ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

JACOBS UNIVERSITY

Influence of sea ice on protist community structure in the Central Arctic Ocean

by

Kristin Hardge

A thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

in Biology

Approved Dissertation Committee:
Dr. Katja Metfies[1, 2]
Prof. Dr. Matthias Ullrich[1]
Dr. Mona Hoppenrath[3]
Dr. Eva-Maria Nöthig[2]

[1] Jacobs University Bremen
[2] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven
[3] Senckenberg Research Institute, German Centre for Marine Biodiversity Research, Wilhelmshaven

Date of Defense: 10 August 2016

“In the case of all things which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something besides the parts, there is a cause.”

Aristotle, Metaphysics

Table of Contents

1 Summary

The sea ice covered Arctic Ocean experiences severe and rapid changes as consequences of atmospheric and ocean warming. This is assumed to impact the protist community diversity and composition (i.e. protist community structure) in the sea ice and water column of the Central Arctic Ocean (CAO). Protists are the main food source for several trophic levels and significantly contribute to the overall productivity and functioning of the Arctic ecosystem. Thus, changes in the protist community structure would have further implications for other trophic levels and carbon sequestration. The main objective of this thesis was to elucidate the influence of sea ice change on protist community structure in several habitats of the CAO using molecular methods. In this regard, three aims were set: **i)** investigating the influence of different sequence processing procedures on sequencing data from protist community structure in environmental samples, **ii)** elucidating the impact of sea ice retreat and sea ice origin on protist communities in sea ice and water column of the CAO, and **iii)** analyzing the habitat-specificity of protists and their exchange between several habitats in the CAO.

The fast development of Next-Generation Sequencing (NGS) involves new chances and challenges for surveillance of protists, which were addressed by the **first aim** of this thesis. With NGS, a large number of samples and protist communities, including small-sized (< 3 µm) and rare (< 1%) species, can be analyzed. There are several sequence processing tools and methods available to enhance sequence quality and to obtain reliable estimates of species diversity and composition (i.e. community structure). However, these procedures differ considerably from each other. The results of **Manuscript I** show that the protist community structure obtained with different sequence processing tools and methods is comparable for high taxonomic levels (e.g. protist phyla) and for abundant taxa (i.e. relative sequence abundance of ≥ 1%). However, a comparison with emphasis on diatom genera suggests that low taxonomic levels and particularly rare taxa (i.e. relative sequence abundance of < 1%) are highly affected by different sequence quality-checking, clustering and annotation methods. **Manuscript I** demonstrates that an appropriate sequence quality improvement and a phylogenetic classification of clustered sequences are the basis for a reliable estimate of the protist community structure. In addition, amplicon-based sequencing should always be accompanied by other molecular or conventional methods, such as light microscopy, to enable a critical evaluation of the results.

As appropriate sequence processing tools and methods were chosen, protist communities in the changing Arctic Ocean were investigated. Within the meaning of the **second aim**,

Manuscript II reports on a large-scale comparative study, which analyzed the impact of sea ice retreat and sea ice origin on protists in the sea ice and water column of the CAO. The results show that sea ice concentration and water mass origin were the main drivers for protists in the water column of the CAO in summer 2012. The sea ice community structure was significantly influenced by water mass, but more pronounced by sea ice origin. A comparison of two summer periods with contrasting sea ice concentrations (2011 and 2012) revealed that protist diversity in sea ice and water column was considerably lower during the sea ice minimum year 2012. In particular, the number of rare sea ice algae was much lower in 2012 compared to 2011, most likely due to a decrease in sea ice concentration and a reduction of diverse microhabitats.

The habitat-specificity and exchange of protists in the deep-chlorophyll maximum water, under-ice water, sea ice and melt pond water were investigated in light of the **third aim**. The results of **Manuscript III** show that the sea ice was the habitat with the highest number of unique taxa. These habitat specific taxa were most likely highly-adapted to the unique environmental conditions. The results suggest that sea ice algae were released into the water column during sea ice melt. Therefore, adjacent habitats, such as under-ice water or melt ponds, play an important role in the habitat exchange. Protist occurrences and interactions between the habitats were strongly influenced by sea ice thickness and other environmental variables measured in the habitats (e.g. temperature, salinity, dissolved inorganic nutrients). The number of taxa found in all habitats (i.e. ubiquitous taxa) was considerably higher when new sea ice was formed in September 2012. In contrast, the number was much lower in August when the sea ice was in an advanced state of melt. Thus, the physical processes during sea ice formation most likely led to an increased protist exchange between the habitats. A further decrease in sea ice concentration will likely lead to a diminished protist exchange. This could reduce the species diversity in several habitats of the CAO. In particular, habitat-restricted and highly specialized species, such as sea ice algae, might be more affected by sea ice thinning and decrease than widely distributed taxa.

Overall, this thesis gives valuable insight into the application and suitability of different sequence processing tools and methods and advises caution in the interpretation of results obtained with Next-generation sequencing. Moreover, it presents detailed analyses on both large-scale and local patterns of protist communities in several habitats of the CAO. Therefore, this thesis is an important and substantial contribution to our understanding of the influence of sea ice loss on protist community structure in the CAO.

2 General Introduction

2.1 Global warming

The natural atmospheric greenhouse effect of Earth has changed. Since 1750, increasing atmospheric concentrations of the greenhouse gases carbon dioxide, methane, and nitrous oxide, result in higher absorption of outgoing heat and warm the Earth's surface (Hartmann et al. 2013). Since the late 19th century, global mean surface temperature has increased of about 0.85°C. The period of 1983 to 2012 was the warmest period of the last 800 years (Masson-Delmotte et al. 2013). Warming surface temperatures are also prominent in the World Oceans, particularly the Northern Hemisphere experiences severe climate related changes in the upper 700 m (1971 - 2010) (Rhein et al. 2013). Heat flux of Atlantic Water (AW) through the Fram Strait into the Central Arctic Ocean (CAO) has increased since 1997 as a consequence of higher surface water temperature (0.06°C yr^{-1}, mean temperature 1997 to 2010) and stronger water flow (Schauer et al. 2004, 2008, Polyakov et al. 2005, Beszczynska-Möller et al. 2012). The increased heat flux in combination with atmospheric warming has negatively impacted the sea ice conditions in the CAO (Comiso 2003, Comiso et al. 2008, Stroeve et al. 2007, 2008, 2012). A nearly ice-free summer is expected to occur between 2020 and 2050 (Kirtman et al. 2013, Overland and Wang 2013) with huge ecological consequences for species associated and adapted to sea ice (see review of Post et al. 2013).

2.2 Changes in Arctic sea ice

Since the last decade, extreme sea ice loss in the CAO was observed due to warming of the atmosphere and ocean. The average sea ice extent is decreasing by about 10% per decade (Comiso et al. 2008, Stroeve et al. 2012). At the same time, the ice gets thinner and fresher with increasing coverage of melt ponds. Finally, less thick multi-year ice (MYI, ~1.5 - 3 m) survives the summer melt period and thin first-year ice (FYI, ~0.5 - 1.5 m) is becoming the pre-dominant ice type in the Arctic Ocean (Maslanik et al. 2011, Laxon et al. 2013, Kwok and Rothrock et al. 2009).

In general, sea ice conditions in the Arctic Ocean show strong seasonal and annual variations. With the beginning of polar day at the end of February/beginning of March, sea ice melts and the productive season in the pelagic realm starts with increasing light intensity (Eilertsen 1993, Arrigo at al. 2008, Perrette et al. 2011). The summer minimum sea ice extent

is usually reached in August or September where large open water areas can be observed (Arrigo at al. 2008). Dramatic sea ice losses were observed in September 2007 and 2012 (**Fig. 1**). The last record sea ice minimum extent occurred in September 2012, with extreme and abrupt ice loss in the Beaufort, Chukchi, and East Siberian Seas (NSIDC; http://nsidc.org). Due to sea ice thinning in the CAO and strong southward geostrophic winds, an increased volume of sea ice is exported through the Fram Strait out of the CAO (Smedsrud et al. 2011, Halvorsen et al. 2015, Krumpen et al. 2015).

In autumn, the formation of new ice occurs when ocean surface temperatures drop below the freezing point of seawater (-1.8°C and a salinity of ~ 34 psu) (Golden et al. 1998, Krembs et al. 2011, reviewed by Arrigo 2014). Sea ice extent and concentration are highest during the dark winter period from the end of October until the end of February. Climate-related changes become more and more apparent during the Arctic winter. The first record seasonal minimum sea ice extent since the winter of 1979 was reached on March 24, 2016 (NSIDC; http://nsidc.org) where large Arctic shelf areas were ice-free.

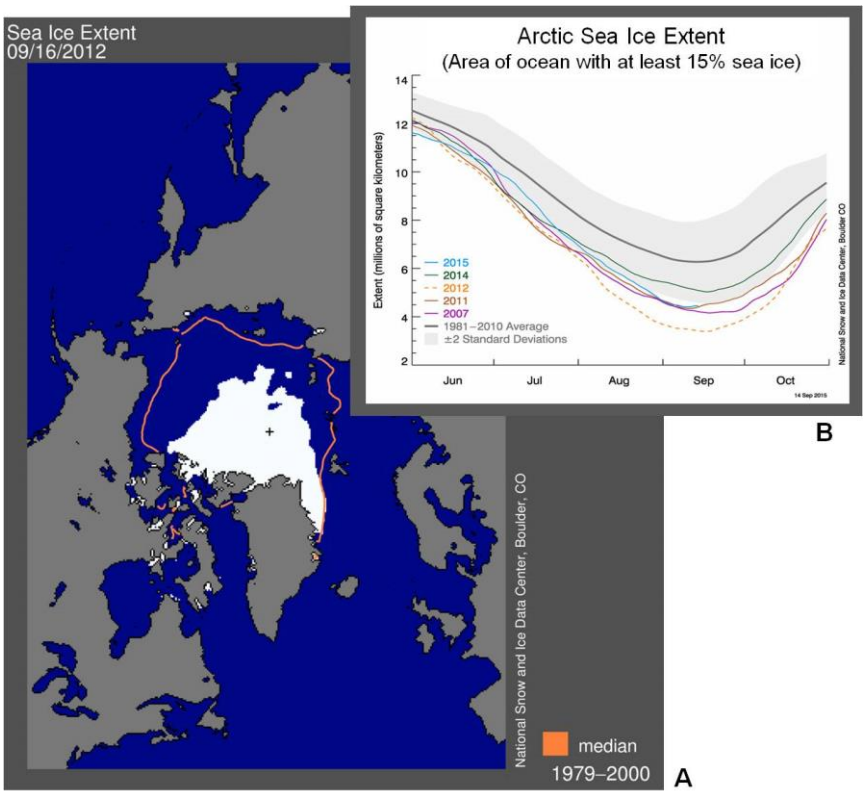


Fig. 1. (**A**) Record minimum Arctic sea ice extent (3.41 million km²) during summer 2012 (shown is September 16, 2012). The orange line shows median extent (1979 to 2000) for that day. (**B**) Sea ice extent for June to October in 2007 (second minimum), 2011 (third minimum), 2012 (record minimum) and 2014. The gray area around the average line (1981 - 2010) shows standard deviation range of the data. (A and B: credit: National Snow and Ice Data Center).

The sea ice of the CAO is unique in physico-chemical conditions and an important habitat for unicellular eukaryotes, fish, crustaceen, and higher trophic levels, such as seals and polar bears. Sea ice decline and thus, habitat loss is assumed to strongly affect these highly adapted Arctic species in their biodiversity (i.e. variety and variability of species including genetic diversity) and ecology. In particular, the various sea ice habitats, the water column, under-ice water, sea ice and melt ponds, (**Fig. 2**) are inhabited by unicellular eukaryotes (protists), which form the base of the marine and benthic food web.

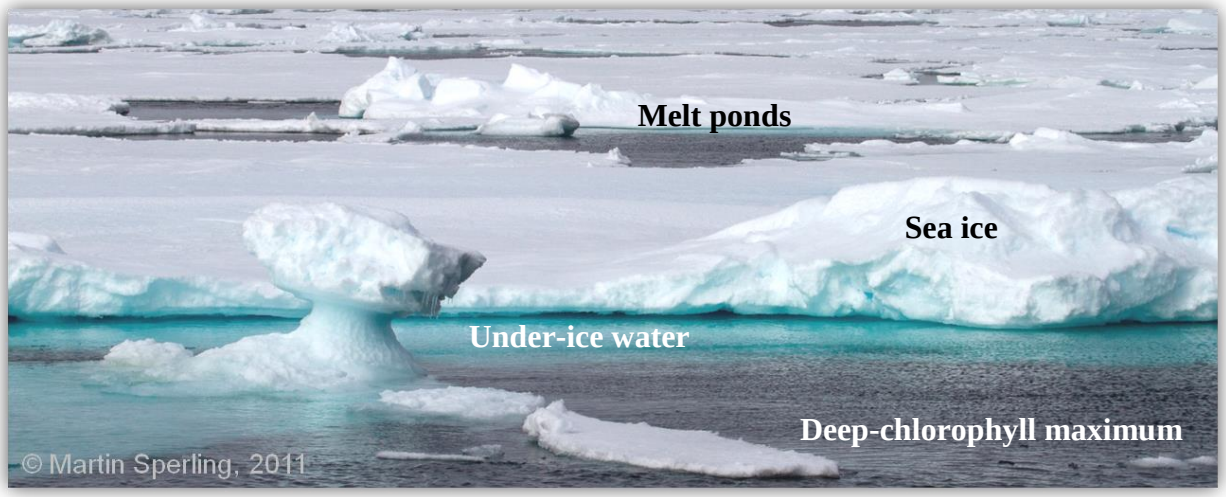


Fig. 2. The major types of habitats of the Arctic Ocean. The deep-chlorophyll maximum depth varies between 10 m and 50 m, whereas the under-ice water is located directly under the sea ice (0 m and 2 m). The sea ice can reach a thickness between a few tens of millimeters and several meters (Arrigo 2014). The skeletal layer and brine drainage system are highly complex and formed during sea ice formation and melt. Surface melt ponds are mostly freshwater or brackish ponds on top of the sea ice, however, some ponds are also connected to the underlying water column (open ponds). All habitats, particularly the sea ice-associated habitats, are highly variable in their physico-chemical conditions and consist of several microhabitats, which are inhabited by complex protist communities.

2.3 Protist biodiversity in the Arctic Ocean

Protists, or unicellular eukaryotes, form a major part of the Arctic ecosystem and are of great importance for diversity and productivity within several water and ice habitats. Autotrophic species, such as phytoplankton or sea ice algae (e.g. diatoms, flagellates) contribute significantly to primary production and are therefore of pivotal importance for energy fluxes and carbon sequestration in the Arctic Ocean (Gosselin et al. 1997, Quillfeldt et al. 2000, Gradinger 2009, Perrette et al. 2011, Boetius et al. 2013, Fernández-Méndez et al. 2015). Phytoplankton and ice algae are grazed by heterotrophic zooplankton (e.g. flagellates, dinoflagellates and ciliates). In addition, algae aggregates and zooplankton fecal pellets carry incorporated carbon into the Arctic deep-sea where they serve as food supply for benthic communities and are preserved in sediments (Bathmann et al. 1990, Bauerfeind et al. 1994, Peinert et al. 2001).

Among the most diverse plankton organisms are Bacillariophyceae (referred here as diatoms, **Fig. 3**) which belong to the supergroup Stramenopila. They are a major group in the water column and the sea ice (reviewed by Arrigo 2014) where they serve as an important food source for heterotrophic flagellates or crustacean. Their size range from 5 to 500 µm and some species can even exceed 1 mm in length. The frustule (cell wall) of diatoms is made of silica which protects diatoms from grazers (Passow 1991, Assmy et al. 2013). If physico-chemical conditions in the pelagic realm become suitable during spring or summer, some diatom species (e.g. *Chaetoceros spp.*, *Thalassiosira spp.* and *Fragilariopsis spp.*, **Fig. 3**) can produce blooms in the open water, ice edge or under sea ice (Sakshaug et al. 1989, Eilertsen 1993, von Quillfeldt 2000, Perrette et al. 2011, Arrigo et al. 2012). Driving factors which trigger bloom situations are for example an enhanced water stratification or nutrients and light availability. During and after the blooms, aggregates of chain-forming diatoms (e.g. *Fragilariopsis, Melosira, Thalassiosira,* **Fig. 3**) or heavy silicified diatoms (e.g. *Fragilariopsis, Coscinodiscus, Melosira*) have a high sinking velocity and contribute significantly to carbon export to deep-sea floor in the Fram Strait and Arctic Ocean (Bauerfeind et al. 2005, 2009, Boetius et al. 2013).

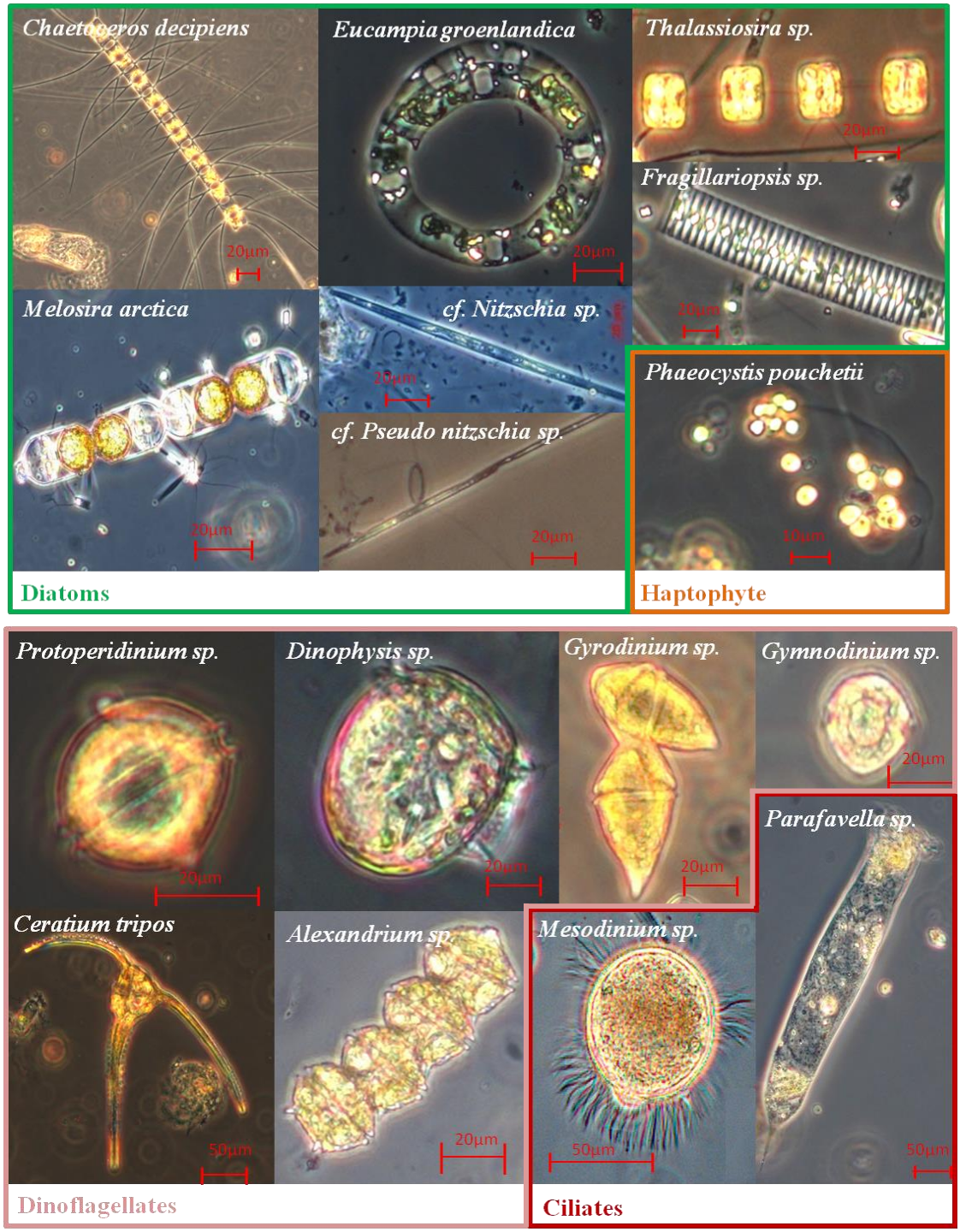


Fig. 3. Most representative species of unicellular micro-eukaryotes found in net-tow samples (surface waters, 20 µm mesh size) across Fram Strait. Diatoms (green-framed) and haptophytes (orange-framed) are autotrophic, dinoflagellates (rose-framed) are mainly mixotrophic and ciliates (red-framed) are mixotrophic (*Mesodinium* sp.) and heterotrophic (*Parafavella* sp.). Pictures were taken with inverted light microscopy during *Polarstern* expedition Ark 26/1 to the Fram Strait in summer 2011.

The supergroup Haptophyta (**Fig. 3**) includes autotrophic flagellates that belong mainly to pico- and nanoplankton with cell sizes ranging from 3 - 20 µm. Some haptophytes (e.g. *Emiliania huxleyi*, *Coccolithus pelagicus*) form calcified scales (coccoliths) which make them even visible from space when the species produce massive blooms. Therefore, coccolithophorids are important carbonate contributors and act as a significant source of calcite carbon on a regional scale (Bramlette 1958, Brown and Yoder 1994). *Phaeocystis*, for example, is a single celled or gelatinous colony forming genus with a cosmopolitan distribution, particularly in warm North Atlantic water in the Fram Strait and Arctic Ocean (Reigstadt and Wassmann 2007, Metfies et al. 2016). In contrast to coccolithophorids, the contribution of *Phaeocystis* (**Fig. 3**) to vertical particulate organic carbon export is rather small (Reigstadt and Wassmann 2007, Wolf et al. 2016). Therefore, these probably non-sinking species serve as important food source for mesozooplankton in the upper ocean (Rousseau et al. 2000).

Chlorophyta are green algae that include multi-cellular and single-cell taxa, commonly found in marine and freshwater phytoplankton assemblages (reviewed by Caron et al. 2012). The smallest eukaryotic marine species (less than 3 µm) can be found in the chlorophyte class Mamiellophyceae (e.g. *Ostreococcus*). The cosmopolitan genus *Micromonas* (Mamiello-phyceae) is an important part of the abundant biosphere (abundance of ≥ 1%) in the Arctic Ocean (Lovejoy et al. 2007, Foulon et al. 2008, Kilias et al. 2014a). More recently, *Micromonas pusilla* was observed in the Arctic halocline of the Nansen and Amundsen Basin where it was mainly associated with AW (Metfies et al. 2016). However, *M. pusilla* includes several genetic lineages or clades which possess a high adaptability for different environments (Foulon et al. 2008, Worden et al. 2009).

Among the supergroup Alveolata, dinoflagellates are possibly the most prominent protozooplankton in the Arctic pelagic realm with sizes ranging between 3 and 200 µm. Heterotrophic dinoflagellates (e.g. *Akashiwo, Prorocentrum, Protoperidinium,* **Fig. 3**) possess the largest cells and are important grazers of phytoplankton and smaller flagellates. Mixotrophic dinoflagellates, such as *Gymnodinium* **(Fig. 3**) and *Karlodinium*, combine phagotrophy (i.e. incorporation of cells) and phototrophy in one cell (Lee et al. 2014, Mitra et al. 2016). Depending on the environmental conditions (in-/organic substrates, light availability, presence of food organisms), mixotrophic species can switch between these modes of nutrition by the regulation of gene expression (reviewed by Matantseva et al. 2013).

Other representatives of the supergroup Alveolata are Ciliophora (**Fig. 3**); mainly heterotrophic, but also mixotrophic species which play an essential role as grazers of small flagellates and diatoms (Posch et al. 2015). Growth rates of ciliates are closely related to abundance, productivity and distribution of phytoplankton communities (Veritiy 1985, Jensen and Hansen, 2000). For example, the mixotrophic ciliates *Mesodinium* (**Fig. 3**) ingest cryptophytes (Gustafson et al. 2000), while tintinnids (e.g. *Acanthostomella*, *Parafavella*, **Fig. 3**, *and Tintinnopsis*) graze for example on *Phaeocystis pouchetii* (Haptophyta, **Fig. 3**) as observed during spring blooms in the Dutch Wadden Sea and the coastal North Sea (Admiraal and Venekamp 1986).

2.4 Protist biodiversity in the changing Arctic Ocean

As response to climate change, biological ranges of species have shifted along three axes: spatial, temporal and in their own physiology (reviewed by Bellard et al. 2012). More precisely, species migrate to more suitable habitats at a local or micro-habitat level (i.e. spatial shift), respond in their life cycle and phenology to the changing environmental conditions (i.e. temporal shift) or adapt to new environmental conditions in their local habitat (alteration of physiological tolerance) (reviewed by Bellard et al. 2012).

In the Arctic Ocean, sea ice decline and rising water temperatures have severe consequences for the biological ranges of sea ice and pelagic protists. Melting of sea ice and warming of sea surface temperatures could trigger the development of true Atlantic species, such as *E. huxleyi*, which is moving northwards via the warm AW masses (Hegseth and Sundfjord 2008, Iida et al. 2012). Early sea ice retreat during summer lead to a shift in the timing and extent of phytoplankton blooms, which most likely changes phytoplankton productivity and influences food web structure and downward matter fluxes (Wassmann and Slagstad 1993, Booth and Horner 1997, Carmack and Wassmann 2006, Hegseth and Sundfjord 2008, Wassmann and Reigstad 2011, Fernández-Méndez et al. 2015). In addition, early sea ice retreat leads to a late freeze-up, because absorption is increased in open water (Perovich and Polashenski 2012, Stroeve et al. 2014). These temporal changes could impact the abundance and community structure (i.e. composition and diversity) of the sea ice and pelagic biota because the winter community is supposed to initiate following spring ice algae bloom (Niemi et al. 2011). Temporal changes in ice-freeze up could also involve spatial

changes, however, nothing is known regarding the influences of regions of sea ice origin on protist abundances or community structure.

Increased melting of sea ice, glacial ice and thawing of permafrost lead to a fresher Arctic Ocean (Serreze et al. 2006). When the Arctic Ocean gets warmer and fresher, a shift towards smaller plankton species (pico- and nanoplankton) is assumed (Li et al. 2009, Morán et al. 2010, Ardyna et al. 2011, Hilligsøe et al. 2011, **Contribution I**). Furthermore, freshwater species are probably introduced into the Arctic Ocean via increased river runoff (Cooper et al. 2008, Yamamoto-Kawai et al. 2011). Freshwater species such as *Chlamydomonas* and *Ochromonas* might also gain higher abundances and importance in melt ponds (Kilias et al. 2014b) if melt pond coverage increases. This could affect the productivity of the CAO, because melt pond and sea ice algae contribute significantly to primary productivity in the Arctic Ocean (Fernández-Méndez et al. 2015). In addition, intensive exchange between melt ponds, sea ice and water column during melting and freezing process could lead to a gradual reorganization of protist communities in these habitats. Despite extensive research, it is still unclear how far the protist biodiversity and composition in the Arctic Ocean will change and whether ice algae and phytoplankton productivity (Fernández-Méndez et al. 2015) as well as species diversity will increase or decrease with ocean warming and sea ice thinning. Therefore, **this thesis** investigated the influences of sea ice on protist community structure in several habitats of the CAO.

2.5 Morphological and molecular methods

2.5.1 Light microscopy

Several taxonomists pointed out the advantages of using molecular techniques but claimed that we still need to collect morphological data (Wiens et al. 2004, Smith and Turner 2005). Recognizable protists can be identified and enumerated using an inverted microscope (Zeiss IM 35) with phase contrast. The phase contrast is used for recognition of cell morphology, as many phytoplankton species are partially transparent and hence are difficult to see under a light microscope. Light microscopy has the advantage over molecular methods that it provides additional, highly valuable ecological information not accessible by molecular methods, such as sequencing. For example, the amount of plant tissue, detritus and zooplankton fecal pellets can indicate the conditions during sampling or the trophic status of the sampled environment. However, using light microscopy, only protists with cell sizes of at least 5 µm are identifiable

because smaller cells lack morphological features. Therefore, for the identification of nano- and picoplankton species other methods, such as scanning electron microscopy or molecular methods, are necessary.

2.5.2 Automated ribosomal intergenic spacer analysis (ARISA)

ARISA is a molecular approach based on the length-heterogeneity of the intergenic spacer region (ITS). The ITS regions are located between the 18S and 5.8S (ITS 1) or 28S (ITS 2) rRNA genes (**Fig. 4**), which encode for the small and large subunit rRNA genes. ITS-fragments are amplified with fluorescent primers which can be visualized on an electropherogram or an agarose gel. The fragment lengths were first assumed to be species specific (Baldwin 1992). However, one or several fragments can correspond to one or several species, making identifying single species impossible (Bent et al. 2007, Caron et al. 2012). Therefore, ARISA can be used to gain an overall picture (a "fingerprint") of the microbial community. It is a cost-effective and a rapid method that can be applied to a high number of samples. Several studies approved the suitability of ARISA for the assessment of protist community patterns and their relationship to environmental parameters (Fechner et al. 2010, Bienhold et al. 2012, Wolf et al. 2013, Gobet et al. 2014, Kilias et al. 2015). ARISA is optimal to be used in conjunction with sequencing techniques, which allow a more accurate identification of the species (Fisher and Triplett 1999, Wolf et al. 2013).

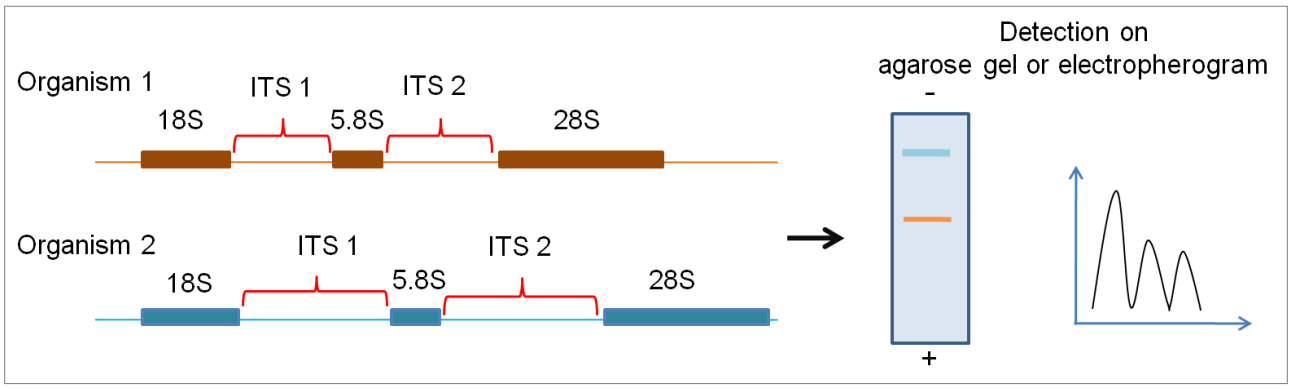


Fig. 4. Simplified illustration of ARISA analysis of different fragment lengths of the ITS regions.

2.5.3 Next-generation sequencing

With the start of the Human Genome Project in 2004 (International Human Genome Sequencing Consortium 2004), next-generation sequencing (NGS) has developed after the standard capillary sequencing methods (i.e. Sanger-Sequencing, Sanger et al. 1977).

The two sequencing techniques, 454 pyrosequencing (Margulies et al. 2005) and Illumina sequencing (MiSeq technology), share basic sequencing steps (**Fig. 5**). In the first step, single stranded library fragments are produced from the original isolated DNA. This sequencing library is done by specific synthetic DNAs (adapters), which bind on the target DNA sequence; in this thesis the V4 hypervariable region of the 18S rRNA gene of eukaryotes. The library fragments are then amplified in emulsify beads (454 pyrosequencing) or attached on a solid surface (Illumina sequencing) by a polymerase-mediated reaction. In a next step, DNA nucleotides are added to the single stranded library fragments and a complementary DNA strand is formed. Every time when a complement nucleotide is integrated, light signals are produced by luciferase (454 pyrosequencing) or digital images are produced with fluorescently labeled nucleotides (Illumina sequencing). These steps are repeated for thousand of sequences simultaneously, why these techniques are also called 'high-throughput' or 'massive parallel' sequencing techniques.

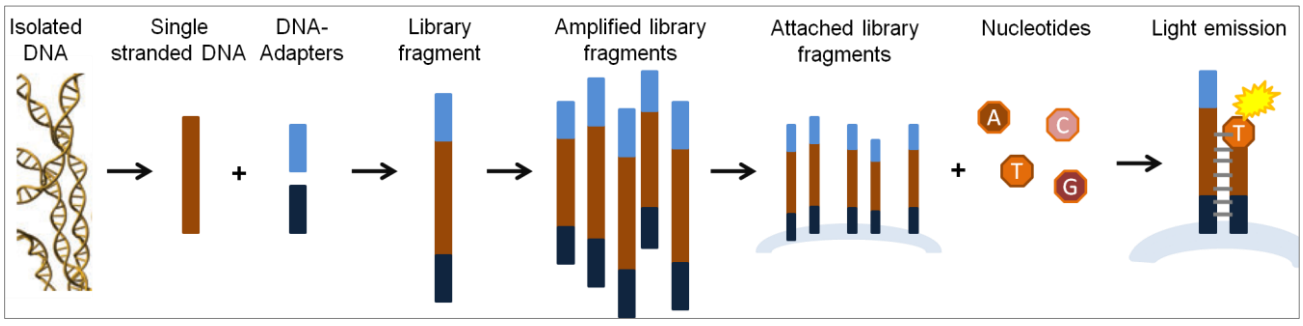


Fig. 5. Simplified illustration of the Next-Generation sequencing methods 454 pyrosequencing and Illumina sequencing.

The sequencing output generated by 454 pyrosequencing and Illumina differs in read length and read quality. 454 pyrosequencing reads are usually around 670 bp long. Whereas Illumina sequence reads are shorter (2x 250 bp). Sequencing errors accumulate toward the distal end, therefore much lower error rates can be found in the first 250 bp (Huse et al. 2007, Schloss et al. 2011). Thus, read quality is higher for Illumina. Furthermore, 454 pyrosequencing is more expensive and time intensive, while Illumina sequencing has relatively low costs and a faster sequencing procedure. Since summer 2015, 454 pyrosequencing is no longer available and was replaced by Illumina. The transition from 454 pyrosequencing to Illumina sequencing platforms involved parameter adjustments for sequence processing but the obtained community structures are largely comparable across the two technologies (Kozich et al. 2013, Mahé et al. 2014).

For data analysis, sequences are processed with bioinformatic tools, for example the open-source software packages QIIME (Caporaso et al. 2010), mothur (Schloss et al. 2009), DNAStar (SeqMan Pro, Lasergene, USA) or PhyloAssigner (Vergin et al. 2013). QIIME and mothur provide SOPs (standard operating procedures) which include recommended parameter settings investigated based on comparative studies and experiences of developers. The first step in sequence processing is always the improvement of sequence quality, so that a reliable estimate of protist community composition and diversity can be obtained. It includes the deletion of primer mismatches, ambiguous bases and homopolymers (repeats of single bases) (**Fig. 6**). Sequences are also trimmed because PCR-induced (e.g. chimeras) and sequencing errors (e.g. homopolymers) are more likely to occur at the distal end of longer sequences (Huse et al. 2007, Schloss et al. 2011). The deletion of PCR-chimeras (hybrid-sequences) is very important because chimeras can lead to an overestimation of species diversity (Behnke et al. 2011). Beyond that, an average quality score can be set so that only high-quality reads are retained. A recommended method for enhancing sequence quality of 454 pyrosequences is denoising, i.e. the removing of sequencing noise. Undenoised 454-datasets tend to overestimate the diversity, since a substantial part of the rare biosphere is assumed to be an artifact of sequencing errors (Reeder and Knight 2010, Quince et al. 2011, Schloss et al. 2011, Bachy et al. 2013, Egge et al. 2013). However, in some cases, the denoising step can alter the sequencing output and therefore the diversity estimates (Gaspar et al. 2013, Majaneva et al. 2015). With the improvement of sequence technology, denoising is not an issue for the Illumina technique MiSeq, because the technique is more robust to homopolymers and thus, has a lower error rate.

After sequence quality-improvement, sequences can be clustered into operational taxonomic units (OTUs) at a predefined similarity threshold (Huse et al. 2010). A similarity threshold of 97% is considered to be adequate to estimate the original eukaryotic diversity (Behnke et al. 2011, Zimmermann et al. 2011). There are several clustering methods available, which differ in their performance and resulting cluster structure. Hence, diversity estimates can differ considerably between the methods (Bonder et al. 2012, Sun et al. 2012, Schloss et al. 2013, Chen et al. 2013). Clustered or unclustered sequences are then assigned to reference taxonomy. The taxonomic assignment can be based on sequence-similarity with curated reference databases, such as SILVA (Pruesse et al. 2007) or Greengenes (DeSantis et al. 2006) or on the evolutionary information in a phylogenetic tree (Matsen et al. 2010, Vergin et al. 2013). The choice of sequence annotation method can significantly influence the interpretation of the community composition (Westcott and Schloss 2015).

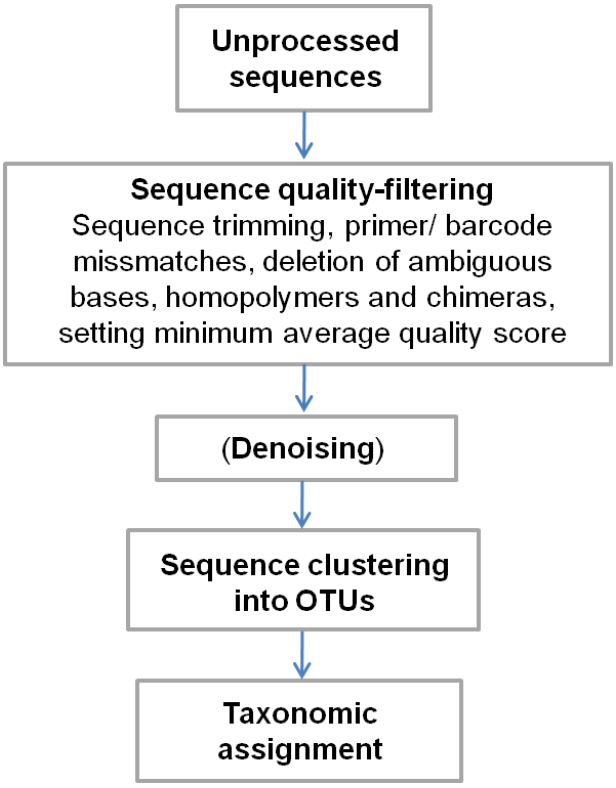


Fig. 6. Schematic diagram of a possible sequence processing pipeline.

3 Aims and Outline

3.1 Aims

Protists form a major part of the Arctic ecosystem and are of great importance for biodiversity and productivity of CAO. The sea ice is considered as one of the main factors that influence protist communities. Current changes in sea ice conditions are assumed to alter protist community biodiversity and composition in ice-influenced habitats with further implications for the overall productivity of the CAO and carbon sequestration to the deep sea. Protist taxa that are adapted to habitat-specific environmental conditions or available food sources might be affected most by the recent sea ice retreat. However, little is known about how drastic sea ice loss might affect biodiversity and composition of sea ice biota and phytoplankton. So far, in-depth knowledge about protist community structure and protist ecology in variable habitats of the CAO is still scarce and mainly based on the analysis of conventional approaches (e.g. light microscopy). Therefore, this thesis analyzed 18S amplicon sequence data in a total of 97 samples collected from the deep-chlorophyll maximum water depth (DCM), under-ice water (UIW), sea ice (ICE) and melt pond water (MW). Samples were collected in the CAO during August and September in 2011 and 2012 to investigate the environmental factors driving protist community structure and the potential consequences of sea ice retreat on protists in the CAO.

The following questions will be addressed:

1. How to assess protists in environmental samples? What are the influences of different sequence processing procedures on the resulting protist diversity estimates obtained from environmental samples?

2. How does sea ice retreat and sea ice origin impact protist communities in the sea ice and water column of the CAO?

3. How habitat-specific are Arctic protist communities and is there any protist exchange between sea ice-influenced habitats in the CAO?

3.2 Outline

To elucidate the influences of climate related changes on protist community structure, it is a fundamental task to find an appropriate method to identify protist taxa. Capturing protist diversity in environmental samples is still a challenging task, particularly due to the fast development of sequencing techniques. Deep-sequencing technologies such as 454 pyrosequencing and Illumina have revolutionized the analysis and interpretation of the unseen diversity of protist organisms. However, careful processing of sequences is crucial for accurate taxonomic characterization of protists. In **Manuscript I,** we analyzed 18S rRNA gene sequences from water samples with 454 pyrosequencing and processed sequences with three open-source software packages: QIIME, mothur and a custom-made pipeline which includes DNAStar and PhyloAssigner. To examine the influences of different sequence processing procedures on resulting protist diversity estimates from environmental samples, different sequence preprocessing (i.e. quality trimming), clustering and annotation methods were applied and the resulting composition and diversity estimates of protist phyla and diatom genera were compared. Diatom diversity was also assessed with light microscopy and used as reference for sequencing. Sequence preprocessing and clustering had a major impact on the resulting protist community structure at a low taxonomic level (e.g. diatom genera) and rare taxa (abundance < 1%). Therefore, the study demonstrates that the choice of sequence processing method is an important step to obtain a reliable estimate of microbial diversity.

The second aim of this study was to elucidate the impacts of sea ice retreat and sea ice origin on protist communities in sea ice and water column of the CAO. **Manuscript II** reports on a comprehensive large-scale study that investigated protist community structure in sea ice and water samples collected in the CAO during two summer periods with contrastive sea ice concentrations (2011 and 2012, the last recorded sea ice minimum year). The analysis is based on a combination of the molecular fingerprinting method ARISA and sequencing of the V4 region of the 18S rRNA gene. For sequence processing, we followed the recommendations given in **Manuscript I**. Possible correlations of environmental parameters, such as water mass, sea ice concentration and sea ice origin, with the water and sea ice community were assessed. We show for the first time that the regions of ice floe origin play a fundamental role in structuring the sea ice community. The water and sea ice communities differed in diversity and composition between the years and were less diverse in 2012, possibly as a consequence of sea ice melting. A reduction in protist diversity was especially true for the sea ice community in 2012, which suggests that sea ice algae might be more vulnerable to climate change.

Manuscript III focuses on the protist community structure and exchange between several habitats which experienced different sea ice conditions. The sampling of DCM, UIW, ICE and MW at several ice-stations provided the unique opportunity to analyze protist communities in a cross-section of the CAO during record sea ice minimum year 2012. Habitat-specific taxa found in one habitat only (unique taxa) and taxa shared between at least two habitats or found in all habitats (ubiquitous taxa) were discussed in an ecological context. The highest abundance and diversity of unique taxa was found in sea ice (mainly pennate diatoms) and the importance of this habitat in the CAO was outlined. Ubiquitous taxa were identified and considered as ecological specialists as they always showed clear habitat preferences. Differences in the degree of habitat exchange during sea ice melt and formation were investigated. As the highest protist exchange was observed when new sea ice was formed, we concluded that a sea ice decrease could lead to an overall lower species exchange, which might reduce the species diversity in all habitats of the CAO. In particular, a substantial part of the sea ice protist community could be endangered because specialized species restricted to sea ice may not be able to react to rapid environmental changes.

4 List of Manuscripts

Manuscript I

***Kristin Hardge**, Stefan Neuhaus, Estelle S. Kilias, Christian Wolf, Katja Metfies & Stephan Frickenhaus*

Impact of Sequence Processing and Taxonomic Classification Approaches on Eukaryotic Community Structure from Environmental Samples with Emphasis on Diatoms, under review in *Molecular Ecology Resources*

Manuscript II

***Kristin Hardge**, Ilka Peeken, Stefan Neuhaus, Thomas Krumpen, Thorsten Stoeck & Katja Metfies*

Sea Ice Origin and Sea Ice Retreat are Major Drivers of Variability in Arctic Marine Protist Composition, under review in *Marine Ecology Progress Series*

Manuscript III

***Kristin Hardge**, Ilka Peeken, Stefan Neuhaus, Benjamin Lange, Alexandra Stock, Thorsten Stoeck, Lea Weinisch & Katja Metfies*

The importance of sea ice for exchange of habitat-specific protist communities in the Central Arctic Ocean, *Journal of Marine Systems*, in press, doi: 10.1016/j.jmarsys.2016.10.004

5 Statement of Contribution

Manuscript I

The concept was planned together with Stefan Neuhaus, Estelle S. Kilias, Christian Wolf and Katja Metfies and performed by Stefan Neuhaus and myself. Stefan Neuhaus and I analyzed the data. Stephan Frickenhaus significantly contributed to the discussion and review process. I wrote the manuscript together with Stefan Neuhaus. Both authors contributed equally to the manuscript.

Manuscript II

The concept was planned together with Ilka Peeken and Katja Metfies. Laboratory work was carried out by Lea Weinisch (Uni Kaiserslautern, DNA isolation) and me (DNA isolation, ARISA-PCR and sequencing-PCR). Data analysis was carried out by Stefan Neuhaus (sequence processing), Thomas Krumpen (sea ice drift data) and me (ARISA and sequence analysis). I wrote the manuscript.

Manuscript III

The concept was planned together with Ilka Peeken, Thorsten Stoeck and Katja Metfies. Laboratory work was carried out by Alexandra Stock, Lea Weinisch (DNA isolation, sequencing-PCR) and me (DNA isolation, sequencing-PCR). Data analysis was carried out by Stefan Neuhaus (sequence processing), Ben Lange (sea ice physics) and myself (sequence analysis). I wrote the manuscript.

6 Further Contributions

Contribution I

Nöthig E.-M., Bracher A., Engel A., Metfies K., Niehoff B., Peeken I., Bauerfeind E., Cherkasheva A., Gäbler-Schwarz S., **Hardge K.**, Kilias E., Kraft A., Mebrahtom Kidane Y., Lalande C., Piontek J., Thomisch K. & Wurst M. 2015. Summertime plankton ecology in Fram Strait - a compilation of long- and short-term observations. *Polar Research* 34, article no. 23349, doi: http://dx.doi.org/ 10.3402/polar.v34.23349

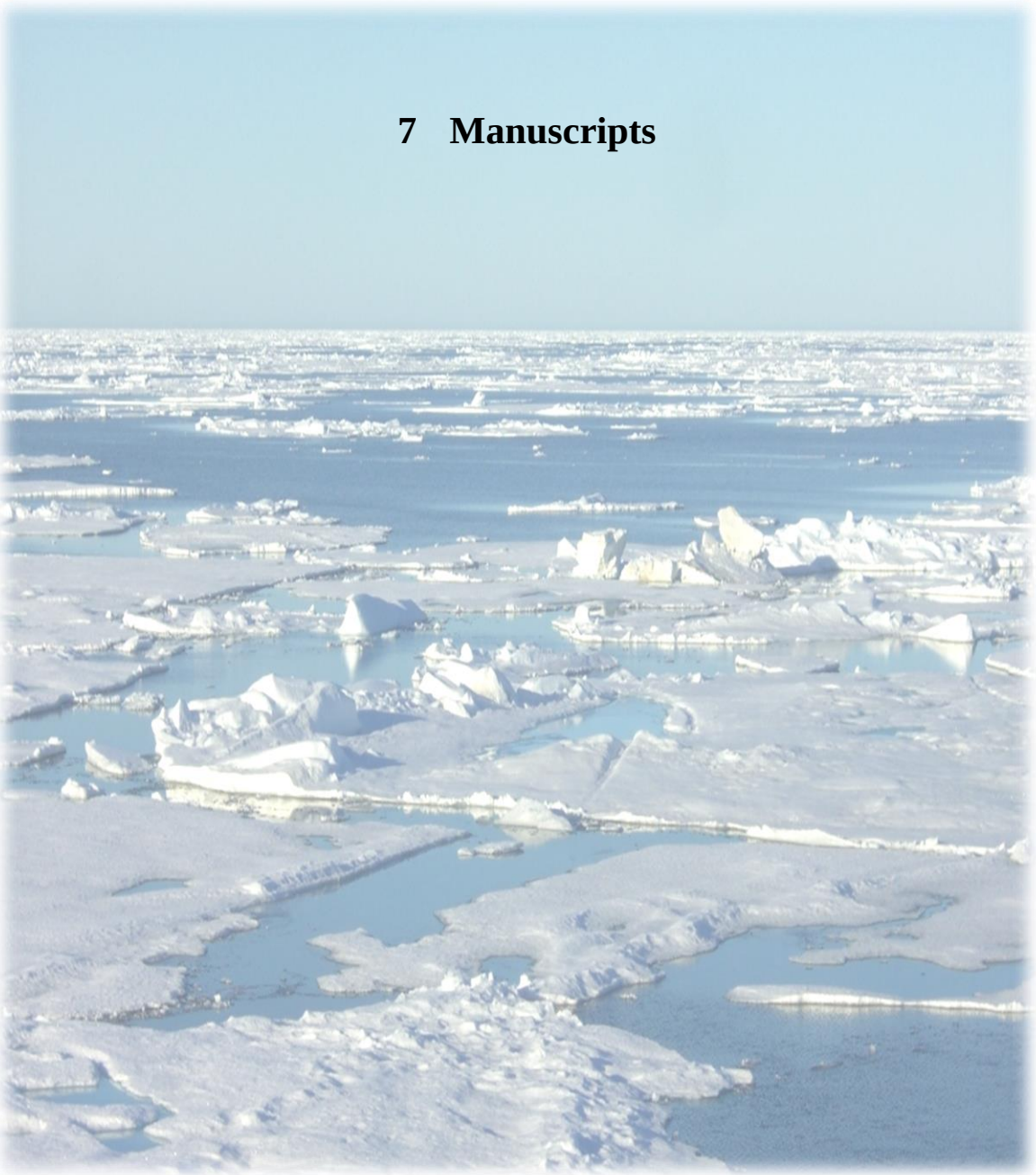
Contribution II

Lalande C., Nöthig E.-M., Bauerfeind E., **Hardge K.**, Besczynszka-Moller A. & Fahl K. 2016. Lateral supply and downward export of particulate matter from upper waters to the seafloor in the deep eastern Fram Strait, accepted in *Deep-Sea Research Part I*

Contribution III

David C., Flores H., **Hardge K.**, Kohlbach D., Lange B., Metfies K., Niehoff B., Peeken I., Sea ice retreat and nutrient limitation go along with a higher degree of planktonic heterotrophy in Arctic post bloom ecosystems, in preparation

7 Manuscripts



7.1 Manuscript I

Impact of sequence processing and taxonomic classification approaches on eukaryotic community structure from environmental samples with emphasis on diatoms[1]

Running title: Assessment of protist communities

K. Hardge[1,2¶,*], S. Neuhaus[1,¶], E.S. Kilias[1,#,a], C. Wolf[1], K. Metfies[1,2] and S. Frickenhaus[1,3]

[1]Department of Bioscience, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany

[2] Jacobs University Bremen, Bremen, Germany

[3] Hochschule Bremerhaven, Bremerhaven, Germany

[#a] Current Address: Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands

* * Corresponding author: kristin.hardge@awi.de

[¶] These authors contributed equally to this work.

Keywords: Next-generation sequencing, amplicon metagenomics, biodiversity, protists, diatoms

Abstract

Next-generation sequencing is a common method for analyzing microbial community diversity and composition. Configuring an appropriate sequence processing strategy within the variety of tools and methods is a non-trivial task and can considerably influence the resulting community characteristics. We analyzed the V4 region of 18S rRNA gene sequences of marine samples by 454-pyrosequencing and generated several data sets with QIIME, mothur and a custom-made pipeline based on DNAStar and the phylogenetic-tree based PhyloAssigner under a variety of processing parameters. Our results revealed strong differences in total number of operational taxonomic units (OTUs), indicating that sequence preprocessing and clustering had a major impact on protist diversity estimates. However, diversity estimates of the abundant biosphere (abundance of $\geq 1\%$) were reproducible for all conducted processing pipeline versions. A qualitative comparison of diatom genera reveals differences between between pipelines using only-sequence-characteristics- (mothur and QIIME) and DNAStar/PhyloAssigner whereas the latter came close to light microscopy-based diatom identification. We conclude that diversity studies using different sequence processing strategies are comparable if the focus is put on higher taxonomic levels, and if abundance thresholds are used to filter out OTUs from the rare biosphere. Because most diversity studies aim at the characterization of microbial communities realistically at the genus level, we recommend that results obtained from a standardized analysis workflow of 18S rRNA sequences should be still be accompanied by conventional methods such as microscopy.

