Biogeography of Arctic Eukaryotic Microbiome: A comparative approach between 18S rRNA gene metabarcoding and microscopic analysis

Marta Isabel Lopes de Sousa

Mestrado em Ciências e Tecnologia do Ambiente, Área de Especialização em Riscos: Avaliação e Gestão Ambiental

Departamento de Geociências, Ambiente e Ordenamento do Território

2019/2020

Orientadora
Catarina Maria Pinto Mora Pinto de Magalhães
Professora Auxiliar Convidada na Faculdade de Ciências da Universidade do Porto e Investigadora no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

Co-orientadora
Maria Paola Tomasino
Investigadora no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Dissertação de candidatura ao grau de Mestre em Ciências e Tecnologia do Ambiente submetida à Faculdade de Ciências da Universidade do Porto.

O presente trabalho foi desenvolvido no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), sob a orientação científica da Doutora Catarina Maria Pinto Mora Pinto de Magalhães e da Doutora Maria Paola Tomasino.

Dissertation for applying to a Master's degree in Environmental Sciences and Technology, submitted to the Faculty of Sciences of the University of Porto.

The present thesis was developed at the Interdisciplinary Center for Marine and Environmental Research (CIIMAR) under the scientific supervision of PhD Catarina Maria Pinto Mora Pinto de Magalhães and PhD Maria Paola Tomasino.

Acknowledgements

I would like to take advantage of this space to thank all of those who, in various ways, have contributed to the accomplishment of the work here presented.

First of all, I would like to express my sincere gratitude to my supervisor Catarina Magalhães, for all the opportunities and responsibilities entrusted to her, and for the continuous attention, guidance and clarification of doubts. I also express my sincere gratitude to my co-supervisor Maria Paola for the immense support, advice, help, knowledge, motivation and patience. I also thank the BYT+ team (Blue Young Talent PLUS) for developing this program and thus allowing me the opportunity to contact to the world of scientific research for the first time.

I must also take the opportunity to thank Agnieskzka Tatarek and Józef Wiktor, collaborators of the Norwegian Polar Institute (NPI) for all the help and clarification of doubts. My thanks to the dedicated multidisciplinary team, which integrates or integrated EcoBioTec, during my dissertation period, especially to Eva Lopes for all the friendship, help and dedication and to Francisco Pascoal for having spent his time helping me with the results and interpretation of the data, I express my enormous gratitude to you both.

I also wish to thank everyone from Room 22, for being there for me, for listening to me, for the immense hours of conversation and for those reasons and more, I am grateful for our friendship. Most importantly, I thank my parents for always supporting me, fighting for my dreams, never letting me down and being my source of inspiration and strength.

Finally, I would like to thank the institutions and projects that financed this work, namely the Portuguese Polar Program, the Norwegian Polar Institute (NPI), the MOSJ (Environmental Monitoring of Svalbard and Jan Mayen) and the Portuguese Foundation for Science and Technology that financed this study through the NITROLIMIT project (PTDC/CTA-AMB/30997/2017).

Resumo

Ambientes extremos como o Oceano Ártico têm enfrentado mudanças severas e drásticas para o nosso planeta. As manifestações mais notáveis das mudanças climáticas são o derretimento do gelo, juntamente com a diminuição da extensão e espessura do gelo marinho do Oceano Ártico. Estas implicações estão a comprometer todo o ecossistema Ártico e mudando a produtividade primária e a biogeoquímica do Ártico.

Este estudo teve como objetivo investigar os padrões biogeográficos da diversidade e distribuição do microplâncton do Ártico ao longo de dois transectos oceanográficos na Zona Marginal do Gelo (MIZ). Ao combinar a abordagem 18S metabarcoding, usando a sequenciação de nova geração (NGS), e a clássica identificação e contagem de células, este estudo também produziu algumas diretrizes para programas de monitoramento de fitoplâncton do Ártico, a fim de preencher lacunas críticas sobre as questões emergentes que têm surgido devido a um Ártico em mudança.

Durante a campanha de monitoramento do Ártico, MOSJ-ICE2016 (*Environmental Monitoring of Svalbard* and *Jan Mayen*), liderado pelo Instituto Polar Norueguês (NPI), a água do mar foi amostrada, à superfície (2m-5m), profundidade máxima da clorofila (17m-50m) e profundidade (50m-1035m), juntamente com dados físicos e biogeoquímicos, durante 25 de Julho a 4 de Agosto de 2016, ao longo de dois transectos oceanográficos na Zona Marginal do Gelo (MIZ): Kongsfjorden e Rijpfjorden, no arquipélago de Svalbard. A contagem e a identificação do fitoplâncton pelo microscópio foram realizadas pelo NPI, enquanto as análises de metabarcoding foram realizadas no CIIMAR (Porto, Portugal), através da análise da sequenciação massiva paralela da região hipervariável V4 do gene da subunidade pequena do RNA ribossomal (SSU rRNA, sigla anglo-saxónica).

Os resultados obtidos por 18S-V4 metabarcoding revelaram que o conjunto de dados eucarióticos compreende uma estrutura de comunidade protista diversificada e altamente complexa, com um padrão biogeográfico acentuado das comunidades protistas ao longo da Zona Marginal do Gelo, Svalbard com uma dependência de profundidade ao longo da coluna de água e variações das massas de água. Fortes ligações entre a distribuição das comunidades protistas e gradientes ambientais também foram identificados. Quanto à comparação metodológica, os resultados mostraram que a identificação e caracterização dos grupos de fitoplâncton foram bastante diferentes quando a análise microscópica e metabarcoding foram utilizadas. Vários grupos taxonómicos foram identificados sob metabarcoding que não foram encontrados em microscopia, e o mesmo se verificou ao contrário. Além disso, 18S-V4

metabarcoding foi capaz de detetar grupos fitoplâncton de maior diversidade, mas com menor recuperação de abundância, já na microscopia, este método detetou menos grupos taxonómicos, mas com maior recuperação de abundância.

No entanto, apesar das discrepâncias de ambos os métodos, este é o primeiro passo para melhorar a comparação para a comunidade protista na região do Ártico e pode servir como base para estudos futuros. Assim, a coleção microbiana do MOSJ-ICE2016 analisada no presente estudo fornece uma comparação pormenorizada entre métodos, oferecendo diferentes percepções ecológicas, ajudando a estar um passo à frente no preenchimento de lacunas críticas da resposta dos produtores primários, de acordo com o drástico declínio do gelo Ártico. Isto é relevante para a aplicação de metodologias eficazes em programas de monitoramento de fitoplâncton de longo prazo, imprescindíveis para antecipar tendências futuras do aquecimento do Ártico.

Palavras-Chave

Oceano Ártico, Fitoplâncton, Protistas, Alterações Climáticas, Morfologia, 18S Metabarcoding, Diversidade, Padrões Biogeográficos

Abstract

Extreme environments like the Arctic Ocean have been facing severe and drastic changes in our planet. The most notable manifestations of climate change are the melting ice, along with the decrease of the Arctic Ocean sea ice extent and thickness. These pronounced implications are compromising the entire Arctic ecosystem and changing Arctic's primary productivity and biogeochemistry.

This study aimed to investigate the biogeographic patterns of Arctic microplankton diversity and distribution along two oceanographic transects in the Marginal Ice Zone (MIZ) around Svalbard. By combining the 18S metabarcoding approach, using next generation sequencing techniques (NGS), and the classic microscope cell identification, this study also produced some guidelines for Arctic phytoplankton monitoring programs in order to fill critical gaps concerning the emerging questions that have been raising due to a changing Arctic.

During the Arctic monitoring campaign, MOSJ-ICE2016 (Environmental Monitoring of Svalbard and Jan Mayen), led by Norwegian Polar Institute (NPI), seawater was collected, at surface (2m-5m), deep chlorophyll maximum (17m-50m) and bottom depth (50m-1035m), together with physical and biogeochemical data, during 25th July and 4th August of 2016, along two oceanographic transects at the Marginal Ice Zone (MIZ): Kongsfjorden and Rijpfjorden transect, in Svalbard archipelago. Microscope phytoplankton counts and identification were performed by NPI while metabarcoding analysis were performed at CIIMAR (Porto, Portugal), through the analysis of the massive sequencing of the hypervariable V4 region of the small subunit ribosomal RNA (SSU rRNA) gene.

The results from 18S-V4 metabarcoding revealed that the eukaryotic dataset comprises highly complex and diverse protists community structure, with a marked biogeographic pattern of the protists communities along the Svalbard Marginal Ice Zone with clear trends of depth-dependency along the water column and water masses variations. Strong links between protists communities distribution and environmental gradients were also identified. As for the methodological comparison, results showed that the identification and characterization of the phytoplankton groups were quite different when using the microscopy analysis and metabarcoding. Several taxonomic groups were identified under metabarcoding that were not found in microscopy, and the same was verified in the other way around. In addition, 18S-V4 metabarcoding was able to detect higher phytoplankton diversity but with lower abundance recovery, as for microscopy, this method detected fewer taxonomic groups but with higher abundance recovery.

Nevertheless, despite the discrepancies of both methods, this is the first step to improve a methodological comparison for the protists community in the Arctic region and could serve as a baseline for further studies. Thus, the microbial collection of MOSJ-ICE2016 analyzed in the present study provides a comprehensive comparison between methods, offering different ecological perceptions, helping to be one step ahead in filling critical gaps of the response of primary producers, according to the drastic decline of the Arctic ice. This is relevant to apply effective methodologies in long-term phytoplankton monitoring programs, essential to anticipate future Arctic warming trends.

Key words

Arctic Ocean, Phytoplankton, Protists, Climate Change, Morphology, 18S Metabarcoding, Diversity, Biogeographic patterns

Table of Contents

List of Communications

The elaboration of this dissertation benefited from the following conference presentations.

- Sousa, M., Tomasino, M., Magalhães, C. – *Metabarcoding vs Microscope Approaches to Study Arctic Eukaryotic Microbiome* (ref 16784) in: Book of Abstracts of the 13th Meeting of Young Researchers of University of Porto, Portugal (**IJUP 2020**). ISBN: 978-989-746-253-5 (*Distinction Best Poster Communication in the panel Biological Sciences*). Available from: https://ijup.up.pt/2020/wp-content/uploads/sites/464/2020/02/IJUP2020_LivroResumos.pdf

- Sousa, M., Tomasino, M., Magalhães, C. - *Biogeographic of Arctic Eukaryotic Microbiome: A combining approach of metabarcoding and microscopic cell counts.* Accepted for submission of the BYT (Blue Young Talent) Journal, Issue 1. (**BYT 2020**). (*Article*)

- Sousa, M., Tomasino, M., Magalhães, C. - *Biogeographic patterns, diversity and distribution of Arctic's protists community.* Accepted for Book of Abstracts of the first meeting, Blue Think Conference, CIIMAR, Matosinhos Porto, Portugal (**BTC 2020**). (*Oral Communication in the Marine* Biotechnology).

- Sousa, M., Tomasino, M., Magalhães, C. - *Metabarcoding vs Microscope Approaches to Study Arctic Eukaryotic Microbiome.* In: Book of Abstracts of the first Annual Student Research Conference (European University Alliance for Global Health, **EUGLOH 2020**). (*Oral presentation in the Environmental challenges: water and food contamination, climate change session*).

List of Figures

List of Tables

List of Abbreviations

°C - Degrees Celsius

µm – micrometre

16S rRNA - 16S ribosomal ribonucleic acid

18S rRNA - 18S ribosomal ribonucleic acid

aka - also known as

AW – Atlantic water

BLAST - Basic Local Alignment Search Tool

bp - base pair

Chl *a* - Chlorophyll *a*

CO_2 - Carbon dioxide

CTD- Conductivity, Temperature and Depth

DCM - Deep chlorophyll maximum

DIC – Differential interference contrast

DNA - Deoxyribonucleic acid

e.g. - exempli gratia

FYI - First-year ice

HTS- High-throughput sequencing

i.e. - id est

L - litre

m – metre

MAW - Modified Atlantic Water

MIT – Microplankton

MIZ- Marginal Ice Zone

MOSJ - Environmental Monitoring of Svalbard and Jan Mayen

MYI - Multi-year ice

N - any base

ND – Not Determined

NGS - Next-generation sequencing

$NH4^+$ - Ammonium

NMDS – Non-metric multidimensional scaling

NO_2^- - Nitrite

NO_3^- - Nitrate

NPI – Norwegian Polar Institute

nt - nucleotide

O_2 - Oxygen

OTUs - Operational taxonomic units

PAR - Photosynthetically active radiation

PCA - Principal component analysis

PCR - Polymerase chain reaction

pH - Potential of hydrogen

PHT - Phytoplankton

PO_4^{3-} - Phosphate

PR_2 - Protist Ribosomal reference database

PSW - Polar surface Water

RV Lance - Research Vessel Lance

SiO_4^- - Silicate

SMHI - Swedish Meteorological and Hydrological Institute

SPAR- Superficial photosynthetically active radiation

SOP - Standard operating procedure

sp. - Species (singular)

SSU - Small SubUnit

TAW – Transformed Atlantic water

THC - *Thermohaline circulation*

v. – version

vs – *versus*

WoRMS - World Register of Marine Species

Introduction

Microbial Biogeography

Understanding what could influence the distribution of species, how organisms are dispersed, what patterns of distribution can be expected, as well comprehend why a specific taxon became restricted to a current area are pertinent questions of microbial biogeography [1].

Over the years, many studies have provided fundamental ecological insights to answer these questions but they remain still unsolved. Indeed, there are many factors that alter at different spatial and temporal scales the distribution of microbial communities, such as: climate change, seasonal variations, changes in temperature, light, salinity, depth, biotic interactions and/or low-abundance taxa and anthropogenic impact [2,3]. Therefore, there are many drivers that can affect the study of microbial biogeography, especially in communities that are inserted in extreme environments, such as the Arctic, where any change in the ecosystem could be reflected in the diversity and dynamics of microbial communities [4].

With the progressive molecular methods and the knowledge along the years through the classic taxonomic classification, it is possible to obtain more precise trends and patterns to get insight into microbial biogeography [5].

Arctic Ecosystem

At the most northernmost region is located the Arctic. The Arctic ecosystem is complex and unique and is mostly covered by water, the Arctic Ocean [6]. Both terrestrial and marine ecosystems are quite complex and climate stability is an important indicator of this balanced system [7]. Seasons in the Arctic are very intense with extreme changes in light and temperature, verified by the long cold and dark winter months and short summer periods [8,9].

The continuous flow of nutrients between Arctic terrestrial and marine ecosystem is crucial to maintain balance in all interconnections of the food chain, the amount of light is essential to maintain the observed patterns in sea ice floating on Arctic Ocean, when the massive spring blooms of algae feed the ecosystem, that grow in the Fall and Winter and melt in the Spring [10]. Any perturbation and stress induced in the ecosystem or in the food web could influence the balance established. Thus, it is crucial to consider how the Arctic ecosystem is responding to climate change, especially the response of microbial communities that might affect both lower and higher trophic levels [11].

Climate Change Impact in Arctic Ocean

Over the years, climate change has been causing severe consequences and implications in the Arctic Ocean [12]. The phrase "*Arctic is experiencing a warming at nearly twice the global rate",* is well documented in the literature [11,13,14]. As a matter of fact, it is in the Arctic region that suffers the most consequences with severe and drastic changes, especially in the Arctic Ocean [13,14].

The impacts of climate change in the Arctic Ocean, is quite well known by now and can be observed in different levels: increasing temperatures in air and waters, sea-level rise, decrease in snow and in ice cover, glacial runoff, shifts from thicker perennial multiyear ice (MYI) for thinner first year ice (FYI), ocean circulation changes, ocean acidification, changes in precipitations patterns, permafrost thawing, increased input of freshwater (from melting ice) and increased input of warmer water (from the North Atlantic Ocean) [15,16]. However, these are the direct implications that we have noticed over the years and we have a better perception of, but what about the indirect implications? What are the feedbacks of these implications for the marine ecosystem and the species that inhabit the Arctic Ocean?

There are many concerns about these consequences and because everything is interconnected between the ecosystem Atmosphere-Ice-Ocean, it is plausible to say that the implications of climate change in the Arctic Ocean has a vicious cycle [17]. By increasing the temperature progressively, due to global warming, there is a drastic reduction of the ice, which later will cause the substitution of MYI for FYI. Consequently, the glacier runoff will cause the sea level rise (along with the melting of the Greenland ice sheet), with a very sharp inflow of fresh water [18,19,20].

In turn, due to the increase of temperature, the oceans become warmer causing the reduction of sea ice cover and the surface of the ocean becomes more and more visible absorbing the sunlight, boosting once again the process described above [21]. Furthermore, the increased ocean load and the entrance of warmer and salty water from the North Atlantic Ocean to the Arctic Ocean, which is less dense and less salty, it's causing interferences in the Thermohaline circulation (THC) between the Atlantic Ocean, Arctic and Pacific Ocean, most likely to influence the ocean circulation and patterns [22].

These changes will also affect the organic matter, nutrients and minerals available in the Arctic Ocean altering the environmental gradients and subsequently the Arctic's primary productivity and biogeochemistry [16,22].

So, on the basis of a changing Arctic sea-ice, these changes and pressures are likely to lead to an Arctic ice-free during Summer by the end of the 21st century and to

an entire biological reorganization of this environment, by forcing Arctic Ocean species to adapt to these new conditions [23].

Implications on Arctic Ocean Phytoplankton Communities and Primary Productivity

Phytoplankton needs to have the right conditions for its growth, which include availability of sunlight, nutrients and minerals [24]. When these microscopic organisms have sufficient amounts of nutrients concentrations and the right amount of sunlight, there is a rapid increase and accumulation of phytoplankton populations, forming blooms, leading to the natural seasonal cycle of the annual phytoplankton spring blooms [25]. An example, is the massive spring blooms that developed in the water column under the Arctic ice [26]. These under sea ice blooms are becoming more frequent as a response to climate change due to the retreat of sea-ice and thinning ice cover [26]. This is an important aspect reported in literature [26,27,28], and an example on how these communities are responding to the Arctic warming trend. On one hand the excessive light compromises the growth and the open water phytoplankton do not have the capacity to photosynthesize as efficiently as before, but on the other hand the ice melting and the substitution of MYI for FYI along with the melt ponds is a potential key-driver to absorb more light and stimulate massive blooms under-ice [26,27,28].

The substitution of MYI for FYI and the increase of sea level also causes stress in the communities and loss of habitat for some species, and there is a need to understand the feedback expected from these communities regarding these threats [18,19].

On top of that, the increase of CO_2 levels boosts ocean acidification and lower the pH, where in fact, there are some studies (mostly performed in single species laboratory experiments) focused on different implications on phytoplankton communities, reporting the effects of ocean acidification enhanced by the increased CO_2 on phytoplankton [29]. However, the main common conclusion of these experiments is that different effects are expected because the biological feedback is specific to each species [29,30,31].

Ocean circulation changes have an influence on Arctic's marine productivity and ecosystem, the entrance of warmer and salty water from the Atlantic Ocean resulted already in shifts of phytoplankton blooms distribution [22,32,33]. In addition, the increased inputs of warmer and freshwater, along with the nutrients transport, will subsequently alter the salinity and other minerals that can facilitate the appearance of invasive species [33,34,35]. Furthermore, even if there is an increase in nutrients generated by melted ice, nutrient depletion will also occur, due to excessive light and

higher exposure to photosynthetic active radiation (PAR), which possibly will alter the vertical stratification of the water masses. This will further boost competitiveness in the food chain due to nutrients limitation [36,37].

Therefore, climate change driven implications and threats have serious repercussions in the phytoplankton communities, and the expected feedback are: loss of habitat, interspecies competitiveness, metabolic and cellular stress, limitations in the growth of phytoplankton when nutrients are limited or uncontrolled increase in growth under the ice due to excessive light and with the right amount of nutrient availability, hence the early spring blooms, or even a second spring bloom. As a consequence of changes in primary productivity dynamics it is also expected shifts in its communities, and force species to adapt to a new regime or, in some cases, extinction because of the "dead zones", as well as changes in the biogeography of phytoplankton species [16,30,38].

Phytoplankton Communities Dynamics in Svalbard Region

The study of phytoplankton in Svalbard region dates back many years, to understand the dynamics and possible changes that may occur due to climate change impact [39]. Efforts have been made concerning the monitoring of phytoplankton in this region, especially in the Kongsfjorden region, where there is an established monitoring site and has been most intensively studied over the years [40,41]. Early studies in this area have provided knowledge about the dynamics of phytoplankton by using traditional microscopic methods [42,43,44,45]. However, both traditional and molecular methods studies highlight temporal trends observed by the phytoplankton and primary productivity. In fact, typical Arctic species are observed, but also invasive and evident Atlantic species, caused mainly by the increase in the introduction of North Atlantic water masses in Svalbard region. These studies reported clear long term trends shifts in the phytoplankton communities [33,46,47,48].

Therefore, the Svalbard region presents a typical Arctic climate, and the phytoplankton communities and primary productivity in this region are able to endure the harsh conditions that characterized this region [49]. However, as mentioned in the sections above the implications of a changing Arctic are extremely severe and well manifested over the years in Svalbard region. Thus, in Svalbard waters, phytoplankton dynamics and composition vary and are dependent on biological, environmental characteristics and hydrological factors, as well the state of sea ice thickness and cover [43].

In table 1, is reported a compilation of the main findings from studies on phytoplankton communities dynamics in Svalbard region, with more focus on Kongsfjorden and Rijpfjorden areas.

Table 1: Compilation of the main findings from studies on phytoplankton communities dynamics in Svalbard region

Region	Main Methods	Main Findings	Climate Drivers Implications	References
Kongsfjorden	Microscopy	• Observations of Atlantic indicator species, in particular the coccolithophore *Coccolithus pelagicus*; • Dominant haptophyte *Phaeocystis pouchetii* and *chrysophyte Dinobryon balticum.*	• Advection of Atlantic Water into Kongsfjorden; • Glacier runoff.	Hegseth et al. [41] Hasle et al. [42] Halldal et al. [50]
Kongsfjorden	SSU rDNA clone libraries (18S rDNA gene)	• *Alveolate* occupied the highest percentage of taxa in the library of surface sea water; • *Dinoflagellates, diatoms,* and *pico-Prasinophytes* were detected as prevalent phytoplankton through the analysis of libraries.	• Kongsfjorden is influenced by both Atlantic and Arctic water masses; • Inputs from large tidal glaciers that create steep environmental gradients in sedimentation and salinity.	Luo et al. [48]
Krossfjorden and Kongsfjorden	Microscopy and sequencing of 18S rRNA genes	• Data revealed that *diatoms* and *Phaeocystis sp.* were replaced by small nano- and picophytoplankton during late spring, coinciding with low nutrient availability; • The innermost stations showed higher relative abundances of nano- and picophytoplankton throughout, notably of *cyanophytes* and *cryptophytes.*	• Glacier influence, mediated by early meltwater input.	Piquet et al. [51]
Kongsfjorden and Rijpfjorden	Microscopy and Molecular techniques	• Both Kongsfjorden and Rijpfjorden showed inter-annual differences in taxonomic composition, abundance and biomass of phytoplankton during summer, but most taxonomic groups were present every year; • Shift in phytoplankton community composition towards dominance of small-sized phytoplankton.	• Temperature increase; • Increased advection of heat by the West Spitsbergen Current.	CAFF [33]
Rijpfjorden	Microscopy	• The number of taxa varied between years. • Protist community dominated by nanoplanktonic *dinoflagellates (gymnodiniales), ciliates* and *prymnesiophytes*; • Postbloom *P. pouchetii* dominance.	• Higher sea and air temperatures and less sea ice; • Variations in Atlantic Water Boundary Current.	Hop et al. [52]
Wijdefjorden and Rijpfjorden	Microscopy	• Dominance of groups by *flagellates (Dinophyceae* and *Cryptophyceae), Prymnesiophyceae* (primarily non-motile stage of *Phaeocystis pouchetii*) and *Flagellates indet.* (most likely the motile stage of *P. pouchetii*). • Planktonic protists succession pattern in the northern Arctic regions showed a clear linkage with seasonal change in the hydrographic conditions.	• Glacier retreat; • Glaciers discharging fresh and turbid waters directly to the sea; • Atlantic-influenced northern fjords.	Trudnowska et al. [53]

Methodologies to Study Phytoplankton Communities

Understanding the best methodological approaches to investigate the response of primary producers to the dramatic decrease of Arctic ice will be of key importance in order to have effective methodologies to be applied in long term monitoring programs, essential to anticipate future trends.

With respect to the microscopy approach, microscopic counting of phytoplankton is one of the traditional determinations used in the assessment of the ecological status of marine environments [54]. According to literature, the most direct approach for a detailed analysis of phytoplankton community structure is microscopic observations of water samples [55]. However, accurate phytoplankton analysis using this methodology is very time-consuming and highly limited to identify small phytoplankton cells [55] (Table 2). In addition, conventional methods have also led to inherent problems when it comes to classifying the protists community, due to several aspects, but mainly to ‘’*undersampling*’’ [5]. Moreover, microscope cell counts of phytoplankton have also limitation in the detection of cryptic species [56,57,58] (Table 2).

Over the years, molecular methods start to emerge due to the need to circumvent the limitations of conventional microscopy methodologies, and have revolutionized the way we look at microbial communities [59]. The advantage of genomics with high-throughput DNA (deoxyribonucleic acid) sequencing provided by next-generation sequencing (NGS) technologies, advances in computational biology and bioinformatic pathways, was the key point to improve phytoplankton research at a broader level [59,60]. Nowadays, genomics approaches are the most innovative source in marine *in situ* monitoring and provide new insights into the diversity of organisms, as well their abundance and dynamic patterns [59,60] (Table 3).

Among them, metabarcoding technique has proved to be a powerful tool for the identification of multiple species from a bulk sample and became the method most commonly used also for the studies of microbial eukaryotes (protists) community [16,56,57] (Table 3). This method relies on a short section of DNA from a specific gene marker (18S rRNA gene for eukaryotes and 16S rRNA for prokaryotes), on techniques such as NGS, PCR (Polymerase chain reaction) and bioinformatic tools [16,60]. This promising approach allowed even more relevant ecological discoveries, as well the status of the marine environment and how the organisms respond to certain disturbances [61]. Also, through this approach, researchers were able to decipher new lineages of protists and expand the current knowledge of protists and diversity through the taxonomic levels, where in methods based on morphology this could be an issue [61].

Nonetheless, despite the advantages and disadvantages of both methodologies, several authors began to point out that the use of conventional methods associated with molecular techniques can offer new insights into the microbial world and unravel some unanswered questions [33,55,56,62].

Taking into account the different power of resolution of the two approaches for studying phytoplankton diversity, in the present research we focused on both techniques: microscope phytoplankton cell counts and identification and metabarcoding approach. More details on the advantages and disadvantages of each methodology are compiled in tables 2 and 3 with respective references.

Table 2: Compilations of the advantages and disadvantages of microscopic approach

Advantages	Disadvantages	References
Morphology, cellular variability, age, growth, shape and general physical state can be observed as they counted.	Time-consuming and possibility of experimental errors.	Lund et al. [63] Lund et al. [64]
Possibility of using counts as a basis for other determinations.	Statistical procedures are only applicable to entities and not to their constituent parts and cells of algae vary so greatly in size.	Lund et al. [64]
The Utermöhl method using an inverted microscope is best known and most widely accepted technique of phytoplankton enumeration and the sedimentation technique permits the investigator great flexibility.	Biologically sparse samples usually contain insufficient numbers of the rarer species, and thereby produce erroneous community evaluations.	Stofan et al. [65]
The Utermöhl method is a standard method and relatively easy to use because there is lot of comparable data with the same or similar protocols.	Requires a lot of training; Depending on desired taxonomic resolution may require high-quality high-powered compound scopes with specialized features.	Phytoplankton Measuring and Culture Techniques working paper [66]
Microscopy based methods and pioneering researchers contributed to phytoplankton knowledge.	Access to older scientific literature is essential and is often necessary for detailed species descriptions, however, these may be difficult to access.	Karlson et al. [67]
Once acquire a high-quality microscope it can remain for many years.	The initial cost is high.	Karlson et al. [67]
Microscopic determination is the only method by which it is possible to acquire information on the whole species composition of phytoplankton samples.	Small picocyanobacteria and pico-eukaryotes may not settle and/or are too small to be identified, resulting in an underestimation of their numbers and total phytoplankton abundance.	Blauw [58] Helcom [68]
Microscopy is the practical choice for larger planktonic groups, as the equipment needed is readily available and not technically complex; Quantifying cell numbers is essential for reporting toxic species, following species succession, and mapping long-term changes in species distribution.	Taxonomic expertise is becoming rare; Operator fatigue and even taxonomic bias (different time series can place a focus on different taxon groups based on the expertise of available analysts) taxonomic records based on microscopy are rarely complete.	CAFF [33]
More valuable ecological and architecture information of phytoplankton.	Many taxa exanimated microscopically is probably new to science and a source of difficulty when establishing robust taxonomic lists.	Rivera et al. [69]
Microscopy techniques command a unique position in analyzing various features such as morphology, chemical composition and structure, topology, interfacial properties, molecular, microstructure, and micromechanical properties, that allow a better perception of other visible criteria that wouldn't be perceived by other methods.	Identification and quantification of rare species and the ability to distinguish morphologically close or identical species (i.e., cryptic species), or poorly characterized juvenile stages of known species.	Danovaro et al. [56] Venkateshaiah et al. [70]

Table 3: Compilations of the advantages and disadvantages of metabarcoding approach

Advantages	Disadvantages	References
Rapid, less expensive and precise identification of field-collected planktonic organisms; Relatively simple, applicable to all life stages of a given species, can be performed on parts of an organism, is culture-independent and is objective; Specialized training is not needed; Revolutionized the understanding of microbial eukaryote diversity and ecology.	The primer choice could significantly affect the results by amplifying non-target groups and target sequences more than others; Using a multiple primer sets, or multiple barcode genes for different target taxonomic groups will help to reduce primer bias, but it is more costly and time-consuming, so that could not be a preferable choice by the researcher.	McManus et al. [71] Stoeck et al. [72]
DNA metabarcoding opens the door to high-throughput biodiversity assessment and has been implemented for several ecological applications; field with an abundant literature available.	DNA metabarcoding applications require short amplicons and robust PCR conditions to achieve unbiased amplification from a mixture of several DNA templates; The selected markers must correspond to a genomic loci flanked by two highly conserved priming sites to simplify PCR conditions and reduce differences in amplification among the different DNA templates.	Coissac et al. [73]
It can be implemented for both modern and ancient environmental samples; Metabarcoding datasets are also taxonomically more comprehensive and less reliant on taxonomic expertise; reliable method for recovering alpha- and beta-diversity information from large-scale, field-collected data set.	Considerable investment needed to build comprehensive taxonomic reference libraries; Need for making progress in DNA sequencing technologies that will allow elimination of a DNA amplification step, and also need for make comprehensive taxonomic reference libraries composed of whole organellar genomes and repetitive ribosomal nuclear DNA	Taberlet et al. [74] Ji et al. [75]
The presence of the 18S rRNA across all eukaryotes, its extensive occurrence in public reference databases and the availability of generalist primers make this gene the best universal marker available to date for eukaryotes.	Intracellular polymorphism, rDNA copy number variation and presence of pseudogenes; The sequencing depth of the next generation technologies, the different copies, pseudogenes and other variants of the 18S rRNA of each organism could consequently lead to inflated diversity metrics by increasing the number of predicted OTUs.	Decelle et al. [76]
Metabarcoding approach, have provided further insights into protist biodiversity with unprecedented detail; the number of OTUs could easily exceed the number of protist taxa identified by traditional methods; Better coverage of the 'rare biosphere'; Can uncover cryptic species.	Could provide an inaccurate or wrong estimation (under/over estimation) of the actual biodiversity of the sample due to variability in primers, PCR conditions, sequencing technology and bioinformatics pathways used; Moreover, cases occurred in which different species shared the same barcoding gene target in V4 region.	Danovaro et al. [56] Piredda et al. [57] Mordret et al. [77]
Metabarcoding technique was shown to be more valuable for assessing pico-phytoplankton and novel phytoplankton diversity than morphological identification due to the in-depth sequencing that is achieved.	When comparing with morphology methods, different data bases should be considered; Metabarcoding don't detect dead cell.	Rivera et al. [69] Huo et al. [78]

Objectives

The thesis presented herein - “Biogeography of Arctic Eukaryotic Microbiome: a comparative approach between 18S rRNA gene metabarcoding and microscopic analysis” - makes part of an international collaboration between CIIMAR and the Norwegian Polar Institute (NPI) through a joint research based on a long term program on Environmental Monitoring of Svalbard and Jan Mayen (MOSJ) led by NPI. For this study microplanktonic samples from the 2016 MOSJ Arctic campaign were used to accomplish two main objectives:

1 - Investigate the biogeographic patterns of Arctic phytoplankton diversity and distribution along two oceanographic transects (covering Fjord, Coastal and Open ocean regions) in the Marginal Ice Zone (MIZ) around Svalbard, using 18S rRNA gene metabarcoding approach.

