA-based community compositions have been reported only in a few studies comparing outcomes from both metabarcoding templates (i.e., Barnard et al., 2013; Zhang et al., 2014; Lüneberg et al., 2018). Yet it seems important to understand which parameters are crucial for shaping the community composition. Our study confirmed that pH is an important edaphic variable (Bååth and Anderson, 2003; Rousk et al., 2010; Mundra et al., 2016a), which explained both rDNA and rRNA-derived community composition. However, the similarities between explanatory power of the most important edaphic variables between the two templates end here. Langenheder and Prosser (2008) found that resource availability (such as organic matter, nitrogen and carbon concentration) explained most variability within rRNA-based results from heterotrophic soil bacteria. Fungi are also heterotrophic organisms that rely on resource availability. Both bacteria and fungi enhance their growth rate and cellular capacity for protein synthesis when metabolically available nitrogen and carbon levels increase (for more on regulation: Broach, 2012; You et al., 2013). Effectively, this means that cells transcribe more rRNA in order to produce more ribosomes for protein synthesis, to use available resources more efficiently.

The level of expressed rRNA is not always equivalent to the level of growing and dividing cells. Instead, the increased number of rRNA may rather be a stress response for handling multiple stressors and in order to do so, cells transcribe more ribosomes than they would for growth without these stressors (Blazewicz et al., 2013). Contrary to some microbial dormant stages, such as bacterial spores, basidiospores of five species of fungi were shown not to contain rRNA (Van der Linde and Haller, 2013), implying that by using rRNA in our study we eliminate the contribution of not only dead cells, but also dormant stages of fungi. However, just before germination when the spores swell, the amount of rRNA increases rapidly and not proportionally to normal cellular growth (Moore et al., 2011), possibly influencing our results to some extent.

The relationship between the number of sequences originating from rDNA and rRNA is complex. The number of rDNA copies in a genome differs between organisms, also between fungal

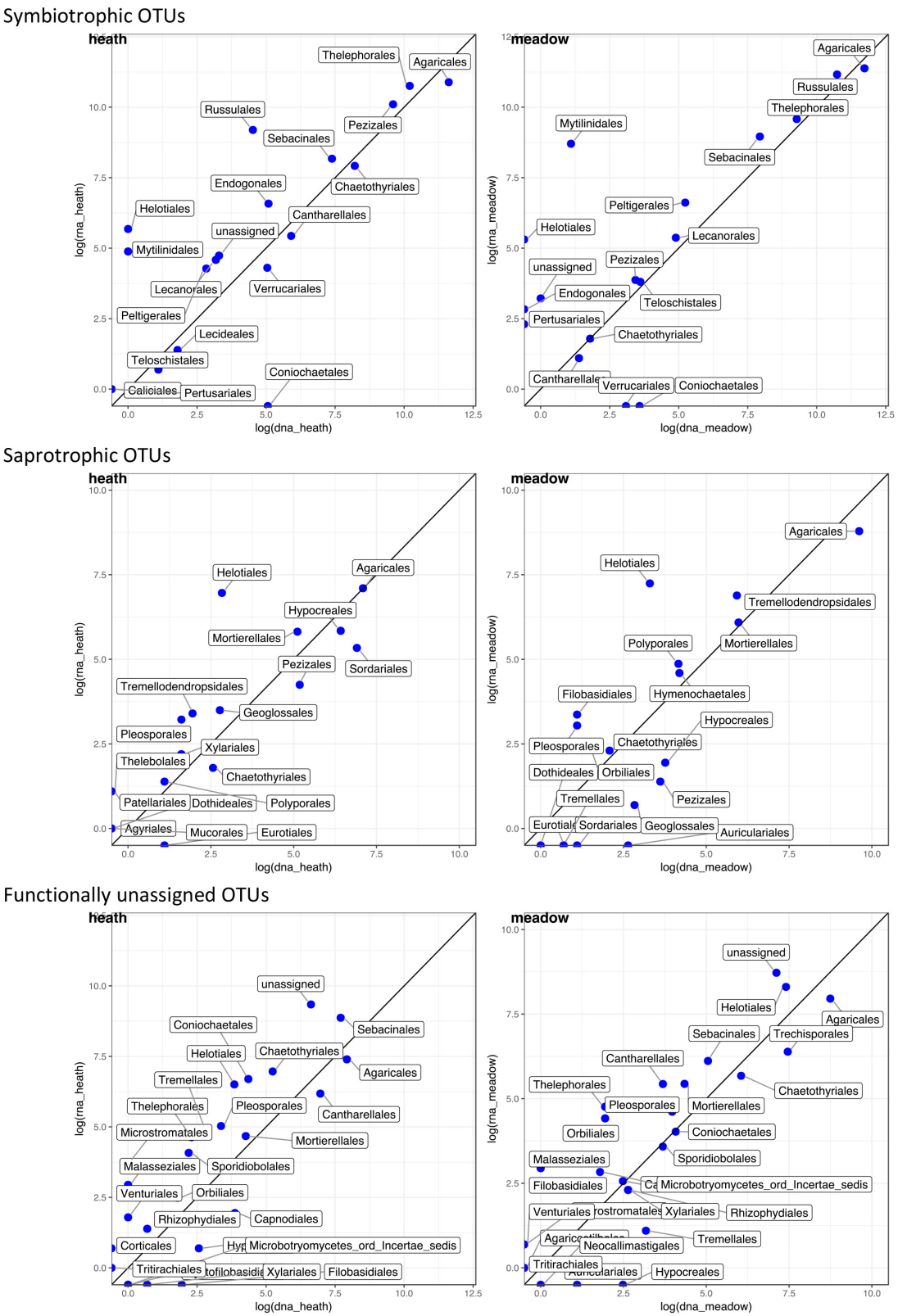


FIGURE 4 | Correlation of rDNA- and rRNA-derived abundances of OTUs grouped in higher taxonomic rank (order) and divided into assigned trophic modes. Abundance data come from 9 pairs of coresponding rDNA and rRNA samples; data were log transformed. Data points above the line show orders which contributed more to rRNA than rDNA pool; and vice versa, data points beneath the line point out orders that contributed more to rDNA than rRNA pool.

species (Torres-Machorro et al., 2010; Black et al., 2013; Das and Deb, 2015; Johnson et al., 2015). Copy numbers of rRNA (ribosomes) can differ depending on conditions and is a result of the synthesis and degradation rates (Blazewicz et al., 2013). However, by targeting the ITS fragments in this study, we eliminated the influence of ribosome degradation rates, since ITS is removed from the rRNA precursor prior to ribosome formation (Schoch et al., 2012). While relationships between cellular growth and rRNA can be measured for cultured organisms in carefully controlled laboratory conditions, it is not known how this ratio is maintained in a complex environment full of interactions and stressors. It is, however, clear that rDNA copy numbers vary less over time or in different conditions than the number of rRNA per cell, making rDNA a rather poor predictor of growth or approximation of biomass content (Blazewicz et al., 2013).

Changes of environmental and edaphic parameters can cause shifts in fungal community compositions or in fungal richness. Strong seasonality in environments, such as in the Arctic tundra, lead to temporal dynamics within fungal communities (Mundra et al., 2015a), which can respond differently to changing conditions depending on their function in the ecosystem (Mundra et al., 2016b). At the same time, cold conditions may delay decomposition or favor preservation of dead biomass (Conant et al., 2011; Ejarque and Abakumov, 2016). In these circumstances, microbial communities monitored only with rDNA-based marker genes reflect not only currently thriving microbes, but also these active in the past, even in a multidecade time frame (Yoccoz et al., 2012). Our study explored differences of tundra soil between total and active fungal communities at only one time point. A study of the temporal dynamics of rDNA and rRNA across all seasons would profoundly enhance our understanding of the possible seasonal differences in microbial community composition, especially after major changes in environmental conditions.

DATA AVAILABILITY STATEMENT

Sequencing data generated for this study is available online: 10.5281/zenodo.146288. Detailed description of bioinformatics pipeline, mapping file for demultiplexing, environmental dataset, OTU table, taxonomic and functional assignments and R code generated for this study is available at https://github.com/magdawutkowska/Dead_or_alive.

AUTHOR CONTRIBUTIONS

MW analyzed the data and wrote the manuscript. AV planned, collected samples, processed samples, analyzed the data, discussed the results, and the manuscript. SM planned, collected samples, discussed the results, and the manuscript. EC designed the sampling site, discussed the results, and the manuscript. PE planned, financed, discussed the results, and wrote the manuscript.

FUNDING

This research was funded by ConocoPhillips and Lundin Petroleum through the Northern Area Program and by the Norwegian Research Council, project number 230970. This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian Metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. The publication charges for this paper have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

ACKNOWLEDGMENTS

We want to acknowledge the help received from the Department for Research Computing at USIT, the University of Oslo IT-department. Moreover, we would like to thank Dorothee Ehrich, Mette M. Svenning, Kim Præbel and the reviewers for valuable comments on the manuscript and Clara Ruiz-González for help with plotting figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.03243/full#supplementary-material

REFERENCES

Anderson, I. C., and Parkin, P. I. (2007). Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *J. Microbiol. Methods* 68, 248–253. doi: 10.1016/j.mimet.2006.08.005

Bååth, E., and Anderson, T. H. (2003). Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. *Soil Biol. Biochem.* 35, 955–963. doi: 10.1016/S0038-0717(03)00154-8

Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., and Tedersoo, L. (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME J.* 10, 885–896. doi: 10.1038/ismej.2015.164

Bálint, M., Schmidt, P. A., Sharma, R., Thines, M., and Schmitt, I. (2014). An Illumina metabarcoding pipeline for fungi. *Ecol. Evol.* 4, 2642–2653. doi: 10.1002/ece3.1107

Barnard, R. L., Osborne, C. A., and Firestone, M. K. (2013). Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *ISME J.* 7, 2229–2241. doi: 10.1038/ismej.2013.104

Barnes, M. A., and Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* 17, 1–17. doi: 10.1007/s10592-015-0775-4

Bass, D., Stentiford, G. D., Littlewood, D. T. J., and Hartikainen, H. (2015). Diverse applications of environmental DNA methods in parasitology. *Trends Parasitol.* 31, 499–513. doi: 10.1016/j.pt.2015.06.013

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 4, 914–919. doi: 10.1111/2041-210X.12073

Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M., and Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* 13, 218–224. doi: 10.1111/1755-0998.12065

Black, J., Dean, T., Byfield, G., Foarde, K., and Menetrez, M. (2013). Determining fungi rRNA copy number by PCR. *J. Biomol. Tech.* 24, 32–38. doi: 10.7171/jbt.13-2401-004

Blazewicz, S. J., Barnard, R. L., Daly, R. A., and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* 7, 2061–2068. doi: 10.1038/ismej.2013.102

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., et al. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29, 358–367. doi: 10.1016/j.tree.2014.04.003

Broach, J. R. (2012). Nutritional control of growth and development in yeast. *Genetics* 192, 73–105. doi: 10.1534/genetics.111.135731

Buckeridge, K. M., and Grogan, P. (2008). Deepened snow alters soil microbial nutrient limitations in arctic birch hummock tundra. *Appl. Soil Ecol.* 39, 210–222. doi: 10.1016/j.apsoil.2007.12.010

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). Correspondence QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Publish. Group* 7, 335–336. doi: 10.1038/nmeth0510-335

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). Biodiversity loss and its impact on humanity. *Nature* 486, 59–67. doi: 10.1038/nature11148

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S., and Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* 2:16242. doi: 10.1038/nmicrobiol.2016.242

Cernansky, R. (2017). Biodiversity moves beyond counting species. *Nature* 546, 22–24. doi: 10.1038/546022a

Chessel, D., Dufour, A. B., and Thioulouse, J. (2004). The ade4 package – I?: one-table methods. *R News* 4, 5–10.

Chu, H., Neufeld, J. D., Walker, V. K., and Grogan, P. (2011). The influence of vegetation type on the dominant soil bacteria, archaea, and fungi in a low arctic tundra landscape. *Soil Sci. Soc. Am. J.* 75:1756. doi: 10.2136/sssaj2011.0057

Clemmensen, K. E., Michelsen, A., Jonasson, S., and Shaver, G. R. (2006). Increased ectomycorrhizal fungal abundance after long-term fertilization and warming of two arctic tundra ecosystems. *New Phytol.* 171, 391–404. doi: 10.1111/j.1469-8137.2006.01778.x

Conant, R. T., Ryan, M. G., Ågren, G. I., Birge, H. E., Davidson, E. A., Eliasson, P. E., et al. (2011). Temperature and soil organic matter decomposition rates – Synthesis of current knowledge and a way forward. *Glob. Change Biol.* 17, 3392–3404. doi: 10.1111/j.1365-2486.2011.02496.x

Cooper, E. J., Dullinger, S., and Semenchuk, P. (2011). Late snowmelt delays plant development and results in lower reproductive success in the High Arctic. *Plant Sci.* 180, 157–167. doi: 10.1016/j.plantsci.2010.09.005

Costello, M. J. (2015). Biodiversity: the known, unknown, and rates of extinction. *Curr. Biol.* 25, R368–R371. doi: 10.1016/j.cub.2015.03.051

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., et al. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods Ecol. Evol.* 7, 1008–1018. doi: 10.1111/2041-210X.12574

Das, S., and Deb, B. (2015). DNA barcoding of fungi using ribosomal ITS marker for genetic diversity analysis: a review. *Int. J. Pure Appl. Biosci.* 3, 160–167.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C. (2016). UCHIME2: improved chimera prediction for amplicon sequencing. *BioRxiv* [preprint]. doi: 10.1101/074252

Ejarque, E., and Abakumov, E. (2016). Stability and biodegradability of organic matter from Arctic soils of Western Siberia: insights from13C-NMR spectroscopy and elemental analysis. *Solid Earth* 7, 153–165. doi: 10.5194/se-7-153-2016

Fodor, E. (2011). Ecological niche of plant pathogens. *Ann. For. Res.* 54, 3–21.

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*, 2nd Edn. Thousand Oaks, CA: Sage.

Gardes, M., and Dahlberg, A. (1996). Mycorrhizal diversity in arctic and alpine tundra: an open question. *New Phytol.* 133, 147–157. doi: 10.1111/j.1469-8137.1996.tb04350.x

Gilichinsky, D. A., Wagener, S., and Vishnevetskaya, T. A. (1995). Permafrost microbiology. *Perm. Perigl. Process.* 6, 281–291. doi: 10.1002/ppp.3430060402

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., et al. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* 7, 1299–1307. doi: 10.1111/2041-210X.12595

Hawksworth, D. L., and Lücking, R. (2017). Fungal diversity revisited?: 2. 2 to 3. 8 million species. *Microbiol. Spect.* 5, 1–17. doi: 10.1128/microbiolspec.FUNK-0052-2016.Correspondence

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition – Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12:e04957. doi: 10.1371/journal.pcbi.1004957

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., et al. (2012). New primers to amplify the fungal ITS2 region – Evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* 82, 666–677. doi: 10.1111/j.1574-6941.2012.01437.x

Johnson, S. M., Carlson, E. L., and Pappagianis, D. (2015). Determination of ribosomal DNA copy number and comparison among strains of coccidioides. *Mycopathologia* 179, 45–51. doi: 10.1007/s11046-014-9820-y

Kaiser, C., Meyer, H., Biasi, C., Rusalimova, O., Barsukov, P., and Richter, A. (2007). Conservation of soil organic matter through cryoturbation in arctic soils in Siberia. *J. Geophys. Res. Biogeosci.* 112, 1–8. doi: 10.1029/2006JG000258

Kohler, A., Kuo, A., Nagy, L. G., Morin, E., Barry, K. W., Buscot, F., et al. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* 47, 410–415. doi: 10.1038/ng.3223

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi: 10.1111/mec.12481

Konopka, A. (2009). What is microbial community ecology. *ISME J.* 3, 1223–1230. doi: 10.1038/ismej.2009.88

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27. doi: 10.1007/BF02289565

Langenheder, S., and Prosser, J. I. (2008). Resource availability influences the diversity of a functional group of heterotrophic soil bacteria. *Environ. Microbiol.* 10, 2245–2256. doi: 10.1111/j.1462-2920.2008.01647.x

Li, R., Tun, H. M., Jahan, M., Zhang, Z., Kumar, A., Fernando, D., et al. (2017). Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water. *Sci. Rep.* 7:5752. doi: 10.1038/s41598-017-02516-3

Liu, H., Økland, T., Halvorsen, R., Gao, J., Liu, Q., Eilertsen, O., et al. (2008). Gradients analyses of forests ground vegetation and its relationships to environmental variables in five subtropical forest areas, S and SW China. *Sommerfeltia* 32, 1–196. doi: 10.2478/v10208-011-0012-6

Louca, S., Jacques, S. M. S., Pires, A. P. F., Leal, J. S., Srivastava, D. S., Parfrey, L. W., et al. (2016). High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1:0015. doi: 10.1038/s41559-016-0015

Lüneberg, K., Schneider, D., Siebe, C., and Daniel, R. (2018). Drylands soil bacterial community is affected by land use change and different irrigation practices in the Mezquital Valley, Mexico. *Sci. Rep.* 8, 1–15. doi: 10.1038/s41598-018-19743-x

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13:31. doi: 10.1186/1471-2105-13-31

Moore, D., Robson, G. D., and Trinci, A. P. (2011). *21st Century Guidebook to Fungi*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511977022

Morgado, L. N., Semenova, T. A., Welker, J. M., Walker, M. D., Smets, E., and Geml, J. (2016). Long-term increase in snow depth leads to compositional changes in arctic ectomycorrhizal fungal communities. *Glob. Change Biol.* 22, 3080–3096. doi: 10.1111/gcb.13294

Morgner, E., Elberling, B., Strebel, D., and Cooper, E. J. (2010). The importance of winter in annual ecosystem respiration in the High Arctic: effects of snow depth in two vegetation types. *Polar Res.* 29, 58–74. doi: 10.1111/j.1751-8369.2010.00151.x

Mundra, S., Bahram, M., and Eidesen, P. B. (2016a). Alpine bistort (*Bistorta vivipara*) in edge habitat associates with fewer but distinct ectomycorrhizal fungal species: a comparative study of three contrasting soil environments in Svalbard. *Mycorrhiza* 26, 809–818. doi: 10.1007/s00572-016-0716-1

Mundra, S., Halvorsen, R., Kauserud, H., Bahram, M., Tedersoo, L., Elberling, B., et al. (2016b). Ectomycorrhizal and saprotrophic fungi respond differently to long-term experimentally increased snow depth in the High Arctic. *Microbiol. Open* 5, 856–869. doi: 10.1002/mbo3.375

Mundra, S., Bahram, M., Tedersoo, L., Kauserud, H., Halvorsen, R., and Eidesen, P. B. (2015a). Temporal variation of *Bistorta vivipara* -associated ectomycorrhizal fungal communities in the High Arctic. *Mol. Ecol.* 24, 6289–6302. doi: 10.1111/mec.13458

Mundra, S., Halvorsen, R., Kauserud, H. H., Müller, E., Vik, U., Eidesen, P. B., et al. (2015b). Arctic fungal communities associated with roots of *Bistorta vivipara* do not respond to the same fine-scale edaphic gradients as the aboveground vegetation. *New Phytol.* 205, 1587–1597. doi: 10.1111/nph.13216

Nguyen, N. H., Smith, D., Peay, K., and Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytol.* 205, 1389–1393. doi: 10.1111/nph.12923

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241–248. doi: 10.1016/j.funeco.2015.06.006

Ogram, A., Sayler, G. S., Gustln, D., and Lewis, R. J. (1988). DNA adsorption to soils and sediments. *Environ. Sci. Technol.* 22, 982–984. doi: 10.1021/es00173a020

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Lagendre, P., McGlinn, D., et al. (2018). *Vegan: Community Ecology Package. R package version 2.5-1*. Available at: https://cran.r-project.org/package=vegan

Olofsson, J., Ericson, L., Torp, M., Stark, S., and Baxter, R. (2011). Carbon balance of Arctic tundra under increased snow cover mediated by a plant pathogen. *Nat. Clim. Change* 1, 220–223. doi: 10.1038/nclimate1142

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C., Haile, J., Hellstrom, M., et al. (2015). Ancient and modern environmental DNA. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20130383. doi: 10.1098/rstb.2013.0383

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351. doi: 10.1038/ismej.2010.58

Saeki, K., and Kunito, T. (2010). Adsorptions of DNA molecules by soils and variable-charged soil constituents GMOs. *Appl. Microbiol.* 1, 188–195.

Saeki, K., Kunito, T., and Sakai, M. (2011). Effect of Tris-HCl buffer on DNA adsorption by a variety of soil constituents. *Microbes Environ.* 26, 88–91. doi: 10.1264/jsme2.ME10172

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1–6. doi: 10.1073/pnas.1117018109

Semenchuk, P. R., Elberling, B., Amtorp, C., Winkler, J., Rumpf, S., Michelsen, A., et al. (2015). Deeper snow alters soil nutrient availability and leaf nutrient status in high Arctic tundra. *Biogeochemistry* 124, 81–94. doi: 10.1007/s10533-015-0082-7

Semenova, T. A., Morgado, L. N., Welker, J. M., Walker, M. D., Smets, E., and Geml, J. J. (2016). Compositional and functional shifts in arctic fungal communities in response to experimentally increased snow depth. *Soil Biol. Biochem.* 100, 201–209. doi: 10.1016/j.soilbio.2016.06.001

Shelton, A. O., O'Donnell, J. L., Samhouri, J. F., Lowell, N., Williams, G. D., and Kelly, R. P. (2016). A framework for inferring biological communities from environmental DNA. *Ecol. Appl.* 26, 1645–1659. doi: 10.1890/15-1733.1

Shi, Y., Xiang, X., Shen, C., Chu, H., Neufeld, J. D., Walker, V. K., et al. (2015). Vegetation-associated impacts on Arctic tundra bacterial and microeukaryotic communities. *Appl. Environ. Microbiol.* 81, 492–501. doi: 10.1128/AEM.03229-14

Soina, V. S., Vorobiova, E. A., Zvyagintsev, D. G., and Gilichinsky, D. A. (1995). Preservation of cell structures in permafrost: a model for exobiology. *Adv. Space Res.* 15, 237–242. doi: 10.1016/S0273-1177(99)80090-8

Thomsen, P. F., and Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18. doi: 10.1016/j.biocon.2014.11.019

Timling, I., and Taylor, D. L. (2012). Peeking through a frosty window: molecular insights into the ecology of Arctic soil fungi. *Fungal Ecol.* 5, 419–429. doi: 10.1016/j.funeco.2012.01.009

Torres-Machorro, A. L., Hernández, R., Cevallos, A. M., and López-Villaseñor, I. (2010). Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? *FEMS Microbiol. Rev.* 34, 59–86. doi: 10.1111/j.1574-6976.2009.00196.x

Van der Linde, S., and Haller, S. (2013). Obtaining a spore free fungal community composition. *Fungal Ecol.* 6, 522–526. doi: 10.1016/j.funeco.2013.10.001

White, T. J., Bruns, S., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phyologenetics," in *PCR Protocols: A Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand, and J. J. Sninsky (New York, NY: Academic Press Inc.).

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., et al. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* 21, 3647–3655. doi: 10.1111/j.1365-294X.2012.05545.x

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., et al. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500, 301–306. doi: 10.1038/nature12446

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G. J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* 23, 2260–2274. doi: 10.1111/mec.12739

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers

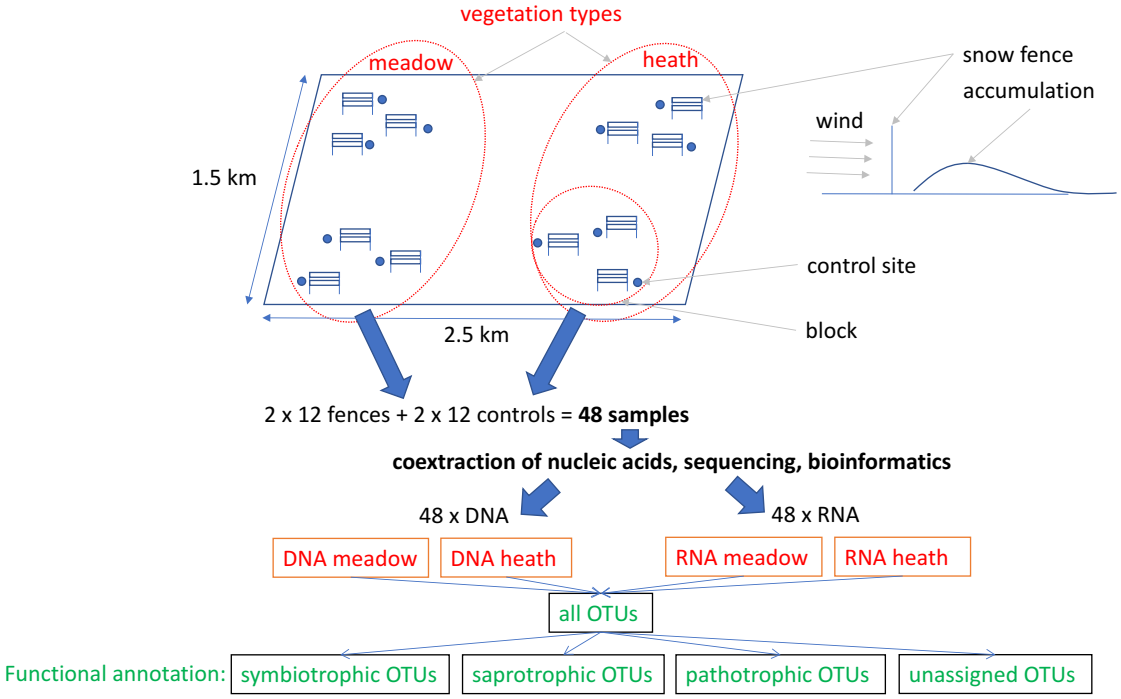
Supplementary Material

Dead or alive; or does it really matter? Level of congruency between trophic modes in total and active fungal communities in High Arctic soil.

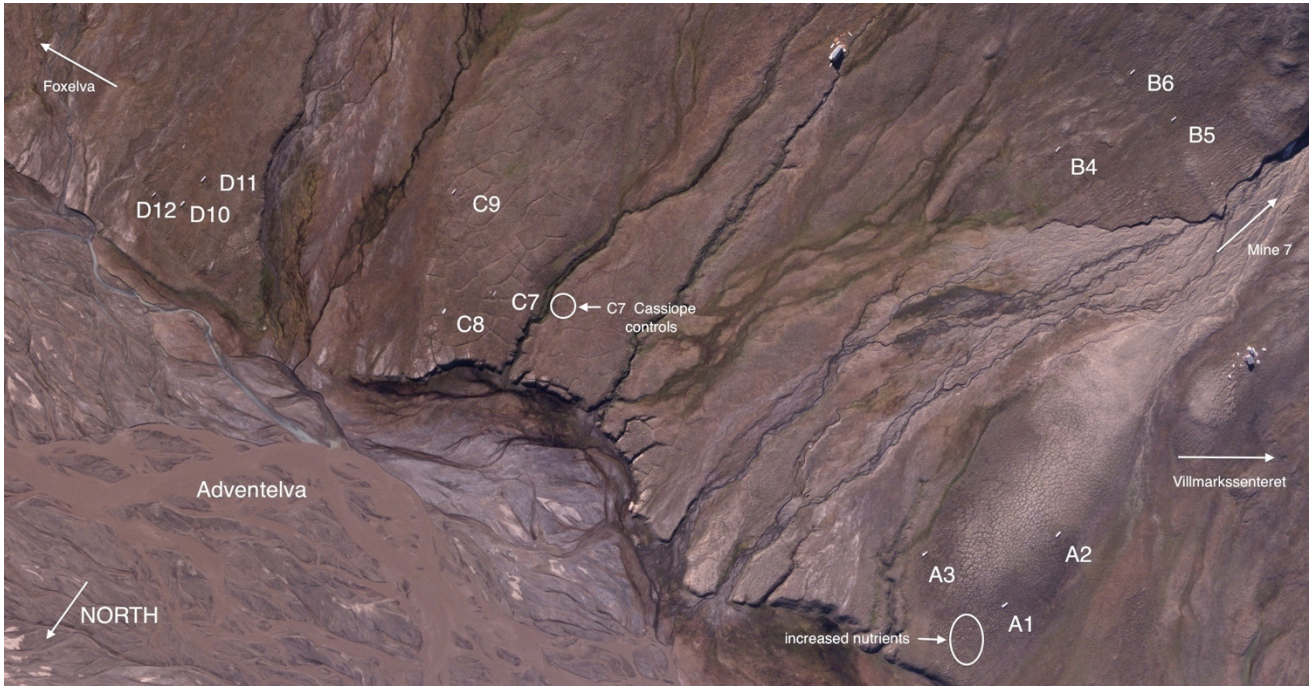
Magdalena Wutkowska* [1, 2], Anna Vader [1], Sunil Mundra [3], Elisabeth J. Cooper [2], Pernille B. Eidesen [1]* **Correspondence:** Corresponding Author: magdalena.wutkowska@unis.no; magda.wutkowska@gmail.com

Supplementary Figures and Tables

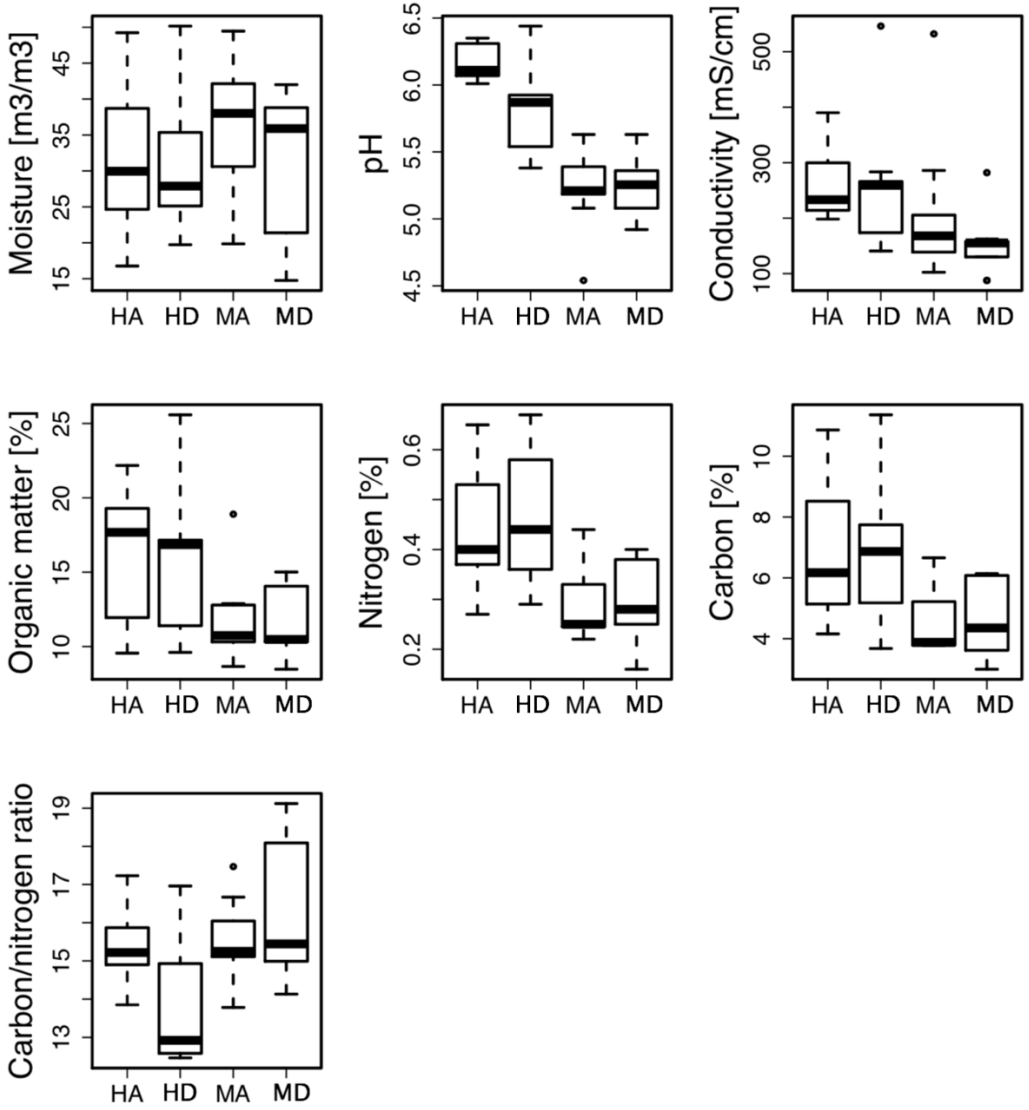
Supplement 1a | Experimental setup and design of the study. Twelve fences were erected, six on each of meadow and heath vegetation. The fences/ambient areas were arranged into blocks of three across the landscape, within an area of approx. 1.5-2.5 km. Heaths were dominated by *Cassiope tetragona* and had faster-draining stony soils and more undulating topography than the flatter meadows dominated by *Salix polaris* and *Luzula arcuata ssp. confusa* (Morgner et al., 2010). Snow depth and resulting date of snowmelt were manipulated using fences placed perpendicular to the winter wind direction. Beside each fence an unmanipulated area was designated with ambient snow conditions. Ambient regimes had 10-35cm snow in winter, whereas deep regimes had up to 150cm snow, and melted out ca. 17 days later (Semenchuk *et al*., 2013).