Introduction

Amplicon-based metagenomic analysis via next-generation sequencing (NGS) enables rapid and in-depth analysis of microbial communities in environmental samples, including protists. Deep-sequencing technologies such as 454-pyrosequencing and Illumina have revolutionized the analysis and interpretation of the unseen parts of the community in several studies (e.g. Stoeck et al. 2010, Wolf et al. 2013, Kilias et al. 2014). A characterization of communities is influenced by several working steps, for example the sampling procedure, DNA isolation, choice of primer set, PCR-program, sequencing technique, and sequence processing. In general, the latter includes quality filtering, OTU-clustering of sequences according to a similarity threshold and finally taxonomic assignment (Schloss et al. 2011, Quince et al. 2011 For example, sequence quality has to be controlled, appropriate sequence clustering and annotation method as well as reference data set have to be chosen with regard to the scientific question. In particular, for protist community analyses, an adapted processing parameter set has to be evaluated, since software configurations are specifically validated for 16S analyses. Therefore, a careful processing of sequences is crucial for the accurate taxonomic characterization of protists.

Several studies demonstrated that if sequences were not quality-checked (e.g. removing primer mismatches, homopolymers, ambiguous bases) or denoised (removing sequencing noise), the species diversity can be considerably overestimated, in particular the rare biosphere (i.e. species with relative abundance < 1%) (Schloss et al. 2011, Huse et al. 2007, Quince et al. 2009, Behnke et al. 2011, Bonder et al. 2012, Bachy et al. 2013). However, it was also shown that under certain circumstances, true rare species were not detectable anymore after denoising. In consequence, diversity estimates and qualitative composition were affected (Gaspar et al. 2013). Sequencing errors were also reduced during pre-clustering of sequences at a defined similarity threshold (Behnke et al. 2011, Kunin et al, 2010). The pre-clustering and subsequent sequence clustering into operational taxonomic units (OTUs) reduce the computational time for downstream sequence analysis, such as the sequence annotation. In addition, diversity estimates are more reliable if sequences are clustered into OTUs based on predefined sequence similarity threshold (Huse et al. 2010), particularly in the case of insufficient reference databases. However, the currently available OTU picking strategies perform with variable results and the discussion about the best strategy is ongoing (Bonder et al. 2012, Sun et al. 2012, Schloss et al. 2013, Chen et al. 2013, Schmidt et al. 2015).

Besides the clustering method, the choice of sequence annotation method can significantly influence the interpretation of the community composition. The taxonomic assignment of sequences can be based on the similarity features that query sequences share with reference sequences, e.g. RDP classifier (Wang et al. 2007), QIIME's UCLUST-based consensus classifier (Edgar et al. 2010) or based on the evolutionary information in a phylogenetic tree, e.g. PhyloAssigner (Matsen et al. 2010). All approaches have in common that the unknown sequences are compared against databases of known and regularly curated reference sequences like SILVA (Pruesse et al. 2007) or Greengenes (DeSantis et al. 2006). However, most of the mentioned studies were performed on prokaryotes using mock communities (i.e. artificial communities) while studies on eukaryotes are rare (Majaneva et al. 2015). Furthermore, diversity estimates from eukaryotic communities in environmental samples might be affected differently than those of mock communities, as they will most probably contain a higher and in particular unknown genetic diversity. Less is known about the comparability of protist diversity estimates in environmental samples analyzed with different processing strategies (Bachy et al. 2013, Morgan et al. 2013, Majaneva et al. 2015). A recent study conducted on environmental communities showed that different 18S amplicon read processing considerably affect diversity estimates and taxonomic composition analysis (Majaneva et al. 2015). The study analyzed the V7, V8 and V9 region of the 18S rRNA gene and revealed that the V4 region is more suitable to analyze protist diversity as it provides higher sequence variability (Nickrent et al. 1991, Stoeck et al. 2010, Behnke et al. 2011, Zimmermann et al. 2011, Luddington et al. 2012, Kermarrec et al. 2013, 2014). Because of this higher complexity, pyrosequencing errors rates are higher for V4 fragments than for V9 fragments (Behnke et al. 2011). Consequently, the diversity estimates might be more affected by the sequence processing using the V4 gene region.

Here, we sequenced the V4 region of the 18S rRNA gene of three water samples containing complex eukaryotic communities with different taxonomic compositions and investigated whether diversity estimates of protists are still comparable if different sequence processing procedures were applied. The samples were part of a study by Kilias et al. (2013), who examined different protist community compositions in the Fram Strait, influenced by water masses and sea ice concentrations. In addition to the protist diversity in general, we were especially interested in diatoms (Bacillariophyceae) as they were highly abundant in the samples and their composition gave valuable information in relation to the environmental conditions of the sampling area (Quillfeldt et al. 2004, Kilias et al. 2013). Furthermore, we used light microscopy as reference method for 454-pyrosequencing to estimate the diatom

diversity and compared the results with the results obtained by NGS. We systematically tested several configurations of sequence analysis pipelines built on three open-source software packages which are widely-used or established in-house: QIIME (Caporaso et al. 2010) mothur (Schloss et al. 2009) and a combination of DNAStar (SeqMan Pro, Lasergene, USA) and PhyloAssigner (Vergin et al. 2013). The pipeline versions differed in the following main sequence-processing steps: quality filtering of sequences, sequence clustering, and taxonomic classification. Their configurations followed several recommendations or standard operating procedures (SOPs/ tutorials) (for QIIME and mothur), or, for DNAStar/PhyloAssigner, an experience- and literature-based parameter set.

Material and methods

Collection of samples

Environmental water samples were taken during *RV Polarstern* cruise (ARK-XXV/2) to the Fram Strait along a transect navigated from 11°58.362′ to 11°5.09′ E longitude at ~78°50′ N latitude in July 2010 **(S1 Figure)**. The water samples (T1, T3 and T9) were collected with Niskin bottles (12 L) attached to a CTD (conductivity, temperature, depth) rosette from the chlorophyll maximum depth (20 – 50 m). For light microscopy, 200 ml of the collected seawater was filled in brown glass bottles, spiked with 5 ml formalin (20%) and stored dark and refrigerated. For sequencing, 2 L of the same collected seawater was filtrated through Isopore Membran Filters (Millipore, Billerica, MA, USA) with pore sizes of 10, 3, and 0.4 µm, respectively, to ensure collection of all protist cell sizes. Filters were stored in Eppendorf tubes at -80°C until further processing (see Kilias et al. 2013 for details).

DNA isolation and PCR amplification

DNA extraction and amplification were conducted according to Kilias et al. (2013): DNA was extracted from filters using E.Z.N.A TM SP Plant DNA Kit Dry Specimen Protocol (Omega Bio-Tek, Norcross, GA, USA), according to the manufacturer's protocol. For subsequent sequencing, the V4 region was amplified using the primer set 528F (GCG GTA ATT CCA GCT CCA A), and 1055R (ACG GCC ATG CAC CAC CAC CCA T) (modified after Elwood et al. 1985). The forward primer 528F attaches approx. 25 bp upstream of the V4 region which has a length range of approx. 230 - 500 bp (Nickrent and Sargent 1991). PCR

products were purified with the Mini Elute PCR Purification Kit (QIAgen, Hamburg, Germany) and pooled in equal volumes for sequencing on a 454 GS FLX sequencer (Roche, Germany, XLR70 Titanium, 800 flows).

Processing the sequencing data

Sequences were processed within four software packages: QIIME (version 1.8.0, Caporaso et al. 2010), mothur (version 1.29.2, Schloss et al. 2009) and DNAStar (SeqMan Pro, Lasergene, USA) in conjunction with PhyloAssigner (Vergin et al. 2013). An overview of the conducted pipeline versions including tested parameters is given in **Table 1** and **Supplement Text S1**. We compared the diversity estimates of six pipeline versions produced with two software packages and a software combination: denoising and not-denoising with mothur and QIIME (including different classifier); OTU-clustering and not-OTU-clustering with DNAStar and PhyloAssigner (**Table 1**).

We processed the sequences with recommended parameter settings for the analysis of 454-pyrosequencing data. For QIIME and mothur, we used the SOPs provided on the respective websites (see Data Accessibility). These SOPs include sequence quality filtering and trimming, chimera detection and denoising prior to the sequence clustering (**Table 1**). Because denoising was shown to alter the processing output spuriously (Gaspar and Thomas 2013), we compared the diversity estimates of not-denoising but clustering pipeline versions processed by QIIME (Q_ud) and mothur (m_ud) additionally to denoising pipeline versions (Q_d and m_d).

Flowgrams in Q_d were denoised using DeNoiser (Reeder and Knight 2010), an implementation of PyroNoise (Quince et al. 2011) within the QIIME package. Subsequently, the QIIME pipeline usearch.qf (Usearch version 5.2.236, Edgar 2010), which incorporates UCHIME (Edgar et al. 2011), was used for chimeric sequence detection and OTU clustering in Q_d and Q_ud. Chimeric sequences were detected with reference-data (Silva 111 SSU Ref NR) and without reference-data (i.e. *de novo*, from the sequence data alone) and removed from the query set if considered as chimeric sequence by both methods. Remaining sequences were clustered and OTUs were determined *de novo* (i.e. without reference sequence which could serve as cluster seeds) at a minimum similarity threshold of 97%. Representative sequences were annotated with QIIME default classifier UCLUST consensus (version 1.2.22q) (Edgar et al. 2010). For sequence annotation, we used the QIIME-prepared 97% clustered version of the Silva SSU Rev NR 111 as reference database. In addition, we

compared the default classifier with two other classifiers: RDP-classifier (version 2.2, Wang et al 2007) at a confidence level of 0.8 and BLASTN algorithm (version 2.2.22, Altschul et al. 1990). This comparison was done, because the sequence annotation method possibly influences the estimates of community composition and diversity of the same preprocessed data set.

Flowgrams in m_d were denoised using the mothur implementation of the PyroNoise algorithm (Quince et al. 2011). PCR-chimeras were detected and removed by applying UCHIME (Edgar et al. 2011) in *de novo* mode in m_d and m_ud. For distance matrix calculation for clustering in mothur, pairwise distances larger than 0.15 were discarded. Sequences were clustered into OTUs at a distance level of 0.03 on basis of the average neighbor clustering. Representative sequences were annotated in mothur using mothur's implementation of the RDP classifier 2.2 (Wang et al 2007) (referred here as "mothur" classifier) at a 0.8 confidence level and the QIIME-prepared 97% clustered version of the Silva SSU Rev NR 111 as reference database.

The quality-checked and clustering DNAStar/PhyloAssigner pipeline version (P_qcc) was specifically assembled for 18S amplicon analyses and well-established in other studies (Kilias et al. 2013, 2014, Wolf et al. 2013). In addition to the clustering DNAStar/PhyloAssigner pipeline version, we processed a quality-checked but not-clustering pipeline version (P_qc) to examine a possible effect of clustering on the diversity estimates. Denoising of pyrosequences was not part of these pipelines. For sequence annotation in PhyloAssigner, the QIIME-prepared reference database was thinned down manually with the goal to harmonize qualities of selected sequences while reducing redundancies in the structure of the phylogenetic tree, i.e. optimizing the tree to a structurally and taxonomically representative set of 4000 leafs. Phylogenetic placement of sequences in a fixed rooted phylogenetic backbone tree was done in pplacer (Matsen et al. 2010) within PhyloAssigner. A more detailed description of all sequence processing pipelines can be found in **S1 Text.**

D*ata analysis and interpretation*

We assessed the sequence characteristics after quality filtering, namely sequence length, number of chimera and number of processed sequences. For comparison of eukaryotic diversity estimates, we assessed the number of OTUs classified as eukaryotic phyla and diatom genera. For subsequent interpretation of the resulting community structures, different relative sequence abundance thresholds were applied on each sample separately, subdividing the protist community into an abundant (≥ 1% and ≥ 5%) and rare biosphere (< 1%).

To qualitatively inspect the dissimilarities of taxonomic compositions obtained from the different pipeline versions, a dendrogram was derived from hierarchical clustering (unweighted pair-group method with arithmetic mean, UPGMA) based on Jaccard-distances including the sample structure (T1, T3, T9) across all pipeline versions. Calculations were performed using PAST software (Hammer et al. 2001).

Light microscopy of diatoms

To obtain diatom identification and abundances for comparison with sequencing data, we used inverted light microscopy to estimate diatom diversity in the water samples. Diatom genera were identified and enumerated by means of the Utermöhl method (Utermöhl 1958) using Zeiss IM 35 with phase contrast. Therefore, 50 ml sampling volume was filled in an Utermöhl chamber (3 ml) and settled down for at least three hours. At least 400 individuals were counted in half of the chamber at a magnification of 160. Cell counting was performed on two replicates per sample and a minimum of 400 cells were counted. Counts were reported in individuals per liter (Ind/L) and relative proportions. Taxonomical identification was mainly based on Hoppenrath et al. (2009), Kraberg et al. (2010) and on the open-access database Plankton*Net (hdl: 10013/de.awi.planktonnet).

Results

Sequence characteristics after quality filtering

The raw sequences (i.e. before processing) of all samples had a mean length of 610 bp with a length varying between 80 - 1193 bp (**Table 2**). After quality filtering, sequences of Q_ud and Q_d had a mean length of 520 bp and 526 bp. In contrast, filtered sequences of the mothur pipeline versions were shorter (mean length of 278 bp on m_ud and m_d) than sequences of the QIIME pipelines. The longest sequences were achieved with the PhyloAssigner pipeline versions with a maximum length of 607 bp and a mean length of 430 bp.

With the de-novo mode of UCHIME in mothur, we detected a total of 491 unique chimeric sequences in m_d and 499 in m_ud (**Table 2**). The combination of *de novo* and reference based chimera detection applied in QIIME, revealed a considerably higher number of chimeras in Q_ud (1347 chimeras) than in Q_d (95 chimeras). In the DNAStar/PhyloAssigner pipeline versions, a total of 108 sequences were found to be

chimeric with the reference-based method. A direct comparison of these numbers is not possible, since the QIIME, mothur and DNAStar/PhyloAssigner pipeline versions applied chimera detection at different sequence preclustered data sets (QIIME at 97% sequence similarity, mothur: ~ 99% sequence similarity and PhyloAssigner no sequence preclustering).

For all samples, the lowest number of sequences entering the OTU-cluster step was observed in Q_d (between 8276 and 11110 seqs.), whereas the highest number of sequences was observed in P_qc (between 15759 and 26783 seqs.) (**Table 3**). The number of sequences differed strongly between Q_ud and Q_d but was similar in m_ud and m_d (**Table 3**). Thus, sequence preprocessing and denoising had a stronger influence on the final sequence number in the QIIME pipelines than in the mothur pipelines.

Estimates of eukaryotic diversity

Number of OTUs

All pipelines found the highest number of OTUs in the water sample T9 and the lowest in water sample T1. However, the number of OTUs differed between the used pipelines and partially between denoised and undenoised data sets (**Table 3**). Applying the USEARCH cluster algorithm in the QIIME pipelines, we obtained a slightly higher number of OTUs in Q_d (maximum of 146 OTUs) compared to Q_ud (maximum of 137 OTUs), in which a minimum cluster size cut-off value of four was applied (**Table 3**). In comparison to QIIME, the number of OTUs was considerably higher in m_ud (maximum of 253 OTUs) and m_d (maximum of 244), in which no cluster size cutoff was used at all, according to the SOP. The number of OTUs was even higher in the DNAStar/PhyloAssigner pipeline (maximum of 1296 OTUs). Overall, the numbers of abundant OTUs (relative sequence abundance of ≥ 1%) were comparable between all pipeline versions (**Table 3**). In contrast, the numbers of rare OTUs (relative sequence abundance of < 1%) were of high variability across the three samples: QIIME showed roughly two times higher OTU numbers compared to mothur, while for DNAStar/PhyloAssigner a factor of ~12 up to ~20 was observed (**Table 3**).

Eukaryotic phyla

All pipeline versions were able to assign between 98% and 100% of the processed sequences to eukaryotic phyla (**Fig 1a**), with only a minor fraction unclassified. The protist composition in T1 was dominated by sequences of Metazoa (38 - 57%), comprising mainly of Crustaceae, and Stramenopila (34 - 45%), comprising mainly of Bacillariophyta (**Fig 1a**). The remaining sequences were classified as Alveolata (2 - 4%), and Haptophyta (4 - 6%). The water samples T3 and T9 were characterized by similar protist compositions (**Fig 1a**). In these, the highest proportion of sequences was classified as Mamiellophyceae within the phylum of Chlorophyta (66 - 75% in T3, 45 - 55% in T9). The second most abundant group was Alveolata (15 - 23% in T3 and 29 - 36% in T9), whereas Haptophyta (6 - 10%) and Stramenopila (1 - 9%) accounted for smaller proportions. Sequences belonging to Metazoa were found in low abundance (relative sequence abundance < 1%) in T3 and showed a relative proportion below 5% in T9.

Diatom genera

Comparison of sequence-based and light microscopy analyses

With inverted light microscopy, we identified six to seven diatom genera per sample, and in total nine different genera. Using the default parameters, we assigned two to four diatom genera with QIIME, two to six genera with mothur and five to 21 diatom genera with the DNAStar/PhyloAssigner pipeline (**Table 4**). The number of OTUs assigned to diatoms varied strongly between the pipeline versions and the highest number was always achieved with DNAStar/PhyloAssigner (**S1 Table, S2 Figure**). The highest relative proportion of diatom sequences and OTU number was observed in T1. In this sample, we also recorded the largest discrepancy of diatom diversity between all pipelines. Whereas only 15 to 17 OTUs were listed in two to four diatom genera using QIIME, 35 to 38 OTUs were listed in six diatom genera using mothur. With the sequence assignment in the clustered DNAStar/PhyloAssigner pipeline, we achieved up to 16 diatom genera corresponding to 512 OTUs in T1.

In the results of all sequence analysis pipelines, the genus *Thalassiosira* dominated the diatom assemblage (47 - 97%, depending on the sample and pipeline) (**Fig 1b**) and was represented by at most 13 OTUs (in T1, m_ud). With the DNAStar/PhyloAssigner pipeline, a large proportion of reads was assigned only to the class-level (*Mediophyceae*) (P_qc). The second most abundant genus in the molecular approach was *Porosira*, in particular in samples

T1 and T3 and more pronounced in the QIIME pipeline versions (up to 16%). Sequences belonging to *Chaetoceros* were also found, but in very low numbers and a maximum abundance of 5% by the Q_d pipeline using UCLUST consensus classifier.

In contrast to the molecular approach, the most abundant genus obtained with light microscopy was *Chaetoceros* spp. (mainly small cells of *C. cf. socialis*) which reached 82% of the total diatom assemblage in T1 (184.43 x 10^3 Ind/L) and with 89% in T3 (113.08 x 10^3 Ind/L) (**Fig 2**). *Thalassiosira* spp. (mainly *T. nordenskioldii*) reached much lower proportions in the microscopy examination, and *Porosira* spp. was rarely observed. Another difference between the molecular and morphological approach is obvious in the diatom composition of water sample T9 (17.04 x 10^3 Ind/L), which mainly consisted of pennate diatoms, such as *Fragilariopsis* spp. (34%, mainly *F. cylindrus)* and *Nitzschia* spp. (~17%), when analyzed by microscopy. However, we could not observe *Fragilariopsis* spp. with the molecular method and P_qc detected *Nitzschia* spp. only in T1 and T9. *Navicula* spp. was present in all microscope samples and detected by all analyses pipeline versions, but not in sample T3. In T3, only P_qc could detect *Navicula* spp. In addition, we encountered a low cell number of *Eucampia zodiacus* in T1 and T3 by microscopy, but this genus was not detected by any of the sequence analysis pipelines at all.

Comparison of classifier in QIIME

Except for some taxa additionally found with BLASTN, an effect of different classifier (UCLUST, rdp, BLASTN) on the protist phyla composition was nearly absent (**Fig 1a**). However, at the diatom genus level, 16 - 22% of the total sequences were not classified in Q_ud using QIIME's UCLUST consensus classifier (**Fig 1b**). In this case, we observed only two diatom genera (*Chaetoceros* and *Thalassiosira*) (**Table 4, S1 Table**), while we detected two more genera if sequences were denoised (T1 and T3). For example, no *Porosira* was detected by the UCLUST consensus classifier in the Q_ud pipeline but by all other pipeline versions. Furthermore, in the water sample T9, some of the sequences were assigned to *Minidiscus* if the BLASTN algorithm was used (Q_d), whereas it was absent in all other QIIME pipelines. With this classifier we detected more diatom genera compared to the other classifier (RDP or UCLUST).

Sample cluster analysis

The sample cluster analysis revealed a grouping in software tools and - within these tools - in samples (**Fig 3**). The water sample T1 clustered separately from T3 and T9 indicating its unique composition of eukaryotic taxa, whereas the composition in T3 and T9 was more similar to each other. Hence, sample-dependent separation was superseded by pipeline-dependence. For T3 and T9, the DNAStar/PhyloAssigner pipelines formed separate clusters near the mothur pipelines' results. The application of denoising or not-denoising caused only minor distances in the mothur pipelines as m_d clustered close to m_ud. In contrast to this, the separation of Q_d and Q_ud was more pronounced.

Discussion

In our study, we compared different sequence processing strategies on the basis of the community composition and diversity estimates of protists in environmental samples. All used pipelines and parameter settings were previously established in publications, mainly on prokaryote communities. Though, in the eukaryotic domain, we observe that the "human factor" in terms of pipeline and parameter selection has major impacts on the conclusions based on the pipeline outcomes. A large set of sequence processing versions was conducted here and some results and conclusions are stable while others are hardly reproducible across the workflows. All generated datasets revealed similar community compositions at the phylum level of eukaryotes but differed strongly at the genus level of diatoms. In addition, the abundant biosphere was very similar for all conducted pipelines, while the number of rare OTUs turned out rather variable.

Comparison of sequencing method and light microscopy

We used light microscopy as reference for the molecular approach and compared the methods qualitatively in terms of occurrences of diatom genera. A quantitative comparison is not possible because relative proportions of sequences cannot be related to relative proportions of individuals in the samples, since the number of target gene copies (i.e. 18S rRNA) may vary strongly between species (Egge et al. 2013, Prokopowich et al. 2003, Godhe et al. 2008, Zhu et al. 2005). In our study, this issue is exemplified by large abundance discrepancies of *Chaetoceros* and *Thalassiosira* between the morphological and molecular approach. With sequencing, *Thalassiosira* dominated the diatom assemblage, whereas it was *Chaetoceros* if

we used light microscopy. As observed with light microscopy, *Thalassiosira* outplayed *Chaetoceros* by cell size and cell condition and possessed presumably more DNA material.

Estimating diatom diversity

Since we have analyzed environmental samples instead of a mock community, we are confined to evaluate the eukaryotic community structure achieved with light microscopy and different sequence processing methods by taking into account the systematic errors of these methods.

The observed diversity achieved with microscopy was more likely an underestimation of the real diversity, because rare species could have been missed due to the small counting volume. This diversity estimate was partly comparable to the diatom diversity estimate achieved with sequencing. In comparison to microscopy, the number of classified diatom genera was generally lower, if sequences were processed in QIIME and mothur, but much higher in the PhyloAssigner pipelines. Thus, we assume that the QIIME and mothur pipeline versions tend to underestimate the diatom diversity, whereas the PhyloAssigner pipeline versions tend to overestimate the diatom diversity. The exceptional high number of genera identified by PhyloAssigner might be the consequence of a less stringent sequence quality filtering in contrast to the quality filtering applied in QIIME and mothur in addition to the different sequence clustering and OTU annotation methods (see **S2 Text** for detailed discussion**)**. However, in contrast to the similarity-based sequence annotation methods used by QIIME and mothur, a phylogenetic placement can unveil close phylogenetic relationships of taxa and therefore improve the capability to depict the complex diatom diversity (Vergin et al. 2013).

The discrepancies between the molecular and morphological method could also be related to possible visual misidentifications of diatoms. Morphological misidentifications cannot be excluded in our study, since identification of some diatom species is only possible if other methods such as scanning electron microscopy are used (Hoppenrath et al. 2007). One example is the centric diatom genus *Attheya* which was not observed by light microscopy but detected by mothur and PhyloAssigner in T1 and T9. The morphological valve structure of some *Attheya* species is very similar to those of *Chaetoceros.* This identification issue is underlined by the fact that the unaccepted synonym of *Attheya septentrionales* (Østrup) R.M.Crawford is *Chaetoceros septentrionales* Østrup, 1895 (Guiry 2015a). The same was true for *Entomoneis alata* (Ehrenberg) Ehrenberg, which was identified by microscopy in T1

and T3 but not classified with pyrosequencing, except of sequence classification using UCLUST classifier in QIIME. The old and not accepted name of this species is *Amphiprora alata* (Ehrenberg) Kützing, 1844, (Guiry 2015b) which corresponding genus was found by PhyloAssigner alone. Taking these misidentification and synonymy into account, the best agreement with light microscopy was most likely obtained with the PhyloAssigner pipelines.

Several studies demonstrated the suitability of the highly variable V4 region of the 18S rRNA and a similarity threshold of 97% for assessing diatom diversity in environmental samples (Zimmermann et al. 2011, Luddington et al. 2012, Kermarrec et al. 2012, 2014). However, this threshold could be too low to distinguish some diatom genera (Luddington et al. 2012). There is strong evidence that distinct taxonomic levels cannot be reflected by one single defined similarity threshold because this can result in over- or underestimation of diversity (Reeder and Knight 2009, Behnke et al. 2011, Schloss and Westcott 2011), depending on the local branching characteristics of the regions within the phylogenetic tree. Therefore, Schloss and Westcott (2011) recommend a calculation of OTUs for different distance thresholds ranging between 0.00 and 0.10 instead of using a single threshold. Estimation of community composition up to genus level or beyond might be more reliable, if single sequences are classified instead of generating sequence clusters, but only if no species are missing in the reference database. If sequences were not clustered with DNAStar (P_qc), additional diatom genera were observed (*Nitzschia*, *Coscinodiscus* and *Discotella*). These genera have been disregarded in the other pipeline versions due to one of the following reasons: i) their sequences were assigned to a cluster which representative sequence belonged to a closely-related genus, ii) they belonged to the rare biosphere and thus, no sequences or not enough sequences passed the quality filtering, denoising or cluster size filtering to form an OTU.

A further issue that might have influenced the diatom diversity estimates is the selection and quality of the reference database, since bad alignment quality (relevant for P_qcc and P_qc) as well as erroneous classification of reference sequences (relevant for all pipelines) can lead to misidentification of species. Missing species in the reference was an apparent issue as well. For example the diatom genus *Eucampia* was identified via the light microscope but not via sequence classification because it was not present in the QIIME-formatted 97% clustered version of the Silva SSU Rev NR 111 reference database.

Impact of taxonomic assignment on community structure

On eukaryotic phyla level, the usage of the BLASTN (Altschul et al. 1990) classifier within Q_d and Q_ud resulted in a higher number of additional eukaryotic phyla not observed with the other classifiers. The taxonomic assignment with BLASTN is based on the "best hit" by means of sequence similarity and could misclassify sequences, particularly if no well-sampled and curated reference database is present. The results of taxonomic assignment by BLASTN best-hit annotation can therefore differ from results obtained from other assignment methods and probably have a higher proportion of misclassification. Other methods have the ability to assign at lower taxonomic levels (e.g. UCLUST consensus classifier and rdp), if the reference database does not contain very similar sequences with unique annotation. For example UCLUST did not detect *Navicula* (Q_d, Q_ud) and *Porosira* (Q_ud). Thus, the UCLUST method might have performed too conservative in these cases and assigned OTUs to a more basal taxonomic group.

One alternative to sequence classification based on only-sequence-similarity is a classification based on phylogenetic relationships. In the PhyloAssigner pipelines, sequences are placed onto a fixed phylogenetic reference tree and the taxonomic label of the last common ancestor (LCA) node at a defined level of maximum uncertainty was assigned (Vergin et al. 2013). Sequences originating from unknown species, i.e. not represented in the reference database, can at least be assigned to a more basal taxonomic group (i.e. phylogenetic LCA).

For most pipelines, *Thalassiosira* was a considerable part of the diatom composition. However, in P_qc the read annotation often only reached the diatom class level Mediophyceae. This might be due to undetected chimeras or to the high diversity and complex taxonomy of *Thalassiosira*. At least 27 species of *Thalassiosira* were found to co-occur in the North Sea (Hoppenrath et al. 2007). It is a paraphyletic genus with species occurring among several groups of the phylogenetic tree, such as the genera *Detonula, Cyclotella, Minidiscus* and *Skeletonema* (Hoppenrath et al. 2007). On the other hand, only with the PhyloAssigner pipeline versions, we were able to detect *Minidiscus* and *Skeletonema* reliably. Therefore, it is possible that these *Thalassiosira*-related genera were not detected by pipelines based on the OTU clustering methods used within the QIIME and mothur pipelines, because the respective sequences were within the 97% similarity radius of a *Thalassiosira*-dominated cluster; or were purely represented to pass a stringent quality-filtering (Huse et al. 2010, Kunin et al. 2010), inclusive denoising and cluster-size cutoff as recommended by the SOPs of QIIME and mothur. In P_qcc, however, *Minidiscus* and *Skeletonema* were detected,

as the higher number of remaining sequences after preprocessing allowed a selection of rare taxa as OTU representatives (Huse et al. 2010, Kunin et al. 2010). These findings illustrates i) that phylogeny-based methods have a great potential for a reliable sequence annotation and ii) that the OTU clustering algorithms which are based on an arbitrary similarity threshold often fail to model diversity accurately.

Summary and future perspective

This study demonstrated that the choice and configuration of sequence processing methods is important for obtaining a reliable estimate of the microbial diversity of environmental samples. We confirmed that i) OTU-clustering methods estimate the real community structures more accurately than not-clustering methods and ii) that a reliable identification and discrimination of low taxonomic levels with NGS remain challenging (e.g. Kilias et al. 2014, Bonder et al. 2012, Chen et al. 2013, Bik et al. 2012, Zimmermann et al. 2015, Majaneva et al. 2015). The estimated microbial community structure is strongly dependent on sequence-quality improvement, choice of clustering and classification method and related parameter settings as well as quality and completeness of the reference database.

In terms of relative abundance of eukaryotic phyla, the results were comparable across all sequence processing strategies. In addition, a good agreement was achieved in terms of OTU numbers of abundant taxa. However, large discrepancies were observed at the genus level, where the methods considerably influenced the estimated community structure, as shown here for diatoms. Therefore, conclusions on ecological questions drawn on the abundant biosphere of high taxonomic levels are comparable among studies but we advise caution if the emphasis is on rare species.

Our findings can be applied to other sequencing technologies as they are based on fundamental issues and challenges of analyzing amplicon sequences. The transition from the 454 to the Illumina sequencing platforms involves parameter adjustments for sequence processing (e.g. high error-rates in homopolymers is not an issue for Illumina reads), the microbial community features obtained are largely comparable (Kozich et al. 2013, Mahé et al. 2014a) across sequencing technologies. In addition, PCR-bias and PCR-chimeras are still important issues and need to be considered to obtain reliable information on community features. These problems can most likely only be tackled in PCR-free methods, like third-generation sequencing technologies (e.g., Nanopore, PacBio). Generally, for taming

systematic errors, it is desirable to use multi-marker based quantification, e.g. analyzing sequenced full meta-genomes directly, and restricting to single copy genes (e.g. PhyloSift, Darling et al. 2014). These approaches demand for a huge enhancement of reference databases, in particular for eukaryotic organisms from the marine realm.

We highlighted that the identification of sequences at different taxonomic levels would be more accurate if unclustered sequences were classified by means of a "perfect" database in terms of completeness and absence of sequencing and classification errors. Since this is utopian for most eukaryotic taxonomic groups, OTU clustering remains essential for target-gene-based assays focusing on community features as it allows quantifying the number of uncharacterized groups. Consequently, diversity can be well-depicted with OTU-clustering methods, but composition analyses remain an issue due to dependency on reference data, which is still sparse and needs further improvement in quality for eukaryotes.

In contrast to conventional OTU approaches, methods like SWARM (Mahé et al. 2014b) offer similarity parameter free clustering which reflect the real diversity more accurate by omitting arbitrary threshold values. In SWARM, clusters are defined as the network of similar sequences connected by edges, each reflecting one single nucleotide difference. In the future, this approach may replace the more arbitrary OTU-picking strategies discussed in this work, since it is mainly based on sequence similarity landscapes, integrating abundance information on the nodes without selecting a possibly inappropriate clustering algorithm. However, this approach will not tackle the problem of normalization with respect to gene copy number.

Furthermore, the placement of sequences (OTUs or full sequence set) onto a fixed rooted phylogenetic backbone tree, constructed of well selected high-quality reference sequences, has the potential to depict the real eukaryotic composition best, especially the details in cryptic or rare species composition (Vergin et al. 2013). A combination of pre-classification based on only-sequence-similarity (e.g. rdp, UCLUST consensus) only on the level of defined supergroups (e.g. diatoms, dinoflagellates) and a more accurate phylogeny-based annotation utilizing group-specific phylogenetic trees has been shown to be well-suited (Stecher et al. 2015).

In contrast to the fast developing sequencing technologies, the reference databases are improving only slowly. Even if reference databases are extensively and continuously curated, as it is the case e.g. for Silva, it is still a challenge to update the fast-changing taxonomy, the nomenclature of species and their phylogeny. Numerous taxa are still missing or underrepresented which is also true for diatoms (Kermarrec et al. 2013, Zimmermann et al.

2015). In particular, the PR2 reference database (Guillou et al. 2012) can be seen as a community effort in curating and extending the 18S reference data and its taxonomic information quality. Similar activities are underway in Silva in the "Eukaryotic Taxonomy Working Group", optimizing phylogeny and taxonomy locally in the eukaryotic domain, e.g. with information from Adl et al. (2012).

For characterizing composition with non-molecular approaches there is a high demand for additional information augmenting the reference databases, such as ecological data of the sampling area, light microscopy pictures or voucher of the investigated specimen (Bik et al. 2012, Zimmermann et al. 2015). Light microscopy provides a pure insight into the species diversity and beyond that valuable ecological information such as cell conditions or inorganic and organic debris. However, some species or even genera are hardly distinguishable from each other, even with expert knowledge. Still, we recommend the inclusion of conventional methods, such as light microscopy, to diversity studies as they give additional sample information and help to critically evaluate the outcomes of molecular surveys in terms of the applied analysis methods.

Tables

Table 1. Overview of software and sequence-processing-parameters. Sequences obtained from environmental samples were processed with QIIME (Q), mothur (m) and DNAStar/PhyloAssigner (P). Quality checking of sequences included trimming, maximum number of allowed primer and barcode mismatches, maximum homopolymer length and ambiguous bases and the deletion of chimeric sequences. Undenoised (ud) and denoised (d) data sets were generated within QIIME and mothur. Sequences were only quality-checked (qc) or quality-checked and clustered (qcc) in DNAStar and classified with PhyloAssigner. For further details see **S1 Text**.

	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc
Trimming	min. 200 - max. 670 bp		450 flows, subsequent min. 200 bp		min. 300 - max. 670 bp	
Max. primer mismatch	2		2		-	
Max. barcode-mismatch	1		1		-	
Max. homopolymer length	8		8		8	
Max. ambiguous bases	0		0		0	
Denoising software	-	DeNoiser	-	PyroNoise	-	-
Chimera detection	UCHIME		UCHIME		UCHIME AssemblePro in DNASTAR	
Clustering	USEARCH		Average neighbor		(Furthest neighbor)	-
OTU min. size cutoff	4	-	-		1	-
Reference taxonomy	Silva SSU Ref NR 111 (97% clustered)		Silva SSU Ref NR 111 (97% clustered)		Silva SSU Ref NR 111 (subset of 97% clustered)	
Classifier	Uclust, RDP, BLASTN		mothur (implementation of RDP)		pplacer	

Table 2. Summary of sequence characteristics. Sequence statistics for undenoised (ud) and denoised (d) data sets within QIIME (Q) and mothur (m) and only quality-checked (qc) and quality-checked and clustered (qcc) data sets within DNAStar/PhyloAssigner (P). The unprocessed (raw data set) sequence characteristics are given as well. Results from samples T1, T3 and T9 were pooled.

	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc	Raw data set
No. of OTU represented sequences	48982	29957	37853	37785	57286	64740	101601
Mean length	520	526	278	278	430	430	610
Min. – max. length	203 - 637	275 - 562	205 - 293	205 - 293	300 - 607	300 - 607	80 - 1193
No. of chimera	1347	95	499	491	108	108	n.a.
Max. ambiguous bases	0	0	0	0	0	0	5
Max. homopolymer length	8	8	8	8	8	8	31

Table 3. Summary of processed sequences and numbers of OTUs. Final number of OTU representative sequences (repr. seqs.), percentage of raw sequences (% raw seqs) and total number of OTUs given for the abundant biosphere (relative OTU abundance of ≥ 5% and ≥ 1%) and the rare biosphere (< 1%). Abbreviations heading the rows: QIIME (Q), undenoised (ud), denoised (d), mothur (m), only quality-checked (qc), quality-checked and additionally clustered (qcc), DNAStar/PhyloAssigner (P). No clustering in P_qc (n.a.). The numbers of sequences before any processing (raw data set) are given as well.

Pipeline	T1						T3						T9					
	repr. seqs.	% raw seqs.	OTUs total	OTUs ≥ 5%	OTUs ≥ 1%	OTUs < 1%	repr. seqs.	% raw seqs.	OTUs total	OTUs ≥ 5%	OTUs ≥ 1%	OTUs < 1%	repr. seqs	% raw seqs.	OTUs total	OTUs ≥ 5%	OTUs ≥ 1%	OTUs < 1%
Q_ud	18489	44.3	67	4	11	56	13354	52.6	115	2	12	103	17139	49.7	137	2	18	119
Q_d	11110	26.6	60	4	8	52	8276	32.6	129	2	8	121	10571	30.67	146	3	11	135
m_ud	14901	35.7	105	3	5	100	10307	40.6	194	2	9	185	12645	36.7	253	4	12	241
m_d	14906	35.7	90	3	5	85	10277	40.4	190	2	9	181	12602	36.6	244	4	12	232
P_qcc	23702	56.8	1056	3	13	1043	14242	56.1	953	2	9	944	19342	56.1	1296	2	17	1279
P_qc	26783	64.2	n.a.	n.a.	n.a.	n.a.	15759	62	n.a.	n.a	n.a	n.a.	22198	64.4	n.a.	n.a.	n.a	n.a.
Raw data set	41750	100	n.a.	n.a	n.a	n.a.	25407	100	n.a.	n.a	n.a	n.a.	34466	100	n.a.	n.a.	n.a	n.a.

Table 4. **Relative proportions of identified diatom genera achieved with sequence processing and light microscopy of water samples T1, T3 and T9**. Only complete diatom cells were taken into consideration in light microscopy (LM). Default classifiers were: uclust for QIIME (Q), RDP-implementation (mothur) for mothur (m) and PhyloAssigner (P). ud undenoised, d denoised, qcc quality-checked and clustered, qc quality-checked.

	T1							T3							T9						
	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc
Amphiprora						0.03	0.01														
Attheya				0.02	0.02	0.01												0.61	0.63	0.54	0.44
Bacillaria						0.1														3	
Chaetoceros	81.8	0.43	0.46	0.02	0.02	0.44	0.5	89.3	2.74	5.26			1.89	3.62	22.1	4.56	3.87			2.45	3.74
Corethron													2.83	1.45	0.47					7.08	6.59
Coscinodiscus							0.02														
Craticula						0.03	0.02														
Discotella							0.01														
Entomoneis	0.02		0.36					0.19		3.51											
Eucampia	2.45							1.87													
Eunotia							0.01														
Fragilariopsis	1.34							1.87							34.3						
Haslea							0.01														
Minidiscus						0.35	0.08													1.09	0.22
Navicula	0.12			0.02	0.02	0.37	0.49	0.14						2.17	0.23			7.32	7.5	7.63	5.71
Nitzschia	1.54						0.4	0.53							16.9						1.76
NPK2-133						0.03														3.54	0.44
Papiliocellulus						15.6	0.77							0.72						0.27	0.66
Pinnularia						0.02	0.12														0.22

Table 4. continued

	T1							T3							T9						
	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc	LM	Q_ud	Q_d	m_ud	m_d	P_qcc	P_qc
Pleurosigma				0.02	0.04		0.04														
Porosira	0.19		12.9	2.13	2.13	7.9	11.1			15.8	7.81	8.62	11.3	8.7				3.05	3.13	1.36	1.98
Raphoneis							0.01														
Sellaphora						0.03	0.06							0.72							0.44
Skeletonema						0.09	0.36						1.89	2.9						1.09	1.98
Stephanodiscus						0.02	0.01														
Thalassiosira	8.44	82.5	82.1	96.6	96.7	39.3	5.37	5.39	75.3	70.2	75	82.8	47.2	5.8	21.6	74.7	72.9	84.2	86.3	36.5	2.64
Zeuk10						0.53	0.19														0.22
Total no. of genera	7	2	4	6	6	16	20	7	2	4	2	2	5	8	6	2	2	4	4	11	14

Figures

Fig 1. Relative proportions of sequences for all analyzed pipelines. (**A**) Eukaryotic phyla; (**B**) diatom genera. Different taxonomic classification/placement methods are specified as suffix to pipeline version abbreviations (for more details see text). Default classifiers were: uclust for QIIME (Q), RDP-implementation (mothur) for mothur (m) and PhyloAssigner classifier (PA) for PhyloAssigner (P). ud undenoised, d denoised, qcc quality-checked and clustered, qc quality-checked.

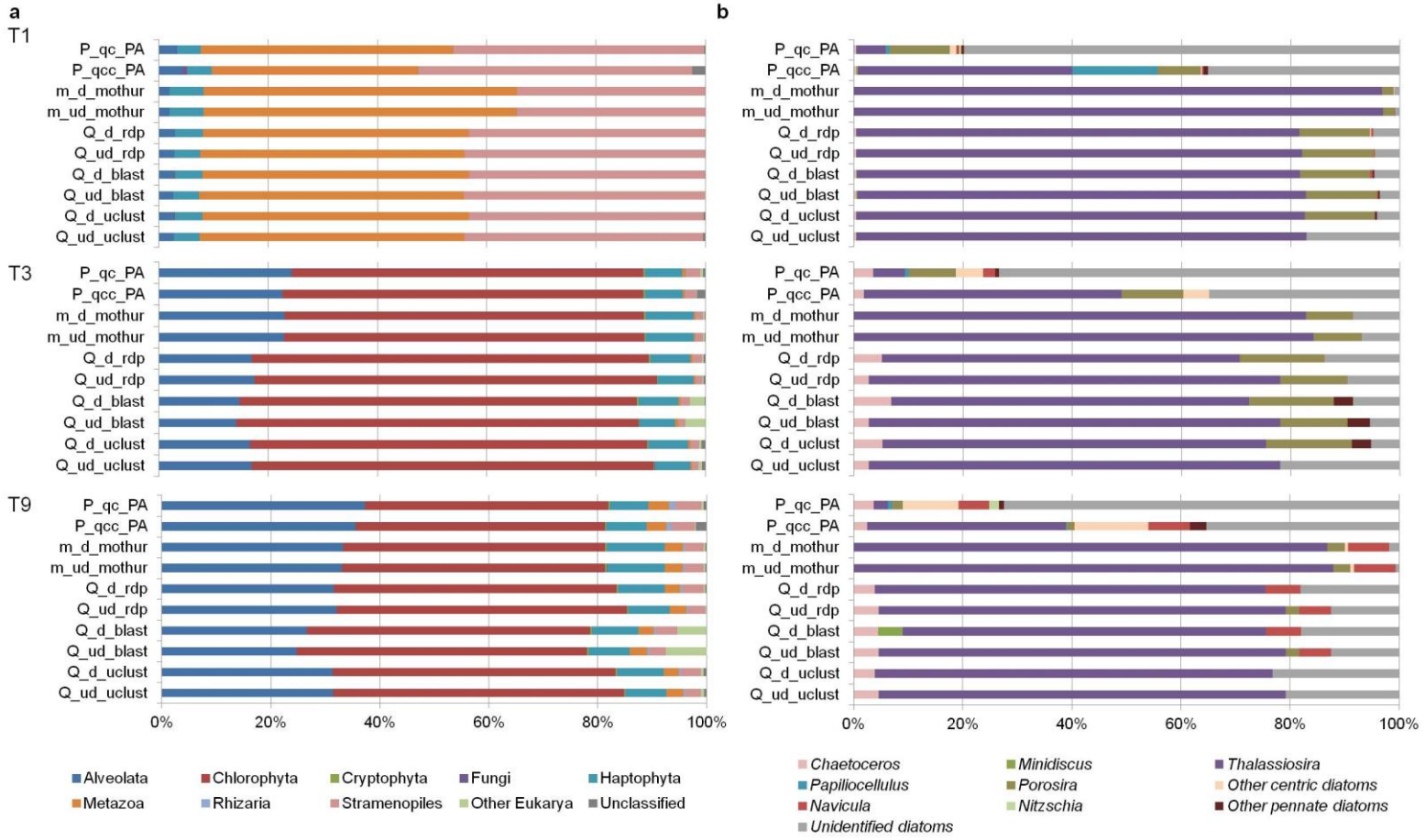


Fig 2. **Relative abundance of diatom genera observed with inverted light microscopy**. Only complete cells were counted, and individuals per liter were recorded by means of the Utermöhl method at 160-fold magnification.

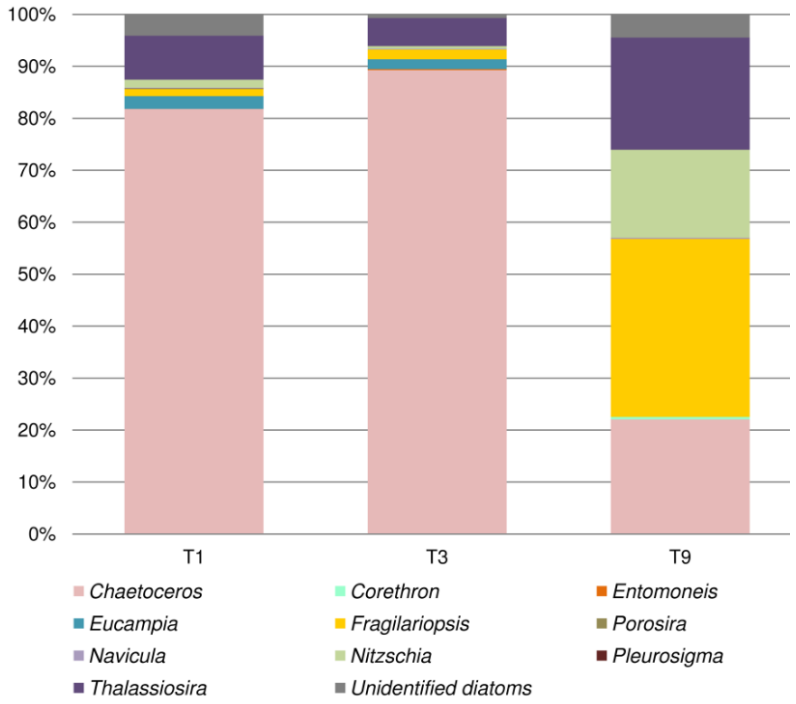
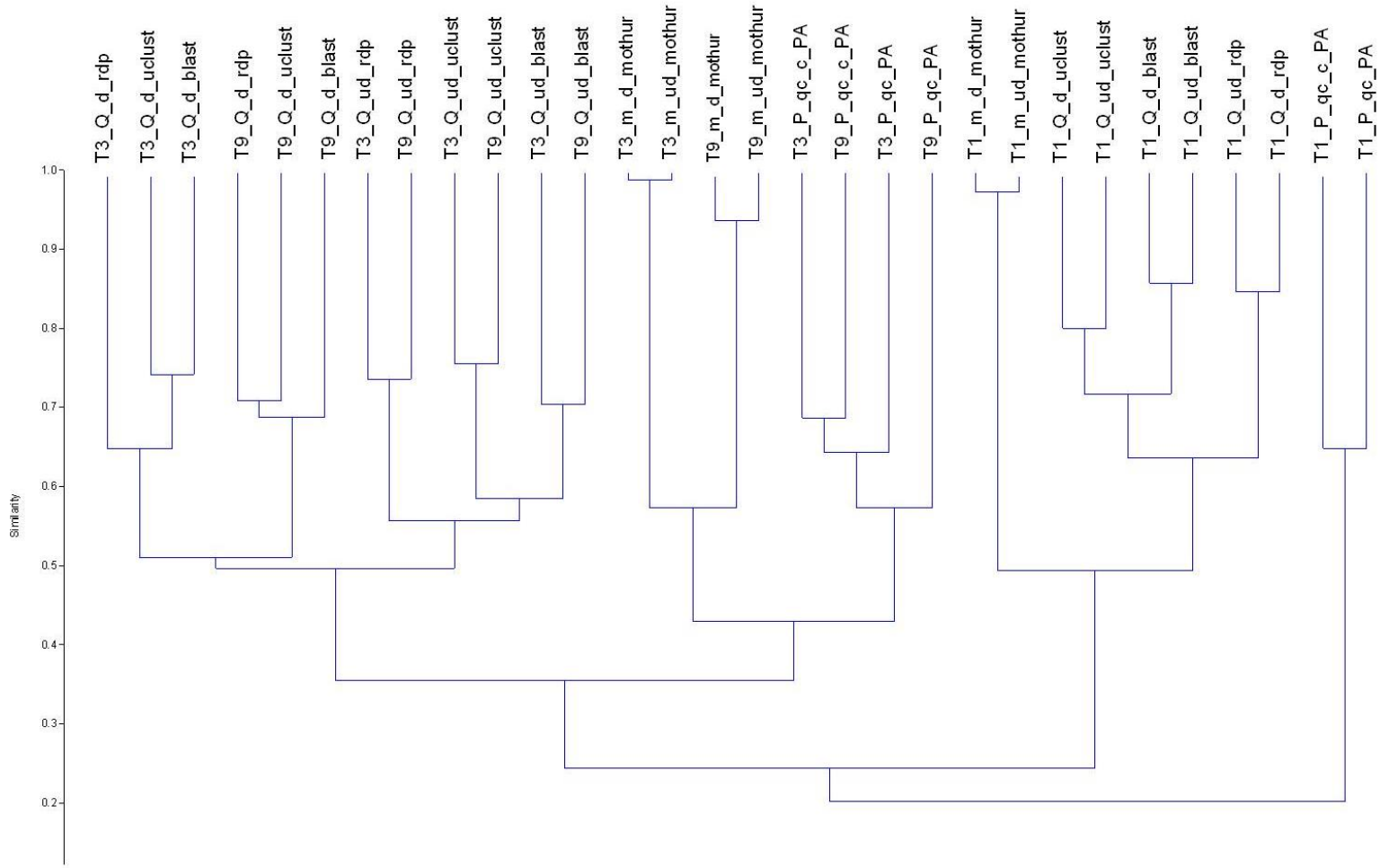


Fig 3 Dendrogram of hierarchical clustering from pairwise Jaccard distances between the different pipeline versions of the three water samples T1, T3 and T9. Default classifiers were: uclust for QIIME (Q), RDP-implementation (mothur) for mothur (m) and PhyloAssigner classifier (PA) for PhyloAssigner (P). ud undenoised, d denoised, qcc quality-checked and clustered, qc quality-checked.



Acknowledgments

We thank the captain and crew of the RV *Polarstern* for their support during the cruise. We also thank want to thank F. Kilpert and B. Beszteri for their bioinformatical support. We are very grateful to K. Oetjen for technical support in the laboratory and E.M. Nöthig for help with microscopic diatom identification. Discussions with T. Jendrossek are especially acknowledged.

References

Adl SM, Simpson AGB, Lane CE *et al.* (2012) The Revised Classification of Eukaryotes. *Journal of Eukaryotic Microbiology*, **59**, 429–514, doi:101111/j1550-7408201200644.