2 - Perform a comprehensive comparison between the 18S rRNA gene metabarcoding approach, using new generation sequencing techniques, and the classic microscope cell counts for phytoplankton analysis.

In this sense, by combining different approaches (microscope analysis and 18S rRNA gene metabarcoding), this study aims to understand the advantages and limitations of each methodology and identify the relevant complementary information to achieve a full understanding of Arctic phytoplankton diversity and to produce methodological guidelines for Arctic phytoplankton monitoring programs, essential to fill critical gaps concerning the emerging questions that has been raising due to a changing Arctic.

Materials and Methods

Study Area

Sampling was conducted along two oceanographic transects (Kongsfjorden – western transect and Rijpfjorden – northern transect) in the Marginal Ice Zone around Svalbard (Figure 1) by a CIIMAR research team during 25 July and 4 August 2016. This sampling program was integrated into a long term international program of Svalbard and Jan Mayen (MOSJ), led by Norwegian Polar Institute (NPI), that focuses on relevant environmental information to understand the effects of climate change in Svalbard region at physical, chemical and biological levels.

Kongsfjorden (79ºN, 12ºE) is a glacial open fjord, located on the west coast of Spitsbergen, which is part of the Svalbard archipelago in the Arctic Ocean. This transect is 27 km long, and the region is influenced by both Atlantic and Arctic water masses, which carries relatively warm and salty water into the west Spitsbergen current [36,40]. Rijpfjorden (80ºN 22º30'E) is a fjord at the northern coast of Svalbard, and is oriented with a wide northward opening towards the cold Arctic Ocean. It is more influenced by Arctic waters and less influenced by warm North Atlantic waters [52].

Figure 1 – Map highlighting the sampling sites of MOSJ-ICE2016 expedition along two oceanographic transects (Kongsfjorden – western transect and Rijpfjorden – northern transect) in the Marginal Ice Zone around Svalbard. Map generated by Orange data mining software (version 3.26.0) [79].

18S Metabarcoding- Sampling Sites: Data acquisition and description

During the MOSJ-ICE2016 expedition, sampling for microplankton metabarcoding analysis was performed by using the RV Lance as a research platform. It includes two oceanographic transects crossing Fjord, Coastal and more Offshore ocean domains along the western (Kongsfjorden) and northern (Rijpfjorden) coast of Svalbard. A total of 33 sample stations were collected at three distinct depths: surface (2m-5m), deep chlorophyll maximum (DCM) (17m-50m) and bottom (50m-1035m). For protists community analysis, between 3 to 5L of water were concentrated through filtration, using 0.22 µm Sterivex filters (Millipore) and preserved (-80ºC) on board to be used for later environmental DNA isolation and 18S metabarcoding analysis using specific methodologies described below [80]. Sampling was performed by a CIIMAR research team. See the summarized description and the IDs of each sample in table 4 (Kongsfjorden) and table 5 (Rijpfjorden).

Table 4: Coordinates, sampling depth and water filtration parameters from Kongsfjorden stations

Sample ID	Latitude	Longitude	Water Depth (m)	Collection filter pore (um)	Volume Filtrated (L)
G1_S	78,89	12,47	5	0,22	NA
G1_M	78,89	12,47	30	0,22	NA
G1_B	78,89	12,47	65	0,22	NA
KB6_S	79,06533333	4,184166667	2	0,22	1
KB6_M	79,06533333	4,184166667	25	0,22	1,4
KB6_B	79,06533333	4,184166667	50	0,22	1,45
KB3_S	79,05166667	11,1075	5	0,22	1,6
KB3_M	79,05166667	11,1075	50	0,22	1,4
KB3_B	79,05166667	11,1075	332	0,22	1,2
KB0_S	79,05166667	11,1075	5	0,22	1,5
KB0_M	78,95691667	11,94705	25	0,22	1,6
KB0_M	78,95691667	11,94705	50	0,22	1,63
KB0_B	78,95691667	11,94705	320	0,22	1,83
V12_S	78,93023333	12,38136667	5	0,22	2,345
V12_M	78,93023333	12,38136667	25	0,22	0,835
V12_B	78,93023333	12,38136667	200	0,22	2,95
V6_S	78,90683333	7,766666667	5	0,22	3,11
V6_M	78,90683333	7,766666667	25	0,22	3,12
V6_B	78,90683333	7,766666667	200	0,22	4,35
HGIV_S	78,97533333	9,485	5	0,22	2,55
HGIV_M	78,97533333	9,485	43	0,22	3
HGIV_M	78,97533333	9,485	50	0,22	3,4

Table 5: Coordinates, sampling depth and water filtration parameters from Rijpfjorden stations

Sample ID	Latitude	Longitude	Water Depth (m)	Collection filter pore (µm)	Volume filtrated (L)
R1_S	80,12558333	22,15266667	5	0,22	1,46
R1_M	80,12558333	22,15266667	17	0,22	2,45
R1_B	80,12558333	22,15266667	168	0,22	2,9
R4_S	80,65566667	22,0855	5	0,22	2,9
R4_M	80,65566667	22,0855	28	0,22	2,4
R4_B	80,65566667	22,0855	124	0,22	1,13
R6_S	81,205	22,11116667	5	0,22	2,8
R6_M	81,205	22,11116667	50	0,22	2,8
R7_S	81,50783333	21,85	5	0,22	2,6
R7_M	81,50783333	21,85	25	0,22	2,4
R7_B	81,50783333	21,85	1035	0,22	4,5

Water Column Physico-Chemical Parameters

Contextualize environmental data was collected at each station by having several sensors attached to the multi-bottle carousel water sampler generating water column profiles of Salinity, Temperature, Fluorescence, Chlorophyll, Photosynthetically Active Radiation (PAR) and Superficial Photosynthetically Active Radiation (SPAR). In addition, water was collected at selected depths to be analysed in laboratory for Ammonium (NH_4^+), Oxygen (O_2), Nitrite (NO_2^-), Nitrate (NO_3^-), Phosphate (PO_4^{3-}) and Silicate (SiO_4^-). These analyses were performed by the Norwegian Ocean Institute research team.

DNA Extraction, Library Preparation, and Sequencing of SSU rRNA Amplicons

The Sterivex™ filters stored at -80°C on board were transported to CIIMAR in frozen conditions. In the lab sterivex filters were defrosted, at room temperature to extract the total DNA using the PowerWater® Sterivex™ DNA Isolation Kit protocol (MO BIO Laboratories, Inc., Portugal) following the manufacturer's instructions.

The target V4 region of 18S rRNA gene was amplified with the primer set described in Stoeck et al. [72], TAReuk454FWD1 (5' – CCAGCASCYGCGGTAATTCC – 3') and TAReukREV3_modified (5' – ACTTTCGTTCTTGATYRATGA – 3'), with the exception of reverse primer (TAReukREV3_modified) which had an additional TGA triplet added at the 3' end compared to the original Piredda et al. [57]. The 18S-V4 rRNA gene region amplified by PCR was used to build Illumina paired-end libraries sequenced on an Illumina MiSeq platform using 2 x 300 bp, V3 Chemistry (Illumina). These steps

were performed by LGC Genomics (LGC Genomics GmbH, Berlin, Germany) and a detailed description is given in Sousa et al. [80].

Bioinformatics Pipeline: Preprocessing and Taxonomic Assignment of V4-18S rRNA Amplicons

Upstream Sequence Analysis

The mothur pipeline was used to pre-process the 18S-V4 rRNA amplicons from the MOSJ-ICE2016 campaign following the MiSeq Standard Operating Procedure (SOP; https://www.mothur.org; Kozich et al. [81]). Individually, the forward and reverse reads of each library were joined from raw Illumina fastq files (mothur v. 1.39.5; Schloss et al. [82]). Merged reads with ambiguities (aka N) and shorter than 300 bp were excluded as well as the ones with homopolymers (n>8).

The remaining sequences were dereplicated (based on 100% similarity). After dereplication step, the unique sequences that differ within 3 base pairs similarity from a more abundant one were clustered together. Chimeric sequences were identified *de novo* and removed with UCHIME [83].

Taxonomic assignment of the unique reads (obtained after chimera removing) was performed using standalone BLAST in BLAST+ suite (Altschul et al. [84]; Camacho et al. [85]) against the PR^2 reference database 4.12.0 for mothur [86]. Undesirable lineages, Metazoa|Fungi|Streptophyta|Ulvophyceae|Rhodophyceae were excluded from the dataset.

Afterward, a distance matrix was built and the sequences clustered into OTUs (Operational Taxonomic Units) using 0.01 cutoff values with vsearch [87]. OTUs containing only one read (singletons) were removed from downstream analyses. Finally, 18S-V4 OTU table at 99% of similarity was generated and used as input in the following downstream analysis.

Downstream Sequence Analysis

Taxonomic Composition

The distribution of protists taxa across MOSJ-ICE2016 campaign was analysed at different taxonomic levels (Phylum, Class, Order, Family, Genus and Specie). Taxonomic distribution plots were performed using several R-packages, namely, Phyloseq, ggplot2, tidyverse and scales [88,89,90,91]. The taxonomic profiles of Kongsfjorden and Rijpfjorden transects were presented in the results section at Phylum, Class, Family and Genus levels. The distribution of protists taxa across MOSJ-ICE2016

collection at phylum, class, order, family, genus, specie and OTU levels is provided in Additional file 1.

Relations between shared OTUs and samples stations

A Venn diagram was performed with the 18S OTU table, to see possible logical relations between shared OTUs. This analysis was run in the R environment [92] using the R-packages vegan and venn [93,94].

Diversity, Structure and Environmental Analysis

Alpha-diversity and Beta-diversity were obtained in the R environment [92]. For alpha diversity, the following estimators Observed (Count of unique OTUs in each sample - Number of observed samples), Chao1 (Abundance-based estimator of species richness) and Shannon-Index (Estimator of species richness and species evenness) were run using the R-package vegan [93], for different sampling depths (surface from 2m-5m; Medium referring to deep chlorophyll maximum from 17m-50m; Bottom from 50m-1035m) within the Rijpfjorden and Kongsfjorden transects.

For beta diversity, the analysis was performed using vegdist function in vegan package. The Bray-Curtis index (based on abundance data) was calculated and then used in subsequent analysis (clustering, ordination and correlation with environmental parameters). Cluster dendrograms presenting hierarchically organized samples were built with the hclust and average method. Also, for multivariate and comparative analysis, the abundance dataset was normalized with a random subsampling to the median value of reads in samples with the “rrarefy” function [95,96].

The ordination and correlation with environmental variables between OTU table were explored with Mantel statistics based on Spearman’s rank correlation and the BIO-ENV analysis [97] was performed using the Bray-Curtis dissimilarity matrix. Within a set of environmental variables, BIO-ENV indicates a subset of environmental variables that shows the highest explanatory values. After the identification of the highest explanatory values, the environmental analysis was performed using non-metric multidimensional scaling (NMDS). In addition, a principal component analysis (PCA) was accomplished, to fully exploit our data in order to analyse a distance matrix to see the distribution of samples without biological association [92].

The environmental and biological correlations were analyzed in the ORCA platform. This platform is an R-based scripts workbench with a graphical user interface that allows to seek meaningful biological and environmental relationships [98]. The rank of statistics used was Spearman correlations with a level of significance of 95%, among interest groups of taxa at class level.

Microscope cell counts and identification

The planktonic protists (PHT and MIT) were collected with the CTD Niskin bottles. After collection, samples were preserved with glutaraldehyde and hexamine buffered formaldehyde in final concentration of 0,1% and 1%, respectively and the subsamples were qualitatively and quantitatively analyzed using protocols described by Utermöhl [99] and modified by Edler [100]. In the laboratory subsamples were settled in a counting chamber (Hydro-Bios, Kiel, Germany) for 8-24 h (Edler et al. [101]) and enumeration and identification were performed at 100–600× magnification using an inverted Nikon Ti-S and Nikon TE-300 light and epifluorescence microscopes, equipped with DIC and KF contrasts and picture acquisition system.

The microplanktonic cells (>20 µm) were counted from the entire chamber surface at 100× magnification, while the nanoplanktonic cells (3–20 µm) were counted at 400-600× magnification by moving the field of view along the length of the transverse transects. Taxa were identified to the lowest possible taxonomic level in accordance with the World Register of Marine Species WoRMS, (http://www.marinespecies.org) and grouped into size-classes. However, when it was not possible to have a detailed identification, specimens were assigned to higher-rank taxa, incertae sedis (i.e., protists, but not determined to higher rank) or three size classes (3, 3– 7, and >7 µm) in case of unidentified flagellates (*Flagellates indet*.) [52].

Afterwards total abundance per liter was calculated based on the investigated area of chamber surface, chamber volume and chamber conversion coefficient for the different magnifications (chamber surface to field of view ratio). For biovolume estimations, taxa were measured and the average cell size was used to calculate the biovolume from equivalent geometrical shapes [102]. When the measurements were not possible, mean biovolumes were taken from available HELCOM database (http://www.helcom.fi/helcom-at-work/projects/phytoplankton) or the Nordic Microalgae web base (http://nordicmicroalgae.org), developed and operated by the SMHI (Swedish Meteorological and Hydrological Institute) developed and operated by the SMHI [52].

The microscopic phytoplankton dataset was divided into two categories: PHT (Phytoplankton) and MIT (Microplankton) with a depth range between 2m and 270m. For the PHT category, the samples were collected directly from the niskin bottles of chosen depths. The amount of sample collected was 190 ml (plus 10 ml of preservatives). This type of samples was counted for qualitative and quantitative data. For MIT samples, the collection was intended to find big rare species, and thus 24L of water were filtered through and concentrated in a volume of 90 ml.

18S Metabarcoding vs Microscopic Analysis - Comparison

A comprehensive comparison between the 18S metabarcoding approach and the classic microscope cell identification was performed to study in depth the diversity and distribution of phytoplankton communities in the Arctic Ocean detected by two different approaches. To this aim, the stations and depths that were not common between the two methods were excluded from the comparison. However, as the stations where samples were collected for PHT were different from the MIT stations, the comparison was made as follows: 18S vs PHT data set (17 common samples) and 18S vs MIT data (16 common samples). Details on which stations and depths were included in each comparison is given in the results section. Also, a volume normalization between the different methods was performed, in order to proceed with the comparison with the equal volumes. In addition, it's important to highlight that for the methodological comparison, microscopic abundance data were normalized for the volume of water sample that was collected during the PHT and in the MIT procedure.

To proceed with the comparison, several steps were followed. First, we identified the total taxa detected by each method and the most abundant phytoplankton groups (with a relative abundance >1%) at different taxonomic levels (Phylum, Class, Family, Genus and Specie). Then, we proceeded with the detection of groups that were present in both methods taking into account all samples, followed by the detection of matching groups within the same samples and comparing differences in the relative abundances of common groups that matched between the same samples.

Finally, we identified the groups that were unique in each methodology at the highest and lowest taxonomic levels. We emphasize that the relative abundances of each taxonomic group were determined taking into account the value of the abundance detected in a given sample station and divided by the total abundance for that same station, multiplying by 100.

Results – Part A

Phytoplankton communities distribution across two transects in Svalbard Coast

18S rRNA Metabarcoding Data Presentation

The 18S rRNA dataset used in this study was composed of 1.998.533 merged raw reads. Filtering procedure generated a final curated dataset including 1.653.809 (236.747 unique) sequences. After removal of metazoa and the other undesirable lineages the number of protist sequences was 802536 (173078 unique). Clustering at 99% similarity produced 25668 OTUs, which after the removal of those OTUs containing one sequence (singletons) were reduced to 5876 (Table 6). Dataset description of all individual samples is provided in Additional file 2.

Table 6: Dataset description of 18S raw table to OTU table

Total Samples	Total merged raw reads of all groups	Total cleaned reads*	Total cleaned reads (protists)	Total protists	Total cleaned reads (protists unique)	*Total OTU (obtained at 99%)	**OTU after removal of singleton
33	1.998.533	1.653.809	236.747	802536	173078	25668	5876

*Total cleaned reads, after removing chimeras.

**OTUs after removal of those containing only one sequence (singletons).

Number of unique and shared protists OTU across transects and ocean features

The Kongsfjorden and the Rijpfjorden transects shared a total of 1057 OTUs, with Kongsfjorden showing the highest number of unique protists OTUs (Figure 2a).

Highlighting the number of protists OTUs between the two transects at different depths, both transects shared 327 OTUs at surface, 287 at deep chlorophyll maximum and 653 at the bottom (Figure 2).

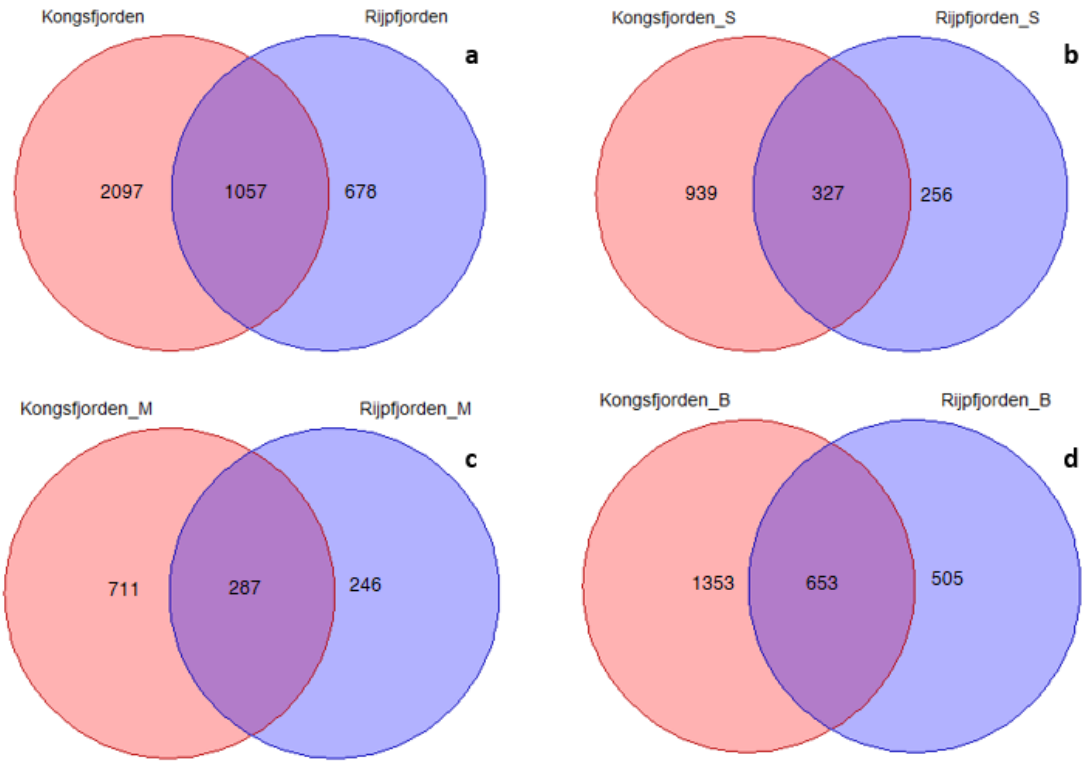

Figure 2 - Venn Diagram analysis. **a)** All samples from Kongsfjorden vs Rijpfjorden transects; **b)** Surface samples from Kongsfjorden vs Rijpfjorden transects; **c)** DCM samples from Kongsfjorden vs Rijpfjorden transects; **d)** Bottom samples from Kongsfjorden vs Rijpfjorden transects.

Regarding the characteristics of oceanic features, namely: Fjord, Coastal and Open Ocean, three venn diagrams were generated (Figure 3). Results showed that the number of unique protists OTUs decreased from Fjord to Open ocean stations for samples from both transects together (Figure 3a). In addition, the number of shared OTUs was higher between Coastal and Open ocean for Kongsfjorden, although an opposite trend was observed for Rijpfjorden (Figure 3b,c).

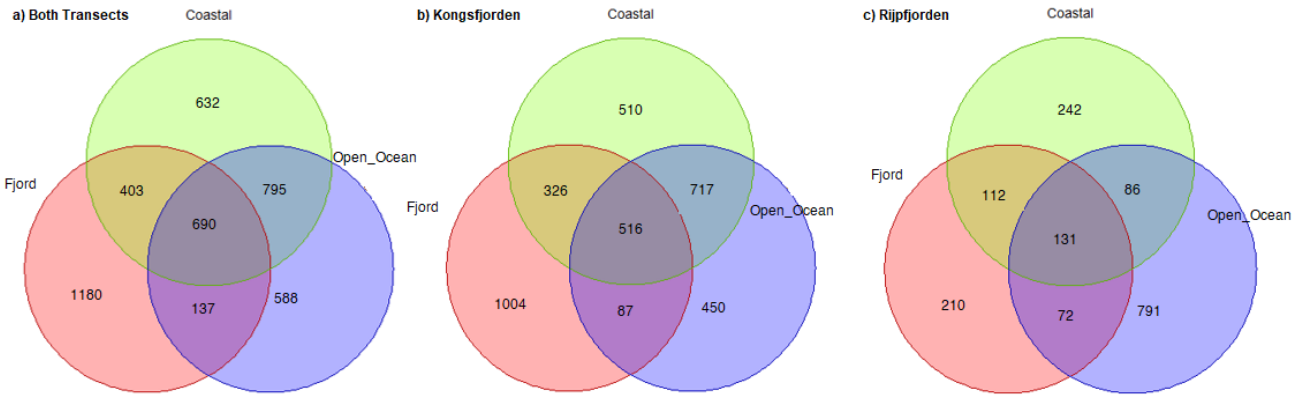

Figure 3 - Venn Diagram analysis. **a)** All features from Kongsfjorden vs Rijpfjorden transects; **b)** Kongsfjorden all features; **c)** Rijpfjorden all features.

Alpha Diversity

Alpha diversity (Observed OTUs, Chao1 Index and Shannon Index) for the protists communities, varied with depth, within Kongsfjorden and Rijpfjorden transects (Figure 4).

The number of observed OTUs decreased from bottom depth to surface in both transects, however in Rijpfjorden transect, the DCM and surface depth presented similar observed OTUs numbers. Regarding the Chao1 index, a similar pattern can be verified, with slight differences between the three depths. As for Shannon measure, it is noticeable a higher diversity at bottom and DCM depth, which suggest a more even distribution at the Kongsfjorden and Rijpfjorden transects at these two depths.

We can also notice that Kongsfjorden transect presented higher values regarding alpha diversity measures, compared with Rijpfjorden transect.

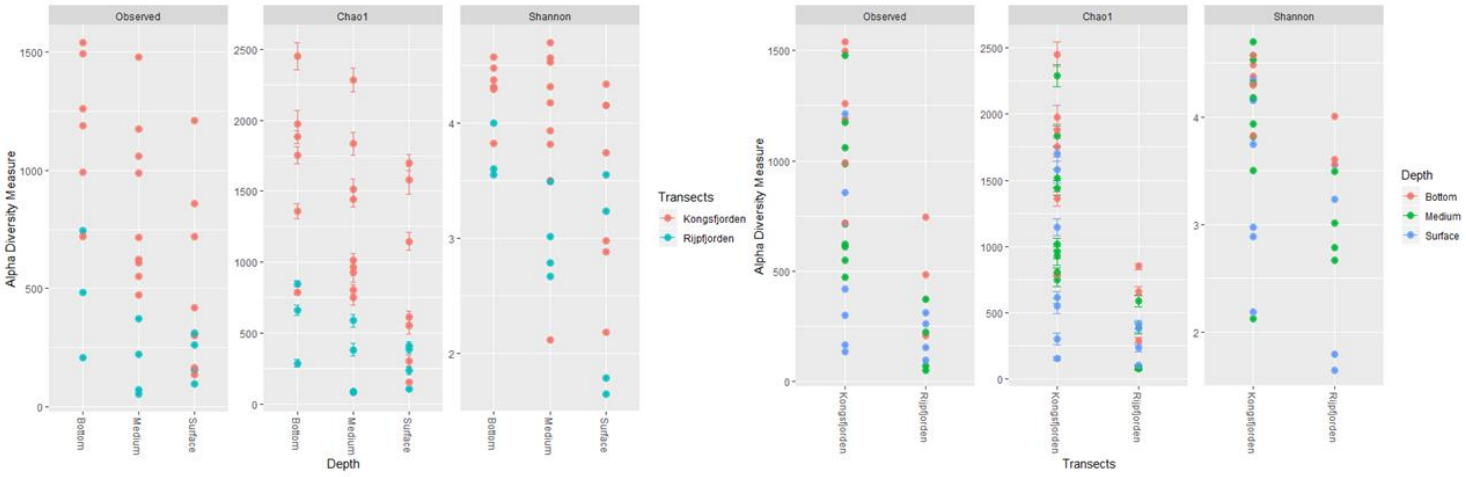

Figure 4 - Alpha Diversity (Observed, Chao1 and Shannon measure) for the different sampling depths (Surface from 2m-5m); (Medium referring to deep chlorophyll maximum from 17m-50m); (Bottom from 50m-1035m) and within the Rijpfjorden and Kongsfjorden transects.

Beta Diversity

Cluster Analysis

A hierarchical cluster analysis with Bray-Curtis distance was performed in order to understand the similarity of protists community structure across the different samples (Figure 5). Results from the hierarchical cluster analysis performed for all samples revealed a similarity pattern between samples that are more offshore (Open ocean), and between Coastal/Fjord samples, where each group of those samples were placed in different clusters showing a divergency in terms of protists communities (Figure 5a).

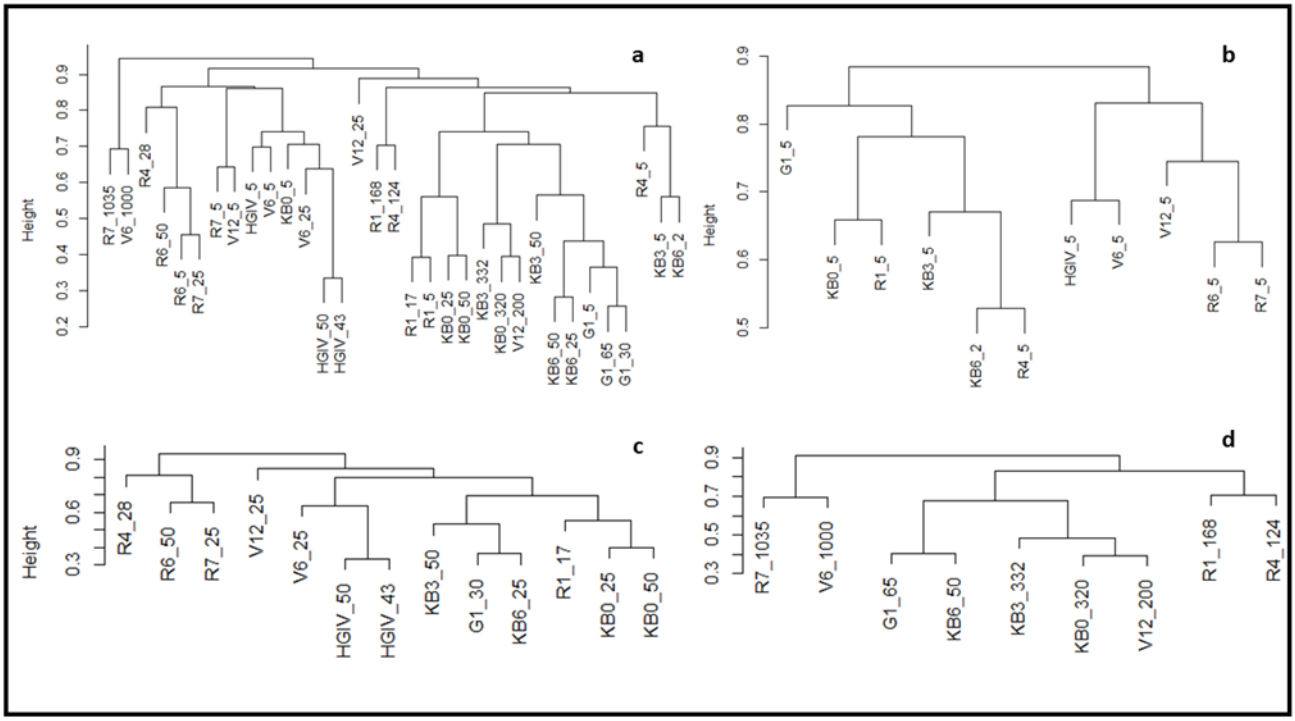

Figure 5 - Hierarchical clustering analysis performed with Bray-Curtis distance. **a)** Cluster Dendrogram with samples from both transects and all depths. **b)** Cluster Dendrogram with samples from both transects at surface depth. **c)** Cluster Dendrogram with samples from both transects at deep chlorophyll maximum depth. **d)** Cluster Dendrograms with samples from both transects at bottom depth.

When samples were analyzed by depths, the same pattern was generally observed (Figure 5b,c,d). At surface, samples from Coastal/Fjord were included in the same cluster, and a second cluster was generated with samples from offshore waters. As for DCM depth, there is a cluster that includes the majority of the samples (R4_28, V12_25, KB0_25, KB0_50) from Coastal waters, and the other samples were mixed together with Open ocean samples and Fjord samples. Finally, analyzing only the bottom samples, three clusters were generated, the first one includes samples from Open ocean (R7 and V6), the second group includes samples from Kongsfjorden transect (G1, Fjord; V12, Coastal) and the last group includes samples from Rijpfjorden transect (R1, Fjord; R4, Coastal). These results suggested that the protists communities are distributed according to sample depths and ocean features.

Taxonomic Analysis

Phylum Level

Metabarcoding dataset from MOSJ-ICE2016 showed that protist communities were distributed within 12 phyla, being *Ciliophora* and *Dinophyceae* the most dominant in all samples across both transects (Figure 6).