Supplement 1b | Satellite image of the snow fence experimental setup in Adventdalen, Svalbard. A1-B3 fences are located in two vegetation types: heath and in meadow.



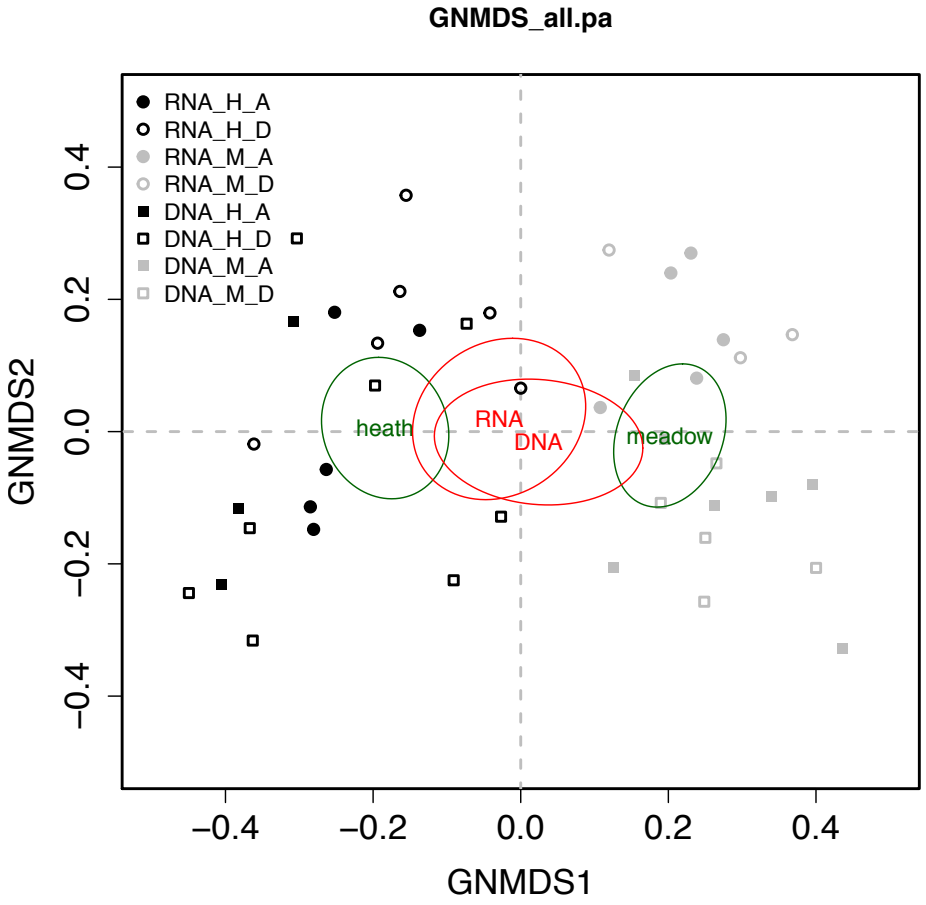
Supplement 2 | Boxplots representing differences in edaphic variables among ambient (A) and deep (D) snow regimes in two vegetation types heath (H) and meadow (M).



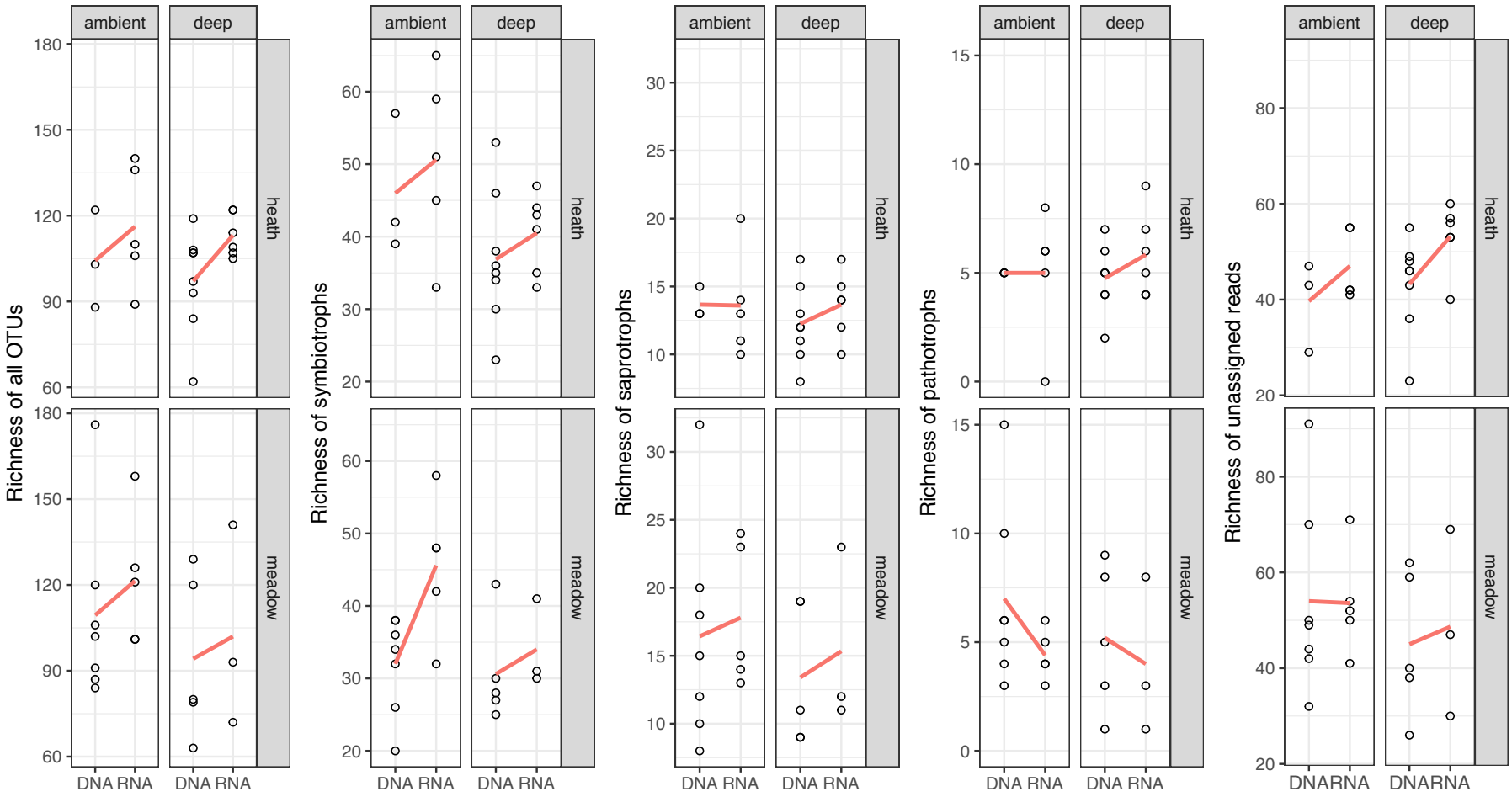
Supplement 3 | Summary information of the dataset.

	all OTUs (n=42)			rDNA_OTUs (n=23)			rRNA_OTUs (n=19)		
	total number of OTUs	mean no of reads per sample ± SD	percentage of total number of reads	total number of OTUs	mean no of reads per sample ± SD	percentage of total number of reads	total number of OTUs	mean number of reads per sample ± SD	percentage of total number of reads
pathotroph	34	102±215	0.2%	27	95±121	0.2%	22	111±296	0.2%
saprotroph	105	2460±4428	5.8%	83	1528±3356	3.6%	83	3588±5334	8.4%
symbiotroph	288	35820±7787	84.3%	237	37981±5236	89.4%	227	33203±9561	78.2%
unassigned	410	4105±4909	9.7%	309	2884±4302	6.8%	306	5585±5298	13.1%
sum	837			656			638		

Supplement 4 | Global non-dimensional scaling of all 42 samples plotted according to template (rDNA/rRNA), vegetation type (H – heath, M - meadow) and snow regime (A – ambient snow regime, D – deep snow regime introduced by snow fences). Figure is based on presence-absence table.



Supplement 5 | OTU richness per experimental factor: template, vegetation type and snow regime. Red lines connects mean values between levels of each factor (with outliers).



PAPER II

Can root-associated fungi mediate the impact of abiotic conditions on the growth of a High Arctic herb?

Magdalena Wutkowska[1,2*], **Dorothee Ehrich**[2], **Sunil Mundra**[1,3,4], **Anna Vader**[1], **Pernille B. Eidesen**[1]

[1] Department of Arctic Biology, The University Centre in Svalbard, Longyearbyen, Norway
[2] Department of Arctic and Marine Biology, UiT – The Arctic University of Norway, Tromsø, Norway
[3] Department of Biosciences, University of Oslo, Oslo, Norway
[4] Department of Biology, College of Science, United Arab Emirates University, Abu Dhabi, United Arab Emirates
*corresponding author: magda.wutkowska@gmail.com

ABSTRACT
Arctic plants are affected by many stressors. Root-associated fungi are thought to influence plant performance in stressful environmental conditions. However, the relationships are not transparent; do the number of fungal partners, their ecological functions and community composition mediate the impact of environmental conditions and/or influence host plant performance? To address these questions, we used a common arctic plant as a model system: *Bistorta vivipara*. Whole plants (including root system) were collected from nine locations in Spitsbergen (n=214). Morphometric features were measured as a proxy for performance and combined with metabarcoding datasets of their root-associated fungi (amplicon sequence variants, ASVs), edaphic and meteorological variables. Seven biological hypotheses regarding fungal influence on plant measures were tested using structural equation modelling. The best-fitting model revealed that local temperature affected plants both directly (negatively aboveground and positively below-ground) and indirectly - mediated by fungal richness and the ratio of symbio- and saprotrophic ASVs. Fungal community composition did not impact plant measurements and plant reproductive investment did not depend on any fungal parameters. The lack of impact of fungal community composition on plant performance suggests that the functional importance of fungi is more important than their identity. The influence of temperature on host plants is therefore complex and should be examined further.

KEY WORDS
plant-microbe interaction, plant performance, root-associated fungi, arctic soil biology, below-ground vegetation

Introduction

Arctic plants are facing many environmental constraints for growth, such as short vegetation season, consistent cold, limitation of nutrients or cyclic physical disturbances, i.e. cryoturbation[1]. These plants have evolved a range of adaptations to cope with the prevailing conditions, including being perennial and allocating most of their biomass below-ground[2–4]. Being perennial provides a resource-saving advantage in nutrient-poor habitats with low temperatures that slow down biochemical reactions and therefore also growth, whereas the benefits of biomass allocation to below ground parts include increased area of nutrient absorption. Because of nutrient scarcity, the interface between plant and soil is of relatively greater importance in the Arctic than in other biomes[3]. A significant part of the soil-plant interface is inhabited by microbes, including roots-associated fungi (RAF). Arctic RAF consist mostly of symbiotrophic fungi, especially ectomycorrhizal fungi[5–7]. These fungi efficiently increase the volume of soil that can be penetrated in search for resources, such as nutrients from seasonally or newly thawed permafrost[8]. The most severe limitations for growth observed in arctic plants are due to low temperatures and resource limitation[1,9], suggesting that the relationship with RAF might play a crucial role in plant survival and growth.

Multiple characteristics of species communities play an essential role in the functioning of ecosystems, such as richness, abundance or community structure[10–13]. Based on previous findings, we may expect that the more diverse the community of RAF, the better for a host plant[14]. However, it is not clear how these characteristics of RAF communities impact their host plants, especially in cold biomes. Symbiotic fungi provide resources and probably additional benefits mitigating possibly harmful effects of environmental stressors enhancing plant growth and productivity[15]. However, releasing root exudates of primary metabolites that can be absorbed by members of its microbiome does come with a cost for a plant[16,17]. In nitrogen-limited tundra in Alaska, 61-88% of plant nitrogen (N) was supplied from mycorrhizal fungi. In exchange, the plants delivered 8-17% of carbon (C) produced photosynthetically to the fungi[18]. A plant could perhaps increase the amount of released nutritious root exudates to attract more species of symbiotrophic fungi that in turn, could potentially increase the amount of nitrogen delivered. However, higher fungal richness would increase competition for limited space in the rhizosphere and possibly for resources, although the mechanism is not yet fully described. Therefore, plants ‘living on the edge’ in the High Arctic may benefit from the selective choice of their members of RAF communities,

favouring the most beneficial fungal partners for plant growth or mediation of stressors[19]. In this scenario, species richness in RAF communities would be irrelevant for plant performance. The presence of specific functional traits rather than their identity could be more important[20]. The vast array of interconnected biotic and abiotic factors occurring in natural systems complicate uncovering if and how plants show preference among their root-associated fungi among the pool of species present in the soil[21].

One approach to disentangle these often confounded factors are controlled experiments. Most of the experiments assessing the impact of RAF diversity on host plant performance have focused on arbuscular mycorrhiza in crops[22]; whereas similar studies on ectomycorrhizal (EcM) plant species come mostly from the pre-high throughput sequencing era and have focussed on trees (e.g. [14]). Several experiments under controlled settings have shown that EcM host plants may clearly benefit from their increased fungal richness, however, the tested level of richness was often incomparable with natural environments, such as an increase from 1 to 4 species of EcM fungi[14]. Some studies, however, did not find any enhancements in plant performance mediated by EcM fungi or concluded that the outcome of EcM species richness on plant productivity is context dependent[23]. RAF diversity was shown to be particularly sensitive to experimental conditions compared to fungi that inhabit space further from the roots in the rhizosphere or bulk soil[24]. Moreover, morphology and physiology of lab-grown plants differ from those in the natural systems, e.g. by increasing growth rate and higher concentrations of nutrients in tissues[25]. All these differences could affect and alter plant-associated organisms, such as RAF. Experimental procedures cannot consider all the complexity of natural systems and their effects do not always reflect those observed in the wild. Thus, observational studies can provide crucial complementary knowledge, in particular for extreme environments like the high Arctic.

Species response to environmental shifts, including ongoing climate changes, is one of the crucial questions in natural sciences. It is a particularly outstanding issue in the Arctic where rates of temperature and precipitation are changing at the fastest pace in the world, and are predicted to continue rising rapidly[26,27]. These changes impact mechanisms that alter biogeochemical cycles and determine critical ecosystem-climate feedback processes, such as the release of organic carbon of which nearly half of the global stock is stored in the Arctic soils[28,29] or increased growth of vascular plants. Such ecosystem feedbacks, which are essential bricks in the understanding of global change, depend on complex relationships

between abiotic and biotic factors in arctic soils[30]. However, the biology of these soils remains at present an understudied 'black box'.

To shed some light onto these soil processes, we used a plant-centric approach to study the impact of the root-associated fungal community on the growth and reproductive investment of a wide-spread arctic plant, *Bistorta vivipara*. We took into account the most important abiotic factors which likely affect the host plant and its RAF community. We used structural equation modelling (SEM) to assess whether the fungal community mediates the effect of abiotic conditions on plant performance and to disentangle direct from indirect effects. We tested the following hypotheses: (i) Plant morphological measurements (considered as a proxy for plant performance) depend both on abiotic conditions and on the fungi community, and (ii) only richness and functional traits, but not the specific species composition of the RAF community affects plant morphology. Moreover, we tested, which measurements of plant parts involved in different processes such as energy storage, energy acquisition and reproduction depended on the RAF community.

Methods

Study system

To test our hypotheses, we selected alpine bistort *Bistorta vivipara* (L.) Delarbre (*Polygonaceae*), a model plant to study root-associated microbial communities in alpine[6,31–35] and arctic habitats[7,19,36–40]. *Bistorta vivipara* is a common, long-lived perennial herb in the northern hemisphere. Its compact root system, combined with the ability to inhabit a range of habitats, makes this species a perfect candidate to study root-associated communities concerning environmental gradients, such as chronosequences[6,38,39] or climate gradients[37].

Datasets

We combined and reanalysed datasets spanning over nine different locations in Spitsbergen, the largest island of the high-arctic archipelago Svalbard, Norway (Table 1; Figure 1). Each dataset consisted of host morphology, molecular descriptions of the RAF community, together with associated edaphic variables (Table 1). Each of the studies established a randomized sampling scheme in the locality of choice, also assuring that

sampled plants are of different age. Whole plants with an intact root system were excavated. To explore the associations between plant performance, allocation patterns and its environment we measured three morphological features of the *B. vivipara* individuals hosting the analysed RAF communities (Supplementary 1): The rhizome is an underground storage organ that accumulates assimilated biomass as nonstructural carbohydrates, therefore here we used it as a proxy for overall plant performance[41]. Rhizome dimensions were measured and used to calculate an approximate volume (RV) by multiplying its length, height and width. Length of the longest stem leaf (LL) was used as a proxy for photosynthetic capabilities of the plant – the longer the leaf, the bigger photosynthetic area. In the upper part of the stem, *B. vivipara* produces flowers and bulbils for sexual and asexual reproduction, respectively. We used the ratio of the length of the stem covered by flowers and bulbils (inflorescence) to the total stem length (I/S), as a proxy for the plant's investment in reproduction.

Meteorological and edaphic variables

Meteorological data were obtained for each sampling point from the high-resolution 1 km-gridded dataset Sval_Imp_v1[42]. We extracted the sum of average monthly precipitation (p) and average July air temperature (t), both from the year of sampling.

Soil samples were collected from the same sampling spot as plants. The following edaphic parameters, representing critical properties of the abiotic environment, were measured in all datasets: pH, soil nitrogen concentration (N) and carbon to nitrogen ratio (C/N; used as an indicator for soil nitrogen availability or soil fertility). Edaphic variables were obtained in the same way for all datasets (described in detail in [7,19,40]).

Fungal data

Bistorta vivipara roots were cleaned within a day from sampling and fixed in a 2% CTAB extraction buffer until DNA extraction (details described in each of the publications; Table 1). All datasets targeted the same fragment of internal transcribed spacer 2 amplified with fITS7a forward primer[43] and reverse primer ITS4[44] and sequenced with Illumina MiSeq (300bp paired-end reads).

Each dataset was a mixture of sequences located in 'forward' and 'reverse' direction. Thus, first, a mapping file with variable length barcodes and primer sequences was used to identify sequences in each location using sabre (https://github.com/najoshi/sabre) and generating separate R1 and R2 files for each read direction. Next, primers were clipped, and sequences

with ambiguous bases (Ns) were removed using cutadapt v. 2.5[45]. Python script FastqCombinePairedEnd.py (https://github.com/enormandeau/Scripts) was used to assure that each sequence had its pair and were in the matching order for further analyses. We used an amplicon sequence variants (ASVs) approach implemented in DADA2 v. 1.11.1[46] and executed in R v. 3.5.2[47] (for details see Supplement 2 and scripts generated for this study). The datasets were analysed using DADA2 ITS workflow (https://benjjneb.github.io/dada2/ITS_workflow.html). Fungal data were produced independently for each study; therefore, they were initially analysed separately due to different error rates for each sequencing run. Separate ASVs tables were then merged. Consensus method was used to remove chimaeras (3759 out of 11243 input sequences). Sequences shorter than 200bp and six samples with a very low number of reads were removed. Due to profound differences in depth of sequencing the ASV table was randomly subsampled (21639 reads per sample; number of detected ASVs before and after subsampling was highly correlated; Kendall's $\tau = 0.95$). Taxonomy was assigned using the RDP naive Bayesian classifier implemented in DADA2 with a full UNITE+INSD reference dataset for fungi[48] (sh_general_release_dynamic_02.02.2019). All the ASVs were functionally annotated using the FUNGuild database[49].

Differences in community composition were summarized through non-metric multidimensional scaling (GNMDS; *vegan* package[50]), and we used the first axis as a proxy for composition in further analyses. We used both presence-absence based metrics and parameters based on read abundance to describe RAF communities: ASV diversity (D), a ratio of symbio- to saprotrophs (Sy/Sa) and GNMDS values for 1st axis as a proxy for community composition (CC; Table 2).

Statistical analyses and model selection

The statistical analyses were performed in R v. 3.5.2[47]. Based on available literature of soil and weather influence on fungi and plant interactions in the Arctic (Table 3), we built seven hypothetical causal path models relating abiotic variables to the three metrics characterizing the fungal community and plant morphological measurements (solid lines in Figure 2). The unbranched rhizome of *B. vivipara* elongates with age, providing space for new roots to stem from its distal end[51] and therefore increasing the richness of recruited RAF[34]. Randomised sampling schemes in each of the studies included in our study excluded the potential influence of plant age on the results. For the full model, we assumed that all three fungal parameters influence all three plant measures, additionally to abiotic factors impacting both fungal and plant variables.

Next, we hypothesized that fungi might not be essential for specific plant measurements. Therefore, in the three subsequent models, we preserved all the relationships omitting only the fungal variables in a specific plant response (I/S, RV or LL does not depend on fungi). In the next models, we, therefore, hypothesized that CC is not an important parameter for any of the plant measurements. Additionally, we combined this last model with the best model obtained from simplifying the relationships between fungi and plants responses.

Finally, to evaluate whether fungal parameters have any impact on plant measurements, we removed all connections between fungal parameters and plant measurements. In the models, we treated edaphic and meteorological variables as independent. We are aware that they can affect each other, but this was not the focus of the study. The most considerable correlation among them was between N and C/N ($r = -0.64$). We did also not hypothesize any causal links between the fungal parameters. Concerning the plant variables, we assumed a causal link between rhizome volume and leaf length, because leaf growth in the start of the season depends on stored resources. Locality was used as a random effect in all the models because fungal community composition usually shows a high spatial variation (e.g.[36]) and because preliminary ordinations showed that in our dataset fungal communities differed between localities.

We applied structural equation modelling (SEM) to carry out an exploratory path analysis of these models, using the psem function in the piecewiseSEM package[52]. The SEM was composed of linear mixed-effects models (LMMs) for each fungi parameter and plant measurement, which was fitted using the lme function in *nlme* package[53]. The fit of the separate LMMs were assessed graphically for normality of the residuals. Residuals clearly deviating from the expected distribution on a quantile-quantile plot with standardised residuals > |3| were considered as outliers and therefore excluded.

The analysis was performed using both presence-absence based and read abundance metrics for the fungal community. Because some of the fungal parameters were correlated, we included non-directed correlations among them in the SEM to make it possible to estimate the paths in our exploratory model. It was the case for CC and Sy/Sa based on presence-absence and for Sy/Sa and D based on read abundance. The distributions of all variables were assessed graphically, and some were log- or logit-transformed to assure roughly normal distributions. All variables were scaled to 0 mean and a standard deviation of 1 to make effect sizes comparable.

A prerequisite for a SEM model to be considered as fitting was Fisher's C p-value > 0.05[54]. The best models among the candidate sets described above were chosen based on the lowest AIC values. Both of these values were calculated within the *psem* function. We used

statistically significant estimates from the best fitting presence-absence model to calculate indirect effects of abiotic factors on plant measures.

The combined dataset consisted of 214 *B. vivipara* plant measurements with associated edaphic data and corresponding RAF data. For the SEM, we excluded all observations with missing values resulting in a final dataset with 188 plants (after excluding outliers presence-absence dataset had 187 and abundance dataset 185 values).

Results

Models based on presence-absence fungal parameters

The best-fitting presence-absence path model (AIC_{min} = 117.97; Table 4) supported the hypothesis that fungal CC does not impact plant measurements, and simultaneously no fungal parameters affect the I/S. The second best-fitting model with a relative difference $\Delta AIC < 1$, supported a related hypothesis that I/S does not depend on any fungal parameters included in this study, but included the effect of CC on other plant parameters.

In the best-fitting and most parsimonious model, fungal community richness and the ratio of symbiotrophic to saprotrophic species were related to plant measurements as follows (Figure 3a): fungal richness with RV (positive path coefficient (PC ± SE = 0.26 ± 0.07, $p < 0.001$); full list of all the effect sizes in Supplement 4a) and Sy/Sa with LL (PC ± SE = -0.20 ± 0.07, $p = 0.004$). Except for the fungal metrics, the RV also showed positive correlations with p (PC ± SE = 0.29 ± 0.11, $p = 0.01$). LL was negatively impacted by N content (PC ± SE = -0.20 ± 0.08, $p = 0.02$) and t (PC ± SE = -0.34 ± 0.08, $p < 0.001$). The highest estimate in our model suggested correlation between RV and LL (PC ± SE = 0.53 ± 0.06, $p < 0.001$).

Meteorological data had a clear effect on fungal parameters: p with Sy/Sa (PC ± SE = 0.44 ± 0.21, $p < 0.04$), and t with fungal CC (PC ± SE = 0.27 ± 0.09, $p = 0.003$) and D (PC ± SE = -0.45 ± 0.13, $p < 0.001$). Based on the best fitting presence-absence model, edaphic variables did not seem to impact any of fungal parameters and plant measurements except the already mentioned N content impact on LL. On the other hand, t correlated with multiple fungal and plant variables.

Among abiotic factors impacting plant measurements, t affected LL over three pathways: direct (negative, PC = -0.34) and two indirect: positive through RV (PC = 0.29 * 0.53 = 0.154) and negative through fungal D (PC = -0.45 * 0.26 = -0.117). The direct effect was therefore the strongest and the two indirect effects were of comparable magnitude, but opposite directions.

Abundance model

The best-fitting path model based on read abundance supported the hypothesis that fungal parameters do not impact any plant measurements (AIC_{min} = 119.28; Table 4). Another model that differed by ΔAIC = 0.25 supported the same hypothesis as the best fitting presence-absence model: fungal CC does not impact plant measurements, and I/S is not affected by other fungal parameters either.

Although the role of fungi in the best model differs fundamentally from the best model based on presence-absence ASV table, they both preserved some of the same statistically significant relationships between environmental variables and plant measurements (Figure 3b, full list of all the effect sizes in from both types of models in Supplement 4). This included correlations between N content and LL (PC ± SE = -0.23 ± 0.08, p = 0.005), t and LL (PC ± SE = -0.37 ± 0.08, p < 0.001), as well as t and CC (PC ± SE = 0.31 ± 0.09, p < 0.001). Also, the relationship between two plant variables, RV and LL, showed the same magnitude as in the best fitting presence-absence model (PC ± SE = 0.54 ± 0.06, p < 0.001). This model supported no indirect effects of abiotic factors mediated by fungal parameters.
The abundance-based model revealed links between edaphic and fungal parameters that were not statistically significant in the presence-absence model. N content and C/N ratio correlated negatively with Sy/Sa (PC ± SE = -0.28 ± 0.10, p = 0.007 and PC ± SE = -0.20 ± 0.10, p < 0.04; respectively). The N content positively impacted fungal diversity (PC ± SE = 0.24 ± 0.11, p < 0.04).

Variance in fungal and plant response variables

In both best fitting models, the variance in plant measurements was on average better explained by fixed factors than the variance in fungal parameters (marginal R^2 = 0.02-0.44 vs 0.07-0.26, Table 5). However, overall the variance explained by fixed factors was rather low.

On the contrary, locality included as a random factor explained on average more variation in fungi than in plants (conditional R^2 - marginal R^2 = 0.03 - 0.58 and 0.01 - 0.33, respectively). The high proportion of variance explained for fungal response variables was especially pronounced in presence-absence compared to the abundance model (conditional R^2 - marginal R^2 = 0.40 - 0.58 and 0.03-0.48, respectively).

Discussion

Establishing functional relationships between biological components, such as a host plant and its root-associated microbiome, taking into account abiotic drivers, could enhance the current understanding of soil carbon pools and decrease associated uncertainties[55,56]. To narrow these gaps, we studied the common arctic host plant *B. vivipara* and its RAF communities in connection with their environment. Here, we linked above- and below-ground plant measurements to fungal parameters, all assumed to be influenced by the same edaphic and meteorological conditions. This exploratory study revealed that measurements of below- and aboveground plant organs responded in opposite ways to temperature, the effects of which were both direct and mediated by parameters of the RAF community. Regarding fungal parameters, both species richness and functional diversity were important for plant performance measurements, but not the specific community composition.

Our study revealed that among the abiotic factors temperature was the most important for biotic elements, which reflects its immense significance in physical constraints for arctic biota[1] and the general tendency of modifying interactions between organisms[57]. However, our results also suggest that the impact of temperature on an arctic host plant is far more complex than previously thought[58,59] and in general, perhaps unpredictable[60]. The mechanism behind fungal mediation of temperature is not clear. Here we looked only into a few parameters associated with RAF communities that impacted the plant both positively and negatively balancing themselves out. However, there are other molecular and physiological characteristics that could explain the influence of fungi on plant performance mechanistically. For instance, secretion of fungal signalling molecules, such as volatile organic compounds[61] or plant-like hormones[62,63], that can be translocated to host plant cells and there elicit a physiological response. Release of these molecules could be temperature-dependent. Similarly, plant-based responses to these signals could also be at least partly temperature-dependent, e.g. release of root exudates[64].

Different influences of temperature on below- and aboveground plant measurements could question current methods of monitoring changes in arctic vegetation, such as the normalized difference vegetation index (NDVI) used as a proxy for plant biomass. This technology advanced the understanding of vegetation biomass dynamics simultaneously over vast and otherwise under-sampled areas of the Arctic (e.g.[65–67]). However, it is based on remote measurements of Earth's surface reflectance, and therefore takes into consideration only aboveground changes in foliage. In these methods plants' below-ground productivity and biomass are omitted, probably resulting in underestimation of the overall impact of increased temperatures on plants, such as *B. vivipara*, which is an ubiquitous species in the Arctic and essential food source for ptarmigans[68], geese[69] and reindeer[70]. Temperature had a direct opposite effect of similar magnitude on LL and RV (-0.34 vs 0.29, respectively), additionally strengthened by indirect fungal effects, which suggests that NDVI can easily underestimate the impact of warming on overall plant biomass and misjudge understanding of carbon stocks dynamics. Presently, there are no tools that could be used to scan below-ground plant biomass at scales similar to NDVI. However, there are some more laborious *in situ* methods, e.g. minirhizotrons, that are used to measure below-ground biomass[71]. Their use significantly enhances our understanding of the dynamics in belowground biomass allocation. Nevertheless, the implications of temperature affecting a host plant through multiple pathways generate major difficulties in projections of the future response of ecosystems to warming.

Negative impact of nitrogen on leaf length was unexpected in the light of previous findings[72]. *Bistorta vivipara* is regarded as a pioneer plant[73], able to cope with severe conditions and resource limitations[32,39]. In a High Arctic nitrogen-rich habitat, such as bird cliffs, where the competition between organisms is high, it is most likely outcompeted by other plants. Additionally, these highly nutritious habitats are characterised by an increased number of plant interactions with herbivores, such as reindeers, that can eliminate foliage.