Altschul SF, Gish W, Miller W *et al.* (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Bachy C, Dolan JR, López-García P *et al*. (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME Journal*, **7**, 244–255.

Behnke A, Engel M, Christen R *et al.* (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.

Bik HM, Porazinska DL, Creer S *et al.* (2012) Caporaso JG, Knight R, Thomas WK. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, **27**, 233–243.

Bonder MJ, Abeln S, Zaura E *et al*. (2012) Comparing clustering and pre-processing in taxonomy analysis. *Bioinformatics*, **28**, 2891–2897.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chen W, Zhang CK, Cheng Y *et al*. (2013) A Comparison of Methods for Clustering 16S rRNA Sequences into OTUs. *PLoS ONE*, **8**, e70837.

Darling AE, Jospin G, Lowe E *et al.* (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, **2**, e243.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a chimera checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.

Edgar RC. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar RC, Haas BJ, Clemente JC *et al.* (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Egge E, Bittner L, Andersen T *et al.* (2013) 454-pyrosequencing to Describe Microbial Eukaryotic Community Composition, Diversity and Relative Abundance: A Test for Marine Haptophytes. *PLoS ONE*, **8**, e74371.

Elwood HJ, Olsen GJ, Sogin ML (1985) The small-subunit ribosomal-RNA gene sequences from the hypotrichous ciliates Oxytrichia nova and Stylonychia pustulata. *Molecular Biology and Evolution*, **2**, 399–410.

Eren AM, Morrison HG, Lescault PJ *et al*. (2014) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, **9**, 968–979.

Gaspar JM, Thomas WK (2013) Assessing the consequences of denoising marker-based metagenomic data. *PLoS ONE*, **8**, e60458.

Godhe A, Asplund ME, Härnström K (2008). Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Applied and Environmental Microbiology*, **74**, 7174–7182.

Guillou L, Bachar D, Audic S (2012) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic acids research*, gks1160.

Guiry M (2015a). *Chaetoceros septentrionalis* Østrup, 1895. In: Guiry MD & Guiry GM (2016). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway (taxonomic information republished from AlgaeBase with permission of M.D. Guiry). Accessed through: World Register of Marine Species at http://www.marinespecies.org/aphia.php/ aphia.php?p=taxdetails&id=149296 on 2016-05-10

Guiry M. (2015b). *Amphiprora alata* (Ehrenberg) Kützing, 1844. In: Guiry MD & Guiry GM (2016). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway (taxonomic information republished from AlgaeBase with permission of M.D.

Guiry). Accessed through: World Register of Marine Species at http://www.marinespecies.org/aphia.php?p =taxdetails&id=149141 on 2016-05-10

Hammer Ø, Harper DAT, Ryan PD (2001) PAST-PAlaeontological STatistics, ver. 1.89. *Palaeontologia electronica*, **4**, 1–9.

Hoppenrath M, Beszteri B, Drebes G *et al.* (2007) Thalassiosira species (Bacillariophyceae, Thalassiosirales) in the North Sea at Helgoland (German Bight) and Sylt (North Frisian Wadden Sea) a first approach to assessing diversity. *European Journal of Phycology*, **42**, 271–288. doi: 10.1080/09670260701352288

Hoppenrath M, Elbrächter M, Drebes G (2009) Marine phytoplankton: Selected microphytoplankton species from the North Sea around Helgoland and Sylt. Kleine Serckenerg – Reihe 49, E. Schweizerbart Science Publishers, Stuttgart, Germany, 264 p.

Huse SM, Huber JA, Morrison HG *et al.* (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, 143–152.

Huse SM, Welch DM, Morrison HG *et al.* (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12**, 1889–1898.

Kermarrec L, Franc A, Rimet F *et al*. (2013) Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Molecular Ecology Resources*, **13**, 607–619.

Kermarrec L, Franc A, Rimet F *et al*. (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, **33**, 349–363.

Kilias E, Wolf C, Nöthig EM *et al*. (2013) Protist distribution in the Western Fram Strait in summer 2010 based on 454-pyrosequencing of the 18S rDNA, *Journal of Phycology*, **49**, 996–1010.

Kilias E, Kattner G, Wolf C *et al.* (2014) A molecular survey of protist diversity through the central Arctic Ocean. *Polar Biology*, **37**, 1271–1287.

Kozich JJ, Westcott SL, Baxter NT *et al*. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbioly*, **79**, 5112–5120.

Kraberg A, Baumann M, Durselen CD (2010) Coastal phytoplankton - Photo guide for northern European seas. Verlag Dr. Fiedrich Pfeil, München, Germany, 204 pp.

Kunin V, Engelbrektson A, Ochman H *et al*. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity etimates. *Environmental Microbiology*, **12**, 118–123.

Luddington IA, Kaczmarska I, Lovejoy C (2012) Distance and character based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). PLoS ONE, **7**, e45664.

Mahé F, Mayor J, Bunge J *et al*. (2014a) Comparing high-throughput platforms for sequencing the V4 region of SSU-rDNA in environmental microbial eukaryotic diversity surveys. *The Journal of Eukaryotic Microbiology*, **62**, 338–345.

Mahé F, Rognes T, Quince C *et al.* (2014b) Swarm: robust and fast clustering method for amplicon-based studies. *Peer J*, **2**, e593.

Majaneva M, Hyytiäinen K, Varvio SL (2015) Bioinformatic Amplicon Read Processing Strategies Strongly Affect Eukaryotic Diversity and the Taxonomic Composition of Communities. *PLoS ONE*, **10**, e0130035. doi:10.1371/journal.pone.0130035

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, **11**, 538.

Morgan MJ, Chariton AA, Hartley DM *et al.* (2013) Improved Inference of Taxonomic Richness from Environmental DNA. *PLoS ONE*, **8**, e71974.

Nickrent DL, Sargent ML (1991) An overview of the secondary structure of the V4-region of eukaryotic small-subunit ribosomal-RNA. *Nucleic Acids Research*, **19**, 227–35.

Plankton*Net Data Provider at the Alfred Wegener Insitute for Polar and Marine Research. hdl: 10013/de.awi.planktonnet.

Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, **46**, 48–50.

Pruesse E, Quast C, Knittel K *et al*. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

Quillfeldt CH von (2004) The diatom Fragilariopsis cylindrus and its potential as an indicator species for cold water rather than for sea ice. *Vie Milieu*, **54**, 137–143.

Quince C, Lanzen A, Curtis TP *et al*. (2009) Accurate determination of microbial diversity from 454-pyrosequencing data. *Nature Methods*, **6**, 639–641.

Quince C, Lanzen A, Davenport RJ *et al*. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.

Reeder J, Knight R (2009) The 'rare biosphere': a reality check. *Nature Methods*, **6**, 636–637.

Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nature Methods*, **7**, 668–669.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7741.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*, **6**, e27310.

Schloss PD, Westcott SL (2011) Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Applied and Environmental Microbiology*, **77**, 3219–3226.

Schloss PD (2013) Secondary structure improves OTU assignments of 16S rRNA gene sequences, *The ISME Journal*, **7**, 457–460.

Schmidt TSB, Matias Rodrigues JF, von Mering C (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environmental Microbiology*, **17**, 1689–1706.

Stecher A, Neuhaus S, Lange B, *et al*. (2015) rRNA and rDNA based assessment of sea ice protist biodiversity from the central Arctic Ocean. *European Journal of Phycology*, doi:10.1080/09670262.2015.1077395

Stoeck T, Bass D, Nebel M *et al*. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, **19**, 21–31.

Sun Y, Cai Y, Huse SM *et al*. (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Briefings in Bioinformatics*, **13**, 107–121.

Vergin KL, Beszteri B, Monier A *et al.* (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME Journal*, **7**, 1322–1332.

Wang Q, Garrity GM, Tiedje JM *et al.* (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Wolf C, Frickenhaus S, Kilias ES *et al*. (2013) Regional variability in eukaryotic protist communities in the Amundsen Sea. *Antarctic Science*, doi:10.1017/S0954102013000229.

Zhu F, Massana R, Not F *et al.* (2005) Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, **52**, 79–92.

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms, Diversity & Evolution*, **11**, 173–192.

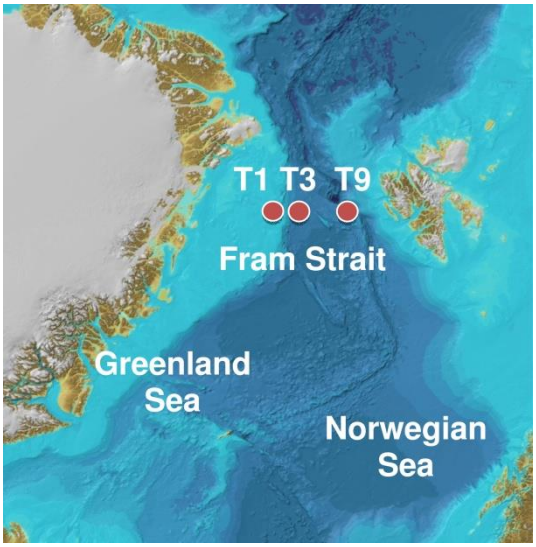
Zimmermann J, Glöckner G, Jahn R *et al*. (2015) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, **15**, 526–542.

Data accessibility

The sequencing data of the analyzed samples are available from Kilias et al. (2013) at Journal of Phycology, DOI: 10.1111/jpy.12109. The programs, user manual and example datasets are available on http://qiime.org/1.9.0/tutorials/index.html (QIIME), http://www.mothur.org/wiki/ 454_SOP (mothur) and https://www.awi.de/en/science/special-groups/scientific-computing/bioinformatics/software.html (PhyloAssigner). The compiled reference database used for sequence's phylogenetic placements in PhyloAssigner is available on request.

Supporting Information

Fig. S1 Sampling position of water samples T1, T3 and T9. Samples were collected during *RV Polarstern* cruise to the Fram Strait with a CTD-Rosette (conductivity, temperature and depth) from the respective chlorophyll maximum layer depth (15 – 35 m) in July 2010. Base of map: http://www.ngdc.noaa.gov/mgg/bathymetry/arctic/



Text S1

Description of sequence processing pipelines

QIIME pipelines

We used QIIME version 1.8.0 (Caporaso et al. 2010) for pyrosequence analysis. For the denoising (referred as Q_d) pipeline version, the quality of reads was checked as follows: reads with one mismatch of barcode and two mismatches of primer, ambiguous bases and more than 8 homo-polymers were discarded (**Table 1**). We retained reads with the default minimum length of 200 bp. The remaining reads were filtered as follows: A 50 bp-sliding-window scan was applied and each read was removed from the dataset which showed a drop of the average phred quality below 25.

Flowgrams of the remaining reads were denoised using DeNoiser (Reeder and Knight 2010), an implementation of PyroNoise (Quince et al. 2009) within the QIIME package. Subsequently, the QIIME pipeline usearch.qf (based on Usearch version 5.2.236 (Edgar 2010) which incorporates UCHIME) was used for chimeric sequence detection and OTU clustering. The usearch.qf pipeline has a pre-clustering step included which is based on the same similarity threshold as the OTU clustering. Chimeric sequences were detected with reference-data (Silva 111 SSU Ref NR) and without reference-data (i.e. *de novo*). In the *de novo* mode, sequences are sorted by abundance and chimeras are deleted under the assumption that they are less abundant than their parental sequences. Sequences were removed if considered to be chimeric by both methods in consensus.

After quality filtering and denoising, remaining sequences were clustered and OTUs were determined *de novo* (from the sequence data alone, i.e., not using a reference data set) at a minimum similarity threshold of 97%. OTUs that represent true singletons in the Q_d are retained under assumption that exclusively high quality reads remained. Representative sequences were assigned using a QIIME-prepared version of Silva 111 SSU Ref NR after 97% similarity clustering (http://www.arb-silva.de /fileadmin/silva_databases/qiime/Silva_111_release.tgz). The default classifier used in QIIME is UCLUST consensus (UCLUST version 1.2.22q), with standard settings, including the best three kmer search hits which have at least 90% similarity to the query sequence. Taxonomic assignment then is the common taxonomic string prefix of the majority of the best hits. For additional comparisons, we used the RDP-classifier version 2.2 at a confidence level of 0.8 and the BLASTN algorithm version 2.2.22 (best hit) for taxonomic assignment as well.

The non-denoising (referred as Q_ud) pipeline version was conducted similar to Q_d, however with less stringent settings, as described on the QIIME-website. Reverse primers were truncated and if not found, the sequences were cut at the expected maximum amplicon length of 670 bp, and sequence output was not altered. In contrast to the retainment of OTUs representing true singletons in Q_d, we retained OTUs comprising more than three sequences in Q_ud.

Mothur pipelines

We processed pyrosequencing data using mothur version 1.29.2. As for QIIME, we constructed a denoised (referred as m_d) and an undenoised (referred as m_ud) data set (**Table 1**). We conducted both mothur pipelines by using the default options following P. Schloss' SOP for 454 (http://www.mothur.org/wiki/454_SOP), with the exception that no denoising was applied in m_ud.

The pipeline steps are briefly described as follows: Flowgrams were trimmed to 450 flows ("trim.flows"), the recommended value for reducing the average error rate which was shown to increase significantly after around 450 flows (Schloss et al. 2011). As for QIIME, one base mismatch of barcode and two base mismatches of the primer as well as a maximum homo-polymer length of 8 were allowed.

We denoised the sequences using the mothur implementation of the PyroNoise algorithm (Quince et al. 2009). Sequences with a minimum length of 200 bp and an average quality score of 35 in a 50-bp sliding window were retained in the data set (trim.seqs) and aligned with "align.seqs" to a reference database (Silva 111 SSU Ref NR) based on mothur implemented version of the multiple sequence alignment algorithm NAST (Schloss et al. 2009). Aligned reads were trimmed at the same position. The subsequent "pre.cluster" command reduced pyrosequencing errors of remaining unique sequences by combining highly similar (up to 2 bp difference from the entire sequence length) sequences into a so called pre-cluster. According to the SOP, PCR-chimeras were detected and removed by applying UCHIME in *de novo* mode.

For distance matrix calculation for clustering ("dist.seqs"), pairwise distances larger than 0.15 were discarded. Sequences were clustered to OTUs at a distance level of 0.03 on basis of the average neighbor clustering ("cluster"). OTUs that represent true singletons were

retained. Representative sequences were annotated using mothur's implementation of the RDP classifier 2.2 (Wang et al. 2010) (referred here as "mothur" classifier) at a 0.8 confidence level ("classify.seqs" followed by "classify.otu") and the QIIME-prepared 97% clustered version of the Silva SSU Rev NR 111.

DNAStar/PhyloAssigner pipelines

For taxonomic assignment with PhyloAssigner version 6.166 (Vergin et al. 2013), we prepared two different data sets: a quality-checked and sequence-clustered data set (referred as P_qcc) and a quality-checked data set without sequence clustering (referred as P_qc) (**Table 1**).The pipeline was configured in the same way for both data sets. Hence, we annotated all quality-checked sequences in P_qc but only OTU-cluster representatives in P_qcc.

Sequences of length below 300 bp and over 670 bp were excluded from further analysis. These settings ensure the inclusion of the complete hypervariable V4 region. Sequences were deleted if they contained ambiguous bases (N) or more than length 8 homo-polymers. Sequences were checked for chimera using UCHIME version 4.2 (Edgar et al. 2011) with the same reference set of sequence as used for taxonomic classification within this pipeline.

The data set P_qcc was obtained using Assemble Pro in DNAStar (furthest neighbor algorithm), SeqManPro (DNAStar Inc., Madison, WI. USA, Version 9.1.1) (assembling match size: 50, minimum match percentage: 97%). Single sequences that were abundant in only one of the three water samples were removed from the clustered data set.

For the PhyloAssigner-step, representative sequences (for P_qcc) or all sequences (for P_qc) were aligned with hmmalign from the HMMER3 package to a reference alignment (Silva 111 SSU Ref NR; see below). The sequence's phylogenetic placement in a fixed rooted phylogenetic backbone tree was done in pplacer (Matsen et al. 2010) within PhyloAssigner. For each sequence the last common ancestor node (LCA) was assigned, using a cumulated likelihood weight ratios cutoff of 0.8, denoted LCA-assignment (Vergin et al. 2013). The LCA can be an inner node without own taxonomic information. Therefore, we extracted the full taxonomic strings of all tree leafs below the LCA node and used their common prefix string for taxonomic assignment of the LCA-assigned sequence. As reference database we

used Silva 111 SSU Ref NR and the corresponding ARB tree. This data set was thinned down manually with the goal to harmonize qualities of selected sequences while reducing redundancies in the structure of the phylogenetic tree, i.e. optimizing the tree to a structurally and taxonomically representative set of 4000 leafs.

References

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7: 335–336.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinform. 2010;26: 2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27: 2194–2200.

Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010;11: 538.

Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, et al. Accurate determination of microbial diversity from 454-pyrosequencing data. Nat Methods. 2009;6: 639–641.

Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods. 2010;7: 668–669.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75: 7537–7741.

Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One. 2011;6: e27310.

Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, et al. High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. ISME J. 2013;7: 1322–1332.

Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiology. 2007;73(16): 5261–5267.

Fig. S2 Relative abundance of operational taxonomic units. 454 pyrosequences were assigned to eukaryotic phyla (**a**) and diatoms (**b**) of the water samples T1, T3 and T9. Compared were undenoised (ud) and denoised (d) data sets within QIIME (Q) and mothur (m) and quality-checked (qc) and clustered (c) data sets within DNAStar/PhyloAssigner (P).

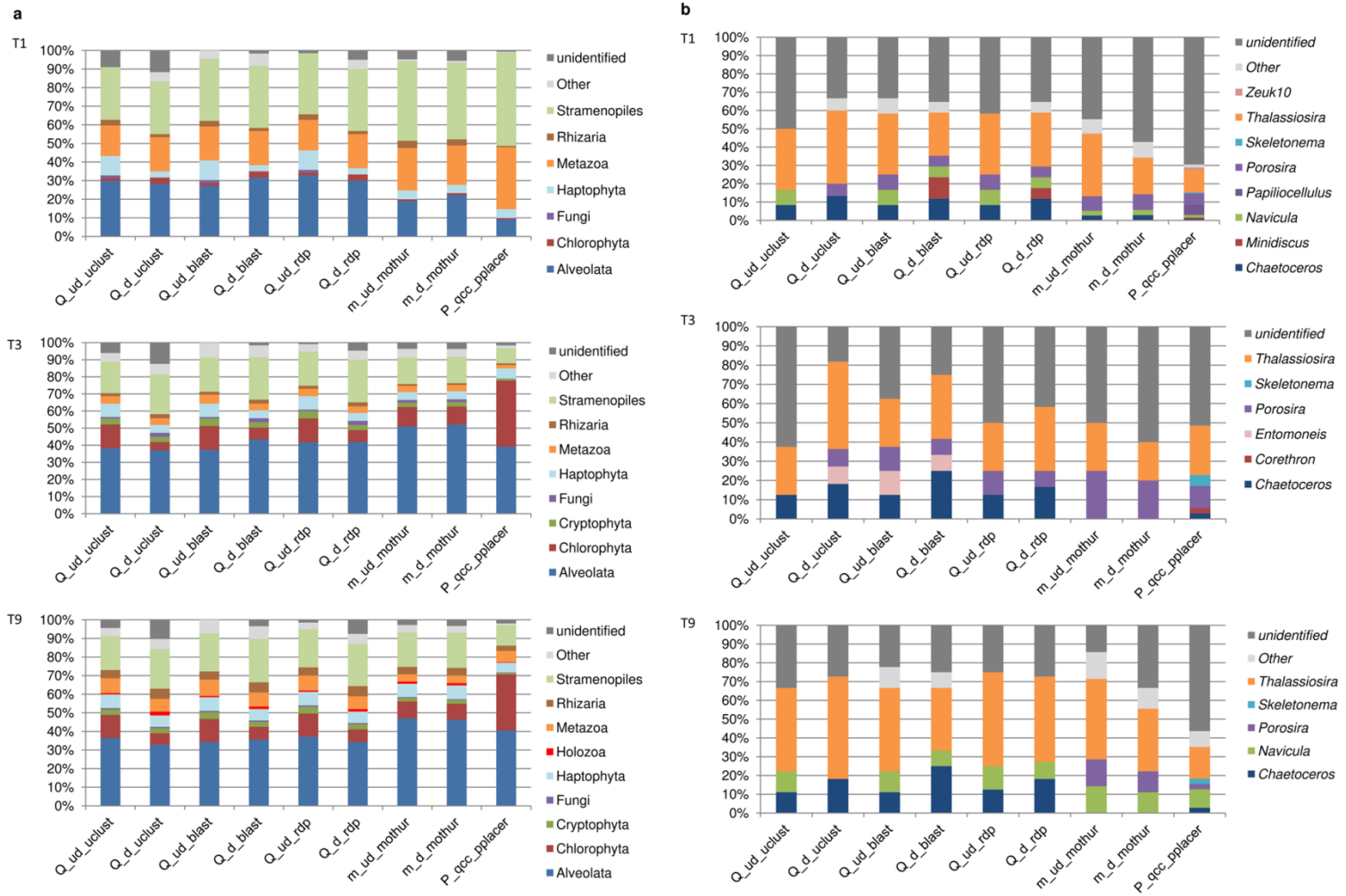


Table S1 Number of operational taxonomic units for eukaryotes and diatoms. Eukaryotic phyla (**a**) and diatoms (**b**) were assigned at a 97% similarity threshold. Compared were undenoised (ud) and denoised (d) data sets within QIIME (Q) and mothur (m) and quality-checked (qc) and clustered (c) data sets within DNAStar/PhyloAssigner (P).

a

T1 Eukarya	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Alveolata	20	17	18	19	22	18	20	20	95
Chlorophyta	1	2	1	2	1	2	1	1	6
Discoba	0	0	1	0	0	0	0	0	0
Fungi	1	0	1	0	1	0	0	0	4
Haptophyta	7	2	7	2	7	2	5	4	50
Holozoa	0	1	0	1	0	1	0	0	2
Kathablepharidophyta	0	1	0	1	0	1	0	0	1
Mantamonas	0	0	0	0	0	0	0	0	1
Metazoa	11	11	12	11	11	11	24	19	352
Rhizaria	2	1	2	1	2	1	4	3	7
Stramenopiles	19	17	22	20	22	20	45	37	531
Telonema	0	1	0	1	0	1	1	1	0
Zeuk77	0	0	3	1	0	0	0	0	0
unidentified	6	7	0	1	1	3	5	5	7
Total	67	60	67	60	67	60	**105**	**90**	**1056**

T3 Eukarya	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Alveolata	44	48	43	56	48	54	99	99	365
Chlorophyta	16	6	16	9	16	9	22	20	361
Cryptophyta	4	4	5	4	5	4	5	5	10
Discoba	0	0	1	0	0	0	0	0	0
Fungi	1	3	1	3	1	3	3	3	1
Haptophyta	9	6	9	6	9	6	9	9	56
Holozoa	0	1	0	1	0	1	2	2	2
Katablepharidophyta	0	0	0	0	0	0	0	0	1
Metazoa	5	5	6	5	5	5	7	7	18
Palpitomonas	1	2	1	2	1	2	1	1	2
Picozoa	1	2	1	2	1	2	3	2	0
Rhizaria	2	3	2	3	2	3	2	2	9
Stramenopiles	21	30	23	32	23	32	30	29	82
Telonema	3	2	3	2	3	2	4	4	11
Zeuk77	1	1	4	2	0	0	0	0	0
unidentified	7	16	0	2	1	6	7	7	15
Total	115	129	115	129	115	129	**194**	**190**	**933**

T9 Eukarya	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Alveolata	50	48	47	52	51	50	119	113	526
Chlorophyta	17	9	17	10	17	10	23	21	391
Cryptophyta	4	4	5	4	5	4	5	5	11
Discoba	0	0	1	0	0	0	0	0	0
Fungi	1	1	1	1	1	1	1	1	2
Haptophyta	10	9	10	9	10	9	18	18	67
Holozoa	1	3	1	2	1	2	3	3	4
Mantamonas	0	0	0	1	0	1	1	1	0
Metazoa	11	10	12	11	11	10	10	10	79
Palpitomonas	1	1	1	1	1	1	0	0	3
Picozoa	1	3	1	4	1	4	5	4	0
Rhizaria	6	8	6	8	6	8	10	10	36
Stramenopiles	25	31	28	34	28	33	47	46	143
Telonema	3	2	3	2	3	2	4	4	9
Zeuk77	1	2	4	2	0	0	0	0	0
unidentified	6	15	0	5	2	11	7	8	25
Total	137	146	137	146	137	146	**253**	**244**	**1296**

b

T1 Diatoms	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Attheya	0	0	0	0	0	0	1	1	1
Amphiprora	0	0	0	0	0	0	0	0	1
Bacillaria	0	0	1	0	0	0	0	0	1
Chaetoceros	1	2	1	2	1	2	1	1	4
Craticula	0	0	0	0	0	0	0	0	1
Entomoneis	0	1	1	1	0	0	0	0	0
Minidiscus	0	0	1	2	0	1	0	0	4
NPK20133	0	0	0	0	0	0	0	0	1
Navicula	0	0	1	1	1	1	1	1	7
Papiliocellulus	0	0	0	0	0	0	0	0	28
Pinnularia	0	0	0	0	0	0	0	0	1
Pleurosigma	0	0	0	0	0	0	2	2	0
Porosira	0	1	1	1	1	1	3	3	32
Odontella	0	0	0	0	0	1	0	0	0
Sellaphora	0	0	0	0	0	0	0	0	1
Skeletonema	0	0	0	0	0	0	0	0	3
Stephanodiscus	0	0	0	0	0	0	0	0	1
Thalassiosira	4	6	3	4	2	5	13	7	64
Zeuk10	0	0	0	0	0	0	0	0	6
unidentified	11	5	8	6	12	6	17	20	356
Total	**16**	**15**	**17**	**17**	**17**	**17**	**38**	**35**	**512**

T3 Diatoms	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Chaetoceros	1	2	1	3	1	2	0	0	1
Corethron	0	0	0	0	0	0	0	0	1
Entomoneis	0	1	1	1	0	0	0	0	0
Porosira	0	1	1	1	1	1	1	1	4
Skeletonema	0	0	0	0	0	0	0	0	2
Thalassiosira	2	5	2	4	2	4	1	1	9
unidentified	5	2	3	3	4	5	2	3	18
Total	**8**	**11**	**8**	**12**	**8**	**12**	**4**	**5**	**35**

T9 Diatoms	Q_ud_uclust	Q_d_uclust	Q_ud_blast	Q_d_blast	Q_ud_rdp	Q_d_rdp	m_ud_mothur	m_d_mothur	P_qcc_pplacer
Attheya	0	0	0	0	0	0	1	1	1
Bacillaria	0	0	0	0	0	0	0	0	1
Chaetoceros	1	2	1	3	1	2	0	0	2
Corethron	0	0	0	0	0	0	0	0	1
Guinardia	0	0	0	0	0	0	0	0	0
Minidiscus	0	0	0	1	0	0	0	0	1
NPK20133	0	0	0	0	0	0	0	0	1
Navicula	0	0	1	1	1	1	1	1	7
Papiliocellulus	0	0	0	0	0	0	0	0	1
Porosira	0	0	1	0	1	0	1	1	2
Skeletonema	0	0	0	0	0	0	0	0	2
Thalassiosira	2	6	2	4	2	5	3	3	12
unidentified	6	3	4	3	4	3	1	3	40
Total	**9**	**11**	**9**	**12**	**9**	**11**	**7**	**9**	**71**

Text S2

Influence of sequence preprocessing on protist composition and diversity estimates

Sequence quality filtering and trimming

The sequence length plays a fundamental role for a reliable taxonomic classification since the potential phylogenetic information is higher the longer the reads are, providing that a sufficient read quality is given. In this study, the sequence information content was probably highest in the QIIME pipeline variants, since pre-processing here yielded the longest sequences (**Table 2**). Counterintuitively, we detected the lowest number of OTUs with QIIME. For Q_d, this was probably the consequence of the more stringent quality filtering applied in addition to denoising (fewer sequences were left) (**Table 3**). In contrast to Q_d, the number of OTUs in Q_ud was primarily not reduced by a stringent quality filtering but by the application of the minimum cluster-size cutoff per OTU (at least 4 sequences). Sequence processing in QIIME considerably reduced the rare biosphere (i.e. relative sequence abundance < 1%) and thus, the eukaryotic diversity was not overestimated (Huse et al. 2010, Kunin et al. 2010). However, poorly represented taxa could have been missed, possibly leading to an underestimation of the actual microbial diversity.

Sequences processed in mothur were the shortest compared to the other pipelines, since flowgrams of all samples were trimmed to the same flowgram length of 450 flows. Consequently, less read information could be used for taxonomic assignment of the sequences than in the QIIME and the PhyloAssigner pipelines. However, compared to Q_d, a higher number of sequences remained after quality-filtering in mothur which might be one reason for the higher number of OTUs and a better retention of the potential rare biosphere in mothur (**Table 3)**.

Besides basic quality filtering, trimming and chimera removal, no threshold for the minimum average quality score was set and no denoising was conducted in the two PhyloAssigner pipeline variants. Thus, many more sequences with lower quality were kept here. On the one hand, studies demonstrated that the possibility to detect true rare taxa in the data set is higher, the more sequences are left for classification (Huse et al. 2010, Kunin et al. 2010). On the other hand, species diversity might be overestimated due to a less stringent quality filtering which allows accumulation of sequencing errors (Schloss et al. 2011, Huse et al. 2007). Therefore, the number of OTUs of abundance <1% was far higher than those achieved with QIIME and mothur (**Table 3**).

Chimeric sequences

The deletion of PCR-chimeras is highly recommended because chimeras can spuriously increase diversity estimates (Behnke et al. 2011). In this study, the varying number of chimeras is difficult to compare as different chimera-checking methods at different sequence aggregation levels (i.e. similarity threshold of pre-clustering) were applied.

UCHIME is a fast chimera detection algorithm which was used in all pipeline variants. It either can be used with a reference database or in *de novo* mode (see Supplement Text S1). If the method is reference-based, as applied in the QIIME (in combination with *de novo*) and PhyloAssigner pipelines (only reference-based), species coverage of the reference database is important since more chimeras can be detected if a comprehensive reference is given. In the PhyloAssigner pipeline, we used a high-quality subset of the 97% clustered Silva 111 SSU Ref NR as reference, which included only 4000 sequences. Because this reference was less comprehensive compared to the reference database used in the other pipeline versions (complete 97% clustered Silva111RefNR) and because chimeras were not detected *de-novo*, fewer chimeras were found in the DNAStar/PhyloAssigner pipeline compared to the other pipelines. Not further compared were the parameters of the respective pipeline-incorporated UCHIME versions.

The occurrence of chimeras was reported to correlate with sequence length (Edgar et al. 2011, Haas et al. 2011, Bik et al. 2012). Even if a higher amount of long sequences remained in Q_d compared to m_d, the stringent quality filtering and denoising already reduced the number of detectable chimeras in Q_d. However, a considerably higher number of chimeras was detected in the longer sequences recovered with Q_ud, because reads were filtered less stringent and chimera inducing PCR errors are more likely to occur at the distal end of longer sequences.

Denoising of sequences

Several studies demonstrated that denoising pyrosequences reduces the sequencing error rate enormously and therefore, the number of spurious OTUs (Schloss et al. 2011, Quince et al. 2009, 2011, Bachy et al. 2013, Egge et al. 2013). Denoising was reported to enhance the clustering accuracy and reduce the number of overestimated biodiversity considerably, especially if subsequent chimera filtering is applied (Bonder et al. 2012). However, one side-effect of denoising can be that sequences from rare taxa are falsely altered and therefore

appear as part of sequence clusters of higher abundance, the so-called ‘accordion effect’ (Gaspar and Thomas 2013).

The denoising step altered the sequencing output (due to the ‘accordion effect’) and therefore the diversity estimates in QIIME, as it was also observed by other studies (Gaspar and Thomas 2013, Majaneva et al. 2015). In our study, the effect of denoising was further confirmed by a clear separation of denoised and undenoised data sets in the cluster dendrogram based on taxa abundances (**Fig. 2**). However, we observed similar numbers of OTUs for Q_ud and Q_d (**Table 3**), which was most likely the effect of different cluster-size cutoffs rather than the effect of denoising. The numbers of OTUs were considerably lower (data not shown), if we applied a cluster size cutoff of four to the processed sequence data set of Q_d, as it was done for Q_ud.

The diversity estimates achieved from mothur differed only marginally between denoised and undenoised data sets. This is strong evidence that the phenomenon of erroneously suppressed taxa does not play an important role in the dataset processed by mothur. Here, the diversity estimate was reduced by only four to 15 OTUs if denoising was applied compared to the data set which was, except for the denoising, equally processed. One reason for this small difference could be the applied flowgram trimming step to 450 flows which resulted in improved sequence quality and reduced the number of spurious OTUs in both data sets. In addition, the trimming to equal flowgram lengths avoided the ‘accordion effect’. Thus, most probably real errors were removed in m_d.

Sequence clustering

A similarity threshold of 97% was shown to be most suitable to reproduce original eukaryotic diversity (Behnke et al. 2011). Clustering has the advantage to reduce the computational time since downstream analyses are performed with the OTU representative sequences only. As it has the effect of absorbing most of the sequencing errors and thus allows a more reliable classification of sequences (Kunin et al. 2010), a higher proportion of diatom genera (65%) was detected in P_qcc compared to P_qc (20%). In addition, the diversity can be reliably estimated and is independent from the taxonomic assignment.

Mothur’s hierarchical clustering algorithm found more OTUs at a 97% sequence similarity than QIIME’s heuristic USEARCH algorithm. Before the actual OTU clustering was carried out in mothur and QIIME, sequences were pre-clustered, meaning that highly

similar sequences were grouped together. The clustering method in mothur is based on a pseudo-single-linkage algorithm (pre-clustering) followed by an average-linkage OTU clustering based on a distance matrix (Schloss 2011). The pre-clustering, however, is performed with the original aligned sequences instead of calculating a distance matrix as it was introduced by Huse et al. (2010). This method is supposed to reduce the OTU richness but not the number of rare OTUs (Schloss 2011). In the QIIME pipelines, sequences were pre-clustered at the identity threshold of 97% in the course of the usearch.qf pipeline, which might already have reduced the number of potential rare OTUs beforehand of the actual OTU clustering at the same similarity threshold. This might have resulted in an underestimation of eukaryotic diversity in our study. An effect of clustering methods on the diversity estimates was also reported by other studies (Bonder et al. 2012, Sun et al. 2012, Schloss 2013, Chen et al. 2013, Schmidt et al. 2015).

Heuristic cluster algorithms, such as USEARCH, require less computation time than hierarchical clustering but grouping of sequences into OTUs has been found defective in some studies (Sun et al. 2012, Schloss 2013, Chen et al. 2013, Schmidt et al. 2015). Schmidt et al. (2015) reported a shift in the diversity estimates and cluster composition if UCLUST/ USEARCH were used, compared to the hierarchical clustering algorithm (average linkage) that was used in the mothur pipeline variants. In terms of OTU assignment accuracy, Schloss (2013) observed that a NAST alignment against an aligned Silva-reference combined with hierarchical clustering outperformed the alignment independent algorithm USEARCH. Using the latter, a higher rate of falsely clustered sequences was observed, i. e. sequences co-occurring in the same OTU which pairwise similarity is below 97%.

References

Bachy C, Dolan JR, López-García P *et al*. (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME Journal*, **7**, 244–255.

Behnke A, Engel M, Christen R *et al*. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.

Bik HM, Porazinska DL, Creer S *et al.* (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in ecology and evolution*, **27**, 233–243.

Bonder MJ, Abeln S, Zaura E *et al*. (2012) Comparing clustering and pre-processing in taxonomy analysis. *Bioinformatics*, **28**, 2891–2897.

Chen W, Zhang CK, Cheng Y *et al*. (2013) Comparison of Methods for Clustering 16S rRNA Sequences into OTUs. *PLoS ONE*, **8**, e70837.

Edgar RC, Haas BJ, Clemente JC *et al*. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Egge E, Bittner L, Andersen T *et al.* (2013) 454-pyrosequencing to Describe Microbial Eukaryotic Community Composition, Diversity and Relative Abundance: A Test for Marine Haptophytes. *PLoS ONE*, **8**, e74371.

Gaspar JM, Thomas WK (2013) Assessing the consequences of denoising marker-based metagenomic data. *PLoS ONE*, **8**, e60458.

Haas BJ, Gevers D, Earl AM *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, **21**, 494–504.

Huse SM, Huber JA, Morrison HG et al. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, 143–152.

Huse SM, Welch DM, Morrison HG et al. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12**, 1889–1898.

Kunin V, Engelbrektson A, Ochman H et al. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity etimates. *Environmental Microbiology*, **12**, 118–123.

Majaneva M, Hyytiäinen K, Varvio SL et al. (2015) Bioinformatic Amplicon Read Processing Strategies Strongly Affect Eukaryotic Diversity and the Taxonomic Composition of Communities. *PLoS ONE*, **10**, e0130035. doi:10.1371/journal.pone.0130035

Quince C, Lanzen A, Curtis TP et al. (2009) Accurate determination of microbial diversity from 454-pyrosequencing data. *Nature Methods*, **6**, 639–641.

Quince C, Lanzen A, Davenport RJ et al. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*, **6**, e27310.

Schloss PD (2013) Secondary structure improves OTU assignments of 16S rRNA gene sequences, The ISME Journal, **7**, 457–460.

Schmidt TSB, Matias Rodrigues JF, von Mering C (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. Environmental *Microbiology*, **17**, 1689–1706.

Sun Y, Cai Y, Huse SM *et al.* (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Briefings in bioinformatics*, **13**, 107–121.

7.2 Manuscript II

Sea Ice Origin and Sea Ice Retreat are Major Drivers of Variability in Arctic Marine Protist Composition[1]

Running title: Influence of sea ice origin and retreat on Arctic protists

---

Kristin Hardge*[a,b], Ilka Peeken[a], Stefan Neuhaus[a], Thomas Krumpen[a], Thorsten Stoeck[c] and Katja Metfies[a, b]

[a] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, Bremerhaven, 27570, Germany

[b] Jacobs University Bremen, Bremen, Germany

[c] University of Kaiserslautern, Faculty of Biology, Ecology Group, Erwin Schrödinger Str. 14, D-67663 Kaiserslautern, Germany

* Corresponding author: kristin.hardge@awi.de

1 This is the submitted and not revised manuscript version. This is a pending publication in Marine Ecology Progress Series.

ABSTRACT

The ongoing decrease of Arctic sea ice thickness and extent is expected to have significant implications for protists in the Arctic Ocean. In this study, we analyzed protist community patterns and diversity in 62 water and 21 sea ice samples to elucidate the impact of sea ice conditions and origin as well as water mass properties. Samples were collected in the Central Arctic Ocean during two summer periods. ARISA analysis and Illumina sequencing revealed highly variable protist community patterns in sea ice, mostly determined by regions of ice floe origin. Patterns of pelagic protists correlated with sea ice concentrations, water masses and sampling regions. Compared to 2011, pelagic and sea ice protist communities were less diverse in 2012, suggesting a community restructuring as a consequence of strong melting processes. In the future, effective changes in Arctic marine protist communities can be expected because of the strong dependence on sea ice origin and the time shift of sea ice growth to less biological active winter months on the shelves. This comprehensive large-scale study serves as a good baseline for future studies as it shows the importance of sea ice origin and conditions for protist biodiversity and gives evidence for a possible change in protist communities due to sea ice decline.

KEY WORDS: 18 S rRNA gene; Arctic Ocean; ARISA; Illumina sequencing; biogeography; ice algae; phytoplankton

INTRODUCTION

The Arctic Ocean with its ice cover, wide shelves, strong freshwater inflow through large rivers and its inflow and outflow gateways, harbors highly variable and complex habitats for biodiversity. Even within the sea ice, unique sub habitats such as snow, melt ponds, brine channels or the ice water interface are present. While studies about protistan communities in melt ponds and snow are rare, it is well-known that bottom ice is usually inhabited by a diverse protist community (Horner 1985; Horner et al. 1992). The ice/water interface plays an important role in structuring the sea ice properties due to the exchange of nutrients and protists during melting and freezing processes (Ackley et al. 1987; Gradinger and Ikävalko 1998; Melnikov et al. 2003; Różańska et al. 2008; Lee et al. 2011). Protistan species entrapped in the new ice are the base for the succession of protistan communities in the following spring (Niemi et al. 2011). However, hardly anything is known about how far the ambient environmental conditions during sea ice formation might impact the standing stocks of e.g. nutrients in sea ice. So far any evidence of a potential link between sea ice origin and its nutrient or protist compositions is still missing.

During sea ice melting, temperature, salinity and nutrient profiles of the sea ice changes which in turn has an impact on succession of e.g. diatoms, the dominant taxa in the sea ice (reviewed by Arrigo 2014). During the melting process, particulate material and ice algae are released into the water column (Ambrose et al. 2005; Juul-Pederson et al. 2008; Boetius et al. 2013) possibly altering the protist community composition in the uppermost pelagic zone. The community in the water column is influenced by water mass properties and sea ice concentration (Hegseth & Sundfjord 2008; Moran et al. 2012; Kilias et al. 2014). Therefore, the current environmental changes, particular the constant loss of sea ice might affect habitat structure and protist community structure in the Arctic water column and sea ice.

While in the past, sea ice was melting predominantly from the top, the combined effects of different environmental factors such as increased heat transport via the North Atlantic Current (Holliday et al. 2009; Beszczynska-Möller et al. 2012), warmer air temperatures (Serreze et al. 2000; Comiso 2003; Stroeve et al. 2012) and strong winds (Parkinson & Comiso 2013; Zhang et al. 2013), now results in an additional melting of sea ice from the bottom. As a consequence, sea ice thickness decreases (reviewed in Meier et al. 2014), melt ponds proliferate (Nicolaus et al. 2012) and sea ice concentration declines (Maslanik et al. 2007; Comiso et al 2008; Maslanik et al. 2011; Comiso 2012). The most recent sea ice minimum was observed in summer 2012. During this summer, net primary

productivity of algae in sea ice, water column and melt ponds exceeded previous estimates and was highly influenced by sea ice melting (Fernández-Mendez et al. 2015). Fast sinking aggregates of the sub-ice diatom *Melosira arctica* were observed on the sea floor with strong implications for the deep-sea community and carbon export (Boetius et al. 2013). Additionally, the lower sea ice concentration strongly influenced the community of under-ice fauna in the Eurasian Basin with a shift to pelagic amphipods at nearby ice-free stations (David et al. 2015). But little is known how the drastic sea ice loss affects the biodiversity of the sea ice biota and the phytoplankton.

Overall, extensive large-scale studies investigating the protist community structure and the environmental impact on their distribution in the Central Arctic Ocean are still scarce because of its limited accessibility and methodological constraints. Conventional methods to investigate the entire protist community such as light- or electron microscopy are very time consuming and require high taxonomic expertise. This is particularly true for pico-eukaryotes (< 3 µm) which have been understudied in the past due to their small size and often limited morphological features. In this case, high-throughput sequencing technologies are appropriate to analyze species with varying sizes and abundances not ascertainable with light microscopy. In addition, the molecular fingerprinting method ARISA (automated ribosomal intergenic spacer analysis) is a quick and cost-efficient method to provide a general overview of differences in protist community structure based on the heterogeneity of the ITS1 gene region. Several studies showed the suitability of this method for assessment of diversity patterns and its relationship to environmental parameters (Fechner et al. 2010; Bienhold et al. 2012; Wolf et al. 2013; Gobet et al. 2014; Kilias et al. 2015). A recent comparison of fingerprint profiles with 454 pyrosequencing showed its validity in the Arctic Ocean (Kilias et al. 2015).

In this study, we aimed at better understanding the impact of variable sea ice conditions and origins on Arctic marine protist community composition. We used ARISA and Illumina sequencing to assess protist community variability, composition and diversity in water and sea ice samples collected in the Central Arctic Ocean during two summer periods with contrasting sea ice concentrations. The unique opportunity of sampling during August and September in 2011 and during the record sea ice low in 2012 allowed us to define possible influences of sea ice origin and sea ice retreat on protist assemblages in sea ice and water column.

MATERIAL AND METHODS

Sample collection

Samples were taken in the Central Arctic Ocean during the *RV Polarstern* expeditions TransArc in 2011 (PS 78; 15 August - 23 September 2011, Fig. 1A) and IceArc in 2012 (PS 80; 5 August - 29 September 2012, Fig. 1B). A large part of the samples was taken in a common region within the Eurasian Basin (Fig. 1C, white box). In total 23 water samples were collected in 2011 (Fig. 1A, Kilias et al. 2014) and 39 water samples in 2012 (Fig. 1B, Tab. 1). The water samples were collected with Niskin bottles (12 liter) attached to a CTD (conductivity, temperature, depth) rosette from the chlorophyll maximum depth which varied between 10 m and 50 m (mostly around 20 m) in both years. For molecular analysis, 2 liter of seawater were filtered through Isopore Membran Filters (Millipore, Billerica, MA, USA) with pore sizes of 10 µm, 3 µm and 0.4 µm to ensure collection of all protist cell sizes. Filters were stored in Eppendorf tubes at -80 °C. The water samples taken in 2011 were analyzed and discussed in a previous publication by Kilias et al. (2014) and used here for an overall comparison of protist community patterns, not only in the water column but also in the sea ice collected in 2011 and 2012.

Eleven sea ice cores (nine first-year ice, FYI, two multi-year ice, MYI, Fig. 1A) were taken during the TransArc expedition in 2011 and ten sea ice cores (seven FYI, three MYI, Fig. 1B) were taken during the IceArc expedition in 2012 (Tab. 1). Ice cores were drilled with a Kovacs 9 cm diameter corer (Kovacs Enterprise, Roseburg, USA). The temperature of the ice was directly measured on the floe by drilling into the ice and determining the temperature every 5 cm with a temperature probe (Testo 720). These ‘physical’ cores were sectioned in 10 cm pieces and the salinity was determined with Salinometer (WTW Cond.3110) after melting. Subsamples were taken for measurements of dissolved inorganic nutrients. Additional ‘biological’ cores were sampled for molecular analysis. The ‘biological cores’ were sectioned in 10 cm intervals and melted in 0.2 µm-filtered sea-water to minimize osmotic stress to protists during the melting process (Miller et al. 2015). Melting of sea ice was conducted under low light conditions at 4 °C for 24 - 48 hours. The further processing (filtering and storage) was done as described for the water samples. While the data set of 2011 only focused on the bottom ice community, the data set of 2012 focused on the community sampled in the entire sea ice cores. However, the main biomass of protists is present in the bottom ice section (Horner 1985; Horner et al. 1992) and thus, the overall protist community patterns observed should not be affected. In general, one sea ice core per station was taken. At stations 335 and 384 of the IceArc cruise, we took two samples per station to gain an insight in the variability

of sea ice habitats (FYI and MYI core at Stn. 335 and FYI and recently formed sea ice at Stn. 384).

Physico-chemical parameters

Sea ice samples

Physico-chemical parameters, such as temperature, salinity and inorganic dissolved nutrients (nitrate, nitrite, ammonium, phosphate, and silicate) were measured in the 'physical cores' taken at each ice station (see Method section "Sample collection) (Kattner, unpublished). Due to restrictions in the research permissions, we were not allowed to measure nutrients for the water samples at stations 267 and 290 taken in 2011. No nutrient data were available for the sea ice cores at stations 203 and 209 in 2011 and station 224 in 2012. The physico-chemical data were determined and compared for the entire ice cores of each year. However, for comparison with the ARISA data, we analyzed the physico-chemical data according to the 'biological cores'; that is for the bottom ice section in 2011 and the entire ice core in 2012.

Sea ice concentration data were obtained from http://www.meereisportal.de (grant: REKLIM-2012-04) based on Advanced Microwave Scanning Radiometer (AMSR-E) data using the ARTIST SeaIce (ASI) algorithm (Spreen et al. 2008). To elucidate the impact of sea ice origin for protist distribution in the Central Arctic Ocean, we determined the sea ice origin based on ice drift information obtained from satellites. In this study, two different sets of ice drift products were used: The first data set, Polar Pathfinder Sea Ice Motion Vectors (Version 2) obtained from the National Snow and Ice Data Center (NSIDC) was chosen because of its year round availability. Below it is used to calculate ice drift trajectories during summer months (June-August). The second data set, sea ice motion provided by the Center for Satellite Exploitation and Research (CERSAT) at the Institut Francais de Recherche pour l'Exploitation de la Mer (IFREMER), shows a good performance on the Siberian shelf (Krumpen et al. 2013) and was therefore used to complement the calculation of ice drift trajectories between September and May. To determine drift trajectories and source areas of sampled sea ice a specific ice area is tracked backwards until: (a) the ice reaches a position next to a coastline, (b) the ice concentration at a specific location reaches a threshold value of (> 15 %) when ice parcels are considered lost, or (c) the tracking time exceeds four years. A more detailed method description is provided in Krumpen et al. (2016).

Water samples

Water temperature, salinity and dissolved inorganic nutrients (phosphate, silicate, nitrite and nitrate) were determined for the Deep Chlorophyll Maximum (DCM) (Ark-XXVI/3, doi:10.1594/PANGAEA.832164, Ark-XXVII/3, Bakker unpublished dataset #834081) according to standard methods (Kattner and Becker 1991; Kerouel and Aminot 1997). Water masses were distinguished based on the combination of water temperature, salinity and nutrient signatures at the DCM and named according to their main signatures in Atlantic Water (AW), Mixed Water I (MW I), Pacific Water (PW) and Mixed Water II (MW II) (Tab. 1). We tested the water mass classification by using the analysis of similarity (ANOSIM) permutation test and used the classification only, if the differences were significant.

DNA isolation and ARISA PCR amplification

Genomic DNA was extracted from filters using the NucleoSpin® Plant II kit (Macherey-Nagel) according to the manufacturer's protocol. For Automated ribosomal intergenic spacer analysis (ARISA) of water samples and sea ice cores, extracted DNA of each size fraction was pooled in equal volumes. We amplified triplicates of the ITS1 (internal transcribed spacer) region which is located between the 18S and 5.8S rRNA genes using the primer-set 1528F (5'-GTA GGT GAA CCT GCA GAA GGA TCA-3') (modified after Medlin et al. 1988) and ITS2 (5'-GCT GCG TTC TTC ATC GAT GC-3') (White et al. 1990). The forward primer was labeled with 6-FAM (6-carboxyfluorescein) at the 5'-end. The PCR (polymerase chain reaction) mixture contained 1 µl of DNA extract, 1 x HotMaster Taq Buffer containing 2.5 mM Mg^{2+} (5 Prime, USA), 0.8 mM dNTP-mix (Eppendorf, Germany), 0.2 mM of each Primer and 0.4 U of HotMaster Taq DNA polymerase (5 Prime, USA) in a final volume of 20 µl. Reactions were carried out in a Mastercycler (Eppendorf, Germany) under the following conditions: initial denaturation at 94 °C for 3 min, 35 cycles of denaturation at 94 °C for 45 sec, followed by annealing at 55 °C for 1 min and extension at 72 °C for 3 min, and a final extension at 72 °C for 10 min. PCR fragments were separated by capillary electrophoresis on an ABI Prism 310 Genetic Analyzer (Applied Biosystems, USA).

ARISA analysis

Analysis of fragment sizes achieved with ARISA was carried out according to Kilias et al. (2015). We analyzed the electropherograms with GeneMapper v. 4.0 software (Applied Biosystems). All fragments smaller than 50 bp were excluded from the analysis because they potentially originated from primers or primer dimers. We used R (R development Core Team 2008) for binning to remove background noise and obtain sample-by-bin operational taxonomic unit tables (Ramette 2009). A given fragment length was considered present if it appeared in at least two of the three replicates. For distance calculations based on Jaccard index the table was converted to a presence/absence matrix. The average number of fragments per habitat and sampling period and their lengths were determined.