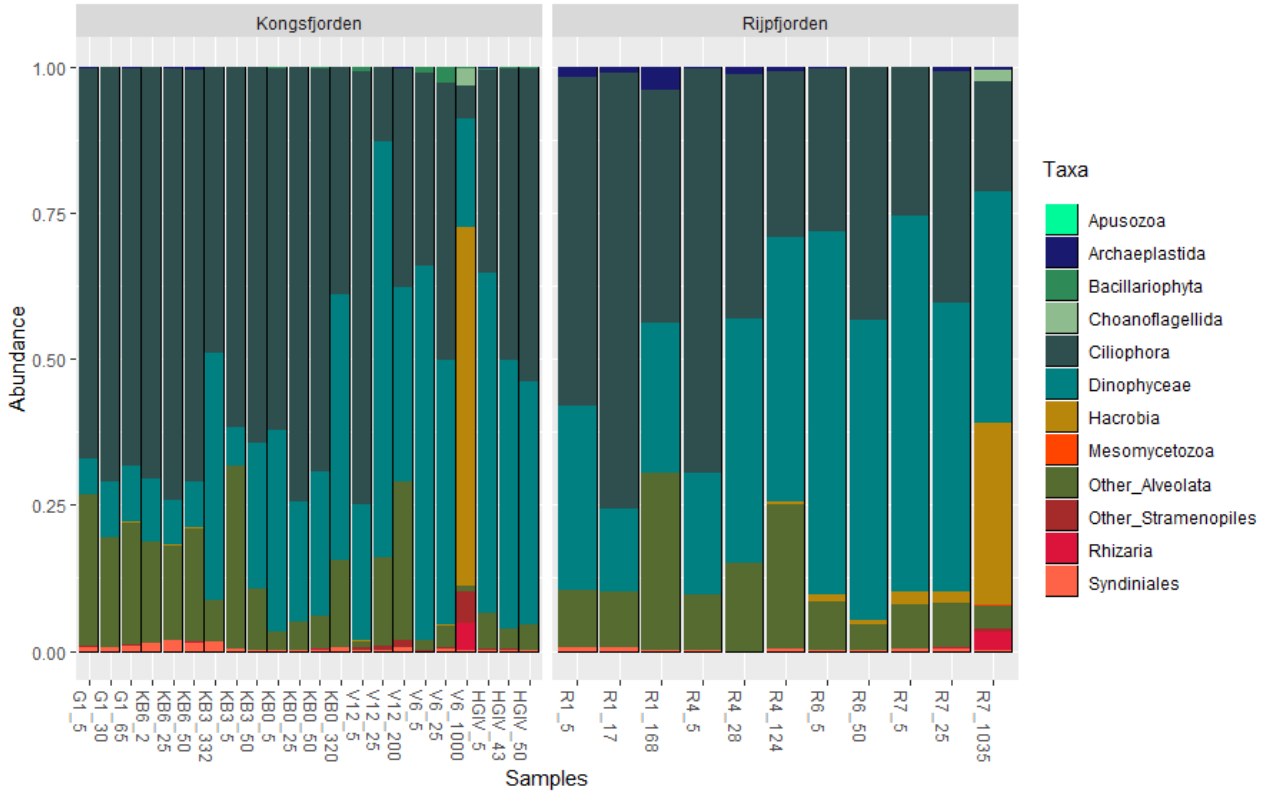

Figure 6 - Taxonomic profile of MOSJ-ICE2016 protists. Relative abundance of protists retrieved from the surface, deep chlorophyll maximum and bottom depth at Kongsfjorden (on the left) and Rijpfjorden transect (on the right), at phylum level retrieved from the OTU table assigned with PR^2 classification.

Phylum *Other_Alveolata*, was also present in all samples of both transects, and *Hacrobia* seems to mark more presence in stations V6_1000 and R7_1035, representative of more offshore waters. As for the remaining phyla the abundance distribution was less perceivable, however a closer detail can confirm that in stations V6_1000 and R7_1035 were the stations with higher number of representative phyla.

Class Level

As for class level, the 18S metabarcoding dataset possesses a repertoire of 22 protists taxa. The different classes were superimposed in a nonmetric multidimensional scaling (NMDS) plot, in order to have a better insight on the distribution of the different classes within the different samples from both transects. When analyzing the NMDS, we can clearly see a differentiation between samples from the two transects, with the exception of the R7_5 station. From the 22 taxa at class level retrieved from the OTU table, *Chlorophyta, Ciliophora* and *Dinoflagellata* dominate and are cosmopolitan distributed within stations from both transects (Figure 7).

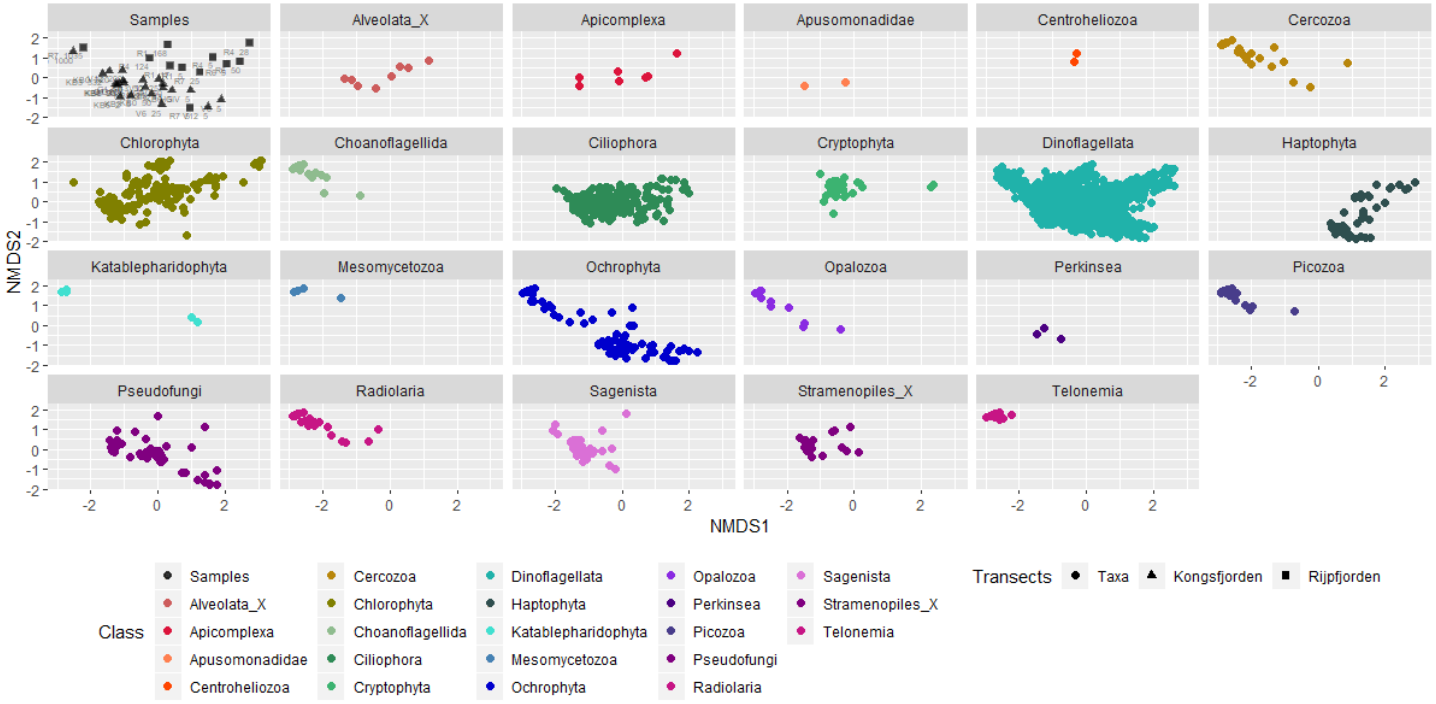

Figure 7 - Taxonomic profile of MOSJ-ICE2016 protists, retrieved from the surface, deep chlorophyll maximum and bottom depth at Kongsfjorden and Rijpfjorden transect, at class level retrieved from the OTU table assigned with PR^2 classification presented with NMDS of individual taxa of protists.

However, when it comes to the remaining class, some are more representative of Kongsfjorden and others from Rijpfjorden transects. For example, *Alveolata_X*, was not very dominant and seems to appear closer to the fjord and the coast in the two transects, but with more representativeness in Kongsfjorden. *Apicomplexa, Apusomonadidae, Centroheliozoa* were also not very dominant, and its presence was more evident in Rijpfjorden transect, with the exception of *Apusomonadidae* which reveal to be more present in Kongsfjorden.

Similar patterns can be observed in certain classes, such as *Cercozoa, Choanoflagellida, Ochrophyta, Opalozoa, Picozoa, Pseudofungi, Radiolaria, Sagenista*

and *Stramenopiles_X*, where they were mainly distributed across Kongsfjorden transect. On the other hand, *Haptophyta* revealed an opposite pattern (mainly distributed in Rijpfjorden) and *Cryptophyta* seems to occur in samples from Kongsfjorden and Rijpfjorden, especially the ones located in Coastal/Fjord stations.

Finally, *Katablepharidophyta, Mesomycetozoa, Perkinsea* and *Telonemia* classes were more dominant in the Open ocean in deeper stations, R7_1035 and V6_1000.

Family Level

Concerning the taxonomic family level, the dataset possesses a repertoire of 80 taxa, and the most dominant families in both transects were: *Dinophyceae, Cyclotrichium_like_organism, Nassophorea, Picozoa_X* and *CONThreeP* (Figure 8).

Once again, stations V6_1000 and R7_1035 were found to be the stations with high number of taxa diversity.

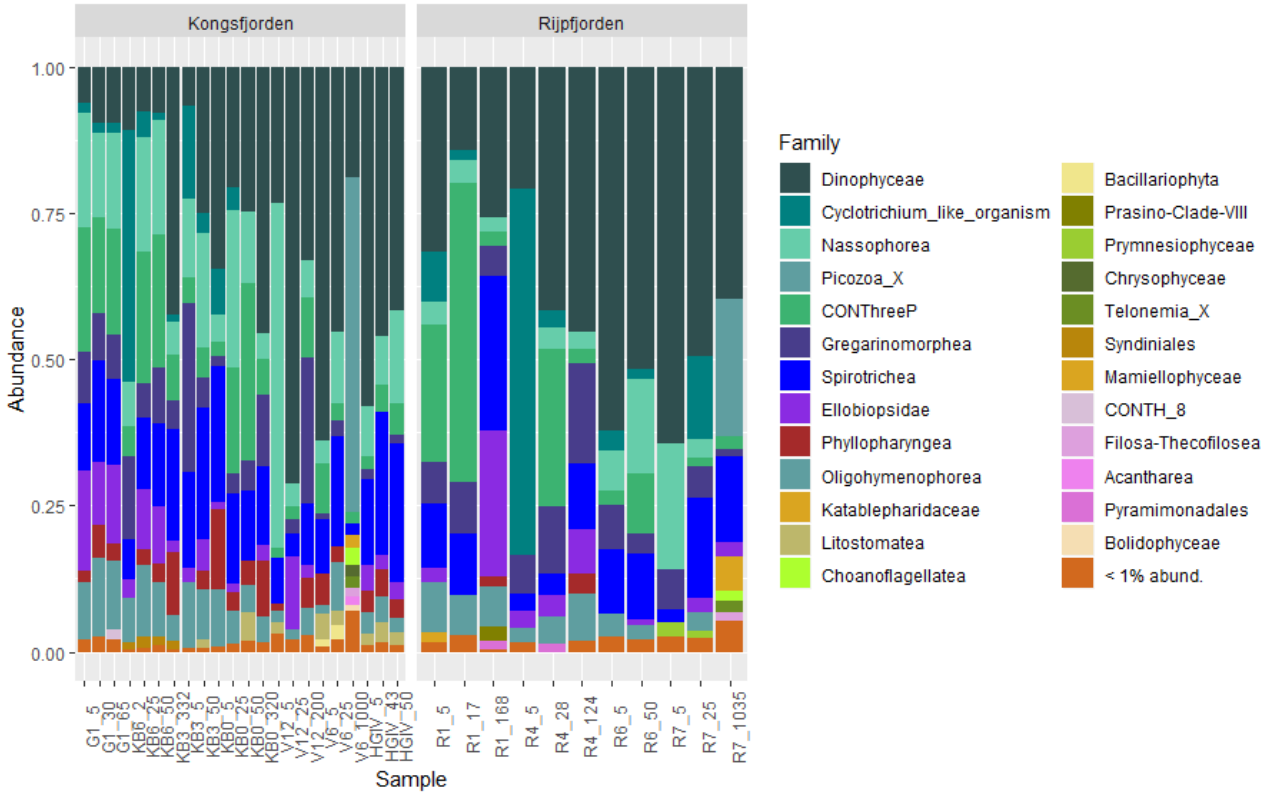

Figure 8 - Taxonomic profile of MOSJ-ICE2016 protists. Relative abundance of protists retrieved from the surface, deep chlorophyll maximum and bottom at Kongsfjorden (on the left) and Rijpfjorden (on the right). Relative abundance of family level condensed into low abundance (<1%) retrieved from the OTU table assigned with PR^2 classification.

Genus Level

Taxonomic analysis at lower taxonomic level identified a total of 407 genus of protists in 18S metabarcoding dataset (Figure 9). *Dinophyceae_XXX* was found to be the most dominant and cosmopolitan genus across all stations from both transects. *Cyclotrichium, NASSO_1, Picozoa_XXX* and *Askenasia* were also found to be highly abundant genus in MOSJ-ICE2016, although not presented in all stations.

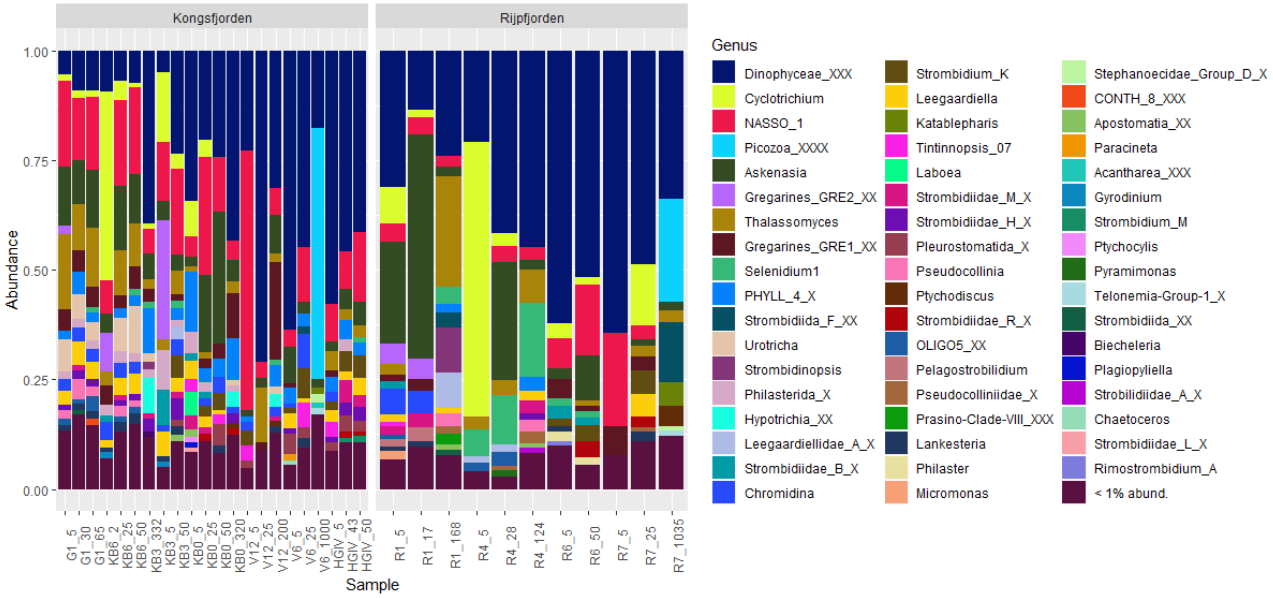

Figure 9 - Taxonomic profile of MOSJ-ICE2016 protists. Relative abundance of protists retrieved from the surface, deep chlorophyll maximum and bottom at Kongsfjorden (on the left) and Rijpfjorden (on the right). Relative abundance of genus level condensed into low abundance (<1%) retrieved from the OTU table assigned with PR^2 classification.

Correlations between the different groups of protists

Spearman's correlations between the different taxonomic groups at phylum and class level are presented in figures 10 and 11, respectively. Interestingly, this analysis showed that all the significant correlations between the different phyla and classes were positive. There are however, certain taxa that are stronger correlated than others (e.g. at phylum level, *Archaeplastida, Ciliophora, Other_Alveolata, Other_Stramenopiles, Rhizaria* and *Syndiniales*); (e.g. at class level, *Apicomplexa, Chlorophyta, Ciliophora, Picozoa* and *Sagenista*).

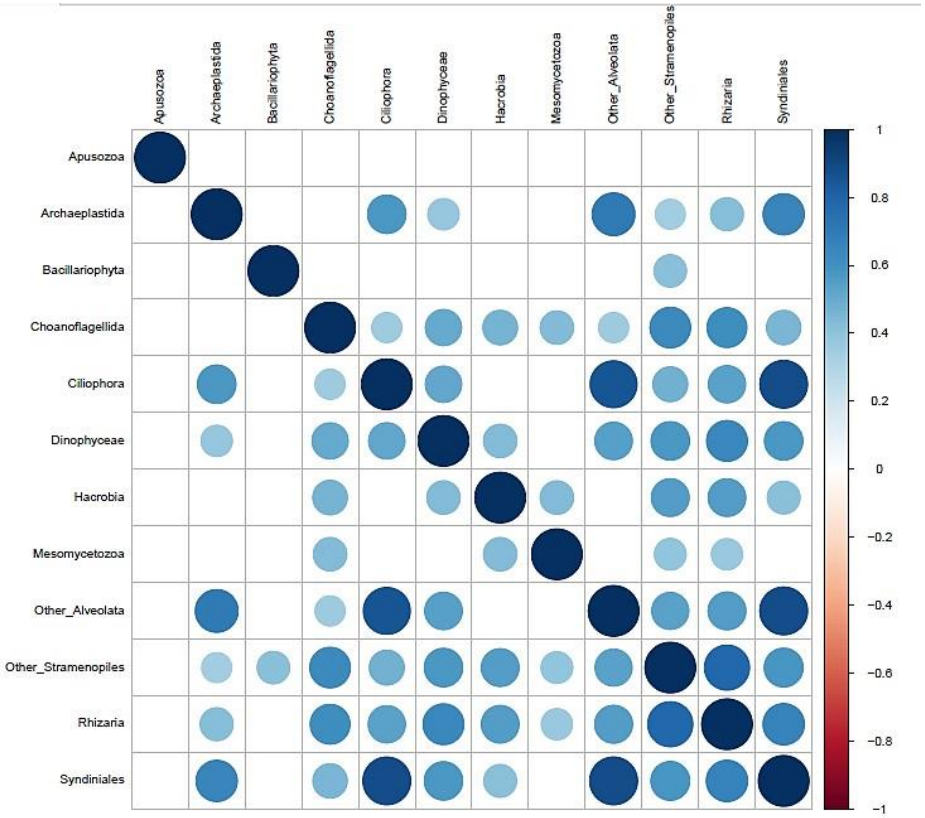

Figure 10 - Spearman's correlations between the different phyla.

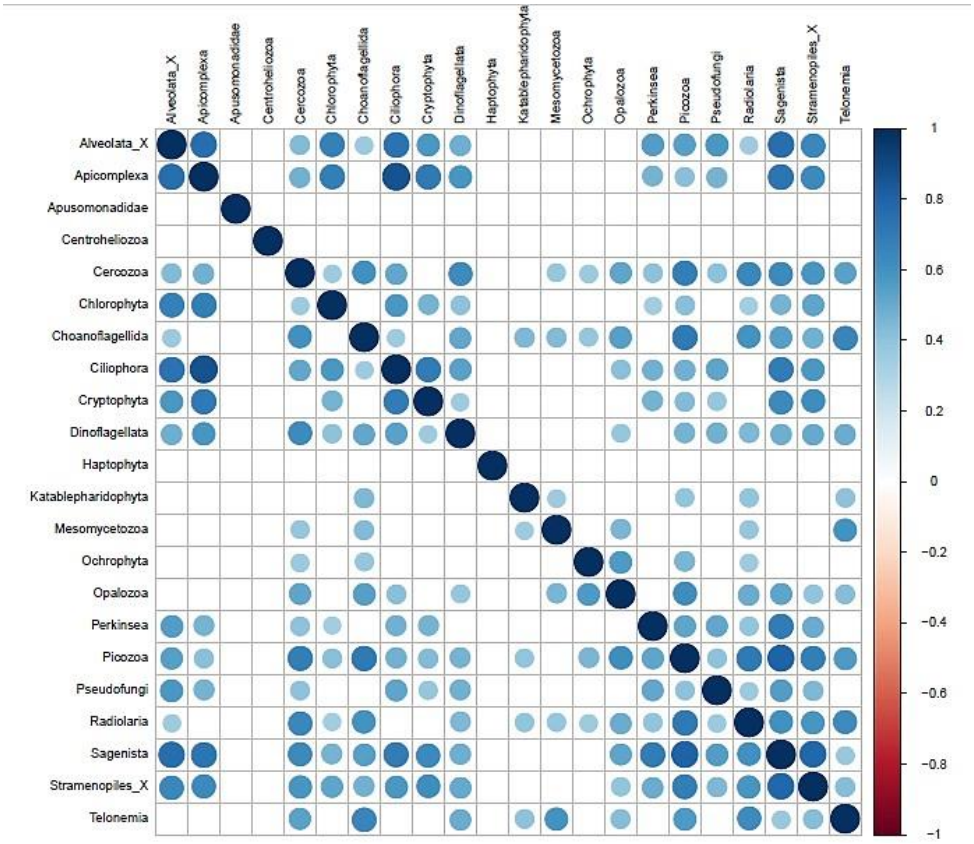

Figure 11 - Spearman's correlations between the different classes.

Environmental controls on phytoplankton community's distribution

Environmental data presentation

Physical and chemical environmental contextualized data for Kongsfjorden and Rijpfjorden stations is given respectively in table 7 and table 8. Starting with the physical conditions, regarding temperature, comparing both transects the Kongsfjorden (western transect) registered higher temperatures than Rijpfjorden (northern transect) and the same can be observed for the salinity parameter, with the exception of KB6_2 (Fjord) which is highly influenced by glacier water inflow presenting the lowest value for salinity.

In the western transect, the station located in Open ocean (HGIV) registered the lowest values for oxygen and at the northern transect the lowest values of oxygen was registered at the Coastal station (R4). In general, the PAR and SPAR values were higher in Kongsfjorden and regarding the fluorescence parameter it's in the Rijpfjorden that highest value were registered, but the rest of the values were found to be lower than in Kongsfjorden.

In what inorganic nutrients is respect, higher availability of Ammonium (NH_4^+) was registered in Rijpfjorden transect, R4_124 and in the Fjord/glacier, G1_5. As for nitrite (NO2−) stations from the western transect, showed higher concentrations at KB3_332 and V6_200, and for the northern transect higher nitrate levels were registered at R1_168. Concerning nitrate (NO3-), in both transects these levels increased with depth, with the exception of G1_5, which was higher at the surface in this glacier station. For, Phosphate (PO_4^{3-}) it was also registered an increased with depth, with the exception of G1_5, in both transects. Silica hydroxide Si(OH)4 increases also with depth in the northern transect and as for the western transect the same trend was observed, with the exception of KB6_2.

Comparing both transects, Chlorophyll a shows generally higher values at surface and medium depths (DCM). The same trend was observed for Rijpfjorden, with the exception of station R7_25, where Chlorophyll a increased as the depth increased. Also, offshore stations presented lower Chlorophyll a concentrations.

Table 7: Physical and Chemical parameters from Kongsfjorden stations

	Sample ID	G1_5	G1_30	G1_65	KB6_2	KB6_25	KB6_50	KB3_5	KB3_50	KB3_332	KB0_5	KB0_25	KB0_50	KB0_320	V12_5	V12_25	V12_1000	V6_5	V6_25	V6_200	HGIV_5	HGIV_43	HGIV_50
Physical	Temperature (°C)	ND	ND	ND	5.27	3.56	3.91	3.98	4.54	1.98	5.74	4.83	4.83	1.86	7.94	5.64	2.9	7.94	5.64	2.92	7.18	5.53	5.32
	Salinity	ND	ND	ND	27.6	33.5	34.7	31.5	34.7	34.9	34.4	34.7	34.7	34.9	34.6	35.01	34.9	34.6	35.01	34.9	34.7	35.02	35.07
	Oxygen (mL.L^{-1})	ND	ND	ND	2.66	2.53	2.47	2.67	2.48	2.49	2.47	2.46	2.46	2.51	2.46	2.37	2.45	2.46	2.37	2.45	2.08	2.09	2.02
	PAR (µE.m^{-2}.s^{-1})	ND	ND	ND	1.710	1.00e-12	1.00e-12	88.54	0.0558	88.54	0.473	0.024	0.024	8.65	7.919	8.8609e-01	1.0000e-12	7.9194	8.8609e-01	1.0000e-12	29.15	1.668	0.851
	SPAR	ND	ND	ND	159.2	162.07	167.04	985	988.2	185.6	53.253	54.076	54.076	59.425	30.386	2.8948e+01	3.1261e+01	30.386	2.8948e+01	3.1261e+01	695.8	697.06	701.43
	Fluorescence (mg.m^{-3})	ND	ND	ND	1.299	0.1043	0.0576	7.1547	0.2362	-0.0963	-0.0800	1.046	1.046	0.455	2.46061	2.37377	2.45731	2.46061	2.37377	2.45731	0.714	2.984	1.468
Biogeochemical	Ammonium (µM)	4.54	2.9	2.75	1.53	2.76	2.20	1.48	1.74	0.75	0.82	1.06	1.06	0.28	0.13	0.09	1.01	0.13	0.09	1.01	0.15	0.12	0.12
	Nitrite (µM)	0.14	0.11	0.1	0.06	0.15	0.15	0.07	0.07	0.55	0.06	0.11	0.11	0.45	0.06	0.06	0.51	0.06	0.06	0.51	0.06	0.09	0.09
	Nitrate (µM)	1.4	0.9	0.9	0.5	1.5	1.4	0.7	1	10.6	1.3	1.5	1.5	10.7	0.5	0.8	9.1	0.5	0.8	9.1	0.5	11.9	11.9
	Phosphate (µM)	0.17	0.14	0.12	0.06	0.18	0.34	0.13	0.24	0.92	0.15	0.19	0.19	0.87	0.08	0.17	0.71	0.08	0.17	0.71	0.09	0.84	0.84
	Silicate (µM)	1.50	1.40	1.7	2.8	1.5	1.5	1.7	1.6	8.1	1.4	1.5	1.5	6.3	0.70	0.70	3.50	0.70	0.70	3.50	1.1	4.5	4.5
	Chl *a* (mg.m^{-3})	ND	ND	ND	0.82	0.11	0.06	0.82	0.17	0.10	0.77	0.52	0.52	1.34	0.34	0.97	0.44	0.34	0.97	0.44	0.41	0.28	0.28
	Phaeopigment (mg.m^{-3})	ND	ND	ND	0.72	0.12	0.10	0.74	0.28	0.23	0.63	0.43	0.43	1.27	0.27	0.61	0.51	0.27	0.61	0.51	0.30	0.33	0.33

ND - Not Determined.

Table 8: Physical and Chemical parameters from Rijpfjorden stations

	Sample ID	R1_5	R1_17	R1_168	R4_5	R4_28	R4_124	R6_5	R6_50	R7_5	R7_25	R7_1035
Physical	Temperature (°C)	5.18	4.66	-1.61	4.14	3.68	1.49	4.14	2.39	1.19	-0.03	-0.55
	Salinity	32.9	34.08	34.9	34.05	34.7	34.8	32.02	34.8	31.6	34.2	34.9
	Oxygen (mL.L^{-1})	2.37	2.38	2.49	2.31	2.24	2.24	2.49	2.40	2.49	2.40	2.49
	PAR (µE.m^{-2}.s^{-1})	11.18	0.924	1.00e-12	52.04	6.995	1.00e-12	60.47	1.9	26.29	5.130	1.00e-12
	SPAR	65.99	66.53	85.68	181.9	203.1	198.9	205.25	189.35	89.30	89.94	90.02
	Fluorescence (mg.m^{-3})	1.001	3.169	0.031	0.219	1.372	-0.046	0.047	0.262	0.065	0.847	-0.117
Biogeochemical	Ammonium (µM)	0.03	0.02	0.87	0.40	1.06	4.27	0.06	1.27	0.24	0.51	0.13
	Nitrite (µM)	0.06	0.06	0.36	0.06	0.06	0.12	0.06	0.1	0.06	0.06	0.06
	Nitrate (µM)	0.5	0.5	9.9	0.5	1	5.9	0.5	5.1	0.5	1.7	14.4
	Phosphate (µM)	0.09	0.11	0.88	0.06	0.15	0.59	0.06	0.56	0.11	0.26	1.10
	Silicate (µM)	1.6	1.6	5.1	0.8	2	3.4	1.2	2.2	1.5	1.8	10.3
	Chl *a* (mg.m^{-3})	1.30	1.44	0.07	0.18	0.07	0.04	0.22	0.08	0.47	0.59	0.12
	Phaeopigment (mg.m^{-3})	0.77	0.74	0.18	0.12	0.07	0.09	0.14	0.08	0.26	0.52	0.13

From the PCA (Principal component analysis) plot (Figure 12), we can verify that the stations of both transects are all very close to each other, especially the stations with bottom depths, except R4_124 and KB0_320. Analyzing carefully, we can verify once again that stations R7_1035 and V6_1000 remained together and concerning the stations from DCM and surface depth they are all mixed together.

The stations from DCM and surface depth seems to be strongly influenced by environmental parameters, such as NH4, SPAR, PAR, Temperature, Fluorescence, Chlorophyll a, and Phaeopigment. The stations at bottom depth seem influenced by Salinity, PO4, Si(OH)4 and NO2, except R4_124 and KB0_320.

From the PCA correlations (Table 9), PC1 (Comp. 1) explains approximately 33% of the variation between the samples distributions, and PC2 (Comp. 2) explains about 17%. In light of the above, the two axes combined explain exactly 50% of the variance. PC1 seems to be more correlated with the environmental variables rather than PC2.

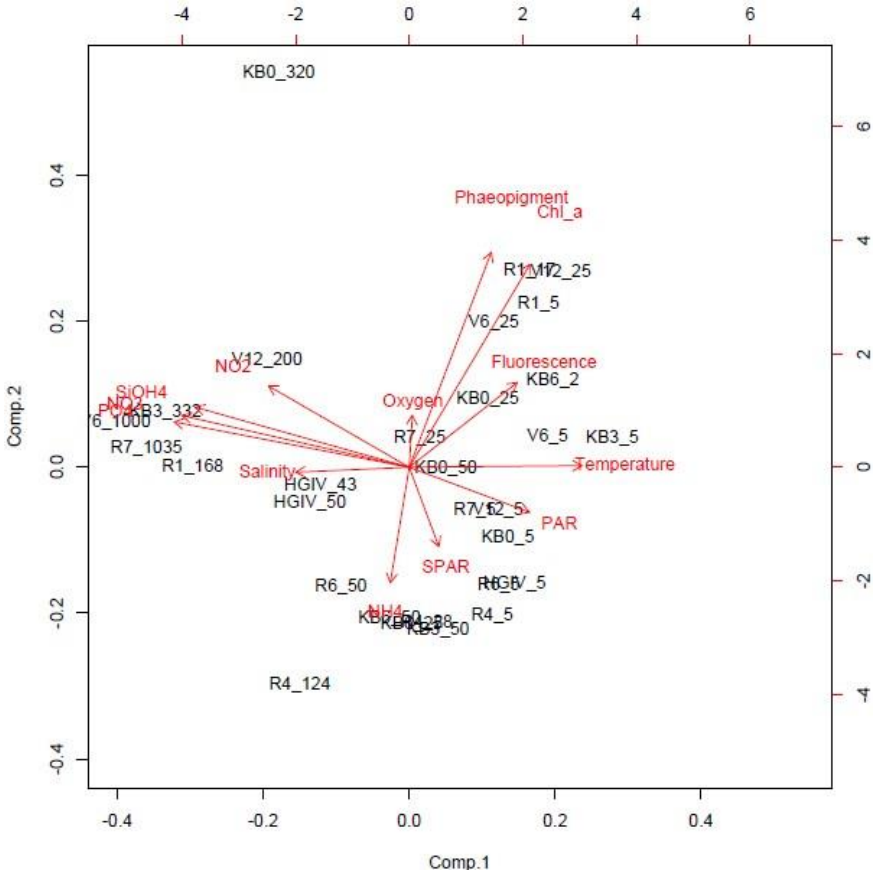

Figure 12 - PCA (Principal Component Analysis) of environmental data of protists collection of all samples.