Almost all symbiotrophic RAF of *B. vivipara* in Svalbard are ectomycorrhizal[7,39]. Since these fungi exchange nitrogen with plants in return for versatile carbon metabolites[18], we hypothesized that in a resource-limiting environment this fungal trophic mode could promote bigger plants[74], therefore bigger leaves. This way, fungi could potentially influence the number and amount of metabolites that the plant could produce in return and share in its rhizosphere. However, our results showed the opposite scenario, where Sy/Sa had a

negative effect on leaf length, which suggests that more fungal partners enhance competition over resources that are scarce[75]. The richness of symbio- and saprotrophs taken into account separately did not show any associations with plant measurements (data not shown); however, the ratio of their richness did, perhaps reflecting the characteristics of soil conditions in different localities. Particularly small ratio of Sy/Sa was found in localities with little organic matter (Supplementary 3), suggesting that this parameter mirrors fertility properties of soil. When soil organic matter content is low, then colonizing plant roots ensures access to an easily accessible pool of carbon from root exudates[76]. Although *B. vivipara* root system is relatively compact and flexible, growing in mineral soils, including some stages of soil development of glacier forefronts[77], could promote longer roots to assure access to quickly drained soil water. Intense disturbance caused by periglacial processes in these habitats may contribute to physical breaks in fine roots or associated fungal mycelium, perhaps leading to an increase in the number of saprotrophic species. Alternatively, saprotrophic fungi could be one of the first organisms in primary community assembly using organic carbon from previously unrecognized heterotrophic communities of invertebrates which feed on allochthonous organic matter now recognized as a crucial step of primary succession before establishment of autotrophs[78–80].

Our finding that fungal community composition did not affect plant measurements could perhaps originate from strong environmental filtering on root-associated fungal communities[36]. High physicochemical heterogeneity of arctic soils corresponds with distinct RAF community composition observed at different scales[5–7,37]. On the one hand, a set of physicochemical conditions that translates into ecological niches selects species that can withstand and thrive in these locality-specific combinations of factors. Among them principally abiotic factors were shown to affect fungal parameters [37,81–83]. Relationships between variables established based on the literature search (Table 3) were, in general, poorly reflected in the results of our models. In most cases, we saw no effects of abiotic drivers identified in the literature on neither plants nor fungi. It was especially pronounced in RAF community composition, suggesting other sources of the differences that are specifically connected to locality[19]. These could be other edaphic factors not included in this study (e.g. phosphorus[84] or heavy metal concentrations[85], competition[75,86] or other factors that historically impacted the community assembly[87]. Nevertheless, the fact that arctic ectomycorrhizal RAF display little or no affinity to host species[88] suggests that the fungal contribution to plants reflects mitigation of effects of locality-specific conditions, rather than individual species needs. Similar conclusions were made in edge soil habitats beyond the

Arctic. For instance, RAF communities in soil characterised by combined effects of poor nutritional and water status[89] or high contamination levels[90] seem to also be host-independent and highly variable among the sites.

To explain discrepancies in results between presence-absence and read abundance models, it is necessary to identify possible sources of variation in read abundances in fungal metabarcoding studies. Fungal species vary in the copy number of ribosomal DNA (14-1442), and this number is independent of genome size or ecological roles, such as guild or trophic mode[91]. Strains of the same fungal species, especially yeast, can exhibit high variation of rDNA copy number[92,93]. Relative abundances of reads are sometimes used as a proxy for the relative biomass contributions of some species[94]. However, a quantitative meta-analysis found only a weak relationship between the two[95]. Read abundance can be profoundly affected by methodological biases at several steps during metabarcoding procedures, starting from the choice of primers through wet-lab methods, including sequencing, to bioinformatic pipelines [96–99]. However, in our study, main pathways affecting plants directly and not through fungal parameters remained present in both best-fitting models. This supports prevalence of a biological signal over methodological biases from abundance data. On the other hand, the abundance-based model in this study showed clear links between fungal parameters and soil fertility (N and C/N) mirroring the stoichiometric state of the environment[100] and temperature that controls the rate of biochemical reactions.

Here we demonstrated that fungal parameters, such as richness and functional diversity, could mediate the influence of abiotic factors on host plants, but it is not clear what are the mechanisms behind this. It is not clear how different fungi contribute to plants' biometrics, how many resources are being exchanged with plants and how that changes with RAF variation in time and space. Not only molecular identification, but also establishing biomass estimations for both fungi and bacteria could help to understand below-ground dynamics. Low proportion of variance explained by fixed factors showed that there is a strong need to obtain and include more abiotic and biotic variables that were not considered in this study, but are of high importance for fungi and plants. Controlled experiments could potentially help to address these uncertainties. Additionally, morphological characterization of multiple plant species, biomass and nutrient concentration measurements in separate plant parts would ensure precise comparisons between plant life strategies in variable habitats and distant locations. Another critical aspect in making these links is to include the host plant genotype to tie its phenotype with the influence of the environment accurately[101]. A comprehensive

interdisciplinary study employing various methods could help to develop a mechanistic understanding of links between above- and below-ground biota, including other taxonomic groups.

References

1. Billings, W. D. Constraints to Plant Growth, Reproduction, and Establishment in Arctic Environments. *Arct. Alp. Res.* **19**, 357–365 (1987).
2. Iversen, C. M. *et al.* The unseen iceberg: plant roots in arctic tundra. *New Phytol.* **205**, 34–58 (2015).
3. Poorter, H. *et al.* Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control: Tansley review. *New Phytol.* **193**, 30–50 (2012).
4. Qi, Y., Wei, W., Chen, C. & Chen, L. Plant root-shoot biomass allocation over diverse biomes: A global synthesis. *Glob. Ecol. Conserv.* **18**, e00606 (2019).
5. Bjorbækmo, M. *et al.* High diversity of root associated fungi in both alpine and arctic Dryas octopetala. *BMC Plant Biol.* **10**, 244 (2010).
6. Blaalid, R. *et al.* Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing: PRIMARY SUCCESSION OF ROOT-ASSOCIATED FUNGI. *Mol. Ecol.* **21**, 1897–1908 (2012).
7. Mundra, S. *et al.* Arctic fungal communities associated with roots of Bistorta vivipara do not respond to the same fine-scale edaphic gradients as the aboveground vegetation. *New Phytol.* **205**, 1587–1597 (2015).
8. Hewitt, R. E. *et al.* Mycobiont contribution to tundra plant acquisition of permafrost-derived nitrogen. *New Phytol.* **226**, 126–141 (2020).
9. Chapin, F. S. & Shaver, G. R. Individualistic Growth Response of Tundra Plant Species to Environmental Manipulations in the Field. *Ecology* **66**, 564–576 (1985).
10. Loreau, M. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science* **294**, 804–808 (2001).
11. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
12. Maestre, F. T., Castillo-Monroy, A. P., Bowker, M. A. & Ochoa-Hueso, R. Species richness effects on ecosystem multifunctionality depend on evenness, composition and spatial pattern. *J. Ecol.* **100**, 317–330 (2012).
13. Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proc. Natl. Acad. Sci.* **109**, 10394–10397 (2012).
14. Baxter, J. W. & Dighton, J. Ectomycorrhizal diversity alters growth and nutrient

acquisition of grey birch (Betula populifolia) seedlings in host-symbiont culture conditions. *New Phytol.* **152**, 139–149 (2001).

15. Wardle, D. A. *et al.* Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629–1633 (2004).

16. Brzostek, E. R., Fisher, J. B. & Phillips, R. P. Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve predictions of retranslocation: Carbon cost of mycorrhizae. *J. Geophys. Res. Biogeosciences* **119**, 1684–1697 (2014).

17. Shi, Y. *et al.* Vegetation-Associated Impacts on Arctic Tundra Bacterial and Microeukaryotic Communities. *Appl. Environ. Microbiol.* **81**, 492–501 (2015).

18. Hobbie, J. E. & Hobbie, E. A. 15N in Symbiotic Fungi and Plants Estimates Nitrogen and Carbon Flux Rates in Arctic Tundra. *Ecology* **87**, 816–822 (2006).

19. Mundra, S., Bahram, M. & Eidesen, P. B. Alpine bistort (Bistorta vivipara) in edge habitat associates with fewer but distinct ectomycorrhizal fungal species: a comparative study of three contrasting soil environments in Svalbard. *Mycorrhiza* **26**, 809–818 (2016).

20. Louca, S. *et al.* High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* **1**, 0015 (2017).

21. Jones, P., Garcia, B. J., Furches, A., Tuskan, G. A. & Jacobson, D. Plant Host-Associated Mechanisms for Microbial Selection. *Front. Plant Sci.* **10**, 862 (2019).

22. Begum, N. *et al.* Role of Arbuscular Mycorrhizal Fungi in Plant Growth Regulation: Implications in Abiotic Stress Tolerance. *Front. Plant Sci.* **10**, (2019).

23. Jonsson, L. M., Nilsson, M.-C., Wardle, D. A. & Zackrisson, O. Context Dependent Effects of Ectomycorrhizal Species Richness on Tree Seedling Productivity. *Oikos* **93**, 353–364 (2001).

24. Almario, J. *et al.* Root-associated fungal microbiota of nonmycorrhizal *Arabis alpina* and its contribution to plant phosphorus nutrition. *Proc. Natl. Acad. Sci.* **114**, E9403–E9412 (2017).

25. Poorter, H. *et al.* Pampered inside, pestered outside? Differences and similarities between plants growing in controlled conditions and in the field. *New Phytol.* **212**, 838–855 (2016).

26. Bintanja, R. & Selten, F. M. Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat. *Nature* **509**, 479–482 (2014).

27. Post, E. *et al.* The polar regions in a 2°C warmer world. *Sci. Adv.* **5**, eaaw9883 (2019).

28. Schuur, E. A. G. *et al.* Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015).

29. Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region: SOIL ORGANIC CARBON POOLS. *Glob. Biogeochem. Cycles* **23**, n/a-n/a (2009).

30. Wookey, P. A. *et al.* Ecosystem feedbacks and cascade processes: understanding their role in the responses of Arctic and alpine ecosystems to environmental change. *Glob.*

Change Biol. **15**, 1153–1172 (2009).

31. Gao, Q. & Yang, Z. L. Diversity and distribution patterns of root-associated fungi on herbaceous plants in alpine meadows of southwestern China. *Mycologia* **108**, 281–291 (2016).

32. Mühlmann, O., Bacher, M. & Peintner, U. Polygonum viviparum mycobionts on an alpine primary successional glacier forefront. *Mycorrhiza* **18**, 87–95 (2008).

33. Ronikier, M. & Mleczko, P. Observations on the mycorrhizal status of Polygonum viviparum in the Polish Tatra Mts. (Western Carpathians). *Acta Mycol.* **41**, 209–222 (2013).

34. Thoen, E. *et al.* A single ectomycorrhizal plant root system includes a diverse and spatially structured fungal community. *Mycorrhiza* **29**, 167–180 (2019).

35. Vik, U. *et al.* Different bacterial communities in ectomycorrhizae and surrounding soil. *Sci. Rep.* **3**, 3471 (2013).

36. Blaalid, R. *et al.* Arctic root-associated fungal community composition reflects environmental filtering. *Mol. Ecol.* **23**, 649–659 (2014).

37. Botnen, S. S. *et al.* Biogeography of plant root-associated fungal communities in the North Atlantic region mirrors climatic variability. *J. Biogeogr.* jbi.13613 (2019) doi:10.1111/jbi.13613.

38. Brevik, A. *et al.* Diversity of fungi associated with Bistorta vivipara (L.) Delarbre root systems along a local chronosequence on Svalbard. *Agarica* **29**, 15–26 (2010).

39. Davey, M. *et al.* Primary succession of Bistorta vivipara (L.) Delabre (Polygonaceae) root-associated fungi mirrors plant succession in two glacial chronosequences: Fungi and plants share successional trajectories. *Environ. Microbiol.* **17**, 2777–2790 (2015).

40. Mundra, S. *et al.* Temporal variation of Bistorta vivipara-associated ectomycorrhizal fungal communities in the High Arctic. *Mol. Ecol.* **24**, 6289–6302 (2015).

41. Hartmann, H. & Trumbore, S. Understanding the roles of nonstructural carbohydrates in forest trees – from what we can measure to what we want to know. *New Phytol.* **211**, 386–403 (2016).

42. Schuler, T. V. & Østby, T. I. *Sval_Imp_v1: A gridded forcing dataset for climate change impact research on Svalbard*. https://www.earth-syst-sci-data-discuss.net/essd-2019-180/ (2020) doi:10.5194/essd-2019-180.

43. Ihrmark, K. *et al.* New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **82**, 666–677 (2012).

44. White, T. *et al.* Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications. (1990).

45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

46. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon

data. *Nat. Methods* **13**, 581–583 (2016).

47. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2018).

48. UNITE Community. Full UNITE+INSD dataset for Fungi. (2019) doi:10.15156/BIO/786347.

49. Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).

50. Oksanen, J. *et al. vegan: Community Ecology Package*. (2019).

51. Diggle, P. K. Extreme Preformation in Alpine Polygonum viviparum: An Architectural and Developmental Analysis. *Am. J. Bot.* **84**, 154–169 (1997).

52. Lefcheck, J. S. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).

53. José Pinheiro (S *et al. nlme: Linear and Nonlinear Mixed Effects Models*. (2020).

54. Shipley, B. The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology* **94**, 560–564 (2013).

55. Zha, J. & Zhuang, Q. Microbial decomposition processes and vulnerable arctic soil organic carbon in the 21st century. *Biogeosciences* **15**, 5621–5634 (2018).

56. Wieder, W. R., Sulman, B. N., Hartman, M. D., Koven, C. D. & Bradford, M. A. Arctic Soil Governs Whether Climate Change Drives Global Losses or Gains in Soil Carbon. *Geophys. Res. Lett.* **46**, 14486–14495 (2019).

57. Bideault, A., Loreau, M. & Gravel, D. Temperature Modifies Consumer-Resource Interaction Strength Through Its Effects on Biological Rates and Body Mass. *Front. Ecol. Evol.* **7**, 45 (2019).

58. Chapin, F. S. Direct and indirect effects of temperature on arctic plants. *Polar Biol.* **2**, 47–52 (1983).

59. De Long, J. R., Kardol, P., Sundqvist, M. K., Veen, G. F. C. & Wardle, D. A. Plant growth response to direct and indirect temperature effects varies by vegetation type and elevation in a subarctic tundra. *Oikos* **124**, 772–783 (2015).

60. De Long, J. R., Fry, E. L., Veen, G. F. & Kardol, P. Why are plant–soil feedbacks so unpredictable, and what to do about it? *Funct. Ecol.* **33**, 118–128 (2019).

61. Schenkel, D., Maciá-Vicente, J. G., Bissell, A. & Splivallo, R. Fungi Indirectly Affect Plant Root Architecture by Modulating Soil Volatile Organic Compounds. *Front. Microbiol.* **9**, 1847 (2018).

62. Han, X. & Kahmann, R. Manipulation of Phytohormone Pathways by Effectors of Filamentous Plant Pathogens. *Front. Plant Sci.* **10**, 822 (2019).

63. Chanclud, E. & Morel, J.-B. Plant hormones: a fungal point of view: Hormones from fungi. *Mol. Plant Pathol.* **17**, 1289–1297 (2016).

64. Canarini, A., Kaiser, C., Merchant, A., Richter, A. & Wanek, W. Root Exudation of

Primary Metabolites: Mechanisms and Their Roles in Plant Responses to Environmental Stimuli. *Front. Plant Sci.* **10**, 157 (2019).

65. Myers-Smith, I. H. *et al.* Complexity revealed in the greening of the Arctic. *Nat. Clim. Change* **10**, 106–117 (2020).

66. Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).

67. Phoenix, G. K. & Bjerke, J. W. Arctic browning: extreme events and trends reversing arctic greening. *Glob. Change Biol.* **22**, 2960–2962 (2016).

68. Steen, J. B. & Unander, S. Breeding Biology of the Svalbard Rock Ptarmigan Lagopus mutus hyperboreus. *Ornis Scand. Scand. J. Ornithol.* **16**, 191–197 (1985).

69. Anderson, H. B., Godfrey, T. G., Woodin, S. J. & Wal, R. van der. Finding food in a highly seasonal landscape: where and how pink footed geese Anser brachyrhynchus forage during the Arctic spring. *J. Avian Biol.* **43**, 415–422 (2012).

70. Bjørkvoll, E., Pedersen, B., Hytteborn, H., Jónsdóttir, I. S. & Langvatn, R. Seasonal and Interannual Dietary Variation During Winter in Female Svalbard Reindeer (Rangifer Tarandus Platyrhynchus). *Arct. Antarct. Alp. Res.* **41**, 88–96 (2009).

71. Wilson, S. D. Below-ground opportunities in vegetation science. *J. Veg. Sci.* **25**, 1117–1125 (2014).

72. Wookey, P. A. *et al.* Differential Growth, Allocation and Photosynthetic Responses of Polygonum viviparum to Simulated Environmental Change at a High Arctic Polar Semi-Desert. *Oikos* **70**, 131–139 (1994).

73. Dormann, C. F., Albon, S. D. & Woodin, S. J. No evidence for adaptation of two Polygonum viviparum morphotypes of different bulbil characteristics to length of growing season: abundance, biomass and germination. *Polar Biol.* **25**, 884–890 (2002).

74. Tedersoo, L., Bahram, M. & Zobel, M. How mycorrhizal associations drive plant population and community biology. *Science* **367**, (2020).

75. Kennedy, P. Ectomycorrhizal fungi and interspecific competition: species interactions, community structure, coexistence mechanisms, and future research directions: Tansley review. *New Phytol.* **187**, 895–910 (2010).

76. Baldrian, P. & Kohout, P. Interactions of saprotrophic fungi with tree roots: can we observe the emergence of novel ectomycorrhizal fungi? *New Phytol.* **215**, 511–513 (2017).

77. Frenot, Y., Gloaguen, J. C., Cannavacciuolo, M. & Bellido, A. Primary succession on glacier forelands in the subantarctic Kerguelen Islands. *J. Veg. Sci.* **9**, 75–84 (1998).

78. Hodkinson, I. D., Coulson, S. J., Harrison, J. & Webb, N. R. What a wonderful web they weave: spiders, nutrient capture and early ecosystem development in the high Arctic – some counter-intuitive ideas on community assembly. *Oikos* **95**, 349–352 (2001).

79. Hodkinson, I. D., Webb, N. R. & Coulson, S. J. Primary community assembly on land – the missing stages: why are the heterotrophic organisms always there first? *J. Ecol.* **90**, 569–577 (2002).

80. Jumpponen, A. Soil fungal community assembly in a primary successional glacier forefront ecosystem as inferred from rDNA sequence analyses. *New Phytol.* **158**, 569–578 (2003).

81. Canini, F. *et al.* Vegetation, pH and Water Content as Main Factors for Shaping Fungal Richness, Community Composition and Functional Guilds Distribution in Soils of Western Greenland. *Front. Microbiol.* **10**, 2348 (2019).

82. Ni, Y., Yang, T., Zhang, K., Shen, C. & Chu, H. Fungal Communities Along a Small-Scale Elevational Gradient in an Alpine Tundra Are Determined by Soil Carbon Nitrogen Ratios. *Front. Microbiol.* **9**, 1815 (2018).

83. Siciliano, S. D. *et al.* Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biol. Biochem.* **78**, 10–20 (2014).

84. Darcy, J. L. *et al.* Phosphorus, not nitrogen, limits plants and microbial primary producers following glacial retreat. *Sci. Adv.* **4**, eaaq0942 (2018).

85. Hanaka, A. *et al.* Relationships between the properties of Spitsbergen soil, number and biodiversity of rhizosphere microorganisms, and heavy metal concentration in selected plant species. *Plant Soil* **436**, 49–69 (2019).

86. Bell, T., Callender, K., Whyte, L. & Greer, C. Microbial Competition in Polar Soils: A Review of an Understudied but Potentially Important Control on Productivity. *Biology* **2**, 533–554 (2013).

87. Nemergut, D. R., Shade, A. & Violle, C. When, where and how does microbial community composition matter? *Front. Microbiol.* **5**, (2014).

88. Botnen, S. *et al.* Low host specificity of root-associated fungi at an Arctic site. *Mol. Ecol.* **23**, 975–985 (2014).

89. Marasco, R. *et al.* Rhizosheath microbial community assembly of sympatric desert speargrasses is independent of the plant host. *Microbiome* **6**, 215 (2018).

90. Gil-Martínez, M. *et al.* Ectomycorrhizal Fungal Communities and Their Functional Traits Mediate Plant–Soil Interactions in Trace Element Contaminated Soils. *Front. Plant Sci.* **9**, 1682 (2018).

91. Lofgren, L. A. *et al.* Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Mol. Ecol.* **28**, 721–730 (2019).

92. Kwan, E. X., Wang, X. S., Amemiya, H. M., Brewer, B. J. & Raghuraman, M. K. rDNA Copy Number Variants Are Frequent Passenger Mutations in *Saccharomyces cerevisiae* Deletion Collections and *de Novo* Transformants. *G3amp58 GenesGenomesGenetics* **6**, 2829–2838 (2016).

93. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).

94. Deagle, B. E. *et al.* Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Mol. Ecol.* **28**, 391–406 (2019).

95. Lamb, P. D. *et al.* How quantitative is metabarcoding: A meta-analytical approach. *Mol.*

Ecol. **28**, 420–430 (2019).

96. Lindahl, B. D. *et al.* Fungal community analysis by high-throughput sequencing of amplified markers - a user's guide. *New Phytol.* **199**, 288–299 (2013).

97. Nguyen, N. H., Smith, D., Peay, K. & Kennedy, P. Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytol.* **205**, 1389–1393 (2015).

98. Song, Z. *et al.* Effort versus Reward: Preparing Samples for Fungal Community Characterization in High-Throughput Sequencing Surveys of Soils. *PLOS ONE* **10**, e0127234 (2015).

99. Taylor, D. L. *et al.* Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. *Appl. Environ. Microbiol.* **82**, 7217–7226 (2016).

100. Elser, J. J. *et al.* Biological stoichiometry from genes to ecosystems. *Ecol. Lett.* **3**, 540–550 (2000).

101. de Villemereuil, P., Gaggiotti, O. E., Mouterde, M. & Till-Bottraud, I. Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* **116**, 249–254 (2016).

102. Wijk, M. T. V., Williams, M., Gough, L., Hobbie, S. E. & Shaver, G. R. Luxury consumption of soil nutrients: a possible competitive strategy in above-ground and below-ground biomass allocation and root morphology for slow-growing arctic vegetation? *J. Ecol.* **91**, 664–676 (2003).

103. Totland, Ø. & Nyléhn, J. Assessment of the effects of environmental change on the performance and density of Bistorta vivipara: the use of multivariate analysis and experimental manipulation. *J. Ecol.* **86**, 989–998 (1998).

104. Billings, W. D. & Mooney, H. A. The Ecology of Arctic and Alpine Plants. *Biol. Rev.* **43**, 481–529 (1968).

105. Baruah, G., Molau, U., Bai, Y. & Alatalo, J. M. Community and species-specific responses of plant traits to 23 years of experimental warming across subarctic tundra plant communities. *Sci. Rep.* **7**, 1–11 (2017).

Acknowledgments

The authors are thankful to all the people that contributed to the terrestrial part of MicroFun project, collected and processed the samples. This research was funded by University Centre in Svalbard, as well as ConocoPhillips and Lundin Petroleum through The Northern Area Program. The Svalbard Science Forum is acknowledged for providing Arctic Field Grant to SM (2012 and 2013), for sampling in Svalbard (project code 220126/E10; RIS ID 5009). We also thank the Governor (Sysselmannen, Longyearbyen) for permitting us to collect the root and soil samples from Svalbard.

Author Contributions Statement

MW, DE, SM, PB - wrote and edited the manuscript; MW, AV - analysed sequencing data; MW, DE - did statistical modelling; SM - collected and processed soil samples, measured abiotic parameters, generated sequencing data, PB - developed the idea and

Additional Information

The already published datasets are available online. All the other datasets used in this study are available at zenodo.com/addproperlink. Scripts generated for bioinformatic and statistical analysis are available at https://github.com/magdawutkowska/bistorta.

Competing Interests

The authors declare no competing interests.

Figures

Figure 1

Bistorta vivipara plants from the four concatenated datasets were collected in nine localities on Spitsbergen.

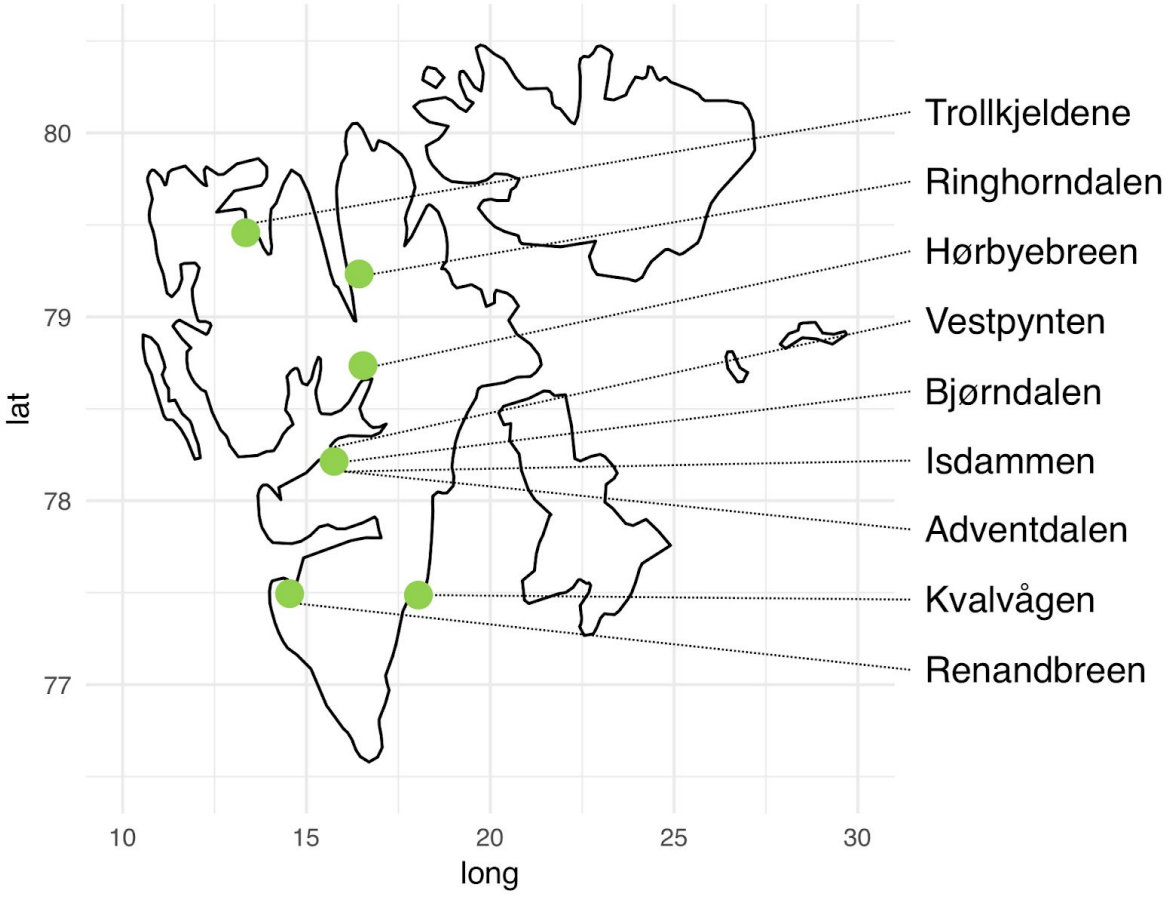


Figure 2

Schematic illustration of a conceptual plant-centric model representing relationships between variables suggested by the literature and tested in this study. Solid lines are associations were researched by studies from the Arctic; dashed lines were described by fewer studies, mainly from other regions. The full model includes all possible links between each abiotic, fungal and plant variable. Abbreviations and symbols: N - soil nitrogen content; C/N - the ratio of soil nitrogen to soil carbon content; p- precipitation; t - temperature; D - diversity ; Sy/Sa - the ratio of symbio- to saprotrophs; CC - fungal community composition; I/S - the ratio of inflorescence to stem length; RV - rhizome volume; LL - leaf length of the longest leaf.

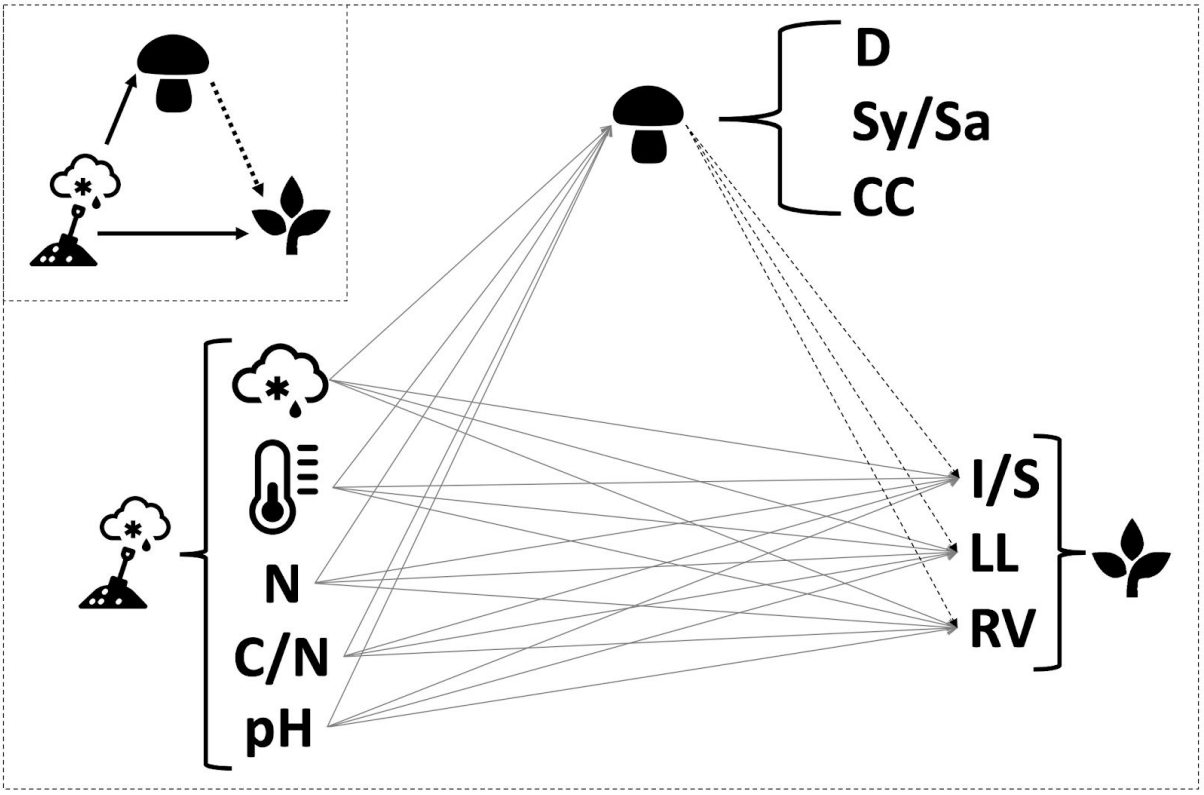
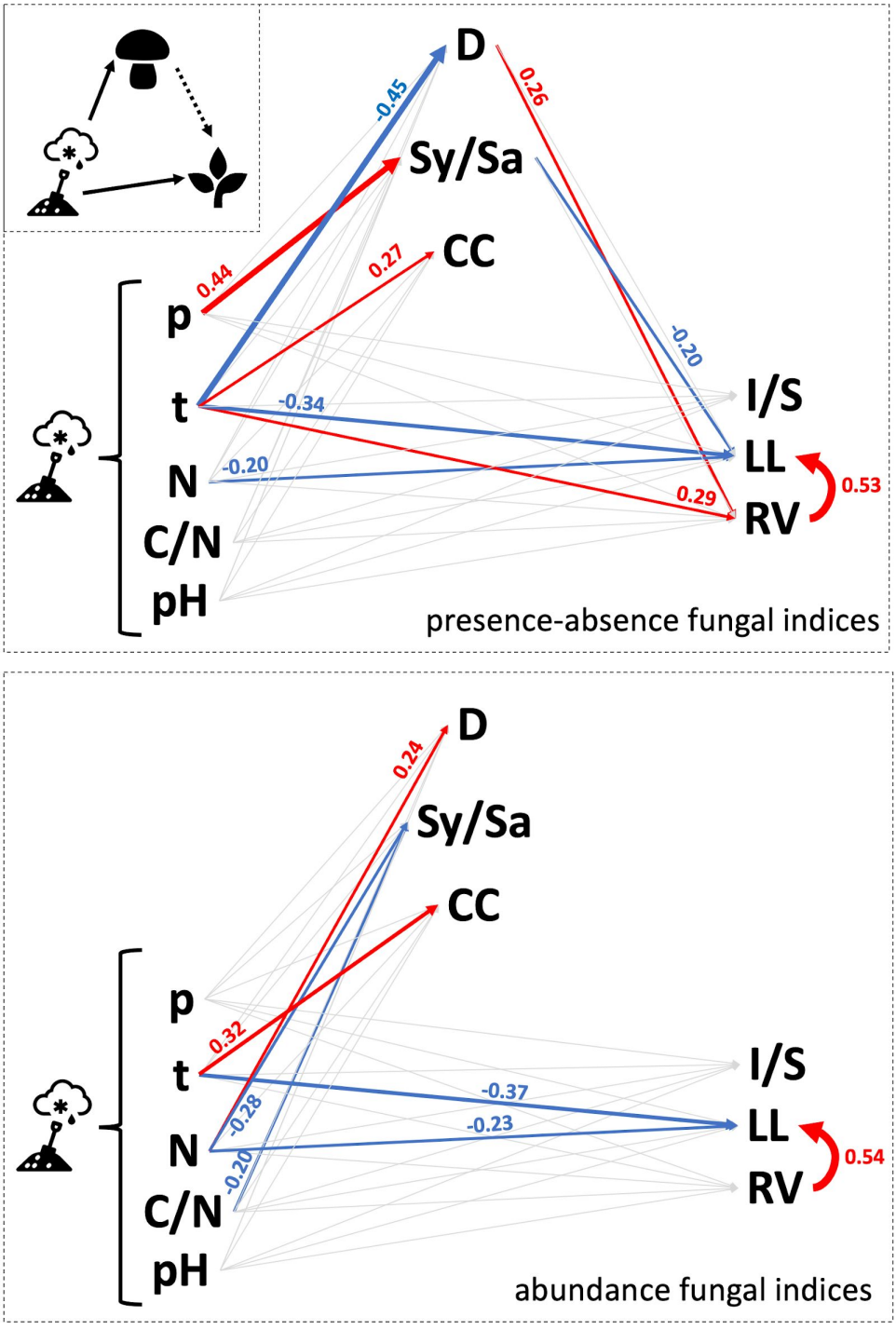


Figure 3

Path diagram showing tested connections between predictor and response variables in the best fitting models. Statistically significant ($p < 0.05$) links are depicted by arrow colours (positive or negative nature of the relationship) and thickness (relationship magnitude); the numbers are estimates from the models. Abbreviations and symbols: N - soil nitrogen content; C/N - the ratio of soil nitrogen to soil carbon content; p - precipitation; t - temperature; D - diversity ; Sy/Sa - the ratio of symbio- to saprotrophs; CC - fungal community composition; I/S - the ratio of inflorescence to stem length; RV - rhizome volume; LL - leaf length of the longest leaf.