Illumina sequencing

A subset of 28 water and sea ice samples was sequenced with Illumina technology (Tab. 1, bold sample numbers). Isolated DNA of each filter size (10 µm, 3 µm and 0.4 µm) was pooled in equal volumes. The V4 region was amplified in triplicates using universal primer set TAReuk454FWD1 (5´-CCA GCA (G/C)C (C/T)GC GGT AAT TCC-3'; *S. cerevisiae* position 565-584) and TAReukREV3 (5`-ACT TTC GTT CTT GAT (C/T)(A/G)A-3`; *S. cerevisiae* position 964-981) (Stoeck et al. 2010). The PCR mixture contained 10 µl 5x Phusion high-fidelity buffer 1 µl dNTP-mix, 1 µl of each primer, 0.5 Phusion Hot Start high-fidelity tag polymerase (New England Biolabs GmbH, Frankfurt/Main, Germany) and 2 µl template DNA in a final volume of 50 µl. DNA amplification was carried out in two rounds using a Mastercycler (Eppendorf, Germany). Initial activation was performed at 98 °C for 30 sec followed by denaturation at 98 °C for 10 sec (10 cycles), annealing at 57 °C for 30 sec and extension at 72 °C for 30 sec. In the second round, denaturation was performed at 98 °C for 10 sec (25 cycles) followed by annealing at 52 °C for 30 sec and extension at 72 °C for 30 sec. The final extension was at 72 °C for 10 min. PCR products were then purified with NucleoSpin® Gel & PCR Clean up kit (Macherey-Nagel) according to manufacturer's protocol. Triplicates were pooled and sequencing was performed on an Illumina MiSeq producing 2x300 paired-end reads.

Sequence analysis

For merging of paired-end reads, the tool fastq-join, which is incorporated in QIIME version 1.8.0 (Caporaso et al. 2010), was used with default settings. An analysis pipeline developed in-house (Stecher et al. 2015), for which the backbone consists of scripts from QIIME version 1.8.0, was used for sequence processing. Reads were quality-filtered according to recommended settings in Bokulich et al. (2013). Only sequences were further processed which fully match the primer sequences at beginning and end of the sequence respectively and which are between 330 and 460 bp in length (length of expected insert 380 bp - 420 bp). Primers were not removed as their degeneracy might contribute to taxonomic differentiation. Sequence numbers were down-sampled to the lowest number of sequences occurring in the dataset (~ 130000 seqs). Subsequently, the QIIME workflow usearch.qf (based on Usearch version 5.2.236 (Edgar 2010) which incorporates UCHIME (Edgar et al. 2011), was used for chimeric sequence detection and OTU clustering. After a pre-clustering step, sequences were checked for chimeras, by both using reference data for comparison (Silva 119 SSU Ref NR) and with utilization of the abundance-sorted query sequence pre-cluster as self-reference (i.e. *de novo*). Sequences were removed if considered to be chimeric by at least one of the two methods. The remaining sequence set was clustered *de-novo* into OTUs at a minimum similarity threshold of 98%. According to Bokulich et al. (2013) all OTUs, consisted of less than 0.005% of processed sequences, were removed.

OTU sequences were classified by phylogenetic placement utilizing PhyloAssigner v. 6.166 (Vergin et al. 2013), which places sequences in a fixed rooted backbone tree and assigns queries to nodes using a last common ancestor (LCA) approach which incorporates placement uncertainties. To obtain a reference set, the ARB tree of Silva 111 SSU Ref NR and the corresponding multiple sequence alignment was thinned down manually by optimizing the tree to a structurally and taxonomically representative set of 4000 leafs. This compiled reference set is available on request. PhyloAssigner uses HMMER v. 3.0 (Eddy 2011) to align the query OTU sequences to the reference alignment and places (pplacer, Matsen et al 2010) them at their most probable positions in the phylogenetic reference tree, which was rooted by means of an outgroup consisting of Opisthokonta. For each sequence its placement likelihood values are reported and the LCA node for each query is determined at an accumulated likelihood weight ratios cutoff of 0.8. LCA nodes can be leafs or inner nodes without taxonomic label. Therefore the full taxonomic string of all leaves below the LCA node was extracted and the common prefix string was used as taxonomic assignment.

Statistics

To visualize protist community patterns obtained with ARISA and sequencing, a non-dimensional metric scaling (NMDS) based on Jaccard index was applied using the statistics software PAST (version 2.17; Hammer et al. 2001). We performed this ordination analysis for all samples and years together and for each habitat separately. The following statistical analysis was performed with R (R development Core Team 2008). To evaluate similarities between ARISA and Illumina data sets, a Mantel test was performed with 999 permutations and the distance measurements Jaccard for ARISA and Jaccard and Bray Curtis for Illumina. As the similarity between the two methods was highly significant using Jaccard and Bray Curtis ($p < 0.001$), we used ARISA profiles to test for correlations with environmental parameters. In addition, sample size was considerably higher for the ARISA analysis, which allows detecting potential relationships with a high confidence level. An ANOSIM (with 999 permutations, R package vegan) was computed to test for differences between predefined grouping of the ARISA profiles and environmental parameters according to water masses and regions. To obtain a detailed picture of the environmental conditions at the sampling stations, we computed principal component analyses (PCA). A PCA of water samples was also conducted in Kilias et al. (2014) with salinity, ice thickness and floe size. Here, we reanalyzed the data with additional environmental variables. A potential correlation of ARISA data with environmental parameters was examined by the application of a Mantel test. This was done with 999 permutations and the distance measures Jaccard for the ARISA profiles and Euclidean for the environmental parameters. If the Mantel test gave significant results, we assessed the best fitting of single environmental variables to the ARISA distances by using the envfit function of the R package vegan.

RESULTS

Environmental conditions of sea ice samples

The sea ice cores sampled during the TransArc expedition in 2011 were characterized by a higher sea ice thickness (on average 138 cm for FYI and 284 cm for MYI) (Tab. 1), in comparison to the relatively thin sea ice (on average 101 cm for FYI and 158 cm for MYI) during the IceArc expedition in 2012. In the ice cores, phosphate concentrations were in a similar range in 2011 and 2012, while nitrate concentrations measured in the entire and bottom ice cores were higher in 2012 than those measured in 2011 (Tab. 1). In contrast, silicate and nitrite concentrations were on average higher in 2011 than in 2012 (in entire and bottom ice section). Sea ice temperatures of FYI and MYI were similar in both years (Tab. 1). Sea ice drift data suggest that sea ice cores sampled in 2011 were older than those sampled in 2012 and originated from both, the Eurasian and Amerasian Basin (Fig. 2, black lines). In contrast, sea ice cores sampled in 2012 were younger and originated from the Eurasian Basin only (Fig. 2, blue lines).

For comparisons with ARISA data, we computed PCA with the physico-chemical parameters in 2011 (bottom sea ice section, Fig. 3A) and 2012 (entire sea ice core, Fig. 3B), according to the origin of the ARISA samples. For the samples collected in 2011, the first axis of the PCA explained 45% of the variance between the samples and was mainly driven by gradients of nitrate, silicate and phosphate (Fig. 3A). The second axis was mainly driven by salinity and sea ice thickness, which were both negatively correlated with temperature. The physico-chemical parameters showed no defined grouping according to the sampling regions (Fig. 3A, colors), even if we tested the nutrients only. However, the nutrient profile was significantly correlated to the regions of sea ice origin (bottom section: R = 0.59, P = 0.02, entire core: R = 0.35, P = 0.08; Fig. 3A, geometrical shapes). For example, ice cores originating from the Makarov Basin (Fig. 3A, circles) were characterized by higher concentrations of silicate, nitrate and phosphate than those from the Amundsen Basin (Fig. 3A, diamonds).

The PCA of the physico-chemical parameters measured in 2012 explained 67% of the total variance in the environmental profile (Fig. 3B). Sea ice thickness and partly nitrate were responsible for the variance in the first PC axis and the second PC axis was mainly driven by gradients of silicate and temperature. Despite different ice types (FYI and MYI) were sampled at station 335, the environmental profiles were similar. The opposite was true for station 384, were we collected a FYI core and a new-ice sample. The new-ice sample was characterized by a high salinity (9 psu) and phosphate concentration (0.11 µmol/l) compared to the FYI

(1.74 psu, 0.01 µmol/l). The nutrient profiles were significantly correlated with the sampling regions (R = 0.47, *P* = 0.01; Fig. 3B, colors) but we observed no correlation with the regions of sea ice origin (Fig. 3A, geometrical shapes). An overview of the results of statistical analysis can be found in Tab. 2.

Environmental conditions of water samples

The water stations were characterized by high sea ice concentrations (on average 81%, Kilias et al. 2014) during the TransArc expedition in 2011, while we observed lower sea ice concentrations (on average 54%, Tab. 1) during the IceArc expedition in 2012. Water temperatures measured at the chlorophyll maximum depth were in a similar range during the two expeditions (2011 average -1.1 °C, 2012 average -1.3 °C, Tab. 1), while salinity was higher in 2012 than in 2011 (2011 average 31.2 psu, 2012 average 32.8 psu, Tab. 1), and dissolved nutrient concentrations, except those of nitrate, were lower in 2012 (Tab. 1) than in 2011.

The PCA of the environmental parameters collected for the water samples in 2011 explained 72.2% of the total variance between the samples (Fig. 4A). As it was presented in Kilias et al. (2014), sea ice conditions explained the environmental distribution of water samples best. In addition, water temperature was an important driver for the sample variance. A more detailed analysis of environmental variables in 2011 can be found in Kilias et al. (2014).

The PCA of the environmental parameters collected in 2012 explained 75.9% of the total variance (Fig. 4B) and illustrates significant regional pattern of the environmental parameters (R = 0.54, *P* = 0.001; Fig. 4B, colors). All stations clustered mainly according to gradients of sea ice concentrations. Sea ice concentrations were low (0 - 53%, Tab. 1) on the Svalbard continental slope, the Laptev Sea and parts of the eastern Amundsen Basin. As sea ice concentration and water temperature were negatively correlated, the water temperature was higher (-1.3 - 1.4 °C, Tab. 1) in these parts of the Arctic Ocean compared to the other regions in this study. The Nansen Basin and the western Amundsen Basin were characterized by high sea ice concentrations (73 - 100%, Tab. 1) and slightly lower water temperatures ranging between -1.4 °C and -1.8 °C.

Protist community structure

The ARISA dataset consisted of, on average, 83 fragments per sample (standard deviation: 14). Overall, we detected 378 different ITS1 fragments with a size range between 50 to 659 bp. In both years, we observed a higher community complexity (i.e. number of fragments per sample) and site-variability (i.e. standard deviation) in the sea ice samples compared to the water samples (Tab. 3). A detailed NMDS ordination analysis based on Jaccard's distances of ITS1 fragments (Fig. 5A) suggests a highly significant clustering of the samples into four groups. Sea ice cores and water samples clustered in four groups, subdivided according to the year of sampling (R = 0.80, *P* = 0.001).

Sequencing resulted in a total of ~ 2.8 million high quality reads (mean lengths 372 - 419 bp) that were assigned to 771 OTUs with on average 482 OTUs per sample. Sea ice samples in 2011 showed higher site-variability than all other samples (Tab. 3). Water communities of both years, were more diverse (number of OTUs) than sea ice communities. NMDS analysis based on the presence/absence of OTUs (Fig. 5B) showed distinct protist communities in sea ice and water samples, subdivided according to the year of sampling (R = 0.65, *P* = 0.001). In both years, sea ice cores were mainly characterized by Chrysophyta (mainly Ochromonadales), Diatomea (mainly Bacillariophyceae) and Cercozoa (mainly Silicofilosea) (Fig. 6). In contrast, typical taxa found in the water column were Chloroplastida (mainly Mamiellophyceae) and Dinoflagellata (mainly Dinophyceae).

Sea ice samples

ARISA profiles of the sea ice protist communities (Fig. 7A) revealed significant differences in protist community composition and variability at annual scale (R = 0.32, *P* = 0.001). These differences were also apparent in the community composition obtained with Illumina sequencing. In 2011, a higher proportion of sequences originated from Holozoa (mainly Choanomonada) were observed in the sea ice core bottom compared to the entire ice cores in 2012 (Fig. 6). Furthermore, *Chlorella* spp. (Chlorophyta, Trebouxiophyceae) was highly abundant in two ice cores of 2011 (212 and 227, other Eukaryota in Fig. 6). In contrast to 2011, higher proportions of *Gymnodinium*-related species ('NPK60-44', Dinophycceae) and *Ochromonas* spp. (e.g. 'CCMP1899', Chrysophyceae) were observed in sea ice of 2012. Overall, the most diverse protist group in sea ice was Bacillariophyta (Diatomea) with up to 23 OTUs, including large proportions of *Nitzschia* spp. and *Navicula* spp. in both years. In 2011, diatom abundance was lower but diversity was higher compared to 2012. However, *Nitzschia* spp. showed higher contributions to sea ice algae community in 2011 (total

abundance 0.2%) than in 2012 (total abundance 0.04%). In contrast, *Melosira* spp. was more frequently observed in 2012 (total abundance 2.7%) than in 2011 (total abundance 1.1%).

Compared to 2012, the variability in sea ice protist community patterns was higher in 2011 as characterized by a higher standard deviation of the average fragment number obtained with ARISA (Tab. 3). The same was observed with Illumina sequencing, where we also found a considerably lower number of OTUs in sea ice 2012 compared to 2011, despite we analyzed entire sea ice cores and not only the bottom section as we did in 2011 (Tab. 3). These differences were mainly visible in a lower number of OTUs belonging to the rare biosphere in 2012.

The ARISA analysis revealed regional differences in community composition. Samples collected in 2011 showed a significant sub-clustering that was best explained by sea ice origin of the samples, which was in Nansen Basin, Amundsen Basin, and the Canada/Makarov Basin (R = 0.34, P = 0.02; Tab. 2, Fig. 7B). In contrast, ARISA profiles of 2012 did not show a significant sub-clustering (Fig. 7C), while the sea ice origin of the samples collected in 2012 was mainly in the Amundsen Basin. The two samples collected at station 384 clustered in close proximity to each other (Fig. 7C) indicating a similar protist community composition of the FYI (384a) and new ice sample (384b). In contrast to this, FYI and MYI cores of station 335 (Fig. 7C) suggest a high within-site variability of different ice types.

Overall, for both years, we found no correlation of the physico-chemical variables of sea ice (i.e. sea ice thickness, temperature, salinity and dissolved nutrients) with ARISA profiles of sea ice protist communities. This was even true if we carried out the Mantel test only with sea ice thickness, salinity and temperature. None of the measured nutrients showed a significant correlation with ranked ARISA distances of the ice samples. An overview of the results of statistical analysis can be found in Tab. 2.

Water samples

In the common sampling regions of the two expeditions, ARISA profiles of pelagic protists differed significantly between the two years (R = 0.66, P = 0.001; Fig. 8A) which was also in accordance with different community compositions. *Micromonas* spp. (*Mamiellophyceae*, Fig. 6) was most common in the water samples of 2011 (total abundance in 2011 18.5%, in 2012 4.6%), while it was a *Gymnodinium*-related dinophyte species ("NPK60-44") in the water samples of 2012 (total abundance in 2011 6.6%, in 2012 9.5%). Diatoms showed a higher contribution to the pelagic community in 2012 than in 2011 (Fig. 6).

This was especially true for the sub-ice diatom *Melosira* spp. (total abundance 0.15% in 2011, 1% in 2012). Overall, Dinoflagellata and Protalveolata were the most diverse protist groups in the water samples with up to 18 different OTUs.

Water samples collected during the record sea ice minimum year in 2012 clustered closer to each other than the samples collected in 2011, suggesting a lower ITS1-lengths variability in 2012 than in 2011 (Fig. 8A). In addition, average number of ITS1 fragments in the water samples 2012 was higher than in 2011 (Tab. 3). In terms of protist diversity, Illumina sequencing revealed on average fewer OTUs in the Eurasian Basin in the year 2012 compared to 2011 (Tab. 3).

The grouping of the samples (Fig. 8B) collected in 2011 was best explained by water masses (R = 0.35, *P* = 0.001; Tab. 2, Fig. 8B and Kilias et al. 2014). Furthermore, we achieved a weak but significant correlation of the protist communities with the sampling regions (R = 0.15, *P* = 0.05; Tab. 2, Fig. 8B, colors, Nansen, Amundsen, Makarov, Canada Basin). The Mantel test revealed no relation of the ARISA dataset with environmental parameters.

In analogy to 2011, protist community composition in water samples collected in 2012 were correlated to ambient water masses (R = 0.29, *P* = 0.001, Tab. 2, Fig. 8C). However, we observed an even stronger regional pattern for these samples. The non-metric cluster analysis revealed five distinct ARISA profiles (R = 0.53, *P* = 0.001; Tab. 2, Fig. 8C, colors). This regional distinction was even more pronounced, if we tested only the samples of the common sampling region (R = 0.71, *P* = 0.001, Tab. 2). The largest cluster contained samples collected in the Nansen Basin. Representative samples were characterized by a lower abundance of diatoms and Protalveolata (mainly Syndiniales) compared to the Amundsen Basin samples (Fig. 6). The Amundsen Basin was further separated in an eastern and western group. The eastern Amundsen group contained also samples collected in the Laptev Sea. However, five Laptev Sea samples located in the western parts showed different ARISA profiles and were comparatively variable among themselves, therefore forming an own cluster. Sequencing of station 311 in the Laptev Sea revealed the highest abundance of diatoms compared to all other water samples. Samples collected on the Svalbard continental slope clustered apart from all other groups, possibly due to higher abundances of Dinoflagellata and Chlorophyta as sequencing of sample 209 revealed. The regional pattern is best explained by ambient environmental conditions, as the clustering of the ARISA profiles was significantly correlated to the collected environmental parameters (Mantel test, r = 0.32, *P* = 0.001). The best fitting of physico-chemical parameters to the NMDS ordination was represented by sea ice

concentration ($r^2 = 0.65$, $P = 0.001$). Further significant correlations were achieved with water temperature ($r^2 = 0.40$, $P = 0.001$), followed by salinity ($r^2 = 0.32$, $P = 0.002$), concentrations of silicate ($r^2 = 0.24$, $P = 0.005$) and nitrate ($r^2 = 0.19$, $P = 0.032$).

DISCUSSION

Sea ice and water mass properties are of pivotal importance for the productivity of protists in the Central Arctic Ocean. Sea ice stimulates the productivity of pelagic protists while increasing the stratification of the surface water layer during the melting process (Eilertsen 1993; Andreassen et al. 1996; Fortiert et al. 2002; Perrette et al. 2011). In September 2011, Arctic sea ice covered a total area of 3.2 million km² whereas in the following summer period the total area was only 2.4 million km², nearly one million less than the year before (based on algorithm after Fetterer and Knowles 2002).

In this study, protist assemblages in Arctic sea ice and water samples collected in 2011 and 2012 during the same summer months (August - September) but with large differences in the sea ice concentrations were analyzed to examine a potential impact of the sea ice decline and sea ice origin on the community structure. We provide a broad overview on the impact of different sea ice conditions on Arctic protist communities originating from the two major habitats, sea ice and water column in different regions of the Arctic Ocean. The large sample size of 83 samples enabled us to detect potential relationships of the ARISA patterns and the physico-chemical parameters with a high confidence level.

Relying exclusively on ARISA profiles to infer ecological patterns is critical, because of limited taxonomic resolution (Bent et al. 2007; Caron et al. 2012, Kilias et al. 2015). Therefore, we used ARISA in combination with Illumina to elucidate community complexity by means of length-heterogeneity of ITS1 and protist community composition and diversity by means of taxonomic assignments of V4 SSU rDNA gene fragments.

Based on ARISA profiles and Illumina sequencing, we found distinct protist communities in sea ice and water which is in agreement with several publications reporting on typical taxa associated with sea ice or water column (Booth and Horner 1997; von Quillfeldt 2000; Ardyna et al. 2011; Arrigo et al. 2012; Poulin et al. 2011; Comeau et al. 2011 and 2013; Kilias et al. 2014). Overall, the sea ice community in 2011 was more diverse (i.e. higher number of fragments and OTUs) and spatially highly variable (high standard deviation) compared to the water community. The higher complexity and site-variability of sea ice communities might be linked to the variable environmental properties in the ice cores. Sea ice

and its environmental properties can be highly variable at small scales of even several meters (Eicken et al. 1991; Granskog et al. 2005). For example, sea ice properties, including algae biomass, of first-year land-fast sea ice in the Baltic Sea showed a horizontal patchiness at scales of >20 m (Granskog et al. 2005). Therefore, sea ice possesses a richness of microhabitats (Horner et al. 1992) being confined small-scale habitats, which can be inhabited by different protist communities. The microhabitats are characterized by several properties such as ice type (e.g. FYI, MYI), ice thickness and topography or snow and melt pond coverage. For example, MYI sites are characterized by thicker ice and a thicker snow depth; both strongly influence the primary production of sea ice algae by reducing the light availability (Arrigo et al. 2008; Popova et al. 2012; Nicolaus et al. 2012; Lange et al. 2015). In contrast, FYI is thinner, resulting in a higher availability of light, which triggers the photosynthetic production of sea ice algae especially at the sea ice bottom (Mundy et al. 2005; Lange et al. 2015). In our study, we observed high abundances of the ice diatoms *Nitzschia* spp., particularly in MYI. *Nitzschia* spp. are pennate diatoms that thrive in water under the ice or in the ice bottom (Horner et al. 1992, Michel et al. 2002, Kaartokallio et al. 2007). The higher abundance *Nitzschia* spp. in the bottom ice in 2011 compared to the entire sea ice core in 2012 emphasizes the importance of MYI and possibly points towards a reduction in ice algae diversity due to reduced diversity of microhabitats or even sea ice loss. This scenario could be one explanation for the observed lower sea ice diversity during record sea ice minimum in 2012 compared to 2011.

Impact of sea ice retreat and origin on Arctic sea ice protist communities

The ARISA-patterns revealed differences in ITS1 fragments of the sea ice communities in 2011 and 2012, while length-variability in sea ice protist community was higher in 2011 than in 2012. Illumina revealed a considerably lower number of OTUs of sea ice algae with low abundances (<1%) in 2012 compared to 2011. This was true, despite the higher ability to detect rare species in 2012, as we analyzed more samples and assessed the protist community in the entire sea ice cores and not only in the bottom ice section, as we did in 2011. One reason for the loss in rare sea ice algae might be the lower sea ice thickness in 2012. The physical properties of the sea ice determine not only the biodiversity within sea ice but also the interactions of the ice with the underlying water column. For example, FYI is more connected to the under ice water via brine channels than MYI, which leads to a higher exchange degree of nutrients and protists between the two habitats (Gradinger and Ikävalko 1998). In the NMDS-analyses performed for ARISA (Fig. 5), sea ice samples of 2012

clustered closer to the pelagic samples than those of 2011, indicating higher similarity between the protist communities of the two habitats in 2012. This might be attributed to an increased exchange of protists between thinning sea ice and water column during this year. Indeed, a higher contribution in terms of abundance and diversity of pelagic dinoflagellates to the sea ice biota in 2012, suggests a stronger exchange between sea ice and water column compared to 2011.

The sea ice cores collected in 2012 reflected not only differences in physical properties; they also differed in nutrient composition, which is known to impact protist biodiversity (Granskog et al. 2003; Gradinger 2009). The differences are probably related to a higher exchange of nutrients between sea ice and water column. For example, nitrate concentrations measured in the water column were on average higher in 2012 than in 2011 and the same trend was observed for nitrate measured in the sea ice. However, it is difficult to investigate patterns in sea ice nutrient profiles or ice algae community without knowing the entire history of the sea ice. Sea ice's nutrients profile reflect the initial water conditions during sea ice formation and change in the course of the seasons due to advection and biological processes (Gradinger and Ikävalko 1998; Granskog et al. 2003; Melnikov et al. 2003). Therefore, regions of sea ice origin may influence the nutrient profile of the ice and determine the protist community structure during the incorporation of particles and protists into newly formed sea ice (Ackley et al. 1987; Różańska et al. 2008). As the data on sea ice origin displayed, sea ice cores collected in 2011 originated from both, the Amerasian and the Eurasian Basin. The different environmental conditions of these basins were reflected in the nutrient profiles. Sea ice originating from the Amerasian Basin showed higher phosphate and silicate concentrations than those originating from the Eurasian Basin as has also been observed during the sea ice record low 2007 by Damm et al. (2010). Therefore, we observed a relatively high correlation of sea ice nutrients with the regions of sea ice origin in 2011.

ITS1 fragments abundance data of sea ice protists in 2011 were significantly correlated with the regions of sea ice origin. In contrast, the sampling area probably had a smaller effect on the protist composition, since the ARISA profiles differed between the Amerasian and Eurasian Basin but a distinction in all oceanographic basins (Canada, Makarov, and Amundsen Basin) was not observed. These findings indicate that the sea ice community was stronger influenced by environmental conditions during sea ice formation than by ambient environmental conditions during sampling. Also it is known that protists are entrapped in newly forming sea ice (Gradinger and Ikävalko 1998; Rozanska et al. 2008) and that winter communities are setting the stage for the sea ice algae bloom in spring (Niemi et

al. 2011), this study is the first that gives evidence for a possible influence of the regions of sea ice origin on the protist community pattern in Central Arctic sea ice.

A division of the sea ice nutrient profiles or of the protists based on sea ice origin was not observed for the samples collected in 2012. This could be explained by the fact that the ice floes originated mainly from nearby shelf regions in the Kara and Laptev Sea (Fig. 2). Here, similar water mass conditions were present during sea ice formation and hence, the effect was probably too weak to be detectable. Furthermore, we observed no effects of sampling regions on the sea ice community pattern. This could either indicate that there simply was no regional effect or that the sample size was too small to detect a possible effect. However, if testing the same number of water samples collected at the sea ice stations or close to the stations, we obtained a correlation coefficient of 0.44 and a *P*-value of 0.03 for the regional distinction in Nansen Basin, eastern and western Amundsen Basin. Therefore, if the protists in the sea ice would show strong regional differences (at least as strong as observed for the water samples), we would be able to detect them despite of the low sample number. Consequently, the sea ice community showed an oppositional response to the distinct environmental regimes within the Eurasian Basin than we observed for the water community during the sea ice record low in 2012. In addition, the distribution pattern of protists in the sea ice was more homogenous among the stations in 2012, maybe already reflecting the impact of fast receding sea ice and the decline of sea ice thickness on the loss of habitats and thus restructuring the sea ice communities in the Central Arctic Ocean.

Impact of sea ice retreat on Arctic pelagic protist communities

Water samples collected from the DCM (around 20 m) during record sea ice minimum year 2012 comprised a slightly higher number of ITS1 fragments than the water samples collected from the same depth range and summer months in 2011. However, this could be attributed to a larger sample size in 2012, as the possibility to find fragments increases with sample size (Jacob et al. 2013). We therefore reassessed our observation by comparing five nearby water stations located in the western Amundsen Basins sampled in 2011 (Stn. 205, 207, 209, 212, 216) and 2012 (Stn. 357, 370, 372, 377, 380). Here, the differences between the years were also apparent and even more pronounced despite the same sample size (on average 63 fragments in 2011, 83 fragments in 2012). This observation indicates a higher pelagic community complexity in 2012 in terms of protists ITS1 fragment lengths. However, in terms of 18S rRNA genes, Illumina sequencing revealed a lower protist diversity in the pelagial of 2012 than in 2011, despite we analyzed a larger sample size.

Diatoms in water samples collected in 2012 showed a higher contribution in terms of sequence abundance and OTU number to the pelagic community than in 2011, which might be one possible explanation for higher community heterogeneity. Light conditions in the pelagial were more favorable for diatoms in 2012, as the predominant ice type was FYI, which transmits more light compared to MYI (Nicolaus et al. 2012). In addition, since our and other studies observed high abundances of diatoms in sea ice (Booth and Horner 1997, von Quillfeldt 1997), ice diatoms could have been released from sea ice into the water column during ice break up and melt. This was also reported for several sea ice associated algae, such as the pennate diatoms *Nitzschia* spp. and *Navicula* spp. (von Quillfeldt 1997). A release of sea ice algae can alter the original composition of the water community, which in turn can have far-reaching implications for the food web structure and the carbon export in the Arctic Ocean. Several studies investigated the effects of sea ice melting and the consequent release of ice associated species into the water column (Melnikov 1997; Ambrose et al. 2005; Boetius et al. 2013). In particular, during the record sea ice minimum in 2012, high amounts of mainly fresh deposits of the sub-ice diatom *Melosira arctica* were observed on the sea floor at the same ice stations as we analyzed in this study (Boetius et al. 2013). Our study confirms these rare massive algae falls in 2012 because the contribution of *Melosira* spp. to the sea ice and especially to the pelagic community was considerably higher in summer 2012 than in 2011.

Pelagic protist communities in Polar Regions are strongly influenced by sea ice concentrations but also by water mass properties (Gradinger and Baumann 1991; Mostajir et al. 2001; Lovejoy et al. 2002; Mundy et al. 2005; Wolf et al. 2013; Kilias et al. 2013 and 2014). An important representative of the water column was the cold adapted chlorophyte *Micromonas*, which are typical pelagic picoplankton in the Arctic Ocean (Lovejoy et al. 2007, Kilias et al. 2014, Metfies et al. 2016). The distribution patterns of *Micromonas* were shown to correlate with Atlantic or Mix II water masses (Kilias et al. 2014). Therefore, the higher abundance of this genus in 2011 could indicate a higher contribution of these water masses in the Eurasian Basin compared to 2012. Sampling of both expeditions, took mainly place in the Eurasian Basin, however, in 2011, sampling was also conducted in the Amerasian Basins (eight stations). Both Arctic basins are known to exhibit characteristic environmental profiles. The Eurasian Basin is mainly influenced by relatively warm, saline and nutrient rich AW entering the Arctic Ocean via the deep Fram Strait or the shallow Barents Sea (Quadfasel 1987; Rudels et al. 1987). In contrast, the Amerasian Basin is characterized by relatively cold, less saline and eventually nutrient poor PW, which originated mainly from the through-flow of the Bering Strait (Swift et al. 1997). Both water masses merge and exit the Arctic Ocean

through the Fram Strait (Aagaard and Carmack 1989; Schauer 2004). Despite of these highly different environmental regimes, the pelagic protists community pattern and composition in the Amerasian Basin in 2011 was similar to those in the Eurasian Basin. A strong regional component was even not observed (Tab. 2), if we focused our comparison on the Eurasian Basin only. This is in contrast to previous studies, where the Amerasian Basin was characterized by flagellates, while the Eurasian basin had a higher proportion of diatoms (Damm et al. 2010). However, we observed a pronounced regional pattern for protists and environmental parameters in the Eurasian Basin in 2012. A highly significant correlation of the environmental variables with the fingerprinting pattern (Mantel test) for this year only, indicates a strong environmental forcing and underlines a possible impact of sea ice retreat on water column protists in the Eurasian Basin during record sea ice minimum in 2012.

Within the Eurasian Basin, we observed distinct pelagic communities separated in Nansen and Amundsen Basin. The Nansen Basin was characterized by high salinities and sea ice concentrations, as well as high nitrate concentrations, whereas the Amundsen Basin showed higher values of the remaining dissolved inorganic nutrients and higher water temperatures. In general, AW enters the Arctic Ocean through the Fram Strait and circulates mainly in the Nansen Basin, whereas the Amundsen Basin is influenced by modified AW of the Barents Sea inflow branch (Schauer et al. 2002; Rudels et al. 2013). This distinct pattern between the two basins could also be seen at higher trophic levels shown by David et al. (2015) who studied macrofaunal under-ice communities during the same cruise in 2012 (David et al. 2015). The authors also identified different environmental regimes in the Nansen and the Amundsen Basin which were characterized by different zooplankton communities and were able to demonstrate a tight pelagic coupling between primary and secondary producers.

Despite of the distinct Nansen and Amundsen regimes, we identified four stations located in the western Amundsen Basin, which showed a high similarity in their environmental characteristics and protist composition with the samples located in the Nansen Basin. This similarity suggests that the unmodified AW not only recirculates in the Nansen Basin (Schauer et al. 2002; Rudels et al. 2013) but also could reach the Amundsen Basin far beyond the Gakkel Ridge which previously was considered to be a natural boundary.

The AW inflow (either through the Fram Strait or through the Barents Sea) could also have influenced the protist community at the continental slope of Svalbard. Their ARISA profiles were clearly separated from the others as the community was influenced by open water with relatively high temperatures and low nutrient concentrations. A nutrient depletion could indicate a phytoplankton post bloom situation at these stations (Bienfang et al. 1982;

Waite et al. 1992; Arrigo et al. 2012). In the Laptev Sea, we observed a similar scenario where open water stations clustered distinct from sea ice stations in the Nansen or Amundsen Basin and were therefore characterized by a different community structure. At these stations, the silicate concentrations exceeded the concentrations measured in the other sampling areas, indicating the strong impact of freshwater from the Lena River (Kattner et al. 1999). The Laptev Sea is known as a highly productive shelf area during summer and fall and inhabited by large proportions of diatoms (Heiskanen and Keck 1996; Kattner et al. 1999; Tuschling et al. 2000) as exemplified by station 311 in our study. The unique setting of the Laptev Sea compared to the Central Arctic is in agreement with a recent study which investigated higher under-ice export fluxes of biogenic matter in the arctic shelf regions, especially in the northern Laptev Sea, compared to the export over the Central Arctic Ocean (Lalande et al. 2014). Furthermore, our results suggest that the Laptev Sea could be subdivided into two different regimes. Based on the NMDS clustering and the environmental characteristics, the samples taken in the Laptev Sea were distinguished in an eastern and a western group (latter were related to the Amundsen Basin-group). This subdivision of the protist community in the Laptev Sea could indicate an influence of river inflow or the effects of confluence and strong mixing of the Fram Strait and Barents Sea branches as described by Rudels et al. (2013). Our observations give evidence that the regions at continental slopes are highly variable areas, which might be affected differently by sea ice retreat in the Arctic Ocean, since they owned distinct protist community structures.

CONCLUSIONS

The results of our large-scale study indicate that sea ice and water column harbor distinct protist communities, with partial exchange and a strong dependency on sea ice and water mass properties. We identified habitat-specific fingerprints, which differed strongly between two years with contrastive sea ice concentrations. Based on ARISA analysis and Illumina sequencing, we observed that protist community complexity and diversity changed between the years. We observed a possible trend towards less complex and diverse communities in both habitats due to sea ice melting. This trend serves as a future scenario for protist communities in the changing Arctic. We conclude that sea ice is a major driver of protist community structure in both, sea ice and water column and that extensive melting of sea ice probably favors the input of sea ice species into the water column where they might alter the community composition with consequences for other trophic levels. Furthermore, we

recognized that sea ice protist biodiversity is determined by sea ice origin. Protist communities in sea ice were mainly structured by environmental conditions several thousand kilometers away during sea ice formation. However, if the sea ice communities on the shelves will still serve as basis for sea ice biodiversity in a changing Arctic with less sea ice is questionable. Particular, recent shifts to a later freeze-up in the winter with less active biota being present to freeze-in suggest a reduction of sea ice biodiversity in the Central Arctic. Overall, sea ice concentration and sea ice origin are major drivers of Arctic protist community composition.

ACKNOWLEDGMENTS

This study was accomplished within the Young Investigator Group PLANKTOSENS (VH-NG-500), funded by the Initiative and Networking Fund of the Helmholtz Association. We thank the captain and crew of the RV Polarstern for their support during the cruise ARKXXVI/3 and ARKXXVII/3. We are grateful to Dr. E.M. Nöthig for valuable comments to the manuscript, G. Kattner for providing unpublished dissolved nutrient data in water and sea ice and the Meereisportal for sea ice concentration data. We thank A. Nicolaus, K. Oetjen and Lea Weinisch (University of Kaiserslautern) for excellent technical support in the laboratory.

LITERATURE CITED

Aagaard K, Carmack EC (1989) The role of sea ice and other fresh waters in the Arctic circulation. J Geophys Res 94:14485–14498

Ackley SF, Dieckmann GS, Shen HT (1987) Algal and foram incorporation into new sea ice. Trans Am Geophys Union 68:1736–1748

Ambrose WG Jr, von Quillfeldt C, Clough LM, Tilney PVR, Tucker T (2005) The sub-ice algal community in the Chukchi sea: large- and small-scale patterns of abundance based on images from a remotely operated vehicle. Polar Biol 28:1–12

Andreassen I, Nöthig E-M, Wassmann P (1996) Vertical particle flux on the shelf of northern Spitsbergen Norway. Mar Ecol Prog Ser 137:215–228

Ardyna M, Gosselin M, Michel C, Poulin M, Tremblay J-E (2011) Environmental forcing of phytoplankton community structure and function in the Canadian High Arctic: contrasting oligotrophic and eutrophic regions. Mar Ecol Prog Ser 442:37–57, doi:103354/meps09378

Arrigo KR, van Dijken G, Pabi S (2008) Impact of a shrinking Arctic ice cover on marine primary production. Geophys Res Lett 35 L19603, doi:101029/2008GL035028

Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE, Mills MM, Palmer MA, Balch WM, Bahr F, Bates NR, Benitez-Nelson C, Bowler B, Brownlee E, Ehn JK, Frey KE, Garley R, Laney SR, Lubelczyk L, Mathis J, Matsuoka A, Mitchell BG, Moore GWK, Ortega-Retuerta E, Pal S, Polashenski CM, Reynolds RA, Schieber B, Sosik HM, Stephens M, Swift JH (2012) Massive phytoplankton blooms under Arctic sea ice. Science 336:1408, doi:101126/science1215065

Arrigo KR (2014) Sea ice ecosystems. Ann Rev Mar Sci 6:439–467, doi:101146/annurev-marine-010213-135103

Bakker K (2014) Nutrients measured on water bottle samples during POLARSTERN cruise ARK-XXVII/3 (IceArc) in 2012. unpublished dataset #834081 Texel

Bent SJ, Peirson JD, Forney LJ (2007) Measuring species richness based on microbial community fingerprints: The emperor has no clothes. Appl Environ Microbiol 73:2399–2401

Beszczynska-Möller A, Fahrbach E, Schauer U, Hansen E (2012) Variability in Atlantic Water temperature and transport at the entrance to the Arctic Ocean 1997-2010. ICES J Mar Sci 69:852–863

Bienfang PK, Harrison PJ, Quarmby LM (1982) Sinking rate response to depletion of nitrate phosphate and silicate in four marine diatoms. Mar Biol 67:295–302

Bienhold C, Boetius A, Ramette A (2012) The energy-diversity relationship of complex bacterial communities in Arctic deep-sea sediments. ISME J 6:724–732

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 10:57–59

Boetius A, Albrecht S, Bakker K, Bienhold C, Felden J, Fernández-Méndez M, Hendricks S, Katlein C, Lalande C, Krumpen T, Nicolaus M, Peeken I, Rabe B, Rogacheva A, Rybakova E, Somavilla R, Wenzhöfer F, RV Polarstern ARK27-3 Shipboard Science Party (2013) Export of algal biomass from the melting Arctic sea ice Science 339:1430–1432, doi:101126/science1231346

Booth BC, Horner RA (1997) Microalgae on the Arctic Ocean section 1994: Species abundance and biomass. Deep Sea Res Part II 44:1607–1622

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012) Marine protistan diversity. Ann Rev Mar Sci 4:467–493

Comeau AM, Li WKW, Tremblay J-AR, Carmack EC, Lovejoy C (2011) Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLoS ONE 6:e27492, doi:101371/journalpone0027492

Comeau AM, Philippe B, Thaler M, Gosselin M, Poulin M, Lovejoy C (2013) Protists in Arctic drift and land-fast sea ice. J Phycol 49:229–240, doi:101111/jpy12026

Comiso JC (2003) Warming trends in the Arctic from clear sky satellite observations. J Clim 16:3498–3510

Comiso JC, Parkinson CL, Gersten R, Stock L (2008) Accelerated decline in the Arctic sea ice cover. Geophys Res Lett 35:L01703, doi:101029/2007GL031972

Comiso JC (2012) Large decadal decline of the Arctic multiyear ice cover. J Clim 25:1176–1193

Damm E, Helmke E, Thoms S, Schauer U, Nothig E, Bakker K, Kiene RP (2010) Methane production in aerobic oligotrophic surface water in the central Arctic Ocean. Biogeosciences 7:1099–1108

David C, Lange B, Rabe B, Flores H (2015) Community structure of under-ice fauna in the Eurasian central Arctic Ocean in relation to environmental properties of sea–ice habitats. Mar Ecol Prog Ser 522:15–32

Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7(10), e1002195.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinform 26:2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.

Eicken H, Lange MA, Dieckmann GS (1991) Spatial variability of sea-ice properties in the northwestern Weddell Sea. J Geophys Res 96:10603–10615

Eilertsen HC (1993) Spring blooms and stratification. Nature 363:24

Fechner LC, Vincent-Hubert F, Gaubert P, Bouchez T, Gourlay-France C, Tusseau-Vuillemin MH (2010) Combined eukaryotic and bacterial community fingerprinting of natural freshwater biofilms using automated ribosomal intergenic spacer analysis. FEMS Microbiol Ecol 74:542–553

Fernández-Méndez M, Rabe B, Katlein C, Nicolaus M, Peeken I, Flores H, Boetius A (2015) Photosynthetic production in the Central Arctic during the record sea-ice minimum in 2012. Biogeosci Discuss 12:2897–2945, doi:105194/bgd-12-2897-2015

Fetterer F, Knowles K (2002) Sea ice index digital media Natl Snow and Ice Data Cent Boulder Colorado

Fortier M, Fortier L, Michel C, Legendre L (2002) Climatic and biological forcing of the vertical flux of biogenic particles under seasonal Arctic sea ice. Mar Ecol Prog Ser 225:1–16

Gobet A, Boetius A, Ramette A (2014) Ecological coherence of diversity patterns derived from classical fingerprinting and Next Generation Sequencing techniques. Environ Microbiol 16:2672–2681

Gradinger RR, Baumann MEM (1991) Distribution of phytoplankton communities in relation to the large-scale hydrographical regime in the Fram Strait. Mar Biol 111:311–32

Gradinger R, Ikävalko J (1998) Organism incorporation into newly forming Arctic sea ice in the Greenland Sea. J Plankton Res 20:871–886

Gradinger R (2009) Sea ice algae: Major contributors to primary production and algal biomass in the Chukchi and Beaufort Sea during May/June 2002. Deep Sea Res Part II 56:1201–1212, doi:101016/jdsr2200810016

Granskog MA, Kaartokallio H, Shirasawa K (2003) Nutrient status of Baltic Sea ice: evidence for control by snow-ice formation ice permeability and ice algae. J Geophys Res 108:3253, doi:101029/2002JC001386

Granskog MA, Kaartokallio H, Kuosa H, Thomas D, Ehn J, Sonninen E (2005) Scales of horizontal patchiness in chlorophyll a chemical and physical properties of landfast sea ice in the Gulf of Finland (Baltic Sea). Polar Biol 28:276–283

Hammer Ø, Harper DAT, Ryan PD (2001) Past: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4:9

Hegseth EN, Sundfjord A (2008) Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. J Mar Syst 74:108–119

Heiskanen AS, Keck A (1996) Distribution and sinking rates of phytoplankton detritus and particulate biogenic silica in the Laptev Sea and Lena River (Arctic Siberia). Mar Chem 53:229–245

Holliday NP, Hughes SL, Beszczynska-Möller A (2009) ICES Report on Ocean Climate 2008. ICES Cooperative Research Report 298 p 66

Horner R (1985) Ecology of sea ice microalgae. In: Horner R (ed) Sea ice biota. CRC Fla p 83–103

Horner RS, Ackley F, Dieckmann GS, Gulliksen B, Hoshiai T, Legendre L, Melnikov IA, Reeburgh WS, Spindler M, Sullivan CW (1992) Ecology of sea ice biota 1: Habitat terminology and methodology. Pol Biol 12:417–427

Jacob M, Soltwedel T, Boetius A, Ramette A (2013) Biogeography of deep-sea benthic bacteria at regional scale (LTER HAUSGARTEN Fram Strait Arctic). PLoS ONE 8:e72779, 101371/journalpone0072779

Juul-Pedersen T, Michel C, Gosselin M, Seuthe L (2008) Seasonal changes in the sinking export of particulate material under first-year sea ice on the Mackenzie Shelf (western Canadian Arctic). Mar Ecol Prog Ser 353:13–25

Kaartokallio H, Kuosa H, Thomas DN, Granskog MA, Kivi K (2007) Biomass composition and activity of organism assemblages along a salinity gradient in sea ice subjected to river discharge in the Baltic Sea. Polar Biol 30:183–197

Kattner G, Becker H (1991) Nutrients and organic nitrogenous compounds in the marginal ice zone of the Fram Strait. J Mar Syst 2:385–394

Kattner G, Lobbes JM, Fitznar HP, Engbrodt R, Nothig EM, Lara RJ (1999) Tracing dissolved organic substances and nutrients from the Lena River through Laptev Sea (Arctic). Mar Chem 65:25–39

Kattner G, Ludwichowski K-U (2014) Inorganic nutrients measured on water bottle samples during POLARSTERN cruise ARK–XXVI/3 (TransArc) Alfred Wegener Institute

Helmholtz Center for Polar and Marine Research Bremerhaven, doi:101594/PANGAEA832164

Kerouel R, Aminot A 1997 Fluorometric determination of ammonia in sea and estuarine waters by direct segmented flow analysis. Mar Chem 57:265–275, doi:101016/s0304-4203(97)00040-6

Kilias E, Wolf C, Nöthig EM, Peeken I, Metfies K (2013) Protist distribution in the Western Fram Strait in summer 2010 based on 454–pyrosequencing of 18S rDNA. J Phycol 49:996–1010

Kilias E, Kattner G, Wolf C, Frickenhaus S, Metfies K (2014) A molecular survey of protist diversity through the central Arctic Ocean. Polar Biol 37:1271–1287, doi:101007/s00300-014-1519-5

Kilias E, Wolf C, Metfies K (2015) Characterizing variability in marine protist communities via ARISA fingerprints - a method evaluation. Limnol Oceanogr: Methods 13:74–80, doi:101002/lom310008

Krumpen T, Janout M, Hodges KI, Gerdes R, Girard-Ardhuin F, Hölemann JA, Willmes S (2013) Variability and trends in Laptev Sea ice outflow between 1992-2011. The Cryosphere 7:349–363, doi:105194/tc-7-349-2013

Krumpen T, Gerdes R, Haas S, Hendricks S, Herber A, Selyuzhenok V, Smedsrud L, Spreen G (2015) Recent summer sea ice thickness surveys in the Fram Strait and associated volume fluxes. The Cryosphere Discuss 9:5171–5202, doi:105194/tcd-9-5171-2015

Lalande C, Nöthig E-M, Somavilla R,Bauerfeind E, Shevchenko V, Okolodkov Y (2014) Variability in under-ice export fluxes of biogenic matter in the Arctic Ocean. Global Biogeochem Cycles 28:571–583, doi:101002/2013GB004735

Lange BA, Michel C, Beckers JF, Casey JA, Flores H, Hatam I, Meisterhans G, Niemi A, Haas C (2015) Comparing Springtime Ice-Algal Chlorophyll a and Physical Properties of Multi-Year and First-Year Sea Ice from the Lincoln Sea. PLoS ONE 10:e0122418, doi:101371/journalpone0122418

Lee SH, McRoy CP, Joo Lee HM, Gradinger R, Cui XH, Yun MS, Chung KH, Kang SH, Kang CK, Choy EJ, Son SH, Carmack E, Whitledge TE (2011) Holes in progressively thinning Arctic sea ice lead to new ice algae habitat. Oceanography 24:302–308

Lovejoy C, Legendre L, Martineau MJ, Bacle J, von Quillfeldt CH (2002) Distribution of phytoplankton and other protists in the North Water. Deep Sea Res Part II Top Stud Oceanogr 49:5027–5047

Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, Potvin M, Massana R, Pedros-Alio C (2007) Distribution phylogeny and growth of cold-adapted picoprasinophytes in arctic seas. J Phycol 43:78–89, doi:101111/j1529–8817200600310

Maslanik JA, Fowler C, Stroeve J, Drobot S, Zwally HJ, Yi D, Emery WJ (2007) A younger thinner ice cover: increased potential for rapid extensive ice loss. Geophys Res Lett 34:L24501, doi:101029/2007GL032043

Maslanik J, Stroeve J, Fowler C, Emery W (2011) Distribution and trends in Arctic sea ice age through spring 2011. Geophys Res Lett 38:L13502, doi:101029/2011GL047735

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S like rRNA coding regions. Gene 71:491–499

Meier WN, Hovelsrud GK, van Oort BEH, Key JR, Kovacs KM, Michel C, Haas C, Granskog MA, Gerland S, Perovich DK, Makshtas A, Reist JD (2014) Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. Rev Geophys 52:185–217, doi:101002/2013RG000431

Melnikov IA (1997) The Arctic sea ice ecosystem. Amsterdam: Gordon and Breach Science Publishers p 204

Melnikov IA, Korneeva GA, Zhitina LS, Shanin SS (2003) Dynamics of ecological-biochemical characteristics of sea ice in coastal waters of the White Sea. Biol Bull 30:164–170

Metfies K, von Appen W-J, Kilias E, Nicolaus A, Nöthig E-M (2016) Biogeography and Photosynthetic Biomass of Arctic Marine Pico-Eukaroytes during Summer of the Record Sea Ice Minimum 2012. PLoS ONE 11:e0148512, doi:101371/journalpone0148512

Michel C, Gissel Nielsen T, Nozais C, Gosselin M (2002) Significance of sedimentation and grazing by ice micro- and meiofauna for carbon cycling in annual sea ice (northern Baffin Bay). Aquat Microb Ecol 30:57–68

Miller LA, Fripiat F, Else BGT, Bowman JS, Brown KA, Collins RE, Ewert M, Fransson A, Gosselin M, Lannuzel D, Meiners KM, Michel C, Nishioka J, Nomura D, Papadimitriou S, Russell LM, Sørensen LL, Thomas DN, Tison J-L, Van Leeuwe MA, Vancoppenolle M, Wolff EW, Zhou J (2015) Methods for biogeo- chemical studies of sea ice: the state of the art caveats and recommendations. Elem Sci Anthropol 3:1–53, doi:1012952/journalelementa000038

Moran SB, Lomas MW, Kelly RP, Gradinger R, Iken K, Mathis JT (2012) Seasonal succession of net primary productivity particulate organic carbon export and autotrophic community composition in the eastern Bering Sea. Deep Sea Res Part II 65:84–97

Mostajir B, Gosselin M, Gratton Y, Booth B, Vasseur C, Garneau M-E, Fouillard E, Vidussi F, Demers S (2001) Surface water distribution of pico- and nanophytoplankton in relation to two distinctive water masses in the North Water northern Baffin Bay during fall. Aquat Microb Ecol 23:205–212

Mundy CJ, Barber DG, Michel C (2005) Variability of snow and ice thermal physical and optical properties pertinent to sea ice algae biomass during spring. J Mar Syst 58:107–120

Nicolaus M, Katlein C, Maslanik J, Hendricks S (2012) Changes in Arctic sea ice result in increasing light transmittance and absorption. Geophys Res Lett 39:L24501, doi:101029/2012GL053738

Niemi A, Michel C, Hille K, Poulin M (2011) Protist assemblages in winter sea ice: Setting the stage for the spring ice algal bloom. Polar Biol 34:1803–1817

Parkinson CL, Comiso JC (2013) On the 2012 record low Arctic sea ice cover: combined impact of preconditioning and an August storm. Geophys Res Lett 40:1356–1361

Perovich DK, Polashenski C (2012) Albedo evolution of seasonal Arctic sea ice. Geophys Res Lett 39:L08501, doi:101029/2012GL051432

Perrette M, Yool A, Quartly GD, Popova EE (2011) Near–ubiquity of ice–edge blooms in the Arctic. Biogeosciences 8:515–524

Poulin M, Daugbjerg N, Gradinger R, Ilyash L, Ratkova T, von Quillfeldt C (2011) The pan Arctic biodiversity of marine pelagic and sea-ice unicellular eukaryotes: A first-attempt assessment. Mar Biodiv 41:13–28

Popova EE, Yool A, Coward AC, Dupont F, Deal C, Elliott S, Hunke E, Jin MB, Steele M, Zhang JL (2012) What controls primary production in the Arctic Ocean? Results from an intercomparison of five general circulation models with biogeochemistry. J Geophys Res 117:C00D12, doi:101029/2011JC007112

Quadfasel D, Gascard JC, Koltermann KP (1987) Large scale oceanography in Fram Strait during MIZEX 84. J Geophys Res 92:6719–6728

Quillfeldt CH von 2000 Common diatom species in arctic spring blooms: their distribution and abundance. Mar Bot 43:499–516

R Development Core Team (2008) R: a language and environment for statistical computing R Foundation for Statistical ComputingVienna

Ramette A 2009 Quantitative community fingerprinting methods for estimating the abundance of operational taxonomic units in natural microbial communities. Appl Environ Microbiol 75:2495–2505, doi:101128/AEM02409-08

Rudels B 1987 On the mass balance of the Polar Ocean with special emphasis on the Fram Strait Norsk Polarinstitutt Skrifter 188 p 53

Rudels B, Schauer U, Björk G, Korhonen M, Pisarev S, Rabe B, Wisotzki A (2013) Observations of water masses and circulation with focus on the Eurasian Basin of the Arctic Ocean from the 1990s to the late 2000s. Ocean Sci 9:147–169, doi:105194/os-9-147-2013

Rózańska M, Poulin M, Gosselin M (2008) Protist entrapment in newly formed sea ice in the coastal Arctic Ocean. J Mar Syst 74:887–901

Schauer U, Rudels B, Jones EP, Anderson LG, Muench RD, Björk G, Swift JH, Ivanov V, Larsson AM (2002) Confluence and redistribution of Atlantic water in the Nansen Amundsen and Makarov basins. Ann Geophys 20:257–273

Schauer U, Fahrbach E, Osterhus S, Rohardt G (2004) Arctic warming through the Fram Strait: Oceanic heat transport from 3 years of measurements. J Geophys Res 109:C06026, doi:101029/ 2003JC001823

Serreze MC, Walsh JE, Chapin III FS, Osterkamp T, Dyurgerov M, Romanovsky V, Oechel WC, Morison J, Zhang T, Barry RG (2000) Observational evidence of recent change in the northern high latitude environment. Clim Change 46:159–207

Spreen G, Kaleschke L, Heygster G (2008) Sea ice remote sensing using AMSR-E 89 GHz channels. J Geophys Res 113:C02S03, doi:101029/2005JC003384

Stecher A, Neuhaus S, Lange B, Frickenhaus S, Beszteri B, Kroth PG, Valentin K (2015) rRNA and rDNA based assessment of sea ice protist biodiversity from the central Arctic Ocean. Eur J Phycol, doi:10.1080/09670262.2015.1077395

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner H, Richard TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19:21–31

Stroeve JC, Maslanik J, Serreze M, Rigor CI, Meier W, Fowler C (2011) Sea ice response to an extreme negative phase of the Arctic Oscillation during winter 2009/2010. Geophys Res Lett 38:L02502, doi:101029/2010GL045662

Stroeve JC, Serreze MC, Holland MM, Kay JE, Masklanik J, Barrett AP (2012): The Arctic's rapidly shrinking sea ice cover: a research synthesis. Clim Change 110:1005–1027

Stroeve JC, Markus T, Boisvert L, Miller J, Barrett A (2014) Changes in Arctic melt season and implications for sea ice loss. Geophys Res Lett 41:1216–1225, doi:101002/(2013)GL058951

Swift JH, Jones EP, Carmack EC, Hingston M, Macdonald RW, McLaughlin FA, Perkin RG (1997) Waters of the Makarov and Canada Basins. Deep-Sea Res Part II 44:1503–1529

Tuschling K, Juterzenka KV, Okolodkov YB, Anoshkin A (2000) Composition and distribution of the pelagic and sympagic algal assemblages in the Laptev Sea during autumnal freeze-up. J Plankt Res 22: 843–864

Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, Kilpert F, Worden AZ, Giovannoni J (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. ISME J 7:1322–1332

Waite AN, Bienfang PK, Harrison PJ (1992) Spring bloom sedimentation in a subarctic ecosystem: I Nutrient sensitivity. Mar Biol 114:119–129

Wang L, Derksen C, Brown R, Markus T (2013) Recent changes in pan-Arctic melt onset from satellite passive microwave measurements. Geophys Res Lett 40:522–528, doi:101002/grl50098

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed) PCR protocols: a guide to methods and applications. Academic Press New York pp 315–322

Wolf C, Frickenhaus S, Kilias ES, Peeken I, Metfies K (2013) Regional variability in eukaryotic protist communities in the Amundsen Sea. Antarctic Science 25:741–751

Zhang J, Lindsay R, Schweiger A, Steele M (2013) The impact of an intense summer cyclone on 2012 Arctic sea ice retreat. Geophys Res Lett 40:720–726, doi:101002/grl50190

Figures

Fig. 1. Station maps of RV Polarstern expeditions to the central Arctic. **A.** Stations sampled during ARK XXVI/3 in 2011 (red dots). **B.** Stations sampled during ARK XXVII/3 in 2012 (black dots). Water samples were taken at each station from the chlorophyll maximum depth. Red marked station numbers indicate sampling of sea ice cores. Sea ice stations were numbered individually in 2012. Sea ice concentrations are given for the day of sea ice minimum extent on 09.09.2011 and 16.09.2012, respectively (data from meereisportal.de). **C.** Overview of expeditions and oceanography of the sampling area. The white box marks the common sampling area of the two expeditions. Red arrows: Atlantic Water, blue arrows: Pacific Water. The Eurasian Basin is subdivided into Nansen Basin (NB) and Amundsen Basin (AB). The Amerasian Basin is subdivided into Makarov Basin (MB) and Canada Basin (CB).