Table 9: PCA correlations of both transects

Environmental Data	Comp. 1	Comp. 2
Temperature	0.336	0
Salinity	-0.221	0
Chl_a	0.235	0.555
Fluorescence	0.210	0.232
Oxygen	0	0.142
PAR	0.234	-0.123
SPAR	0	-0.216
NH4	0	-0.248
NO2	-0.274	0.222
NO3	-0.442	0.141
PO4	-0.458	0.124
Si(OH)4	-0.417	0.164
Phaeopigment	0.160	0.587

Also, principal component analysis was performed for each transect to identify the environmental parameters that better explained the variation between stations. Results revealed that bottom samples from Kongsfjorden transect, were more influenced by NO3, PO4, Si(OH)4 and NO2, while samples collected at DCM and surface depths were more influenced by the remaining environmental parameters (Figure 13a). In this analysis PC1 (Comp.1) explains approximately 33% of the variations between samples and PC2 (Comp. 2) explains about 17% (Table 10).

Regarding the PCA plot for Rijpfjorden stations (Figure 13b), once again the bottom samples were more influenced by PO4, NO3, Si(OH)4, with exception of the sample R4_124 where the DCM and surface depths were more influenced by the remaining variables. In the PCA analysis for Rijpfjorden transect, PC1 (Comp. 1) explains 44% and PC2 (Comp.2) explains about 23% of the variation between samples (Table 11).

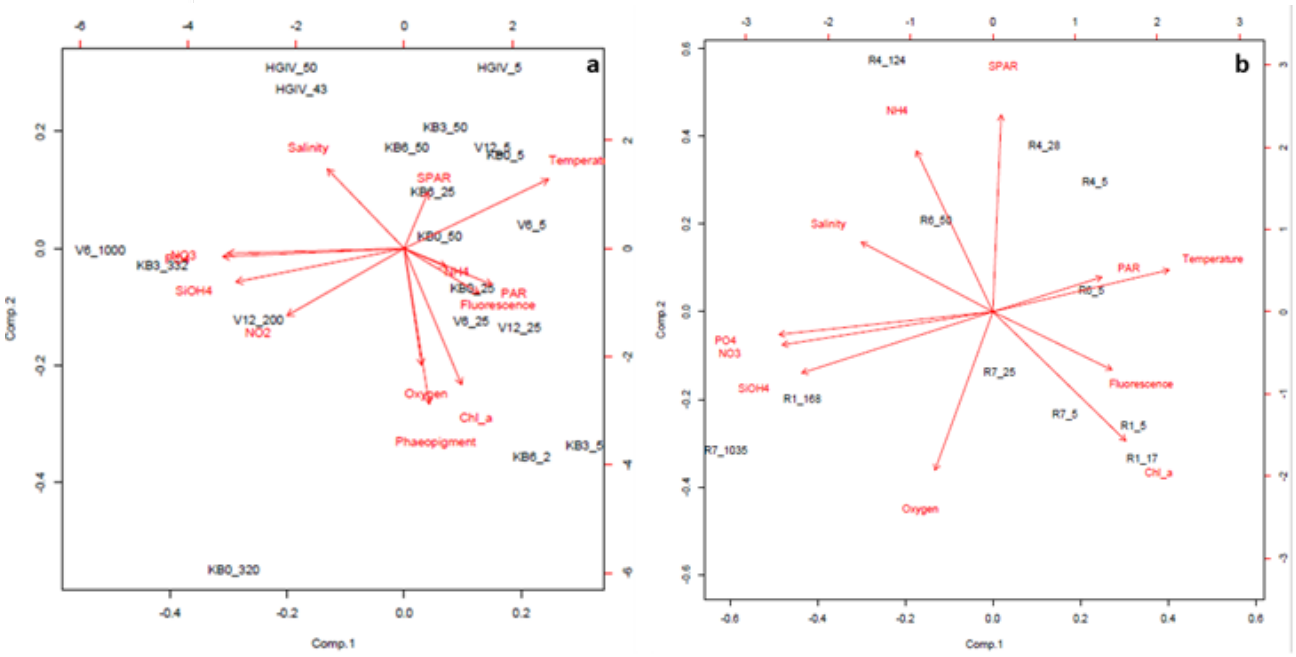

Figure 13 - a) PCA (Principal Component Analysis) of environmental data for Kongsfjorden transect. **b)** PCA (Principal Component Analysis) of environmental data from Rijpfjorden transect.

Table 10: PCA correlations analysis of Kongsfjorden

Environmental Data	Comp. 1	Comp. 2
Temperature	0.369	0.245
Salinity	-0.197	0.281
Chl_a	0.147	-0.481
Fluorescence	0.192	-0.158
Oxygen	0	-0.413
PAR	0.225	-0.128
SPAR	0	-0.686
NH4	0.108	0
NO2	-0.300	-0.237
NO3	-0.452	0
PO4	-0.464	0.109
Si(OH)4	-0.430	-0.118
Phaeopigment	0	-0.550

Table 11: PCA correlations analysis of Rijpfjorden

Environmental Data	Comp. 1	Comp. 2
Temperature	0.369	0.119
Salinity	-0.276	0.198
Chl_a	0.278	-0.368
Fluorescence	0.249	-0.166
Oxygen	-0.121	-0.452
PAR	0.228	0
SPAR	0	0.561
NH4	-0.160	0.459
NO3	-0.441	0
PO4	-0.448	0
Si(OH)4	-0.400	-0.175

Relationships between environmental data and protists distribution

Regarding the NMDS analysis, surface and DCM samples were grouped together and seems to be influenced by Salinity, Temperature, NH4, NO2 and Oxygen and the bottom samples seems to be more influenced by PO4, NO3 and Si(OH)4 (Figure 14).

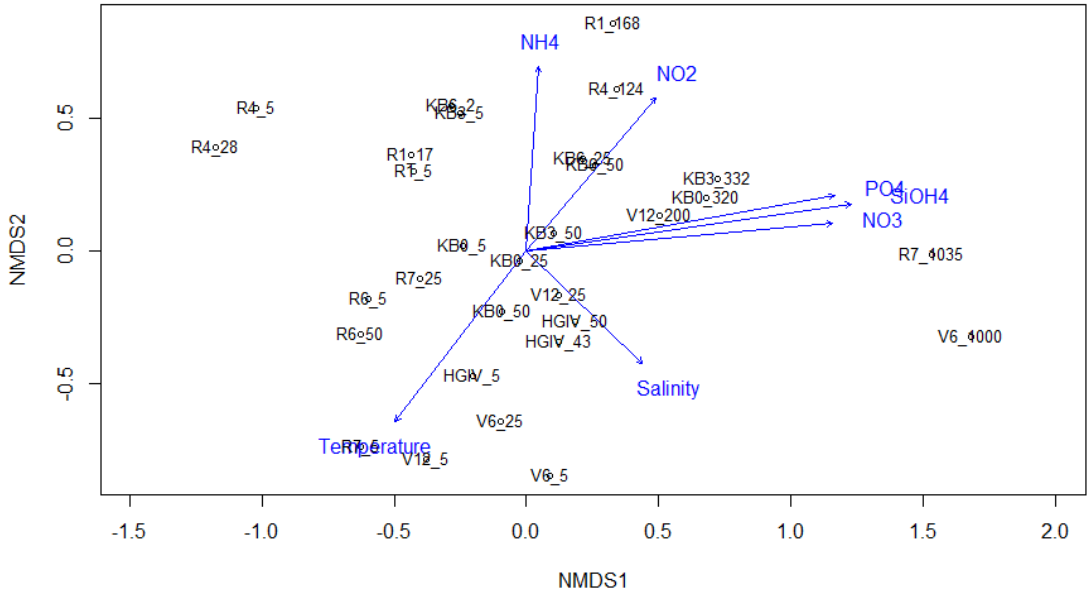

Figure 14 - NMDS (nonmetric multidimensional scaling) of protists collection for all samples.

The NMDS analysis performed only for Kongsfjorden stations (Figure 15a) showed the same pattern registered for all samples together (Figure 14) with the surface and DCM samples being influenced by Temperature, Salinity, NH4 and Oxygen and the bottom samples influenced by NO3, PO4, Si(OH)4. In parallel, NMDS analysis performed for Rijpfjorden transect (Figure 15b) samples showed that bottom and DCM samples seems to be influenced by NO3, PO4 and Si(OH)4, and regarding some of the surface and DCM samples were influenced by Oxygen and Temperature.

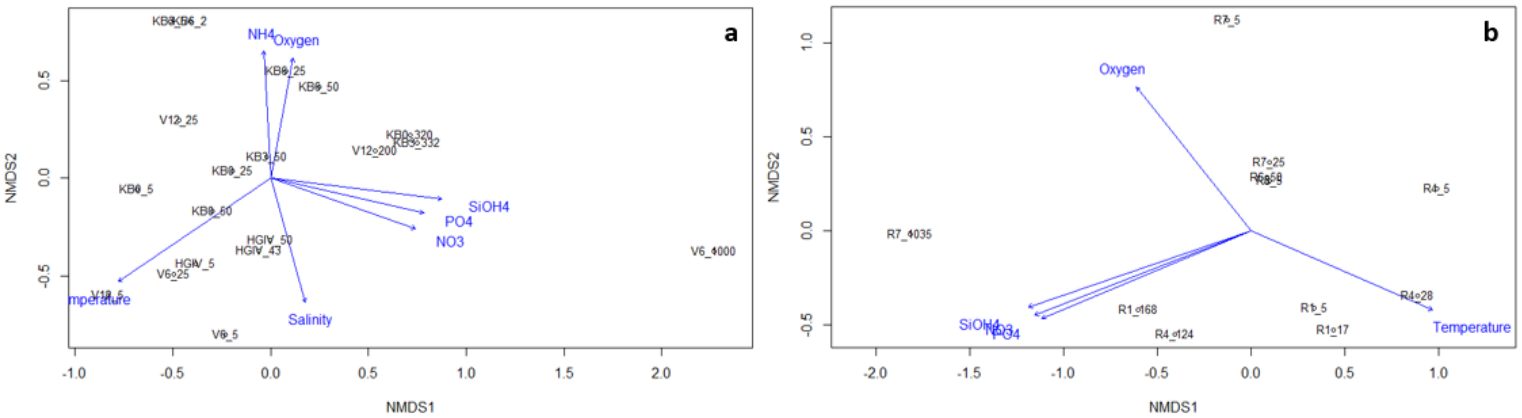

Figure 15 - a) NMDS (nonmetric multidimensional scaling) of protists collection for Kongsfjorden transect; **b)** NMDS (nonmetric multidimensional scaling) of protists collection for Rijpfjorden transect.

In order to understand the protists distribution association with the environmental gradients, Spearman correlations between the different protists groups at class level and the environmental variables were made (Figure 16). Results demonstrated that in general the different protists groups were significantly and positively correlated within the availability of nutrients and with the water column oxygen concentrations. As for PAR, we can verify that was negatively correlated with almost all protists classes. The same was true for Fluorescence, Temperature, Chl_a and Phaeopigment (Figure 16), where no significant trends were observed or significant negative correlations were registered, according with the protist group.

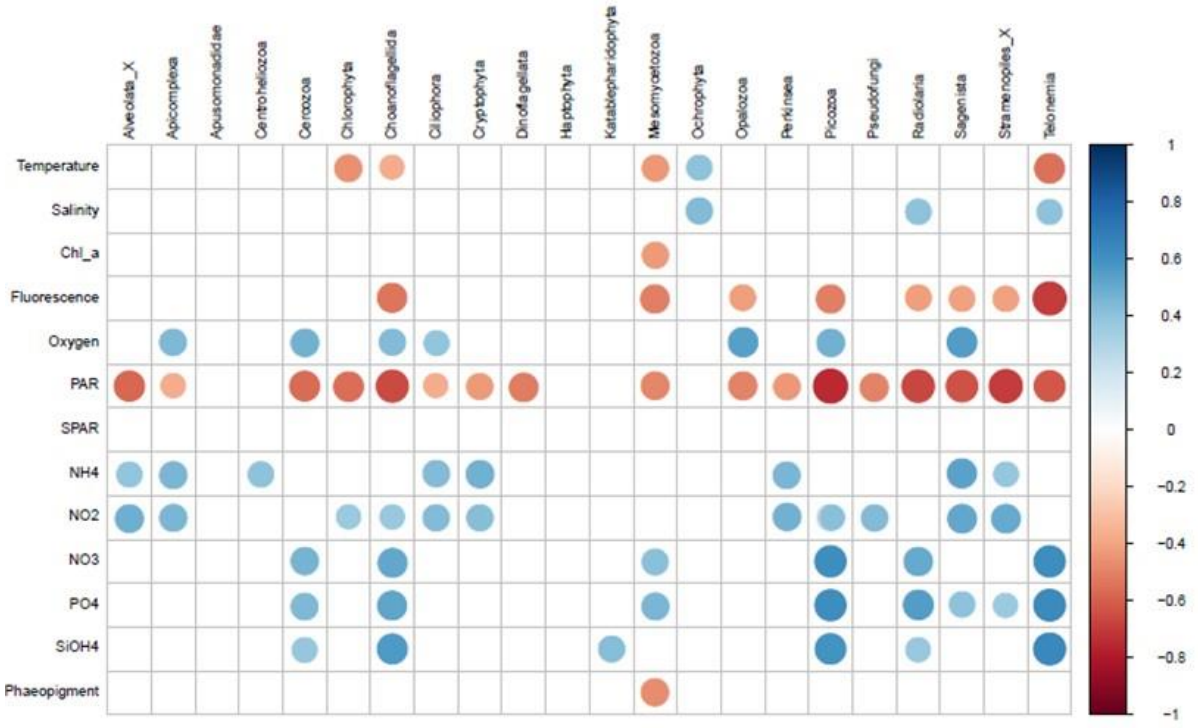

Figure 16 - Spearman's correlations between the different classes and environmental parameters.

Results of environmental and protists correlations also revealed that there were some protist classes that were strongly positively influenced by nutrients and oxygen and negatively correlated by the remaining parameters, although another group of protists classes seems to be apparently less influenced by these variables (Figure 17).

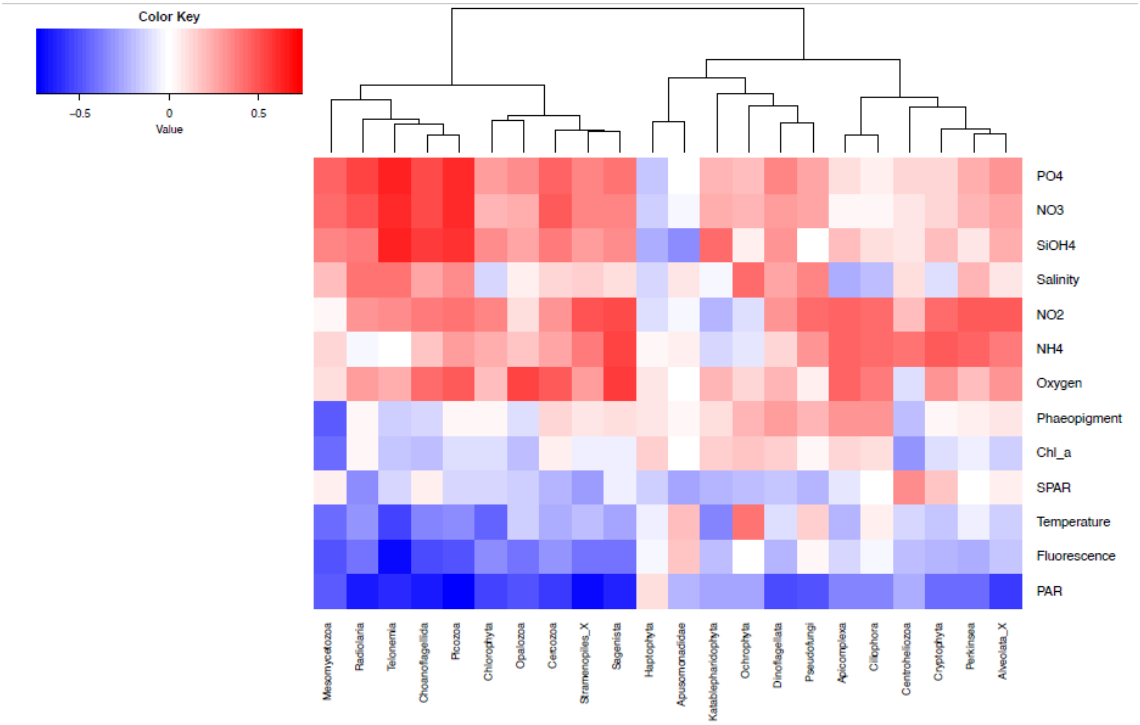

Figure 17 - Heatmap correlations between class level and environmental parameters.

Discussion – Part A

Protists Diversity and Composition across Arctic Marginal Ice Zone using Metabarcoding Approach

Protists diversity across Arctic environments at the Marginal Ice Zone (MIZ) revealed the existence of about 5876 different OTUs at 99% similarity, in a total 33 samples collected across different water column depths.

OTU-based approach is used to differentiate species or taxa based on similarity to a specific sequence and then grouping them under similarity cutoffs. In literature, different authors highlighted distinct thresholds for OTU clustering, using commonly the cutoff range from 95 to 99% [76,103,104,105]. Previous studies mentioned that the OTUs threshold at 95% of similarity, was often deliberately chosen to produce a more conservative estimator of species richness for eukaryotes [57,103,106]. Traditionally, a 97% cut-off threshold has been used in metabarcoding studies and commonly applied for OTU clustering in protists [76,105,107,108]. Higher thresholds, as 98% and 99%, are more stringent and more capable of recovering rare taxa detection by providing a better resolution of species-level assigning sequences [104,107,109]. However, it should also be considered that cut-off ranges depend on the variability of the chosen marker tag between taxa, and when it possible they need to be linked with ecological and physiological traits of species for proper identification [110].

The protist-focused metabarcoding approach used in this study targeting the V4 region of the 18S rRNA gene, using PR^2 reference database, revealed that the Arctic eukaryotic dataset comprises highly complex and diverse protists community structure. Numerous authors emphasize in their studies that the V4-18S and V9-18S region are the most suited regions in metabarcoding surveys for assessing the diversity of microbial eukaryotes [61,76,111,112]. Notwithstanding, some studies pointed out that the V4 region contributed to a greater knowledge of metabarcoding advances for biogeography and ecology studies with aim to uncover protistan diversity and decode their process and patterns of environmental eukaryotic diversity [113,114]. Particularly for studies that aim to access ciliates, diatoms and dinoflagellates the V4 region is a promising candidate, especially at higher level [61,77,115,116].

Moreover, other studies that compare the V4 and V9 region, although reporting similar patterns for the two marker tags, highlighted that the V4 region, that is the longest variable region in the rRNA gene (350 to 450 nt), provides an higher taxonomic resolution often to species or genus-level and it allows also an accurate taxonomic placement for unassigned HTS-amplicon sequences [57,112,117].

The unique and shared protists OTUs between Kongsfjorden and Rijpfjorden transects were detected, through venn diagrams, and results showed that samples from

Kongsfjorden transect presented higher number of OTUs compared with Rijpfjorden and bottom samples were the ones that had higher shared OTUs. When the same venn analysis was performed for the different oceanic features (Fjord, Coastal and Open Ocean), it was detected a decrease of unique OTUs from fjord to open ocean when all samples were analysed together and in Kongsfjorden samples. Although in Rijpfjorden open ocean showed higher unique OTUs. So, overall, when analysing both transects, fjord samples were found to include more unique protists OTUs, which was also evident when only Kongsfjorden samples were included in the venn diagram. In addition, higher number of unique protists OTUs was recorded in Kongsfjorden coastal waters than in Rijpfjorden coastal waters which could reflect that the Kongsfjorden transect is much more dynamic in temporal terms, because the fjord glaciers are regressing at a very accelerated rate and are also influenced by the infiltration of warmer and denser North Atlantic waters [36]. Also, the influence of distinct water masses in Kongsfjorden and Rijpfjorden region as well the influence of different melting glacier rates in both fjords could explain the distribution of unique OTUs adapted to the different transect studied and/or oceanic features.

In fact, three distinct water masses, regarding temperature and salinity properties, can be differentiated in our investigated area [80]: a surface layer of less saline and colder polar surface water (PSW), an intermediate layer of modified Atlantic water (MAW), separated from the fresher PSW by a sharp halocline, and lastly, a deeper layer of Atlantic Water (AW). Thus, the extreme stressors caused by global warming which is more evident in the Kongsfjorden transect [45] could support our results in terms of different unique OTUs observed between the two transects regions studied. Also, in Kongsfjorden coastal and open ocean waters the number of shared OTUs was higher than coastal and fjord waters, leading to support the idea that open ocean and coastal stations shared more similar water masses. A recent study performed by Zheng et al. [118] mentioned that the AW (Atlantic water) brought many species into the intermediate waters where they had the largest species numbers and in TAW (transformed Atlantic Water), where they found greatest biodiversity [118]. The situation in the Rijpfjorden transect was different, the highest number of unique protists OTUs was detected in open ocean waters and coastal and fjord stations shared a highest number of protists OTUs. This can be explained by the high stability of the glaciers in Rijpfjorden, and a more uniform influence of polar water masses [52,119,120].

Overall, in light of the above, these results suggest that in Kongsfjorden transect the protists community are more influenced by different water masses (including North Atlantic waters) and also by the strong influence of glacier materials and freshwater input

in the fjord, which appears to result in overall high numbers of OTUs. In Rijpfjorden transect, the protists community seems to be more connected with Kongsfjorden transect in open ocean stations, mainly in the deeper station where the connectivity is evident due to the higher number of shared OTUs, suggesting higher stability in terms of environmental characteristics between these two stations. This can be explained by the fact that both open ocean stations from Rijpfjorden and Kongsfjorden are influenced by the same water mass (North Atlantic water mass), characterised by warmer and salty waters, with associated more similar protists communities.

Regarding the alpha diversity results, the observed OTUs, Chao1 and Shannon indexes increased with depth, with more richness observed at DCM and bottom depths. Beta diversity analysis revealed that samples were grouped according with depth and oceanic features, with only a few overlaps between coastal and fjord samples. While there is a lack of research relating the protists community distribution across water column depth [80], there are some studies demonstrating a general trend towards higher diversity of microeukaryotes in deeper water column samples [80,121,122,123]. Given the above, our results show strong evidence of a diverse community structure, with a marked biogeographic pattern of the protists communities along the Svalbard MIZ with a clear trend of depth-dependency along the water column regarding both transects.

The most dominant phylum, recovered from V4-18S metabarcoding, was *Ciliophora*, followed by *Dinophyceae*. These phyla were present in all samples from Rijpfjorden and Kongsfjorden transects. The dominance of these phyla, belonging to the superphyla *Alveolata* are in agreement with results from previous studies focused on protists diversity, that indicated a great dominance of *Alveolata*-related sequences identified as *Ciliophora* and *Dinophyceae* in the Arctic Ocean [51,124,124]. In fact, a report of the state of Arctic marine diversity (source of MOSJ) mentioned that these particular taxa often show a consistent pattern of dominance in both Kongsfjorden and Rijpfjorden transect [33]. This result is also supported by different studies performed around Svalbard waters that reported *Ciliophora* and *Dinophyceae* as dominant groups [33,53,125].

Regarding class level, there was a prevalence of *Chlorophyta, Ciliophora* and *Dinoflagellata* in the two Svalbard fjords studied. Some taxa tend to be more concentrated in Kongsfjorden transect than in Rijpfjorden, like *Haptophyta*. This suggests that there is a tendency of a dominant distribution of some taxa in Kongsfjorden transect rather than Rijpfjorden. A possible explanation for this different trend, could be that the influx of North Atlantic waters is more intense in Kongsfjorden region [33]. Likewise, the protists community from Kongsfjorden are subject to less environmental

stability because of the seasonal hydrographic variations triggering protists communities to react to these climate-driven changes [125].

As for family and genus level, *Dinophyceae* and unclassified *Dinophyceae_XXX* remained the dominant groups in the eukaryotic libraries of MOSJ-ICE2016 collection. This taxonomic profile is quite similar to Sousa et. al [80] which also detected that phylotypes associated with *Dinophyceae* dominated most of their 18S rRNA libraries [80]. Also, Mordret et. al [77] referred that the taxonomic annotation of dinoflagellates at family level is problematic because there are still unresolved phylogenetic relationships within genera. Deeper clarification about this issue can be seen in this previous study which provides a taxonomically curated 18S rRNA reference database of dinoflagellates [77].

Interestingly, the significant correlations between the identified taxonomic groups at phylum and class level showed that all the significant relationships identified were positive, suggesting that those protists groups are dependent on each other and/or respond in the same way to the environmental gradients of the studied region. These results also highlight an absence of taxa exclusion due for example to competition. Consequently, results together suggest that protists from MIZ have somehow the same distribution dynamics where biological and physical/environmental gradients may role its distribution in the same way [126].

Environmental parameters analysis between both transects, showed that Kongsfjorden transect stations are characterized by higher values of temperatures and salinity than Rijpfjorden, supporting that the western transect is more influenced by the "*atlantification*" process, rather than the northmost transect, which marks a more Arctic habitat environment, with cold Polar waters, less dense and salty [52,120]. From the PCA and NMDS analysis, the environmental gradients and the nutrients became relevant in this context. A clear separation of bottom samples with DCM and surface depths was noticed, where the biogeochemical gradients (nutrients) seems to influence more the protists distribution in deeper samples and the physical gradients (e.g. temperature, salinity) seems to influence the distribution of protists communities at the surface and DCM samples. As for correlations between taxa and environmental parameters is respect, our results revealed clear positive and significant correlations with nutrients and negative correlations with PAR. This suggests that nutrients availability somehow seems to stimulate the occurrences of several taxonomic groups and PAR can decrease the occurrences of many protists groups that have been identified. This is in agreement with previous studies that point out that during the melting season in summer, the glacial runoff and the entrance of freshwater from the melting of Greenland ice sheet provokes

a large incoming of nutrients that could influence and stimulate the primary production [36,126]. For PAR, our results may be explained by the fact that most primary producers are affected by excessive light intensity, requiring optimal light intensity conditions for photosynthesis to happen [127].

Another important finding is that the deeper stations from Kongsfjorden and Rijpfjorden (V6_1000 and R7_1035), seems to be connected in terms of the taxonomic community profile, where some taxa tend to be more concentrated in those samples. Indeed, our analysis revealed that these samples are characterized by a greater diversity of protists, shared a high number of protists groups, and showed separate trends in terms of protists community composition when compared with DCM and surface stations. As stated above the protists communities similarities between stations V6_1000 and R7_1035 is most likely related with the influence of North Atlantic Water masses in both regions, which in the Rijpfjorden is only evident at R7 bottom station. Higher nutrients availability at bottom depths and more environmental stability in the deep Arctic ocean may also explain the increase of protists diversity with depth [128,129,130].

Results - Part B

Comparison between 18S Metabarcoding and Microscopic cell counts and identification

Methodological comparison between microscopic PHT and 18S metabarcoding

A total of 17 samples were included for methodological comparison between 18S metabarcoding and microscopy protists identification. Ten of which belong to Kongsfjorden transect and seven belong to Rijpfjorden. The depth range of these samples was from 2m to 50m, mostly surface and DCM samples (Figure 18).

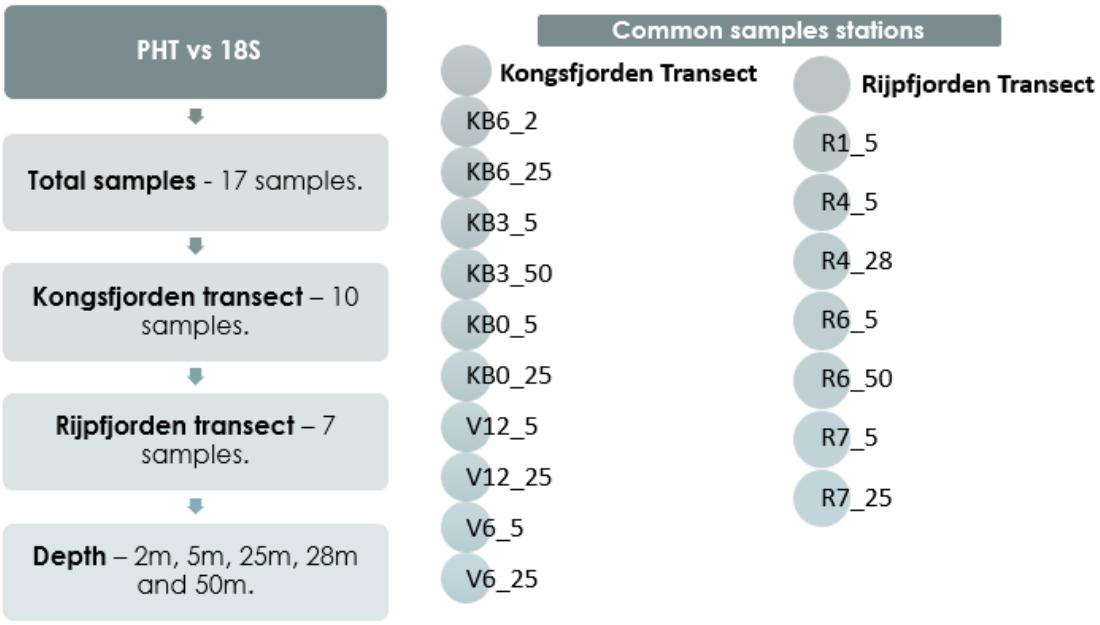

Figure 18 - Data information of the common samples stations and depths used for methodological comparison between 18S metabarcoding and microscopic PHT for community phytoplankton analysis.

Groups of phytoplankton identified in both methodologies

Within the microscopy dataset from the 17 samples, a total of 12 taxa were identified at phylum and class levels, 60 taxa identified at genus level and 101 taxa at species level. As for 18S metabarcoding, a total of 91 taxa were detected at higher level (phylum, class and family levels), 273 at genus level and 342 at species level (Figure 19). All the identification of phytoplankton groups in both methodologies, as well the percentage of abundance recovery is provided in Additional file 3.

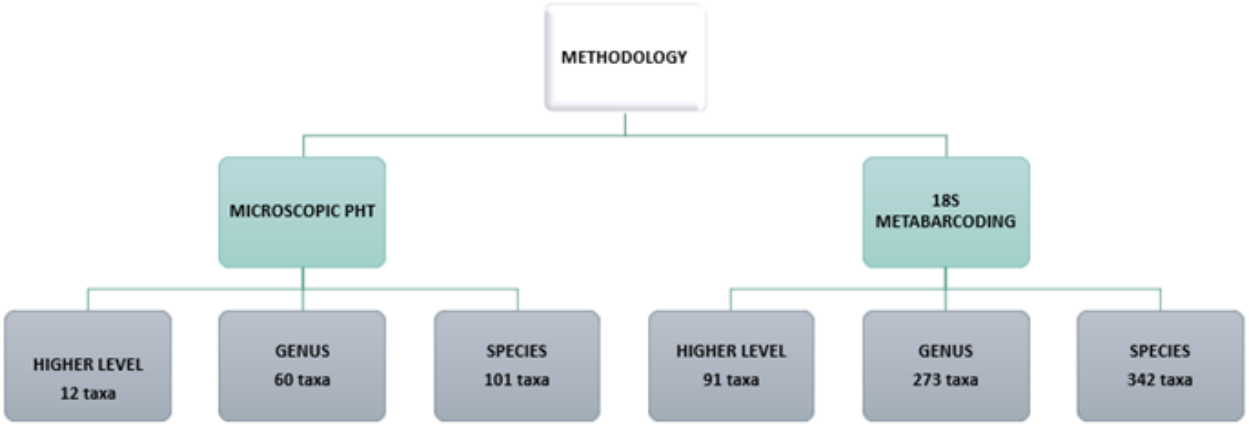

Figure 19 - Diagram with the total number of detected phytoplankton groups in the different methodologies at higher and lower taxonomic levels (18S vs PHT).