Tables

Table 1

Overview of the data included in this study. Each dataset was generated to investigate specific topics regarding *Bistorta vivipara* root-associated fungi (RAF). References are given for previously published data.

Specific topic	**Number of localities /number of plants**	**Variables: edaphic**	***B.v.* RAF**	***B.v.* morphology**
temporal variation	1 / 72	[40]	[40]	this study
marginal habitats	3 / 58	[19]	[19]	this study
large spatial scale variability	5 / 38	this study	this study	this study
fungal response to increased snow	1 / 46	*in prep.*	*in prep.*	this study

Table 2

Metrics used to describe the fungal community used in this study for presence absence data and number of reads, respectively. All the parameters were calculated using a rarefied table containing amplicon sequence variants (ASVs).

Fungal parameter	Presence-absence table	Abundance table
Diversity (D)	richness (number of ASV)	Shannon-Wiener (H') index
$\frac{Symbio-}{Saprotrophs}$ (Sy/Sa)	ratio of ASVs	ratio of reads
Community composition (CC)	GNMDS 1st axis score	GNMDS 1st axis score

Table 3

Relationships between abiotic factors and root-associated fungi or plant metrics documented in the literature. Some of the relationships have been demonstrated generally for arctic plants and arctic fungi, and have not been specifically shown in *B. vivipara*. Abbreviations: N - soil nitrogen content; C/N - ratio of soil nitrogen to soil carbon content; p - precipitation, t - temperature, *B.v.* - whether the study was specifically conducted on *B. vivipara* plants or *B. vivipara* root-associated fungal communities.

Causal variable	**Assumed association**	**Response variable**	**in a study / from**	***B.v.***
SOIL:		PLANTS:		
N	positive	below-ground biomass allocation	[102]/Low Arctic	no
N & C/N	positive	leaf (length, width, area), corm dry weight, spike length, number of bulbils per spike, individual bulbil dry weight	[72]/Svalbard	yes
pH	negative	plant performance	[103]/alpine tundra, Norway	yes
CLIMATE:				
p	positive	leaf area	[72]/Svalbard	yes
t	positive	metabolism rate (growth, productivity etc.)	[104]/circumpolar & alpine	no
	positive	sexual reproduction	[104]/circumpolar & alpine	no
	positive	spike length	[72]/Svalbard	yes
	positive	leaf length and plant height in tussock tundra; leaf width and plant height in *Dryas* heath	[105]/subarctic, Sweden	yes
	negative	leaf length in *Dryas* heath and wet meadow	[105]/subarctic, Sweden	yes
SOIL:		FUNGI:		
N	negatively	richness and community composition	[83]/circumpolar	no
C/N	negatively	richness	[82]/alpine tundra	no
pH	negatively	community composition and richness	[83]/circumpolar	no
	positively	abundance of ectomycorrhizal fungi	[81]/Greenland	no
CLIMATE:				
p	positively	community composition and richness	[37]/ *	yes
t	positively	community composition and richness	[37]/ *	yes

* Austria, Scotland, Mainland Norway, Iceland, Jan Mayen and Svalbard.

Table 4

Summary of the models and statistics used for best fitting model selection. Each model reflects a separate hypothesis. The full model includes all possible links between each fungal variable and each plant variable. Subsequent models exclude some of the links, as indicated in the name of each model. Abbreviations: I/S - ratio of inflorescence to stem length; RV - rhizome volume; LL - leaf length; CC - root-associated fungal community composition.

Model	Fisher's C	p	AIC
Presence-absence			
Full	3.2	0.780	121.23
I/S does not depend on fungi	8.9	0.837	118.86
RV does not depend on fungi	*28.0*	*0.014*	*138.00*
LL does not depend on fungi	22.3	0.073	132.31
Fungal CC not important	10.3	0.739	120.32
Fungal CC not important + no I/S	**12.0**	**0.849**	**117.97**
No effect of fungi on plants	*42.9*	*0.020*	*140.86*
Abundance			
Full	7.1	0.529	123.07
I/S does not depend on fungi	11.7	0.632	121.68
RV does not depend on fungi	13.6	0.482	123.58
LL does not depend on fungi	13.4	0.492	123.44
Fungal CC not important	13.5	0.491	123.45
Fungal CC not important + no I/S	13.5	0.759	119.53
No effect of fungi on plants	**21.3**	**0.728**	**119.28**

The best model for each approach is highlighted in **bold**.
Models which don't fit based on the test of directed separation are in *italics*.

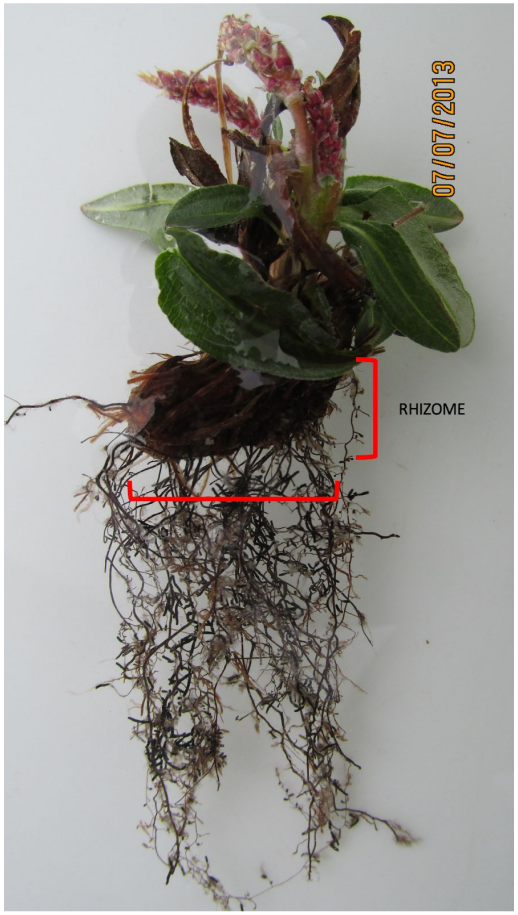
Table 5

Proportion of variance explained without (marginal R^2) and with random factors (conditional R^2). Locality was used as a random factor in all of the models. Abbreviations: D - diversity; Sy/Sa - the ratio of symbio- to saprotrophs; CC - fungal community composition; I/S - the ratio of inflorescence to stem length; RV - rhizome volume; LL - leaf length of the longest stem leaf.

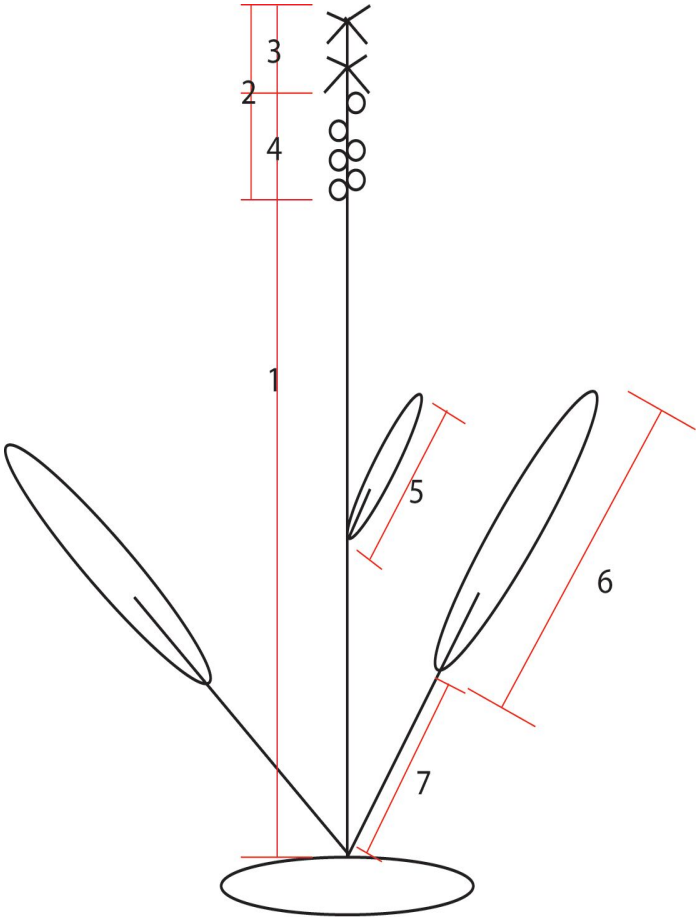
	Presence-absence model **CC does not impact plants + no I/S**		Abundance model **No effect of fungi on plants**	
Response	Marginal R^2	Conditional R^2	Marginal R^2	Conditional R^2
Fungal:				
D	0.11	0.51	0.07	0.32
Sy/Sa	0.16	0.56	0.08	0.11
CC	0.26	0.84	0.18	0.66
Plant:				
I/S	0.02	0.34	0.02	0.33
RV	0.24	0.39	0.15	0.24
LL	0.44	0.46	0.42	0.43

Supplementary 1

Morphological characteristics of *Bistorta vivipara* measured in this study: rhizome volume (RV; panel A), leaf length (LL; panel B, number 6) and a ratio of inflorescence to stem length (I/S; panel B, ratio of number 2 to 1). Photo: Sunil Mundra.



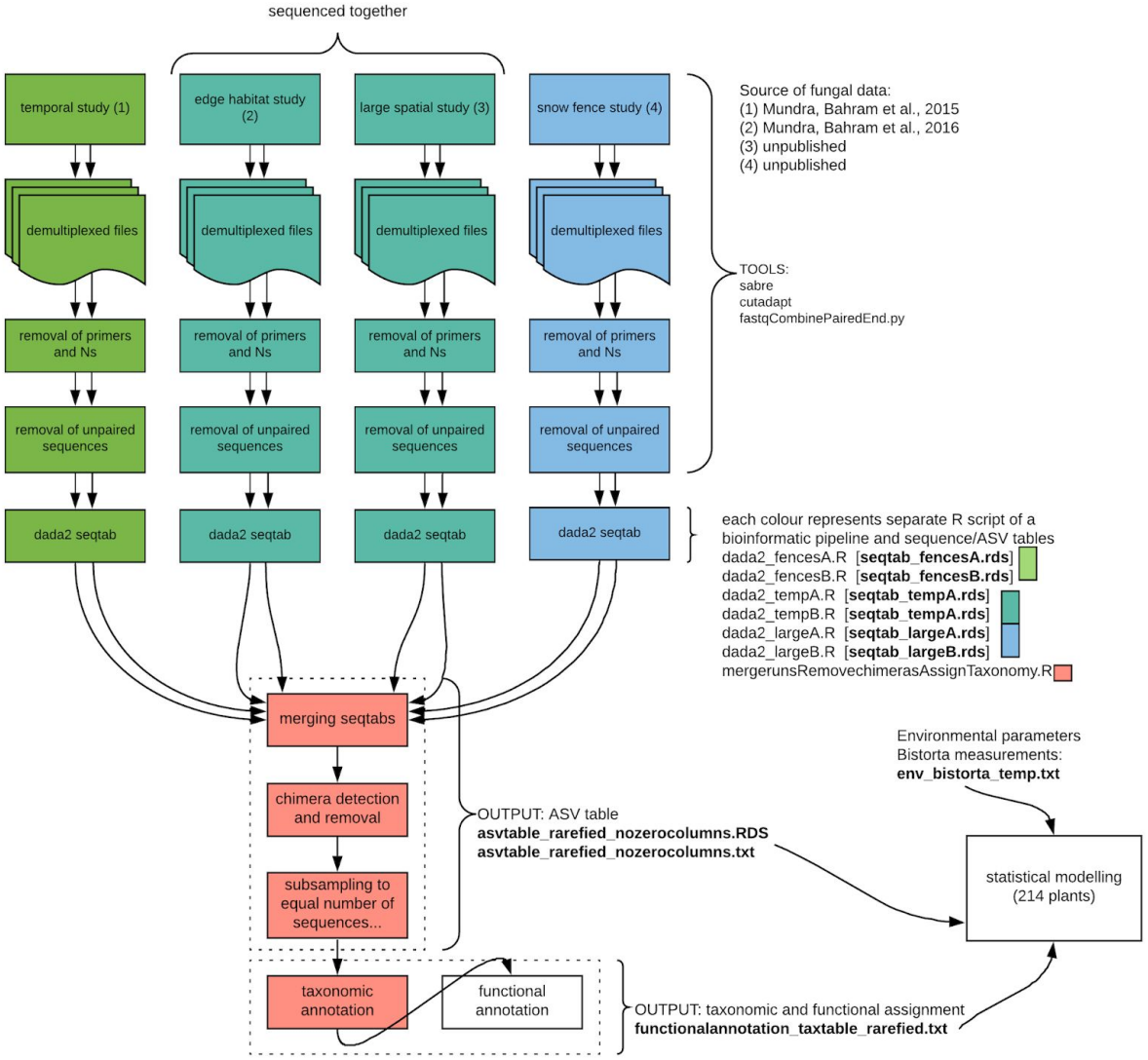
A



B

Supplementary 2

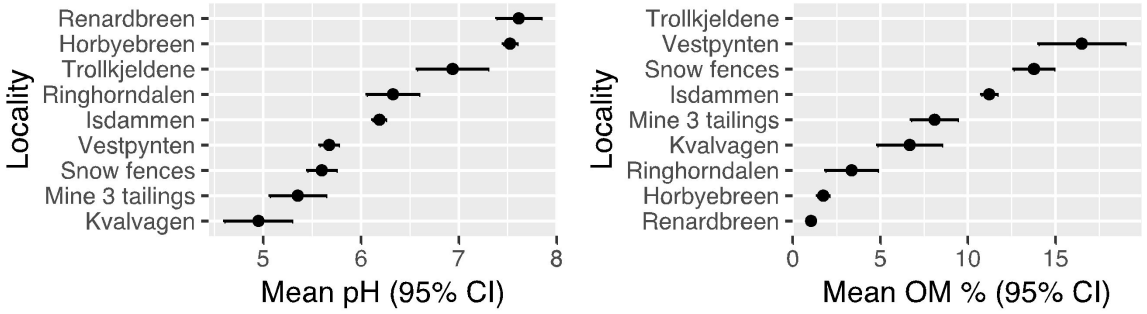
The overview of bioinformatics pipeline analysing fungal data.

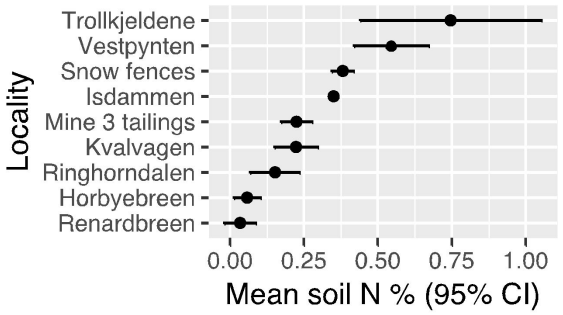


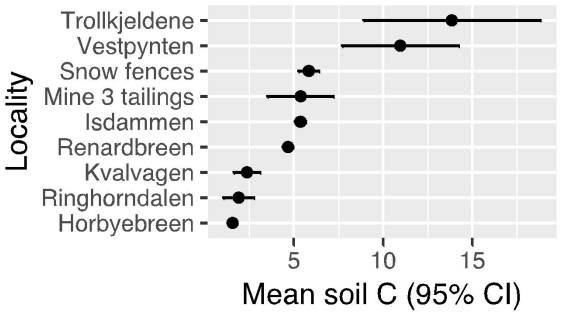
Supplementary 3
Characteristics of localities used in this study

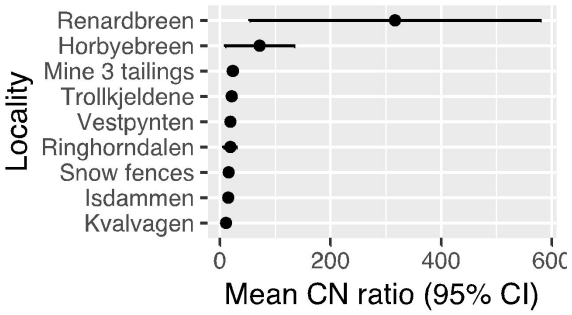
Localities	Description
Renardbreen	glacier forefront
Hørbyebreen	glacier forefront
Trollkjeldene	hot springs
Ringhorndalen	arctic steppe
Isdammen	natural tundra
Vestpynten	nutrient-rich tundra
Adventdalen (Snow fences)	natural tundra
Bjørndalen (Mine 3 tailings)	nutrient-rich mine-contaminated site
Kvalvågen	hydrocarbon-rich site

a. **How did the localities differ in terms of edaphic variables?**

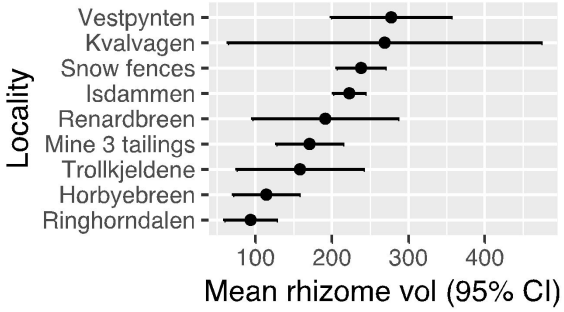


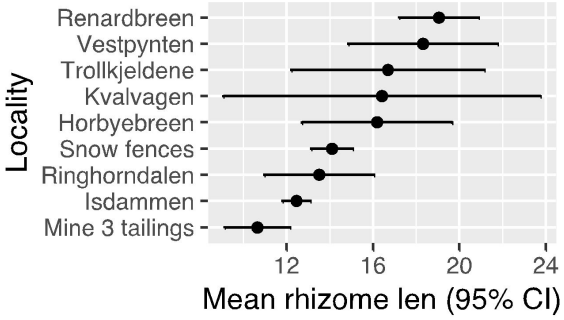


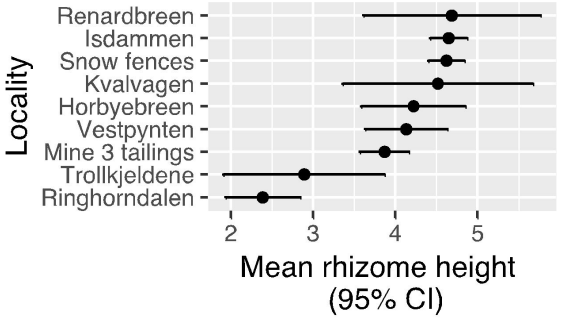


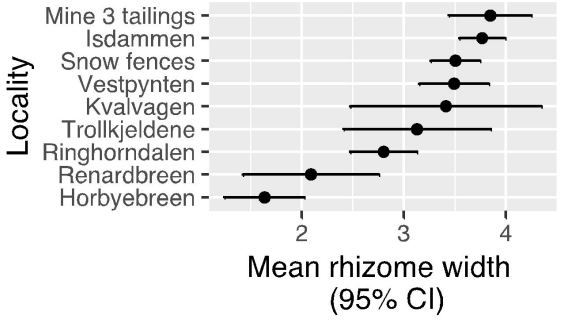


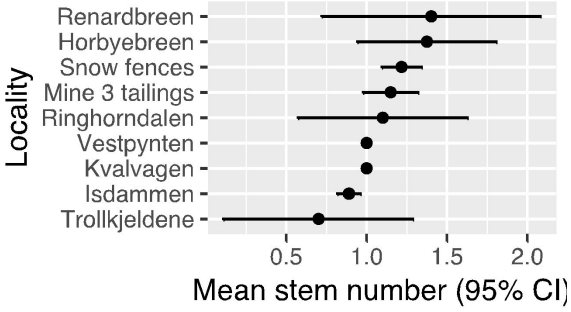
b. **How did plant measurements differ in studied localities?**

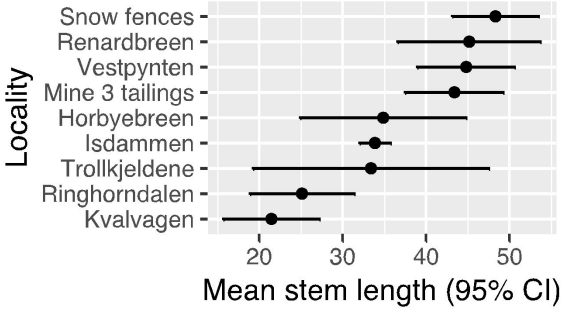


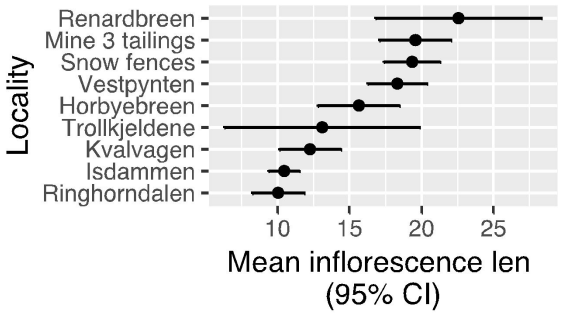


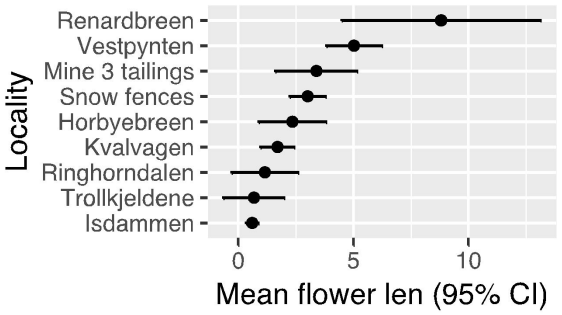


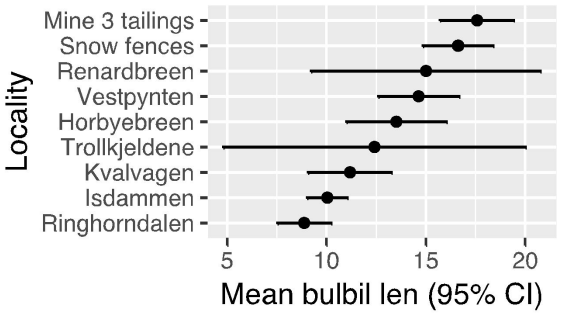


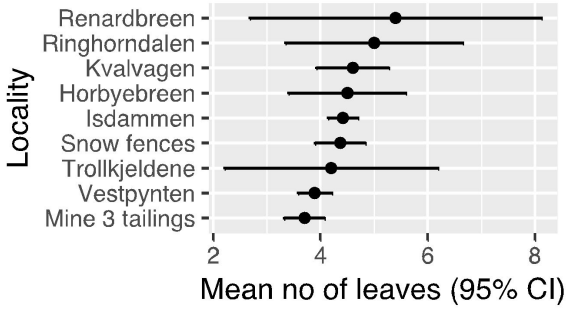


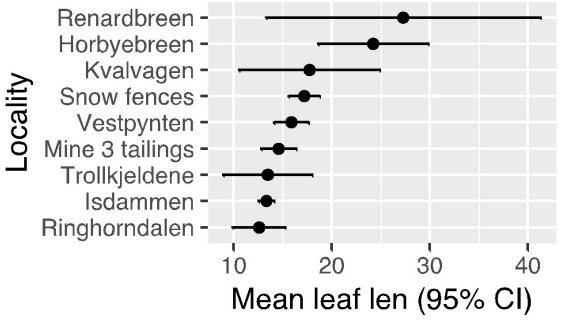


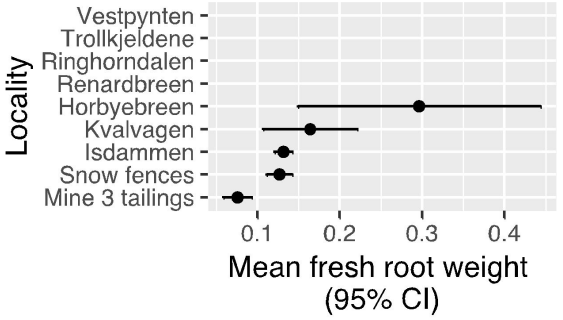

Locality
Renardbreen
Mine 3 tailings
Snow fences
Vestpynten
Horbyebreen
Trollkjeldene
Kvalvagen
Isdammen
Ringhorndalen
10
15
20
25
Mean inflorescence len (95% CI)

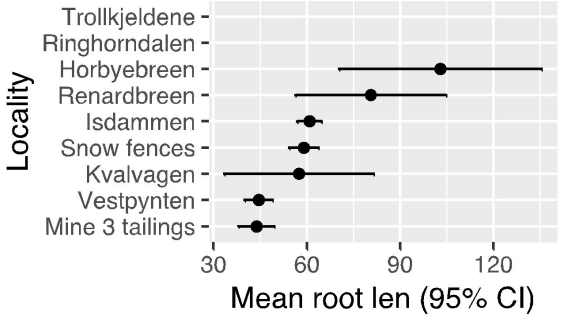

Locality
Renardbreen
Vestpynten
Mine 3 tailings
Snow fences
Horbyebreen
Kvalvagen
Ringhorndalen
Trollkjeldene
Isdammen
0
5
10
Mean flower len (95% CI)

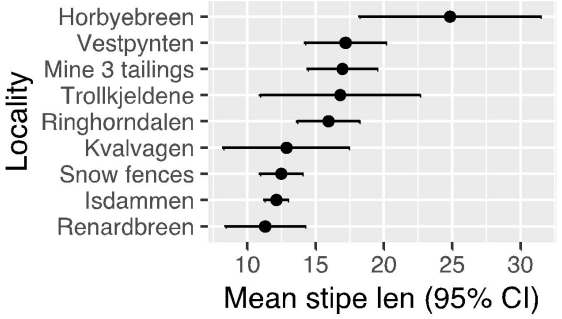

Locality
Mine 3 tailings
Snow fences
Renardbreen
Vestpynten
Horbyebreen
Trollkjeldene
Kvalvagen
Isdammen
Ringhorndalen
5
10
15
20
Mean bulbil len (95% CI)


Locality
Renardbreen
Ringhorndalen
Kvalvagen
Horbyebreen
Isdammen
Snow fences
Trollkjeldene
Vestpynten
Mine 3 tailings
2
4
6
8
Mean no of leaves (95% CI)

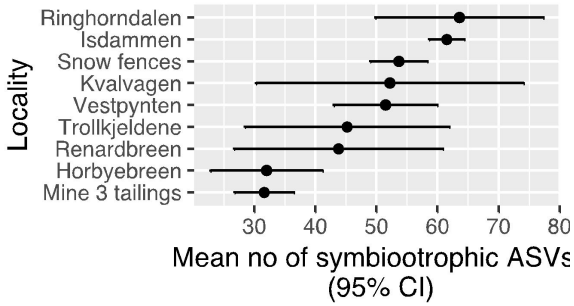

Locality
Renardbreen
Horbyebreen
Kvalvagen
Snow fences
Vestpynten
Mine 3 tailings
Trollkjeldene
Isdammen
Ringhorndalen
10
20
30
40
Mean leaf len (95% CI)

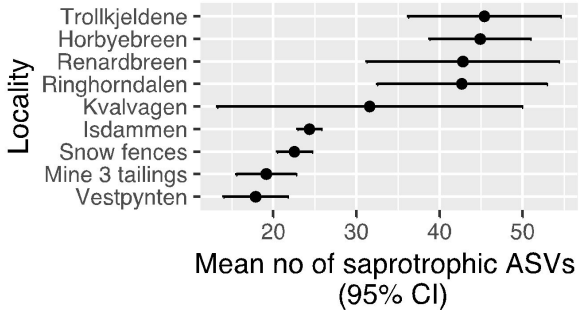

Locality
Vestpynten
Trollkjeldene
Ringhorndalen
Renardbreen
Horbyebreen
Kvalvagen
Isdammen
Snow fences
Mine 3 tailings
0.1
0.2
0.3
0.4
Mean fresh root weight (95% CI)

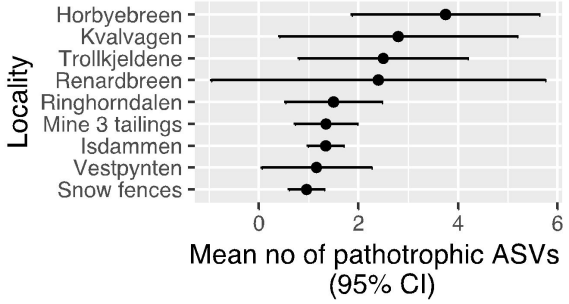

Locality
Trollkjeldene
Ringhorndalen
Horbyebreen
Renardbreen
Isdammen
Snow fences
Kvalvagen
Vestpynten
Mine 3 tailings
30
60
90
120
Mean root len (95% CI)