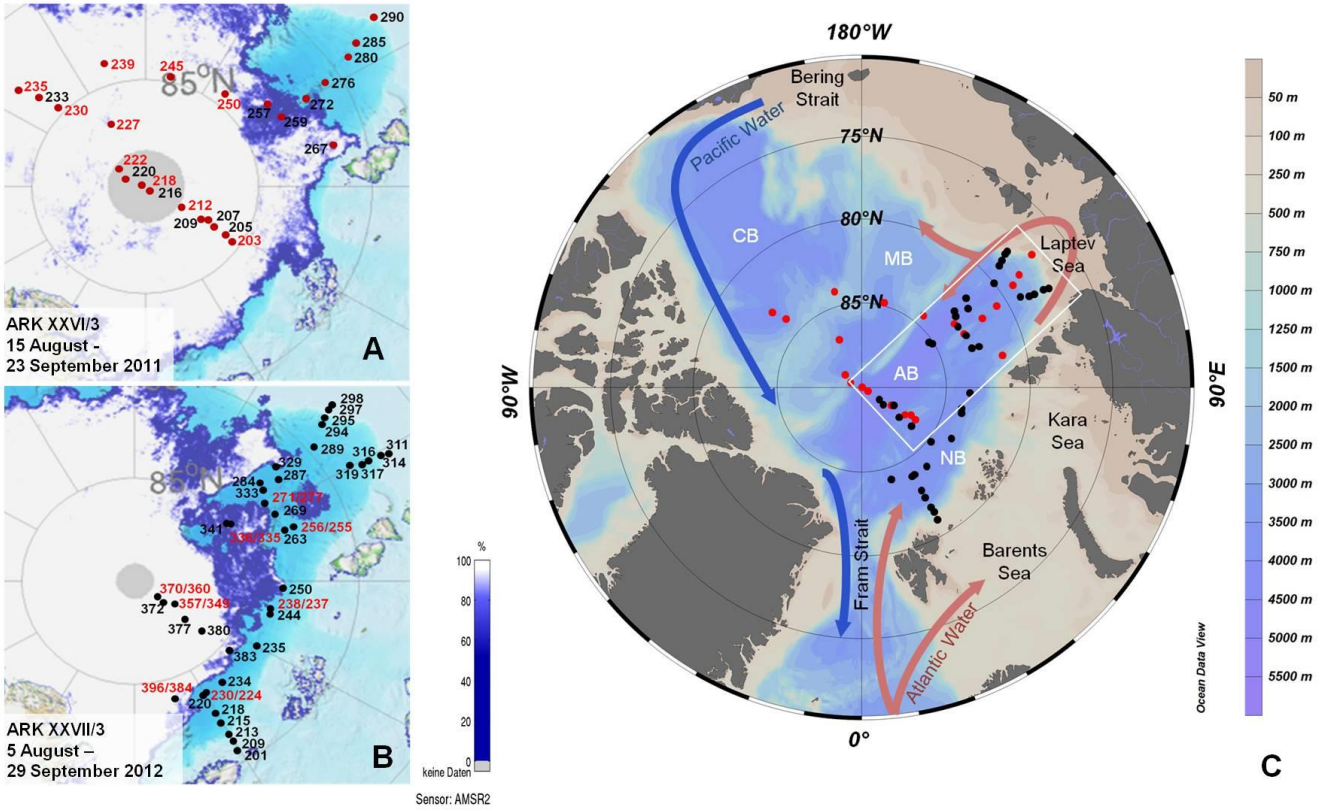


Fig. 2. Sea ice drift information of sea ice cores collected in 2011 (black lines) and 2012 (blue lines). Data obtained from CERSAT/IFREMER data and NSIDC. The black dots refer to the date of ice formation (September 21st).

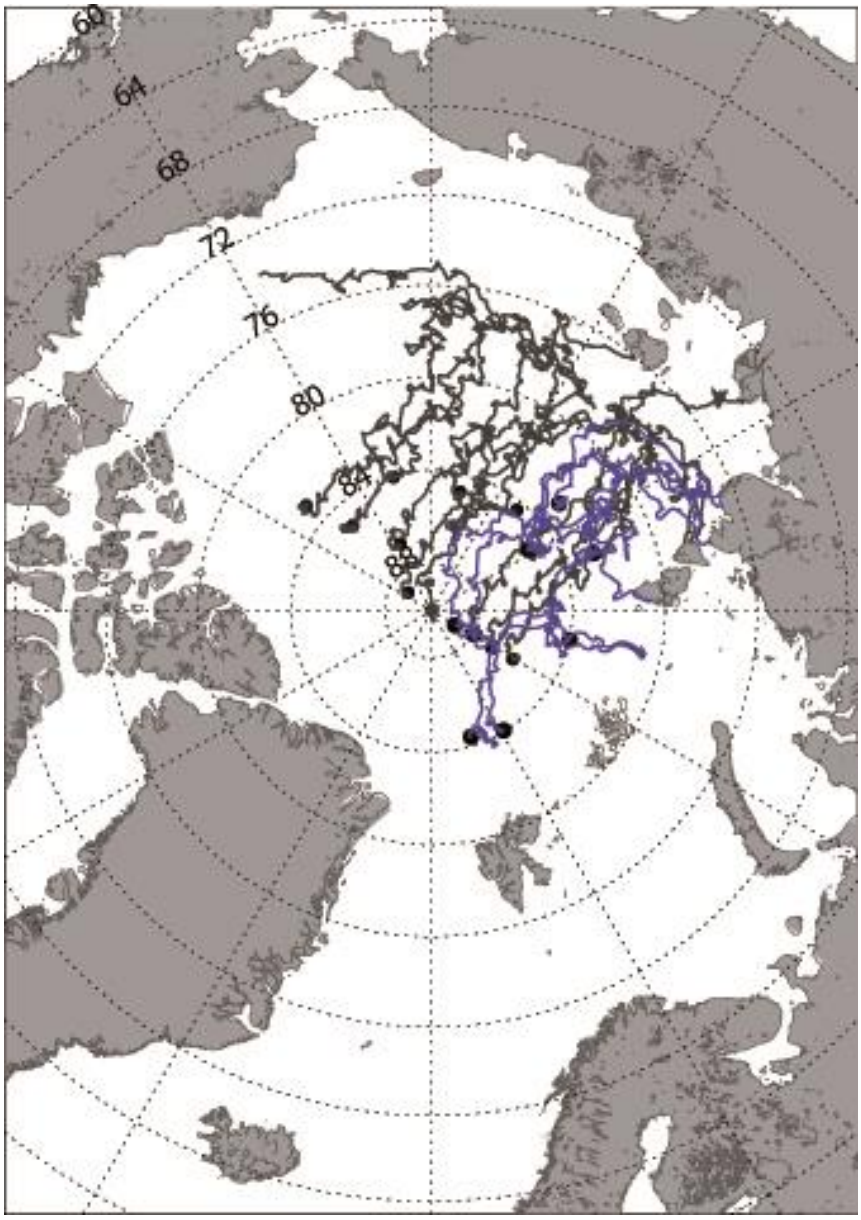


Fig. 3. PCA of environmental parameters for ice cores. **A**. Sea ice samples collected in 2011. **B**. Sea ice samples collected in 2012. Color and symbol codes are given in the legend of (A).

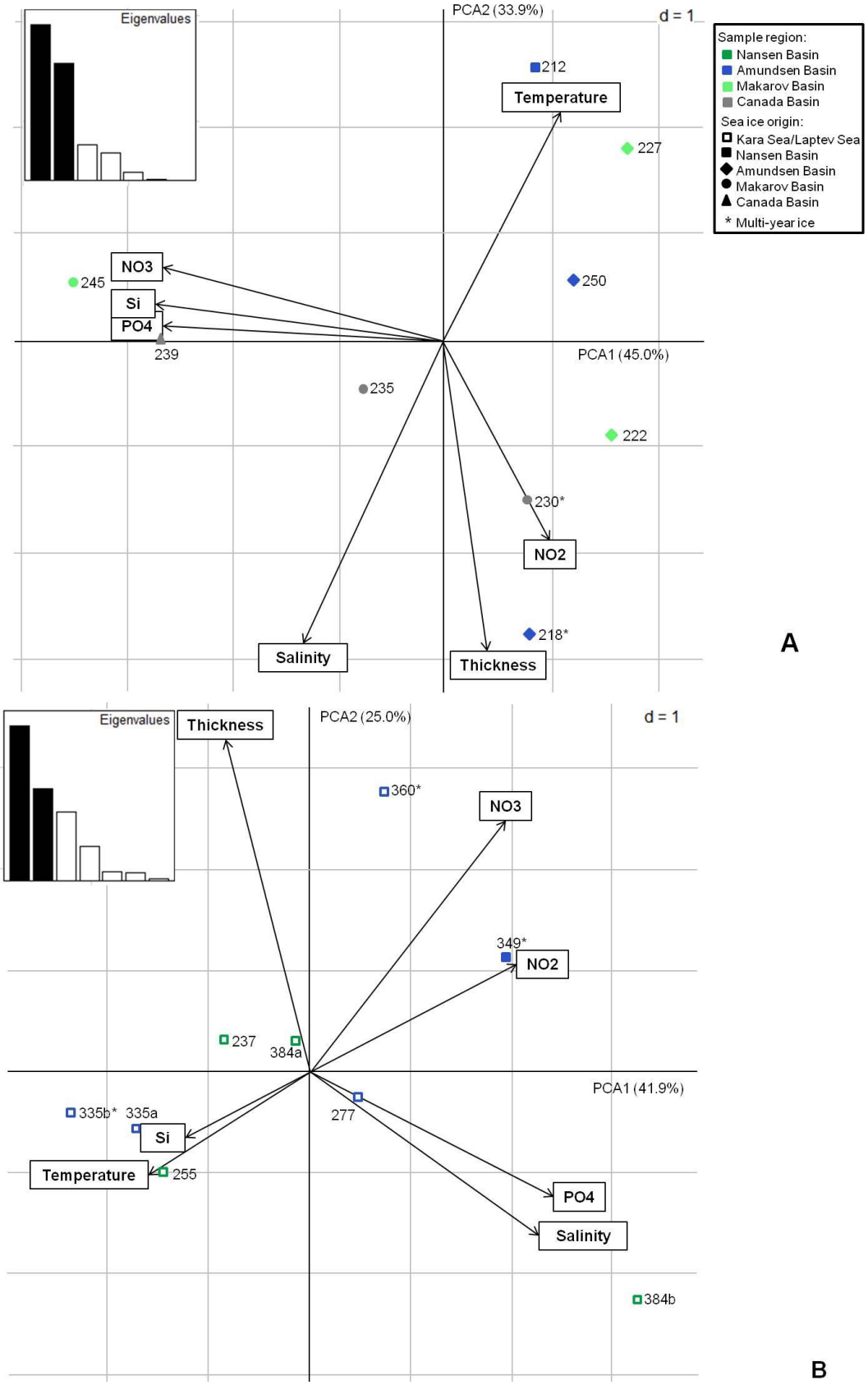


Fig. 4. PCA of environmental parameters for water samples. **A**. Water samples collected in 2011. **B**. Water samples collected in 2012. Color and symbol codes are given in the legend of (A). Ice conc., ice concentration.

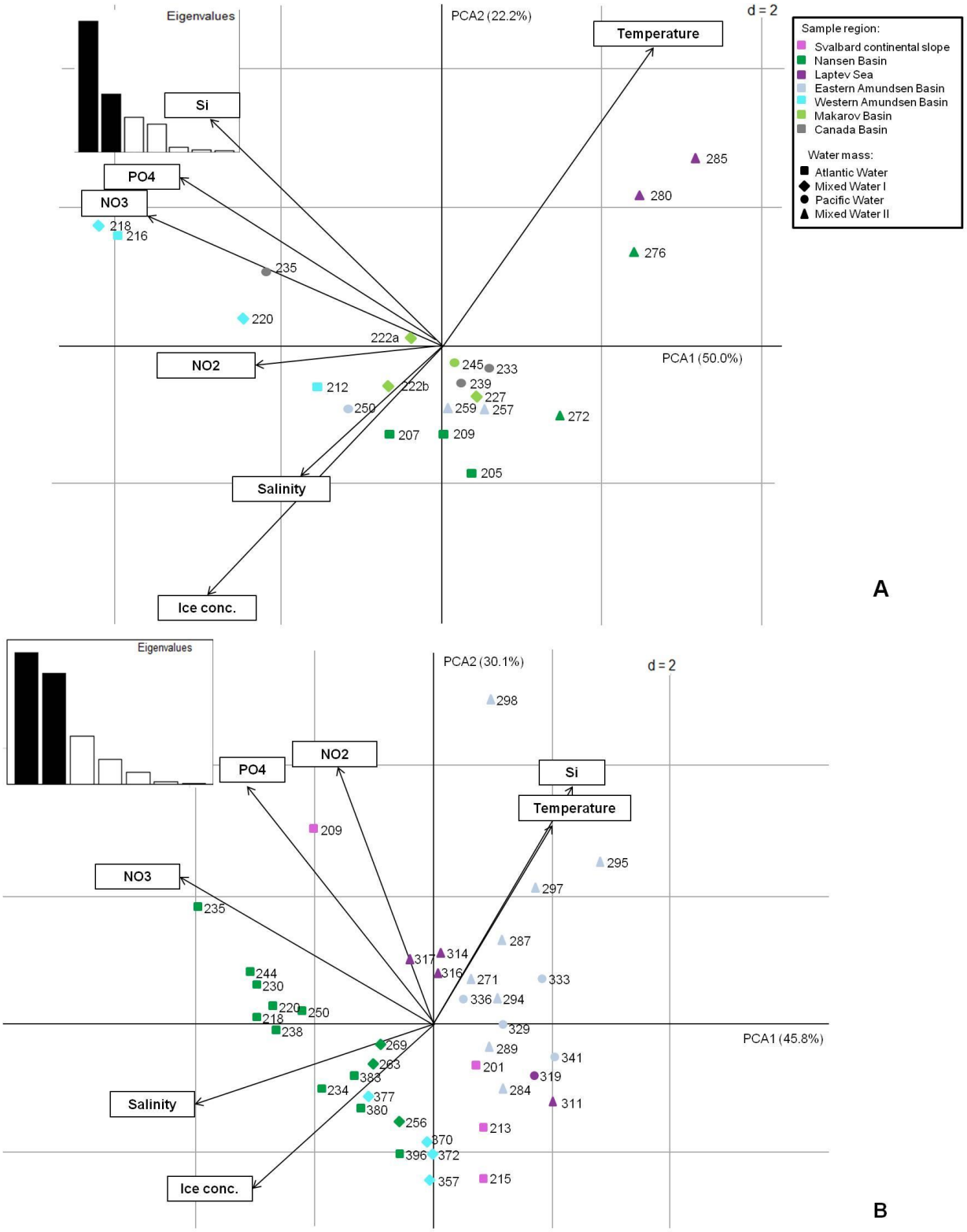


Fig. 5. NMDS plot based on Jaccard index of water and sea ice samples collected in 2011 and 2012. **A.** ARISA fragment profiles. **B**. Illumina sequences of water and sea ice samples. Color code is given in the legend of graph (A).

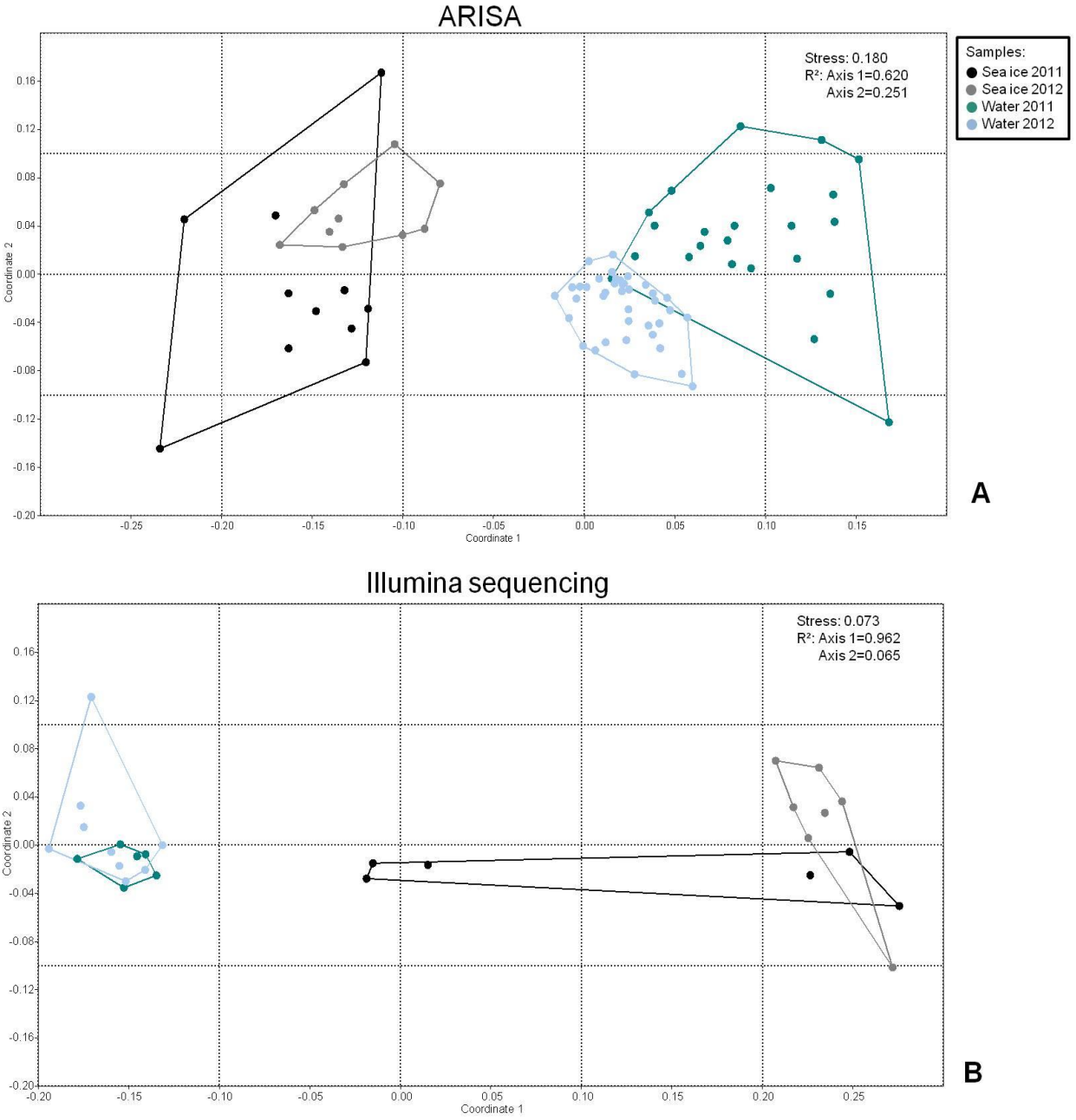


Fig. 6. Protist community composition obtained with sequencing. Stacked bar charts of relative sequence abundance of protists in sea ice (i; fyi, first-year ice; myi, multi-year ice) and water (w) samples collected during TransArc expedition in 2011 (PS78) and IceArc expedition in 2012 (PS80).

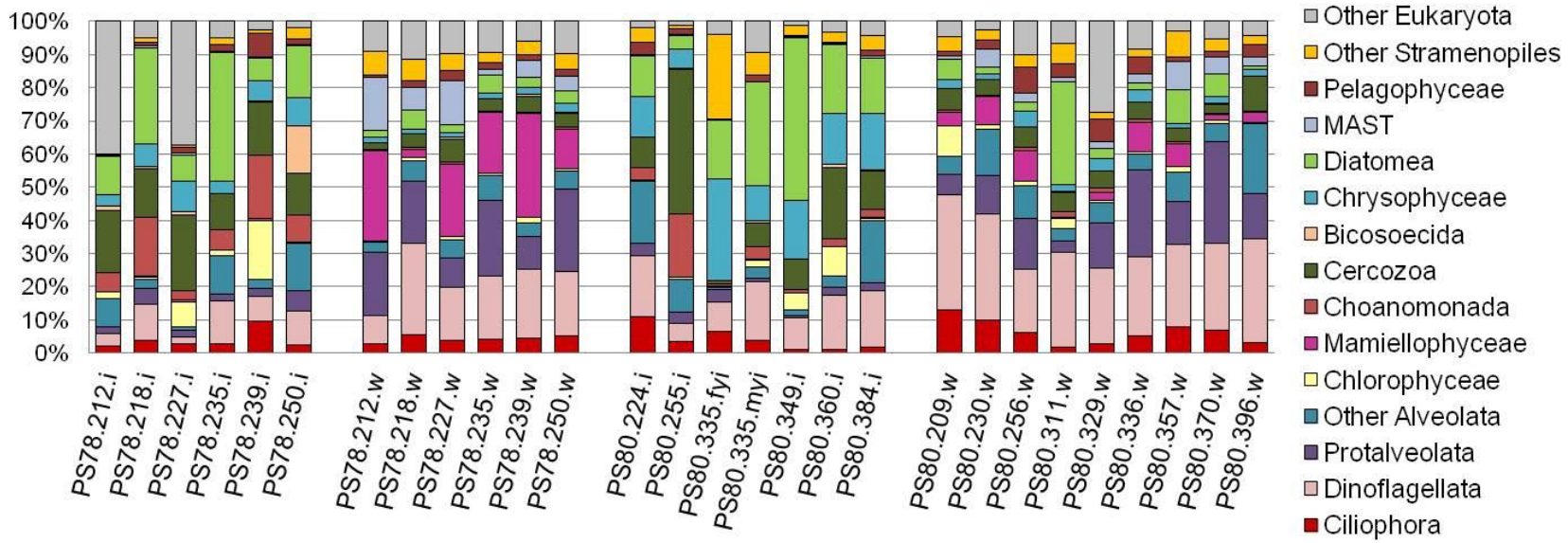


Fig. 7. NMDS plots of the ARISA fragment profiles of sea ice samples. **A**. Sea ice samples collected in 2011 and 2012. **B.** Sea ice samples collected in 2011. **C.** Sea ice samples collected in 2012. Legend of (C) can be found in graph (B).

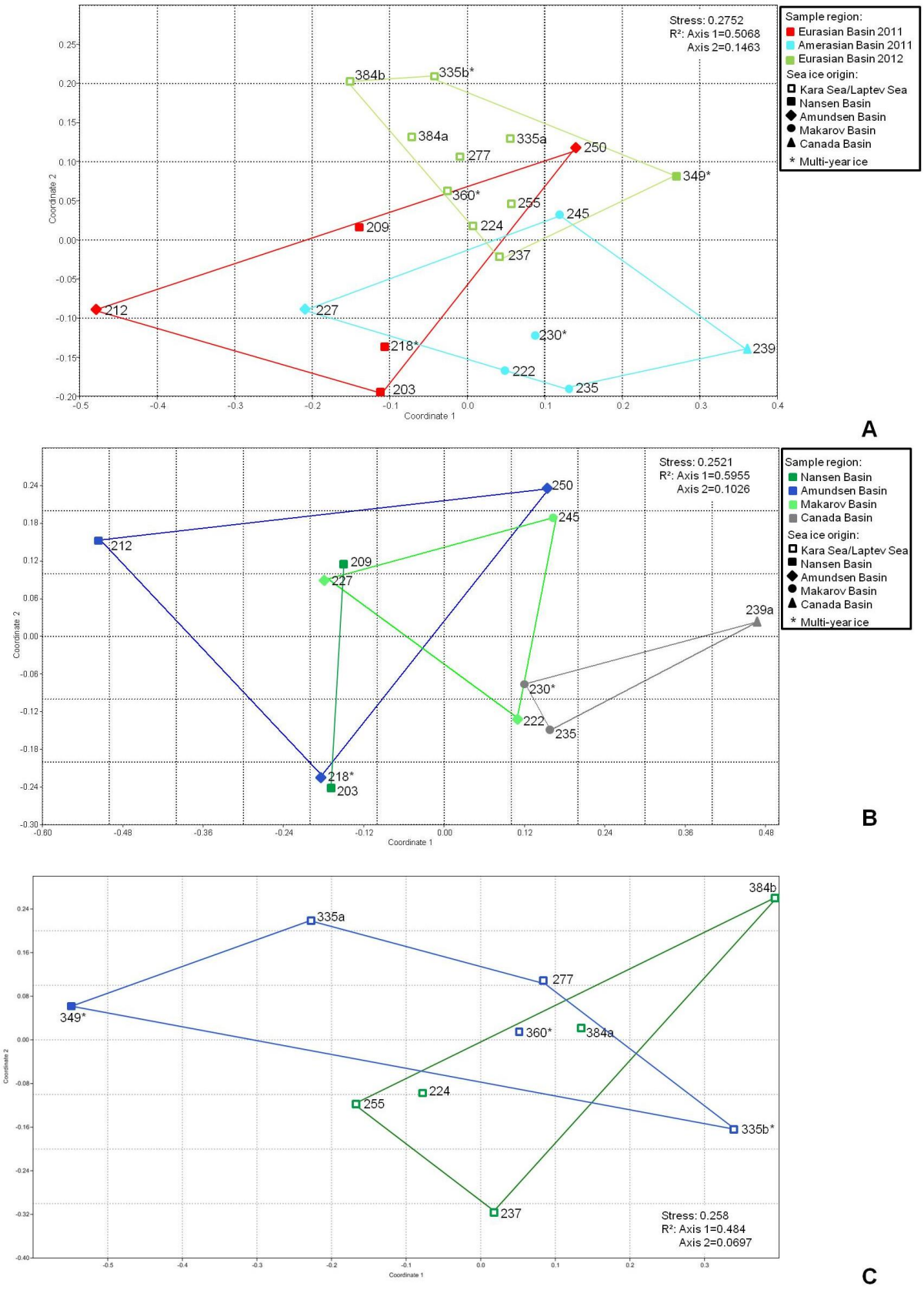
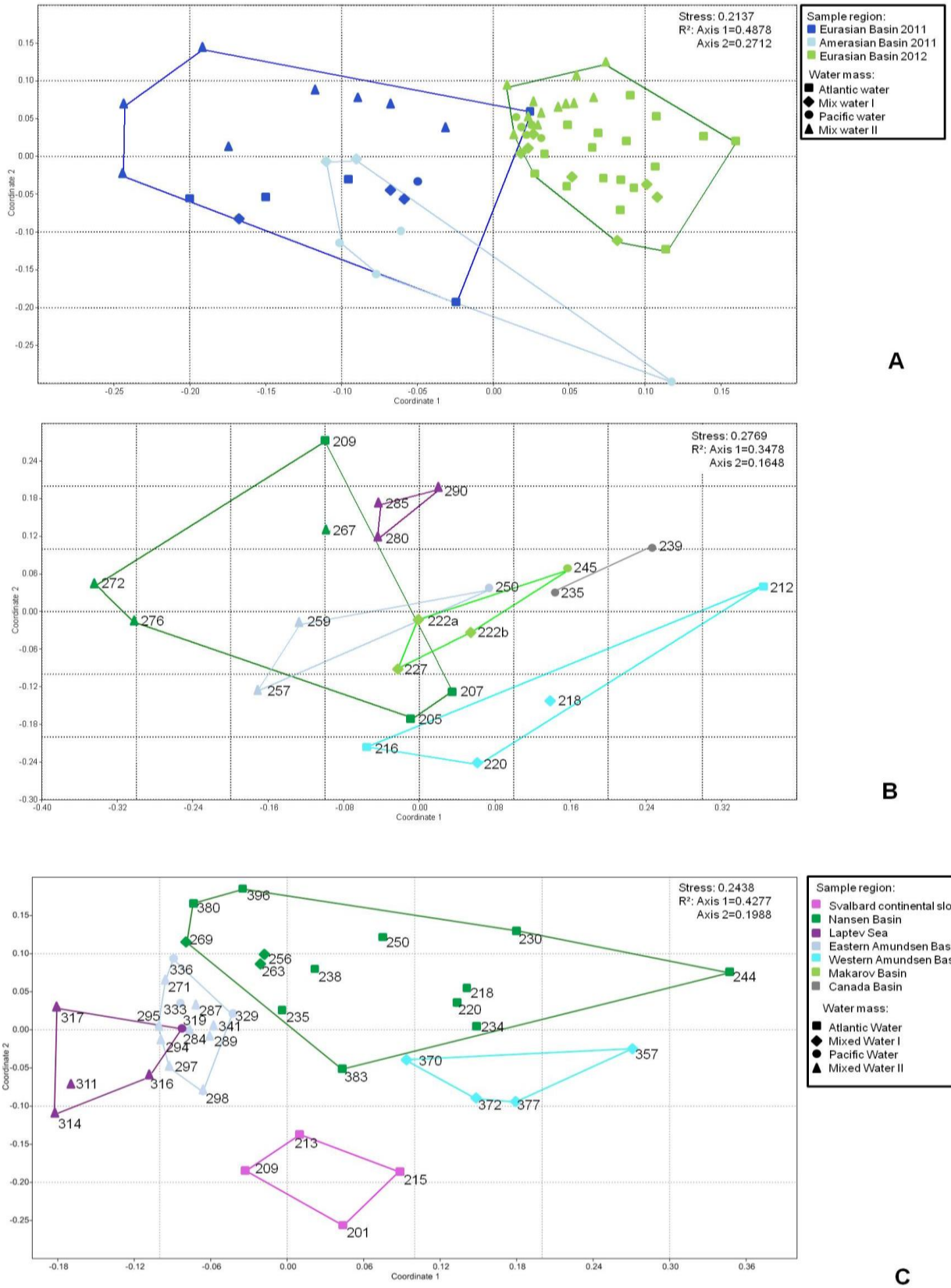


Fig. 8. NMDS plot of the ARISA fragment profiles of water samples. **A**. Water samples collected in 2011 and 2012. **B.** Water samples collected in 2011. **C.** Water samples collected in 2012. Legend of (C) can be found in graph (B).



Tables

Table 1. Stations sampled during RV Polarstern expedition ARK XXVI/3 and ARK XXVII/3 to the Central Arctic Ocean in 2011 and 2012 in chronological order. Stations of 2012 that were part of the common sampling region with 2011 are marked with a star (*). Samples analyzed with Illumina sequencing are in bold. Water samples were taken at each station from the chlorophyll *a* maximum depth measured by the fluorescence probe. Salinity, dissolved nutrients, water temperature, sea ice concentration and water mass are given for each water station. Atlantic Water and Pacific Water formed mixed water masses with different nutrient regimes (Mix I and Mix II). One ice core per ice station was sampled, except of the Stn. 335, where first-year ice (FYI) and multi-year ice (MYI) and Stn. 384, where FYI and newly formed ice (NI) was sampled. Averages of salinity, temperature and dissolved organic nutrients phosphate (PO_4) silicate (Si), nitrite (NO_2) and nitrate (NO_3), were calculated for the sea ice bottom in 2011 and for the entire sea ice core in 2012. Table of water samples taken in 2011 can be found in Kilias et al. (2014). Water samples of 2011 analyzed with Illumina sequencing: 212, 218, 227, 235, 239, 250.

Water samples

Station number	Date (month/day/ year)	Latitude °N	Longitude °E	Depth (m)	Salinity (psu)	PO_4 (µmol/l)	Si (µmol/l)	NO_2 (µmol/l)	NO_3 (µmol/l)	Temp. (°C)	Ice Conc. (%)	Water mass
201	08/05/2012	81.004	29.982	21	34.16	0.21	0.34	0.06	1.49	0.63	0	Atlantic Water
209	08/06/2012	81.493	30.172	32	34.27	0.49	2.41	0.20	5.25	-0.04	0	
213	08/06/2012	81.836	29.954	18	33.7	0.20	1.25	0.03	0.72	-1.16	0	
215	08/07/2012	82.495	30.003	17	33.26	0.17	0.99	0.01	0.46	-1.11	52	
218	08/07/2012	82.99	30.054	50	34.25	0.45	1.90	0.08	5.57	-1.75	91	
220	08/08/2012	83.999	30.021	50	34.19	0.38	1.37	0.17	4.21	-1.78	96	
230	08/11/2012	84.022	31.219	50	34.18	0.41	1.54	0.18	4.73	-1.77	96	
234	08/12/2012	83.99	39.474	22	34.04	0.35	1.36	0.06	3.95	-1.52	94	
235	08/13/2012	83.923	60.655	56	34.18	0.54	2.97	0.15	7.57	-1.68	98	
238	08/14/2012	83.985	78.09	30	34.15	0.41	1.65	0.11	4.79	-1.72	100	
244	08/16/2012	83.918	75.971	30	34.17	0.51	2.67	0.08	6.64	-1.58	90	
250	08/18/2012	83.588	87.452	30	34.13	0.41	1.69	0.12	4.61	-1.68	73	
256*	08/20/2012	82.674	109.59	20	33.74	0.22	1.23	0.11	1.13	-1.68	93	Mixed Water I

Table 1. continued

Station number	Date (month/day/ year)	Latitude °N	Longitude °E	Depth (m)	Salinity (psu)	PO_4 (µmol/l)	Si (µmol/l)	NO_2 (µmol/l)	NO_3 (µmol/l)	Temp. (°C)	Ice Conc. (%)	Water mass
263*	08/22/2012	83.079	110.15	20	33.09	0.28	1.68	0.14	2.21	-1.67	94	
269*	08/23/2012	83.124	116.934	25	33.19	0.32	2.39	0.11	2.49	-1.62	82	
271*	08/24/2012	83.277	122.443	25	31.43	0.34	4.54	0.09	2.27	-1.39	53	Mixed Water II
284*	08/26/2012	82.894	129.769	10	31.18	0.24	3.62	0.06	0.59	-1.55	95	Mixed Water II
287*	08/27/2012	82.166	126.97	20	31.38	0.30	4.44	0.15	1.45	-1.37	10	
289*	08/28/2012	80.004	128.484	25	32.92	0.24	2.65	0.11	0.65	-1.52	0	
294*	08/29/2012	79.05	131.78	18	32.04	0.30	4.01	0.08	1.81	-1.46	0	
295*	08/29/2012	78.745	132.326	20	30.21	0.32	6.53	0.11	1.10	0.11	0	
297*	08/29/2012	78.373	133.196	20	31.22	0.36	3.71	0.11	1.79	0.72	0	
298*	08/29/2012	78.134	133.342	25	31.27	0.54	7.00	0.15	3.98	1.34	0	
311*	09/01/2012	77.397	118.196	20	32.1	0.19	1.28	0.03	0.21	-0.27	0	
314*	09/01/2012	77.716	118.316	20	33.16	0.35	2.58	0.14	1.91	-0.66	0	
316*	09/02/2012	78.35	118.6	18	33.28	0.35	3.75	0.10	2.05	-1.31	0	
317*	09/02/2012	78.666	118.743	20	32.98	0.35	3.16	0.13	2.90	-1.54	0	
319*	09/02/2012	79.162	119.785	18	31.84	0.22	2.75	0.06	0.45	-1.46	0	Contribution of
329*	09/05/2012	81.876	130.878	20	31.04	0.28	3.42	0.10	0.90	-1.48	49	Pacific or river
333*	09/06/2012	83.003	127.179	20	31.04	0.28	4.65	0.10	1.00	-1.5	0	water
336*	09/07/2012	85.094	122.266	20	31.47	0.34	4.68	0.10	1.71	-1.55	80	
341*	09/09/2012	85.159	123.359	19	29.97	0.25	4.69	0.05	0.54	-1.54	80	
357*	09/19/2012	87.924	61.125	15	33.11	0.23	1.52	0.02	1.00	-1.8	100	Mixed Water I
370*	09/23/2012	88.771	55.927	20	32.93	0.28	2.23	0.02	1.68	-1.79	100	
372*	09/24/2012	88.408	52.33	20	33.09	0.27	2.01	0.03	1.00	-1.78	100	
377*	09/25/2012	87.211	51.843	20	33.15	0.35	1.91	0.05	2.96	-1.79	100	
380*	09/26/2012	86.318	52.192	20	33.72	0.31	1.53	0.08	2.42	-1.71	100	Atlantic Water
383	09/27/2012	84.802	52.105	20	32.9	0.37	1.72	0.05	3.59	-1.8	92	
396	09/29/2012	84.346	17.815	10	32.78	0.28	1.15	0.02	2.18	-1.79	100	

Table 1. continued

Sea ice cores

Station number	Date (month/day/ year)	Latitude °N	Longitude °E	Ice type	Ice core length (cm)	Salinity (psu)	Temp. (°C)	PO_4 (µmol/l)	Si (µmol/l)	NO_2 (µmol/l)	NO_3 (µmol/l)	Water mass
203*	08/17/2011	85.974	59.424	FYI	104	4	-1.4	n.d.	n.d.	n.d.	n.d.	Atlantic
209*	08/17/2011	86.987	58.503	FYI	130	2.6	-0.7	n.d.	n.d.	n.d.	n.d.	
212*	08/19/2011	88.018	59.953	FYI	112	0.1	-0.2	0.06	2.68	0.03	0.54	
218*	08/22/2011	89.965	146.631	MYI	318	4.2	-1.3	0.05	0.96	0.23	0.00	MixI
222	08/26/2011	88.736	-128.249	FYI	160	2.0	-1.0	0.00	0.38	0.55	0.25	
227	08/29/2011	86.861	-155.045	FYI	130	0.7	-0.3	0.01	0.25	0.02	0.04	
230	08/31/2011	85.064	-137.235	MYI	249	3.5	-1.1	0.06	1.44	0.18	0.13	Pacific
235	09/02/2011	83.029	-130.035	FYI	209	3.0	-1.0	0.05	7.58	0.13	0.81	
239	09/05/2011	84.074	-164.202	FYI	160	2.9	-1.3	0.36	8.54	0.05	1.04	Mix II
245	09/08/2011	84.795	166.415	FYI	120	3.2	-1.2	0.31	12.91	0.06	1.37	Pacific
250*	09/11/2011	84.372	139.787	FYI	121	0.8	-1.5	0.00	0.50	0.02	0.00	Mix II
224	08/09/2012	84.051	31.114	FYI	120	2.05	-0.66	n.d.	n.d.	n.d.	n.d.	Atlantic
237	08/13/2012	83.987	78.103	FYI	140	1.89	-0.82	0.02	0.34	0.02	0.25	
255*	08/19/2012	82.671	109.590	FYI	92	1.89	-0.71	0.03	0.41	0.01	0.25	MixI
277*	08/24/2012	82.883	130.130	FYI	87	1.66	-0.67	0.06	0.19	0.02	0.74	Mix II
335a*	09/06/2012	85.102	122.245	FYI	84	0.94	-0.66	0.01	0.11	0.01	0.17	Pacific
335b*	09/06/2012	85.102	122.245	MYI	140	1.25	-0.72	0.02	0.60	0.01	0.05	
349*	09/17/2012	87.934	61.217	MYI	139	3.19	-1.94	0.09	0.10	0.02	1.55	MixI
360*	09/21/2012	88.828	58.864	MYI	194	2.39	-1.65	0.02	0.29	0.02	2.31	
384a	09/27/2012	84.375	17.454	FYI	85	1.74	-2.75	0.01	0.09	0.01	0.62	Atlantic
384b	09/27/2012	84.375	17.454	NI	2	9.00	-1.70	0.11	0.29	0.02	0.99	

Table 2. Overview of statistical analysis. An ANOSIM was carried out to test for correlations between ARISA distance matrixes and predefined grouping of water masses, sampling regions or regions of sea ice origin. A Mantel test was performed to test for correlations between ARISA (Jaccard) and environmental (Euclidean) distance matrixes. For comparisons between the years, we repeated the statistical analysis for the common sampling region with water samples. For this region, the number of sea ice samples was too small for adequate statistics. Environmental data comprises sea ice concentration, water temperature, salinity and dissolved nutrients. N.r., not relevant; minus (-), no significance.

	TransArc 2011		**IceArc 2012**		**TransArc 2011**		**IceArc 2012**		**Common sampling region**			
	Sea ice		**Sea ice**		**Water**		**Water**		**Water 2011**		**Water 2012**	
	Nutri-ents	ARISA	Nutri-ents	ARISA	Environ-ment	ARISA	Environ-ment	ARISA	Environ-ment	ARISA	Environ-ment	ARISA
Water masses	n.r.	R = 0.32 *P* = 0.04	n.r.	-	n.r.	R = 0.35 *P* = 0.001	n.r.	R = 0.29 *P* = 0.001	n.r.	R = 0.36 *P* = 0.003	n.r.	R = 0.34 *P* = 0.001
Sampling region	-	R = 0.28 *P* = 0.03	R = 0.47 *P* = 0.01	-	-	R = 0.15 *P* = 0.05	R = 0.54 *P* = 0.001	R = 0.53 *P* = 0.001	R = 0.45 *P* = 0.02	-	R = 0.41 *P* = 0.001	R = 0.71 P = 0.001
Sea ice origin	R = 0.59 *P* = 0.02	R = 0.34 *P* = 0.02	-	-	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
Mantel test	-		-		-		r = 0.32, *P* = 0.001		-		r = 0.33, *P* = 0.001	

Table 3. Overview of ARISA and Illumina analysis. Given are number (no.) of water and sea ice samples collected during the *RV Polarstern* expeditions TransArc in 2011 and IceArc in 2012 to the Central Arctic Ocean. For each region and habitat, average number of ITS1 fragments or OTUs including standard deviations are shown. For Illumina sequencing, a further division in rare (<1%) and abundant (≥1%) biosphere was made. A common sampling region was calculated for the samples collected in 2012 that were collected in the same regions of the Eurasian Basin as in 2011.

			ARISA			Illumina sequencing				
Expedition	Region	Habitat	No. samples	Average no. fragments	Standard deviation	No. samples	Average no. OTUs	Standard deviation	Average no. OTUs rare biosphere	Average no. OTUs abundant biosphere
TransArc 2011	Eurasian Basin	Water	16	73	13	3	556	40	536	20
		Sea ice	5	91	16	3	436	139	418	18
	Amerasian Basin	Water	7	66	12	3	544	30	527	17
		Sea ice	6	89	21	3	535	136	514	21
IceArc 2012	Eurasian Basin	Water	39	86	11	9	508	50	488	20
		Sea ice	10	94	11	7	387	30	366	21
	With 2011 common sampling region	Water	25	87	11	6	527	52	507	20
		Sea ice	6	92	12	5	378	32	359	19

7.3 Manuscript III

The importance of sea ice for exchange of habitat-specific protist communities in the Central Arctic Ocean[1]

Kristin Hardge*[a, b], Ilka Peeken [a], Stefan Neuhaus [a], Benjamin A. Lange [a, c], Alexandra Stock [d], Thorsten Stoeck [d], Lea Weinisch [d], and Katja Metfies [a, b]

a Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Polar Biological Oceanography, Bremerhaven, Germany

b Jacobs University Bremen, Bremen, Germany

c University of Hamburg, Zoological Institute and Zoological Museum, Biocenter Grindel, Hamburg, Germany

d University of Kaiserslautern, Faculty of Biology, Ecology Group, Kaiserslautern, Germany

* Corresponding author: kristin.hardge@awi.de

1 This is the submitted and not revised manuscript version. The accepted publication is available online:

Hardge, K. et al. (2016). The importance of sea ice for exchange of habitat-specific protist communities in the Central Arctic Ocean. *Journal of Marine Systems*, http://dx.doi.org/10.1016/j.jmarsys.2016.10.004.

Abstract

Sea ice is one of the main features influencing the Arctic marine protist community composition and diversity in sea ice and sea water. We analyzed protist communities within sea ice, melt pond water, under-ice water and deep-chlorophyll maximum water at eight sea ice stations sampled during summer of the 2012 record sea ice minimum year. Using Illumina sequencing, we identified characteristic communities associated with specific habitats and investigated protist exchange between these habitats. The highest abundance and diversity of unique taxa were found in sea ice, particularly in multi-year ice (MYI), highlighting the importance of sea ice as a unique habitat for sea ice protists. Melting of sea ice was associated with increased exchange of communities between sea ice and the underlying water column. In contrast, sea ice formation was associated with increased exchange between all four habitats, suggesting that brine rejection from the ice is an important factor for species distribution in the Central Arctic. Ubiquitous taxa (e.g. *Gymnodinium*) that occurred in all habitats still had habitat-preferences. This demonstrates a limited ability to survive in adjacent but different environments. Our results suggest that the continued reduction of sea ice extent, and particularly of MYI, will likely lead to diminished protist exchange and subsequently, could reduce species diversity in all habitats of the Central Arctic Ocean. An important component of the sea ice protist community could be endangered because specialized taxa restricted to this habitat may not be able to adapt to rapid environmental changes.

Keywords: 18 S rRNA gene; Arctic Ocean; Illumina sequencing; ice algae; phytoplankton; protist exchange

Introduction

Sea ice is a major factor that structures the polar ecosystem, which is characterized by strong seasonality. Large amounts of sea water undergo a yearly cycle of freezing and melting leading to variations in physical and biogeochemical properties of the surface ocean and the sea ice. Thus, polar organisms must be well adapted to this strong seasonality in terms of their life cycle, ecology and physiology. Unicellular eukaryotes that live within the water column or sea ice habitats are the primary energy source for all trophic levels in the marine ecosystem (Soreide et al., 2013, Kohlbach et al., in press) and account for a large proportion of total primary production and carbon flux (Andreassen et al., 1996; Michel et al., 2002; Riedel et al., 2008; Fernández-Méndez et al., 2015).

The main habitats of the Central Arctic Ocean (CAO) have been impacted by current climate induced changes with large implications for biological processes and energy fluxes. In particular, the sea ice decline (Comiso 2003; Comiso et al., 2008) and a thinner, younger ice pack (Maslanik et al., 2011; Laxon et al., 2013; Kwok and Rothrock et al., 2009) has major implications for the Arctic ecosystem. These changes resulted in more light being transmitted through sea ice (Nicolaus et al., 2012) which enhanced the photosynthetic production of sea ice algae and phytoplankton (Arrigo et al., 2014a; Arrigo and Van Dijken 2015). Due to earlier sea ice retreat, the timing of spring phytoplankton blooms has shifted (Leu et al., 2011). Because phytoplankton is incorporated in the sea ice, a time shift may also alter the sea ice community structure of the following spring sea ice bloom (Niemi, et al. 2011). Thus, a thinning sea ice cover and loss of multi-year ice (MYI, sea ice that survived at least one summer) (Comiso 2012) can influence sea ice and pelagic community structure and habitat exchange with further implications for species diversity, food web dynamics (Carmack and Wassmann 2006; Wassmann 2008; Hilligsøe et al., 2011; David et al., 2015; Hardge et al., subm.) and carbon sequestration (Boetius et al., 2013). Furthermore, due to sea ice thinning and melt pond proliferation, the abundance and diversity of freshwater genera (e.g. *Chlamydomonas* and *Ochromonas*), could increase in sea ice and melt ponds (Kilias et al., 2014b), and thus, contribute significantly to primary productivity in the Arctic Ocean (Fernández-Méndez et al., 2015). Shifts in community composition were observed in the Canadian Arctic Ocean and Fram Strait, where large plankton cells were displaced by small cells possibly as a result of Ocean freshening and warming (Li et al., 2009; Tremblay et al., 2009; Nöthig et al., 2015).