Regarding the microscopy methodology at higher level, of these 12 detected taxa, 3 were assigned as *uncertain taxa* or *incertae sedis*, respectively*, Dinoflagellata incertae sedis; Incertae taxa; Eukaryota incertae sedis*. The most abundant taxa according to the total relative abundance in all samples, belongs to: *Eukaryota incertae sedis (25,53%), Dinophyceae (24,20%), Prymnesiophyceae (19,78%), Cryptophyta (8,47%)* and *Pyramimonadophyceae (7,03%)*. In 18S metabarcoding methodology in the same 17 samples, 11 phyla were identified, being the most abundant: *Ciliophora (61,21%); Dinophyceae (22,77%)* and *Other_Alveolata (14,80%)* (Additional file 3).

Bearing this in mind, the taxonomic distribution was made for the two transects and for all samples for both methods, however only the distribution at higher level will be presented in this section, that is, in microscopy method the distribution of class/phylum and in metabarcoding only the phylum level (Figure 20). The distribution of the protists groups within the remaining taxonomic levels at higher level between 18S and microscopic PHT is provided in Additional file 4.

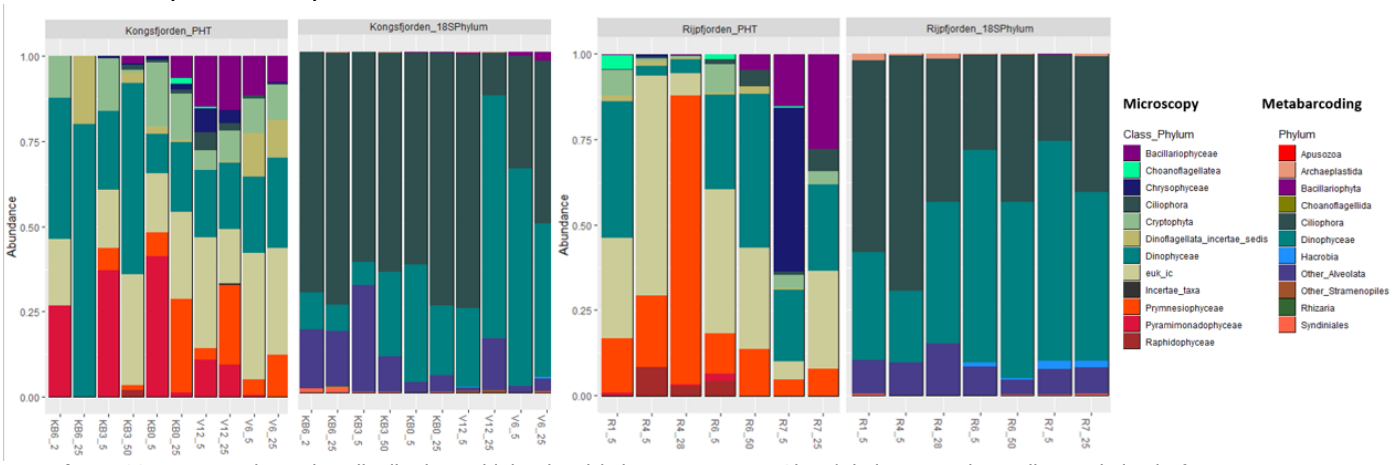

Figure 20 - Taxonomic protists distribution at higher level (Microscopy PHT - Class/Phylum; Metabarcoding – Phylum) of both methods divided by transects. Kongsfjorden on the left and Rijpfjorden on the right.

Continuing with the metabarcoding approach, at class level, 20 taxa were recovered, in which the classes identified with the highest relative abundances were*: Ciliophora (61,21%); Dinoflagellata (23,53%); Apicomplexa (10,05%)* and *Alveolata_X (4,75%).* And at family level, 60 taxa were detected in the 17 common samples and the highest taxa were: *Dinophyceae (22,77%); Cylotrichium_like_organism (18,05%); Nassophorea (12,86%); Spirotrichea (11,77%); Gregarinomorphea (10,05%); CONThreeP (8,71%); Oligophymenophorea (7,71%); Ellobiopsidae (4,75%)* and *Phyllopharngea (1,57%) (*Additional file 3*).*

In microscopy approach, the genera with highest relative abundances were: *Phaeocystis (15.45%); Gymnodinium (11.46%); Pyramimonas (7.03%); Heterocapsa (4.08%); Protodinium (3.50%); Dinobryon (2.85%), Teleaulax (2.85%); Gyrodinium (2.37%); Prorocentrum (2.13%), Rhizosolenia (2.02%)*. In 18S metabarcoding analysis a total of 273 taxa were detected and the genera with total relative abundances were: *Dinophyceae_XXX (25,30%); Cyclotrichium (16,53%); NASSO_1 (12,11%); Askenasia (6,79%); Thalassomyces (5,44%); Gregarines_GRE2_XX (5,33%); Chromidina (2,86%); Gregarines_GRE1_XX (2,51%); Philasterida_X (2,43%); Leegaardiella (2,29%); Strombidiidae_B_X (1,49%); PHYLL_4_X (1,33%); Urotricha (1,18%); Lankesteria (1,18%); Strombidium_K (1,09%)* and *Strombidiiae_H_X (1,07%) (*Additional file 3*).*

Finally, for species level, in microscopy approach a total of 101 taxa were detected, being 60 species identified and the remaining 41 were descripted with ‘’*sp*” or ‘’*indet”*. The more relevant species in terms of relative abundances were: *Phaeocystis pouchetii (15,31%); Heterocapsa rotundata (3,80%); Protodinium simplex (3,70%); Dinobryon balticum (2,85%); Prorocentrum cordatum (2,10%); Leucocryptos marina (1,97%); Gymnodinium galeatum (1,97%); Gymnodinium gracilentum (1,34%)* and *Rhizosolenia hebetata (1,19%).* For metabarcoding, a total of 342 taxa were recovered at taxonomic species level, being 137 species identified and the remaining 205 descripted with ‘’sp’’. From these 137 species, the ones that presented highest relative abundances were: *Thalassomyces_fagei (1,29%); Lankesteria_cytodytae (1,23%)* and *Strombidium_capitatum (1,16%)* (Additional file 3).

Overall, the microscopic approach identified *Dinophyceae* as the dominant higher level taxa and *Ciliophora* was identified as the dominant taxa by the 18S metabarcoding approach. At genus level, the most abundant group detected in microscopy was *Phaeocystis* and in metabarcoding was *Cyclotrichium.* Finally, for species level the most abundant in microscopy was *Phaeocystis pouchetii* while in metabarcoding, *Thalassomyces_fagei* dominates.

In light of the above, these results demonstrate that the identification and characterization of the phytoplankton groups was quite different when using these two different methodologies. Several groups were identified under metabarcoding that were not found in microscopy, as for the other way around. However, there were groups that matched between methodologies that will be presented in the following topic.

Groups of phytoplankton that were identified in microscopic PHT and 18S metabarcoding methodologies

Taking all samples into account and identified groups, between the 12 taxa found in higher level of microscopic analysis and of 91 taxa found in 18S metabarcoding, the number of matching taxonomic groups was 8. At lower level, between the 60 taxa identified at genus level by microscopic analysis and 273 genus level using 18S metabarcoding, the number of matching presences was 29. In what species is respect, the number of matching presences between both methodologies was 10 (Figure 21).

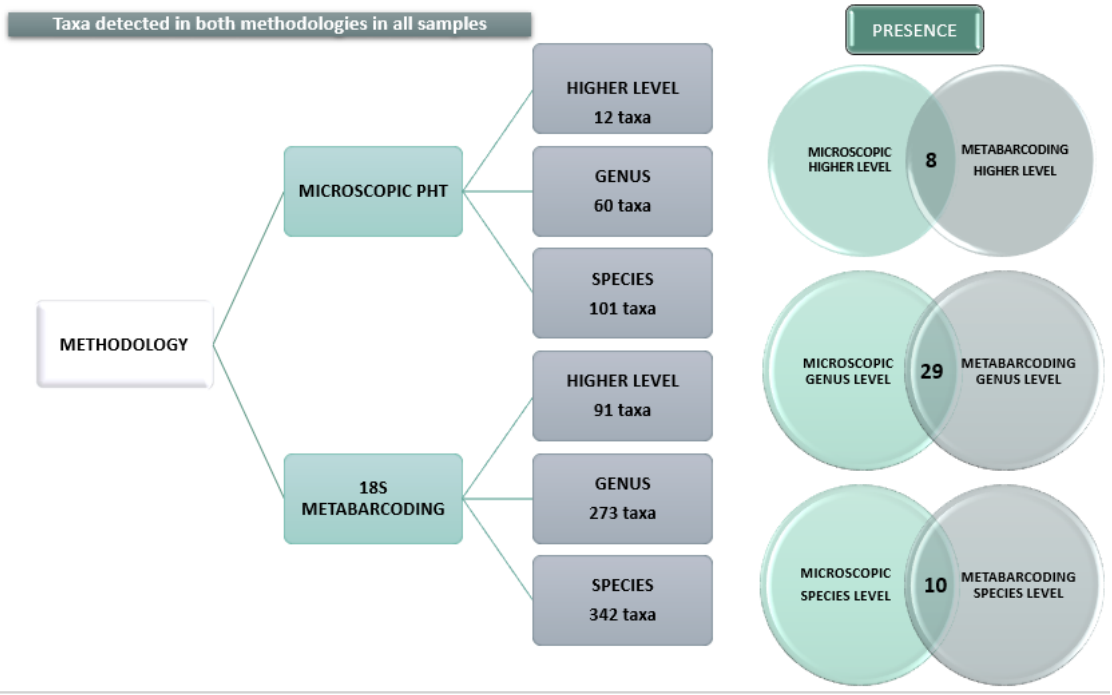

Figure 21 - Diagram showing the number of the detection of protists groups at different taxonomic levels and number of matching taxa that were present in both methods and in all samples (18S vs PHT).

Analyzing in detail the common highest taxonomic level groups registered with 18S metabarcoding and microscopic methodologies (Table 12) we can observe that *Bacillariophyceae* was detected in microscopic analysis, and also in 18S metabarcoding at phylum and family levels *(Bacillariophyta). Choanoflagellatea* and *Chrysophyceae* was detected in 18S metabarcoding at family level. *Ciliophora* was detected at phylum

and class level, and *Cryptophyta* only at class level in 18S metabarcoding method. *Dinophyceae* can be detected at phylum and family level, and finally, with regard to *Prymnesiophyceae* and *Pyramimonadophyceae* only were detected at family level *(Pyramimonadales).*

Table 12: Presence/Absence of higher level groups in both methods

Microscopic_PHT	18S Metabarcoding and Taxonomic levels		
Higher Level	**18S Metabarcoding_Phylum**	**18S Metabarcoding_Class**	**18S Metabarcoding_Family**
Bacillariophyceae	✓	×	✓
Choanoflagellatea	×	×	✓
Chrysophyceae	×	×	✓
Ciliophora	✓	✓	×
Cryptophyta	×	✓	×
Dinophyceae	✓	×	✓
Prymnesiophyceae	×	×	✓
Pyramimonadophyceae	×	×	✓

At lower level, as we can see from the results above, the number of matching presences between the two methodologies were 29 regarding genus level, and 10 regarding species level (Table 13).

Table 13: Taxa that were present in both methods at lower level

Genus		Species
Algirosphaera	**Navicula**	***Algirosphaera robusta***
Amphidoma	**Nitzschia**	***Chaetoceros decipiens***
Azadinium	**Pelagostrobilidium**	***Chaetoceros tenuissimus***
Chaetoceros	**Phaeocystis**	***Coccolithus pelagicus***
Chrysochromulina	**Prorocentrum**	***Cylindrotheca closterium***
Coccolithus	**Pseudo-nitzschia**	***Gyrodinium fusiforme***
Corethron	**Pyramimonas**	***Heterocapsa rotundata***
Cylindrotheca	**Rhizosolenia**	***Laboea strobila***
Dinophysis	**Rhodomonas**	***Phaeocystis pouchetii***
Eucampia	**Strombidium/Strombidium_K/Strombidium_M**	***Pseudo-nitzschia granii***
Gymnodinium	**Synedropsis**	
Gyrodinium	**Teleaulax**	
Heterocapsa	**Thalassiosira**	
Laboea	**Tintinnopsis/Tintinnopsis_05/Tintinnopsis_07**	
Leegaardiella		

Note: Strombidium* and Tintinnopsis* were the genera detected in both methods, however in metabarcoding was assigned as Strombidium_K; Strombidium_M; Tintinnopsis_05/07 from PR^2.

Relative abundance of phytoplankton groups that were detected in microscopic PHT and 18S metabarcoding methodologies in the same samples

Between the 12 phytoplankton taxa found at higher taxonomic level with microscopy analysis and 91 taxa found with 18S metabarcoding method, the number of matching groups within the same samples were only 7. In addition, between the 60 genus identified in microscopic analysis and the 273 genus identified with 18S metabarcoding, only 19 were found to recover in the same samples, as for species, the number of the same matching groups were only 5 (Table 14; Figure 22).

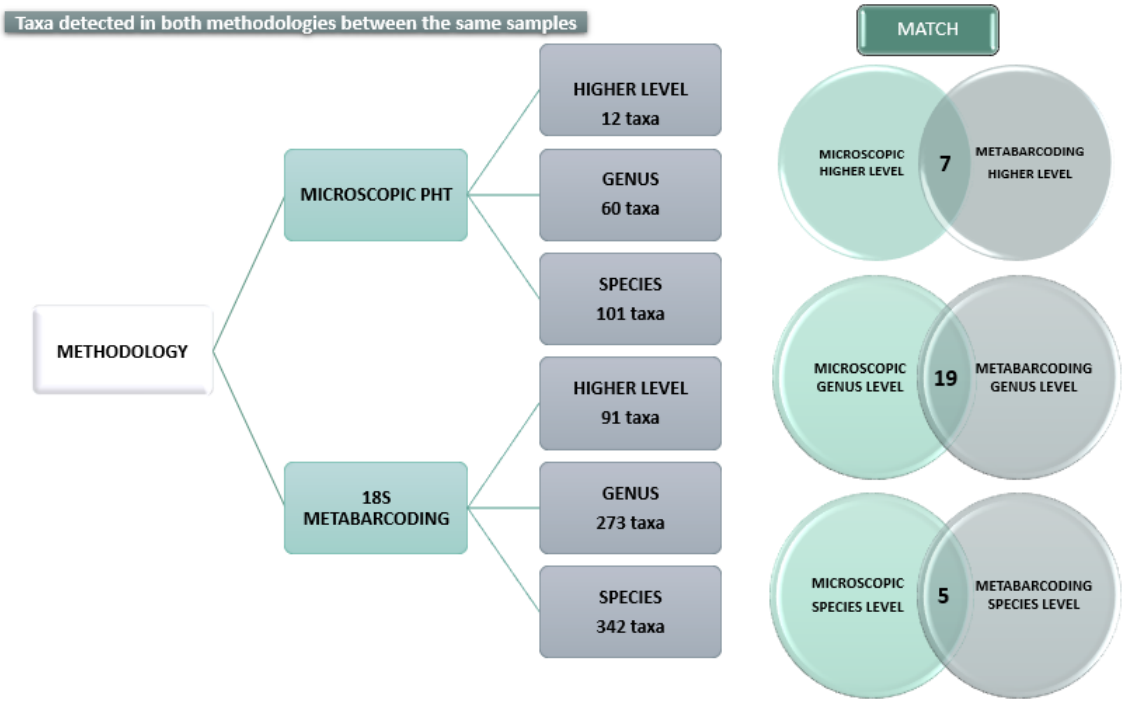

Figure 22 - Diagram showing the number of the detection of protists groups at different taxonomic levels and number of taxa that matched within the same samples (18S vs PHT).

Table 14: Common groups that matched within the same samples in both methods at higher taxonomic level

Microscopic_PHT		**Metabarcoding**	
Higher Level	**[a]Abundance**	**Higher Level**	**[a]Abundance**
Bacillariophyceae	5,24%	**[b]Bacillariophyta**	0,07%
Chrysophyceae	3,49%	**Chrysophyceae**	0,002%
Ciliophora	0,98%	**Ciliophora**	61,21%
Cryptophyta	8,47%	**Cryptophyta**	0,01%
Dinophyceae	24,20%	**Dinophyceae**	22,77%
Prymnesiophyceae	19,78%	**Prymnesiophyceae**	0,08%
Pyramimonadophyceae	7,03%	**[c]Pyramimonadales**	0,01%

[a]Total relative abundance of all samples; [bc]Different names due to the designation assigned according to the PR[2] database.

The matching taxonomic groups between both methodologies within the same samples were: *Bacillariophyceae/Bacillariophyta, Chrysophyceae, Ciliophora, Cryptophyta, Dinophyceae, Prymnesiophyceae* and *Pyramimonadophyceae/Pyramimonadales.*

However, total representativeness of each taxonomic group differs in terms of abundances, between metabarcoding and microscopic approaches. In metabarcoding the relative abundance of those groups was found to be extremely low, with the exception of *Ciliophora* and *Dinophyceae*, which cover greater abundance recovery by metabarcoding. In microscopy, a greater recovery of *Dinophyceae, Prymnesiophyceae* and *Cryptophyta* was found compared with metabarcoding. Generally, the groups with high relative abundances in microscopy are lower represented in metabarcoding, with the exception of *Ciliophora* that was almost negligible in microscopy. As for *Dinophyceae* the abundance recovery is similar in both methodologies.

Nevertheless, despite the importance of the total representativeness of each group, its representativeness by individual sample should be highlighted (Figure 23).

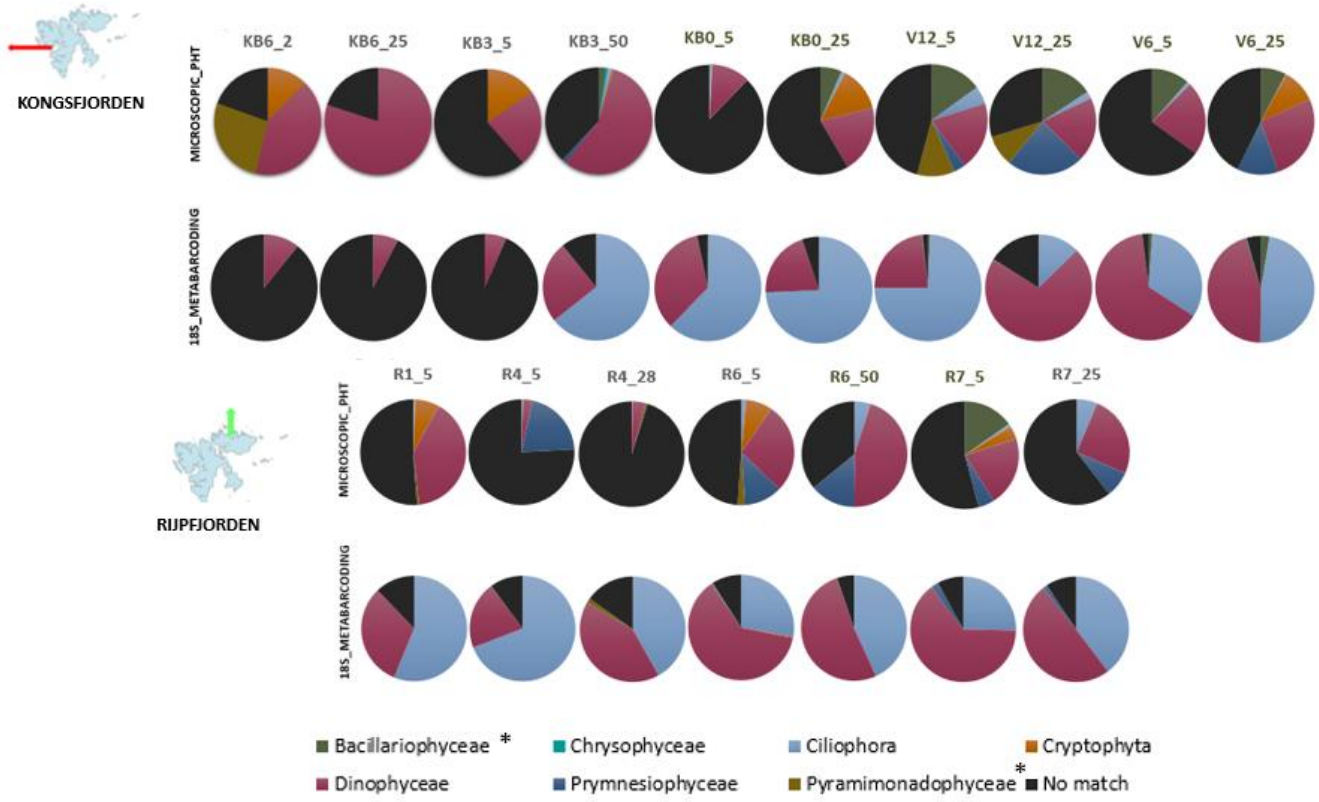

Figure 23 - Pie charts of matching groups (18S vs PHT) at higher level taxonomy of each methodology within the same samples and separated by transects; *Bacillariophyceae was the taxa detected in microscopy that matched with Bacillariophyta in metabarcoding; *Pyramimonadophyceae was the taxa detected in microscopy that matched with Pyramimonadales in metabarcoding; The black color represents the percentage of protists that were only identified by one method in a given sample. The percentages of matching groups within the samples is provided in Additional file 5.

In figure 23, it is possible to perceive the presence/dominance of each taxonomic group that matched within the same samples at a higher level for both methodologies. However, because the relative abundances of the common groups in metabarcoding is much lower than in microscopy, the visualization of the different taxonomic groups in each sample with 18S metabarcoding method is difficult.

Concerning the Kongsfjorden transect, there were noticeable differences in relative abundances of the identified taxa between both methods. The taxon *Dinophyceae* is the most dominant in microscopy analysis, especially in KB6_2, KB6_25 and KB3_50 (Fjord/Coastal waters), whereas in metabarcoding this taxon was the second most dominant and its occurrence was higher in different stations (V12_25 to V6_25; Coastal/Offshore waters).

As for *Ciliophora*, despite not being present in all samples, it dominates most samples analyzed by 18S metabarcoding, (from KB3_50 to V6_25; Coastal/Offshore waters). Although, using microscopic analysis even though it was present in almost samples, presented much lower abundance when compared to 18S metabarcoding.

Microscopic analysis also identified *Bacillariophyceae*, Cryptophyta, Prymnesiophyceae* and *Pyramimonadophyceae** (See legend of figure 23) as relative abundant taxa, while in metabarcoding these groups presented very low abundances. As for *Chrysophyceae* it only matches for both methods in KB3_50 station but abundances were low, therefore hard to visualize.

Secondly, in table 15, we presented the matching protists groups (19) within the same samples at genus level identified by both methods. In terms of abundance recovery, microscopy remained the method with higher abundance values and in metabarcoding the abundance values were extremely lower, when compared with microscopy. Given the above, the dominant genus in microscopy belongs to *Phaeocystis*, while in 18S metabarcoding the dominant genus belongs to *Leegaardiella*. The distribution between samples can be observed in following figure 24.

Table 15: Common groups that matched within the same samples in both methods at genus level

Microscopic_PHT		Metabarcoding	
Genus level	**[a]Abundance**	**Genus level**	**[a]Abundance**
Amphidoma	0,04%	**Amphidoma**	0,002%
Chaetoceros	1,05%	**Chaetoceros**	0,02%
Eucampia	0,04%	**Eucampia**	0,001%
Gymnodinium	11,46%	**Gymnodinium**	0,02%
Gyrodinium	2,37%	**Gyrodinium**	0,59%
Heterocapsa	4,08%	**Heterocapsa**	0,05%
Laboea	0,02%	**Laboea**	0,25%
Leegaardiella	0,05%	**Leegaardiella**	2,29%
Navicula	0,004%	**Navicula**	0,001%
Pelagostrobilidium	0,01%	**Pelagostrobilidium**	0,31%
Phaeocystis	15,45%	**Phaeocystis**	0,03%
Prorocentrum	2,13%	**Prorocentrum**	0,15%
Pseudo-nitzschia	0,16%	**Pseudo-nitzschia**	0,001%
Pyramimonas	7,03%	**Pyramimonas**	0,01%
Rhizosolenia	2,02%	**Rhizosolenia**	0,01%
Strombidium	0,17%	**[b]Strombidium K** **[b]Strombidium M**	1,09% 0,04%
Teleaulax	2,85%	**Teleaulax**	0,0005%
Thalassiosira	0,39%	**Thalassiosira**	0,01%
Tintinnopsis	0,002%	**[c]Tintinnopsis 07**	0,27%

[a]Total *relative abundance of all samples;* [bc]*Different names due to the designation assigned according to the PR[2] database*

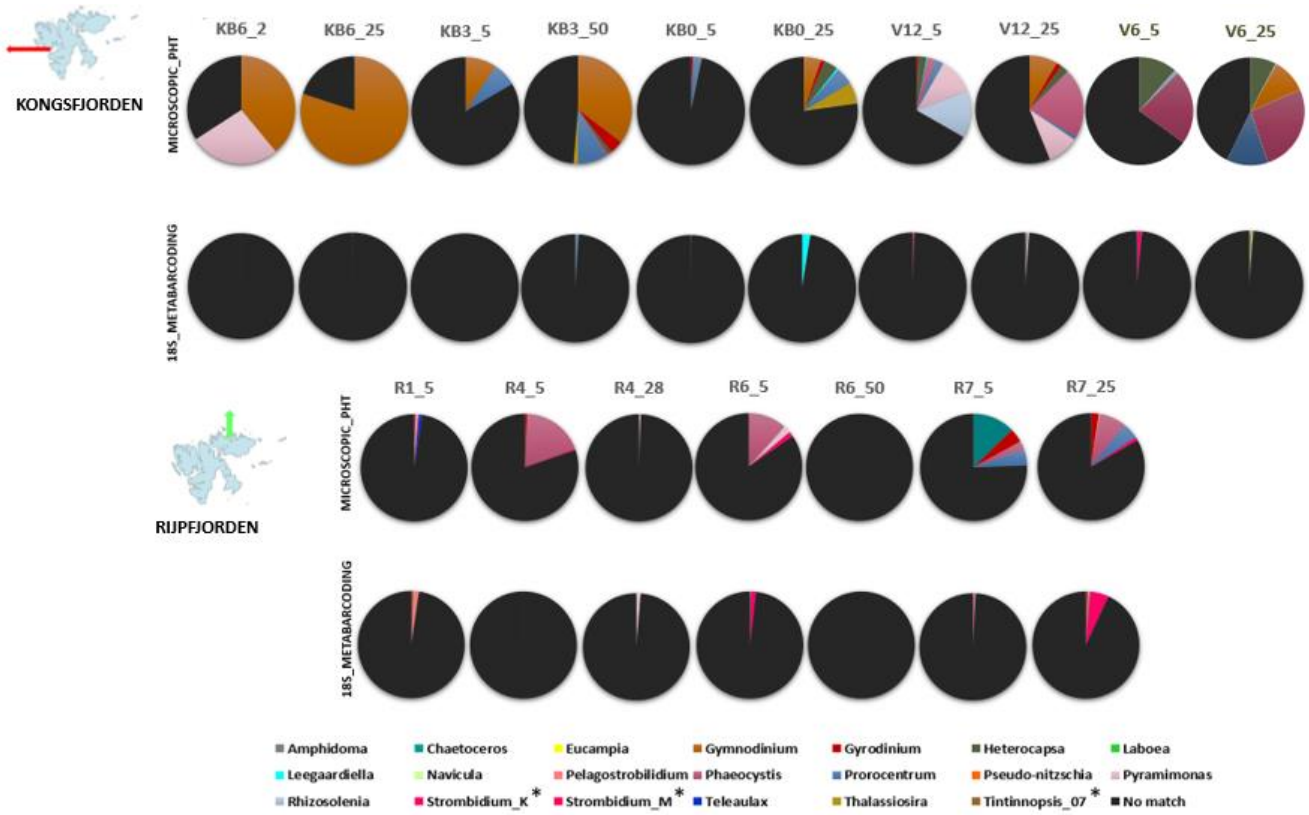

Figure 24 - Pie charts of matching groups (18S vs PHT) at genus level of each methodology within the same samples and separated by transects. *Strombidium_K and Strombidium_M were the taxa detected in metabarcoding that matched with Strombidium in microscopic and Tintinnopsis_07 matched with Tintinnopsis in microscopy; The black color represents the percentage of protists that were only identified by one method in a given sample. The percentages of matching groups within the samples is provided in Additional file 5.

Looking at figure 24, it is clear that the common genera registered in the same samples varied in relative abundance according with the method used. For example, in Kongsfjorden, we can we can observe higher presence of *Gymnodinium* in KB6_2, KB6_25 and KB3_50 using microscopic analysis but using 18S metabarcoding the representativeness of this genus was extremely low (Figure 24). The same was true for *Phaeocystis* at station KB0_25. *Leegaardiella* and *Strombidium_K** were the most dominant genera in 18S metabarcoding analysis that were also identified using microscopic analysis.

These results demonstrated high variability in terms of genus dominance between both methods used for phytoplankton community characterization.

Finally, at the lowest taxonomy level only 5 species matched within samples using both tested methodologies (Table 16; Figure 25). *Phaeocystis pouchetii* was the specie most dominant in microscopy method, where in metabarcoding was *Gyrodinium fusiforme*. The second most dominant in microscopy was *Heterocapsa rotundata* while in 18S metabarcoding was *Laboea strobila*. However, most of the species identified by 18S metabarcoding and microscopic analysis were unique to each method.

Table 16: Common groups that matched within the same samples in both methods at species level

Microscopic_PHT		**Metabarcoding**	
Species Level	**[a]Abundance**	**Species Level**	**[a]Abundance**
Gyrodinium fusiforme	0,28%	***Gyrodinium fusiforme***	0,36%
Heterocapsa rotundata	3,80%	***Heterocapsa rotundata***	0,02%
Laboea strobila	0,02%	***Laboea strobila***	0,27%
Phaeocystis pouchetii	15,31%	***Phaeocystis pouchetii***	0,02%
Pseudo-nitzschia granii	0,07%	***Pseudo-nitzschia granii***	0,0005%

[a]Total relative abundance of all samples.

Microscopic analysis showed that *Phaeocystis pouchetii* was most dominant, especially in stations V12_25 (Kongsfjorden), R4_5, R6_5 and R7_25 (Rijpfjorden), located on Coastal/Open ocean. As for metabarcoding, while the relative percentage of occurrence was much lower, Phaeocystis *pouchetii* was also presented in greater relative abundance in more offshore waters, especially in stations R7_5 and R7_25. As for *Gyrodinium fusiforme,* both methods showed higher presence of this specie in Kongsfjorden samples rather than Rijpfjorden ones (Figure 25).