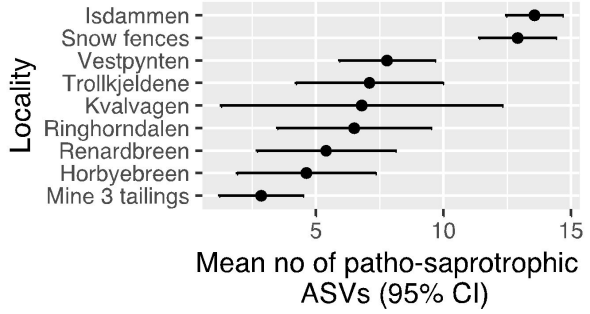

Locality
Horbyebreen
Vestpynten
Mine 3 tailings
Trollkjeldene
Ringhorndalen
Kvalvagen
Snow fences
Isdammen
Renardbreen
10
15
20
25
30
Mean stipe len (95% CI)

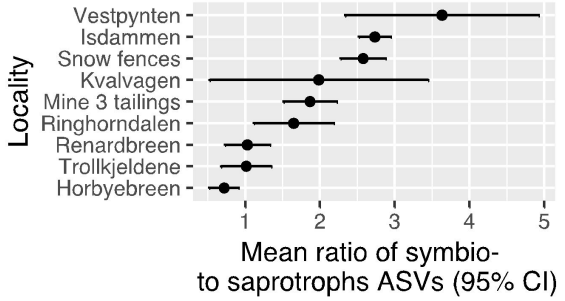
c. **How do fungi differ in studied localities?**











Supplementary 4

Complete list of coefficients and estimates calculated in the two best fitting models. Abbreviations: **N** - soil nitrogen content; **CN** - the ratio of soil carbon and nitrogen content; **p** - precipitation; **t** - temperature; **D** - fungal richness (in presence-absence model: number of fungal amplicon sequence variants (ASVs); in abundance model - Shannon-Wiener index); **Sy/Sa** - in presence-absence model: the ratio of symbio- and saprotrophic ASVs and in abundance model: the ratio of symbio- to saprotrophic reads; **CC** - community composition proxy based on presence-absence or read abundance table, respectively; **I/S** - the ratio of inflorescence length to stem length; **RV** - rhizome volume; **LL** - leaf length; statistical significance is coded such as '***' indicate p-value = 0 - 0.001, '**' - 0.001 - 0.01 and '*' 0.01 - 0.05.

a) presence-absence model (Fungal CC not important + no I/S = community composition does not impact plants and no effect of fungi on I/S)

Response	Predictor	Estimate	Std.Error	DF	Crit.Value	P.Value	Std.Estimate	
D	N	0.0980	0.1230	173	0.7967	0.4267	0.0981	
D	CN	0.1390	0.1151	173	1.2083	0.2286	0.1429	
D	pH	-0.0095	0.1499	173	-0.0635	0.9495	-0.0066	
D	p	-0.0217	0.2350	173	-0.0922	0.9266	-0.0199	
D	**t**	**-0.4482**	**0.1266**	**173**	**-3.5410**	**0.0005**	**-0.4048**	***
Sy/Sa	N	0.0713	0.1060	173	0.6727	0.5020	0.0723	
Sy/Sa	CN	0.0376	0.0991	173	0.3789	0.7052	0.0391	
Sy/Sa	pH	-0.0627	0.1292	173	-0.4853	0.6281	-0.0442	
Sy/Sa	**p**	**0.4381**	**0.2093**	**173**	**2.0931**	**0.0378**	**0.4072**	*
Sy/Sa	t	0.0382	0.1099	173	0.3476	0.7286	0.0349	
CC	N	0.0125	0.0679	173	0.1844	0.8539	0.0129	
CC	CN	-0.0256	0.0636	173	-0.4025	0.6878	-0.0271	
CC	pH	0.0372	0.0829	173	0.4489	0.6541	0.0267	
CC	p	0.4397	0.2373	173	1.8525	0.0657	0.4147	
CC	**t**	**0.2698**	**0.0879**	**173**	**3.0708**	**0.0025**	**0.2504**	**
I/S	N	-0.1020	0.1057	173	-0.9644	0.3362	-0.1188	
I/S	CN	-0.0611	0.0992	173	-0.6162	0.5385	-0.0731	
I/S	pH	0.1317	0.1292	173	1.0195	0.3094	0.1068	
I/S	p	-0.0064	0.1726	173	-0.0371	0.9704	-0.0068	
I/S	t	0.0761	0.1062	173	0.7167	0.4746	0.0800	
RV	N	0.0835	0.1112	171	0.7512	0.4535	0.0857	

RV	CN	-0.1119	0.1046	171	-1.0696	0.2863	-0.1180	
RV	pH	0.0224	0.1362	171	0.1644	0.8696	0.0160	
RV	p	0.2854	0.1552	171	1.8389	0.0677	0.2686	
RV	**t**	**0.2869**	**0.1132**	**171**	**2.5352**	**0.0121**	**0.2656**	*
RV	**D**	**0.2632**	**0.0716**	**171**	**3.6782**	**0.0003**	**0.2698**	***
RV	Sy/Sa	-0.0983	0.0836	171	-1.1756	0.2414	-0.0996	
LL	**RV**	**0.5255**	**0.0607**	**170**	**8.6629**	**0.0000**	**0.5252**	***
LL	CN	0.1178	0.0761	170	1.5477	0.1236	0.1241	
LL	pH	-0.0283	0.0980	170	-0.2889	0.7730	-0.0202	
LL	p	0.0504	0.0896	170	0.5631	0.5741	0.0474	
LL	**t**	**-0.3429**	**0.0816**	**170**	**-4.2005**	**0.0000**	**-0.3173**	***
LL	D	-0.0063	0.0580	170	-0.1086	0.9137	-0.0065	
LL	**Sy/Sa**	**-0.1975**	**0.0672**	**170**	**-2.9377**	**0.0038**	**-0.1999**	**
~~CC	~~Sy/Sa	0.1560	NA	187	2.1428	0.0167	0.1560	*

b) abundance model (no effect of fungi on plants)

Response	**Predictor**	**Estimate**	**Std.Error**	**DF**	**Crit.Value**	**P.Value**	**Std.Estimate**	
D	**N**	**0.2355**	**0.1108**	**171**	**2.1258**	**0.0350**	**0.2470**	*
D	CN	-0.0124	0.1043	171	-0.1190	0.9054	-0.0134	
D	pH	-0.1292	0.1366	171	-0.9459	0.3455	-0.0944	
D	p	-0.0515	0.1670	171	-0.3083	0.7582	-0.0496	
D	t	-0.0288	0.1105	171	-0.2607	0.7946	-0.0273	
Sy/Sa	**N**	**-0.2792**	**0.1015**	**171**	**-2.7513**	**0.0066**	**-0.2967**	**
Sy/Sa	**CN**	**-0.2012**	**0.0973**	**171**	**-2.0686**	**0.0401**	**-0.2199**	*
Sy/Sa	pH	0.2166	0.1239	171	1.7482	0.0822	0.1603	
Sy/Sa	p	0.1372	0.1084	171	1.2654	0.2075	0.1338	
Sy/Sa	t	0.0542	0.1014	171	0.5349	0.5934	0.0521	
CC	N	-0.0341	0.0844	171	-0.4039	0.6868	-0.0347	

CC	CN	-0.0165	0.0790	171	-0.2087	0.8349	-0.0173	
CC	pH	-0.2032	0.1037	171	-1.9595	0.0517	-0.1442	
CC	p	0.1709	0.1979	171	0.8634	0.3891	0.1598	
CC	**t**	**0.3168**	**0.0917**	**171**	**3.4558**	**0.0007**	**0.2916**	***
I/S	N	-0.1016	0.1059	171	-0.9598	0.3385	-0.1184	
I/S	CN	-0.0572	0.0995	171	-0.5754	0.5658	-0.0686	
I/S	pH	0.1153	0.1304	171	0.8846	0.3776	0.0936	
I/S	p	-0.0119	0.1716	171	-0.0694	0.9448	-0.0127	
I/S	t	0.0734	0.1062	171	0.6913	0.4903	0.0773	
RV	N	0.1342	0.1084	171	1.2383	0.2173	0.1385	
RV	CN	-0.0520	0.1034	171	-0.5034	0.6153	-0.0553	
RV	pH	0.0183	0.1344	171	0.1359	0.8921	0.0131	
RV	p	0.2171	0.1324	171	1.6395	0.1029	0.2056	
RV	t	0.1636	0.1081	171	1.5131	0.1321	0.1526	
LL	**RV**	**0.5431**	**0.0603**	**170**	**9.0090**	**0.0000**	**0.5395**	***
LL	**N**	**-0.2272**	**0.0799**	**170**	**-2.8434**	**0.0050**	**-0.2331**	**
LL	CN	0.1374	0.0756	170	1.8167	0.0710	0.1449	
LL	pH	0.0113	0.0953	170	0.1184	0.9059	0.0081	
LL	p	-0.0245	0.0822	170	-0.2978	0.7662	-0.0230	
LL	**t**	**-0.3739**	**0.0788**	**170**	**-4.7443**	**0.0000**	**-0.3464**	***
~~Sy/Sa	~~D	-0.4151	NA	185	-6.1557	0.0000	-0.4151	***

PAPER III

Linking extreme seasonality and gene expression in arctic marine protists

Magdalena Wutkowska[1, 2*], **Anna Vader**[1], **Ramiro Logares**[3], **Eric Pelletier**[4], **Tove M. Gabrielsen**[1, 5]

[1] Department of Arctic Biology, The University Centre in Svalbard, Longyearbyen, Norway
[2] Department of Arctic and Marine Biology, UiT – The Arctic University of Norway, Tromsø, Norway
[3] Institute of Marine Sciences (ICM), CSIC, Barcelona, Catalonia, Spain
[4] CEA – Institut de Génomique, Genoscope, Evry, France
[5] Department of Natural Sciences, University of Agder, Kristiansand, Norway
*corresponding author: magda.wutkowska@gmail.com

ABSTRACT

In high latitudes, strong seasonality, especially manifested as dramatic differences in light availability, affects various marine organisms and restricts the timing of ecosystem processes. Marine protists are key players in Arctic aquatic ecosystems, yet little is known about their ecological roles over yearly cycles. In particular, even though the myth that the polar night is devoid of biological activity has been debunked recently, we barely understand the ecological role of aquatic protists during this period. Here, we explore community-level gene expression patterns in protist assemblages featuring cell sizes between 0.45-10 µm that populate an Arctic marine time series. We found that transcript diversity and evenness was higher during the polar night than during the polar day. Community gene expression was correlated with seasonality, with light as the main driving factor. As expected, light-dependent functions had higher levels of expression during the polar day than during the polar night, except phototransduction. Among the most expressed genes, 64% could not be annotated functionally by any of the databases. However, up to 78% of them were identified in samples from the expedition *Tara* Oceans, especially from the Arctic Ocean, suggesting a genetic makeup distinct from other oceans. Our study increases our understanding of the links between extreme seasonality and biological processes in pico- and nanoplanktonic protists. Furthermore, our results may set the ground for future monitoring studies investigating climate change in the High Arctic.

KEY WORDS

metatranscriptomics, polar night, polar day, microbial eukaryotes, unicellular eukaryotes

Introduction

Solar radiation is a dominant energy source for life on Earth, and an important driver of evolution (Judson, 2017). In the ocean, phytoplankton, mostly cyanobacteria and photosynthetic microbial eukaryotes, contribute half of the net primary production on Earth (Field et al., 1998). Light availability in the ocean declines with depth and forces a vertical distribution of species, with phototrophic organisms dwelling in the epipelagic zone (<200 m depth). The further from the equator, the more pronounced the annual changes in light regime; at high latitudes are the strongest environmental drivers of marine plankton phenology (Boyce et al., 2017). During polar night the sun does not rise above the horizon for 4-6 months, and the opposite happens during polar day, when the sun stays above the horizon for an equally long period. Extreme seasonality introduces profound limitations to biological processes in polar regions, and for centuries researchers perceived polar night as a period devoid of biological activity with limited physical accessibility for sampling. Recent studies reported substantial biological activity during the polar night; however, most of these studies focused on macroorganisms, mainly zooplankton (reviewed in: Berge, Daase, et al., 2015; Berge, Renaud, et al., 2015; Błachowiak-Samołyk et al., 2015).

Our understanding of communities of marine microbial eukaryotes in the Arctic is mainly based on studies limited to a single sampling time point or cruises sampling along transects once or infrequently. Time series stations sampled at regular intervals are important for disentangling the dynamics of changing community composition of organisms (Bunse & Pinhassi, 2017; Moreira & López-García, 2019). The world's northernmost marine time station (IsA) was established in Adventfjorden, Isfjorden, Svalbard, and continuously sampled from December 2011 (Figure 1; Buttigieg et al., 2018). This endeavour generated metabarcoding-based knowledge regarding which marine microbial eukaryotes are present and active throughout the year (Marquardt et al., 2019; Marquardt et al., 2016; Wiedmann et al., 2016). Seasonal dynamics of microbial eukaryotes can be analysed through many ecologically important measures and indices, such as diversity, biomass, cell counts, functions etc. In general, cell counts and biomass of microbial eukaryotes during polar night are lower compared to polar day (Kubiszyn et al., 2017), whereas OTU diversity is inversely proportional to this trend being higher during polar night (Marquardt et al., 2016). Proportion of plastid-bearing to heterotrophic cells is lower during polar night (reviewed in Berge, Renaud, et al., 2015).

Studies on the response of natural polar microbial communities to light/dark cycles are rare and usually shorter than the duration of polar night (Kvernvik et al., 2018). Typical studies on dark survival of photosynthetic unicellular organisms are performed in laboratory conditions on single species cultures (e.g. Smayda & Mitchell-Innes, 1974; Vaulot et al., 1986; Walter et al., 2017). Some of the key arctic microeukaryotic phototrophs were found 'ribosomally active' during polar night (Marquardt et al., 2016; Vader et al., 2014). Since most of the primary production in the Arctic Ocean is performed by marine microbial eukaryotes when enough solar radiation is available (Metfies et al., 2016; Sherr et al., 2003), outside that period these cells are assumed to use accumulated resources (Schaub et al., 2017), decrease their metabolism (Smayda & Mitchell-Innes, 1974; Toseland et al., 2013) or remain dormant (reviewed in McMinn & Martin, 2013). However, many species of microbial eukaryotes instead of passively surviving prolonged darkness might switch their feeding strategy (McKie-Krisberg & Sanders, 2014; Sanders & Gast, 2012), as it may happen with mixotrophs (Stoecker & Lavrentyev, 2018).

Pico- and nanoeukaryotes play important roles in the marine environment, including photo-, heterotrophy or parasitism, and some species can switch between these trophic modes (Caron et al., 2017; de Vargas et al., 2015; Worden et al., 2015). Their gene expression helps to understand what molecular processes they use to respond to environmental heterogeneity (Caron et al., 2017; Keeling et al., 2014). Despite high stochasticity in gene expression of individual unicellular organisms ensuring survival in the times of stressful conditions (reviewed in Kærn et al., 2005), community-level gene expression obtained by 'omics' methods was demonstrated as an effective predictor of current marine biogeochemistry (Coles et al., 2017).

We targeted the 0.45-10 µm size fraction of the microbial eukaryotic community from the IsA time series station to determine the dynamics of gene expression throughout the year. Previous studies described higher diversity of microbial eukaryotes during polar night; thus, we hypothesise that the transcript diversity follows this trend. Given that light is a structuring force of community composition (Boyce et al., 2017), we hypothesise that the light regime plays an essential role in controlling cellular processes in microbial eukaryotes. The presence of active phototrophic microbial eukaryotes during the polar night and their quick ecophysiological response to return of light was confirmed by several studies (Kvernvik et al., 2018; Marquardt et al., 2016; Vader et al., 2014; van der Poll et al., 2020). Hence, we hypothesise that genes involved in light-dependent processes (such as light-harvesting etc.)

are expressed during polar night. To our knowledge, this is the first study to analyse the year-round seasonality of gene expression in pico- and nanoeukaryotes populating Arctic marine waters.

MATERIALS AND METHODS

Study site and sampling

The biological and environmental samples were collected at local noon at 11-time points between 14 December 2011 and 10 January 2013 from the Isfjorden Adventfjorden Time Series Station (IsA); located in the west coast of Spitsbergen, Svalbard (N 78°15.6, E 15°31.8, Figure 1). At each of the 11 sampling dates, 30 l of seawater were collected from 25m depth using a 10 l Niskin bottle (KC Denmark). Samples were kept in the dark and cold conditions while prefiltered by gravity through 10 µm nylon mesh (KC Denmark) and then onto 8-12 47 mm 0.45 µm Durapore filters (Millipore) using a vacuum pump. Filters were fixed in a 600 µl LB buffer (RNAqueous Total RNA Isolation Kit, Invitrogen, Thermo Fisher Scientific) 5-20 min after sampling, and subsequently, flash-frozen in liquid nitrogen and stored at -80°C.

At each sampling date, a vertical profile, up to 85m depth, of environmental variables were obtained using a handheld SAIV 204 STD/CTD probe. Photosynthetically active radiation (PAR), size-fractionated chlorophyll a and nutrient concentrations (nitrate/nitrite, phosphate, silicate), were obtained as described in Marquardt et al., 2016.

mRNA extraction and amplification

Total RNA was extracted with the RNAqueous Total RNA Isolation Kit (Invitrogen, Thermo Fisher Scientific). Samples were thawed on ice, vortexed and kept on ice in-between each step of RNA extraction. The thawed lysate was pipetted out into a tube with 200 µm molecular biology grade zirconium beads from pre-filled tubes. This step was repeated by adding 600 µl of lysis buffer into cryo-tubes with filters, vortexing, and pipetting to another tube with zirconium beads to maximise the capture of filter content. The bead beating step took 2 min with 1/22 s frequency. All centrifuge steps used 13,000 RCF. The total DNA and RNA that was bound to the membrane was eluted in two steps: first by adding 50 µl then 10 µl of preheated elution buffer and each time centrifuged. To detect and inspect fragment sizes of extracted DNA and RNA we used 3 µl of elution liquid from each tube with 1 µl 6x loading dye and run it on 0.7% agarose gel in 70 V current for 45 min in 1x TAE buffer. The

remaining 57 µl of eluted liquid from each sample was stored at -80°C. Content of the eluted liquid from the same sampling date were pooled together into three separate tubes. We removed DNA using TURBO DNA-free Kit (Invitrogen, Thermo Fisher Scientific). A total of 27 µl of elution liquid containing both DNA and RNA was mixed with 3 µl of 10x DNase buffer and 1 µl TURBO DNase and incubated at 37°C for 30 min. Treatment was stopped using room temperature incubation with 3 µl of DNase Inactivation Reagent. The tubes were centrifuged at 10 000 g for 1.5 min; the supernatant transferred to a new sterile tube and immediately stored at -80°C.

To test for the presence of PCR inhibitors, we used reverse transcription reaction of DNase-treated RNA samples. First, we denatured RNA molecules by incubating at 65°C for 5 min. in a mix of 1 µl of DNase-treated RNA, 1 µl of Random Hexamer Primer (at 100 µM concentration, Invitrogen, Thermo Fisher Scientific), 1 µl of dNTP mix (10 mM concentration each). Then we synthesised cDNA within reactions containing 4 µl 5x First Strand Buffer, 1 µl 0.1 M DTT, 1 µl RNase inhibitor (RNaseOUT, 40U/µl, Invitrogen, Thermo Fisher Scientific), 1 µl of SuperScript III Reverse Transcriptase (200U/µl, Invitrogen, Thermo Fisher Scientific) and 13 µl of denatured RNA samples. This reaction was incubated first for 5 min at 25°C, then for 45 min at 50°C and finally for 15 min at 70°C to inactivate reverse transcriptase.

Amplification inhibitors DNase-treated samples were removed by precipitation in 5 M ammonium acetate and absolute ethanol, using glycogen as an RNA carrier. RNA was amplified using the MessageAMP II aRNA Amplification Kit (Invitrogen, Thermo Fisher Scientific) according to manufacturer's recommendations, extending the in vitro transcription step to 14 h. Amplified samples were dissolved in 100 µl of nuclease-free water and frozen in -80°C. Amplified mRNA was sent to GATC (Constance, Germany) where the libraries were prepared and sequenced on Illumina HiSeq 2500/4000, producing 150 bp paired-end reads.

Data processing

Data processing of generated sequences was divided into four steps: pre-processing, metatranscriptomes co-assembly, mapping of reads from individual metatranscriptomes onto the assembly and finally annotation of assembled transcript isoforms (Figure 2). Quality control of the data was assessed with FastQC (0.11.5, Andrews, 2010). The pre-processing step aimed to remove unwanted sequences from the metatranscriptomes. First, Illumina adapters were removed using BBDuk v. 37.36 (Bushnell, 2017). Overrepresented sequences in each metatranscriptome reported by FastQC, consisting predominantly of

poly(A) and poly(T) fragments, were removed with BBDuk. The same software was used to remove PhiX control reads. Although we used poly(A) selection to capture only eukaryotic mRNA during sample preparation, rRNA may remain in the samples (Zhao et al., 2014). Thus, we used SortMeRNA 2.0 (Kopylova et al., 2012) to remove sequences that mapped to rRNA. Lastly, BBDuk was used to remove sequences of quality score <20 and reads <25 bp (because the next step by default uses k-mers of that length).

Detailed statistics on the initial library size of each metatranscriptome, and its change after each of the pre-processing steps can be found in Supplement 1. The 11 metatranscriptomes, containing jointly ~328 million read pairs, were de novo co-assembled using Trinity (Grabherr et al., 2011; Haas et al., 2013). Digital normalization step removed 10 million read pairs with a median k-mer abundance of <2 (--min_cov 2) and >50 (--max_cov 50) prior to the co-assembly. The initial assembly step of Trinity – Inchworm, ran on 199 million of all read pairs with no further normalisation. The assembled output ran through the remaining part of the co-assembly, first constructing de Bruijn graphs (Chrysalis step) and then resolving them (Butterfly step). Expression levels were estimated by mapping clean reads against the co-assembly in RSEM 1.3.0 (B. Li & Dewey, 2011). Due to varying numbers of reads in each of the metatranscriptomes (Supplement 2) and to assure between-sample comparison (Conesa et al., 2016) we used a relative measure of transcripts per million reads (TPM).

Annotations

De novo assembly produced 12 245 433 transcript isoforms, with clean reads mapping at least once to 11 010 859 isoforms. Most transcripts were characterized by low sum of relative abundance across samples (8 transcripts with >10 000 TPM, 154 with >1000 TPM, 3483 with >100 TPM, 68 166 with > 10 TPM and 2 390 862 with >1TPM; Supplementary 3). To increase the robustness of analyses and avoid stochasticity due to low abundance transcripts, further analyses were carried out on a core dataset that contained 68166 of the most abundant transcript isoforms for which the sum of TPM in all the samples was greater than 10 (Figure 2; from now on we will refer to transcript isoforms as transcripts).

Coding regions were predicted using TransDecoder 5.1.0. The core dataset was functionally annotated using Trinotate 3.3.1 with default parameters (Bryant et al., 2017). Similarities between de novo assembled transcripts/predicted coding regions were assessed against the UniProt database UniProt (The UniProt Consortium, 2017) using blastx/blastp, both with max_target_seq = 1 and e-value 1e-3 (BLAST+; Camacho et al., 2009). Protein domains were identified with HMMER3 (Mistry et al., 2013) against the Pfam database (31.0 release;

Finn et al., 2016). Functional annotations were retrieved based on blast or Pfam results and identified protein domains using eggNOG 3.0 (Powell et al., 2012), The Gene Ontology (GO; (Ashburner et al., 2000) and KEGG (Kanehisa & Goto, 2000; Kanehisa et al., 2016) within Trinotate. We focused on the most abundant GO terms dataset corresponding to biological processes, molecular functions and cellular compartments featuring an arbitrary value of > 5000 TPM for each GO term.

Taxonomy was assigned to clean reads using the TaxMapper search tool and corresponding database with default settings (Beisser et al., 2017). Reads were mapped to two taxonomic levels: seven main eukaryotic lineages (supergroups; e.g. Alveolata) and 28 groups within these lineages (e.g. Apicomplexa, Chromerida, Ciliophora, Dinophyta and Perkinsea within Alveolata supergroup).

To validate the process of de novo assembly, we mapped transcripts in core dataset against metatranscriptomics data from *Tara* Oceans expeditions, including *Tara* Oceans Polar Circle sampled in 2013 (Carradec et al., 2018, unpublished). The reads mapping pipeline used is the same as described in Carradec et al. 2018. Briefly, reads from each *Tara* metatranscriptomic readset were mapped onto transcript isoforms in the core dataset using bwa (H. Li, 2013) using 95% identity over at least 80% of the length of the read, random best match mode).

Statistical methods

All statistical analyses were performed in R v3.5.2 (R Core Team, 2018), and most of the data were visualised using *tidyverse* v1.2.1 (Wickham 2019). Principal component analysis (PCA) was calculated on centred and scaled data with *prcomp* function (*stats* package v3.5.3) and visualised using *factoextra* v1.0.5. To explore differences between transcript abundances a Bray-Curtis dissimilarity matrix (*vegdist* function in *vegan* package v2.5-4; Oksanen et al., 2019) was constructed and clustered using a 'complete' method within *hclust* function (*stats* package v3.5.3). *Pvclust* package was used to assign support to the clustering topology (Suzuki & Shimodaira, 2006). To identify the strongest contribution of individual transcript isoforms to clustering patterns, we applied a *simper* function on the transcript matrix.

To explore GO annotations, for each metaT, we summarised relative counts for each transcript that was assigned to a specific GO term. We explored each of the three categories of GO terms: molecular functions, biological processes and cellular components. For each category, a Bray-Curtis dissimilarity matrix of GO abundance tables was used to calculate global non-metric multidimensional scaling (GNMDS; Kruskal, 1964). Then *envfit* function

(*vegan* package) was used to fit environmental parameters onto the GNMDS ordination. Analysis of similarities (*ANOSIM*; *vegan* package) was used to test if there are differences among our samples associated with light (polar day and polar night). The simper function (*vegan* package) was used with 999 permutations to elucidate GO terms that contributed the most to the difference between polar night and polar day within the 3 GO categories. In this analysis, the September sample was excluded due to mixed light conditions in the middle of a transition between polar day and polar night. *Simper* analysis identified GO terms that differed between polar day and polar night; we called them overrepresented if the differences in means were statistically significant. Subsequently, all the GO annotations within biological processes containing strings ‘light’ or ‘photo’ (except ‘flight’, ‘flight response’ and ‘nonphotochemical quenching’) were extracted together with their counts and summarised for polar day and polar night samples. Scripts used to analyse and visualise the data are available at https://github.com/magdawutkowska/metaT.

RESULTS

Seasonality

Our study spanned over 13 months and included two polar nights (three and two samples respectively), one polar day (five samples) and one sample from September coinciding with the transition period between polar day and polar night. Environmental parameters showed a seasonal pattern (Table 1, Supplement 2), that is representative for the IsA time series station in a yearly recurrent pattern (data not shown). Photosynthetically active radiation (PAR) at 25 m depth was only detectable between April and September 2012. The highest values were measured in April and beginning of May 2012, followed by the lowest detected values in the end of May and June. Hydrography of arctic fjords can be influenced by water masses originating from distinct sources and thus displaying different physiochemical properties and categorised mainly based on temperature and salinity according to Svendsen et al., 2002 and Nilsen et al., 2008. Locally formed cold (< 1° C) water masses (LW) were present in the first half of the year (December 2011 to May 2012) with warmer (> 1° C) intermediate water (IW) influenced by land runoff and oxygen-rich Atlantic water in the second half (Jun 2012 to Jan 2013). The coldest temperature was in January 2012 and the warmest in September 2012. Overall, nutrient concentrations (nitrate/nitrite, phosphate and silicates) were heavily depleted from the onset of spring bloom until the end of polar day (from May to August; Table 1). Silicates, however, started to be depleted already in April (Table1). Chlorophyll a was detectable throughout the year with a peak value in May and a second smaller peak in August. In all samples except those collected in May, most of the chlorophyll a was produced by the small phytoplankton fraction (< 10 µm). Detailed description of the system, based on enhanced frequency of sampling can be found in (Brandner et al., 2017; Kubiszyn et al., 2017; Marquardt et al., 2016; Stübner et al., 2016).

Seasonal transcript diversity

The diversity and evenness of transcript isoforms in a total dataset was higher during polar night (n=5) than during polar day (n=5), (Figure 3). The mean number of transcripts collected during polar day was similar to the value obtained from mixed light regime in September (μ_{PD}=1 178 988, σ_{PD}=273 108 and 1 272 116, respectively), whereas average diversity during polar night was 2.7 times higher. However, the January 2012 sample outlied significantly from the other polar night metatranscriptomes, containing 1.6 million transcript isoforms, a similar value to samples from polar day and September. Both the September 2012 and

January 2012 samples that had low numbers of transcripts also had significantly lower depth of sequencing than the other samples (Supplement 1).

We found a clear difference between metatranscriptomes from polar day and polar night with the mixed light regime sample in September clustering with the polar night samples with high support of > 99% of both unbiased and bootstrap probability (Figure 4, not all data presented). The polar day formed distinct subclusters (Figure 4). Subsampled transcript datasets showed similar or identical clustering patterns, indicating these were not altered by the high contribution of rare transcripts, therefore further functional descriptions were conducted using the core dataset (Figure 2 and 5).

We identified the transcripts with the strongest contribution to the differences between the main clusters (Figure 4). Ten of the transcripts contributing to the difference between polar night and polar day were also the most abundant transcripts in our dataset. Only the most abundant transcript out of ten got a functional annotation, and was classified as cytochrome b (Supplement 3).

Taxonomic composition

The ratio of reads that could be assigned to taxonomic groups was similar throughout the year, ranging from 33 to 42% of all reads in each metaT. This left the majority of reads without an annotation (58-67%). The proportion of unannotated reads was independent of light regime and number of sequences per sample. The most represented supergroup in each sample was Alveolata, predominantly Dinophyceae and Ciliophora (Figure 8). Dinophyceae dominated in metatranscriptomes from polar night (32% on 17th January 2012 up to 49% on 14th December 2011) and September (33%), while ciliates were more abundant during polar day (18-34% versus 8-10% in polar night). Many taxonomic groups had low relative transcript abundance throughout the year, never exceeding 2% of the taxonomically assigned reads (Apusozoa, Bigyra, Cercozoa, Chromerida, Euglenozoa, Fornicata, Fungi, Glaucocystophyceae, Heterolobosea, Parabasalia, Perkinsea, Pseudofungi and Rhodophyta).

Activity of expressed genes in a seasonal perspective

Annotation of the core dataset against the GO database resulted in 24,643 transcripts with at least one annotation (36% of core dataset.) Environmental variables fitted into biological processes (GO category) dissimilarity matrix confirmed the importance of light as a structuring parameter (i.e. day length (R^2_{GNMDS}=0.88, p=0.019), declination (R^2_{GNMDS}=0.85, p=0.025) and PAR (R^2_{GNMDS}=0.54, p=0.082).On the other hand, the analysis did not support

water mass (R^2_{GNMDS}=0.04, p=0.974) and temperature (r^2=0.20, p=0.475) as important explanatory variables in structuring biological processes.