In-depth knowledge about protist community structure and exchange between the variable habitats of the CAO is still lacking, particularly in melt ponds (Kilias et al., 2014b; Fernández-

Méndez et al., 2015) and UIW (Arrigo et al., 2012; Laney et al., 2014). The two main habitats of the CAO, the water column and sea ice, are inhabited by different protist communities. While mainly Dinophyceae are common in the water column (Booth and Horner 1997; Jensen and Hansen 2000; Kilias et al., 2014a), the sea ice is mainly dominated by Bacillariophyceae (Poulin et al., 2011; Comeau et al., 2013). However, little is known about the overall degree of protist exchange and the driving forces that trigger exchange between the habitats (Niemi et al., 2011). There is first indication that the habitats are connected with each other via brine channels in the sea ice matrix. Sea ice algae can be released into the under-ice water (UIW) during sea ice melt (Quillfeldt 2000; Boetius et al., 2013) and pelagic protists can be incorporated into the sea ice matrix during sea ice formation (Ackley et al., 1987; Gradinger and Ikävalko 1998; Różańska et al., 2008; Niemi et al., 2011). In addition, the melting of snow during the Arctic summer lead to large melt ponds on the ice floes (Sankelo et al. 2010), which generally are not connected to the under-ice water (UIW). However, with the tendency toward FYI, the under-ice or sea ice communities can evolve into communities of open or frozen melt ponds (Lee et al., 2011). Therefore, the spatial dynamics of protist communities in the Arctic Ocean are affected by sea ice thinning, as they are primarily controlled by changes in the surrounding physical environment rather than by active immigration or emigration. However, existing knowledge about habitat-restriction or habitat-exchange is only based on the analysis of conventional approaches, such as: light microscopy (e.g. Ackley et al., 1987; Gradinger and Ikävalko 1998; Różańska et al., 2008; Lee et al., 2011; Niemi et al., 2011; Poulin et al., 2011). Furthermore, it is largely unknown whether introduced species could survive in a different habitat they are not adapted to. This knowledge is especially important in the context of climate change, as species restricted to one habitat might be more vulnerable to climate change (Myers et al., 2000; Lovejoy and Potvin 2011), because they are highly adapted to a specific living environment in terms of their life cycle and physiology. In contrast, widely distributed species might be less affected, because of their ability to adapt to fast changing environmental conditions. The assessment of rare and habitat-specific taxa, as well as the protist exchange between the habitats (i.e. number of unique or shared taxa) is best analyzed with molecular methods. The fast improvement of next-generation sequencing allows the identification of rare and small-sized species, which lack morphological features, and are both often overlooked using light microscopy.

In this study, we sequenced protist communities in deep-chlorophyll maximum water (DCM), under-ice water (UIW), sea ice (ICE) and melt pond water (MW) sampled in the Central Arctic Ocean (CAO) during summer of the record sea ice minimum year 2012. We provide a comprehensive overview of habitat-specific protist community composition and elucidated the

exchange of protists between the habitats, as well as the driving forces of exchange. We tested the following hypotheses:

1) Different habitats in the ice-covered Arctic Ocean harbor unique, habitat-specific protist taxa.
2) Ubiquitous species that occur in all habitats still have habitat-preferences.
3) Sea ice melting triggers protist exchange between various habitats of the CAO.

Material and Methods

Sampling

Seven ice stations were sampled in the Eurasian Basin of the CAO during *RV Polarstern* expedition "IceArc" in 2012 (PS80, 5 August to 29 September 2012) (**Fig. 1**, **Tab. 1**). A total of seven seawater samples from the DCM, seven UIW samples, eight ice cores and seven MW samples were collected in different water masses (**Tab. 1**).

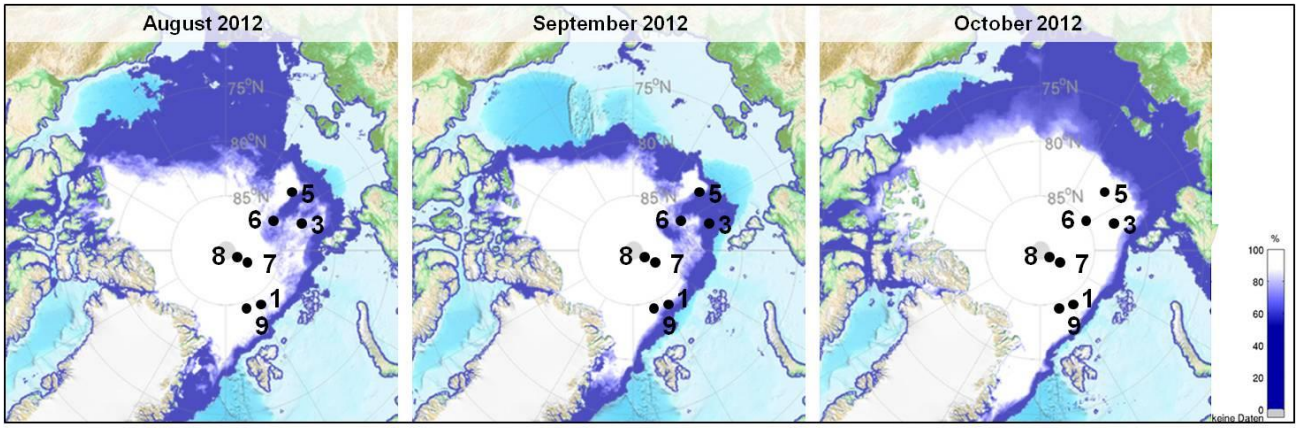


Fig. 1. Position of ice stations sampled during the RV Polarstern expedition IceArc to the central Arctic in 2012. For exact sample positions see Tab. 1. At each ice station, sampling of deep-chlorophyll maximum water layer depth, under-ice water, sea ice cores and melt pond water took place. The maps show monthly mean sea ice concentrations for August, September and October (map from meereisportal.de). October is presented to illustrate the increase in sea ice concentration at the end of the season.

Tab. 1. Overview of ice stations sampled during RV Polarstern expedition IceArc (PS80) to the Central Arctic Ocean in 2012 in chronological order. Given are the ice station numbers and ship station numbers (for CTD and ice core drilling). At each ice station, sampling of the deep-chlorophyll maximum water, under-ice water, sea ice cores and melt ponds water took place. Ice thickness, ice type (first-year ice, FYI; and multi-year ice, MYI) and water mass are given as well. Atlantic Water and Pacific Water formed mixed water masses with different nutrient regimes (Mix I and water masses with enhanced contribution of Pacific or River Water) were sampled.

Station number	Ship station number PS80	Date (month/day/ year)	Latitude °N	Longitude °E	Ice thickness (cm)	Ice type	Water mass	Deep-chlorophyll maximum depth (m)
Ice 1	230 (CTD) 224 (ICE)	09/29/2012 - 08/11/2012	84.051	31.114	120	FYI	Atlantic	50
Ice 3	256 (CTD) 255 (ICE)	08/19/2012 - 08/20/2012	82.671	109.59	92	FYI	MixI	20
Ice 5	329 (CTD) 323 (ICE)	09/03/2012 - 09/05/2012	81.876	130.878	81	FYI	Contribution of Pacific or river water	20
Ice 6_fyi	335 (CTD) 336 (ICE)	09/06/2012 - 09/07/2012	85.102	122.245	84	FYI	Contribution of Pacific or river water	20
Ice 6_myi	335 (CTD) 336 (ICE)	09/06/2012 - 09/07/2012	85.102	122.245	140	MYI	Contribution of Pacific or river water	20
Ice 7	357 (CTD) 349 (ICE)	09/17/2012 - 09/19/2012	87.934	61.217	139	MYI	MixI	15
Ice 8	370 (CTD) 360 (ICE)	09/21/2012 - 09/23/2012	88.828	58.864	194	MYI	MixI	20
Ice 9	396 (CTD) 384 (ICE)	09/27/2012 - 09/29/2012	84.375	17.454	85	FYI	Atlantic	10

Water samples were collected with Niskin bottles (12 L) attached to a CTD (conductivity, temperature, depth) rosette from the DCM, which varied between 10 m and 50 m (**Tab. 1**). Two liters subsamples were taken in PVC bottles. UIW samples were collected during the 2012 IceArc expedition with a Masterflex® E/S™ portable sampler pump. The tube was marked every meter and was placed with a weight one meter under the ice. For molecular analysis, 2 L of water was filtered through Isopore Membran Filters (Millipore, Billerica, MA, USA) with pore sizes of 10 µm, 3 µm and 0.4 µm to ensure collection of all protist cell sizes. Filters were stored in Eppendorf tubes at -80°C.

First-year ice (FYI) and multi-year ice (MYI) cores (**Tab. 1**) were drilled with a Kovacs 9 cm inner diameter corer (Kovacs Enterprise, Roseburg, USA). Four ice cores were drilled per station: two 'physical cores' for determining physico-chemical parameters, one 'biological cores' for analyzing biological communities and one "biogeochemical core" for nutrient analyses. The 'biological cores' were sectioned in 10 to 20 cm intervals and melted in 0.2 µm filtered sea-water to minimize osmotic stress to protists during the melting process (Miller et al., 2015). Melting of sea ice was conducted under low light conditions at 4 °C for 24 - 48 hours. The samples were pooled from the entire core and further processing (filtering and storage) was done as described for the water samples. Sea ice temperature was measured on the first physical core immediately after extraction by drilling into the ice and determining the temperature every 10 cm with a temperature probe (Testo 720). The second physical core was cut into 10 cm sections and melted onboard for further analyses. Biogeochemical cores were sampled and analyzed as described in Fernández-Méndez et al., (2015).

At each ice station, a melt pond was randomly chosen and the water was sampled by placing a tube of the Masterflex® E/S™ portable sampler in the middle of the pond. Collected water was transported to the ship and further processing (filtering and storage) was done as described for the water samples.

Physico-chemical parameters

Water temperature, salinity (Rabe et al., 2013) and dissolved inorganic nutrients (phosphate, silicate, nitrite and nitrate) (Bakker 2014a) were acquired from the PANGAEA database. The temperature and salinity of the other habitats were determined using a temperature probe (Testo 729) and a Salinometer (WTW Cond.3110). Bulk salinity was measured on each melted 10 cm section of the physical cores. Nutrient analyses on the biogeochemical cores were described in Fernández-Méndez et al., (2015). MW and ICE nutrients were taken from PANGAEA database (Bakker 2014b) and

measured according to standard methods (Kattner and Becker 1991). No nutrient data was available for UIW at ice station 9 and for ICE at ice station 1. Sea ice concentration (0 - 100% sea ice coverage) data were obtained from sea ice portal (www.meereisportal.de, grant: REKLIM-2012-04) based on Advanced Microwave Scanning Radiometer (AMSR-E) data using the ARTIST SeaIce (ASI) algorithm (Spreen et al., 2008). Daily average incoming photosynthetically active radiation (incoming PAR; µmol photons m^{-2} s^{-1}) were obtained from Fernández-Méndez et al. (2015).

Illumina sequencing

For Illumina sequencing, isolated DNA of each filter size (10 µm, 3 µm and 0.4 µm) was pooled in equal volumes. The V4 region was amplified in triplicates using universal primer set TAReuk454FWD1 (5´-CCA GCA SCY GCG GTA ATT CC-3'; *S. cerevisiae* position 565-584) and TAReukREV3 (5`-ACT TTC GTT CTT GAT YRA-3`; *S. cerevisiae* position 964-981) (Stoeck et al., 2010). The PCR mixture contained 10 µl 5x Phusion high-fidelity buffer 1 µl dNTP-mix, 1 µl of each primer, 0.5 Phusion Hot Start high-fidelity tag polymerase (New England Biolabs GmbH, Frankfurt/Main, Germany) and 2 µl template DNA in a volume of 50 µl. The DNA amplification was carried out in two rounds using a Mastercycler (Eppendorf, Germany). The first round started with an initial activation at 98°C for 30 sec followed by 10 cycles of denaturation at 98°C for 10 sec, annealing at 57°C for 30 sec and extension at 72°C for 30 sec. The second round was 25 cycles of denaturation at 98°C for 10 sec followed by annealing at 52°C for 30 sec and extension at 72°C for 30 sec. The final extension was at 72°C for 10 min. Resulting PCR products were purified with NucleoSpin® Gel & PCR Clean up kit (Macherey-Nagel) according to the manufacturer's protocol.

Sequence analysis

Sequences were processed with QIIME version 1.8.0 (Caporaso et al., 2010) and quality-filtered according to recommended settings in Bokulich et al. (2013). Primers were retained and used for taxonomic differentiation. Further processing and analyses were only conducted on sequences which fully matched the primer sequences at the beginning and the end of the sequence, and which were are between 330 and 460 bp in length (length of expected insert: 380 bp to 420 bp). For statistical analysis and comparison of protist community structure between the samples, sequence numbers were down-sampled to the lowest number of sequences (~130000 seqs). Subsequently, chimeric sequence detection and OTU clustering was done with the QIIME workflow usearch.qf (based on Usearch version 5.2.236 (Edgar 2010) which incorporates UCHIME (Edgar et al., 2011). Sequences were checked for chimeras, by both using reference data for comparison (Silva 119 SSU Ref NR) and with abundance-sorted query sequence pre-cluster as a self-reference (i.e. *de novo*). Remaining high

quality sequences were clustered *de-novo* into OTUs at a minimum similarity threshold of 98%. Following Bokulich et al., (2013) all OTUs, which consisted of less than 0.005% of the processed sequences, were removed.

OTU representative sequences were classified by phylogenetic placement utilizing PhyloAssigner v. 6.166 (Vergin et al., 2013). PhyloAssigner places sequences in a fixed rooted backbone tree and assigns queries to a last common ancestor (LCA). For the reference set, the ARB tree of Silva 111 SSU Ref NR with corresponding multiple sequence alignment was thinned down manually by optimizing the tree to a structurally and taxonomically representative set of 4000 leafs (available on request). Query OTU sequences were aligned to the reference alignment with HMMER v. 3.0 (Eddy 2011) and placed in the phylogenetic reference tree with pplacer (Matsen et al., 2010). The tree was rooted with representatives of Opisthokonta as an outgroup. Placement likelihood values are reported for each sequence and the LCA node was determined at an accumulated likelihood weight ratio cutoff of 0.8. LCA nodes can be leafs or inner nodes without taxonomic label and thus, the full taxonomic string of all leaves below the LCA node was extracted and the common prefix string was used as taxonomic assignment.

To address the above mentioned hypotheses, we classified the protist community into unique taxa (found in one habitat only), shared taxa (found in two or three habitats) and ubiquitous taxa (found in all four habitats).

Statistical analysis

To visualize similarity patterns of protist communities in the samples, we applied non-dimensional metric scaling (NMDS) based on Jaccard and Bray Curtis index using R package ‘vegan’ (R version 3.2.3, R Core Team 2015). To obtain a detailed picture of the nutrient regime in all habitats at the sampling stations, we computed principal component analyses (PCA) with R package ‘ade4’. In order to identify the driving environmental forces influencing protist communities in the habitats, a Mantel test was performed with R package ‘vegan’ to assess possible correlations of physico-chemical parameters with occurrences of the most prominent protist groups. This was done with 999 permutations, and the distance measures Jaccard for protist communities and Euclidean for environmental parameters.

Results

Protist community composition in the habitats

The community structure of protists in DCM and UIW differed from the community structure observed in ICE and MW. Communities in DCM and UIW were relatively homogenous, whereas communities in ICE and MW were highly variable in composition and especially in diversity (**Fig. 2**). However, in terms of diversity, protist communities in MW samples 3, 8 and 9 were more similar to UIW and DCM than to ICE (**Fig. 2A**) but not in terms of abundance (**Fig. 2B**).

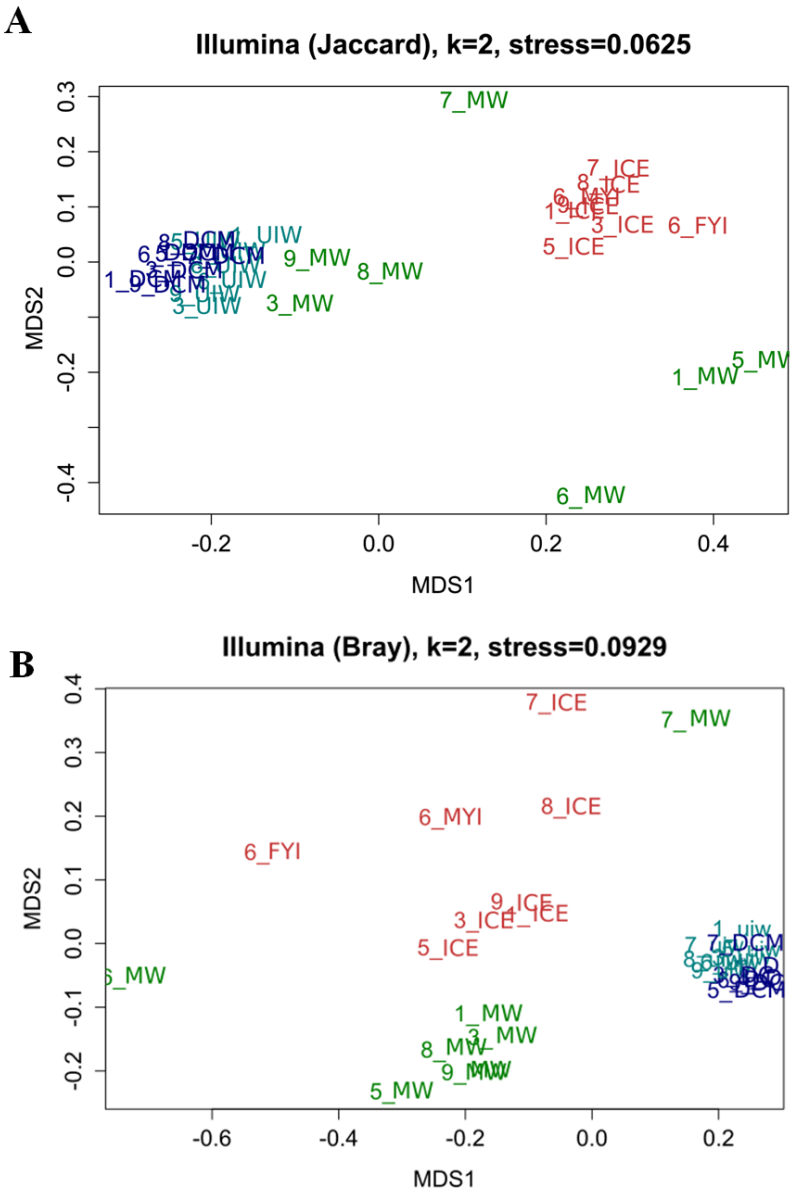


Fig. 2. Protist community structure at the ice stations. NMDS plots based on (**A**) Jaccard index (community diversity) and (**B**) Bray Curtis index (community composition) of protist Illumina sequences analyzed in deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during IceArc expedition.

The highest number of OTUs was observed in the DCM (on average 526 OTUs) and UIW (on average 516 OTUs), whereas the number of OTUs was generally lower in ICE (on average 391 OTUs) and MW (on average 350 OTUs) **(Fig. 3A**). At stations 1, 5 and 6, we observed only 193 - 221 OTUs in MW but the number was considerably higher at the other stations (298 - 555) (**Fig. 3A**). The main protist representatives in DCM and UIW were Dinophyceae (e.g. *Gymnodinium*, *Karlodinium*), and Protalveolata (Syndiniales). In UIW, Chlorophyceae (e.g. *Micromonas*) were more frequent, while Protalveolata were less frequent compared to the DCM (**Fig. 3B**). The ICE samples were mainly characterized by Bacillariophyceae (e.g. *Navicula*, *Nitzschia*, *and Melosira*), Chrysophyceae (e.g. *Ochromonas, Spumella*) and Cercozoa (e.g. NAMAKO-15, *Cryothecomonas*) (**Fig. 3B**). In contrast, MW showed higher abundances of Chrysophyceae (e.g. *Ochromonas*) and at one station (ice station 7) we observed high contribution of Bacillariophyceae-sequences to the total sequence number. Furthermore, Ciliophora (e.g. *Didinium*, *Paramecium*) were highly abundant in two MW (ice station 8 and 9).

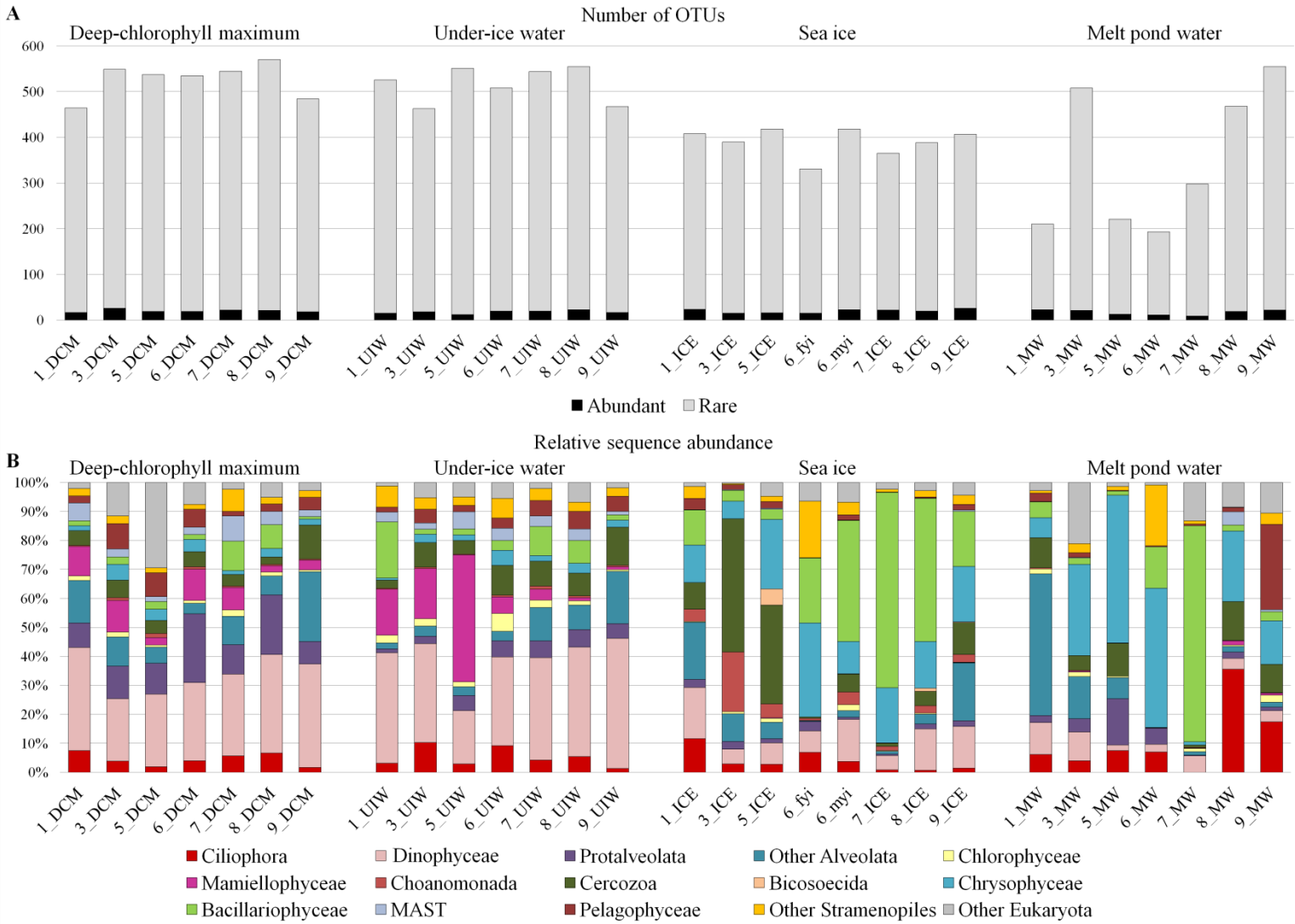


Fig. 3. Protist community structure obtained with Illumina sequencing. (**A**) Number of OTUs belonging to the abundant biosphere (> 1%) and rare biosphere (≤1%) and (**B**) relative sequence abundance of protists analyzed in deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE), inclusive first-year (fyi) and multi-year ice (myi) and melt pond water (MW) collected during IceArc expedition.

Unique OTUs per habitat

At ice stations 1, 5 and 6, the number of unique OTUs (**Fig. 4A**, **Tab. 2**) and OTU representative sequences (**Fig. 4B**) were generally high, indicating that protist exchange between the habitats was overall low at these stations. In contrast, the numbers were lower at ice stations 3, 7, 8 and 9 **(Fig. 4**, **Tab. 2**), suggesting a high-exchange between the habitats. In general, we found the highest proportion (number of OTUs and sequence abundance) of unique taxa in DCM and ICE (**Fig. 4**, **Tab. 2**). Particularly at the MYI stations 6 (20% of total sequences) and 8 (12% of total sequences), the ICE samples showed the highest abundance of unique taxa with *Nitzschia* as main representative (**Fig. 4**). Unique OTUs in DCM were mainly assigned to Protalveolata (mainly Syndiniales) and uncultured marine stramenopiles (MAST) (**Fig. 4**). UIW always showed a lower proportion of unique taxa than DCM and ICE (**Tab. 2**, **Fig. 4**). Representatives were mainly Ciliophora (e.g. *Strombidium*), Bacillariophyceae (e.g. *Porosira* and Fragilariales) and marine phytoflagellates (Pedinellales, at ice station 3, **Fig. 4B**). Compared to all other habitats, MW samples had the lowest number and proportion of unique OTUs (**Tab. 2**, **Fig. 4**), which were mainly assigned to Ciliophora (e.g. *Stokesia* and *Didinium*) (**Fig. 4**). A listed taxonomy and number of unique OTUs assessed for all stations can be found in **Table A1**.

Tab. 2. Overview of protist exchange at the ice stations. (**A**) Number and (**B**) percentage of total number of unique and shared OTUs found at the ice stations. Samples were collected from deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) during IceArc expedition. OTUs found in one habitat are considered as unique and OTUs found in all habitats as ubiquitous OTUs. Stations were classified in low-exchange and high-exchange stations based on the number of unique and shared/ ubiquitous OTUs. Average and standard deviations (Stdev.) were calculated for low-exchange, high-exchange and all stations. The color code indicates the minimum (light blue) and maximum (dark blue) exchange for each habitat or habitat interaction.

A Number of OTUs

Degree of exchange	Ice station	DCM	UIW	ICE	MW	DCM + UIW	DCM + ICE	DCM + MW	UIW + ICE	UIW + MW	ICE + MW	DCM + UIW + ICE	DCM + UIW + MW	DCM + ICE + MW	UIW + ICE + MW	All habitats
Low -exchange																
	Ice 1	70	42	57	7	373	215	145	302	180	209	196	136	140	172	132
	Ice 5	50	41	45	8	448	278	152	295	170	219	240	133	149	161	130
	Ice 6 FYI	76	34	70	6	428	192	154	208	173	158	164	148	110	130	106
	Ice 6 MYI	71	19	83	6	428	254	154	276	173	174	220	148	129	145	124
	Average	67	34	64	7	419	235	151	270	174	190	205	141	132	152	123
	Stdev.	11	11	16	1	32	39	4	43	4	29	33	8	17	18	12
High-exchange																
	Ice 3	61	13	67	12	403	241	402	228	367	276	187	338	212	202	178
	Ice 7	49	38	43	3	465	255	259	265	259	249	230	239	203	203	188
	Ice 8	35	16	40	21	481	267	375	274	381	304	234	334	234	243	213
	Ice 9	19	7	49	46	395	229	392	255	386	333	203	330	213	238	194
	Average	41	19	50	21	436	248	357	256	348	291	214	310	216	222	193
	Stdev.	18	14	12	19	43	17	66	20	60	36	22	48	13	22	15
All stations	Average	54	26	57	14	428	241	254	263	261	240	209	226	174	187	158
	Stdev.	20	14	15	14	36	28	118	32	101	62	26	96	47	42	40

B Percentage of OTUs

Degree of exchange	Ice station	DCM	UIW	ICE	MW	DCM + UIW	DCM + ICE	DCM + MW	UIW + ICE	UIW + MW	ICE + MW	DCM + UIW + ICE	DCM + UIW + MW	DCM + ICE + MW	UIW + ICE + MW	All habitats
Low-exchange																
	Ice 1	10	6	8	1	24	2	0	9	1	4	9	1	1	6	19
	Ice 5	7	6	6	1	28	3	0	3	1	5	15	0	3	4	18
	Ice 6 FYI	11	5	10	1	31	3	0	3	0	3	8	6	1	3	15
	Ice 6 MYI	10	3	11	1	25	4	0	5	1	3	13	3	1	3	17
	Average	9	5	9	1	27	3	0	5	1	4	11	3	1	4	17
	Stdev.	2	1	2	0	3	1	0	3	0	1	3	3	1	1	2
High-exchange																
	Ice 3	8	2	9	2	8	3	4	2	1	6	1	22	5	3	25
	Ice 7	7	5	6	0	26	1	1	3	1	4	6	7	2	2	27
	Ice 8	5	2	5	3	17	2	3	1	2	5	3	16	3	4	29
	Ice 9	3	1	7	6	8	1	6	1	2	10	1	19	3	6	27
	Average	6	3	7	3	15	2	3	2	1	6	3	16	3	4	27
	Stdev.	3	2	2	3	9	1	2	1	1	3	2	6	1	2	2
All stations	Average	7	4	8	2	21	2	2	3	1	5	7	9	2	4	22
	Stdev.	3	2	2	2	9	1	2	3	1	2	5	9	1	1	5

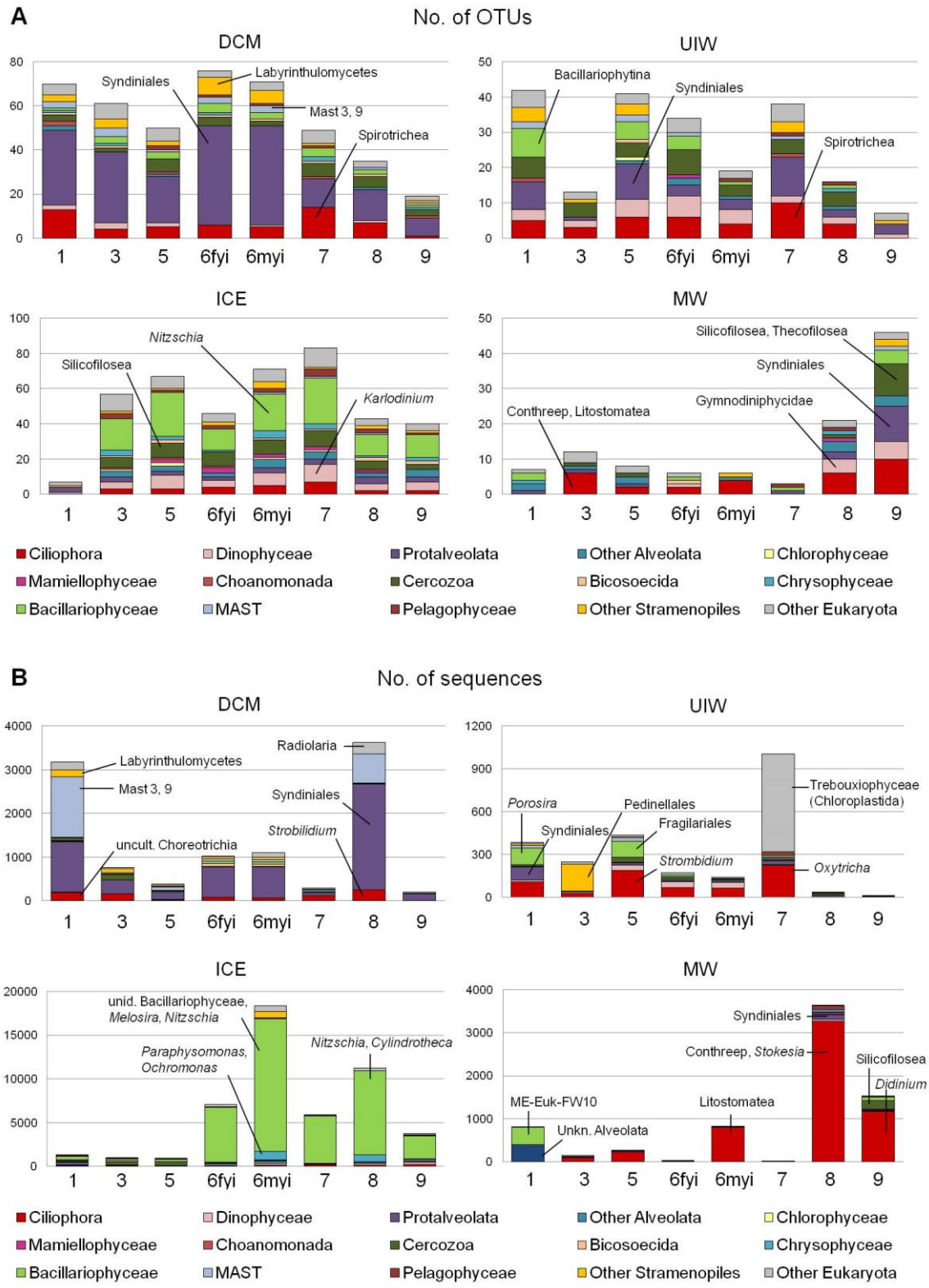


Fig. 4. Unique taxa observed at the ice stations. (**A**) Number (No.) of unique OTUs of protist taxa and taxa groups and (**B**) number of OTU representative sequences found in the deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during the IceArc expedition. Specific examples of most abundant taxa are given. Note the different dimensions between axes (e.g. highest abundances of unique taxa were found in ICE, lowest in UIW).

Shared OTUs between the habitats

With respect to the number of OTUs shared by two local communities, the strongest interaction was observed between protist assemblages in DCM and UIW (**Tab. 2**). The most diverse taxa shared between these water depths were Dinophyceae and Protalveolata (Syndiniales). ICE was slightly more connected to UIW than to DCM (**Tab. 2**). This connectivity was mainly characterized by Dinophyceae, Syndiniales and Bacillariophyceae (e.g. *Melosira*, *Navicula*) and was highest at station 1 and station 5 (about 43%), where Bacillariophyceae showed a high diversity (32 and 42 OTUs). DCM and UIW also shared a large number of OTUs with the MW samples, mainly Bacillariophyceae, Dinophyceae and Protalveolata. At the ice stations 3, 7, 8 and 9, we observed a distinct interaction between UIW, DCM and MW with up to 338 shared OTUs (station 3) (**Tab. 2A**). These four stations were also characterized by a high connectivity between sea ice cores and MW samples (mainly Bacillariophyceae and Dinophyceae). Based on the number of unique and shared OTUs, ice stations 3, 7, 8 and 9 were classified as 'high-exchange stations', and ice stations 1, 5, and 6 were classified as 'low-exchange stations'.

Ubiquitous OTUs at ice stations

At high-exchange stations, we observed the highest number and proportion of ubiquitous taxa (**Tab. 2**, **Fig. 5**), which were mainly classified as Alveolata (49%) and Stramenopila (36%) (**Fig. 5**). Number (**Fig. 5A**) and abundance (**Fig. 5B**) of ubiquitous Bacillariophyceae and Dinophyceae OTUs were higher at high-exchange stations but lower at low-exchange stations. The opposite pattern was observed for ubiquitous Chrysophyceae, which had similar community diversity at all stations (**Fig. 5A**) but their abundance was higher at the low-exchange stations 5, 6, **Fig. 5B**). In the overall sequence data set, we found 488 different OTUs that were shared between all four habitats of the ice stations. A listed taxonomy and number of ubiquitous OTUs assessed for all stations can be found in **Table A2**.

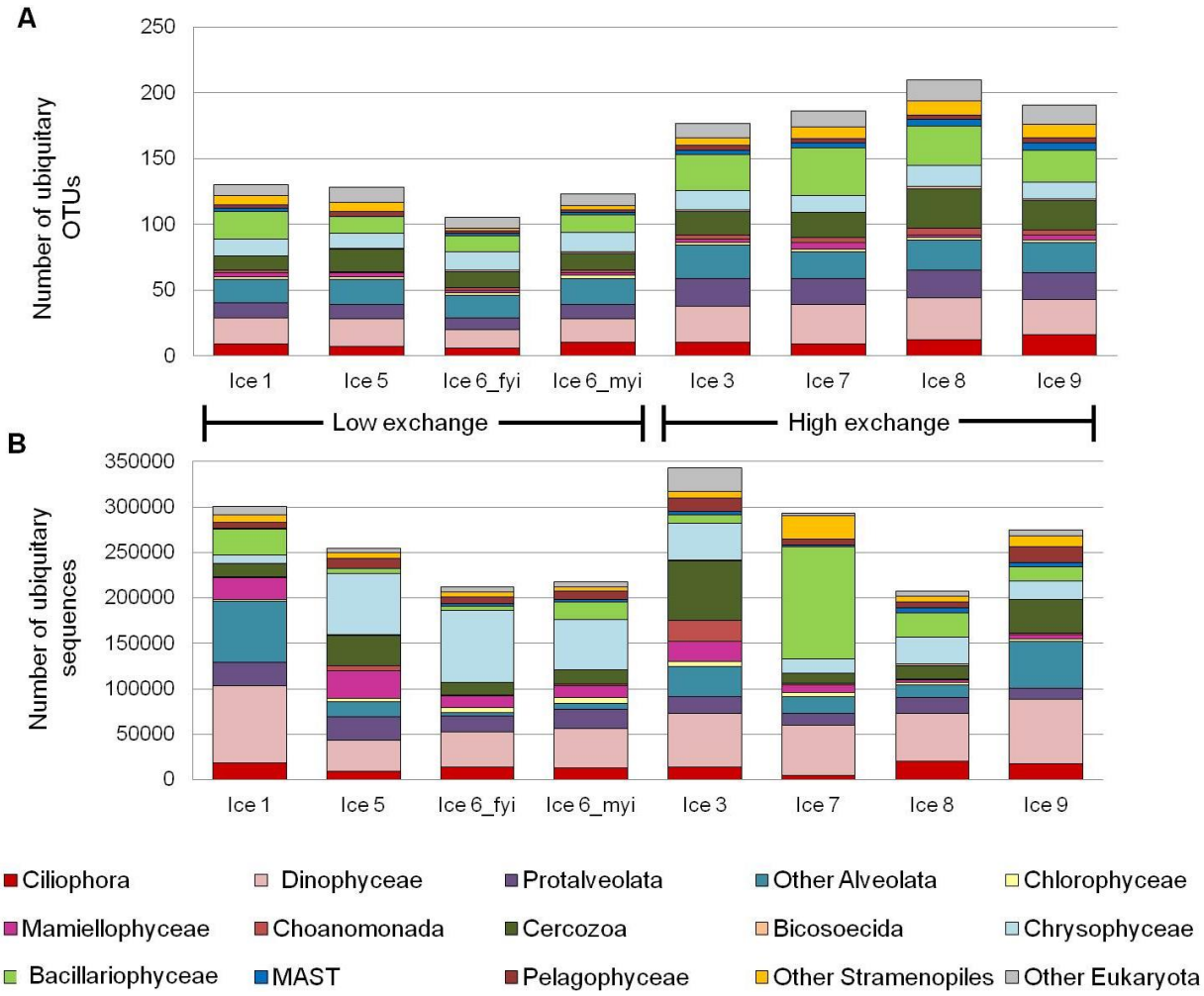


Fig. 5. Number of sequences (**A**) and number of OTUs (**B**) of ubiquitous protist taxa found in the deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during the IceArc expedition. Based on the number of shared OTUs (**Tab. 2**), the stations were classified as 'low-exchange' or 'high-exchange' stations.

Most ubiquitous genera showed clear habitat preferences (**Fig. 6**, **Fig. A1**). Two of the most dominant representatives were *Gymnodinium* (Dinophyceae) and *Ochromonas* (Chrysophyceae). *Gymnodinium* was characteristic for DCM, particularly at high-exchange stations. The same was true for another dinoflagellate genus (*Karlodinium*). In contrast, *Ochromonas* was highly abundant in MW, mainly at the FYI station 6, where the habitat interaction was relatively low (**Fig. 6**). Other examples for ubiquitous genera were *Cochlodinium* (Dinophyceae) in MW and *Micromonas* (Chlorophyceae) in UIW. The latter was highly abundant in UIW at low-exchange stations while its contribution to the under-ice community was minor at the other stations.

In the ICE samples, the diatom genera *Melosira*, *Navicula* and a Naviculaceae-strain "CCMP2297" were ubiquitous. OTUs of the sub-ice diatom *Melosira* were classified in all four habitats of the ice stations, except of ice station 5 and 6 (**Fig. 6**). The highest abundance of *Melosira* was observed at the MYI stations 7 and 8. At station 7, we identified comparable sequence proportions of this centric diatom genus in all four localities, however, at station 8, it was mainly detected in DCM, which indicates the sinking of this genus. The Naviculaceae strain "CCMP2297" was mainly found in ICE, particularly in the MYI of station 6 (**Fig. 6**). In contrast, other *Navicula* species were more abundant in DCM with no clear abundance maximum.

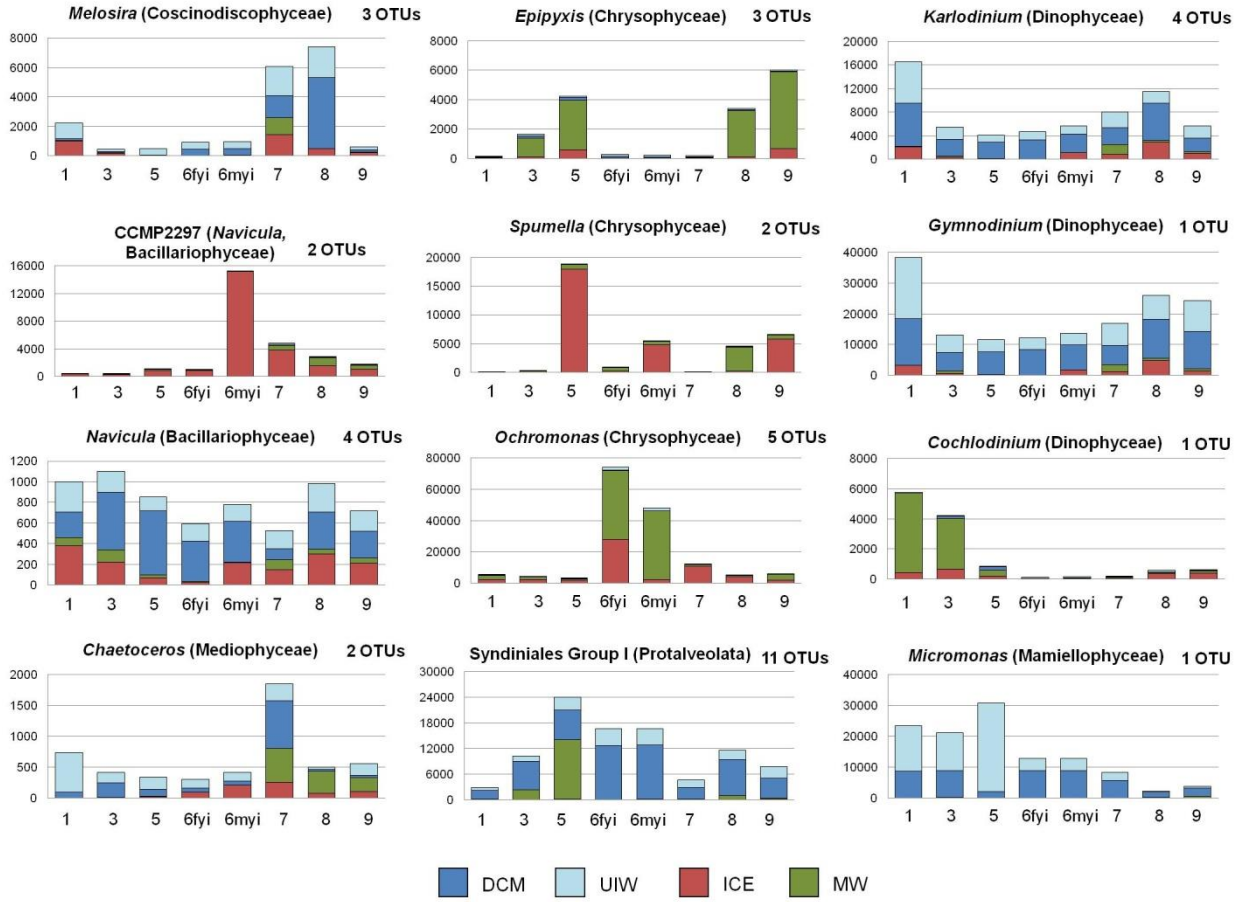


Fig. 6. Sequence abundances and number of OTUs of some ubiquitous taxa or taxa groups found in the deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during the IceArc expedition. Note the different dimensions between axes. An overview of all ubiquitous protist groups per station can be found in Fig. A1.

Physico-chemical conditions at the ice stations in 2012

During the IceArc expedition, we sampled five FYI cores between 81 and 120 cm in length and three MYI cores between 139 and 194 cm in length (**Tab. 1**). Sea ice concentrations were above 80% at all stations except of ice 5 (49%) while reaching 100%, at ice stations 7, 8 and 9 sampled at the end of the productive season in September. Water temperatures measured in UIW and DCM varied between -1.8°C and -1.5°C (**Tab. 3**), with lowest average temperatures at the high-exchange stations. ICE temperatures, averaged for the entire core, were also lowest at the high-exchange stations (**Tab. 3**). Atmospheric freezing conditions were experienced, at the high-exchange stations and station 5, which was indicated by very low ice surface temperatures (-8.5°C to -2.3°C in the upper 10 cm). Bottom ice temperatures, however, indicated sea surface freezing conditions at only the high-exchange stations. Salinity measured in ICE was rarely above 3 psu with maximum values at low-exchange station 1, and the high-exchange stations 7 and 8 (**Tab. 3**). For all nutrients measured in the habitats, the highest values were generally observed in the DCM. The nutrient profiles were similar for the high-exchange stations, which was indicated by the clustering of these stations in the PCA (**Fig. 7**).

Tab. 3. Stations sampled during IceArc expedition. Values of salinity, dissolved nutrients and water temperature measured for deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) are given. For ICE, averages of salinity, temperature and dissolved organic nutrients phosphate (PO_4) silicate (Si), nitrite (NO_2) and nitrate (NO_3), were calculated for entire sea ice cores. In addition, average (Avg.) ICE salinity and temperature are given for the entire sea ice cores, the top (first 10 cm) and bottom (bottom 10 cm). For all physico-chemical parameters, average and standard deviation (Stdev.) were calculated for the low-exchange stations (Ice 1, 5 and 6) and high-exchange stations (Ice 3, 7, 8 and 9), as classified based on number of unique and shared OTUs between the habitats.

	Temperature (°C)	**Salinity (psu)**	**PO_4 (µmol/l)**	**Si (µmol/l)**	**NO_2 (µmol/l)**	**NO_3 (µmol/l)**
Ice 1_DCM	-1.77	34.18	0.41	1.54	0.18	4.73
Ice 3_ DCM	-1.68	33.74	0.22	1.23	0.11	1.13
Ice 5_ DCM	-1.48	31.04	0.28	3.42	0.10	0.90
Ice 6_ DCM	-1.55	31.47	0.34	4.68	0.10	1.71
Ice 7_ DCM	-1.80	33.11	0.23	1.52	0.02	1.00
Ice 8_ DCM	-1.79	32.93	0.28	2.23	0.02	1.68
Ice 9_ DCM	-1.79	32.78	0.28	1.15	0.02	2.18
Avg. (Stdev.) low-exchange	-1.6 (0.15)	32.23 (1.70)	0.34 (0.07)	3.21 (1.58)	0.13 (0.05)	2.45 (2.02)
Avg. (Stdev.) high-exchange	-1.77 (0.06)	33.14 (0.42)	0.25 (0.03)	1.53 (0.49)	0.04 (0.05)	1.50 (0.54)
Ice 1_UIW	-1.57	32.99	0.28	1.15	0.04	2.95
Ice 3_UIW	-1.59	32.61	0.09	1.08	0.01	0.01
Ice 5_UIW	-1.62	30.54	0.18	3.45	0.01	0.04
Ice 6_UIW	-1.54	30.00	0.20	4.80	0.00	0.04
Ice 7_UIW	-1.80	33.17	0.23	1.52	0.02	0.97
Ice 8_UIW	-1.80	33.07	0.23	1.75	0.03	0.64
Ice 9_UIW	-1.80	32.83	n.d.	n.d.	n.d.	n.d.
Avg. (Stdev.) low-exchange	-1.58 (0.04)	31.18 (1.59)	0.22 (0.06)	3.13 (1.85)	0.02 (0.02)	1.01 (1.68)
Avg. (Stdev.) high-exchange	-1.75 (0.11)	32.92 (0.25)	0.18 (0.08)	1.45 (0.34)	0.02 (0.01)	0.54 (0.49)
Ice 1_ICE	Avg.: -0.66 Top: 0 Bot.: -1.4	Avg.: 2.05 Top: 1.6 Bot.: 2.8	n.d.	n.d.	n.d.	n.d.
Ice 3_ICE	Avg.: -0.71 Top: -0.3 Bot.: -1.6	Avg.: 1.89 Top: 1.3 Bot.: 2.5	0.03	0.41	0.01	0.25
Ice 5_ICE	Avg.: -0.71 Top: -2.3 Bot.: -1.2	Avg.: 1.36 Top: 3.7 Bot.: 1.2	0.04	0.19	0.01	0.20
Ice 6_FYI	Avg.: -0.66 Top: -0.2 Bot.: -1.3	Avg.: 0.94 Top: 0.2 Bot.: 2.1	0.01	0.11	0.01	0.17
Ice 6_MYI	Avg.: -0.72 Top: -0.7 Bot.: -1.1	Avg.: 1.25 Top: 0.3 Bot.: 2	0.02	0.60	0.01	0.05
Ice 7_ICE	Avg.: -1.94 Top: -2.9 Bot.: -1.9	Avg.: 3.19 Top: 3 Bot.: 3.5	0.09	0.10	0.02	1.55
Ice 8_ICE	Avg.: -1.65 Top: -5 Bot.: -1.8	Avg.: 2.39 Top: 2.5 Bot.: 3.2	0.02	0.29	0.02	2.31

Tab. 3 continued

	Temperature (°C)	Salinity (psu)	PO_4 (µmol/l)	Si (µmol/l)	NO_2 (µmol/l)	NO_3 (µmol/l)
Ice 9_ICE	Avg.: -2.75 Top: -8.5 Bot.: -1.7	Avg.: 1.74 Top: 0.3 Bot.: 2.7	0.01	0.09	0.01	0.62
Average (Stdev.) low-exchange	Avg: -0.69 (0.03) Top: -0.80 (0.90) Bot: -1.25 (0.11)	1.40(0.41) 1.45 (1.41) 2.03 (0.57)	0.02 (0.01)	0.15 (0.26)	0.01 (0)	0.18 (0.08)
Average (Stdev.) high-exchange	Avg: -1.76 (0.73) Top: -4.18 (3.0) Bot: -1.75 (0.11)	2.30 (0.57) 1.78 (1.05) 2.98 (0.40)	0.04 (0.04)	0.22 (0.16)	0.02 (0.01)	1.18 (0.93)
Ice 1_MW	-1.10	2.50	0.06	0.51	0.01	0.80
Ice 3_MW	0.20	0.40	-0.02	0.43	0.07	0.63
Ice 5_MW	-0.10	0.45	0.04	0.11	0.04	0.65
Ice 6_MW	0.00	0.00	0.29	3.92	0.01	0.27
Ice 7_MW	0.00	31.00	0.19	0.35	0.01	0.19
Ice 8_MW	0.00	3.40	0.12	1.18	0.02	0.99
Ice 9_MW	-0.10	0.80	0.08	0.09	0.03	0.74
Average (Stdev.) low-exchange	-0.40 (0.61)	0.98 (1.33)	0.13 (0.14)	1.51 (2.10)	0.02 (0.01)	0.57 (0.28)
Average (Stdev.) high-exchange	0.03 (0.13)	8.90 (14.79)	0.09 (0.09)	0.51 (0.47)	0.03 (0.03)	0.64 (0.33)

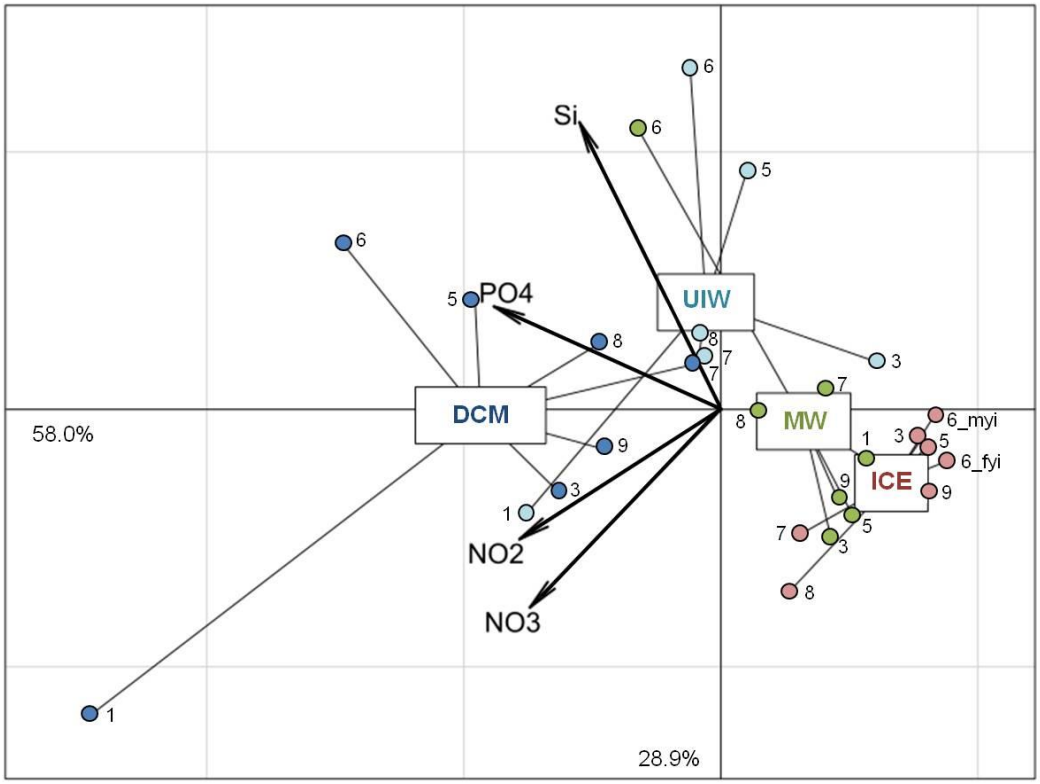


Fig. 7. PCA of dissolved inorganic nutrients measured in the deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during the IceArc expedition. No nutrient data were available for UIW at ice station 9 and for ICE at ice station 1.