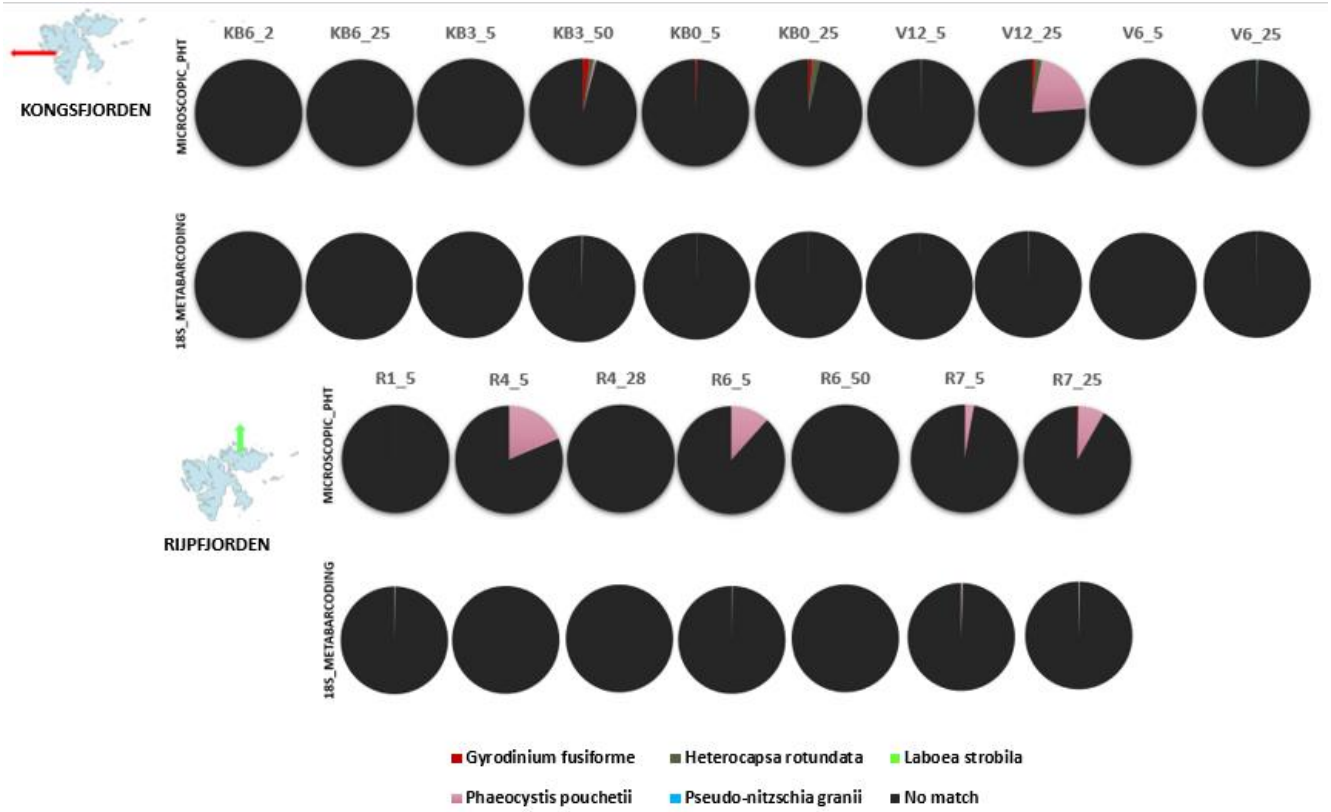

Figure 25 - Pie charts of matching groups (18S vs PHT) at species level of each methodology within the same samples and separated by transects. The black color represents the percentage of protists that were only identified by one method in a given sample. The percentages of matching groups within the samples is provided in Additional file 5.

Groups of phytoplankton that were only identified in microscopy analysis or in 18S metabarcoding

Unique protists groups detected in microscopic PHT analysis

Raphidophyceae (0,76%) was the only protist class that was identified in microscopic analysis and not detected in 18S metabarcoding (Figure 26).

At genus level, 30 were unique of microscopic analysis, being the top 10 most abundant: *Cryptomonas (1,03%); Dinobryon (2,85%); Fragilariopsis (0,41%); Gonyaulax (0,68%); Leucocryptos (1,97%); Monosiga (0,96%); Olisthodiscus (0,76%); Oxyrrhis (1,90%), Protodinium (3,50%)* and *Telonema (1,22%).*

As for species level, there were a total of 49 species that were detected only using microscopic analysis, being the top 10 most abundant: *Dinobryon balticum (2,85%); Gonyaulax gracilis (0,87%); Gymnodinium galeatum (1,97%); Gymnodinium gracilentum (1,34%); Leucocryptos marina (1,97%); Monosiga marina (0,96%); Prorocentrum cordatum (2,10%); Protodinium simplex (3,70%); Rhizosolenia hebetata (1,19%)* and *Rhizosolenia setigera (0,82%).*

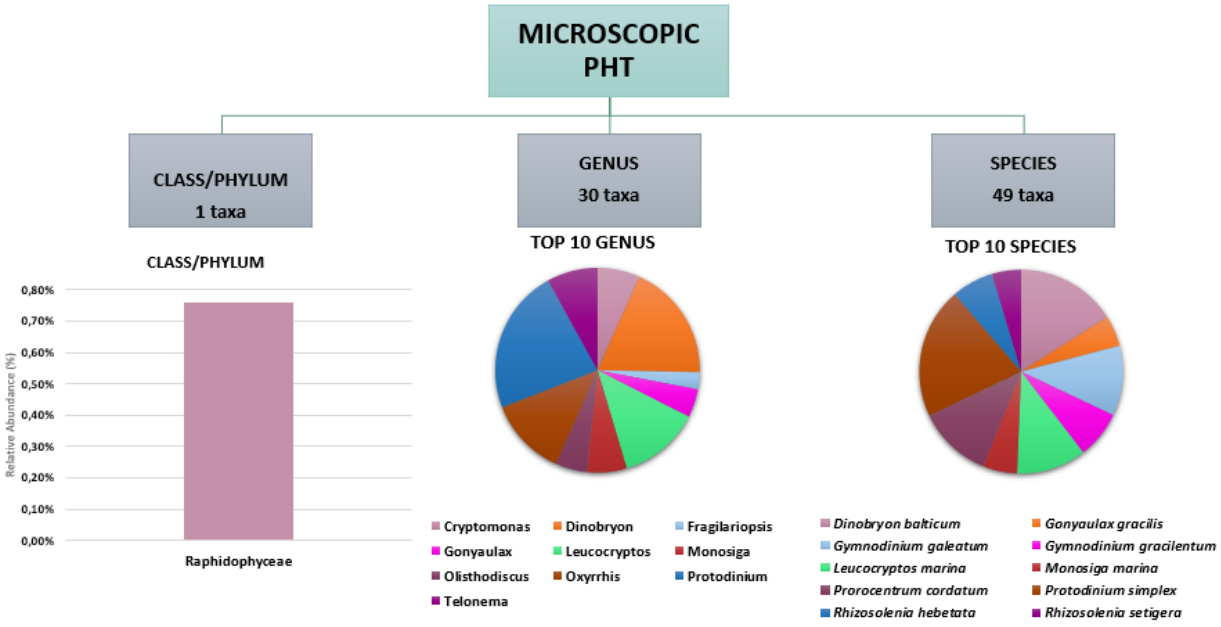

Figure 26 – Number of unique groups of microscopy PHT method at higher and lower level and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 3.

Unique protists groups detected in 18S metabarcoding analysis

A total of 8 phyla, 18 classes and 54 families were exclusively detected using the 18S metabarcoding method (Figure 27). The unique phyla were: *Apusozoa (0,001%); Archaeplastida (0,17%); Choanoflagellida (0,001%); Hacrobia (0,10%); Other_Alveolata (14,80%;) Other_Stramenopiles (0,10%); Rhizaria (0,01%)* and *Syndiniales (0,76%).*

At class level the unique top 10 in terms of most dominant were: *Alveolata_X (4,75%), Apicomplexa (10,05%); Cercozoa (0,01%); Chlorophyta (0,17%); Dinoflagellata (23,53%); Haptophyta (0,08%); Ochrophyta (0,07%); Picozoa (0,01%); Pseudofungi (0,08%)* and *Sagenista (0,01%).* Finally, at family level the unique top 10 most dominant were: *CONThreeP (8,71%); Cylotrichium_like_organism (18,05%); Ellobiopsidae (4,75%); Gregarinomorphea (10,05%); Litostomatea (0,44%); Nassophorea (12,86%); Oligohymenophorea (7,71%); Phyllopharngea (1,57%); Spirotrichea (11,77%)* and *Syndiniales (0,76%).*

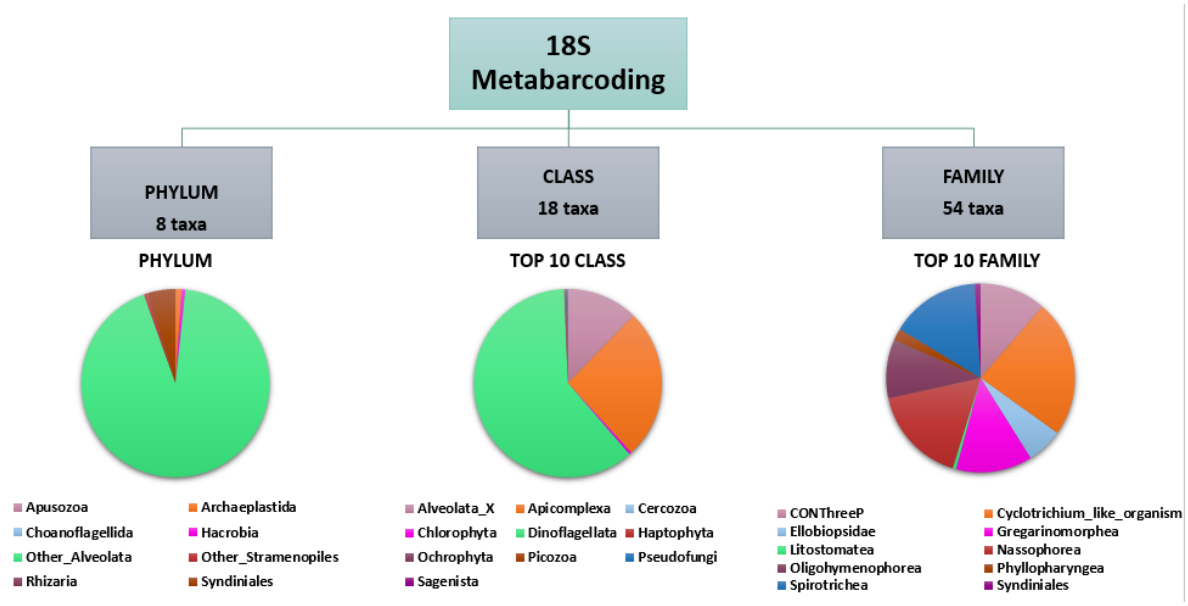

Figure 27 - Number of unique taxonomic groups identified by 18S metabarcoding method at phylum, class and family levels and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 3.

With respect to lower taxonomic level, a much higher number of unique protists groups were observed for 18S metabarcoding analysis, including 241 genera and 129 species.

At genus level the unique top 10 genera were: *Askenasia (6,79%); Chromidina (2,86%); Cyclotrichium (16,53%); Dinophyceae_XXX (25,30%); Gregarines_GRE1_XX (2,51%); Gregarines_GRE2_XX (5,33%); NASSO_1 (12,11%); Philasterida_X (2,43%); Strombidiidae_B_X (1,49%)* and *Thalassomyces (5,44%).*

For species level, the unique top 10 species were: *Gyrodinium_helveticum (0,36%); Gyrodinium_heterogrammum (0,12%); Lankesteria_cystodytae (1,23%); Paracineta_limbata (0,16%); Plagiopyliella_pacifica (0,21%); Pseudocollinia_beringensis (0,32%); Pseudocollinia_similis (0,19%); Selenidium1_serpulae (0,41%); Strombidium_capitatum (1,16%)* and *Thalassomyces_fagei (1,29%).*

The most dominant genus and species in 18S metabarcoding were *Cyclotrichium* and *Thalassomyces_fagei*, respectively (Figure 28).

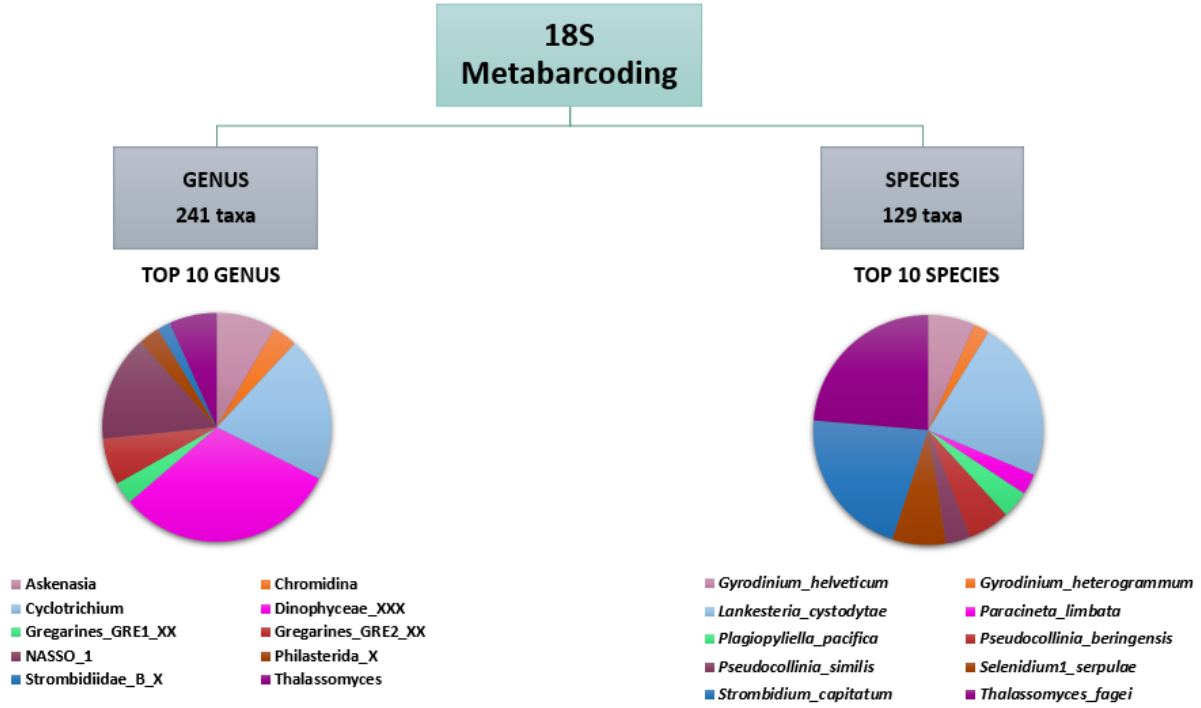

Figure 28 - Number of unique taxonomic groups identified by 18S metabarcoding method at genus and specie levels and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 3.

Methodological comparison between microscopic MIT and 18S metabarcoding

In this comparison a total of 16 samples were included for which 18S metabarcoding and microscopic MIT was performed. Ten of these samples belong to Kongsfjorden transect and six belong to Rijpfjorden. The depth range of these common samples was between 5m to 50m, mostly surface and DCM samples (Figure 29).

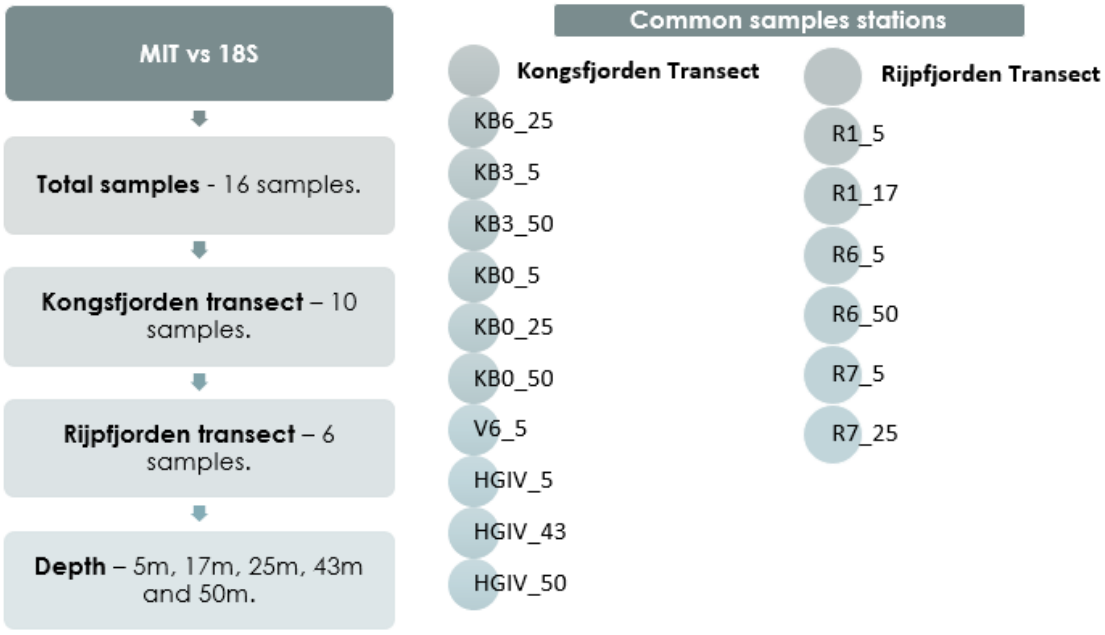

Figure 29 - Data information of the common samples stations and depths used for methodological comparison between 18S metabarcoding and microscopic MIT for community phytoplankton analysis.

Groups of phytoplankton identified in both methodologies

In microscopy MIT data, in the compared 16 samples, the microscopic MIT methodology identified 3 taxa at phylum and class levels, 31 genera and 48 species. As for 18S metabarcoding, a total of 90 taxa were detected at higher level, respectively phylum, class and family level, then 265 genus and 322 species (Figure 30). All the identification of phytoplankton groups in both methodologies, as well the percentage of abundance recovery is provided in Additional file 6.

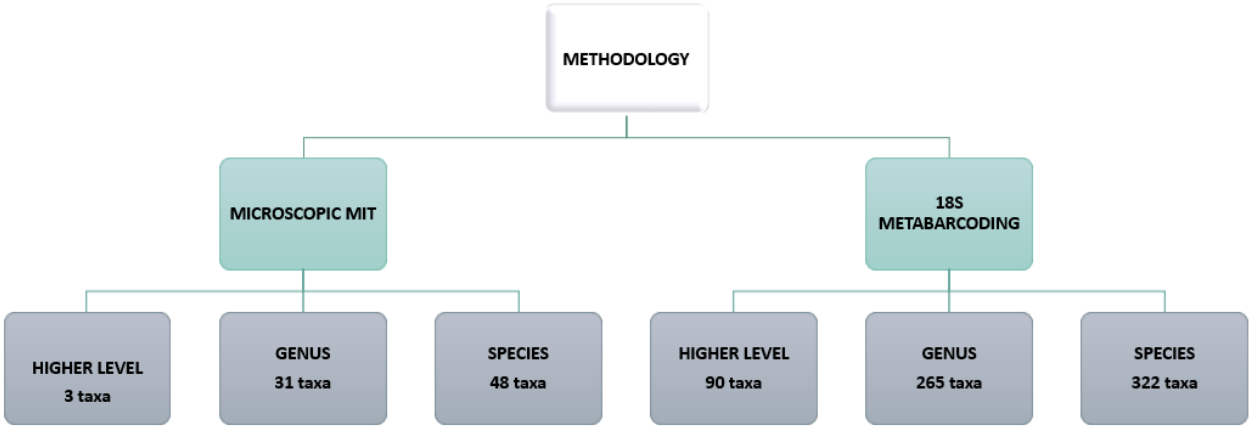

Figure 30 - Diagram with the total number of detected phytoplankton groups in the different methodologies at higher and lower taxonomic levels (18S vs MIT).

Regarding microscopy MIT methodology at higher level, *Bacillariophyceae (74,29%)* was the dominant taxa, being the second most dominant *Dinophyceae (18,03%).* In 18S metabarcoding methodology in the same 16 samples, 11 phyla were identified, being the most abundant: *Ciliophora (64,39%); Dinophyceae (21,71%)* and *Other_Alveolata (12,84%)* (Figure 31; Additional file 6). The distribution of the protists groups within the remaining taxonomic levels at higher level between 18S and microscopic MIT is provided in Additional file 7.

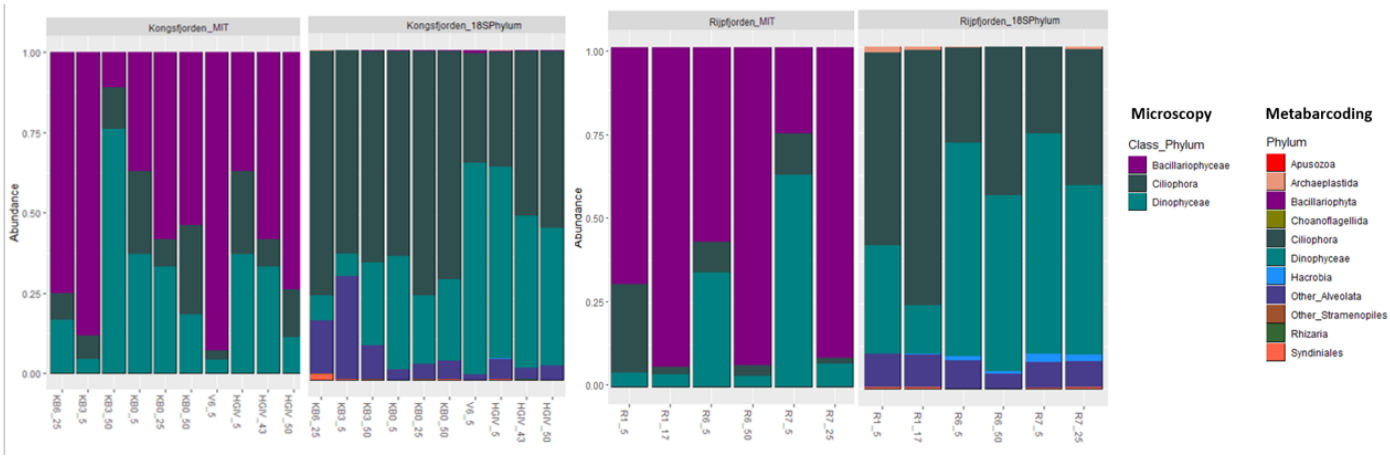

Figure 31 - Taxonomic protists distribution at higher level (Microscopy MIT - Class/Phylum; Metabarcoding - Phylum level) of both methods divided by transects. Kongsfjorden on the left and Rijpfjorden on the right.

Proceeding with 18S, at class level, of the 20 taxa detected, the most dominants class were: *Ciliophora (64,39%); Dinoflagellata (22,27%); Apicomplexa (8,77%)* and *Alveolata_X (4,06%)* and in family level, 59 taxa were detected in the 16 common samples with microscopy and the highest taxa belongs to: *Dinophyceae (21,70%);*

Nassophorea (16,10%); Spirotrichea (15,79%); CONThreeP (14,48%); Gregarinomorphea (8,77%%); Oligophymenophorea (8,44%); Cylotrichium_like_organism (5,78%); Ellobiopsidae (4,06%); Phyllopharngea (2,69%) and *Listomatea (1,00%)* (Additional file 6).

In microscopy approach, the genera with highest relative abundances were: *Rhizosolenia (64,60%); Protoperidinium (11,69%); Chaetoceros (7,70%); Acanthostomella (2,08%); Phalacroma (1,46%); Gyrodinium (1,22%); Strombidium (1,09%)* and *Thalassiosira (1,04%)*. In metabarcoding, a total of 265 taxa were detected and the genera with total highest relative abundances were: *Dinophyceae_XXX (20,86%); NASSO_1 (16,10%); Askenasia (12,63%); Cyclotrichium (5,78%); Gregarines_GRE2_XX (4,99%); Thalassomyces (4,06%); Leegaardiella (3,09%); Chromidina (2,93%); Philasterida_X (2,80%); PHYLL_4_X (2,43%); Strombidium_K (2,04%); Gregarines_GRE1_XX (1,94%); Strombidiidae_B_X (1,93%); Strombidiiae_H_X (1,73%); Urotricha (1,69%); Strombidiidae_M_X (1,42%); Lankesteria (1,38%)* and *Pleurostomatida_X (1,00%)* (Additional file 6).

Finally, for species level, in microscopy approach a total of 48 taxa were detected, being 35 species identified and the remaining 13 were descripted with ‘’sp’ or ‘’indet’’. The more relevant species in terms of relative abundances were: *Rhizosolenia_hebetata (64,61%); Protoperidinium_pellucidum (7,00%); Chaetoceros_decipiens (4,14%); Protoperidinium_pallidum (3,88%); Acanthostomella_norvegica (2,08%); Chaetoceros_convolutus (1,64%); Phalacroma_rotundatum (1,46%)* and *Gyrodinium_fusiforme (1,22%).* For metabarcoding, a total of 322 taxa were recovered at taxonomic species level, being 131 species identified and the remaining 191 descripted with ‘’sp’’. From these 131 species, the ones that presented highest relative abundances were: *Strombidium_capitatum (2,04%)* and *Lankesteria_cytodytae (1,38%)* (Additional file 6).

Groups of phytoplankton that were identified in microscopic MIT and 18S metabarcoding methodologies

Taking all 16 samples into account a total of three protists groups from higher taxonomic level were found to be common to both methods and a total of 12 genera and 4 species were also identified in both methodologies (Figure 32).

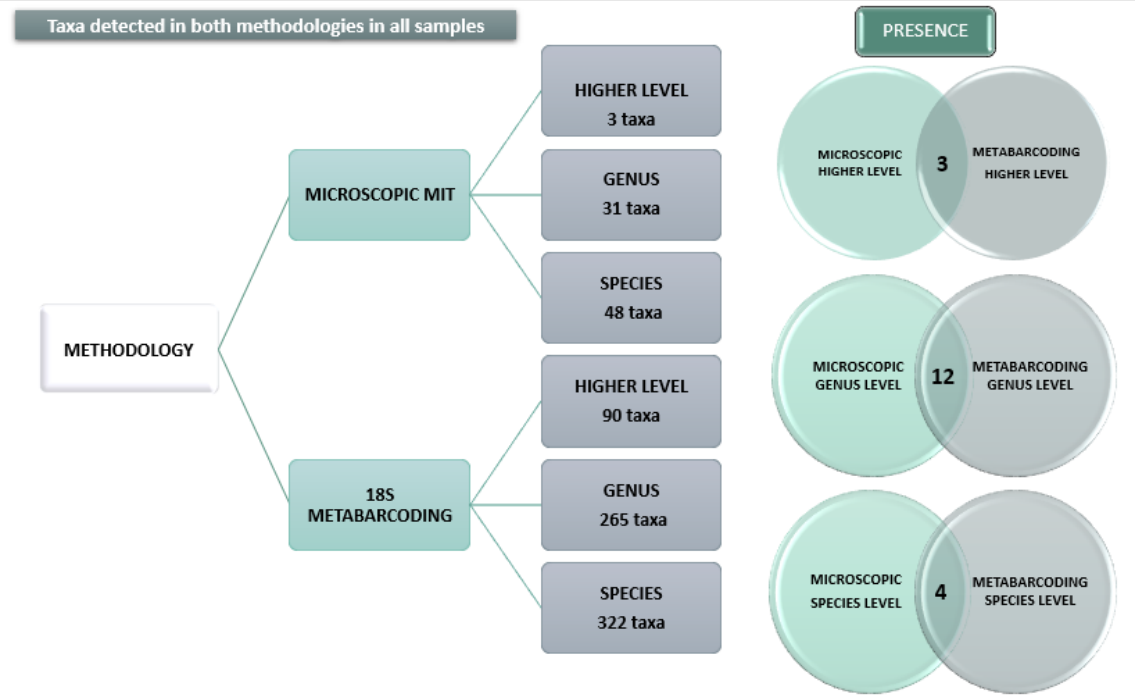

Figure 32 - Diagram showing the number of the detection of protists groups at different taxonomic levels and number of matching taxa that were present in both methods and in all samples (18S vs MIT).

The common higher level taxa were: *Bacillariophyceae, Ciliophora* and *Dinophyceae.* And the common genera and species are represented in table 17.

Table 17: Taxa that were present in both methods at higher and lower level

Higher Level	Genus	Species
Bacillariophyceae/ Bacillariophyta **Ciliophora** **Dinophyceae**	**Chaetoceros** **Dinophysis** **Gyrodinium** **Laboea** **Leegaardiella** **Parafavella** **Ptychocylis** **Rhizosolenia** **Strombidium/Strombidium_K/Strombidium_M** **Thalassiosira** **Tintinnopsis/Tintinnopsis_05/Tintinnospis_07** **Tripos**	***Chaetoceros decipiens*** ***Chaetoceros socialis*** ***Gyrodinium fusiforme*** ***Laboea strobila***

Note: Strombidium* and Tintinnopsis* were the genera detected in both methods, however in metabarcoding was assigned as Strombidium_K; Strombidium_M; Tintinnopsis_05/07 from PR^2.

Relative abundance of phytoplankton groups that were detected in microscopic MIT and 18S metabarcoding methodologies in the same samples

The taxonomic groups that were identified in both 18S metabarcoding and microscopic MIT methodologies that matched within the same samples are represented in figure 33 (3 taxa at higher taxonomic level, 10 genera and 1 specie).

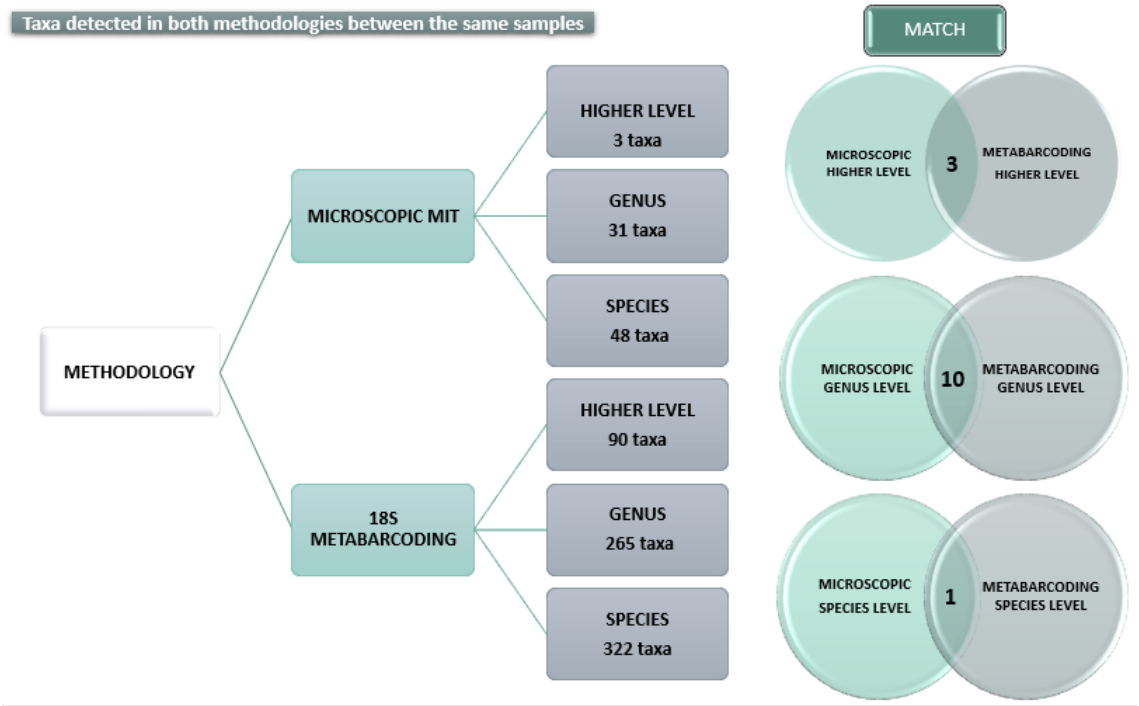

Figure 33 - Diagram showing the number of the detection of protists groups at different taxonomic levels and number of taxa that matched within the same samples (18S vs MIT).

Table 18: Common groups that matched within the same samples in both methods at higher taxonomic level

Microscopic_MIT		Metabarcoding	
Higher Level	**[a]Abundance**	**Higher Level**	**[a]Abundance**
Bacillariophyceae	74,29%	**[b]Bacillariophyta**	0,05%
Ciliophora	7,68%	**Ciliophora**	64,39%
Dinophyceae	18,03%	**Dinophyceae**	21,70%

[a]Total relative abundance of all samples; [b]Different names due to the designation assigned according to the PR² database.

In microscopic MIT analysis the most dominant taxa were, *Bacillariophyceae* followed by *Dinophyceae*. In 18S metabarcoding analysis, *Ciliophora* was the most dominant followed by *Dinophyceae* and represented in much lower relative abundance *Bacillariophyta* (Table 18). The distribution representativeness by individual sample is highlighted in figure 34.