The most abundant GO terms within **biological processes** belonged to housekeeping genes encoding proteins involved in translation, microtubule-based process, respiratory electron transport chain or protein folding etc. (Figure 5A). The majority of the most abundant biological processes were overrepresented in polar day samples, such as respiratory electron transport chain or cytoplasmic translation (Figure 5A). Some of the GO terms were more uniformly distributed throughout the year, such as cell or mitotic nuclear division (Figure 5A). Finally, a few of the most abundant GO terms were overrepresented during polar night. This was the case for one-carbon metabolic processes (mean number of TPM in polar night samples, μ_{PN}=1974, in polar day μ_{PD}=1134, p=0.03), response to stress (μ_{PN}=1482 in polar night samples, while in polar day μ_{PD}=498, p=0.01) and phototransduction (μ_{PN}=936 in polar night samples, while in polar day μ_{PD}=323, p=0.03). The majority of transcripts within one-carbon metabolic process mapped to adenosylhomocysteinase and S-adenosylmethionine synthase. The latter catalyses hydrolysis of L-methionine into S-adenosyl-L-methionine which is an essential source of different chemical groups, e.g. methyl groups used for epigenetic modifications including DNA methylation (Cantoni, 1975; Fontecave et al., 2004). Whereas adenosylhomocysteinase catalyses one of the next reactions in methionine metabolism: hydrolysis of S-adenosyl-L-homocysteinase to adenosine and L-homocysteine (De La Haba & Cantoni, 1959) and connected to silicon (Hildebrand et al., 1993) and vitamin starvation in diatoms (Alexander et al., 2015). All transcripts in response to stress mapped to chaperone proteins, most (451 out of 456) mapped to different types of heat shock proteins, especially HSP90 (406 transcript isoforms).

Almost all light-dependent biological processes were relatively more abundant in polar day samples (Figure 7). This was especially true for GO terms connected to photosynthesis. However, most of the categories were also present during polar night albeit in low numbers. Three of the terms were more abundant in polar night, such as eye photoreceptor cell development, phototaxis and predominantly phototransduction. Phototransduction contained 208 transcripts mapping to green- and blue-light absorbing proteorhodopsins.

The majority of transcripts contributing to less abundant GO terms, but overrepresented during polar night (Figure 6), mapped to multipurpose proteins, mainly chaperones (HSP72 and HSP71 in protein refolding, HSP72 in negative regulation of cellular response to heat or response to virus). Phagocytosis and response to other organism categories consisted

mostly of transcripts assigned to calreticulin, a multipurpose protein acting as calcium-level regulator and chaperone in endoplasmic reticulum (Michalak et al., 1999). Pathogenesis contained mostly tripeptidyl-peptidase transcripts and acidic proteases probably involved in virulence (Reichard et al., 2006). Response to cycloheximide, a naturally occurring fungicide, contained transcripts mapping to 60S ribosomal protein L44.

Most GO terms within **molecular functions** were overrepresented in polar day (Figure 5B). Analyses indicated light, but not water masses, to be an important structuring factor of the most abundant molecular functions of the community ($R2_{GNMDS}$=0.795, p=0.005 versus $R2_{GNMDS}$=0.017, p=0.897, respectively). Only DNA binding (μ_{PN}=6766 and μ_{PD}=4714, p=0.024), adenosylhomocysteinase activity (μ_{PN}=1585 and μ_{PD}=889, p=0.017), photoreceptor (μ_{PN}=936 and μ_{PD}=315, p=0.055) and light-activated ion channel activity (μ_{PN}=883 and μ_{PD}=247, p=0.025) were overrepresented in polar night. DNA binding is a broad category of gene products that were identified as reacting selectively in a non-covalent manner with DNA. We identified 1651 transcripts containing mostly major basic nuclear proteins, histones, cold shock proteins etc. Light-activated ion channel and photoreceptor consisted mostly of identified proteorhodopsins; additionally, photoreceptor contained also transcripts mapping to centrins. Centrins are calcium-binding proteins involved in centrosome and microtubule functioning (Satisbury, 1995), as well as regulation of signalling and molecular translocation (Gießl et al., 2006). Among less abundant molecular functions overrepresented during polar night, we found that the transcripts mapped mostly to multipurpose proteins, similarly to biological processes. Chromatin binding consisted of diverse proteins, and the majority of transcripts mapped to 60S ribosomal protein L29. Fumarate reductase (NADH) activity term consisted of transcripts mapping to an enzyme that catalyses reversible anaerobic reduction of succinate to fumarate that generates NADH and protons (Tielens & Van Hellemond, 1998). Sialic acid, phospholipase and oligosaccharide binding contained transcripts mapping mainly to e-selectin, protein involved in inflammatory response that change properties of cell surface (Bhatia et al., 2003). Mapping to heat shock-related 70kDa proteins was found in glycolipid binding, whereas nucleolin in nucleosome binding. Nucleolins are also plurifuctional proteins playing important roles in viral infections (Hiscox, 2002). 17-beta-hydroxysteroid dehydrogenase (NAD+) activity and 3alpha,7alpha,12alpha- trihydroxy-5beta-cholest-24-enoyl-CoA hydratase activity contained the same transcript isoforms that mapped to peroxisomal multifunctional enzymes taking part in β-oxidation of lipids (Winkler et al., 2003) but could also be necessary in fungal pathogenesis (Klose & Kronstad, 2006).

The strongest differences in transcripts within the three GO categories between polar day and polar night was found in cellular compartment (R_{ANOSIM}=0.928, p=0.01); whereas the differences were less pronounced in biological processes (R_{ANOSIM}=0.792, p=0.008) and molecular functions: (R_{ANOSIM}=0.892, p=0.009). At the same time, we did not find any cellular compartments featuring the most abundant transcripts, that would be overrepresented during polar night (Figure 5C). Less abundant GO terms pointed out compartments overexpressed during polar night that are connected to cytoskeleton (spindle pole, cell cortex, cortical skeleton, filopodium) or cell membrane (coated pit, cell cortex, caveola; Figure 6). Some names of categories are misleading, such as male germ cell nucleus (containing the peroxisomal multifunctional enzymes mentioned in molecular functions) or blood microparticle (containing mostly actins or signalling molecules). The spindle pole term consisted of transcripts that mapped to centrins, mentioned in the photoreceptor term in molecular functions. Cell cortex region lies beneath the cell membrane and contained transcripts that mapped to myosins, 14-3-3-like proteins, profilins and other cytoskeleton related proteins.

Transcript novelty

Levels of functional annotation were overall low, regardless of the database used. Mapping to UniProt (with blastp), Pfam, TmHMM, GO (based on Pfam) resulted in <10% of transcript annotation, while eggnog and Kegg gave 10-20% annotation of success. Only UniProt (with blastx) and GO (with blastp) annotated 38% and 36% of transcripts, respectively. Assembled transcripts mapped against *Tara* Oceans datasets showed that most of our transcripts had hits, matching especially samples from the Arctic (Figure 9). Up to 75% of our transcript isoforms mapped to surface samples (station 196, north of Alaska), up to 78% mapped to the deep chlorophyll maximum layer (station 173, north east of Novaya Zemlya), and up to 74% to the mesopelagic zone (station 201 in west part of Baffin Bay). Mean proportion of transcripts mapping to surface samples from *Tara* Oceans stations located above 60°N differed from these in temperate and tropical regions ($\mu_{\uparrow 60N}$=64%, $\sigma_{\uparrow 60N}$=9% and $\mu_{\downarrow 60N}$=21%, $\sigma_{\downarrow 60N}$=8%, respectively). Mean proportion of transcripts mapping to deep chlorophyll maximum depth was also higher above 60°N compared to these below 60°N ($\mu_{\uparrow 60N}$=69%, $\sigma_{\uparrow 60N}$=12% and $\mu_{\downarrow 60N}$=23%, $\sigma_{\downarrow 60N}$=9%, respectively). In case of mesopelagic depth 67% ($\sigma_{\uparrow 60N}$=6%) of transcript isoforms mapped to samples from high latitudes.

DISCUSSION

Climate change is already influencing Arctic marine ecosystems (Wassmann et al., 2011) and different scenarios of the development of Arctic marine ecosystems have been suggested. However, predictions of its influence on polar ecosystems are challenging without a deep understanding of both structure and function of its parts (Murphy et al., 2016). Thus, responses of microbial communities to these changes cannot be predicted without understanding which biological and molecular activities are taking place and how they impact biogeochemical cycles. Differences in gene expression could change the outcome of trophic interactions in an ecosystem, potentially altering the energy and nutrient flow to higher trophic levels (Murphy et al., 2016; Worden et al., 2015). In this study we went beyond reporting detected species or its molecular proxies to look into community-level molecular engagement in biological activities. Our study offers a first description of the key processes performed by the microbial eukaryotic community over seasons.

The strong seasonality at high latitudes affects microbial eukaryotes by influencing cells counts, biomass distribution, community composition, dominating carbon acquisition mode and biodiversity measures (Kubiszyn et al., 2017; Marquardt et al., 2016), and as a consequence the overall pool of available genes and their products. Higher OTU richness of marine protists during polar night compared to polar day seems to be a panarctic phenomenon (Marquardt et al., 2016; Onda et al., 2017), and was also observed for marine bacteria and archaea (Ladau et al., 2013). In line with these findings, we showed that diversity and evenness of transcripts were higher during polar night, possibly reflecting either higher taxonomic diversity of microbial eukaryotes or an increased functional diversity, i.e. expression of different genes needed for survival. The proportions of transcripts belonging to predominantly photosynthetic protists such as diatoms, haptophytes and chlorophytes, were consistently lower during polar night, confirming lower representation in the community and perhaps also lower overall activity (Marquardt et al., 2016). Despite high diversity of OTUs and transcripts during polar night, cell counts, and therefore also biomass of protists at that time are low (Kubiszyn et al., 2017; Rokkan Iversen & Seuthe, 2011; Seuthe et al., 2011).

During polar night the contribution of single species to the overall low pool of biomass is more equal than at any other times of the year (Marquardt et al., 2016). This includes crucial primary producers, such as *Micromonas polaris*, which were encountered as active at

different depths of Arctic marine habitats during polar night (Vader et al., 2014). Persistence of low levels of light-dependent biological processes in primary producers during polar night is likely due to the persistence and perhaps maintenance of a functional photosynthetic apparatus ready to be activated once the light comes back (Kvernvik et al., 2018; van der Poll et al., 2020). Therefore an overrepresentation of eukaryotic proteorhodopsins during polar night was rather unexpected, as they contribute to an alternative pathway to photosynthesis being the main contributors to harnessing solar energy in the ocean by bacteria (Gómez-Consarnau et al., 2019). It is not clear what is their function in microbial eukaryotes such as dinophytes (Slamovits et al., 2011). However, recently, it was suggested that they are involved in G protein-coupled receptor-based signalling in Dinophyta (Mojib & Kubanek, 2020).

Gene expression during polar night is probably more strictly controlled in many organisms due to overall lower availability of energy in the ecosystem (Berge, Renaud, et al., 2015). Increased expression of histones or major binding nuclear proteins or similar genes could serve as a way to control gene expression by binding and therefore preventing DNA from being transcribed (Prado et al., 2017). On the other hand, it might also point towards cellular division and the need to produce new histones for new cells (Salomé & Merchant, 2019). GO term classification of polar night overrepresented transcripts mostly covers response to stress, cellular signalling, modifications in cytoskeleton, pathogenesis etc through multifunctional proteins. Multifunctionality might be an efficient strategy of effective use of resources that could be limited for some groups of organisms during polar night. Other polar night overrepresented functions that involve adenosylhomocysteinase could play an important role in increasing lifespan of microbial eukaryotes by controlling the concentration of methionine (Parkhitko et al., 2019). In general, biochemical reactions involved in methionine degradation are the main source of methyl groups used in gene silencing by DNA methylation which could possibly be another argument for strict control of gene expression during polar night (Parkhitko et al., 2019). Overrepresentation of different types of chaperon and heat shock protein transcripts during polar night may be connected to high importance of energy conservation by assuring correct assembly and maintenance and stability of protein structure (Balchin et al., 2016; Frydman, 2001). Moreover, heat shock proteins could influence increased survival by several mechanisms able to attenuate apoptosis (Verbeke, 2001).

Among the most expressed transcripts in our study were few functionally annotated sequences reaching up to 38% of the total number of transcripts which coincides with similar numbers of taxonomically annotated reads in our study. Metatranscriptomic studies often report low levels of functional annotations (down to 19%) that might be a result of various factors, such as complexity of the studied environment (Jiang et al., 2016), available reference databases (Carradec et al., 2018), choice of algorithms, bioinformatic tools and parameters used for data analysis (e.g. Celaj et al., 2014), etc. To date, the biggest metatranscriptomics marine global survey examining expressed eukaryotic genes based on *Tara* Oceans 2009-2012) reported 51.2% unannotated clusters of expressed genes (Carradec et al., 2018). Although overall rates of annotation in our study were low, the data mapped successfully against the *Tara* Oceans dataset (including Polar Circle campaign in 2013) by identifying up to 78% of transcript isoforms, specifically in polar regions (Figure 9). Therefore, we concluded that *de novo* assembled transcripts in our bioinformatic pipeline were robust and 22% of novel genes that are less likely to be found in lower latitudes (Figure 9).

The similarity of average difference between proportion of transcripts mapping to different depths in *Tara* Oceans metatranscriptomes suggests a distinct genetic makeup of microbial eukaryotes' in the Arctic. It corresponds with classification of global marine bacterial biogeography, where high latitude and deep waters were more similar than mid-latitude surface samples (Ghiglione et al., 2012) and moreover, similar to biogeography of DNA viruses and their biodiversity hotspot in the Arctic Ocean (Gregory et al., 2019). Viruses and mobile elements carried by them are known as powerful agents of evolution in all living cells (Koonin, 2016), thus they could potentially contribute to increased diversification of genes in microbial communities in the Arctic (Anesio & Bellas, 2011). Several categories of biological processes overrepresented during polar night, such as DNA recombination, phagocytosis/engulfment, response to other organisms and viruses, could perhaps suggest that viral infections are of high importance particularly at that time of the year. Regardless of possible links between distinct populations of viruses and microbial eukaryotes, we hypothesize that the proportion of successfully mapping transcripts in our study would have been higher if the *Tara* Oceans campaign in the Arctic was extended beyond June-October to collect samples during polar night.

Polar night seems to work as a reset stage for Arctic marine environments, possibly enforcing shifts to heterotrophy in the absence of light and allowing protist survival as low

biomass populations. Moreover, changing proportions of transcripts annotated to taxonomic groups as well as fluctuating abundances of functional categories point out that community-level metabolic state changes together with shifting community composition. The two polar nights in our study showed a striking similarity in taxonomic and functional composition of transcripts that might reflect a specific, recurrent impact of environmental filtering imposed by seasonal light regime. Perhaps the different genetic makeup of eukaryotic communities in high latitudes could reflect necessary adaptations to Arctic seasonality that are not present in potential invasive microbial eukaryotes moving northwards with climate change induced shifts in environmental gradients. A long-term monitoring of taxonomic and transcriptional dynamics could evaluate to which extent other factors such as inflow of warmer water masses or arrival of species moving northwards, influence the strength of light regime filtering and development of future eukaryotic communities.

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. *Proceedings of the National Academy of Sciences*, *112*(44), E5972–E5979. https://doi.org/10.1073/pnas.1518165112

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anesio, A. M., & Bellas, C. M. (2011). Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends in Microbiology*, *19*(2), 52–57. https://doi.org/10.1016/j.tim.2010.11.002

Balchin, D., Hayer-Hartl, M., & Hartl, F. U. (2016). In vivo aspects of protein folding and quality control. *Science*, *353*(6294), aac4354. https://doi.org/10.1126/science.aac4354

Beisser, D., Graupner, N., Grossmann, L., Timm, H., Boenigk, J., & Rahmann, S. (2017). TaxMapper: An analysis tool, reference database and workflow for metatranscriptome analysis of eukaryotic microorganisms. *BMC Genomics*, *18*(1), 787. https://doi.org/10.1186/s12864-017-4168-6

Berge, J., Daase, M., Renaud, P. E., Ambrose, W. G., Darnis, G., Last, K. S., Leu, E., Cohen, J. H., Johnsen, G., Moline, M. A., Cottier, F., Varpe, Ø., Shunatova, N., Bałazy, P., Morata, N., Massabuau, J.-C., Falk-Petersen, S., Kosobokova, K., Hoppe, C. J. M., … Callesen, T. A. (2015). Unexpected Levels of Biological Activity during the Polar Night Offer New Perspectives on a Warming Arctic. *Current Biology*, *25*(19), 2555–2561. https://doi.org/10.1016/j.cub.2015.08.024

Berge, J., Renaud, P. E., Darnis, G., Cottier, F., Last, K., Gabrielsen, T. M., Johnsen, G., Seuthe, L., Weslawski, J. M., Leu, E., Moline, M., Nahrgang, J., Søreide, J. E., Varpe, Ø., Lønne, O. J., Daase, M., & Falk-Petersen, S. (2015). In the dark: A review of ecosystem processes during the Arctic polar night. *Progress in Oceanography*, *139*, 258–271. https://doi.org/10.1016/j.pocean.2015.08.005

Bhatia, S. K., Swers, J. S., Camphausen, R. T., Wittrup, K. D., & Hammer, D. A. (2003). Rolling Adhesion Kinematics of Yeast Engineered To Express Selectins. *Biotechnology Progress*, *19*(3), 1033–1037. https://doi.org/10.1021/bp025756b

Błachowiak-Samołyk, K., Wiktor, J. M., Hegseth, E. N., Wold, A., Falk-Petersen, S., & Kubiszyn, A. M. (2015). Winter Tales: The dark side of planktonic life. *Polar Biology*, *38*(1), 23–36. https://doi.org/10.1007/s00300-014-1597-4

Boyce, D. G., Petrie, B., Frank, K. T., Worm, B., & Leggett, W. C. (2017). Environmental structuring of marine plankton phenology. *Nature Ecology & Evolution*, *1*(10), 1484–1494. https://doi.org/10.1038/s41559-017-0287-3

Brandner, M. M., Stübner, E., Reed, A. J., Gabrielsen, T. M., & Thatje, S. (2017). Seasonality of bivalve larvae within a high Arctic fjord. *Polar Biology*, *40*(2), 263–276. https://doi.org/10.1007/s00300-016-1950-x

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., … Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports*, *18*(3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Bunse, C., & Pinhassi, J. (2017). Marine Bacterioplankton Seasonal Succession Dynamics. *Trends in Microbiology*, *25*(6), 494–505. https://doi.org/10.1016/j.tim.2016.12.013

Bushnell, B. (2017). *BBMap*. sourceforge.net/projects/bbmap/

Buttigieg, P. L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., & Boetius, A. (2018). Marine microbes in 4D — using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Current Opinion in Microbiology*, *43*, 169–185. https://doi.org/10.1016/j.mib.2018.01.015

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Cantoni, G. L. (1975). Biological Methylation: Selected Aspects. *Annual Review of Biochemistry*, *44*(1), 435–451. https://doi.org/10.1146/annurev.bi.44.070175.002251

Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy, C., Bell, C. J., Bharti, A., Dyhrman, S. T., Guida, S. M., Heidelberg, K. B., Kaye, J. Z., Metzner, J., Smith, S. R., & Worden, A. Z. (2017). Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews Microbiology*, *15*(1), 6–20. https://doi.org/10.1038/nrmicro.2016.160

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M.-A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., … Wincker, P. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*, *9*(1), 373. https://doi.org/10.1038/s41467-017-02342-1

Celaj, A., Markle, J., Danska, J., & Parkinson, J. (2014). Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome*, *2*(1), 39. https://doi.org/10.1186/2049-2618-2-39

Coles, V. J., Stukel, M. R., Brooks, M. T., Burd, A., Crump, B. C., Moran, M. A., Paul, J. H., Satinsky, B. M., Yager, P. L., Zielinski, B. L., & Hood, R. R. (2017). Ocean biogeochemistry modeled with emergent trait-based genomics. *Science*, *358*(6367), 1149–1154. https://doi.org/10.1126/science.aan5712

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13. https://doi.org/10.1186/s13059-016-0881-8

De La Haba, G., & Cantoni, G. L. (1959). The enzymatic synthesis of S-adenosyl-L-homocysteine

from adenosine and homocysteine. *The Journal of Biological Chemistry*, *234*(3), 603–608.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605–1261605. https://doi.org/10.1126/science.1261605
Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237
Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344
Fontecave, M., Atta, M., & Mulliez, E. (2004). S-adenosylmethionine: Nothing goes to waste. *Trends in Biochemical Sciences*, *29*(5), 243–249. https://doi.org/10.1016/j.tibs.2004.03.007
Frydman, J. (2001). Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annual Review of Biochemistry*, *70*(1), 603–647. https://doi.org/10.1146/annurev.biochem.70.1.603
Ghiglione, J.-F., Galand, P. E., Pommier, T., Pedros-Alio, C., Maas, E. W., Bakker, K., Bertilson, S., Kirchman, D. L., Lovejoy, C., Yager, P. L., & Murray, A. E. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences*, *109*(43), 17633–17638. https://doi.org/10.1073/pnas.1208160109
Gießl, A., Trojan, P., Rausch, S., Pulvermüller, A., & Wolfrum, U. (2006). Centrins, gatekeepers for the light-dependent translocation of transducin through the photoreceptor cell connecting cilium. *Vision Research*, *46*(27), 4502–4509. https://doi.org/10.1016/j.visres.2006.07.029
Gómez-Consarnau, L., Raven, J. A., Levine, N. M., Cutter, L. S., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J. A., Gasol, J. M., & Sañudo-Wilhelmy, S. A. (2019). Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances*, *5*(8), eaaw8855. https://doi.org/10.1126/sciadv.aaw8855
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., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883
Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., … Wincker, P. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, *177*(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040
Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084
Hildebrand, M., Higgins, D. R., Busser, K., & Volcani, B. E. (1993). Silicon-responsive cDNA clones isolated from the marine diatom Cylindrotheca fusiformis. *Gene*, *132*(2), 213–218. https://doi.org/10.1016/0378-1119(93)90198-C
Jiang, Y., Xiong, X., Danska, J., & Parkinson, J. (2016). Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome*, *4*(1), 2. https://doi.org/10.1186/s40168-015-0146-x

Judson, O. P. (2017). The energy expansions of evolution. *Nature Ecology & Evolution*, *1*(6), 0138. https://doi.org/10.1038/s41559-017-0138

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, *6*(6), 451–464. https://doi.org/10.1038/nrg1615

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., … 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 Biology*, *12*(6), e1001889. https://doi.org/10.1371/journal.pbio.1001889

Klose, J., & Kronstad, J. W. (2006). The Multifunctional β-Oxidation Enzyme Is Required for Full Symptom Development by the Biotrophic Maize Pathogen Ustilago maydis. *Eukaryotic Cell*, *5*(12), 2047–2061. https://doi.org/10.1128/EC.00231-06

Koonin, E. V. (2016). Viruses and mobile elements as drivers of evolutionary transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1701), 20150442. https://doi.org/10.1098/rstb.2015.0442

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Kubiszyn, A. M., Wiktor, J. M., Wiktor, J. M., Griffiths, C., Kristiansen, S., & Gabrielsen, T. M. (2017). The annual planktonic protist community structure in an ice-free high Arctic fjord (Adventfjorden, West Spitsbergen). *Journal of Marine Systems*, *169*, 61–72. https://doi.org/10.1016/j.jmarsys.2017.01.013

Kvernvik, A. C., Hoppe, C. J. M., Lawrenz, E., Prášil, O., Greenacre, M., Wiktor, J. M., & Leu, E. (2018). Fast reactivation of photosynthesis in arctic phytoplankton during the polar night [1]. *Journal of Phycology*, *54*(4), 461–470. https://doi.org/10.1111/jpy.12750

Ladau, J., Sharpton, T. J., Finucane, M. M., Jospin, G., Kembel, S. W., O'Dwyer, J., Koeppel, A. F., Green, J. L., & Pollard, K. S. (2013). Global marine bacterial diversity peaks at high latitudes in winter. *The ISME Journal*, *7*(9), 1669–1677. https://doi.org/10.1038/ismej.2013.37

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Marquardt, M, Skogseth, R., Wiedmann, I., Vader, A., Reigstad, M., Cottier, F., & Gabrielsen, T. (2019). Vertical export of marine pelagic protists in an ice-free high-Arctic fjord (Adventfjorden, West Spitsbergen) throughout 2011-2012. *Aquatic Microbial Ecology*, *83*(1), 65–82. https://doi.org/10.3354/ame01904

Marquardt, Miriam, Vader, A., Stübner, E. I., Reigstad, M., & Gabrielsen, T. M. (2016). Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). *Applied and Environmental Microbiology*, *82*(6), 1868–1880. https://doi.org/10.1128/AEM.03208-15

McKie-Krisberg, Z. M., & Sanders, R. W. (2014). Phagotrophy by the picoeukaryotic green alga Micromonas: Implications for Arctic Oceans. *The ISME Journal*, *8*(10), 1953–1961. https://doi.org/10.1038/ismej.2014.16

McMinn, A., & Martin, A. (2013). Dark survival in a warming world. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755), 20122909. https://doi.org/10.1098/rspb.2012.2909

Metfies, K., von Appen, W.-J., Kilias, E., Nicolaus, A., & Nöthig, E.-M. (2016). Biogeography and

Photosynthetic Biomass of Arctic Marine Pico-Eukaroytes during Summer of the Record Sea Ice Minimum 2012. *PLOS ONE*, *11*(2), e0148512. https://doi.org/10.1371/journal.pone.0148512

Michalak, M., Corbett, E. F., Mesaeli, N., Nakamura, K., & Opas, M. (1999). Calreticulin: One protein, one gene, many functions. *The Biochemical Journal*, *344 Pt 2*, 281–292.

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, *41*(12), e121–e121. https://doi.org/10.1093/nar/gkt263

Mojib, N., & Kubanek, J. (2020). Comparative transcriptomics supports the presence of G protein-coupled receptor-based signaling in unicellular marine eukaryotes. *Limnology and Oceanography*, *65*(4), 762–774. https://doi.org/10.1002/lno.11345

Moreira, D., & López-García, P. (2019). Time series are critical to understand microbial plankton diversity and ecology. *Molecular Ecology*, *28*(5), 920–922. https://doi.org/10.1111/mec.15015

Murphy, E. J., Cavanagh, R. D., Drinkwater, K. F., Grant, S. M., Heymans, J. J., Hofmann, E. E., Hunt, G. L., & Johnston, N. M. (2016). Understanding the structure and functioning of polar pelagic ecosystems to predict the impacts of change. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1844), 20161646. https://doi.org/10.1098/rspb.2016.1646

Nilsen, F., Cottier, F., Skogseth, R., & Mattsson, S. (2008). Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Svalbard. *Continental Shelf Research*, *28*(14), 1838–1853. https://doi.org/10.1016/j.csr.2008.04.015

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wegner, H. (2019). *vegan: Community Ecology Package* (Version 2.5-4) [Computer software]. https://CRAN.R-project.org/package=vegan

Onda, D. F. L., Medrinal, E., Comeau, A. M., Thaler, M., Babin, M., & Lovejoy, C. (2017). Seasonal and Interannual Changes in Ciliate and Dinoflagellate Species Assemblages in the Arctic Ocean (Amundsen Gulf, Beaufort Sea, Canada). *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00016

Parkhitko, A. A., Jouandin, P., Mohr, S. E., & Perrimon, N. (2019). Methionine metabolism and methyltransferases in the regulation of aging and lifespan extension across species. *Aging Cell*, *18*(6). https://doi.org/10.1111/acel.13034

Prado, F., Jimeno-González, S., & Reyes, J. C. (2017). Histone availability as a strategy to control gene expression. *RNA Biology*, *14*(3), 281–286. https://doi.org/10.1080/15476286.2016.1189071

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reichard, U., Léchenne, B., Asif, A. R., Streit, F., Grouzmann, E., Jousson, O., & Monod, M. (2006). Sedolisins, a New Class of Secreted Proteases from Aspergillus fumigatus with Endoprotease or Tripeptidyl-Peptidase Activity at Acidic pHs. *Applied and Environmental Microbiology*, *72*(3), 1739–1748. https://doi.org/10.1128/AEM.72.3.1739-1748.2006

Rokkan Iversen, K., & Seuthe, L. (2011). Seasonal microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): I. Heterotrophic bacteria, picoplankton and nanoflagellates. *Polar Biology*, *34*(5), 731–749. https://doi.org/10.1007/s00300-010-0929-2

Salomé, P. A., & Merchant, S. S. (2019). A Series of Fortunate Events: Introducing Chlamydomonas as a Reference Organism. *The Plant Cell*, *31*(8), 1682–1707. https://doi.org/10.1105/tpc.18.00952

Sanders, R. W., & Gast, R. J. (2012). Bacterivory by phototrophic picoplankton and nanoplankton in Arctic waters. *FEMS Microbiology Ecology*, *82*(2), 242–253. https://doi.org/10.1111/j.1574-6941.2011.01253.x

Satisbury, J. L. (1995). Centrin, centrosomes, and mitotic spindle poles. *Current Opinion in Cell Biology*, *7*(1), 39–45. https://doi.org/10.1016/0955-0674(95)80043-3

Schaub, I., Wagner, H., Graeve, M., & Karsten, U. (2017). Effects of prolonged darkness and temperature on the lipid metabolism in the benthic diatom Navicula perminuta from the Arctic Adventfjorden, Svalbard. *Polar Biology*, *40*(7), 1425–1439. https://doi.org/10.1007/s00300-016-2067-y

Seuthe, L., Rokkan Iversen, K., & Narcy, F. (2011). Microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): II. Ciliates and dinoflagellates. *Polar Biology*, *34*(5), 751–766. https://doi.org/10.1007/s00300-010-0930-9

Sherr, E. B., Sherr, B. F., Wheeler, P. A., & Thompson, K. (2003). Temporal and spatial variation in stocks of autotrophic and heterotrophic microbes in the upper water column of the central Arctic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *50*(5), 557–571. https://doi.org/10.1016/S0967-0637(03)00031-1

Slamovits, C. H., Okamoto, N., Burri, L., James, E. R., & Keeling, P. J. (2011). A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nature Communications*, *2*(1), 183. https://doi.org/10.1038/ncomms1188

Smayda, T. J., & Mitchell-Innes, B. (1974). Dark survival of autotrophic, planktonic marine diatoms. *Marine Biology*, *25*(3), 195–202. https://doi.org/10.1007/BF00394965

Stoecker, D. K., & Lavrentyev, P. J. (2018). Mixotrophic Plankton in the Polar Seas: A Pan-Arctic Review. *Frontiers in Marine Science*, *5*, 292. https://doi.org/10.3389/fmars.2018.00292

Stübner, E. I., Søreide, J. E., Reigstad, M., Marquardt, M., & Blachowiak-Samolyk, K. (2016). Year-round meroplankton dynamics in high-Arctic Svalbard. *Journal of Plankton Research*, *38*(3), 522–536. https://doi.org/10.1093/plankt/fbv124

Suzuki, R., & Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, *22*(12), 1540–1542. https://doi.org/10.1093/bioinformatics/btl117

Svendsen, H., Beszczynska-Møller, A., Hagen, J. O., Lefauconnier, B., Tverberg, V., Gerland, S., Ørbøk, J. B., Bischof, K., Papucci, C., Zajaczkowski, M., Azzolini, R., Bruland, O., Wiencke, C., Winther, J.-G., & Dallmann, W. (2002). The physical environment of Kongsfjorden?Krossfjorden, an Arctic fjord system in Svalbard. *Polar Research*, *21*(1), 133–166. https://doi.org/10.1111/j.1751-8369.2002.tb00072.x

The UniProt Consortium. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

Tielens, A. G. M., & Van Hellemond, J. J. (1998). The electron transport chain in anaerobically functioning eukaryotes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1365*(1–2), 71–78. https://doi.org/10.1016/S0005-2728(98)00045-0

Toseland, A., Daines, S. J., Clark, J. R., Kirkham, A., Strauss, J., Uhlig, C., Lenton, T. M., Valentin, K., Pearson, G. A., Moulton, V., & Mock, T. (2013). The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nature Climate Change*, *3*(11), 979–984. https://doi.org/10.1038/nclimate1989

Vader, A., Marquardt, M., Meshram, A. R., & Gabrielsen, T. M. (2014). Key Arctic phototrophs are widespread in the polar night. *Polar Biology*, *38*(1), 13–21. https://doi.org/10.1007/s00300-014-1570-2

van der Poll, W. H., Abdullah, E., Visser, R. J. W., Fischer, P., & Buma, A. G. J. (2020). Taxon‐specific dark survival of diatoms and flagellates affects Arctic phytoplankton composition during the polar night and early spring. *Limnology and Oceanography*, *65*(5), 903–914. https://doi.org/10.1002/lno.11355

Vaulot, D., Olson, R. J., & Chisholm, S. W. (1986). Light and dark control of the cell cycle in two marine phytoplankton species. *Experimental Cell Research*, *167*(1), 38–52. https://doi.org/10.1016/0014-4827(86)90202-8

Verbeke, P. (2001). Heat shock response and ageing: Mechanisms and applications. *Cell Biology International*, *25*(9), 845–857. https://doi.org/10.1006/cbir.2001.0789

Walter, B., Peters, J., & van Beusekom, J. E. E. (2017). The effect of constant darkness and short

light periods on the survival and physiological fitness of two phytoplankton species and their growth potential after re-illumination. *Aquatic Ecology*, *51*(4), 591–603. https://doi.org/10.1007/s10452-017-9638-z

Wassmann, P., Duarte, C. M., Agustí, S., & Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem: FOOTPRINTS OF CLIMATE CHANGE. *Global Change Biology*, *17*(2), 1235–1249. https://doi.org/10.1111/j.1365-2486.2010.02311.x

Wickham, H., & RStudio. (2019). *tidyverse: Easily Install and Load the 'Tidyverse'* (Version 1.3.0) [Computer software]. https://CRAN.R-project.org/package=tidyverse

Wiedmann, I., Reigstad, M., Marquardt, M., Vader, A., & Gabrielsen, T. M. (2016). Seasonality of vertical flux and sinking particle characteristics in an ice-free high arctic fjord—Different from subarctic fjords? *Journal of Marine Systems*, *154*, 192–205. https://doi.org/10.1016/j.jmarsys.2015.10.003

Winkler, U., Säftel, W., & Stabenau, H. (2003). A New Type of a Multifunctional β-Oxidation Enzyme in Euglena. *Plant Physiology*, *131*(2), 753–762. https://doi.org/10.1104/pp.013151

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, *347*(6223), 1257594–1257594. https://doi.org/10.1126/science.1257594

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D., & Perou, C. M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*, *15*(1), 419. https://doi.org/10.1186/1471-2164-15-419

Acknowledgments

The authors are thankful to all the people that contributed to the marine part of MicroFun project, collected and processed the samples. This research was funded by University Centre in Svalbard, as well as ConocoPhillips and Lundin Petroleum through The Northern Area Program and by the Norwegian Research Council, project number 230970. The cost of sequencing was covered through Jan Christiansen Legacy. This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian Metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. The publication charges for this paper have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Additional Information

The already published datasets are available online. All the other datasets used in this study are available at zenodo.com/addproperlink. Scripts generated for bioinformatic and statistical analysis are available at https://github.com/magdawutkowska/metaT.