Correlation of environmental parameters with protists

Using the Mantel test, we tested for potential correlations between the most common protist groups (Bacillariophyceae, Dinophyceae and Chrysophyceae) and physico-chemical parameters measured in the respective habitats. A statistical summary is presented in **Tab. 4**. Bacillariophyceae occurrences in DCM were best explained by ICE thickness, temperature and salinity as well as nutrient concentrations in ICE. Bacillariophyceae in UIW had a weak correlation with ICE-nutrients. Bacillariophyceae and Dinophyceae in UIW samples, showed a strong significant correlation with incoming PAR and ice thickness, while those sampled from DCM showed a possible trend with incoming PAR and ice thickness. In contrast to UIW-Bacillariophyceae, UIW-Dinophyceae had a strong significant correlation with DCM nutrient concentrations. This was also observed for Chrysophyceae, where we additionally observed a correlation with UIW nutrient concentrations.

UIW salinity and temperature was an important driver for the occurrences of ICE algae (Bacillariophyceae, Dinophyceae and Chrysophyceae), as all tested groups showed significant correlations with these variables. Bacillariophyceae in MW had a significant correlation with ICE temperature and salinity. In addition, Bacillariophyceae and Dinophyceae in MW had significant correlations with MW salinity and temperature, while Chrysophyceae in MW had a trend for correlation ($p=0.07$) with these parameters. However, when only testing the occurrences of ubiquitous OTUs belonging to the Chrysophyceae *Spumella*, *Epipyxis* and *Ochromonas* ("CCMP1899"), we observed a significant correlation with MW nutrient concentrations ($r = 0.55$, $p = 0.03$, not shown in Tab. 4) as well as MW salinity and temperature ($r = 0.51$, $p = 0.03$, not shown in Tab. 4).

Tab. 4. Statistical summary of the Mantel test performed to test for correlations between occurrences (Jaccard index) of the most prominent ubiquitous taxa and environmental (Euclidean) distance matrices calculated for the physico-chemical parameters of the four habitats sampled during the IceArc expedition. Nuts., dissolved inorganic nutrients; Temp., temperature; Sal., salinity; PAR, incoming photosynthetically active radiation; Thickn., ICE thickness; n.s., no significance.

		Bacillariophyceae				**Dinophyceae**				**Chrysophyceae**			
		DCM	**UIW**	**ICE**	**MW**	**DCM**	**UIW**	**ICE**	**MW**	**DCM**	**UIW**	**ICE**	**MW**
DCM	Nuts.	n.s.	n.s.	n.s.	n.s.	n.s.	r = 0.65, p = 0.01	n.s.	n.s.	n.s.	r = 0.72, p = 0.01	n.s.	n.s.
	Temp. + Sal.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	r = 0.46, p = 0.03	n.s.	n.s.	n.s.	n.s.
UIW	Nuts.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	r = 0.63, p = 0.04	n.s.	n.s.
	Temp. + Sal.	n.s.	n.s.	r = 0.41, p = 0.05	n.s.	n.s.	n.s.	r = 0.56, p = 0.001	r = 0.57, p = 0.02	n.s.		r = 0.42, p = 0.05	n.s.
ICE	Nuts.	r = 0.70, p = 0.01	r = 0.66, p = 0.06	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	Temp. + Sal.	n.s.	n.s.	n.s.	r = 0.48, p = 0.03	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	Temp. + Sal. + Thickn.	r = 0.66, p = 0.004	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	PAR + Thickn.	r = 0.52, p = 0.07	r = 0.83, p = 0.02	n.s.	n.s.	r = 0.39, p = 0.06	r = 0.79, p = 0.03	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
MW	Nuts.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	r = 0.45, p = 0.02	n.s.
	Temp. + Sal.	n.s.	n.s.	n.s.	r = 0.59, p = 0.05	n.s.	n.s.	n.s.	n.s.	r = 0.64, p = 0.01	n.s.	n.s.	r = 0.49, p = 0.07

Discussion

During the last decades, severe climate related changes have been observed in the Arctic Ocean, with important consequences for ecosystem functioning (e.g. primary production, consumption and biogeochemical fluxes). Current changes in sea ice conditions include a reduction in sea ice thickness and proportion of MYI (Maslanik et al., 2011; Comiso 2012) and increased melt pond coverage of the predominantly FYI covered Arctic Ocean (Rosel and Kaleschke 2012). In September 2012, the lowest sea ice extent was measured since the onset of satellite observations (Parkinson and Comiso 2013). Several models have projected a summerly ice-free Arctic Ocean, lasting for at least five consecutive summers, before the middle of the 20th century (Kirtman et al., 2013). These sea ice related changes have already had an impact on pelagic phytoplankton and ice algal communities with alterations to trophic food webs and biogeochemical cycling (e.g. Wassmann 2011; Arrigo et al., 2008; 2014a, Leu et al., 2011). To detect current and predict ongoing changes to Arctic ecosystems, it is indispensable to investigate habitat-specific protist biodiversity and their exchange between the habitats, in particular the mechanisms of protist community incorporation in sea ice.

The CAO harbors highly-adapted habitat-specific communities

Protist community composition associated with the water column (UIW and DCM) and sea ice (ICE and MW) differed considerably from each other. Our results demonstrate that communities in UIW and DCM were mainly characterized by Dinophyceae and Protalveolata (Jensen and Hansen 2000; Monier et al., 2013; Kilias et al., 2014a), while communities in ICE and MW were mainly characterized by Bacillariophyceae, Cercozoa and Chrysophyceae, which are consistent to previous studies (Booth and Horner 1997; Poulin et al., 2011; Kilias et al., 2014b; Fernández-Méndez et al., 2015). To further understand protist habitat-preferences, we identified the taxa that were abundant in one habitat by assessing the number of unique OTUs per habitat at each ice station. This number refers to protist taxa that potentially prefer one habitat over the others and thus, might be an indication for habitat-preference or restriction due to physico-environmental constraints or barriers. Furthermore, some unique OTUs may correspond to protist taxa that are highly adapted to the physico-chemical conditions of their living environment (Horner 1985; Arrigo 2014b). The identification of habitat restricted and specialized species is an important issue in the context of climate change. Species restricted to one habitat or specific environmental conditions, might be

endangered as they are not able to react to habitat loss or fast environmental changes in their living environment (Myers et al., 2000; Lovejoy and Potvin 2011).

We identified sea ice as the habitat with the highest diversity and abundance of unique taxa, which were mainly classified as Bacillariophyceae. This highlights sea ice as a unique habitat and hotspot of protist biodiversity. In particular, at MYI station 6, 17% of the entire sea ice community sequences were exclusive to sea ice. A high proportion of unique OTUs corresponded to the pennate diatom genus *Nitzschia.* These typical sea ice algae have a high morphologic diversity and phylogenetic complexity (Lundholm et al., 2002; Comeau et al., 2013), and were also found in bottom ice sections or UIW in other studies (Horner et al., 1992; Michel et al., 2002; Kaartokallio et al., 2007). Considered as a true ice alga, for example, is *Nitzschia frigida* (Syvertsen 1991). Recent studies demonstrated that *N. frigida* is very sensitive to changes in light conditions, which influence biomass (Enberg et al., 2015) and the sedimentation characteristics of the species (Aumach and Juhl 2015). Aumach and Juhl (2015) found that *N. frigida* assemblages sink slower under low light conditions. As this could also be true for other *Nitzschia* species, this may explain why we observed high proportions of unique *Nitzschia* OTUs in the MYI, where the under ice light field is typically limited (Nicolaus et al., 2012), but also could be the result of lower incoming PAR at these stations due to latitude and time of the year (Fernández-Méndez et al., 2015). In addition, high concentrations of exopolysaccharides produced by many sea ice algae (Meiners et al., 2003) can slow down the rejection of brine water and ice algae from the brine channels into the water column (Krembs et al., 2011). These mechanisms could have resulted in reduced sedimentation of *Nitzschia* and other sea ice algae at the analyzed ice stations. Therefore, a further reduction of sea ice thickness could lead to increased ice-algal sedimentation and subsequently a lower proportion of sea ice algae. Overall, it remains uncertain how sea ice algae biomass will be impacted by continued loss of MYI (Lange et al., 2015). However, the declining sea ice cover is expected to result in the reduction of diversity (Melnikov 2009) and activity (Stecher et al., 2016) of sea ice flora. In this regard, rare species, and species restricted to sea ice are probably more affected than abundant and widely distributed species. One important factor influencing the diversity of sea ice algae, is the incorporation of algal cells from the water column into the sea ice (Syvertsen 1991; Niemi et al., 2011). As a consequence of climate change, sea ice is melting earlier, which leads to a later freeze-up in autumn because absorption is increased in open water (Perovich and Polashenski 2012; Stroeve et al., 2014). A time shift of sea ice formation towards the less productive autumn

could affect sea ice algae, for example by reducing their abundance and biomass (Niemi et al., 2011), particularly of species entrapped in the sea ice matrix during ice formation.

In DCM, unique taxa were mainly represented by Syndiniales and MAST (marine stramenopiles), while these taxa were rarely in the UIW habitat. Both groups recently gained more attention with the development of high-throughput sequencing technologies, which revealed high phylogenetic diversity (Massana et al., 2004; Not et al., 2009; Lin et al., 2012; Kilias et al., 2014a; Wolf et al., 2015). Syndiniales and MAST play an important role in the marine food web. Syndiniales infect dinoflagellates, fish and crustacean (Guillou et al., 2008) and members of MAST are considered as parasites or epiphytes of Bacillariophyceae (e.g. *Solenicola,* MAST-3) (Gómez et al., 2011).). They also tend to dominate winter communities in the Arctic fjords (Marquardt et al. 2016). Because these unique taxa were rarely found in the other habitats, we suggest that their abundance was highly influenced by the abundance or conditions of hosts in the DCM.

The UIW differed considerably from the DCM. We found much less unique OTUs in UIW most of which were classified as Ciliophora rather than Dinophyceae, which were typical for the DCM. Ciliophora were also part of the unique protist taxa found in MW. Similar to other unique OTUs found in our study, these taxa were most likely well adapted to the conditions under sea ice, e.g. available food sources. Ciliophora play an essential role as grazers of pico- and nanoplankton that live in UIW, MW or algae aggregates at the bottom of the melt ponds (Assmy et al., 2013; Kilias et al., 2014b). *Strombidium* for example (e.g. station 5), is an aloricate tintinnid ciliate genus with size classes of 40 to 60 µm, which has also been observed in sea ice and under-ice water of the Baltic Sea (Kaartokallio et al., 2007; Rintala et al., 2010). Its importance in the food webs has been shown during spring blooms in the Dutch Wadden Sea and the coastal North Sea (Admiraal and Venekamp 1986). *Strombidium* was reported to feed on particulate matter that sedimented from ice (Michel et al., 2002; Kaartokallio et al., 2007), which could explain its unique occurrence in the UIW. Another example is the mixotrophic genus *Stokesia* found in the MW at MYI station 8. *Stokesia* is a large (> 100 µm) ciliate, which contains endosymbiotic green algae and has been found during spring blooms in oligo-mesotrophic lakes (Amblard et al., 1993; Posch et al., 2015; Przytulska et al., 2016).

Taken together, the results confirmed our first hypothesis that different habitats in the sea ice-covered Arctic Ocean harbor unique, habitat-specific protist communities.

Ubiquitous taxa still have habitat-preferences

High-exchange stations were associated with a high number of ubiquitous OTUs, which could indicate potential generalists of the Arctic habitats. However, ubiquitous taxa found in all four habitats, still showed distinct habitat preferences, which indicates their need for specific environmental conditions. Thus, we suggest that all ubiquitous protist taxa found in this study were ecological specialists. In contrast to ecological generalists, ecological specialists have a narrow niche (breadth of adaptation) (reviewed by Kassen 2002) and thus, are most likely highly adapted to the physico-chemical conditions of their living environment (Horner 1985; Arrigo 2014b).

A considerably high proportion of ubiquitous OTUs was classified as Dinophyceae, which mainly dominated the water column, e.g. *Gymnodinium*. This observation is in contrast to a study conducted in Arctic fjords, where *Gyrodinium* spp. dominated the water column community (Marquardt et al 2016). Overall, the ubiquitous distribution of Dinophyceae showed less dependency on abiotic factors compared to autotrophic groups. This was indicated by the lack of correlation for the occurrence of Dinophyceae in all samples with nutrients and light conditions measured in the habitats (with the exception of UIW Dinophyceae, which correlated with DCM nutrients). *Gymnodinium* and other marine Alveolates were also identified as generalists in subsurface samples collected in the eastern English Channel (Genitsaris et al., 2015). Our observation is also in agreement with a comparative study, which found the highest number of shared OTUs between the Arctic and Antarctic Ocean for Alveolates, while the number was lower for chlorophytes, haptophytes and stramenopiles (Wolf et al., 2015). The authors suggested that the alveolates have a bipolar distribution because they are less influenced by abiotic factors than autotrophic protists. However, as Dinophyceae are mixotrophic and heterotrophic, they depend on the abundance of their food source in DCM and UIW (e.g. small flagellates and phytoplankton).

Another important ubiquitous representative was the golden algae *Ochromonas* (Chrysophyceae); a freshwater genus that was mostly found in MW by this and other studies (Bursa 1963; Kilias et al., 2014b). In contrast to *Ochromonas*, another Chrysophyceae genus, *Spumella*, was mostly found in ICE, demonstrating different habitat-preferences within the same family (Chromulinaceae). Chrysophyceae occurrences of both ICE and MW groups, however, were mainly associated with melt ponds as habitat because they significantly correlated with the MW physico-chemical parameters.

In summary, we could demonstrate that ubiquitous taxa found in all habitats still had distinct habitat preferences as their occurrences were closely related to habitat-specific environmental conditions or available food sources.

Sea ice melting and formation trigger protist exchange in the CAO

To assess which driving forces trigger protist exchange between the habitats, we analyzed the number of unique, shared and ubiquitous OTUs at the ice stations. We assumed that a low number of unique taxa and a high number of shared taxa were indicative of a high protist exchange between the habitats. Given the fact that biogeochemical and physical processes of sea ice are highly complex and vary strongly between and even within the ice types, this study gives valuable insights into the mechanisms of protist exchange.

Based on our observations, we present the most-likely scenarios of protist exchange and propose two major exchange-types. The first type of exchange commenced not later than at the beginning of the sampling period in August, when sea ice was in a state of advanced melt and reached a record minimum extent in mid-September (Parkinson and Comiso 2013). In general, this type was characterized by an enhanced protist exchange between ICE, UIW and DCM. However, the overall protist exchange among all habitats was low and we therefore classified these stations as low-exchange stations (**Fig. 8**). Yet, a high exchange was the case for the second type, which occurred at the end of September, during the onset of freeze-up when bottom ice and sea surface waters were near the sea water freezing point (-1.8°C and a salinity of ~ 34 psu) (Golden et al., 1998) and new ice was formed. In contrary to our initial assumption, this type of exchange was characterized by an overall high protist exchange between all habitats. Particularly MW-associated protists contributed significantly to the overall exchange (**Fig. 8**).

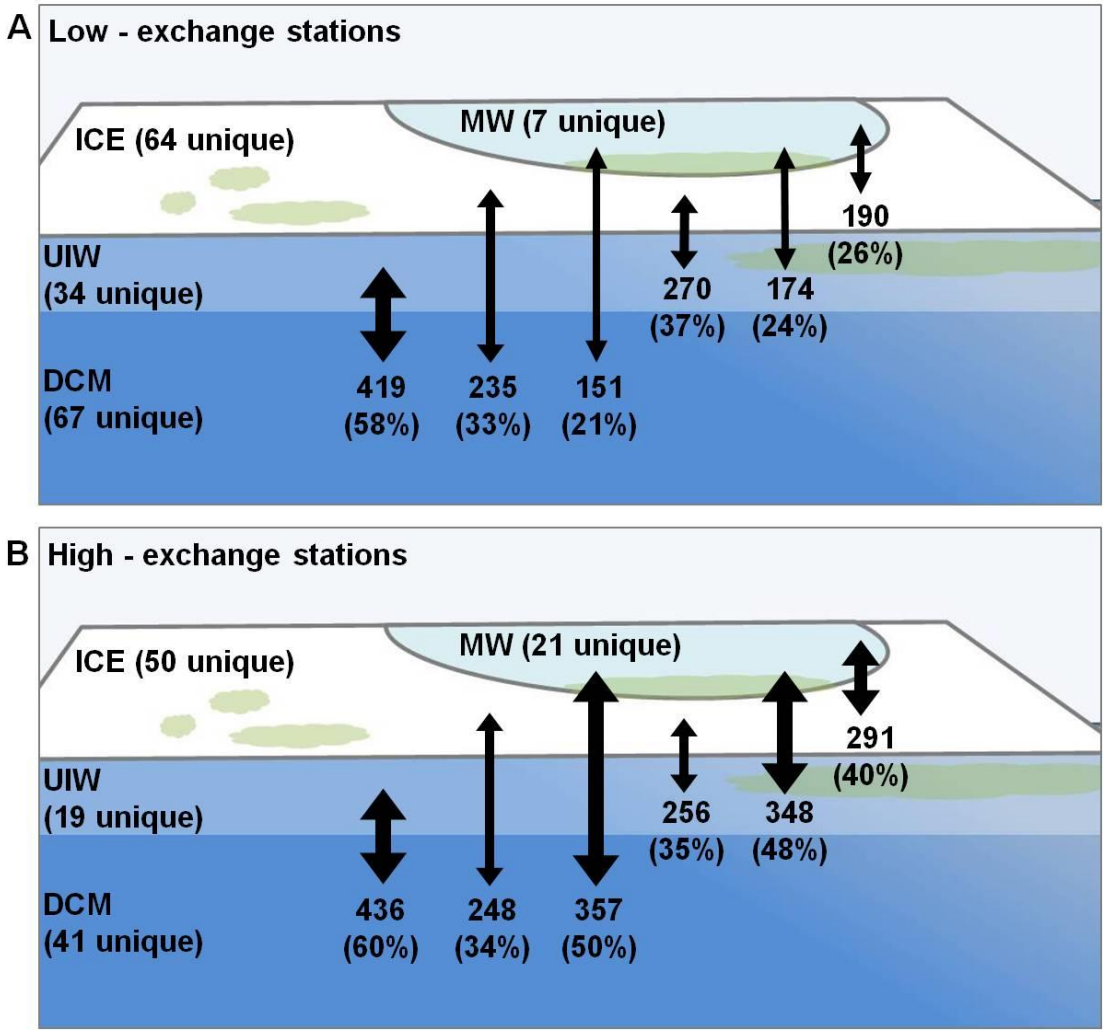


Fig. 8. Schematic drawing of habitat-interactions of protists at the ice stations sampled in the CAO during summer 2012. Given are the number of unique OTUs found in the deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) sampled at (**A**) low-exchange stations (Ice 1, 5 and 6) and (**B**) high-exchange stations (Ice 3, 7, 8, and 9). The arrows indicate the degree of interaction between two habitats. Numbers under the arrows indicate average number and, in brackets, proportion of shared OTUs observed at the ice stations. Total average number of OTUs found at the ice stations was 721. Number and proportion of unique and shared OTUs of each station respectively can be found in **Tab. 3**.

Protist exchange during sea ice melting

UIW and DCM always showed the highest number of shared OTUs between all habitats (**Fig. 8**). This exchange was always higher than the exchange between the adjacent habitats UIW and ICE, indicating that the UIW community mostly derived from the DCM instead of ICE. However, as a consequence of sea ice melt and release of sea ice algae (Ambrose et al., 2005; Juul-Pederson et al., 2008; Boetius et al., 2013), UIW also shared a remarkable number of taxa with ICE, particularly at the low-exchange stations 1 and 5 (**Fig. 8**). An example for a single specific genus could be observed at ice stations 1, 7 and 8, where high abundances of the ubiquitous under-ice algae *Melosira* was attached to the sea ice bottom. As a consequence of the sea ice melting, the species were released into the water column below the ice, which resulted in an unusually rapid sedimentation event of *Melosira arctica* to the deep sea floor in 2012 (Boetius et al., 2013).

The strong influence of sea ice melt on pelagic protist diversity was further confirmed by the close relationship of Bacillariophyceae occurrences with sea ice conditions. Bacillariophyceae in DCM, and partially in UIW, were significantly correlated with ICE temperature, thickness and dissolved inorganic nutrient concentrations in ICE as well as incoming PAR. Thus, their growth was possibly triggered by the more suitable light conditions (Perrette et al. 2011) present under thinner sea ice (e.g. thinner ice equals more light) at the beginning of the sampling (Nicolaus et al., 2012). Therefore, phytoplankton diversity will likely increase and thereby contribute significantly to protist exchange between the habitats in the future CAO. However, sea ice melt leads to a strong water stratification (Eilertsen 1993) which was possibly the main reason for the suppressed connectivity between all habitats and thus, overall low protist exchange. This assumption is in agreement with a recent modeling study which investigated the responses of phytoplankton to a seasonally ice-free Arctic Ocean (Lawrence et al., 2015). The authors found, that the depth-integrated net production will shift to deeper water layers and increase by about 30% by the end of the century (Lawrence et al., 2015). These changes are mainly triggered by reduced surface nitrate concentrations and increasing light conditions (Lawrence et al., 2015). Therefore, a shift of the chlorophyll maximum to deeper water layers will likely lead to a reduced interaction between water and sea ice habitats.

Protist exchange during sea ice formation

In late September, we observed an increase in protist diversity and abundance (**Fig. 8**), which was associated with an increased interaction of protists between the habitats as new sea ice formed. In the Arctic Ocean, sea surface water begins to freeze at a temperature of around -1.8°C and a salinity of around 34 psu (Golden et al., 1998). During ice formation, high salinity (high density) water from brine channels is rejected into the water column and new ice crystals are formed (Golden et al., 1998; 2007; Krembs et al., 2011). This downward transport of brine water, which is driven by its higher density, and growth of ice crystals might also have trapped MW protists and enhanced their downward transport into the ICE and finally into the UIW and DCM (Melnikov 1997; Lee et al., 2011). The incorporation of melt pond algae into the sea ice during freezing of surface ice was also reported by Lee et al., (2011) and was apparent at ice station 9. Furthermore, for some taxa we also observed connectivity between UIW and ICE at the high-exchange stations. For example, the ubiquitous genera *Chaetoceros* (Bacillariophyceae), *Karlodinium* and *Gymnodinium* (Dinophyceae), which were prominent in UIW and DCM, showed increased abundances in ICE during late-September. This suggests that protists from UIW were entrapped into the newly formed ice (Ackley et al., 1987; Melnikov 1997; Gradinger and Ikävalko 1998). The incorporation of pelagic protists into the sea ice (Syvertsen 1991; Niemi et al., 2011) is an important factor influencing the diversity and composition of sea ice algae communities. As a consequence of climate change, sea ice is melting earlier, which leads to a later freeze-up in autumn because absorption is increased in open water (Perovich and Polashenski 2012; Stroeve et al., 2014). A time shift of sea ice formation towards the less productive autumn could affect sea ice algae, for example by reducing their abundance and biomass, particularly of species entrapped in the sea ice matrix during ice formation (Niemi et al., 2011).

In summary, our hypothesis, that sea ice melt was associated with high protist exchange, was partially confirmed. Sea ice melting lead to an increased exchange of UIW and ICE but the overall protist exchange between the habitats was low and possibly not as pronounced as it was at the beginning of the melt season (Boetius et al., 2013). Yet, the exchange was mainly triggered by sea ice formation, where physico-biochemical processes during sea ice growth lead to increased habitat-connectivity by enhancing transport of protist cells through the ice matrix. This underlines our second hypothesis that the CAO still has unique habitats with highly adapted communities. Species introduced in new habitats might therefore form cysts to survive the unfavorable conditions (e.g. Dinophyceae) or, if they are not able to react to fast-changing conditions, might be extinguished.

Conclusions

This study is a significant contribution to our understanding of how current and future changes in sea ice conditions will impact protist communities in different habitats of the CAO. We identified unique and ubiquitous protist taxa, which were used to examine potential protist habitat-restrictions and interactions. Overall, our findings point towards a potential restructuring of protist communities and reduction of protist diversity in both, water column and sea ice as the sea ice cover continues to decline. We observed the highest abundance of unique taxa (mainly diatoms) in the sea ice, particularly MYI, highlighting the importance of this unique habitat in the CAO for the overall biodiversity.

Sea ice conditions were of pivotal importance for protist community structure and habitat interactions in the CAO. Our data suggest that sea ice melt prevents the overall-exchange of protists between the habitats, despite the increased protist exchange observed between UIW and ICE. In contrast, sea ice formation led to high habitat interactions and thus, is a major process for the overall biodiversity exchange of protists in the CAO. Therefore, a further reduction or loss of sea ice may lead to a decreased exchange rate between the habitats. Furthermore, the habitat-specific sea ice community could be affected by a time shift in sea ice formation. As a consequence of climate change, an earlier summer sea ice retreat and increased absorption of solar radiation due to more open water can lead to later freeze-up in autumn, when phytoplankton productivity in the CAO is lower. These temporal changes in the onset of melt and freeze-up have also large spatial variability, which are likely to have a large impact on protist community structure in the CAO. In particular, highly adapted and sea ice algae species with low abundances might be affected most. Therefore large-scale biogeographical and ecophysiological studies are required to further identify highly habitat adapted species in the CAO. In addition, future research should not only focus on habitat structure and dynamics in the CAO during the sea ice melt season but also during the polar night. In addition, more emphasis should be placed on protist communities in Arctic continental shelf regions, where a substantial part of sea ice is formed and thus, a high degree of protist exchange can be expected.

Acknowledgements

This study was accomplished within the Young Investigator Group PLANKTOSENS (VH-NG-500), funded by the Initiative and Networking Fund of the Helmholtz Association. We thank the captain and crew of the RV Polarstern for their support during the cruise ARKXXVII/3. We thank A. Nicolaus and K. Oetjen for excellent technical support in the laboratory.

References

Ackley, S.F., Dieckmann, G.S. Shen, H.T., 1987. Algal and foram incorporation into new sea ice. Trans. Am. Geophys. Union 68, 1736–1748.

Admiraal, W., Venekamp, L.A.H., 1986. Significance of tintinnid grazing during blooms of *Phaeocystis pouchetii* (Haptophyceae) in Dutch coastal waters. J. Sea Res. 20, 61–66.

Amblard, C., Simengando, T., Rachiq, S., Bourdier, G., 1993. Importance of ciliated protozoa in relation to the bacterial and phytoplanktonic biomass in an oligo–mesotrophic lake, during the spring diatom bloom. Aquat. Sci. 55, 1–9

Ambrose, W.G.Jr., von Quillfeldt, C., Clough, L.M., Tilney, P.V.R. Tucker, T., 2005. The sub–ice algal community in the Chukchi sea: large– and small–scale patterns of abundance based on images from a remotely operated vehicle. Polar Biol. 28, 1–12.

Andreassen, I., Nöthig, E.–M., Wassmann, P., 1996. Vertical particle flux on the shelf of northern Spitsbergen, Norway. Mar. Ecol. Prog. Ser. 137, 215–228.

Arrigo, K.R., van Dijken, G., Pabi, S., 2008. Impact of a shrinking Arctic ice cover on marine primary production. Geophys. Res. Lett. 35 (19).

Arrigo, K.R., Perovich, D.K., Pickart, R.S., Brown, Z.W., van Dijken, G.L., Lowry, K.E., Mills, M.M., Palmer, M.A., Balch, W.M., Bahr, F., Bates, N.R., Benitez–Nelson, C., Bowler, B., Brownlee, E., Ehn, J.K., Frey, K.E., Garley, R., Laney, S.R., Lubelczyk, L., Mathis, J., Matsuoka, A., Mitchell, B.G., Moore, G.W.K., Ortega–Retuerta, E., Pal, S., Polashenski, C.M., Reynolds, R.A., Schieber, B., Sosik, H.M., Stephens, M., Swift, J.H., 2012. Massive phytoplankton blooms under Arctic sea ice. Science 336, 1408, doi:10.1126/science.1215065.

Arrigo, K.R., Perovich, D.K., Pickart, R.S., Brown, Z.W., van Dijken, G.L., Lowry, K.E., Mills, M.M., Palmer, M.A., Balch, W.M., Bates, N.R., Benitez–Nelson, C.R., Brownlee, E., Frey, K.E., Laney, S.R., Mathis, J., Matsuoka, A., Mitchell, B.G., Moore, G.W.K., Reynolds, R.A., Sosik, H.M., Swift, J.H., 2014a. Phytoplankton blooms beneath the sea ice in the Chukchi Sea. Deep Sea Res. Part II Top. Stud. Oceanogr. 105, 1–16.

Arrigo, K.R., 2014b. Sea ice ecosystems. Ann. Rev. Mar. Sci. 6, 439–467.

Arrigo, K.R., van Dijken, G.L., 2015. Continued increases in Arctic Ocean primary production. Prog. Oceanogr.136, 60–70.

Assmy, P., Ehn, J.K., Fernandez–Mendez, M., Hop, H., Katlein, C., Sundfjord, S., Bluhm, K., Daase, M., Engel, A., Fransson, A., Granskog, M.A., Hudson, S.R., Kristiansen, S., Nicolaus, M., Peeken, I., Renner, A.H.H., Spreen, G., Tatarek, A., Wiktor, J., 2013. Floating ice–algal aggregates below melting Arctic sea ice. PLOS ONE 8. DOI:10.1371/journal.pone.0076599.

Aumack, C.F., Juhl, A.R., 2015. Light and nutrient effects on the settling characteristics of the sea ice diatom Nitzschia frigida. Limnol. Oceanogr. 60(3), 765–776.

Bakker 2014a; Nutrients measured on water bottle samples during POLARSTERN cruise ARK–XXVII/3, IceArc in 2012; doi:10.1594/PANGAEA.834081

Bakker 2014b; Nutrients measured on melt pond water samples during POLARSTERN cruise ARK–XXVII/3 (IceArc) in 2012; Unpublished dataset #834282

Boetius, A., Albrecht, S., Bakker, K., Bienhold, C., Felden, J., Fernández–Méndez, M., Hendricks, S., Katlein, C., Lalande, C., Krumpen, T., Nicolaus, M., Peeken, I., Rabe, B., Rogacheva, A., Rybakova, E., Somavilla, R., Wenzhöfer, F. RV Polarstern ARK27–3 Shipboard Science Party. 2013. Export of algal biomass from the melting Arctic sea ice. Science 339, 1430–1432. doi:10.1126/science.1231346.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality–filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods 10, 57–59.

Booth, B.C., Horner, R.A., 1997. Microalgae on the Arctic Ocean section, 1994: Species abundance and biomass. Deep–Sea Research II 44 (8), 1607–1622.

Bursa, A., 1963. Phytoplankton in Coastal Waters of the Arctic Ocean at Point Barrow, Alaska. Arctic 16 (4), 239–262.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., 2010. QIIME allows analysis of high–throughput community sequencing data. Nat. Methods 7, 335–336.

Carmack, E., Wassmann, P., 2006. Food webs and physical–biological coupling on pan–Arctic shelves: unifying concepts and comprehensive perspectives. Prog. Oceanogr.71, 446–477.

Comeau, A.M., Philippe, B., Thaler, M., Gosselin, M., Poulin, M., Lovejoy, C., 2013. Protists in Arctic drift and land-fast sea ice. J. Phycol. 49, 229–240. doi:101111/jpy12026.

Comiso, J.C., 2003. Warming Trends in the Arctic from Clear Sky Satellite Observations. J. Clim. 16 (21), 3498–3510.

Comiso, J.C., Parkinson, C.L., Gersten, R., Stock, L., 2008. Accelerated decline in the Arctic sea ice cover. Geophys. Res. Lett. 35, L01703. doi:10.1029/2007GL031972.

Comiso, J.C., 2012. Large Decadal Decline of the Arctic Multiyear Ice Cover. J. Clim. 25, 1176–1193.

David, C., Lange, B., Rabe, B. Flores, H., 2015. Community structure of under–ice fauna in the Eurasian central Arctic Ocean in relation to environmental properties of sea–ice habitats. Mar. Ecol. Prog. Ser. 522, 15–32.

Eddy, S.R., 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7 (10), e1002195.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinform. 26, 2460–2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinform. 27, 2194–2200.

Eilertsen, H.C., 1993. Spring blooms and stratification. Nature 363, 24.

Enberg, S., Piiparinen, J., Majaneva, M., Vähätalo, A.V., Autio, R., Rintala, J.M., 2015. Solar PAR and UVR modify the community composition and photosynthetic activity of sea ice algae. FEMS Microbiol. Ecol. 91 (10), fiv102.

Fernández–Méndez, M., Katlein, C., Rabe, B., Nicolaus, M., Peeken, I., Bakker, K., Flores, H. , Boetius, A., 2015. Photosynthetic production in the CAO during the record sea–ice minimum in 2012. Biogeosciences 12 (11), 3525–3549.

Genitsaris, S., Monchy, S., Viscogliosi, E., Sime–Ngando, T., Ferreira, S., Christaki, U., 2015. Seasonal variations of marine protist community structure based on taxon–specific traits using the eastern English Channel as a model coastal system. FEMS Microbiol. Ecol. 91 (5), fiv034.

Golden, K.M., Ackley, S.F., Lytle, V.I., 1998. The percolation phase transition in sea ice. Science 282, 2238–2241.

Golden, K.M., Eicken, H., Heaton, A.L., Miner, J., Pringle, D.J., Zhu, J., 2007. Thermal evolution of permeability and microstructure in sea ice. Geophys. Res. Lett. 34, L16501. doi:10.1029/2007GL030447.

Gómez, F., Moreira, D., Benzerara, K., López-García, P., 2011. Solenicola setigera is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. Environ. Microbiol. 13 (1), 193–202.

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., Massana, R., Scanlan, D.J., Worden, A.Z., 2008. Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environ. Microbiol. 10, 3349–3365.

Gradinger, R., Ikävalko, J., 1998. Organism incorporation into newly forming Arctic sea ice in the Greenland Sea. J. Plankton Res. 20, 871–886.

Hardge, K., Peeken, I., Neuhaus, S., Krumpen, T., Stoeck, T. Metfies K., Sea Ice Origin and Sea Ice Retreat are Major Drivers of Variability in Arctic Marine Protist Composition. submitted.

Hilligsøe, K.M., Richardson, K., Bendtsen, J., Sørensen, L.L., Nielsen, T.G., Lyngsgaard, M.M., 2011. Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux. Deep–Sea Research I 58, 826–838.

Horner, R., 1985. Ecology of sea ice microalgae. In: Horner, R. (ed) Sea ice biota. CRC, Fla. p. 83–103.

Horner, R.S., Ackley, F., Dieckmann, G.S., Gulliksen, B., Hoshiai, T., Legendre, L., Melnikov, I.A., Reeburgh, W.S., Spindler, M. Sullivan, C.W., 1992. Ecology of sea ice biota, 1: Habitat, terminology, and methodology. Pol. Biol. 12, 417–427.

Jensen, F., Hansen, W.B., 2000. Ciliates and heterotrophic dinoflagellates in the marginal ice zone of the central Barents Sea during spring. J. Mar. Biol. Assoc. 80, 45–54.

Juul–Pedersen, T., Michel, C., Gosselin, M., Seuthe L., 2008. Seasonal changes in the sinking export of particulate material under first–year sea ice on the Mackenzie Shelf (western Canadian Arctic). Mar. Ecol. Prog. Ser. 353, 13–25.

Kaartokallio, H., Kuosa, H., Thomas, D.N., Granskog, M.A., Kivi, K., 2007. Biomass, composition and activity of organism assemblages along a salinity gradient in sea ice subjected to river discharge in the Baltic Sea. Polar Biol. 30, 183–197.

Kattner, G., Becker, H., 1991. Nutrients and organic nitrogenous compounds in the marginal ice zone of the Fram Strait. J. Mar. Syst. 2, 385–394.

Kassen, R., 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. 15 (2), 173–190.

Kilias, E., Kattner, G., Wolf, C., Frickenhaus, S. Metfies, K., 2014a. A molecular survey of protist diversity through the CAO. Polar Biol. 37, 1271–1287. doi:10.1007/s00300–014–1519–5.

Kilias, E.S., Peeken, I., Metfies, K., 2014b. Insight into protist diversity in Arctic sea ice and meltpond aggregate obtained by pyrosequencing. Polar Res. 33. doi:http://dx.doi.org/10.3402/polar.v33.23466.

Kirtman, B., Power, S.B., Adedoyin, J.A., Boer, G.J., Bojariu, R., Camilloni, I., Doblas-Reyes, F.J., Fiore, A.M., Kimoto, M., Meehl, G.A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G.J., Vecchi, G., Wang, H.J., 2013. Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F, Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kohlbach, D., Graeve, M., Lange, B.A., David, C., Peeken, I., Flores H., (in press), The importance of ice algae-produced carbon in the central Arctic Ocean ecosystem: food web relationships revealed by lipid and stable isotope analyses. Limnol. Oceanogr, doi: 10.1002/lno.10351, in press.

Krembs, C., Eicken, H., Deming, J.W., 2011. Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic. Proc. Natl. Acad. Sci. USA 108, 3653–58.

Kwok, R., Rothrock, D.A., 2009. Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. Geophys. Res. Lett. 36, L15501. doi:10.1029/2009GL039035.

Laney, S.R., Sosik, H.M., 2014. Phytoplankton assemblage structure in and around a massive under–ice bloom in the Chukchi Sea. Deep Sea Res. Part 2 Top Stud. Oceanogr. 105, 30–41.

Lange, B.A., Michel, C., Beckers, J.F., Casey, J.A., Flores, H., Hatam, I. Meisterhans, G., Niemi, A. Haas, C., 2015. Comparing Springtime Ice–Algal Chlorophyll a and Physical Properties of Multi–Year and First–Year Sea Ice from the Lincoln Sea. PLoS ONE 10, e0122418. doi:10.1371/journal.pone.0122418.

Lawrence, J., Popova, E., Yool, A., Srokosz, M., 2015. On the vertical phytoplankton response to an ice-free Arctic Ocean, J. Geophys. Res. Oceans 120, 8571–8582, doi:10.1002/2015JC011180.

Laxon, S.W., Giles, K.A., Ridout, A.L., Wingham, D.J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krish, R., Kurtz, N., Farrell, S., Davidson, M., 2013. CryoSat-2 estimates of Arctic sea ice thickness and volume. Geophys. Res. Lett. 40, 1–6. doi:10.1002/GRL.50193.

Lee, S.H., Mcroy, C.P., Joo, H.M., Gradinger, R., Cui. X.H., Yun, M.S., Chung, K.H., Kang, S.H., Kang, C.K., Choy, E.J., Son, S.H., Carmack, E., Whitledge, T.E., 2011. Holes in Progressively Thinning Arctic Sea Ice Lead to New Ice Algae Habitat. Oceanography 24, 302–308.

Leu, E., Soreide, J.E., Hessen, D.O., Falk–Petersen, S., Berge, J., 2011. Consequences of changing sea–ice cover for primary and secondary producers in the European Arctic shelf seas: Timing, quantity, and quality. Prog. Oceanogr. 90, 18–32.

Li, W.K.W., Mclaughlin, F.A., Lovejoy, C., Carmack, E.C., 2009. Smallest Algae Thrive As the Arctic Ocean Freshens. Science 326, 539–539.

Lin, Y.C., Campbell, T., Chung, C.C., Gong, G.C., Chiang, K.P., Worden, A.Z., 2012. Distribution patterns and phylogeny of marine stramenopiles in the North Pacific Ocean. Applied and Environ. Microbiol. 78 (9), 3387–3399.

Lovejoy, C., Potvin, M., 2011. Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. J. Plankt. Res. 33 (3), 431–444. doi:10.1093/plankt/fbq124.

Lundholm, N., Daugbjerg, N., Moestrup, Ø., 2002. Phylogeny of the Bacillariaceae with emphasis on the genus Pseudo–nitzschia (Bacillariophyceae) based on partial LSU rDNA. Eur. J. Phycol. 37 (01), 115–134.

Marquardt, M., Vader, A., Stubner, E.I., Reigstad, M., Gabrielsen, T.M., 2016. Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). Appl. Environ. Microb. 82, 1868–1880.

Maslanik, J., Stroeve, J., Fowler, C., Emery, W., 2011. Distribution and trends in Arctic sea ice age through spring 2011. Geophys. Res. Lett. 38, L13502. doi:10.1029/2011GL047735.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., Pedrós–Alió, C., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. 70 (6), 3528–3534.

Matsen, F.A., Kodner, R.B., Armbrust, E.V., 2010. pplacer: linear time maximum–likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11, 538.

Melnikov, I.A. 1997. The Arctic Sea Ice Ecosystem. Gordon and Breach Science Publishers, Amsterdam, 204 pp.

Melnikov, I.A., Nihoul, J.C.J., Kostianoy, A.G., 2009. Recent sea ice ecosystem in the Arctic Ocean: a review. In: Influence of Climate Change on the Changing Arctic and Sub-Arctic Conditions. ed. Nihoul, J.C.J., Kostianoy, A.G., pp. 57–71. NATO Sci. Peace Secur. Ser. C. Dordrecht: Springer.

Meiners, K., Gradinger, R., Fehling, J., Civitarese, G., Spindler, M., 2003. Vertical distribution of exopolymer particles in sea ice of the Fram Strait (Arctic) during autumn. Mar. Ecol. Prog. Ser. 248, 1–13.

Michel, C., Gissel Nielsen, T., Nozais, C., Gosselin, M., 2002. Significance of sedimentation and grazing by ice micro– and meiofauna for carbon cycling in annual sea ice (northern Baffin Bay). Aquat. Microb. Ecol. 30, 57–68.

Miller, L.A., Fripiat, F., Else, B.G.T., Bowman, J.S., Brown, K.A., Collins, R.E., Ewert, M., Fransson, A., Gosselin, M., Lannuzel, D., Meiners, K.M., Michel, C., Nishioka, J., Nomura, D., Papadimitriou, S., Russell, L.M., Sørensen, L.L., Thomas, D.N., Tison, J.–L., Van Leeuwe, M.A., Vancoppenolle, M., Wolff, E.W., Zhou, J., 2015. Methods for biogeo–chemical studies of sea ice: the state of the art, caveats, and recommendations. Elem. Sci. Anthropol. 3, 1–53. doi:10.12952/journal.elementa.000038.

Monier, A., Terrado, R., Thaler, M., Comeau, A., Medrinal, E., Lovejoy, C., 2013. Upper Arctic Ocean water masses harbor distinct communities of heterotrophic flagellates. Biogeosciences 10 (6), 4273–4286.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 6772, 853–858.

Nicolaus, M., Katlein, C., Maslanik, J. Hendricks, S., 2012. Changes in Arctic sea ice result in increasing light transmittance and absorption. Geophys. Res. Lett. 39, L24501. doi:10.1029/2012GL053738.

Niemi, A., Michel, C., Hille, K. Poulin, M., 2011. Protist assemblages in winter sea ice: Setting the stage for the spring ice algal bloom. Polar Biol. 34, 1803–1817.

Not, F., del Campo, J., Balagué, V., de Vargas, C., Massana, R., 2009. New insights into the diversity of marine picoeukaryotes. PLoS One 4 (9), e7143.

Nöthig E.–M., Bracher A., Engel A., Metfies K., Niehoff B., Peeken I., Bauerfeind E., Cherkasheva A., Gäbler–Schwarz S., Hardge K., Kilias E., Kraft A., Mebrahtom Kidane Y., Lalande C., Piontek J., Thomisch K. Wurst M., 2015. Summertime plankton ecology in Fram Strait—a compilation of long– and short–term observations. Polar Res. 34, 23349. doi: http://dx.doi.org/10.3402/polar.v34.23349.

Parkinson, C.L., Comiso, J.C., 2013. On the 2012 record low Arctic sea ice cover: combined impact of preconditioning and an August storm. Geophys. Res. Lett. 40, 1356–1361.

Perovich, D.K., Polashenski, C., 2012. Albedo evolution of seasonal Arctic sea ice. Geophys. Res. Lett. 39, L08501. doi:101029/2012GL051432.

Perrette, M., Yool, A., Quartly, G.D. Popova, E.E., 2011. Near–ubiquity of ice–edge blooms in the Arctic. Biogeosciences 8, 515–524.

Posch, T., Eugster, B., Pomati, F., Pernthaler, J., Pitsch, G., Eckert, E.M., 2015. Network of Interactions Between Ciliates and Phytoplankton During Spring. Front. Microbiol. 6, 1289. doi: 10.3389/fmicb.2015.01289.

Poulin, M., Daugbjerg, N., Gradinger, R.., Ilyash, L., Ratkova, T., Quillfeldt, C. von, 2011. The pan–Arctic biodiversity of marine pelagic and sea–ice unicellular eukaryotes: A first–attempt assessment. Marine Biodiversity 41, 13–28.

Przytulska, A., Comte, J., Crevecoeur, S., Lovejoy, C., Laurion, I., and Vincent, W.F., 2016. Phototrophic pigment diversity and picophytoplankton in permafrost thaw lakes. Biogeosciences 13, 13–26. doi:10.5194/bg–13–13–2016.

Quillfeldt, C.H. von, 2000. Common diatom species in Arctic spring blooms: their distribution and abundance. Botanica Marina 43, 499–516.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R–project.org/.

Rabe et al., 2013. Physical oceanography measured on water bottle samples during POLARSTERN cruise ARK–XXVII/3 (IceArc). doi:10.1594/PANGAEA.819452.

Riedel, A., Michel, C., Gosselin, M., LeBlanc, B., 2008. Winter–spring dynamics in sea–ice carbon cycling in the coastal Arctic Ocean. J. Mar. Syst. 74 (3), 918–932.

Rintala, J.M., Piiparinen, J., Uusikivi, J., 2010. Drift–ice and under–ice water communities in the Gulf of Bothnia (Baltic Sea). Polar Biol. 33 (2), 179–191.

Rosel, A., L. Kaleschke., 2012. Exceptional melt pond occurrence in the years 2007 and 2011 on the Arctic sea ice revealed from MODIS satellite data. J. Geophys. Res. Oceans 117 (C5).

Różańska, M., Poulin, M. Gosselin, M., 2008. Protist entrapment in newly formed sea ice in the coastal Arctic Ocean. J. Mar. Syst. 74, 887–901.

Sankelo, P., Haapala, J., Heiler, I., Rinne, E.,. 2010. Melt pond formation and temporal evolution at the drifting station Tara during summer 2007. Polar Res. 29, 311–321.

Søreide, J.E., Carroll, M.L., Hop, H., Ambrose Jr, W.G., Hegseth, E.N., Falk–Petersen, S., 2013. Sympagic–pelagic–benthic coupling in Arctic and Atlantic waters around Svalbard revealed by stable isotopic and fatty acid tracers. Mar. Biol. Res. 9, 831–850.

Spreen, G., Kaleschke, L. Heygster, G., 2008. Sea ice remote sensing using AMSR–E 89 GHz channels. J. Geophys. Res. 113, C02S03. doi:10.1029/2005JC003384.

Stoeck,T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H., Richard, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19, 21–31.

Stroeve, J.C., Markus, T., Boisvert, L., Miller, J., Barrett, A., 2014. Changes in Arctic melt season and implications for sea ice loss. Geophys. Res. Lett. 41, 1216–1225. doi:101002/(2013)GL058951.

Syvertsen, E.E., 1991. Ice algae in the Barents Sea: types of assemblages, origin, fate and role in the ice-edge phytoplankton bloom. Polar Res. 10 (1), 277–288.

Tremblay, G., Belzile, C., Gosselin, M., Poulin, M., Roy, S., Tremblay, J.E., 2009. Late summer phytoplankton distribution along a 3500 km transect in Canadian Arctic waters: strong numerical dominance by picoeukaryotes. Aquat. Microb. Ecol. 54, 55–70.

Vergin, K.L., Beszteri, B., Monier, A., Thrash, J.C., Temperton, B., Treusch, A.H., Kilpert, F., Worden, A.Z., Giovannoni, J., 2013. High–resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time–series Study site by phylogenetic placement of pyrosequences. ISME J. 7, 1322–1332.

Wassmann, P., 2008. Impact of global warming on Arctic pelagic ecosystems and processes, p. 113–140. In C. M. Duarte [ed.], Impacts of global warming on polar ecosystems. Fundación BBVA.

Wassmann, P., 2011. Arctic marine ecosystems in an era of rapid climate change. Prog. Oceanogr. 90 (1), 1–17.

Wolf, C., Kilias, E., Metfies, K., 2015. Protists in the polar regions: comparing occurrence in the Arctic and Southern oceans using pyrosequencing. Polar Res. 34. doi: 10.3402/polar.v34.23225.

Appendix

Table A.1. Taxonomy and number of OTUs found in one habitat (unique OTUs) and found in two habitats of all ice station sampled during IceArc expedition in 2012.