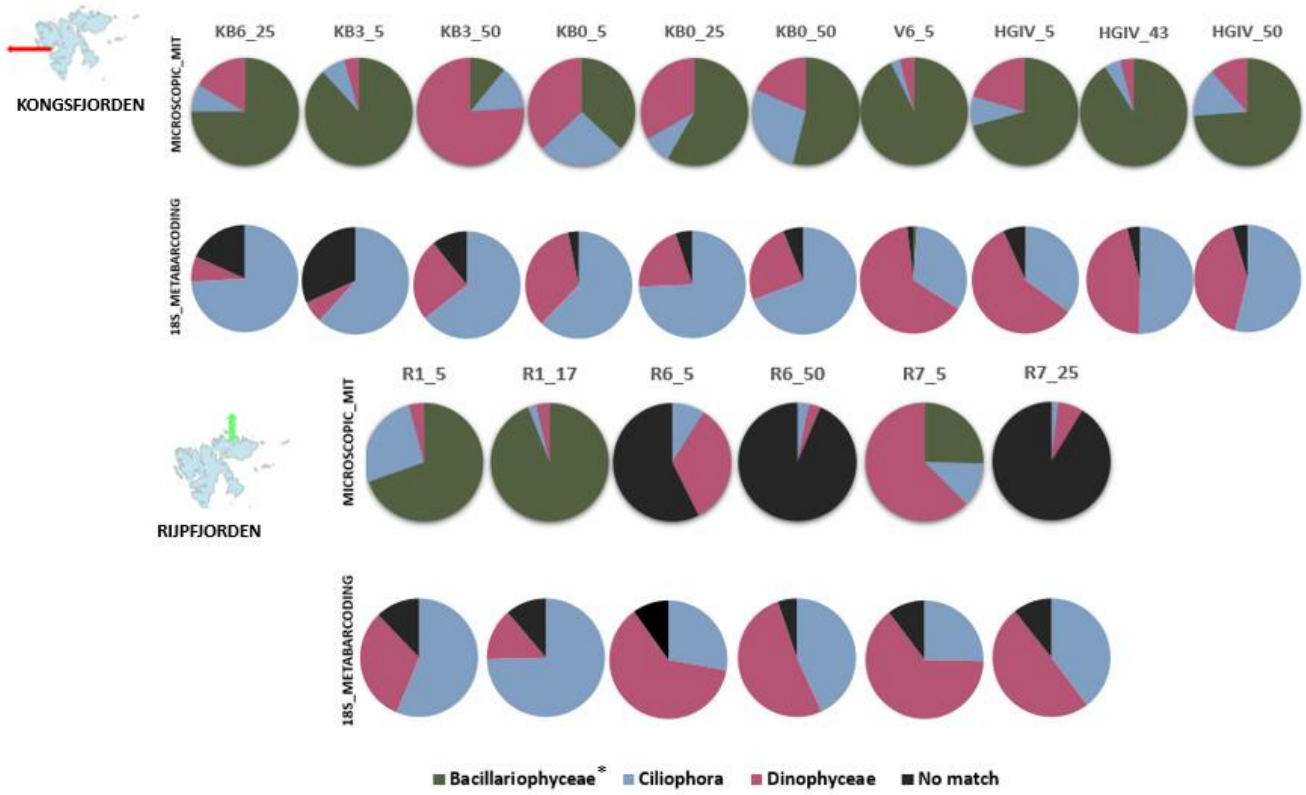

Figure 34 - Pie charts of matching groups (18S vs MIT) at higher level taxonomy of each methodology within the same samples and separated by transects; *Bacillariophyceae was the taxa detected in microscopy that matched with Bacillariophyta in metabarcoding; The black color represents the percentage of protists that were only identified by one method in a given sample. The percentages of matching groups within the samples is provided in Additional file 8.

Bacillariophyceae was highly abundant in almost all samples, especially in Kongsfjorden transect, only in stations R6_5, R6_50 and R7_25 this taxon was absent, when samples were analyzed using microscopic MIT method. In 18S metabarcoding method, *Ciliophora* remained the dominant taxa in both transects, followed by *Dinophyceae* (Figure 34).

Secondly, in table 19, we presented the matching protists groups (10) within the same samples at genus level identified by both methods. Given the above, the dominant genus in microscopy MIT method belongs to *Rhizosolenia* and in metabarcoding method belongs to *Leegaardiella*. The distribution representativeness by individual sample is highlighted in figure 35.

Table 19: Common groups that matched within the same samples in both methods at higher taxonomic level

Microscopic_MIT		Metabarcoding	
Genus Level	**[a]Abundance**	**Genus Level**	**Abundance**
Chaetoceros	7,70%	**Chaetoceros**	0,02%
Dinophysis	0,28%	**Dinophysis**	0,0004%
Gyrodinium	1,22%	**Gyrodinium**	0,36%
Leegaardiella	0,35%	**Leegaardiella**	3,09%
Parafavella	0,87%	**Parafavella**	0,02%
Ptychocylis	0,90%	**Ptychocylis**	0,21%
Rhizosolenia	64,60%	**Rhizosolenia**	0,003%
Strombidium	1,09%	**[b]Strombidium_K**	2,04%
		[b]Strombidium_M	0,10%
Thalassiosira	1,04%	**Thalassiosira**	0,004%
Tintinnopsis	0,70%	**[c]Tintinnopsis_07**	0,50%
		[c]Tintinnopsis_11	0,001%

[a]Total *relative abundance of all samples;* [bc]*Different names due to the designation assigned according to the PR[2] database*

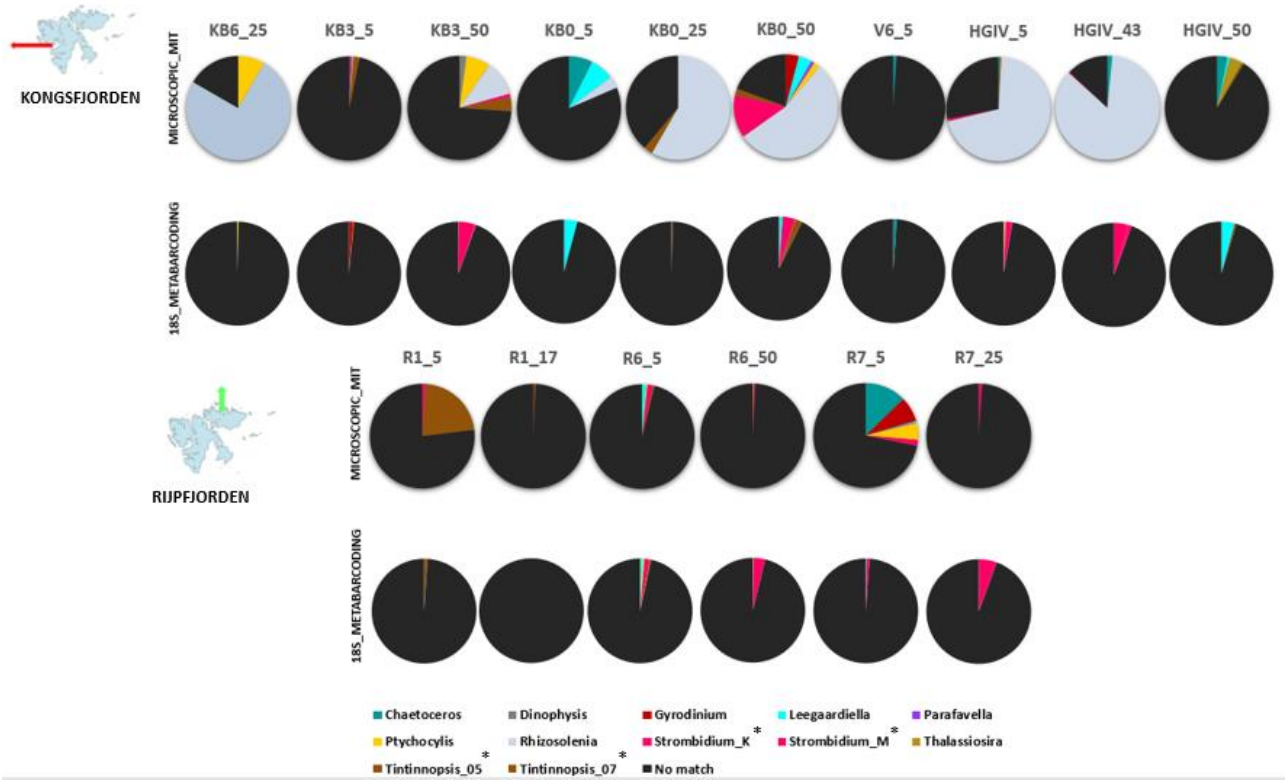

Figure 35 - Pie charts of matching groups (18S vs MIT) at genus level taxonomy of each methodology within the same samples and separated by transects; *Strombidium_K and Strombidium_M were the taxa detected in metabarcoding that matched with Strombidium in microscopic and Tintinnopsis_05/07 matched with Tintinnopsis in microscopy; The black color represents the percentage of protists that were only identified by one method in a given sample. The percentages of matching groups within the samples is provided in Additional file 8.

From the microscopy MIT, the abundant distribution of *Rhizosolenia* across samples from Kongsfjorden is very notable, especially in KB6_25, KB0_25, HGIV_5, HGIV_43 and in Rijpfjorden this genus was absent. In Rijpfjorden transect the genera, *Tintinnopsis, Strombidium, Leegaardiella, Ptychocylis, Chaetoceros* and *Gyrodinium*, were more abundant. In what 18S metabarcoding is respect to the genus *Strombidium_K* and in some stations *Strombidium_M* were more frequently detected at higher relative percentages when comparing with microscopic MIT methodology.

As for species level, only one specie was common between both methods, *Gyrodinium fusiforme*, identified at stations KB3_5 and R7_5. In microscopy method, the total relative abundance was 1,22% and in metabarcoding was 0,22% (Additional file 6).

Groups of phytoplankton that were only identified in microscopy analysis or in 18S metabarcoding

Unique protists groups detected in microscopic MIT analysis

All the higher level taxonomic groups identified using microscopic MIT were also detected using 18S metabarcoding. However, a total of 18 unique genera were identified in microscopic MIT method, being the top 10 most abundant: *Acanthostomella (2,08%); Alexandrium (0,14%); Corethron (0,70%); Didinium (0,21%); Fossula (0,16%); Lohmanniella (0,56%); Phalacroma (1,56%); Polarella (0,67%), Pronoctiluca (0,23%)* and *Protoperidinium (11,69%).*

As for species level, a total of 32 unique species were identified in microscopic MIT method, being the top 10 most abundant: *Acanthostomella norvegica (2,08%); Chaetoceros convolutus (1,64%); Parafavella obtusangula (0,87%); Phalacroma rotundatum (1,46%); Protoperidinium brevipes (0,78%); Protoperidinium pallidum (3,88%); Protoperidinium pellucidum (7,00%); Ptychocylis obtusa (0,90%); Rhizosolenia hebetata (64,61%)* and *Thalassiosira gravida/antarctica (0,90%)*. The unique genus most dominant was *Protoperidinium*, and the specie most dominant was *Rhizosolenia hebetata* (Figure 36).

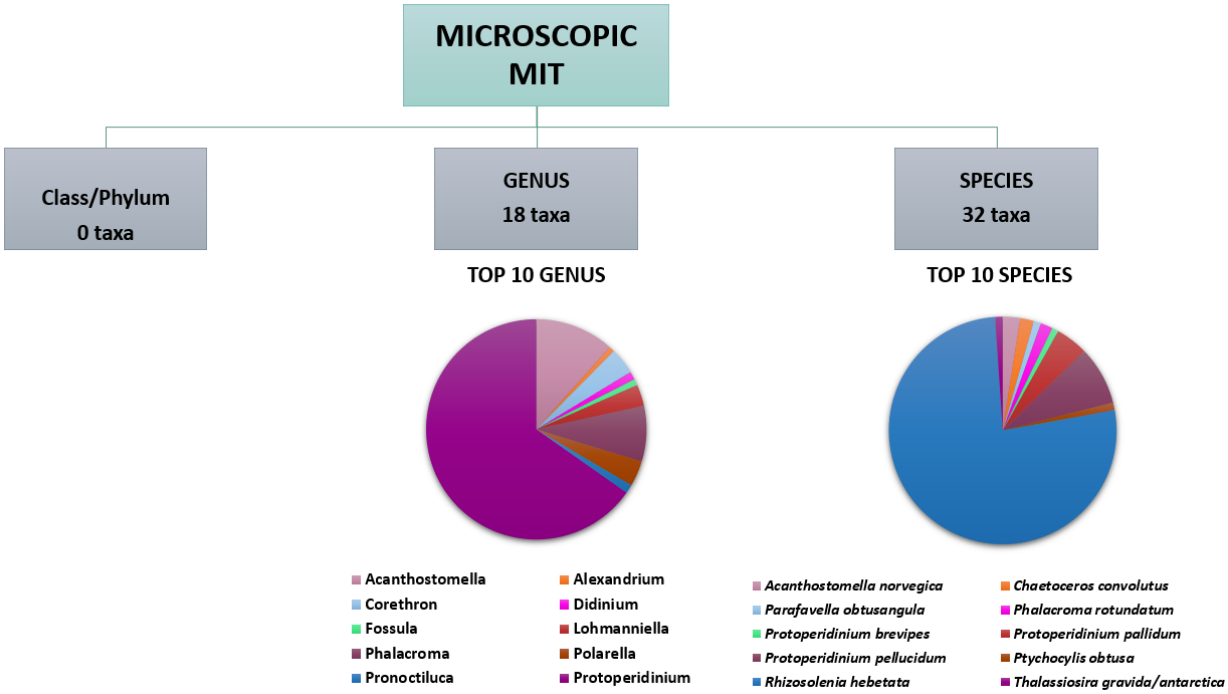

Figure 36 - Number of unique taxonomic groups identified by microscopy MIT method at higher and lower levels and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 6.

Unique protists groups detected in 18S metabarcoding analysis

A total of 8 unique phyla, 19 unique classes and 57 unique families were only detected in 18S metabarcoding. At phylum level the unique taxa were: *Apusozoa (0,004%); Archaeplastida (0,26%); Choanoflagellida (0,001%); Hacrobia (0,13%); Other_Alveolata (12,84%;) Other_Stramenopiles (0,04%); Rhizaria (0,01%)* and *Syndiniales (0,57%).* At class level the unique top 10 class most dominant were: *Alveolata_X (4,06%), Apicomplexa (8,77%); Cercozoa (0,01%); Chlorophyta (0,26%); Cryptophyta (0,02%); Dinoflagellata (22,27%); Haptophyta (0,11%); Ochrophyta (0,05%); Picozoa (0,01%); Pseudofungi (0,02%)* and *Sagenista (0,02%).*

At family level the unique top 10 most dominant were: *CONThreeP (14,48%); Cylotrichium_like_organism (5,78%); Ellobiopsidae (4,06%); Gregarinomorphea (8,77%); Litostomatea (1,00%); Nassophorea (16,10%); Oligohymenophorea (8,44%); Phyllopharngea (2,69%); Spirotrichea (15,79%)* and *Syndiniales (0,57%).*

The unique phylum most dominant was *Other_Alveolata*, the most dominant class was *Dinoflagellata* and the most dominant family was *Nassophorea* (Figure 37).

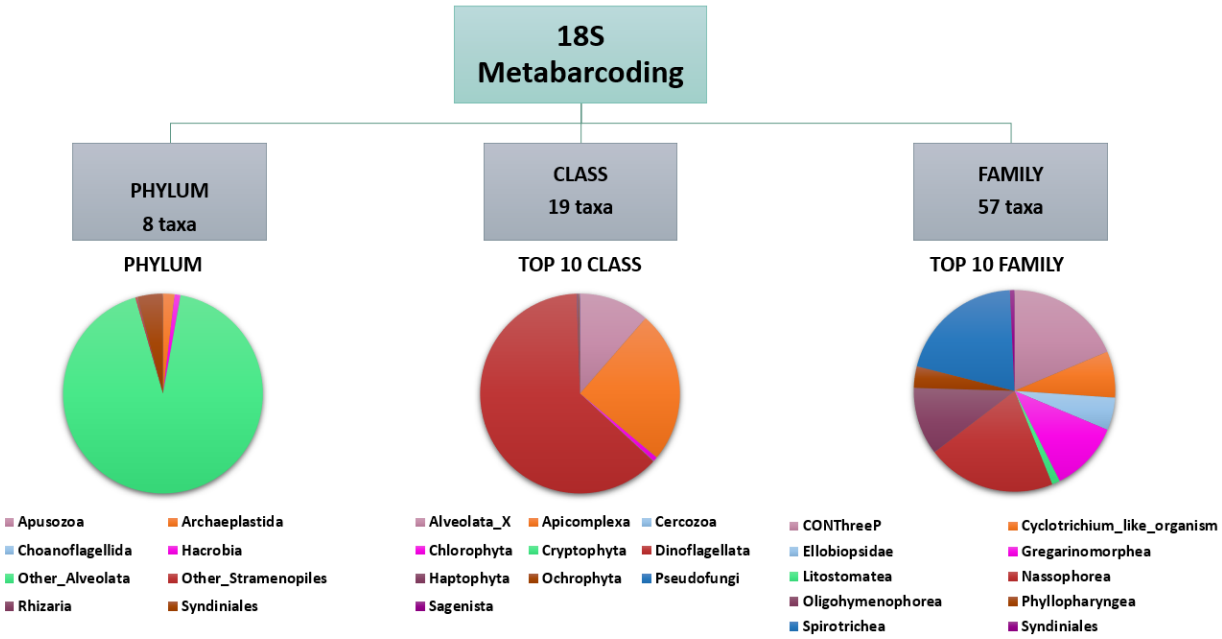

Figure 37 - Number of unique taxonomic groups identified by 18S metabarcoding method at phylum, class and family levels and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 6.

For the lowest taxonomic level, a huge number of unique genera and species were identified with 18S metabarcoding method (250 and 125, respectively) (Figure 38). The most dominant, top 10 genera were: *Askenasia (12,63%); Chromidina (2,93%); Cyclotrichium (5,78%); Dinophyceae_XXX (20,86%); Gregarines_GRE1_XX (1,94%); Gregarines_GRE2_XX (4,99%); NASSO_1 (16,10%); Philasterida_X (2,80%); PHYLL_4_X (2,43%)* and *Thalassomyces (4,06%).*

For species level, the unique top 10 species were: *Lankesteria_cystodytae (1,38%); Micromonas_polaris (0,17%); Paracineta_limbata (0,25%); Plagiopyliella_pacifica (0,22%); Pseudocollinia_beringensis (0,42%); Pseudocollinia_similis (0,28%); Selenidium1_serpulae (0,47%); Strombidium_capitatum (2,04%)* and *Thalassomyces_fagei (0,35%).* The genus most dominant was *NASSO_1* and the specie most dominant was *Strombidium_capitatum*.

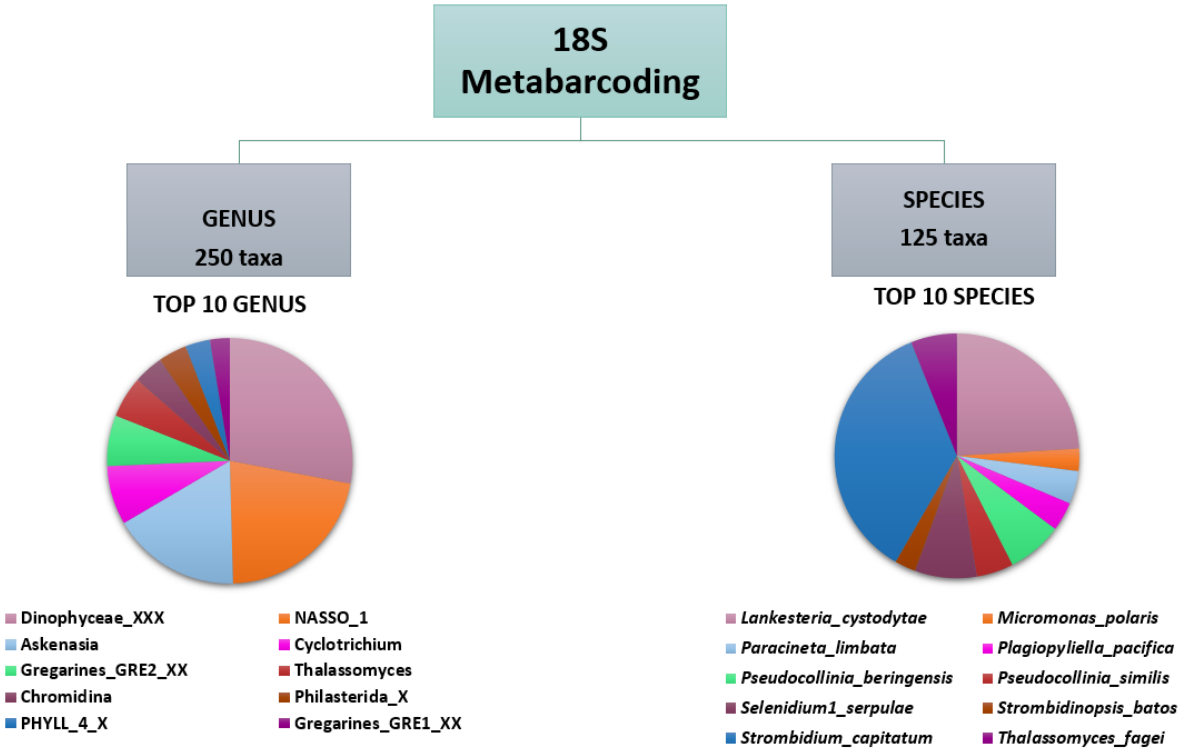

Figure 38 - Number of unique taxonomic groups identified by 18S metabarcoding method at genus and specie levels and top 10 most abundant taxa. The percentage of the remaining unique groups is provided in Additional file 6.

Discussion – Part B

Comparison between 18S rRNA Metabarcoding and Microscopic cell counts to study protists diversity

The comparison between morphological and molecular approaches for studying microplankton communities by analysing the results obtained from each method is a big challenge. In fact, many studies aimed to describe this comparison, suggesting that there is a need to systematically review these two methodologies, understand their limitations and find out what could be improved in order to have a "*successful marriage*'' by these two methodologies as mentioned by McManus et al. [71]. In this section, we will discuss the results observed by comparing both methodologies in samples from Arctic MIZ, relevant for understanding the protists ecological insights provided by each methodology.

Our results revealed that the databases of the protist collection from MOSJ-ICE2016 generated by the different methodological approaches presented clear differences. In our study, the same samples were processed by both methods and the comparison was made in order to identify common and unique taxonomic groups that were able to be detected by 18S metabarcoding and microscopy. In 18S rRNA metabarcoding, the database greatly exceeded the number of protist taxa compared to microscopic analysis. Although protists identification by microscopy allowed to detect lower abundance groups, especially in samples where the MIT sampling approach was applied, the 18S metabarcoding approach was able to detect much more low abundance taxa. However, there are different factors that need to be considered when performing a methodological comparison between genomic and morphological approach to study protists. This includes a precise standardization of the sampling volume analysed by each methodology and mostly importantly a careful understanding of the complexity of the databases and the taxonomic nomenclature that each methodology provides. For example, the 18S metabarcoding approach and the taxonomic nomenclature provided by PR^2 database is different and much more complex than the nomenclature used in classic microscopic classification. The 18S metabarcoding approach detected numerous sequences affiliated to Clades, several Dino-Groups, MAST (Marine Stramenopiles), MOCH (Marine OCHrophyta), among others [86,131]. On the other hand, in microscopy, the taxonomic assignation presents several complexities related for example with numerous unidentified *flagellates (Flagellates indet)*, under which many unidentified species grouped into different size classes. In this approach when it was not possible to identify a particular specie, it was designated as *"incertae sedis"*. Moreover, in both methods, it was not possible to identify certain groups in the different taxonomic levels, which remained with uncertain assignment (e.g., "sp"; "indet. cyst"; "_X"), that depended

on the respective database and methodology used. We suggest the reader to review Hop et al. [52) and Guillou et al. [86]. for better understanding of the assignment of the taxonomic nomenclature for each methodology.

So, in light of the above, the comparative analysis between 18S metabarcoding and microscopic identification across the Svalbard protists community, showed significant differences in terms of both detection and abundance of different taxa recovered by each approach. Morphology-based identification recovered by microscopy identified fewer taxonomic groups and catched lower protists diversity at higher and lower taxonomic levels than 18S metabarcoding method. In fact, it was not surprising that 18S metabarcoding recovered much higher diversity and identified more protists groups, since it has the capacity to detect the small size fraction of protists communities, to provide unprecedented insights of the cryptic communities and better coverage to identify rare taxa within an ecosystem [56,57]. Nevertheless, the abundance recovery in microscopy method, both for PHT and MIT sampling approach was found to be higher than in metabarcoding. This was evident in the protists groups that were identified by using both approaches, where it can be noticed great abundances in microscopy, even after standardizing the sampling volume between the different methodologies. There are only a few exceptions where metabarcoding exceeded microscopy in terms of the abundance of common taxa groups (*Ciliophora*, *Gyrodinium fusiforme* and *Laboea Strobila*).

The most similar abundance recoveries can be found in *Dinophyceae* (metabarcoding 22,77% and microscopy 24,20% - PHT vs 18S) and (metabarcoding 21,70% and microscopy 18,03% - MIT vs 18S). In addition, the distribution of taxa between both methods varies depending on the samples.

The unique taxonomic groups detected by each methodology represent another important result of the comparison. Each approach was able to detect unique groups, both at the highest and lowest taxonomic level, suggesting that microscopy and molecular methodologies could complement each other in order to have a better recovery of the natural planktonic protists diversity. On one hand, our results showed that both methodologies provided different ecological perspectives/insights and discoveries, since there are unique taxa detected by each methodology. On the other, our results highlighted that, when carrying out a particular research using molecular or morphological methods and, in order to have the actual protists community structure in a particular planktonic environment, limitations and awareness of the specific methodologies should be considered.

The taxonomic distribution of common groups within the same samples revealed, as mentioned above, great discrepancies in terms of diversity and relative abundance recovery. In microscopy, the abundance recovery of the different groups was higher and metabarcoding was extremely low, making the comparison by sample quite difficult. Also, when a given common taxonomic group is the most dominant in one method is extremely low in the other and this applies to both PHT and MIT when compared to 18S.

Between PHT and 18S, at higher taxonomic level, *Ciliophora* revealed to be the dominant group in 18S metabarcoding, where in microscopy *Dinophyceae* was the dominant one. At genus level, *Leegaardiella* was the dominant genus in metabarcoding, and in microscopy was *Phaeocystis*, however the distribution of these genera were less perceivable within different samples, where it can be noticed that other genera were well represented (i.e., *Gymnodinium* in microscopy and *Strombidium_K* in metabarcoding). Finally, for species level, *Phaeocystis pouchetii* was the dominant specie in microscopy in terms of abundance and *Gyrodinium fusiforme* was the dominant one in 18S metabarcoding. In contrast to PHT, the comparison between MIT and 18S revealed different out-comes, as we can verify that in higher level *Bacillariophyceae* proved to be the dominant one and *Rhizosolenia* revealed to be the most dominant genus. In metabarcoding, *Ciliophora* (at higher level) and *Leegaardiella* (at lower level genus) remained the dominant groups. At species level, only one specie matched, i.e., *Gyrodinium fusiforme*.

These out-comes are in agreement with studies from microscopy records conducted in Arctic region, where Dinoflagellates (*Dinophyceae*) and Diatoms (*Bacillariophyceae*) were extensively studied and demonstrated to be crucial members of protists community, with *Bacillariophyceae,* being an important group in spring bloom production [52,125,132]. Furthermore, the predominance of the specie *Phaeocystis pouchetii* has been mentioned in several articles since early studies [41,42,133]. These evidences from early to recent studies, may suggest that *Phaeocystis pouchetii* adapted to environmental stressors and is a resilient species to climate change [41,42,133]. In MIT data, the genus *Rhizosolenia* showed great abundance recovery, and the specie more dominant was in fact, *Rhizosolenia hebetata.* This marine large diatom was also identified in other studies from Atlantic-Arctic waters and from Wijdefjord with traces of North Atlantic water [39,41,134]. This supports the fact that the North Atlantic waters are enhancing protists diversity in the Arctic Ocean, influencing the biogeography of this genus which in our study was recovered in both transects studied.

Contrary to microscopy data, in metabarcoding, *Ciliophora* remained the dominant higher taxonomic level group, and regarding to genus level *Leegaardiella* was

the dominant genus both for 18S vs PHT/MIT. This dominant genus makes sense, since according to its taxonomic classification, it belongs to the *Ciliophora* phylum [135].

Concerning the common identified species between 18S metabarcoding and microscopy using the MIT sampling approach, only one specie was identified in both methods, *Gyrodinium fusiforme.* This specie, according to Algae Base, was found in Arctic waters in early studies and still marks presence according to recent studies, suggesting that it is able to manage survival skills to different environmental conditions [125,136,137,138].

The unique protists groups identified in each methodology provide deeper ecological insights and even more awareness of the comparison limitations. Each method was able to detect groups that were not present in the other, and in metabarcoding approach the capability of detection taxonomic groups not identified by using the microscopy was much higher. Indeed, comparative studies of these two methods already reported the existence of unique protists groups of each methodology [56,57,71,78]. The discrepancies of relative abundance taxa between different methods could have to do with the limitations and biases of molecular methodologies [56,57,69,78,139]. It should be noted that some advantages and disadvantages were highlighted in the introductory section (tables 4 and 5), however in order to discuss the discrepancies between the two methodologies, some aspects will be highlighted in this section. For instance, using the 18S metabarcoding approach we were able to detect cryptic species, and also to catch an enormous diversity of picophytoplankton that wasn't visible to microscopy, due to the small size of these organisms, in optical/fluorescence microscopy. On the contrary, molecular detection doesn't take into account the life stages of protists (the juvenile and adult stage), that could be an important aspect when studying the dynamics of natural protists communities [56]. Moreover, significant gaps are still found in sequence reference databases and the catalogue of protists morphologically identified [56]. This could explain the identification of taxa in microscopy not detected in 18S metabarcoding approach. Also, in 18S metabarcoding approach, other possible methodological biases that could compromise the comparison analysis, has to do with the sequencing depth that could inflate the diversity detection [76,140]. The primer choice could significantly affect the results by amplifying non-target groups and target sequences more than others [72]. Furthermore, using multiple primer sets, or multiple barcode genes for different target taxonomic groups could help to reduce primer bias, but it is more costly and time-consuming, so that could not be a preferable choice by the researcher [72]. Regarding the copy number, there may be an association

between cell size, biovolume and gene copies of the SSU rDNA, that could partially explain the disparities in abundances between these methods [141,142].

Finally, it should also be highlighted that the V4 region as a barcoding marker has some features that need to be deeper explored. For instance, as reported in the study by Mordret et al. [77], in some cases it is unable to discriminate species that are close relatives and therefore there is the possibility of ''*shared species within the same region*'' (e.g. there are cases within the *Dinoflagellates* in which sequences belonging to closely related species, shared identical V4 regions and discrimination failed). Indeed, it could represent a limitation in the detection of species by metabarcoding [77]. In these cases, more/different markers or techniques should be used to solve their discrimination. However, in the present study this issue was not explored in detail. Further studies focused for example within a specific taxonomic group of the MIZ database will aim to shed light and get detailed knowledge on this topic.

As for microscopy biases, there's also limitations that could affect the comparison, for example, introduced biases in laboratory procedures, possible contamination, cell losses in the fixation procedures, the taxonomic designation sometimes is difficult because of the life stage and identification of cryptic species and therefore it could be difficult to reach to a lower taxonomic level [56,69,78,139]. Also, there is a need to catch up with the taxonomy field and this applies for implementation of 18S rRNA metabarcoding and microscopy approaches. It is important to confirm the status of each taxonomic group on the respective databases, and when possible continuous communication with experts of each methodology in order to have a more precise comparison. Lastly, considering all these factors it is difficult to answer the question of what will be the best method for studying phytoplankton communities, an issue that is also well expressed in the literature and most of these studies indicate that one method will complement the other.