Figure 1
Location of the Isfjorden Adventfjorden (IsA) time series station in Svalbard.

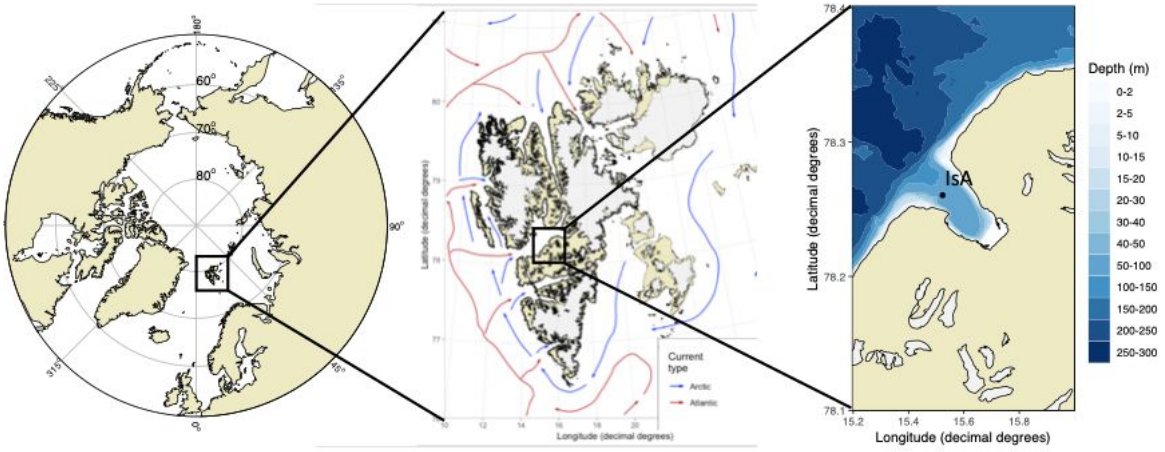


Figure 2
Sequencing data processing workflow. A separate metatranscriptome (metaT) was generated from each sample.

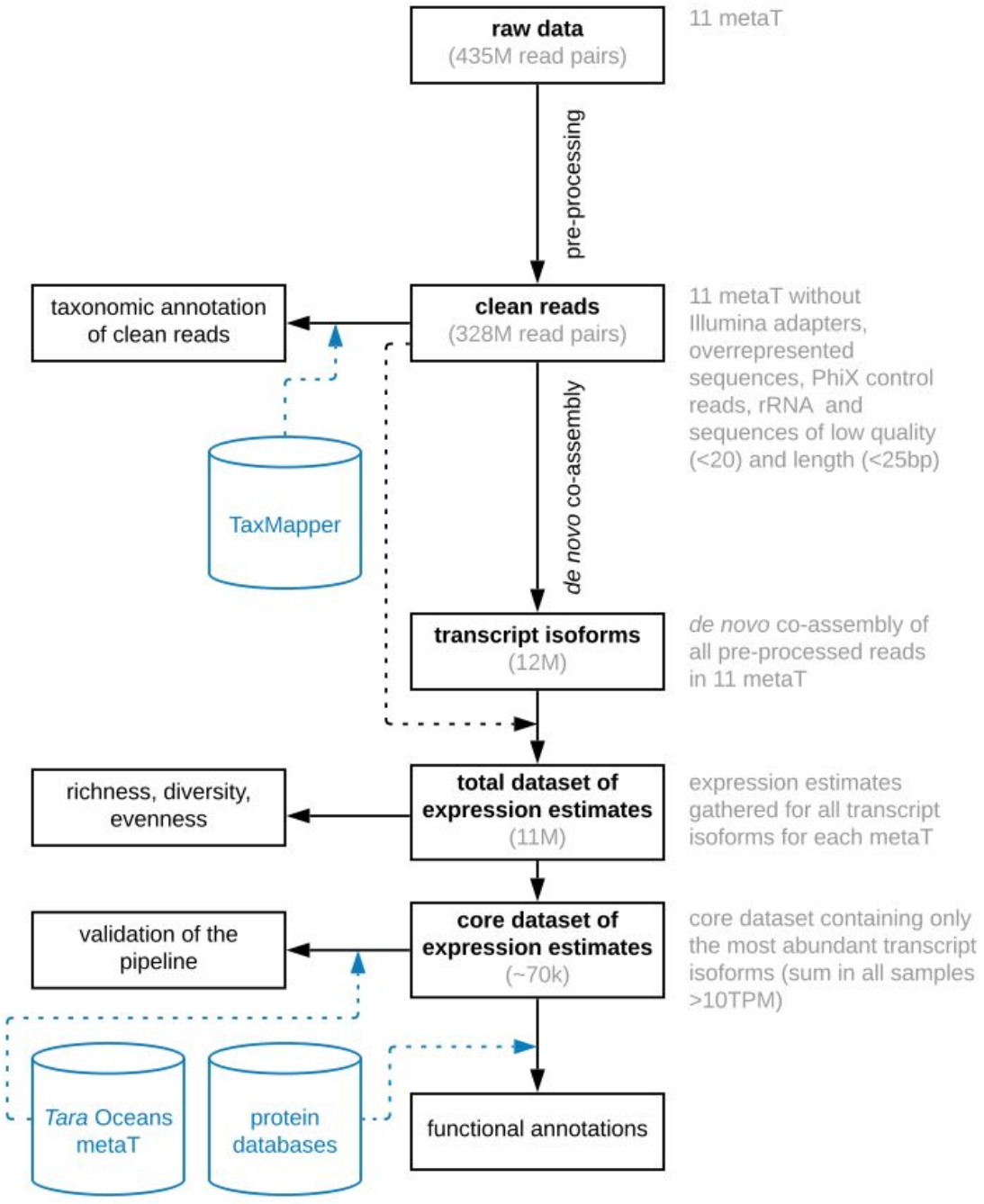


Figure 3
Diversity of transcript isoforms per sample during polar day (n=5) and polar night (n=5). The September sample was excluded.

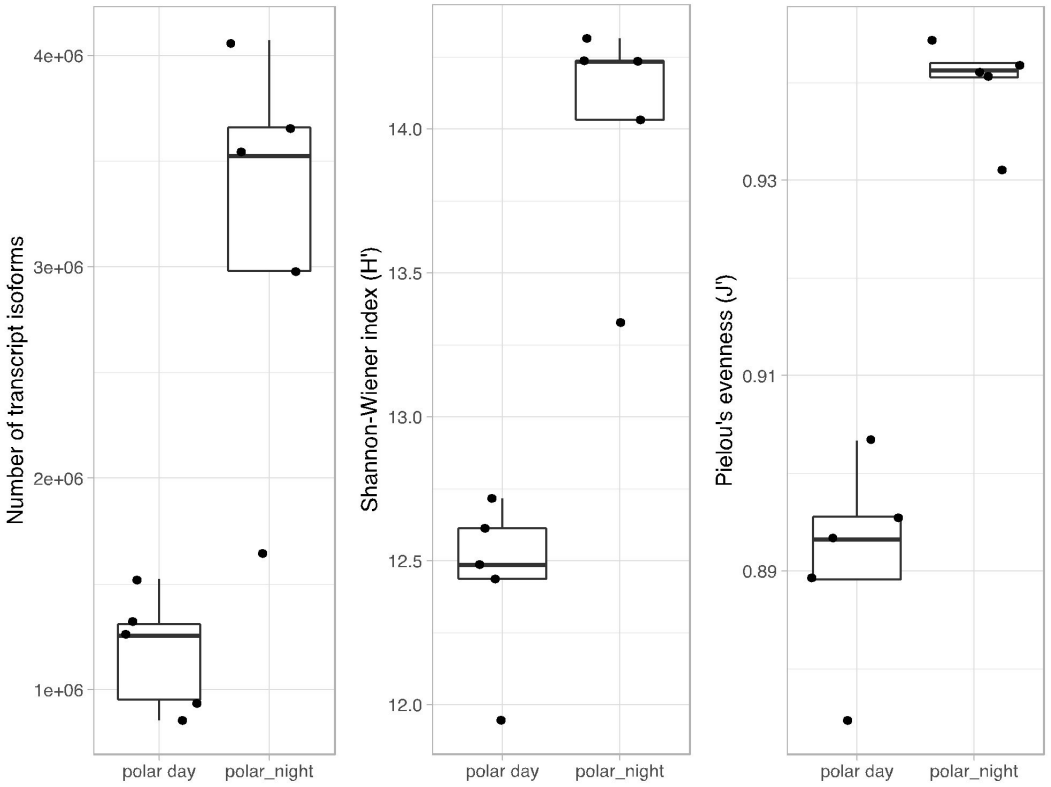


Figure 4
Grouping of the samples according to similarity in their transcript composition based on the core dataset. Approximately unbiased (au) and bootstrap probability (bp) values strongly support the clustering (au and bp > 80). Note that two main highly supported groups are delineated according to the light regime: polar day and times of the year with night present, i.e. polar day and September. The polar day cluster was divided into two groups with strong support.

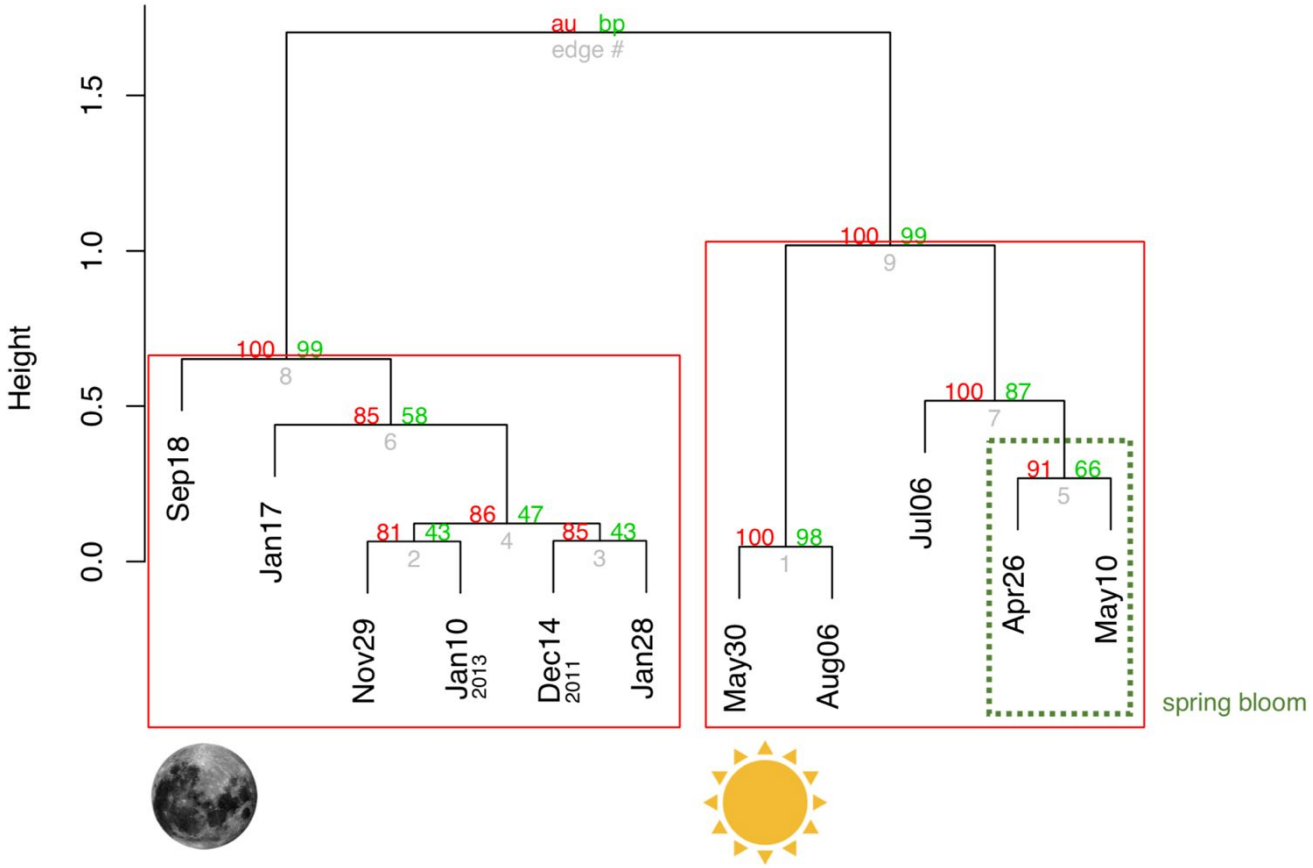
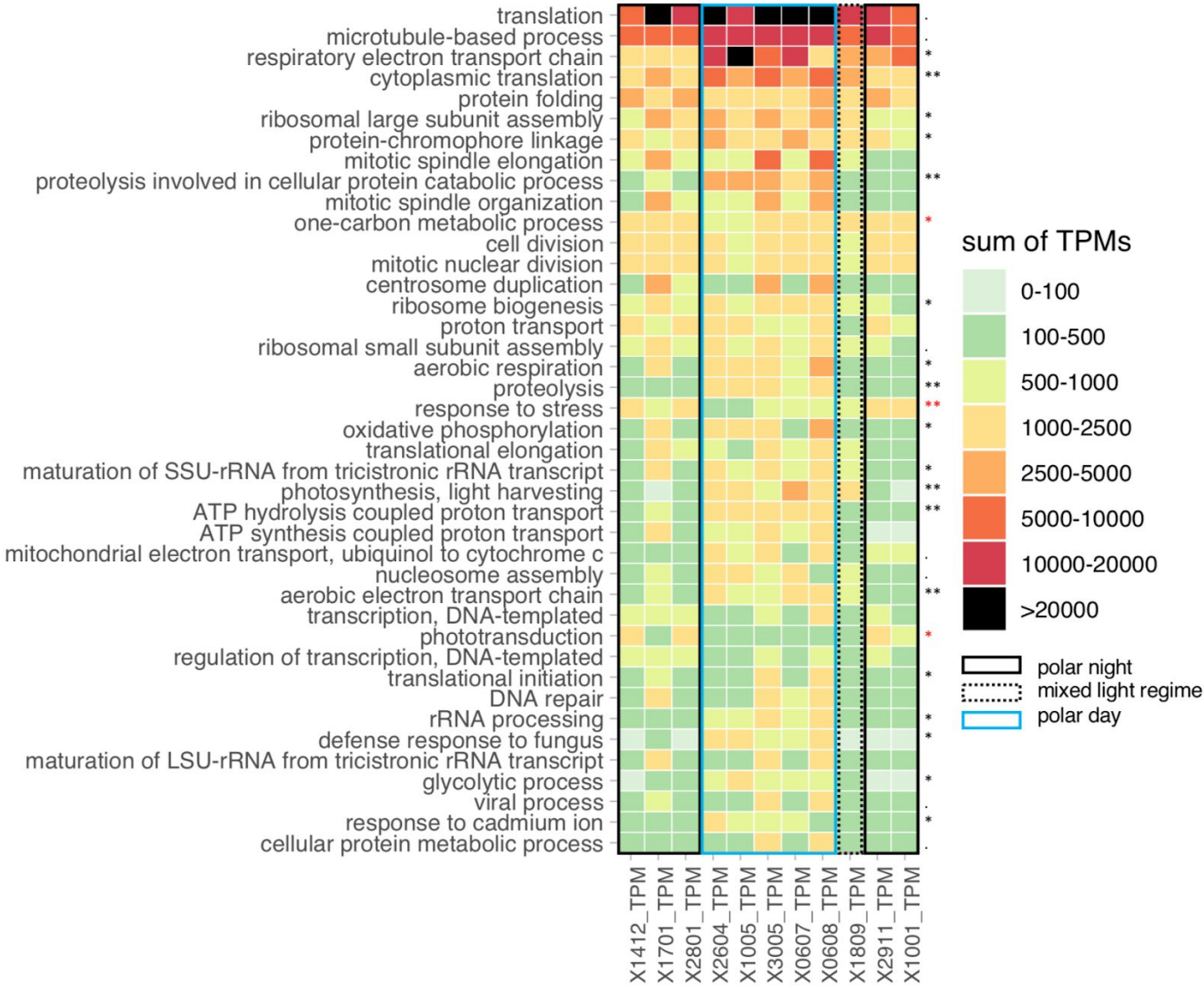
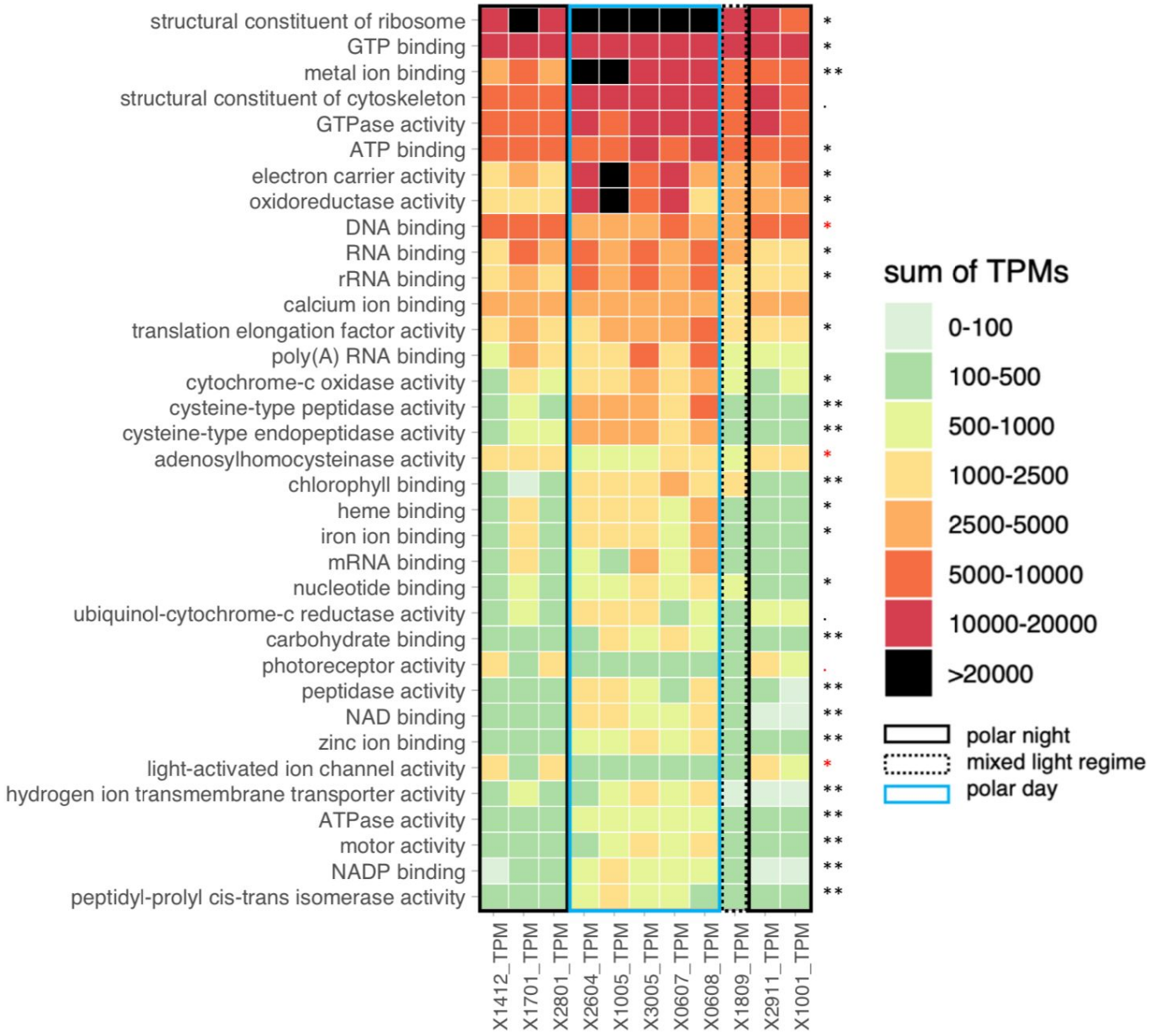


Figure 5
The most abundant GO terms within the core dataset corresponding to biological processes (A), molecular functions (B) and cellular compartments (C) featuring >5000 TPM for each GO term. Asterisks indicate functions that were discriminated between polar day (red asterisk) and polar night (black asterisk) using *simper* function (September was excluded from this analysis). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

(A) Biological processes



(B) Molecular functions



(C) Cellular compartments

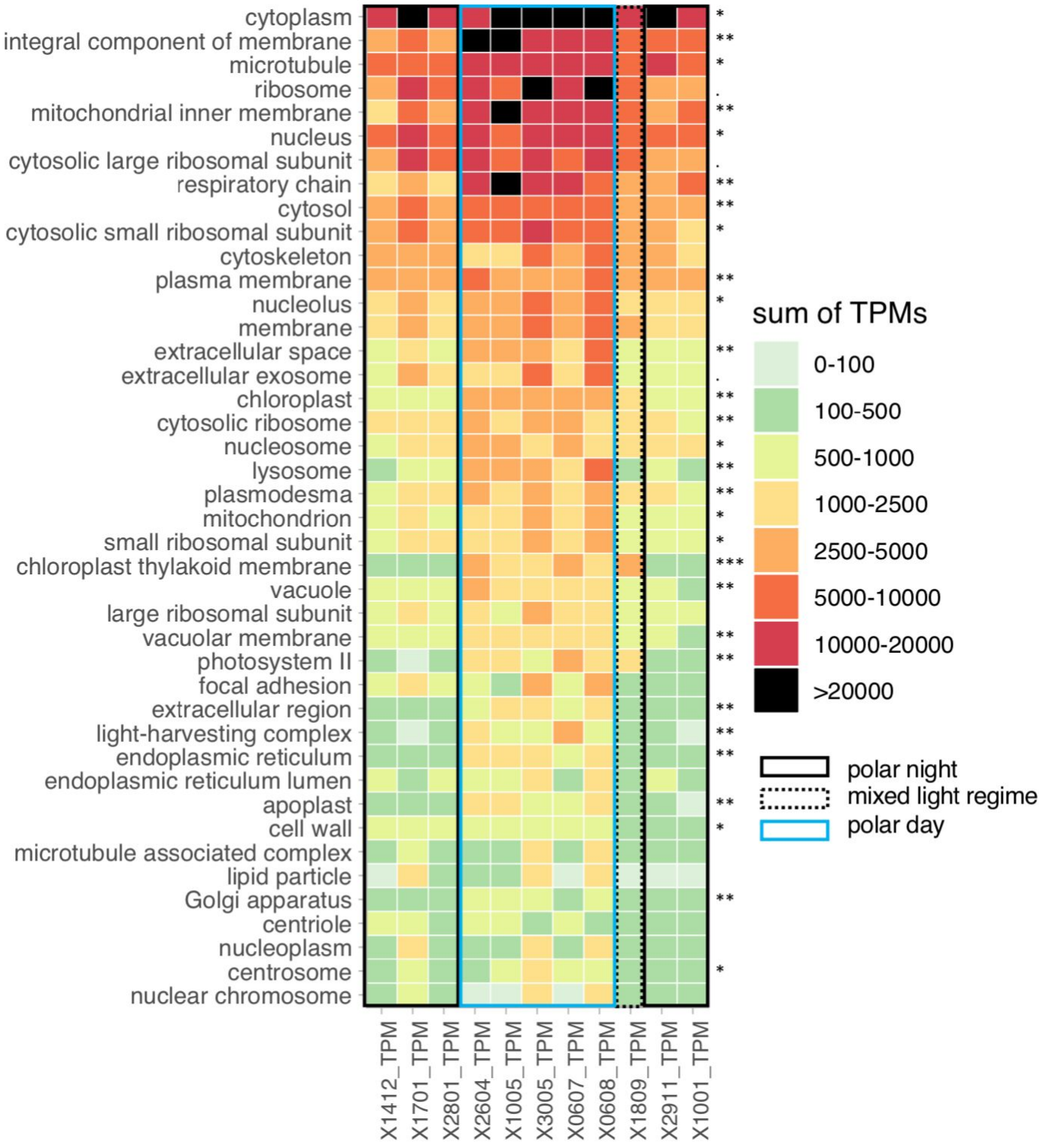


Figure 6
Top 10 less abundant GO terms (sum of TPM in all metatranscriptomes < 5000) with mean number of transcripts higher during polar night than during polar day (p > 0.05).

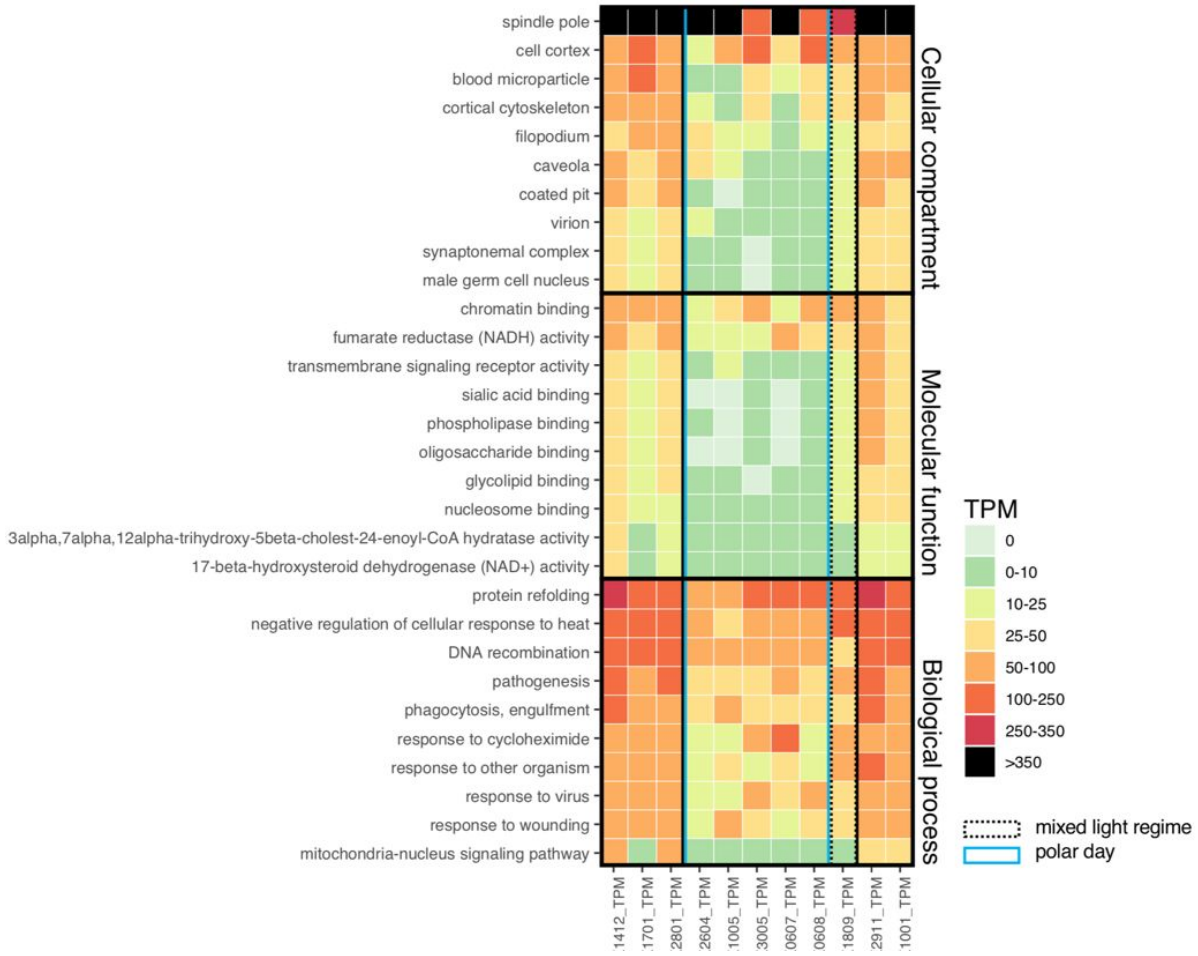


Figure 7
Seasonal abundances of transcripts associated with light-dependent biological processes (GO terms), shown as sum of transcripts per million (TPM) for samples from polar night (n=5) and polar day (n=5).