Final Considerations and Future Perspetives

The MOSJ-ICE2016 eukaryote datasets offer new enrichment insights and perspectives to study diversity and structure of the protists community along two oceanographic transects in the Marginal Ice Zone around Svalbard by performing a comprehensive comparison between the 18S metabarcoding approach, using new generation sequencing techniques, and the classic microscope cell identification.

These results show strong evidence of a diverse community structure, with a marked biogeographic pattern of the protists communities along the Svalbard MIZ with a clear trend of depth-dependency along the water column, water masses variations and nutrients availability.

Comparing both transects studied, Kongsfjorden has a more ‘’*Atlantic*’’ environment signature, with higher temperatures, fjords glaciers regressing at a very accelerated rate and with infiltration of warmer, saltier and denser waters from the North Atlantic. As for Rijpfjorden, this transect is less dynamic in temporal terms, the fjord glaciers are more intact and it is more influenced by polar waters, therefore presenting a more ‘’*Arctic*’’ environment signature. Due to climate change, increasing temperatures and consequently sea ice melt are affecting Arctic’s primary productivity and biogeochemistry by the result of higher organic matter values and nutrients availability, probably enhanced by the upwelling effects at bottom depths where the protists community seem to be more connective with higher diversity.

Methodological comparison showed significant differences on protists abundance and diversity recovery, suggesting that 18S metabarcoding and microscopic counts are complementary methodologies to study the dynamics of phytoplankton. Metabarcoding approach was able to detect higher diversity phytoplankton groups, while microscopy detected lower diversity but still recovered protists groups not identified in 18S metabarcoding, suggesting that this method is probably more realistic when analyzing the abundant patterns of the common protists groups.

These significant methodological differences demonstrate that there is a need to improve the techniques or use different complementary approaches to recover the phytoplankton diversity in natural planktonic ecosystems. However, despite the discrepancies this is the first step to improve the comparison for the protists community in the Arctic region and could serve as a baseline for further protists studies.

For next steps and future work, it is important to understand the full limitations of each methodology by trying to fill the gaps, to improve the methodological comparison process that could serve as a baseline for future monitoring programs. Moreover, it is urgent to fill critical gaps concerning the response of the main Arctic primary producers

to climate driven changes using 18S metabarcoding and microscopic counts as complementary methodologies at an extended time scale.

References

[1] J. B. H. Martiny *et al.*, “Microbial biogeography: Putting microorganisms on the map,” *Nat. Rev. Microbiol.*, vol. 4, no. 2, pp. 102–112, 2006.

[2] K. M. Meyer, H. Memiaghe, L. Korte, D. Kenfack, A. Alonso, and B. J. M. Bohannan, “Why do microbes exhibit weak biogeographic patterns?,” *ISME J.*, vol. 12, no. 6, pp. 1404–1413, 2018.

[3] P. E. Galand, E. O. Casamayor, D. L. Kirchman, and C. Lovejoy, “Ecology of the rare microbial biosphere of the Arctic Ocean,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 52, pp. 22427–22432, 2009.

[4] J. T. Staley and J. J. Gosink, “Poles apart: Biodiversity and biogeography of sea ice bacteria,” *Annu. Rev. Microbiol.*, vol. 53, no. October, pp. 189–215, 1999.

[5] W. Foissner and D. Hawksworth, *Protist Diversity and Geographical Distribution*, no. 1. 2009.

[6] GreenFacts, “Facts on Arctic Climate Change,” 2004.

[7] H. Ahlenius, K. Johnsen, and C. Nellemann, *Vital Arctic Graphics - People and global heritage on our last wild shores*. 2005.

[8] S. Lundgren, “Polar Ecology,” Nuuk, 2001.

[9] WWF-Norway, “Effects of climate change on arctic marine mammal health,” Oslo, Norway, 2008.

[10] E. Kintisch and M. Brangan, “What melting sea ice means for life in the Arctic,” *Vox*, 2018. [Online]. Available: https://www.vox.com/2018/4/10/17210240/melting-arctic-sea-ice-research. [Accessed: 05-Jun-2020].

[11] O. Müller, L. Seuthe, G. Bratbak, and M. L. Paulsen, “Bacterial response to permafrost derived organic matter input in an Arctic Fjord,” *Front. Mar. Sci.*, vol. 5, no. AUG, 2018.

[12] O. Anisimov, B. Fitzharris, J. O. Hagen, R. Jefferies, and D. G. Marchant, H Nelson, F Vaughan, “Polar regions (arctic and antarctic),” 2007.

[13] “Climate change in the Arctic,” *Norwegian Polar Institute*, 2020. [Online]. Available: https://www.npolar.no/en/themes/climate-change-in-the-arctic/#tab-id-2. [Accessed: 05-Jun-2020].

[14] W. A. Programme, “Arctic Climate Change,” *WWF - World Wide Fund For Nature*, 2020. [Online]. Available: https://arcticwwf.org/work/climate/. [Accessed: 18-Jun-2020].

[15] CAFF, “Arctic Biodiversity Trends 2010 – Selected indicators of change,” Akureyri, 2010.

[16] M. Jacob, “Influence of Global Change on microbial communities in Arctic sediments,” Universitat Bremen, 2014.

[17] European Environment Agency, “The Melting Arctic,” 2012.

[18] M. A. Granskog, P. Assmy, S. Gerland, G. Spreen, H. Steen, and L. H. Smedsrud, “Arctic research on thin ice: Consequences of Arctic sea ice loss,” *Eos (Washington. DC).*, vol.

97, no. 5, pp. 22–26, 2016.

[19] P. Duarte *et al.*, “Sea ice thermohaline dynamics and biogeochemistry in the Arctic Ocean: Empirical and model results,” *J. Geophys. Res. Biogeosciences*, vol. 122, no. 7, pp. 1632–1654, 2017.

[20] D. Dahl-Jensen *et al.*, “The Greenland Ice Sheet in a Changing Climate,” 2009.

[21] M. Sommerkorn and S. J. Hassol, “Arctic climate feedbacks: global implications,” Oslo, 2009.

[22] W. Curry, “The Arctic and Climate Change,” 2007.

[23] J. Boé, A. Hall, and X. Qu, “September sea-ice cover in the Arctic Ocean projected to vanish by 2100,” *Nat. Geosci.*, vol. 2, no. 5, pp. 341–343, 2009.

[24] N. Diersing, “Phytoplankton Blooms: The Basics,” 2009.

[25] S. Baldursson, “Module 8 Life in the Ocean,” 2004.

[26] K. R. Arrigo *et al.*, “Massive phytoplankton blooms under arctic sea ice,” *Science (80-.).*, vol. 336, no. 6087, p. 1408, 2012.

[27] P. Assmy *et al.*, “Leads in Arctic pack ice enable early phytoplankton blooms below snow-covered sea ice,” *Sci. Rep.*, vol. 7, no. January, pp. 1–9, 2017.

[28] K. M. Lewis *et al.*, “Photoacclimation of Arctic Ocean phytoplankton to shifting light and nutrient limitation,” *Limnol. Oceanogr.*, vol. 64, no. 1, pp. 284–301, 2019.

[29] Arctic Monitoring and Assessment Programme (AMAP), “AMAP Assessment 2018: Arctic Ocean Acidification,” 2018.

[30] R. Cavicchioli *et al.*, “Scientists’ warning to humanity: microorganisms and climate change,” *Nat. Rev. Microbiol.*, vol. 17, no. 9, pp. 569–586, 2019.

[31] J. Beardall, S. Stojkovic, and S. Larsen, “Living in a high CO2 world: Impacts of global climate change on marine phytoplankton,” *Plant Ecol. Divers.*, vol. 2, no. 2, pp. 191–205, 2009.

[32] C. M. Duarte *et al.*, “Tipping elements in the Arctic marine ecosystem,” *Ambio*, vol. 41, no. 1, pp. 44–55, 2012.

[33] CAFF, “State of the Arctic Marine Biodiversity Report,” 2017.

[34] J. E. Tesdal, R. P. Abernathey, J. I. Goes, A. L. Gordon, and T. W. N. Haine, “Salinity trends within the upper layers of the subpolar North Atlantic,” *J. Clim.*, vol. 31, no. 7, pp. 2675–2698, 2018.

[35] Nordic Council of Ministers, *Marine invasive species in the Arctic*. 2013.

[36] L. Halbach, A. Wulff, and P. Assmy, “Mechanisms Determining Phytoplankton Dynamics in Front of Active Tidewater Glaciers in,” University of Gothenburg, 2018.

[37] A. S. Brierley and M. J. Kingsford, “Impacts of Climate Change on Marine Organisms and Ecosystems,” *Curr. Biol.*, vol. 19, no. 14, pp. R602–R614, 2009.

[38] G. M. Hallegraeff, “Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge,” *J. Phycol.*, vol. 46, no. 2, pp. 220–235, 2010.

[39] C. H. Von Quillfeldt, “Common diatom species in Arctic spring blooms: Their distribution and abundance,” *Bot. Mar.*, vol. 43, no. 6, pp. 499–516, 2000.

[40] H. Hop *et al.*, “The marine ecosystem of Kongsfjorden, Svalbard,” *Polar Res.*, vol. 21, no. 1, pp. 167–208, 2002.

[41] E. N. Hegseth *et al.*, “Phytoplankton Seasonal Dynamics in Kongsfjorden, Svalbard and the Adjacent Shelf,” Cham, Swit., Tromsø, 2019, pp. 173–227.

[42] G. Rytter Hasle and B. Riddervold Heimdal, “The net phytoplankton in Kongsfjorden, Svalbard, July 1988, with general remarks on species composition of Arctic phytoplankton,” *Polar Res.*, vol. 17, no. 1, pp. 31–52, 1998.

[43] G. Owrid *et al.*, “Spatial variability of phytoplankton, nutrients and new production estimates in the waters around Svalbard,” *Polar Res.*, vol. 19, no. 2, pp. 155–171, 2000.

[44] I. A. Melnikov, E. G. Kolosova, H. E. Welch, and L. S. Zhitina, “Sea ice biological communities and nutrient dynamics in the Canada Basin of the Arctic Ocean,” *Deep. Res. Part I Oceanogr. Res. Pap.*, vol. 49, no. 9, pp. 1623–1649, 2002.

[45] H. Hop and C. Wiencke, *The Ecosystem of Kongsfjorden, Svalbard*, Springer N. Advances in Polar Ecology, 2019.

[46] J. Wiktor and K. Wojciechowska, “Differences in taxonomic composition of summer phytoplankton in two fjords of West Spitsbergen, Svalbard,” *Polish Polar Res.*, vol. 26, no. 4, pp. 259–268, 2005.

[47] C. Lovejoy, R. Massana, and C. Pedrós-Alió, “Diversity and distribution of marine microbial eukaryotes in the arctic ocean and adjacent seas,” *Appl. Environ. Microbiol.*, vol. 72, no. 5, pp. 3085–3095, 2006.

[48] W. Luo, H. Li, M. Cai, and J. He, “Diversity of microbial eukaryotes in kongsfjorden, Svalbard,” *Hydrobiologia*, vol. 636, no. 1, pp. 233–248, 2009.

[49] R. Stange, “Svalbard: climate and weather.” [Online]. Available: https://www.spitsbergen-svalbard.com/spitsbergen-information/climate.html. [Accessed: 11-Aug-2020].

[50] H. P and H. K, “Phytoplankton, chlorophyll, and submarine light conditions in Kings Bay, Spitsbergen,” *Nor J Bot*, 1973.

[51] A. M. T. Piquet, W. H. Van De Poll, R. J. W. Visser, C. Wiencke, H. Bolhuis, and A. G. J. Buma, “Springtime phytoplankton dynamics in Arctic Krossfjorden and Kongsfjorden (Spitsbergen) as a function of glacier proximity,” *Biogeosciences*, vol. 11, no. 8, pp. 2263–2279, 2014.

[52] H. Hop *et al.*, “Pelagic ecosystem characteristics across the atlantic water boundary current from Rijpfjorden, Svalbard, to the Arctic ocean during summer (2010-2014),” *Front. Mar. Sci.*, vol. 6, no. APR, pp. 1–21, 2019.

[53] E. Trudnowska, A. M. Dąbrowska, R. Boehnke, M. Zajączkowski, and K. Blachowiak-Samolyk, “Particles, protists, and zooplankton in glacier-influenced coastal svalbard waters,” *Estuar. Coast. Shelf Sci.*, p. 106842, 2020.

[54] K. Salonen, P. Salmi, H. Högmander, and J. Keskitalo, “An improved approach to phytoplankton counting based on real-time statistics,” 2015.

[55] J. Wollschläger, A. Nicolaus, K. H. Wiltshire, and K. Metfies, “Assessment of North Sea phytoplankton via molecular sensing: A method evaluation,” *J. Plankton Res.*, vol. 36, no. 3, pp. 695–708, 2014.

[56] R. Danovaro *et al.*, “Implementing and innovating marine monitoring approaches for assessing marine environmental status,” *Front. Mar. Sci.*, vol. 3, no. NOV, pp. 1–25, 2016.

[57] R. Piredda *et al.*, “Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site,” *FEMS Microbiol. Ecol.*, vol. 93, no. 1, pp. 1–14, 2017.

[58] A. N. Blauw, “Monitoring and prediction of phytoplankton dynamics in the North Sea,” 2015.

[59] K. B. Heidelberg, J. A. Gilbert, and I. Joint, “Marine genomics: At the interface of marine microbial ecology and biodiscovery: Minireview,” *Microb. Biotechnol.*, vol. 3, no. 5, pp. 531–543, 2010.

[60] S. J. Bourlat *et al.*, “Genomics in marine monitoring: New opportunities for assessing marine health status,” *Mar. Pollut. Bull.*, vol. 74, no. 1, pp. 19–31, 2013.

[61] J. Pawlowski *et al.*, “Eukaryotic richness in the abyss: Insights from pyrotag sequencing,” *PLoS One*, vol. 6, no. 4, pp. 0–10, 2011.

[62] K. Stachura-Suchoples, N. Enke, C. Schlie, I. Schaub, U. Karsten, and R. Jahn, “Contribution towards a morphological and molecular taxonomic reference library of benthic marine diatoms from two Arctic fjords on Svalbard (Norway),” *Polar Biol.*, vol. 39, no. 11, pp. 1933–1956, 2016.

[63] J. Lund and F. Tallin, “Botanical limnological methods with special reference to the algae,” *Bot. Rev*, 1957.

[64] J. W. G. Lund, C. Kipling, and E. D. Le Cren, “The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting,” *Hydrobiologia*, vol. 11, no. 2, pp. 143–170, 1958.

[65] P. Stofan and G. Grant, “Guidelines for zooplankton sampling in quantitative baseline

and monitoring programs," *Ecol. Res. Ser.*, vol. 83, no. 85, 1978.

[66] "Phytoplankton Measuring and Culture Techniques," 2007.

[67] UNESCO, "Methods for Quantitative," 2010.

[68] HELCOM, "Guidelines for monitoring phytoplankton species composition, abundance and biomass," 2014.

[69] S. F. Rivera *et al.*, "DNA metabarcoding and microscopic analyses of sea turtles biofilms: Complementary to understand turtle behavior," *PLoS One*, vol. 13, no. 4, pp. 1–20, 2018.

[70] A. Venkateshaiah, V. V. T. Padil, M. Nagalakshmaiah, S. Waclawek, M. Černík, and R. S. Varma, "Microscopic techniques for the analysis of micro and nanostructures of biopolymers and their derivatives," *Polymers (Basel).*, vol. 12, no. 3, 2020.

[71] G. McManus and L. Katz, "Molecular and morphological methods for identifying plankton: what makes a successful marriage?," *J. Plankton Res.*, vol. 31, no. 10, pp. 1119–1129, 2009.

[72] T. Stoeck *et al.*, "Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water," *Mol. Ecol.*, vol. 19, no. SUPPL. 1, pp. 21–31, 2010.

[73] E. Coissac, T. Riaz, and N. Puillandre, "Bioinformatic challenges for DNA metabarcoding of plants and animals," *Mol. Ecol.*, vol. 21, no. 8, pp. 1834–1847, 2012.

[74] P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev, "Towards next-generation biodiversity assessment using DNA metabarcoding," *Mol. Ecol.*, vol. 21, no. 8, pp. 2045–2050, 2012.

[75] Y. Ji *et al.*, "Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding," *Ecol. Lett.*, vol. 16, no. 10, pp. 1245–1257, 2013.

[76] J. Decelle, S. Romac, E. Sasaki, F. Not, and F. Mahé, "Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing," *PLoS One*, vol. 9, no. 8, 2014.

[77] M. Solenn, R. Piredda, D. Vaulot, M. Montresor, W. H. C. F. Kooistra, and D. Sarno, "dinoref: A curated dinoflagellate (Dinophyceae) reference database for the 18S rRNA gene," *Mol. Ecol. Resour.*, vol. 18, no. 5, pp. 974–987, 2018.

[78] S. Huo, X. Li, B. Xi, H. Zhang, C. Ma, and Z. He, "Combining morphological and metabarcoding approaches reveals the freshwater eukaryotic phytoplankton community," *Environ. Sci. Eur.*, vol. 32, no. 1, 2020.

[79] Z. B. Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, MilutinoDemsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan Bvic M, Mozina M, Polajnar M, Topla, "Orange: Data Mining Toolbox in Python," *J. Mach. Learn. Res.*, p.

2349−2353., 2013.

[80] A. G. G. Sousa *et al.*, “Diversity and Composition of Pelagic Prokaryotic and Protist Communities in a Thin Arctic Sea-Ice Regime,” *Microb. Ecol.*, vol. 78, no. 2, pp. 388–408, 2019.

[81] J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss, “Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform,” *Appl. Environ. Microbiol.*, vol. 79, no. 17, pp. 5112–5120, 2013.

[82] P. D. Schloss *et al.*, “Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities,” *Appl. Environ. Microbiol.*, vol. 75, no. 23, pp. 7537–7541, 2009.

[83] R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, and R. Knight, “UCHIME improves sensitivity and speed of chimera detection,” *Bioinformatics*, vol. 27, no. 16, pp. 2194–2200, 2011.

[84] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, 1990.

[85] C. Camacho *et al.*, “BLAST+: Architecture and applications,” *BMC Bioinformatics*, vol. 10, pp. 1–9, 2009.

[86] L. Guillou *et al.*, “The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy,” *Nucleic Acids Res.*, vol. 41, no. D1, pp. 597–604, 2013.

[87] S. L. Westcott and P. D. Schloss, “OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units,” *mSphere*, vol. 2, no. 2, pp. 1–11, 2017.

[88] P. J. McMurdie and S. Holmes, “Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data,” *PLoS One*, vol. 8, no. 4, 2013.

[89] H. Wickham, *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York, 2009.

[90] H. Wickham *et al.*, “Welcome to the Tidyverse,” *J. Open Source Softw.*, vol. 4, no. 43, p. 1686, 2019.

[91] “Scale Functions for Visualization.” 2020.

[92] R. D. C. Team, “A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing.

[93] A. J. Oksanen *et al.*, “Vegan: Community Ecology Package.” 2017.

[94] A. Dusa, “Draw Venn Diagrams.” 2020.

[95] A. Ramette, “Multivariate analyses in microbial ecology,” *FEMS Microbiol. Ecol.*, vol. 62,

no. 2, pp. 142–160, 2007.

[96] P. L. Buttigieg and A. Ramette, “A guide to statistical analysis in microbial ecology: A community-focused, living review of multivariate data analyses,” *FEMS Microbiol. Ecol.*, vol. 90, no. 3, pp. 543–550, 2014.

[97] K. R. Clarke and M. Ainsworth, “A method of linking multivariate community structure to environmental variables,” *Mar. Ecol. Prog. Ser.*, vol. 92, no. 3, pp. 205–219, 1993.

[98] C. Leite, “Domain Oriented Biclustering.” Master Degree in Computer Sciences from FCUP, University of Porto, Portugal - concluded in 9th December 2016, 2016.

[99] H. Utermöhl, “Zur Vervollkomnung der quantitativen Phytoplankton-Methodik,” *Ver. ther. angew*, 1958.

[100] “Edler L., M. E. in Microscopic and Molecular Methods for Quantitative Phytoplankton Analysis IOC. Manuals and guides (ed C. Cusack B. Karlson, and E. Bresnan) 13-20 (UNESCO, 2010).,” 1979.

[101] “Edler, Lars, and Malte Elbrächter. ‘The Utermöhl method for quantitative phytoplankton analysis.’ Microscopic and molecular methods for quantitative phytoplankton analysis 110 (2010).”

[102] “Menden-Deuer, S., and Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protest plankton. Limnol. Oceanogr. 45, 569-579.”

[103] D. A. Caron *et al.*, “Defining DNA-based operational taxonomic units for microbial-eukaryote ecology,” *Appl. Environ. Microbiol.*, vol. 75, no. 18, pp. 5797–5808, 2009.

[104] A. Dell’Anno, L. Carugati, C. Corinaldesi, G. Riccioni, and R. Danovaro, “Unveiling the biodiversity of deep-sea nematodes through metabarcoding: Are we ready to bypass the classical taxonomy?,” *PLoS One*, vol. 10, no. 12, pp. 1–18, 2015.

[105] A. Tesi and D. Upf, “Metabarcoding of the 18S rRNA gene to uncover new molecular biodiversity in Metazoa and unicellular Opisthokonta,” 2019.

[106] K. Tapolczai, V. Vasselon, A. Bouchez, C. Stenger-Kovács, J. Padisák, and F. Rimet, “The impact of OTU sequence similarity threshold on diatom-based bioassessment: A case study of the rivers of Mayotte (France, Indian Ocean),” *Ecol. Evol.*, vol. 9, no. 1, pp. 166–179, 2019.

[107] U. von Ammon *et al.*, “Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[108] R. Massana *et al.*, “Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing,” *Environ. Microbiol.*, vol. 17, no. 10, pp. 4035–4049, 2015.

[109] J. M. Maritz *et al.*, “An 18S rRNA Workflow for Characterizing Protists in Sewage, with a

Focus on Zoonotic Trichomonads," *Microb. Ecol.*, vol. 74, no. 4, pp. 923–936, 2017.

[110] K. Deiner, O. Id, A. Lacoursi, O. Id, and I. Reviews, "Environmental DNA metabarcoding: Transforming how we survey animal and plant communities," pp. 0–2.

[111] K. Hadziavdic, K. Lekang, A. Lanzen, I. Jonassen, E. M. Thompson, and C. Troedsson, "Characterization of the 18s rRNA gene for designing universal eukaryote specific primers," *PLoS One*, vol. 9, no. 2, 2014.

[112] S. Geisen, D. Vaulot, F. Mahé, E. Lara, C. de Vargas, and D. Bass, "A user guide to environmental protistology: primers, metabarcoding, sequencing, and analyses," 2019.

[113] S. M. Janosik, "Metabarcoding advances for ecology and biogeography of Neotropical protists: what do we know, where do we go?," *NASPA J.*, vol. 42, no. 4, p. 1, 2005.

[114] F. Mahé *et al.*, "Comparing high-throughput platforms for sequencing the V4 region of SSU-rDNA in environmental microbial eukaryotic diversity surveys," *J. Eukaryot. Microbiol.*, vol. 62, no. 3, pp. 338–345, 2014.

[115] M. Dunthorn, J. Klier, J. Bunge, and T. Stoeck, "Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity," *J. Eukaryot. Microbiol.*, vol. 59, no. 2, pp. 185–187, 2012.

[116] I. A. Luddington, I. Kaczmarska, and C. Lovejoy, "Distance and Character-Based Evaluation of the V4 Region of the 18S rRNA Gene for the Identification of Diatoms (Bacillariophyceae)," *PLoS One*, vol. 7, no. 9, 2012.

[117] F. Mahé *et al.*, "Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests," *Nat. Ecol. Evol.*, vol. 1, no. 4, pp. 1–8, 2017.

[118] F. Zhang, S. Cao, Y. Gao, and J. He, "Distribution and environmental correlations of picoeukaryotes in an arctic fjord (Kongsfjorden, Svalbard) during the summer," *Polar Res.*, vol. 38, pp. 1–11, 2019.

[119] M. Daase *et al.*, "Timing of reproductive events in the marine copepod Calanus glacialis: A pan-Arctic perspective," *Can. J. Fish. Aquat. Sci.*, vol. 70, no. 6, pp. 871–884, 2013.

[120] C. Wang *et al.*, "Spring sea-ice evolution in Rijpfjorden (80° N), Svalbard, from in situ measurements and ice mass-balance buoy (IMB) data," *Ann. Glaciol.*, vol. 54, no. 62, pp. 253–260, 2013.

[121] A. R. Meshram, A. Vader, S. Kristiansen, and T. M. Gabrielsen, "Microbial eukaryotes in an arctic under-ice spring bloom north of Svalbard," *Front. Microbiol.*, vol. 8, no. JUN, 2017.

[122] D. Xu, "Distribution and Diversity of Microbial Eukaryotes in Bathypelagic Waters of the South China Sea," *J. Med. Virol.*, no. Ii, pp. 0–3, 2016.

[123] C. R. Giner *et al.*, "Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean," *ISME J.*, vol. 14, no. 2, pp. 437–449, 2020.

[124] C. Lovejoy, “Changing views of Arctic protists (marine microbial eukaryotes) in a changing Arctic,” *Acta Protozool.*, vol. 53, no. 1, pp. 91–100, 2014.

[125] M. Marquardt, “Marine microbial eukaryotes in Svalbard waters :,” 2016.

[126] J. A. Gilbert *et al.*, “Defining seasonal marine microbial community dynamics,” *ISME J.*, vol. 6, no. 2, pp. 298–308, 2012.

[127] Z. T. Smoła, A. M. Kubiszyn, M. Różańska, A. Tatarek, and J. M. Wiktor, “Protists of arctic sea ice,” *GeoPlanet Earth Planet. Sci.*, no. 9783319717876, pp. 133–146, 2018.

[128] B. Portugal, C. Magalhães, and M. Tomasino, “Exploring the Nitrogen Cycle in the Arctic Ocean,” 2018.

[129] P. Duarte, P. Assmy, H. Hop, G. Spreen, S. Gerland, and S. R. Hudson, “The importance of vertical resolution in sea ice algae production models,” *J. Mar. Syst.*, vol. 145, pp. 69–90, 2015.

[130] L. Meire *et al.*, “Marine-terminating glaciers sustain high productivity in Greenland fjords,” *Glob. Chang. Biol.*, vol. 23, no. 12, pp. 5344–5357, 2017.

[131] R. Massana, J. Del Campo, M. E. Sieracki, S. Audic, and R. Logares, “Exploring the uncultured microeukaryote majority in the oceans: Reevaluation of ribogroups within stramenopiles,” *ISME J.*, vol. 8, no. 4, pp. 854–866, 2014.

[132] A. M. Kubiszyn and J. M. Wiktor, “The Gymnodinium and Gyrodinium (Dinoflagellata: Gymnodiniaceae) of the West Spitsbergen waters (1999–2010): biodiversity and morphological description of unidentified species,” *Polar Biol.*, vol. 39, no. 10, pp. 1739–1747, 2016.

[133] M. Degerlund and H. C. Eilertsen, “Main Species Characteristics of Phytoplankton Spring Blooms in NE Atlantic and Arctic Waters (68-80° N),” *Estuaries and Coasts*, vol. 33, no. 2, pp. 242–269, 2010.

[134] U. John, “R / V Heincke (HE492) Report POLFJORD : Contrasting plankton diversity and interactions with organic matter flux in Arctic fjord systems subject differential climatic forcing mechanisms along the coast of Svalbard,” *Cruise No. HE492 July*, 2017.

[135] “J. Xu, K.; Agatha, S. & Dolan, ‘World Ciliophora Database,’ Worms, 2020. [Online]. Available: http://www.marinespecies.org/aphia.php?p=taxdetails&id=101179.” .

[136] “K. & Swezy, ‘Gyrodinium fusiforme,’ 1921.” .

[137] C. Caroppo, P. Pagliara, F. Azzaro, S. Miserocchi, and M. Azzaro, “Late Summer Phytoplankton Blooms in the Changing Polar Environment of the Kongsfjorden (Svalbard, Arctic),” *Cryptogam. Algol.*, vol. 38, no. 1, pp. 53–72, 2017.

[138] J. T. Bhaskar, B. V. Parli, and S. C. Tripathy, “Spatial and seasonal variations of dinoflagellates and ciliates in the Kongsfjorden, Svalbard,” *Mar. Ecol.*, vol. 41, no. 3, pp. 1–12, 2020.

[139] D. Mora *et al.*, “Morphology and metabarcoding: A test with stream diatoms from Mexico highlights the complementarity of identification methods,” *Freshw. Sci.*, vol. 38, no. 3, pp. 448–464, 2019.

[140] A. Alberdi, O. Aizpurua, M. T. P. Gilbert, and K. Bohmann, “Scrutinizing key steps for reliable metabarcoding of environmental samples,” *Methods Ecol. Evol.*, vol. 9, no. 1, pp. 134–147, 2018.

[141] F. Zhu, R. Massana, F. Not, D. Marie, and D. Vaulot, “Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene,” *FEMS Microbiol. Ecol.*, vol. 52, no. 1, pp. 79–92, 2005.

[142] A. Godhe, M. E. Asplund, K. Härnström, V. Saravanan, A. Tyagi, and I. Karunasagar, “Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR,” *Appl. Environ. Microbiol.*, vol. 74, no. 23, pp. 7174–7182, 2008.

Additional Files

Note: Additional files are the same as supplementary information. They are given as 'figshare private links' to be accessed online through the figshare web service. They can be displayed online; however, the relative abundance of taxa cannot be displayed in percentage. These specific results can be displayed online as relative abundance, ranging between 0-1; or the files can be downloaded, and the values observed in percentage. The following eight 'Additional files' are given as 'figshare private links' (follow the link):

Additional file 1: Distribution of eukaryotic taxa across MOSJ-ICE2016 collection at phylum, class, order, family, genus, specie and OTU levels. Percentage of taxa at given taxonomic level. (figshare private link: https://figshare.com/s/0179ddd1a6c4d2b6c929)

Additional file 2: Description of 18S rRNA libraries from MOSJ-ICE2016 project. (figshare private link: https://figshare.com/s/784e0b56daf49b480fff)

Additional file 3: All information about the comparison of 18S metabarcoding vs microscopic PHT dataset. Identification of common groups (presence); Identification of matching groups within the same samples; Identification of unique groups in both methods. Total percentage of all identified groups in all common samples of both methodologies. (figshare private link: https://figshare.com/s/6f4f591396fc267a969a)

Additional file 4: Taxonomic protists distribution at higher level between microscopy PHT and 18S metabarcoding at Kongsfjorden and Rijpfjorden transect. (figshare private link: https://figshare.com/s/39201117f20f97f50e8c)

Additional file 5: Percentage of matching groups between the same samples at higher and lower level between 18S metabarcoding and microscopy PHT dataset. (figshare private link: https://figshare.com/s/4d272d5aec6f56e1a110)

Additional file 6: All information about the comparison of 18S metabarcoding vs microscopic MIT dataset. Identification of common groups (presence); Identification of matching groups within the same samples; Identification of unique groups in both methods. Total percentage of all identified groups in all common samples of both methodologies. (figshare private link: https://figshare.com/s/a8ec44f044ef8337f8ab)

Additional file 7: Taxonomic protists distribution at higher level between microscopy MIT and 18S metabarcoding at Kongsfjorden and Rijpfjorden transect. (figshare private link: https://figshare.com/s/a62dd98aa0fb507b240b)

Additional file 8: Percentage of matching groups between the same samples at higher and lower level between 18S metabarcoding and microscopy MIT dataset. (figshare private link: (https://figshare.com/s/773b7e43ce0dd192d365)