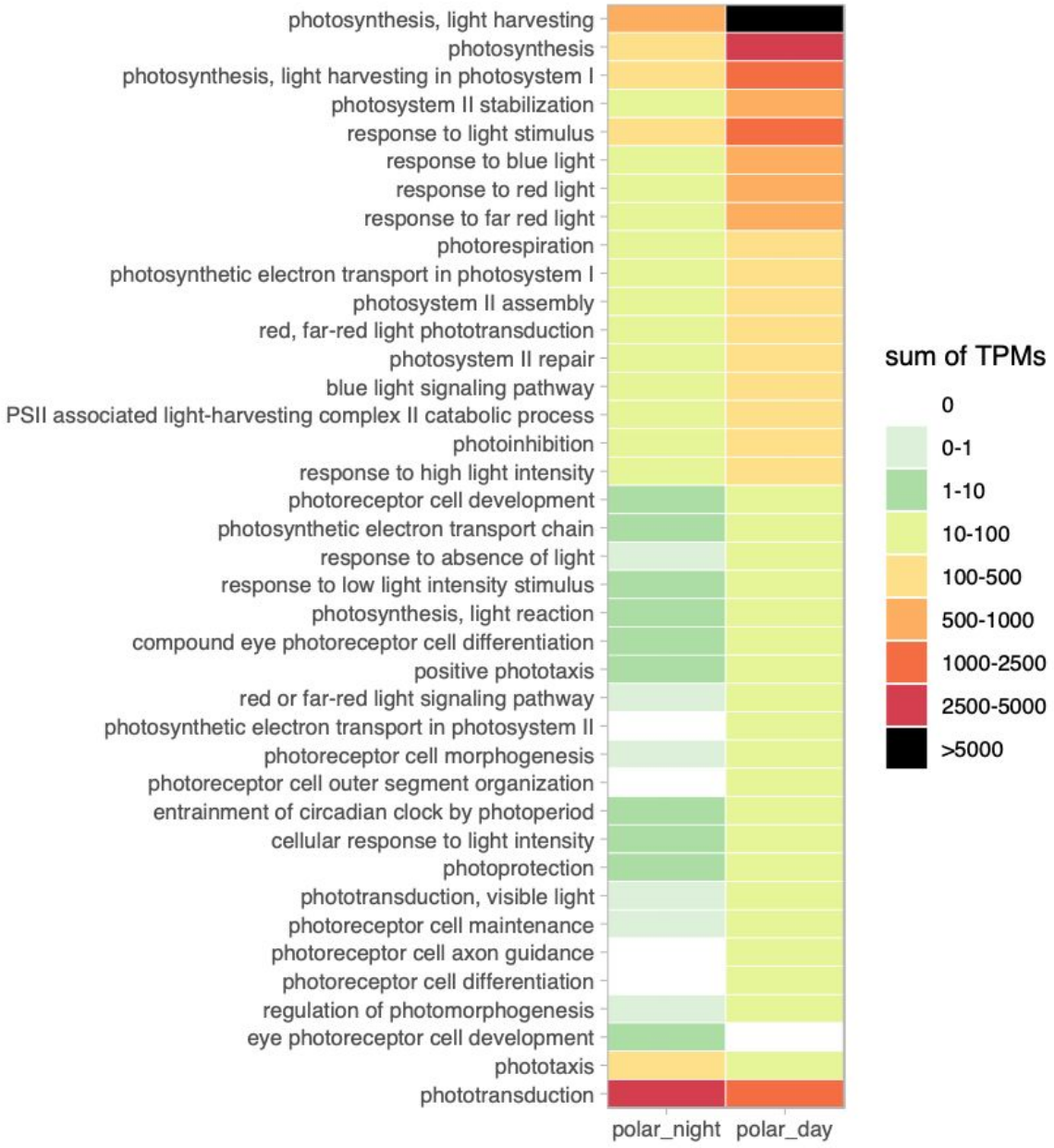


Figure 8
Taxonomic assignment shown as the proportion of clean reads assigned to a taxonomic group with Taxmapper. Each dot represents the proportion of reads in one sample.

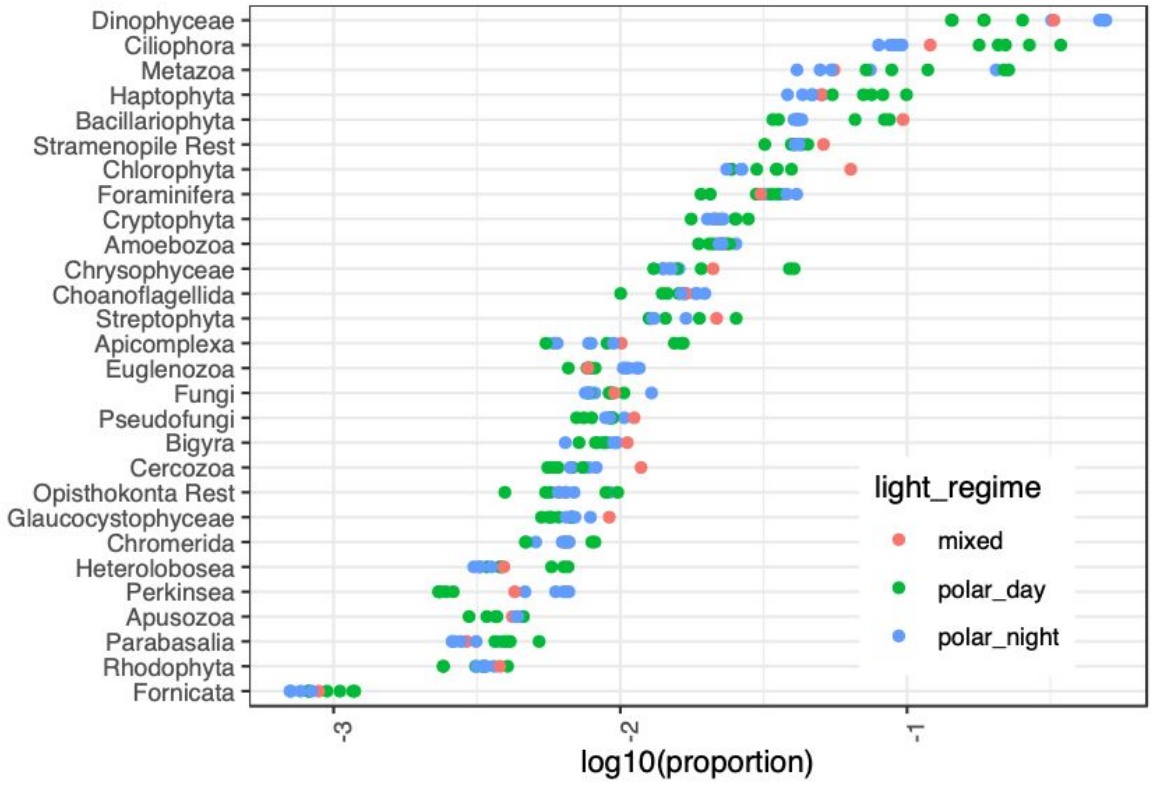


Figure 9
Proportion of core dataset transcript isoforms from IsA (red circle) detected by *Tara* Oceans data on 0.8-2000 µm plankton size fraction from different depths (SUR – surface waters, DCM – deep chlorophyll maximum and MES – mesopelagic waters).

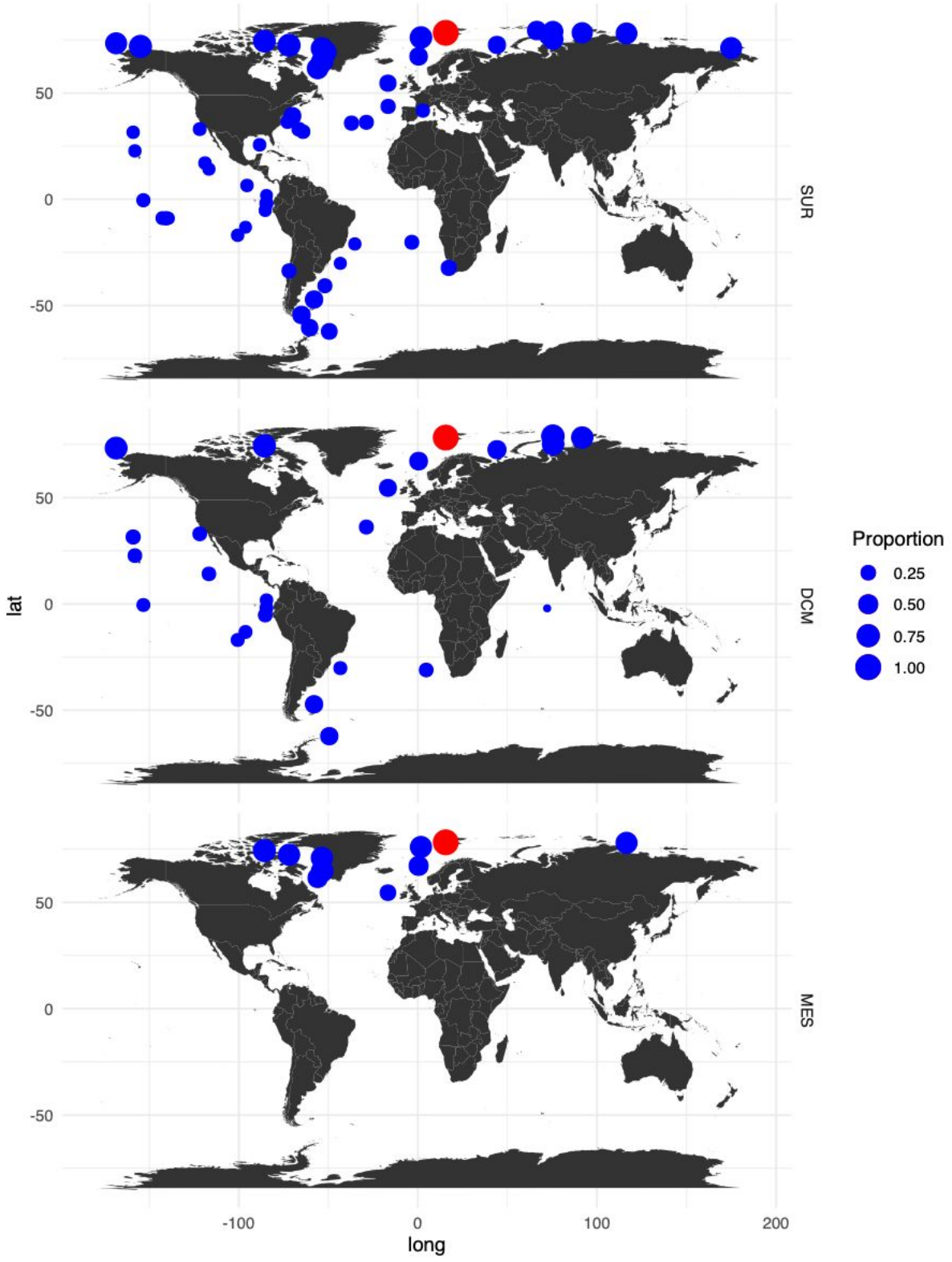


Table 1

Environmental parameters corresponding to each metatranscriptome were sampled at 25 m depth. PAR (photosynthetically active radiation) was measured as close to local noon as possible; declination was calculated for local noon and day length refers to the number of hours when the sun is above the horizon. Chlorophyll a biomass is reported for 2 size fractions: > 0.7 µm (filtered on GF/F glass microfiber filters (Whatman, England) or > 10 µm (filtered on Isopore membrane polycarbonate filters (Millipore, USA)). Water masses: LW – local water and IW – intermediate water. Other abbreviations: BD – below detection, NA – not available. The data were originally published in Kubiszyn et al., 2017; Marquardt et al., 2016.

	Light parameters			Chlorophyll *a* biomass		Nutrient concentrations			Physical parameters		
Sampling date	PAR [µmol m^{-2} s^{-1}]	declination [°]	day length [h]	chl *a* >0.7µm [µg/l]	chl *a* >10 µm [µg/l]	NO_3 and NO_2 [µM]	PO_4 [µM]	$Si(OH)_4$ [µM]	Water mass	Temp [°C]	Salinity
14-Dec-11	BD	-23.20	0	0.024	0.003	2.81	0.15	1.51	LW	0.9	34.32
17-Jan-12	BD	-20.81	0	0.04	0.01	NA	NA	NA	LW	-1.3	34.18
28-Jan-12	BD	-18.29	0	0.036	0.016	7.22	0.3	4.48	LW	-0.8	34.24
26-Apr-12	6.1	13.72	24	1.42	1.261	4.49	0.38	1.65	LW	-0.2	34.56
10-May-12	6.9	17.79	24	3.794	3.163	1.54	0.24	0.28	LW	0.3	34.57
30-May-12	0.5	21.86	24	NA	NA	BD	0.23	0.85	LW	0.3	34.43
06-Jul-12	0.3	22.62	24	0.449	0.039	0.21	0.08	0.43	IW	2.1	34.24
06-Aug-12	3.4	16.51	24	1.04	0.024	0.25	0.12	1.42	IW	2.9	34.12
18-Sep-12	1.7	1.62	13	0.206	0.042	2.6	0.3	2.45	IW	3.8	34.36
29-Nov-12	BD	-21.58	0	0.047	0.009	5.42	0.45	3.03	IW	1.8	34.39
10-Jan-13	BD	-21.89	0	0.02	0.007	5.41	0.57	2.3	IW	1.5	34.67

Supplement 1

Pre-processing statistics of the sequenced samples.

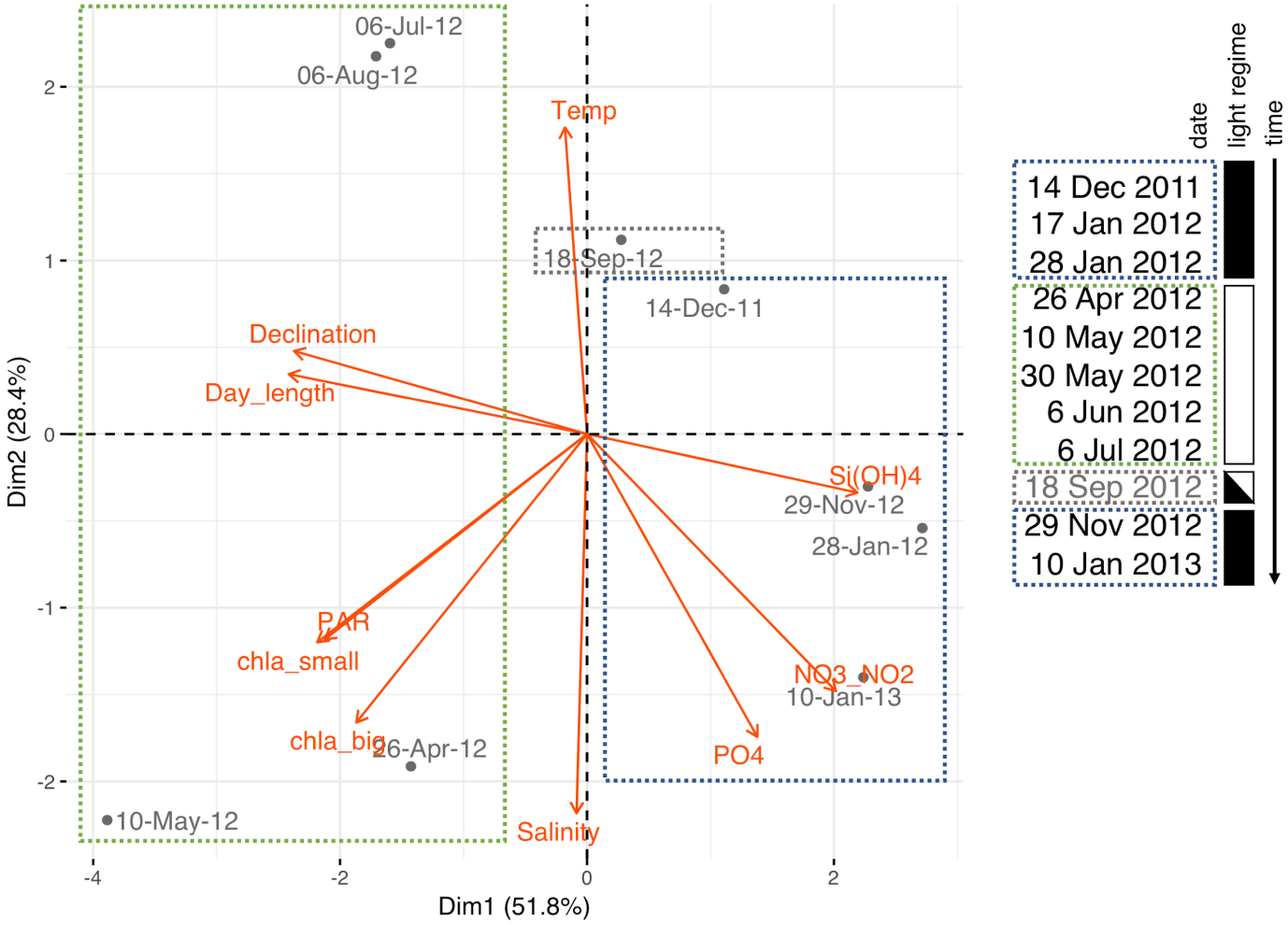
date	**06.08.2012F**	**06.08.2012R**	**10.01.2013F**	**10.01.2013R**	**14.12.2011F**	**14.12.2011R**	**26.04.2012F**	**26.04.2012R**
library ID	178928	178928	178931	178931	178921	178921	178924	178924
GC%	44	46	49	51	53	54	42	39
reads per file	53095430	53095430	29217339	29217339	53507834	53507834	42187756	42187756
input_reads		106190860		58434678		107015668		84375512
input_bp		16034819860		8823636378		16159365868		12740702312
total_removed_adapters (reads)	3594274 reads (3.38%)		1008282 reads (1.73%)		710068 reads (0.66%)		2964822 reads (3.51%)	
total_removed_adapters (bp)	767197458 bases (4.78%)		250592800 bases (2.84%)		252485450 bases (1.56%)		617320950 bases (4.85%)	
after removing _adapters (reads)		102596586		57426396		106305600		81410690
after removing _adapters (bp)		15267622402		8573043578		15906880418		12123381362
total_removed_overrepresented (reads)	12727556 reads (12.41%)		4339668 reads (7.56%)		2156654 reads (2.03%)		14902534 reads (18.31%)	
total_removed_overrepresented (bp)	1908980870 bases (12.50%		650943022 bases (7.59%)		323493264 bases (2.03%)		2235360402 bases (18.44%	
after removing overrepresented sequences (polyAT) (reads)		89869030		53086728		104148946		66508156
after removing overrepresented sequences (polyAT) (bp)		13358641532		7922100556		15583387154		9888020960
total_contaminants_removed (reads)	22 reads (0.00%)		10 reads (0.00%)		24 reads (0.00%)		32 reads (0.00%)	
total_contaminants_removed (bp)	3300 bases (0.00%)		1500 bases (0.00%)		3600 bases (0.00%)		4800 bases (0.00%)	
after removing contamination (phiX) (reads)		89869008		53086718		104148922		66508124
after removing contamination (phiX) (bp)		13358638232		7922099056		15583383554		9888016160
% of the sequences removed by SortMeRNA	12941569 (14.40%)		574047 (1.08%)		1259344 (1.21%)		1083885 (1.63%)	
Qtrimmed (reads)	38996457 reads (50.96%)		28274738 reads (54.02%)		65542976 reads (63.94%)		7800603 reads (11.96%)	
total_removed_in_trimming <q20 <50bp (reads)	4552988 reads (5.95%)		3340956 reads (6.38%)		7721716 reads (7.53%)		58760 reads (1.32%)	
total_removed_in_trimming <q20 <50bp (bp)	1823066072 bases (16.04%		1361719455 bases (17.43%		3191879880 bases (20.81%		231178829 bases (2.38%)	
after quality trimming <q20 (reads) and removing <25bp		71977304		48999030		94780860		64352188
after quality trimming <q20 (bp) and removing <25bp		9540035522		6448676667		12144896024		9463284051
q20								
preprocessing left XX% of the input reads		**67.78107268**		**83.85265681**		**88.56727409**		**76.26879704**
preprocessing left XX% of the input base pairs		**59.49574492**		**73.08411624**		**75.15700878**		**74.27600001**

date	**28.01.2012F**	**28.01.2012R**	**29.11.2012F**	**29.11.2012R**	**30.05.2012F**	**30.05.2012R**	**10.05.2012F**	**10.05.2012R**
library ID	178923	178923	178930	178930	178926	178926	197614	197614
GC%	52	54	51	52	40	43	43	35
reads per file	44223545	44223545	36511980	36511980	66139964	66139964	33405001	33405001
input_reads		88447090		73023960		132279928		66810002
input_bp		13355510590		11026617960		19974269128		10088310302
total_removed_adapters (reads)	853122 reads (0.96%)		2985896 reads (4.09%)		2882120 reads (2.18%)		4833014 reads (7.23%)	
total_removed_adapters (bp)	256674520 bases (1.92%)		609093454 bases (5.52%)		671812558 bases (3.36%)		855073868 bases (8.48%)	
after removing _adapters (reads)		87593968		70038064		129397808		61976988
after removing _adapters (bp)		13098836070		10417524506		19302456570		9233236434
total_removed_overrepresented (reads)	3437922 reads (3.92%)		3745474 reads (5.35%)		21528038 reads (16.64%)		17322844 reads (27.95%)	
total_removed_overrepresented (bp)	515681928 bases (3.94%)		561804262 bases (5.39%)		3229154020 bases (16.73%)		2598424624 bases (28.14%)	
after removing overrepresented sequences (polyAT) (reads)		84156046		66292590		107869770		44654144
after removing overrepresented sequences (polyAT) (bp)		12583154142		9855720244		16073302550		6634811810
total_contaminants_removed (reads)	14 reads (0.00%)		28 reads (0.00%)		28 reads (0.00%)		122 reads (0.00%)	
total_contaminants_removed (bp)	2058 bases (0.00%)		4200 bases (0.00%)		4140 bases (0.00%)		18300 bases (0.00%)	
after removing contamination (phiX) (reads)		84156032		66292562		107869742		44654022
after removing contamination (phiX) (bp)		12583152084		9855716044		16073298410		6634793510
% of the sequences removed by SortMeRNA	788344 (0.94%)		761703 (1.15%)		4361208 (4.04%)		920961 (2.06%)	
Qtrimmed (reads)	48298698 reads (58.12%)		35887947 reads (54.92%)		53213423 reads (51.71%)		16744756 reads (38.49%)	
total_removed_in_trimming <q20 <50bp (reads)	6267376 reads (7.54%)		4650486 reads (7.12%)		6634722 reads (6.45%)		1284764 reads (2.95%)	
total_removed_in_trimming <q20 <50bp (bp)	2438453708 bases (19.63%)		1774945335 bases (18.27%)		2549316689 bases (16.63%)		578366329 bases (8.95%)	
after quality trimming <q20 (reads) and removing <25bp		76829564		60690026		96267868		42218506
after quality trimming <q20 (bp) and removing <25bp		9986095094		7938701707		12782570045		5884564151
q20								
preprocessing left XX% of the input reads		**86.86499918**		**83.10974371**		**72.77586967**		**63.19189453**
preprocessing left XX% of the input base pairs		**74.77134645**		**71.99579904**		**63.99518282**		**58.3305229**

date	**06.07.2012F**	**06.07.2012R**	**17.01.2012F**	**17.01.2012R**	**18.09.2012F**	**18.09.2012R**
library ID	197615	197615	178922	178922	178929	178929
GC%	44	36	43	44	39	41
reads per file	45126210	45126210	16733154	16733154	14583344	14583344
input_reads		90252420		33466308		29166688
input_bp		13628115420		5053412508		4404169888
total_removed_adapters (reads)	8082018 reads (8.95%)		780138 reads (2.33%)		305550 reads (1.05%)	
total_removed_adapters (bp)	1402519158 bases (10.29%		177937372 bases (3.52%)		87869456 bases (2.00%)	
after removing _adapters (reads)		82170402		32686170		28861138
after removing _adapters (bp)		12225596262		4875475136		4316300432
total_removed_overrepresented (reads)	24853874 reads (30.25%)		5134692 reads (15.71%)		6540160 reads (22.66%)	
total_removed_overrepresented (bp)	3728078590 bases (30.49%		770199344 bases (15.80%)		981023728 bases (22.73%)	
after removing overrepresented sequences (polyAT) (reads)		57316528		27551478		22320978
after removing overrepresented sequences (polyAT) (bp)		8497517672		4105275792		3335276704
total_contaminants_removed (reads)	90 reads (0.00%)		234 reads (0.00%)		296 reads (0.00%)	
total_contaminants_removed (bp)	13500 bases (0.00%)		35100 bases (0.00%)		44400 bases (0.00%)	
after removing contamination (phiX) (reads)		57316438		27551244		22320682
after removing contamination (phiX) (bp)		8497504172		4105240692		3335232304
% of the sequences removed by SortMeRNA	892946 (1.56%)		481828 (1.75%)		502888 (2.25%)	
Qtrimmed (reads)	20566426 reads (36.59%)		9731912 reads (36.14%)		9171496 reads (42.28%)	
total_removed_in_trimming <q20 <50bp (reads)	1539676 reads (2.74%)		1524976 reads (5.66%)		1386468 reads (6.39%)	
total_removed_in_trimming <q20 <50bp (bp)	702210113 bases (8.43%)		493699194 bases (12.31%)		465521314 bases (14.36%)	
after quality trimming <q20 (reads) and removing <25bp		54665634		25403058		20307194
after quality trimming <q20 (bp) and removing <25bp		7629747557		3518411592		2775863824
q20						
preprocessing left XX% of the input reads		**60.56971547**		**75.90636529**		**69.62461422**
preprocessing left XX% of the input base pairs		**55.9853459**		**69.62446835**		**63.02808235**

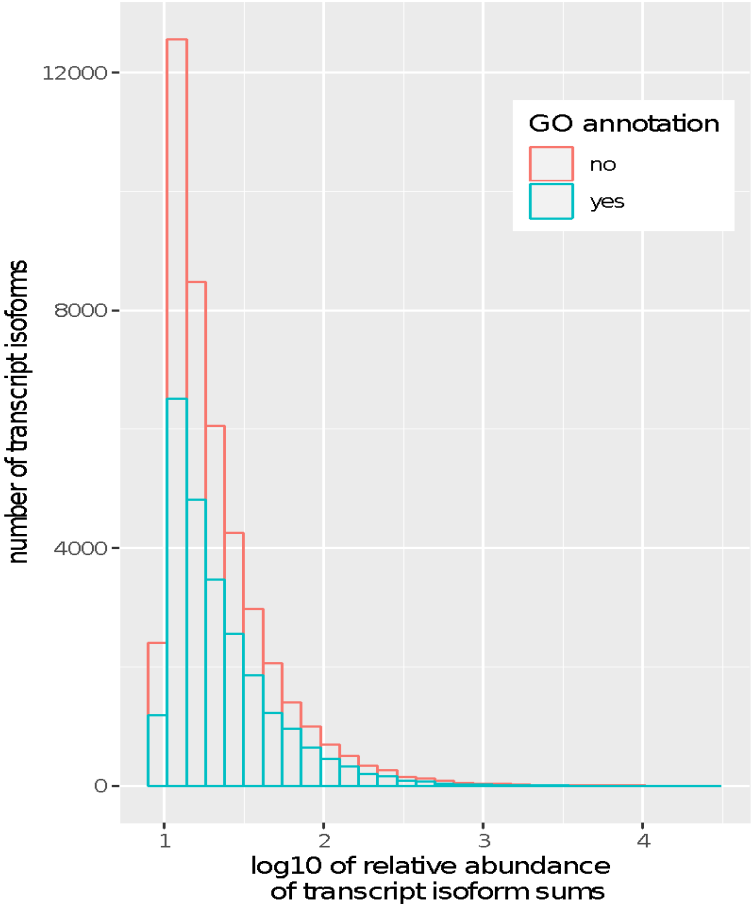
Supplement 2

Principal component analysis of environmental data. 17 January 2012 and 30 May 2012 are not included in the figure, due to some missing values.



Supplement 3

Most transcript isoforms were characterized by low sum of relative abundance across samples: 8 transcripts with >10 000 TPM, 154 with >1000 TPM, 3483 with >100 TPM, 68 166 with > 10 TPM and 2 390 862 with >1TPM. Most of the least abundant transcripts were unannotated with the Gene Ontology (based on blast results from mapping transcript isoforms to UniProt).



Supplement 3

Summary of results from SIMPER analysis of Bray-Curtis dissimilarity matrix of reduced dataset. These are the 10 top individual transcript isoforms contributing to differences between grouping factors (transcript in bold was the only one among the top 10 in section a, b, c that was functionally annotated by at least one database; yellow highlight – transcript isoforms present in a, b and c).

a. grouping factor: polar night vs. polar day (without September)

	average	sd	ratio	ava	avb	cumulative sum
TRINITY_DN3567435_c13_g1_i1	**0.0091068**	**7.643e-03**	**1.1916**	**519.77**	**5236.75**	**0.01020**
TRINITY_DN3531545_c4_g1_i13	0.0057171	4.039e-03	1.4154	645.86	3522.82	0.01661
TRINITY_DN3383058_c4_g1_i10	0.0049891	4.020e-03	1.2409	210.17	2793.45	0.02219
TRINITY_DN3531545_c4_g1_i11	0.0044515	2.692e-03	1.6536	572.16	2830.47	0.02718
TRINITY_DN3528005_c14_g2_i1	0.0033457	1.859e-03	1.7995	51.95	1762.08	0.03093
TRINITY_DN3567090_c10_g1_i2	0.0032716	2.241e-03	1.4599	411.10	2057.50	0.03459
TRINITY_DN3565854_c4_g3_i5	0.0032060	4.265e-03	0.7518	2.79	1660.42	0.03818
TRINITY_DN3417822_c3_g2_i11	0.0031592	2.095e-03	1.5081	271.12	1859.83	0.04172
TRINITY_DN3313835_c6_g2_i3	0.0027240	2.927e-03	0.9306	45.65	1396.98	0.04477
TRINITY_DN3417822_c3_g2_i2	0.0024838	2.791e-03	0.8898	397.70	1458.13	0.04756

b. absence vs. presence of night

	average	sd	ratio	ava	avb	cumulative sum
TRINITY_DN3567435_c13_g1_i1	**0.0089522**	**0.0075762**	**1.1816**	**635.64**	**5236.74**	**0.01014**
TRINITY_DN3531545_c4_g1_i13	0.0055438	0.0039526	1.4026	780.39	3522.82	0.01642
TRINITY_DN3383058_c4_g1_i10	0.0048617	0.0040257	1.2077	279.26	2793.45	0.02193
TRINITY_DN3531545_c4_g1_i11	0.0042184	0.0026293	1.6044	739.23	2830.47	0.02670
TRINITY_DN3528005_c14_g2_i1	0.0033084	0.0018580	1.7807	74.10	1762.08	0.03045

TRINITY_DN3565854_c4_g3_i5	0.0032128	0.0042582	0.7545	2.32	166.42	0.03409
TRINITY_DN3567090_c10_g1_i2	0.0031939	0.0022026	1.4501	469.78	2057.50	0.03771
TRINITY_DN3417822_c3_g2_i11	0.0030590	0.0020509	1.4915	343.66	1859.83	0.04117
TRINITY_DN3313835_c6_g2_i3	0.0027182	0.0029221	0.9302	51.29	1396.98	0.04425
TRINITY_DN3417822_c3_g2_i2	0.0024484	0.0027053	0.9050	478.65	1458.13	0.04702

c. bloom vs. post-bloom

	average	sd	ratio	ava	avb	cumulative sum
TRINITY_DN3567435_c13_g1_i1	**0.0085442**	**4.672e-03**	**1.829**	**7807.2**	**1381.03**	**0.01159**
TRINITY_DN3565854_c4_g3_i5	0.0056244	1.861e-03	3.022	11.80	4133.35	0.01922
TRINITY_DN3531545_c4_g1_i13	0.0048821	1.976e-03	2.471	4963.8	1361.30	0.02585
TRINITY_DN3383058_c4_g1_i10	0.0039800	2.688e-03	1.481	3964.9	1036.28	0.03125
TRINITY_DN3561515_c5_g1_i4	0.0035883	2.845e-04	12.613	64.05	2689.11	0.03612
TRINITY_DN3455354_c3_g3_i2	0.0031505	3.037e-04	10.376	24.96	2331.49	0.04039
TRINITY_DN3531545_c4_g1_i11	0.0031379	1.345e-03	2.334	3757.6	1439.78	0.04465
TRINITY_DN3565854_c4_g3_i2	0.0028920	7.801e-04	3.707	7.66	2126.66	0.04857
TRINITY_DN3567090_c10_g1_i2	0.0028720	1.036e-03	2.772	2911.7	776.19	0.05247
TRINITY_DN3417822_c3_g2_i2	0.0028412	1.986e-03	1.4303	2269.4	241.16	0.05633