Unique OTUs found in one habitat	**DCM**	**UIW**	**ICE**	**MW**
Alveolata; Ciliophora; Conthreep; Phyllopharyngea; Aporthotrochilia	0	0	0	4
Alveolata; Ciliophora; Litostomatea	0	0	6	0
Apusomonadidae; Amastigomonas	0	0	5	0
Chloroplastida; Prasinophytae	0	0	5	0
Rhizaria; Cercozoa; Glissomonadida; Heteromita	1	0	0	0
Rhizaria; Cercozoa; NovelClade2	5	0	0	0
Rhizaria; Cercozoa; Silicofilosea; Marimonadida; NAMAKO-15	0	0	3	0
Rhizaria; Cercozoa; Silicofilosea; Thaumatomonadida; Allas	0	0	1	0
Stramenopiles	0	0	7	0
Stramenopiles; Diatomea_1	0	0	7	0
Stramenopiles; Diatomea_2	0	0	6	0
Stramenopiles; Diatomea_3	0	0	4	0
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; Nitzschia	0	0	8	0
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; Sellaphora	0	0	4	0

OTUs found in two habitats	**DCM**	**UIW**	**ICE**	**MW**
Unknown eukaryote	0	0	8	1
Alveolata_1	0	0	6	1
Alveolata_2	3	0	2	0
Alveolata_3	0	0	1	4
Alveolata_4	0	0	8	1
Alveolata_5	0	0	7	1
Alveolata_6	0	0	4	1
Alveolata; Ciliophora; Conthreep; Phyllopharyngea	1	0	1	0
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; uncultured_1	2	1	0	0
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; uncultured_2	3	2	0	0
Alveolata; Dinoflagellata	0	0	4	5
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Gyrodinium	0	0	8	1
Alveolata; Dinoflagellata; Dinophyceae; Peridiniphycidae	7	7	0	0
Alveolata; OLI11255	0	0	6	6
Alveolata; Protalveolata; Syndiniales_1	0	1	0	3
Alveolata; Protalveolata; Syndiniales_2	3	3	0	0
Alveolata; Protalveolata; Syndiniales_3	0	0	8	6
Alveolata; Protalveolata; Syndiniales; Amoebophrya_1	6	6	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_2	5	1	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_3	5	5	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_4	7	4	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_5	7	3	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_6	4	4	0	0

Table A.1. continued

OTUs found in two habitats	**DCM**	**UIW**	**ICE**	**MW**
Alveolata; Protalveolata; Syndiniales; Amoebophrya_7	5	1	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_8	6	1	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_9	0	0	6	1
Alveolata; Protalveolata; Syndiniales; Amoebophrya_10	6	3	0	0
Alveolata; Protalveolata; Syndiniales; Amoebophrya_11	4	3	0	0
Alveolata; Protalveolata; Syndiniales; Duboscquella	5	3	0	0
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupI	5	4	0	0
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupII	5	1	0	0
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupII	7	4	0	0
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupII	5	3	0	0
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupV	6	5	0	0
Chloroplastida; Prasinophytae; Pyramimonas	0	0	8	1
Chloroplastida; Trebouxiophyceae; Chlorella	1	1	0	0
Discosea; Flabellinia; Vannellida; Protosteliopsis	0	0	1	3
Rhizaria; Cercozoa; NovelCladeGran-3	0	2	4	0
Rhizaria; Cercozoa; Silicofilosea	0	3	1	0
Rhizaria; Cercozoa; Thecofilosea; Ebriacea; Ebria	0	1	7	0
Rhizaria; Cercozoa; Thecofilosea; WHOI-LI1-14	2	4	0	0
Rhizaria; Radiolaria	5	0	0	1
Rhizaria; Radiolaria; Polycystinea; Spumellaria	5	1	0	0
Rhizaria; Radiolaria; RADC	2	3	0	0
Stramenopiles; Bicosoecida; Cafeteriidae	3	1	0	0
Stramenopiles; Bicosoecida; Cafeteriidae; Symbiomonas	0	0	8	4
Stramenopiles; Bolidomonas	0	0	7	2
Stramenopiles; Diatomea_1	1	0	8	0
Stramenopiles; Diatomea_2	0	1	7	0
Stramenopiles; Diatomea_3	0	0	6	1
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae	0	0	7	6
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; CCMP2297	0	0	7	6
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; CCMP2297	0	0	5	1
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; CCMP2297	0	0	8	3
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae; Chaetoceros	7	1	0	0
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae; Chaetoceros	0	0	6	5
Stramenopiles; Labyrinthulomycetes; Thraustochytriaceae; AB3F14RJ3E10	5	4	0	0
Stramenopiles; MAST-3	4	1	0	0
Stramenopiles; Pelagophyceae	1	0	7	0

Table A.2. Taxonomy and number of OTUs found in all habitats of all ice station (ubiquitous taxa) sampled during IceArc expedition in 2012.

Ubiquitous taxa	No. OTUs
Unknown eukaryote	8
Alveolata	31
Alveolata; Ciliophora	1
Alveolata; Ciliophora; Conthreep	3
Alveolata; Ciliophora; Conthreep; Oligohymenophorea	2
Alveolata; Ciliophora; Conthreep; Oligohymenophorea; Homalogastra	2
Alveolata; Ciliophora; Conthreep; Oligohymenophorea; Paramecium	1
Alveolata; Ciliophora; Conthreep; Oligohymenophorea; Porpostoma	1
Alveolata; Ciliophora; Conthreep; Oligohymenophorea; Stokesia	1
Alveolata; Ciliophora; Conthreep; Phyllopharyngea	1
Alveolata; Ciliophora; Conthreep; Phyllopharyngea; Aporthotrochilia	3
Alveolata; Ciliophora; Conthreep; Plagiopylea	2
Alveolata; Ciliophora; Heterotrichea; Peritromus	1
Alveolata; Ciliophora; Litostomatea	3
Alveolata; Ciliophora; Litostomatea; Haptoria	2
Alveolata; Ciliophora; Litostomatea; Haptoria; Didinium	3
Alveolata; Ciliophora; Litostomatea; Haptoria; Epiphyllum	1
Alveolata; Ciliophora; Litostomatea; Haptoria; Loxophyllum	1
Alveolata; Ciliophora; Litostomatea; Haptoria; Phialina	2
Alveolata; Ciliophora; Spirotrichea; Choreotrichia	3
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; Codonella	1
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; Metacylis	1
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; Salpingella	1
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; Strobilidium	1
Alveolata; Ciliophora; Spirotrichea; Choreotrichia; uncultured	7
Alveolata; Ciliophora; Spirotrichea; Hypotrichia; Anteholosticha	1
Alveolata; Ciliophora; Spirotrichea; Hypotrichia; Oxytricha	1
Alveolata; Ciliophora; Spirotrichea; Oligotrichia	10
Alveolata; Ciliophora; Spirotrichea; Oligotrichia; Pseudotontonia	1
Alveolata; Ciliophora; Spirotrichea; Oligotrichia; uncultured	1
Alveolata; Dinoflagellata	14
Alveolata; Dinoflagellata; Dinophyceae	1
Alveolata; Dinoflagellata; Dinophyceae; D244	2
Alveolata; Dinoflagellata; Dinophyceae; Dinophysiales	1
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae	10
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Chytriodinium	1
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Cochlodinium	3
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; FV18-2D9	4
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Gyrodinium	1
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Karlodinium	5
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Lepidodinium	2
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Pelagodinium	1
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Spiniferodinium	3

Table A.2. continued

Ubiquitous taxa	No. OTUs
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Symbiodinium	1
Alveolata; Dinoflagellata; Dinophyceae; Gymnodiniphycidae; Woloszynskia	2
Alveolata; Dinoflagellata; Dinophyceae; NPK60-44	2
Alveolata; Dinoflagellata; Dinophyceae; Peridiniphycidae; Alexandrium	1
Alveolata; Dinoflagellata; Dinophyceae; Peridiniphycidae; Protoperidinium	1
Alveolata; Dinoflagellata; Dinophyceae; Peridiniphycidae; Scrippsiella	1
Alveolata; Dinoflagellata; Dinophyceae; Prorocentrales; Exuviaella	1
Alveolata; Dinoflagellata; Dinophyceae; SCM16C67	1
Alveolata; Dinoflagellata; Dinophyceae; SL163A10	2
Alveolata; Dinoflagellata; Haplozoon	1
Alveolata; H67	1
Alveolata; OLI11255	3
Alveolata; Protalveolata; Syndiniales	11
Alveolata; Protalveolata; Syndiniales	1
Alveolata; Protalveolata; Syndiniales; Amoebophrya	25
Alveolata; Protalveolata; Syndiniales; Duboscquella	4
Alveolata; Protalveolata; Syndiniales; SyndinialesGroup	1
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupI	14
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupII	17
Alveolata; Protalveolata; Syndiniales; SyndinialesGroupIII	4
Alveolata; SCM37C52	2
Ancyromonadida; Ancyromonas	1
Chloroplastida; Chlorophyceae	2
Chloroplastida; Chlorophyceae; Carteria	1
Chloroplastida; Chlorophyceae; Fasciculochloris	1
Chloroplastida; Mamiellophyceae	2
Chloroplastida; Mamiellophyceae	1
Chloroplastida; Mamiellophyceae; Bathycoccus	1
Chloroplastida; Mamiellophyceae; DSGM-81	2
Chloroplastida; Mamiellophyceae; Micromonas	2
Chloroplastida; Prasinophytae; Pyramimonas	4
Chloroplastida; Trebouxiophyceae; AN1-3	3
Chloroplastida; Trebouxiophyceae; Chlorella	3
Chloroplastida; uncultured	3
Cryptophyceae; Cryptomonadales	2
Cryptophyceae; Cryptomonadales; Cryptomonas	1
Cryptophyceae; Cryptomonadales; FV18-2G7	1
Cryptophyceae; Cryptomonadales; Geminigera	1
Cryptophyceae; Cryptomonadales; Teleaulax	1
Cryptophyceae; Goniomonas	1
Haptophyta; Prymnesiophyceae	1
Haptophyta; Prymnesiophyceae; Phaeocystis	1
Haptophyta; Prymnesiophyceae; Prymnesiales	3
Haptophyta; Prymnesiophyceae; Prymnesiales; Imantonia	1

Table A.2. continued

Ubiquitous taxa	No. OTUs
Holozoa	1
Holozoa; Choanomonada; Acanthoecida	1
Holozoa; Choanomonada; Acanthoecida; Acanthocorbis	2
Holozoa; Choanomonada; Acanthoecida; Diaphanoeca	5
Holozoa; Choanomonada; Acanthoecida; Stephanoeca	1
Lobosa; Tubulinea; Euamoebida; Hartmannella	1
Rhizaria; Cercozoa	3
Rhizaria; Cercozoa; Glissomonadida; Heteromita	2
Rhizaria; Cercozoa; NovelClade4	1
Rhizaria; Cercozoa; NovelCladeGran-3	2
Rhizaria; Cercozoa; Silicofilosea	2
Rhizaria; Cercozoa; Silicofilosea; 7-5.4	2
Rhizaria; Cercozoa; Silicofilosea; Chlorarachniophyta; BC52	1
Rhizaria; Cercozoa; Silicofilosea; Chlorarachniophyta; NOR26	1
Rhizaria; Cercozoa; Silicofilosea; Marimonadida; Auranticordis	1
Rhizaria; Cercozoa; Silicofilosea; Marimonadida; NAMAKO-15	2
Rhizaria; Cercozoa; Silicofilosea; p15D09	2
Rhizaria; Cercozoa; Silicofilosea; Thaumatomonadida	4
Rhizaria; Cercozoa; Silicofilosea; Thaumatomonadida; Allas	1
Rhizaria; Cercozoa; Silicofilosea; Thaumatomonadida; D6	1
Rhizaria; Cercozoa; Thecofilosea	7
Rhizaria; Cercozoa; Thecofilosea; Cryomonadida; Cryothecomonas	6
Rhizaria; Cercozoa; Thecofilosea; Ebriacea; Ebria	7
Rhizaria; Cercozoa; Thecofilosea; NIF-3A7	3
Rhizaria; Cercozoa; Thecofilosea; NOR26	1
Rhizaria; Cercozoa; Thecofilosea; uncultured	2
Rhizaria; Radiolaria	1
Rhizaria; Radiolaria; Acantharia; Chaunocanthida	2
Rhizaria; Radiolaria; Polycystinea; Nassellaria; Pseudocubus	1
Stramenopiles	9
Stramenopiles; Bicosoecida; Cafeteriidae; BCI5F15RM3E05	1
Stramenopiles; Bicosoecida; Cafeteriidae; Cafeteria	1
Stramenopiles; Bicosoecida; LG08-10	1
Stramenopiles; Bicosoecida; Siluaniidae; Caecitellus	2
Stramenopiles; Bolidomonas	4
Stramenopiles; Chrysophyceae	12
Stramenopiles; Chrysophyceae; CCMP1899	6
Stramenopiles; Chrysophyceae; Chromulinales; Spumella	2
Stramenopiles; Chrysophyceae; CladeC	1
Stramenopiles; Chrysophyceae; LG01-09	1
Stramenopiles; Chrysophyceae; Ochromonadales; Epipyxis	3
Stramenopiles; Chrysophyceae; Ochromonadales; Paraphysomonas	3
Stramenopiles; Diatomea	16
Stramenopiles; Diatomea; 3b-F4	1

Table A.2. continued

Ubiquitous taxa	**No. OTUs**
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae	19
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; CCMP2297	5
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; Navicula	4
Stramenopiles; Diatomea; Bacillariophytina; Bacillariophyceae; Nitzschia	3
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae	1
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae; Attheya	1
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae; Chaetoceros	3
Stramenopiles; Diatomea; Bacillariophytina; Mediophyceae; Porosira	1
Stramenopiles; Diatomea; Coscinodiscophytina; Fragilariales	2
Stramenopiles; Diatomea; Coscinodiscophytina; Melosirids; Melosira	3
Stramenopiles; Diatomea; ME-Euk-FW10	1
Stramenopiles; Dictyochophyceae; Dictyochales; Dictyocha	1
Stramenopiles; Dictyochophyceae; NIF-1D10	2
Stramenopiles; Dictyochophyceae; Pedinellales	3
Stramenopiles; Labyrinthulomycetes; D2P04F01	1
Stramenopiles; Labyrinthulomycetes; D52	1
Stramenopiles; Labyrinthulomycetes; Thraustochytriaceae; AB3F14RJ3E10	1
Stramenopiles; MAST	1
Stramenopiles; MAST-1; MAST-1A	2
Stramenopiles; MAST-1; MAST-1C	2
Stramenopiles; MAST-2	1
Stramenopiles; MAST-3	4
Stramenopiles; MAST-8	1
Stramenopiles; Pelagophyceae	4
Stramenopiles; Pelagophyceae; Sarcinochrysidales; SS1-E01-69	4
Stramenopiles; Xanthophyceae; Tribonematales	1
Telonema; IncertaeSedis	7
Total	**488**

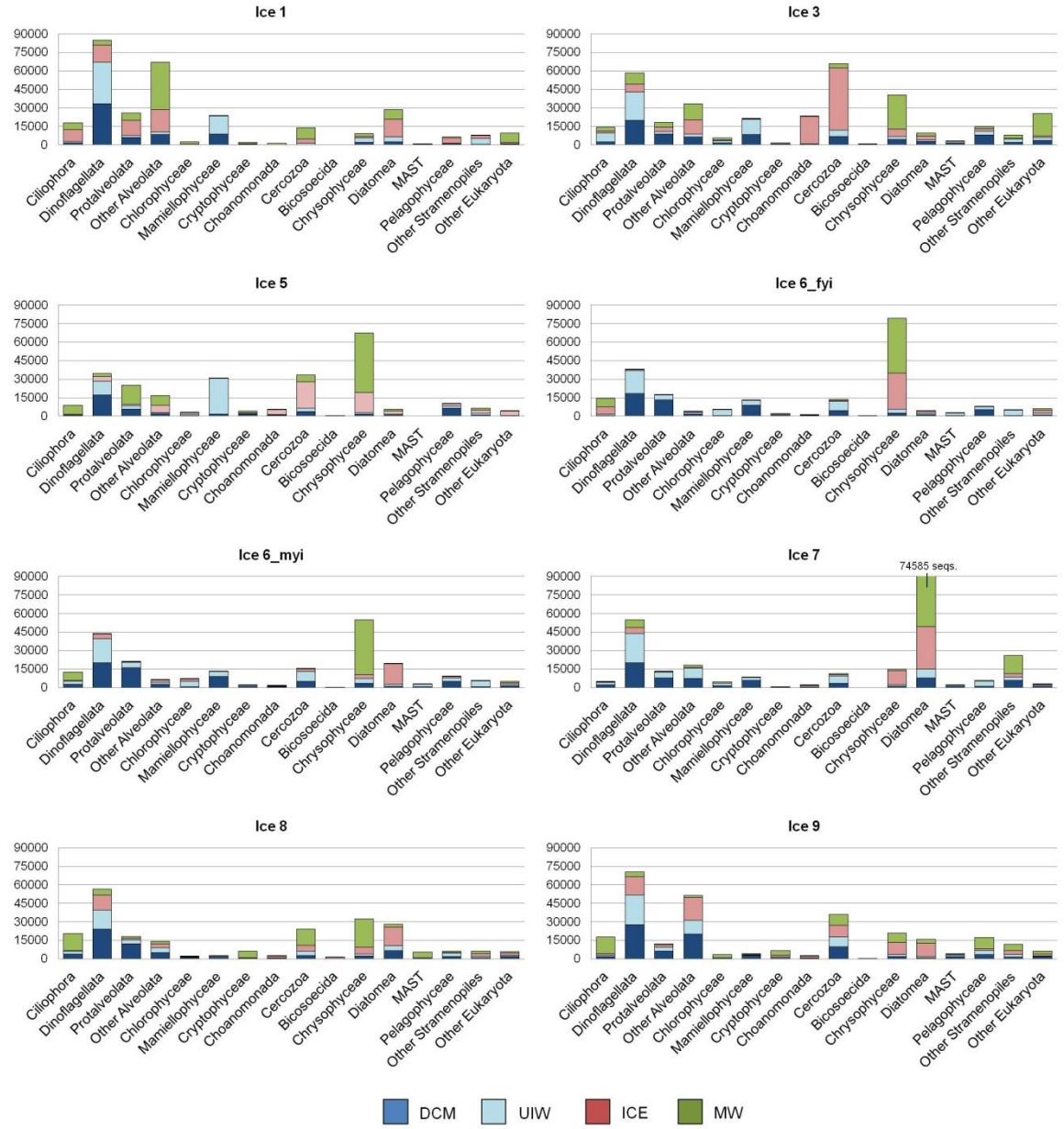


Fig. A.1. Sequence abundances of ubiquitous protist groups observed in deep-chlorophyll maximum layer depth (DCM), under-ice water (UIW), sea ice cores (ICE) and melt pond water (MW) samples collected during IceArc expedition.

8 Synoptic Discussion

In the Arctic Ocean, protistan plankton in water and sea ice serves as a baseline for higher trophic levels and ecosystem functioning. This thesis investigated prevailing protist community compositions and diversities in the changing Arctic Ocean using molecular methods as they allow for comprehensive analyses of protist communities. One important step in investigating complex protist community patterns and their relationship with environmental conditions is to accurately identify protist taxa. Thus, one has to get familiar with the most up-to-date molecular method, next-generation sequencing (NGS). In this regard, the **first aim** of this thesis was to investigate the influences of different sequence processing procedures on resulting protist diversity estimates obtained from environmental samples (**Manuscript I**). In a next step, sequencing was combined with ARISA to investigate the possible impacts of sea ice retreat and sea ice origin on Arctic protist community structure in ICE and DCM (**second aim, Manuscript II**). While **Manuscript II** analyzed the influence of sea ice on large-scale protist community patterns, **Manuscript III** focused on small-scale protist community patterns. In this connection, the habitat-specificity of Arctic protist communities and the local exchange between sea ice-influenced habitats was examined (**third aim**). A comprehensive overview of protist community structure in several habitats was given and sea ice as driving force that triggers the protist exchange between these habitats was discussed.

8.1 Applied methods for the assessment of protists in environmental samples

In this thesis, different conventional and molecular methods were applied to provide a comprehensive and detailed analysis of protist communities in the CAO. Light microscopy was used as a reference method for comparing protist diversity with the estimates obtained from sequencing. ARISA was used to uncover large-scale biogeographic patterns, while correlations of protist occurrences with physico-biogeochemical parameters were confirmed with high confidence. Finally, high-throughput sequencing allowed the characterization of protist taxa, particularly the identification of unique and habitat-specific or rare and widely distributed protists in the CAO. In the following section, the method applicability including pros and cons will be discussed briefly.

8.1.1 Light microscopy

A differentiation of species according to their morphology is the first and most obvious method to assess species community structures. One important advantage of light microcopy is that protist cells and conditions during sampling (fecal pellets of zooplankton, debris, and sediments) can be directly observed. For example, cell observation with light microscopy in **Manuscript I** revealed that the cells of *Chaetoceros* were much smaller and the plasma of individual cells seemed to have been shrunk. Thus, the cells looked less healthy than cells of *Thalassiosira* (**Manuscript I**). These differences in cell conditions have most likely influenced the efficiency of DNA isolation and sequencing. This could have led to a much lower sequence abundance of *Chaetoceros* compared to *Thalassiosira* (**Manuscript I**). In addition, the number of 18S rRNA gene copies may vary strongly between species and correlate mainly with cell size (Egge et al., 2012, Prokopowich et al 2003, Godhe et al. 2008). However, due to the additional information gained with microscopy, we found that *Chaetoceros* actually had a higher contribution to the community composition than sequencing suggested (**Manuscript I**).

On the other hand, light microscopy analysis of numerous samples can be very time consuming. In **Manuscript I**, counting and identification of species was done for 2 x 50 ml (two replicates per sample, i.e. 100 ml per sample). This counting volume is not comparable to the filtered water volume of ~2 liter used for sequencing. Consequently, low abundant diatom genera that were detected with sequencing (e.g. *Haslea* or *Skeletonema*), were most likely missed with light microscopy. In addition, light microscopy requires the analysis of distinct species-specific morphological features that are unambiguously distinguishable from morphological features of other species, which is not necessarily the case. In many cases, morphological features of unicellular organisms are hard to recognize or can be missed in very small species such as pico- and nanoplankton. Therefore, even if an excellent expertise in taxonomy is given, an accurate identification of species via light microscopy is not always possible (Quillfeldt 2001, Hoppenrath et al. 2007). In this thesis, the identification of species was done based on preserved water samples without any additional preparation of cells, which complicated species identification. A comparison of microscopy with sequencing revealed that possible misidentifications might have occurred for the similar-looking centric diatom genera *Thalassiosira* and *Porosira* or *Chaetoceros* and *Attheya* (**Manuscript I**).

Cryptic species or species complexes can share the same or highly similar morphology while having different genetic characters, which was also reported for diatoms (Amato et al. 2007, Poulíčková et al. 2010, Kermarrec et al. 2013). Unveiling these complex phylogenic relationships within or between species is important for the understanding of ecosystem diversity and its functioning, as each species or strain could possess different ecophysiological adaptations (Lovejoy et al. 2006, Foulon et al. 2008, Pfandl et al. 2009, Worden et al. 2009, Metfies et al. 2016).

8.1.2 Sequencing and molecular fingerprinting

When morphological character traits are missing or are not distinct enough to enable reliable taxonomic identification of protists, molecular methods are helpful. In this thesis, this was particularly demonstrated for diatom genera. Light microscopy and sequencing revealed that the centric diatom genus *Thalassiosira* was a significant part of the protist community in the water samples (**Manuscript I, II and III**). *Thalassiosira* is a paraphyletic genus, which species occur in several groups of the phylogenetic tree, such as *Detonula, Cyclotella, Minidiscus* and *Skeletonema* (Hoppenrath et al. 2007). The complex phylogenetic relationship between these similar-looking genera can be best assessed with sequencing of the hyper-variable V4 region of the 18S rRNA gene (Zimmermann et al. 2011, Luddington et al. 2012, Kermarrec et al. 2013, 2014). This was also shown in **Manuscript I**, where *Thalassiosira, Minidiscus* and *Skeletonema* were only detected if sequence processing was done with the in-house established sequence processing pipeline (DNAStar/PhyloAssigner), which included a sequence assignment based on phylogeny. In contrast, *Minidiscus* and *Skeletonema* were missed if sequences were processed with QIIME and mothur. Both pipelines used different parameter settings than DNAStar/PhyloAssigner and a sequence annotation based on sequence similarity.

The choice of sequence processing methods significantly affected the resulting species composition and diversity estimates. **Manuscript I** showed that a reasonable sequence quality trimming is important for a reliable protist community diversity estimate. The resulting number of OTUs (i.e. the diversity) was considerably higher when applying the in-house established PhyloAssigner pipeline (Vergin et al. 2013) for sequence processing, compared to the conducted QIIME (Caporaso et al. 2010) and mothur (Schloss et al. 2009) pipelines. However, protist diversity was most likely overestimated in the DNAStar/PhyloAssigner-based pipelines because a less stringent quality filtering was applied (no denoising) and no

sequence average quality score was set. In addition, sequences were longer after the quality-filtering and therefore, accumulation of sequencing errors toward the sequence end most likely reduced the sequence quality (Schloss et al. 2011, Huse et al. 2007) in the PhyloAssigner pipelines. This probably lead to a higher number of spurious OTUs which showed mainly low abundances (< 1%) (Gaspar and Thomas 2013) compared to the other pipelines. In contrast, sequence processing in QIIME and mothur resulted in an underestimation of protist diversity because quality-filtering (e.g. sequence quality score, denoising) and additional OTU-cluster size cutoff (not applied in the DNAStar/PhyloAssigner pipeline versions) was most likely too stringent. Therefore, it is necessary to strike a balance between a low sequence quality with potential overestimation of diversity, and a high sequence quality with potential underestimation of diversity, especially if 454 pyrosequences are analyzed.

During the course of this thesis, a transition from 454 pyrosequencing to Illumina sequencing has occurred, and 454 pyrosequencing is no longer available since summer 2015. However, the resulting community features are comparable in terms of diversity (Mahé et al. 2014) and curated Illumina sequence data sets were shown to have a better quality than those of 454 pyrosequencing (Kozich et al. 2013). Improvement of sequence quality and denoising was more important for 454 pyrosequencing as it is now for Illumina sequencing, where the sequencing error rate is significantly lower.

Based on the findings of **Manuscript I**, Illumina sequences of the following studies (**Manuscript II and III)** were processed with QIIME. Although the sequence processing with mothur or DNAStar/PhyloAssigner has resulted in a more realistic diversity estimate of diatoms, we used QIIME because it can better manage large sequencing datasets and sequence processing is faster. In doing so, some parameter settings for preprocessing of Illumina sequences were adjusted (e.g. number of homopolymers) and OTUs were constructed *de novo* at a minimum similarity threshold of 98% in QIIME. We used this similarity threshold because the analysis of diatoms in **Manuscript I** revealed a high species diversity which was better detected if a high threshold was applied (Kermarrec et al. 2014). After constructing OTUs in QIIME, protist sequences were classified based on their phylogenetic relationships by placing OTU representative sequences into a phylogenetic tree within PhyloAssigner. This procedure was most likely the best to estimate protist diversity and composition in the samples and demonstrates that different sequence processing methods and software tools can be combined.

In addition to the assessment of protist community structure by sequencing, large-scale biogeographic patterns of protist communities were analyzed in a high number of samples by ARISA. In **Manuscript II,** a total of 83 samples were analyzed with ARISA, which is a suitable method for an assessment of protist diversity patterns and potential correlation to environmental samples (Fechner et al. 2010, Bienhold et al. 2012, Wolf et al. 2013, Gobet et al. 2014, Kilias et al. 2015). Due to the large sample size, statistical correlations between protist communities and sampled regions or respective physico-chemical parameters of the habitats were obtained with high confidence. Because identification of species is not possible with ARISA (Bent et al. 2007, Caron et al. 2012), we additionally sequenced a representative subset of the samples to analyze the species composition and diversity. The combination of Illumina and ARISA (**Manuscript II and III)** was an appropriate strategy to investigate possible influences of environmental change on protists analyzed in a large number of samples and thus, to achieve **aim 2 and 3**.

8.2 Protist communities in the changing Arctic Ocean

In general, the Arctic summer season (May - September) is considered to be the most productive season for phytoplankton and sea ice associated algae (Arrigo et al. 2008, Nicolaus et al. 2012, Popova et al. 2012, Fernández-Méndez et al. 2015). Sea ice melt leads to an enhanced water stratification and higher light availability, which triggers photosynthetic production of sea ice algae and phytoplankton (Eilertsen 1993, Arrigo et al. 2008, Nicolaus et al. 2012). Because sea ice algae and pelagic protists are highly adapted in their life cycle and physiology to the presence and condition of sea ice (Horner 1985, Arrigo 2014), sea ice retreat is expected to have severe implications for protists in the CAO.

Changes in spatial extent and conditions of sea ice (e.g. ice thickness, brine drainage system, snow and melt pond coverage) are mainly the result of changing environmental conditions, such as water mass properties, air temperature and wind speed and direction (Comiso et al. 2008, Stroeve et al. 2008, 2012, 2014, Arrigo 2014). In particular, warm surface water temperatures and strong winds resulted in unusual low sea ice extent and concentration in the past years 2007, 2011 and 2012. Due to an early melt onset and late freeze-up in Arctic shelf regions (e.g. Kara and Laptev Sea), more heat is stored in the upper ocean, which lengthen the summer melt season (Stroeve et al. 2014), impacting pelagic and ice algal communities (Wassmann 2011, Arrigo et al. 2008, 2014, Leu et al. 2011). The

second aim of this thesis was to investigate the impact of sea ice retreat and sea ice origin on protist communities living in sea ice and water column of the CAO during summer. The **third aim** addressed the community structure and local exchange of protists in water and sea ice habitats, including under-ice and melt pond water.

8.2.1 Impact of sea ice changes on protist communities

Sea ice is one of the most important factors that influence Arctic protist communities in several ways (e.g. Eilertsen 1993, Fortier et al. 2002, Granskog et al. 2003, Arrigo et al. 2008, Comeau et al. 2011, Niemi et al. 2011, Arrigo 2014, Boetius et al. 2013, Fernández-Méndez et al. 2015). The presented **Manuscript II** and **III** and **Contribution I**, **II** and **III** demonstrate that the presence and conditions of sea ice can change the biogeochemistry of water and plankton community composition (**Contribution I, II, III**) as well as trophic interactions in sea ice and water column (**Contribution III**). In particular, during the sea ice minimum record in September 2012 (**Fig. 3** in General Introduction), strong regional patterns of pelagic and sea ice protists were observed in the Eurasian Basin (**Manuscript II**). These patterns were most likely the consequence of different environmental conditions because pelagic protists were significantly correlated with dissolved inorganic nutrients and particularly with sea ice concentration and water temperature. In 2012, large parts of the Amundsen Basin were ice free due to extensive melting. In contrast, the Nansen Basin was characterized by high sea ice concentrations as the ice drifted out of the CAO into the Fram Strait. Besides of lower sea ice concentrations, the DCM of the Amundsen Basin also showed lower salinities and higher silicate concentrations compared to the Nansen Basin. The pelagic community in the Amundsen Basin was characterized by high abundances of diatoms, whereas the abundances in the Nansen Basin were lower (**Manuscript II**). In addition, we observed higher abundances of protalveolates (mainly Syndiniales) in the Amundsen Basin. This suggests a higher heterotrophy level in this region (**Manuscript II, Contribution III)**, as syndiniales are mainly parasites of dinoflagellates, fish and crustacean. The reason for the different community compositions was most likely the better light conditions in the upper ocean due to ice melt, which led to an increase in diatom growth in the Amundsen Basin (Arrigo et al. 2008, Popova et al. 2012, Nicolaus et al. 2012). A similar pattern was also observed for the sea ice community, which was dominated by diatoms in the Amundsen Basin, again probably due to better light conditions present under thinner sea ice (**Manuscript II**). The influence of changing sea ice concentrations was also observed at higher trophic levels, for example the

under-ice fauna, such as copepods and amphipods (David et al. 2015**, Contribution III**). The under-ice fauna composition differed between the Amundsen and Nansen Basin with a dominance of pelagic amphipods at nearby ice-free stations (David et al. 2015).

In contrast to 2012, a regional distinction of protists communities was not observed during the same summer months in 2011, a year with higher sea ice concentrations and extent (Kilias et al. 2014a, **Manuscript II**). A comparison of sea ice and pelagic protists between the two summer periods with contrasting sea ice concentrations revealed strong differences in community diversity and composition. ITS1 fragment number and standard deviations calculated for ARISA and Illumina datasets revealed that the sea ice community of 2012 was less variable among the sites and less diverse than the sea ice community in 2011 (**Manuscript II**). In particular, the diversity of rare sea ice algae was considerably reduced. These observations were in good agreement with the results presented in **Manuscript III**. Here, a detailed analysis of protist communities in the CAO showed that the sea ice harbors a significant high number of unique species, which are most likely highly adapted to the unique environmental conditions of their habitat (Horner 1985, Arrigo 2014). Therefore, the abundance of sea ice algae could decrease over the course of climate change (Melnikov 2009), with rare species possibly more affected than abundant species.

8.2.2 Influence of sea ice origin on protist communities in the sea ice

In fall (end of September) and winter (October to end of February), new sea ice is formed when atmospheric and water temperatures decrease (reviewed by Arrigo 2014). During ice formation, bacteria, heterotrophic protists (Weissenberger and Grossmann 1998) and phytoplankton (Gradinger and Ikävalko 1998, Niemi et al. 2011) are incorporated into the sea ice by physico-biochemical processes (Ackley et al. 1987, Melnikov 1997, Gradinger and Ikävalko 1998). This "seeding" of the sea ice community was for example observed by Niemi et al. (2011), who found similar protist diversity in winter and spring sea ice of the Canadian Beaufort Sea and suggested that timing of sea ice formation may impact protist abundances in winter sea ice and the following spring ice algae bloom (Niemi et al. 2011). Because temporal changes of sea ice formation may come along with spatial changes, the regions of ice formation (i.e. sea ice origin) most likely influence protist communities, as well.

Manuscript II demonstrated that the sea ice community correlated with water masses and regions of sea ice origin, while the pelagic community was influenced by prevailing water

masses and sea ice concentrations in the sampling area. In addition to sea ice protists, sea ice nutrients were closely related to the sea ice origin, which indicates that surface waters most likely influenced sea ice nutrients and protists composition during ice formation and/or sea ice drift (Gradinger and Ikävalko 1998; Granskog et al. 2003; Melnikov et al. 2003). The influence of sea ice origin was only observed for the sea ice cores collected in 2011 but not for those collected in 2012. In 2011, the ice cores originated from both, the Eurasian and the Amerasian Basin, which allowed a strong discrimination of the environmental profiles. In contrast to this, sea ice cores of 2012 were collected only in the Eurasian Basin and originated from adjacent shelf regions of the Kara and Laptev Sea. Thus, environmental conditions during ice formation were probably less pronounced in 2012 and a possible influence of sea ice origin on sea ice protist community structure was not detectable.

In **Manuscript III**, we investigated how sea ice formation possibly changed the protist community structure, namely due to an increased exchange of protists between the water column, sea ice and melt ponds when new ice was formed in early autumn 2012.

8.2.3 Importance of sea ice for exchange of habitat-specific protists

While **Manuscript II** focused mainly on the biogeography of protists and analyzed how protist community patterns are impacted by environmental changes, **Manuscript III** focused on protist community structure and local exchange between various habitats of the CAO. In the course of **aim 3**, samples from the deep-chlorophyll maximum water depth (DCM), under-ice water (UIW), sea ice (ICE) and melt pond water (MW) were analyzed to determine protist exchange between the habitats, i.e. the number of shared OTUs assessed with Illumina sequencing. In addition, possible influences of the sea ice conditions (e.g. ice thickness, temperature, salinity) on the occurrences of protist taxa were investigated. The community was grouped in unique taxa (found in one habitat only), shared taxa (found in two or three habitats) and ubiquitous taxa (found in all four habitats).

In general, the sea ice community was highly site-variable and mainly inhabited by Diatomea, Chrysophyceae and Cercozoa, whereas the chlorophyll-maximum community was mainly characterized by Dinophyceae and Protalveolata (**Manuscript II and III**). This differentiation between pelagic and sea ice communities is well known (e.g. Booth and Horner 1997, Ardyna et al. 2011, Poulin et al. 2011, Comeau et al. 2011) but less is known about under-ice water (Arrigo et al. 2012, Laney et al. 2014) and melt ponds (Kilias et al. 2014b, Fernández-Méndez et al. 2015). The under-ice community was highly similar to the

community at the chlorophyll-maximum but was characterized by higher abundances of Mamiellophyceae and lower abundances of Protalveolata. The highly variable melt pond community was mainly characterized by Chrysophyceae, Bacillariophyceae and Ciliophora (**Manuscript III**).

The unique biosphere (i.e. unique OTUs) of the different habitats mostly consisted of the groups that were the dominant contributors of the respective habitats. However, the unique biosphere in under-ice and melt pond water mostly consisted of Ciliophora (e.g. *Strombidium, Oxytricha* and *Didinium*). These taxa were probably grazing on particulate matter (Michel et al. 2002, Kaartokallio et al. 2007), phytoplankton or ice algae (Amblard et al. 1993, Posch et al. 2015, Przytulska et al. 2016) and were thus, restricted to the ice-water interface or melt ponds. We observed the highest number and abundance of unique taxa in the sea ice cores, particularly at MYI stations, which were sampled at the beginning of autumn when atmospheric freezing conditions were experienced. Here, diatoms such as *Nitzschia*, *Melosira*, and *Cylindrotheca* were the main representatives. This observation emphasizes the importance of sea ice as a unique habitat and hotspot of protist diversity. Overall, the combination of the results obtained in **Manuscript II and III** suggest that sea ice algae could be endangered if sea ice is further retreating (Melnikov 2009), in particular if they are rare. Sea ice algae are very likely highly adapted to the unique physico-chemical conditions of sea ice (Horner 1985, Arrigo 2014) and hence, probably not able to react to habitat loss or fast environmental changes in their living environment.

Changes in sea ice conditions could also change the degree of protist exchange between the habitats. The results of **Manuscript II** suggested a higher exchange degree of pelagic and sea ice protists as consequence of sea ice melt, because higher abundances of the pelagic dinoflagellate *Gymnodinium* were found in the sea ice of 2012 compared to 2011. Indeed, analysis in **Manuscript III** revealed that the number of shared species between sea ice and under-ice water was highest at stations, which were sampled in early August and early September when the sea ice was melting. Furthermore, compared to 2011, we observed higher abundances of freshwater taxa, such as *Ochromonas* and *Epipyxis* (Chrysophyceae) in the sea ice of 2012. As these genera were mainly found in melt ponds (**Manuscript III**, Kilias et al. 2014b), their abundance in sea ice most likely evidenced the exchange of protists between the sea ice and melt pond habitats. The protist exchange between the habitats came along with an overall high number of ubiquitous taxa found in all habitats. Among the most dominant representatives were *Gymnodinium* (Dinophyceae) and *Ochromonas* (Chrysophyceae). Despite their appearance in all habitats, ubiquitous taxa showed clear habitat preference. This

demonstrates the adaptation of widely distributed taxa to the unique biotic and abiotic parameters of their habitat. A particularly high exchange was observed when new sea ice was formed in mid- and late-September. Accordingly, protist exchange between the habitats of the CAO was determined by sea ice conditions and thus, mainly triggered by physical changes in the living environment. Therefore, the findings of **Manuscript II and III** suggest that the diversity of sea ice algae, particularly diatoms, could decrease with decreasing sea ice thickness and age. The ongoing reduction of sea ice could result in a lower extent of protist exchange between the habitats as less new sea ice is formed. Finally, less protist species are incorporated into the sea ice, which has further consequences for the community structure of the following ice algae spring bloom (Niemi et al. 2011).

9 Outlook

This thesis demonstrated that the last summer record sea ice minimum extent observed in September 2012, probably lead to less diverse and complex communities in sea ice and water column and to a lower exchange of protists between the habitats. Specialized and rare taxa could be endangered if sea ice is further retreating as they are highly adapted to the living environment in their life cycle, ecology and physiology. Therefore, more large-scale studies are needed to identify specialized taxa in order to elucidate the impact of climate change on their abundance and diversity. Besides biogeographical studies, ecophysiological studies are required to assess the responses of Arctic specialists to environmental changes compared to Arctic generalists with a wide distribution.

The results gave evidence that particularly the conditions during sea ice formation in autumn and winter have severe implications for protist community structure and exchange. Due to climate change, melting of sea ice starts earlier but sea ice formation appears later (Perovich and Polashenski 2012, Stroeve et al. 2014), which is assumed to impact the abundance of the sea ice community and the following spring bloom (Niemi et al. 2011). Thus, the timing and region of sea ice formation influence protist assemblages in the sea ice and water column. To assess possible influences of temporal and/or spatial changes in sea ice formation, more studies conducted during autumn and winter are needed. In particular, more attention should be given to protist community structures and dynamics on Arctic shelf regions, especially the Chukchi and Laptev Sea, which harboured a protist community different from the CAO (**Manuscript II**). These regions are important in the face of climate change because a large part of sea ice is formed there. An immense decrease in summer sea ice concentration and delay in autumn freeze-up was already observed in these shelf regions (Stroeve et al. 2014). Since winter records began in 1979, the first record seasonal minimum in sea ice extent was reached on March 24, 2016 (NSIDC; http://nsidc.org), possibly leading to a next sea ice low during summer 2016. Hence, a further change in protist community composition and decrease of diversity could be possible in summer 2016.

This thesis further recommends that future studies investigating the impact of climate change on protist communities with Amplicon-based sequencing, should use standardized analyses procedures and sequence processing pipelines to guarantee a better comparability of studies. The results emphasized that different sequence processing methods should be compared and applied with caution. Conclusions of ecological questions drawn on the abundant biosphere (relative abundance $\geq$ 1%) and high taxonomic levels (e.g. phyla) are

comparable among studies but caution is advised if the conclusions are drawn on rare species (relative abundance < 1%). In addition, molecular approaches should be combined with conventional or other molecular methods to minimize bias and uncertainties of each method and to critically evaluate the outcome of each method.

10 References

Ackley SF, Dieckmann GS, Shen HT (1987) Algal and foram incorporation into new sea ice. *Transactions American Geophysical Union*, **68**, 1736–1748.

Admiraal W, Venekamp LAH (1986) Significance of tintinnid grazing during blooms of *Phaeocystis pouchetii* (Haptophyceae) in Dutch coastal waters. *Netherlands Journal of Sea Research*, **20**, 61–66.

Amato A, Kooistra WHCF, Levialdi Ghiron JH, Mann DG, Pröschold T, Montresor M (2007) Reproductive isolationamong sympatric cryptic species in marine diatoms. *Protist*, **158**, 193–207.

Amblard C, Simengando T, Rachiq S, Bourdier G (1993) Importance of ciliated protozoa in relation to the bacterial and phytoplanktonic biomass in an oligo–mesotrophic lake, during the spring diatom bloom. *Aquatic sciences*, **55**, 1–9

Ardyna M, Gosselin M, Michel C, Poulin M, Tremblay JE (2011) Environmental forcing of phytoplankton community structure and function in the Canadian High Arctic: contrasting oligotrophic and eutrophic regions. *Marine Ecology Progress Series*, **442**, 37–57. doi:10.3354/meps09378.

Arrigo KR, van Dijken G, Pabi S (2008) Impact of a shrinking Arctic ice cover on marine primary production. *Geophysical Research Letters*, **35**, 1–6.

Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE, Mills MM, Palmer MA, Balch WM, Bahr F, Bates NR, Benitez–Nelson C, Bowler B, Brownlee E, Ehn JK, Frey KE, Garley R, Laney SR, Lubelczyk L, Mathis J, Matsuoka A, Mitchell BG, Moore GWK, Ortega–Retuerta E, Pal S, Polashenski CM, Reynolds RA, Schieber B, Sosik HM, Stephens M, Swift JH (2012) Massive phytoplankton blooms under Arctic sea ice. *Science*, **336**, 1408. doi:10.1126/science.1215065.

Arrigo KR (2014). Sea ice ecosystems. *Annual review of marine science*, **6**, 439–467. doi:10.1146/annurev–marine–010213–135103.

Assmy P, Smetacek V, Montressor M, Klaas C, Henjes J, Strass VH, Arrieta JM, Bathmann U, Berg GM, Breitbarth E, Cisewski B, Friedrichs L, Fuchs N, Herndl, Jansen S, Krägefsky S, Latasa M, Peeken I, Röttgers R, Scharek R, Schüller SE, Steigenberger S, Webb A, Wolf-Gladrow D (2013) Thick–shelled, grazer–protected diatoms decouple ocean carbon and silicon cycles in the iron–limited Antarctic Circumpolar Current. *Proceedings of the National Academy of Sciences*, **110** (51), 20633–20638.

Bachy C, Dolan JR, López–García P, Deschamps P, Moreira D (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME Journal*, **7**, 244–255.

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. *Molecular Phylogenetic Evoutionl*, **1**, 3–16. doi:10.1016/1055–7903(92)90030–K.

Bathmann UV, Noji TT, von Bodungen B (1990) Copepod grazing potential in late winter in the Norwegian Sea – a factor in the control of spring phytoplankton growth? *Marine Ecology Progress Series*, **60**, 225–233.

Bauerfeind E, von Bodungen B, Arndt K, Koeve W (1994) Particle flux, and composition of sedimenting matter, in the Greenland Sea. *Journal of Marine Systems*, **5**, 411–423.

Bauerfeind E, Leipe T, Ramseier RO (2005) Sedimentation at the permanently ice–covered Greenland continental shelf (74 degrees 57.70N/12 degrees 58.70W): significance of biogenic and lithogenic particles in particulate matter flux. *Journal of Marine Systems*, **56**, 151–166.

Bauerfeind E, Nöthig EM, Beszczynska A, Fahl K, Kaleschke L, Kreker K, Klages M, Soltwedel T, Lorenzen C, Wegner J (2009) Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: Results from the Arctic long–term observatory HAUSGARTEN. *Deep–Sea Research I*, **56**, 1471–1487.

Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology letters*, **15**, 365–377.

Bent SJ, Peirson JD, Forney LJ (2007) Measuring species richness based on microbial community fingerprints: The emperor has no clothes. *Applied and environmental microbiology*, **73**, 2399–2401.

Beszczynska–Möller A, Fahrbach E, Schauer U, Hansen E (2012) Variability in Atlantic Water temperature and transport at the entrance to the Arctic Ocean, 1997–2010. *ICES journal of marine science : journal du conseil*, **69**, 852–863.

Bienhold C, Boetius A, Ramette A (2012) The energy–diversity relationship of complex bacterial communities in Arctic deep–sea sediments. *The ISME Journal*, **6**, 724–732.

Boetius A, Albrecht S, Bakker K, Bienhold C, Felden J, Fernández–Méndez M, Hendricks S, Katlein C, Lalande C, Krumpen T, Nicolaus M, Peeken I, Rabe B, Rogacheva A, Rybakova E, Somavilla R, Wenzhöfer F, RV Polarstern ARK27–3 Shipboard Science Party (2013) Export of algal biomass from the melting Arctic sea ice. *Science*, **339**, 1430–1432. doi:10.1126/science.1231346.

Bonder MJ, Abeln S, Zaura E, Brandt BW (2012) Comparing clustering and pre–processing in taxonomy analysis. *Bioinformatics*, **28**, 2891–2897.

Booth BC & Horner RA (1997) Microalgae on the Arctic Ocean section, 1994: Species abundance and biomass. *Deep Sea Research Part II*, **44**, 1607–1622.

Bramlette MN (1958) Significance of coccolithophorids in calcium–carbonate deposition. *Bulletin of the Geological Society of America*, **69**, 121–126.

Brown CW, Yoder JA (1994) Coccolithophorid Blooms in the Global Ocean. *Journal of Geophysical Research Oceans*, **99**, 7467–7482.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez Pena A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high–throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Carmack E, Wassmann P (2006) Food webs and physical–biological coupling on pan–Arctic shelves: unifying concepts and comprehensive perspectives. *Progress in Oceanography*, **71**, 446–477.

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012) Marine protistan diversity. *Annual Review of Marine Science*, **4**, 467–493.

Chen W, Zhang CK, Cheng Y, Zhang S, Zhao H (2013) A Comparison of Methods for Clustering 16S rRNA Sequences into OTUs. *PLoS ONE*, **8** (8): e70837.

Comeau AM, Li WKW, Tremblay JAR, Carmack EC, Lovejoy C (2011) Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. *PLoS ONE*, **6**, e27492. doi:10.1371/journal.pone.0027492.

Comiso JC (2003) Warming trends in the Arctic from clear sky satellite observations. *Journal of climate*, **16**, 3498–3510.

Comiso JC, Parkinson CL, Gersten R, Stock L (2008) Accelerated decline in the Arctic sea ice cover. *Geophysical Research Letters*, **35**, L01703. doi:10.1029/2007GL031972.

Cooper LW, McClelland JW, Holmes RM, Raymond PA, Gibson JJ, Guay CK, Peterson BJ (2008) Flow–weighted values of runoff tracers (d18O, DOC, Ba, alkalinity) from the six largest Arctic rivers. *Geophysical Research Letters*, **35**, L18606. doi:10.1029/2008GL035007.

David C, Lange B, Rabe B, Flores H (2015) Community structure of under–ice fauna in the Eurasian central Arctic Ocean in relation to environmental properties of sea–ice habitats. *Marine Ecology Progress Series*, **522**, 15–32.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu, P, Andersen GL (2006) Greengenes, a chimera checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, **72**, 5069–5072.

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B (2013) 454–pyrosequencing to Describe Microbial Eukaryotic Community Composition, Diversity and Relative Abundance: A Test for Marine Haptophytes. *PLoS ONE*, **8** (9), e74371.

Eilertsen HC (1993) Spring blooms and stratification. *Nature*, **363**, 24.

Fechner LC, Vincent-Hubert F Gaubert P, Bouchez T, Gourlay-France C, Tusseau-Vuillemin MH (2010) Combined eukaryotic and bacterial community fingerprinting of natural freshwater biofilms using automated ribosomal intergenic spacer analysis. *EMS microbiology ecology*, **74**, 542–553.

Fernández-Méndez M, Rabe B, Katlein C, Nicolaus M, Peeken I, Bakker K, Flores H, Boetius, A (2015) Photosynthetic production in the Central Arctic during the record sea-ice minimum in 2012. *Biogeoscience Discussions*, **12**, 2897–2945. doi:10.5194/bgd–12–2897–2015

Fisher MM, Triplett E (1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied Environmental Microbiology*, **65**, 4630–4636.

Fortier M, Fortier L, Michel C, Legendre L (2002) Climatic and biological forcing of the vertical flux of biogenic particles under seasonal Arctic sea ice. *Marine Ecology Progress Serie*, **225**, 1–16.

Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N (2008) Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. *Environmental Microbiology*, **10** (9), 2433–2443.

Gaspar JM, Thomas WK (2013) Assessing the consequences of denoising marker–based metagenomic data. *PLoS ONE*, **8**, e60458.

Gobet A, Boetius A, Ramette A (2014) Ecological coherence of diversity patterns derived from classical fingerprinting and Next Generation Sequencing techniques. *Environmental microbiology*, **16**, 2672–2681.

Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar A (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real–time PCR. *Applied and environmental microbiology*, **74** (23), 7174–7182.

Golden KM, Ackley SF, Lytle VI (1998) The percolation phase transition in sea ice. *Science*, **282**, 2238–2241.

Gosselin M, Levasseur M, Wheeler PA, Horner RA, Booth BC (1997) New measurements of phytoplankton and ice algal production in the Arctic Ocean. *Deep–Sea Research*, **44**, 1623–1644.

Gradinger R, Ikävalko J (1998) Organism incorporation into newly forming Arctic sea ice in the Greenland Sea. *Journal of Plankton Research*, **20**, 871–886.

Gradinger R (2009) Sea ice algae: Major contributors to primary production and algal biomass in the Chukchi and Beaufort Sea during May/June 2002. *Deep Sea Res. Part II*, **56**, 1201–1212. doi:10.1016/j.dsr2.2008.10.016.

Granskog MA, Kaartokallio H, Shirasawa K (2003) Nutrient status of Baltic Sea ice: evidence for control by snow–ice formation, ice permeability, and ice algae. *Journal of geophysical research*, **108**, 3253. 10.1029/2002JC001386.

Gustafson Jr DE, Stoecker D K, Johnson MD, Van Heukelem WF, Sneider K (2000) Cryptophyte algae are robbed of their organelles by the marine ciliate *Mesodinium rubrum*. *Nature*, **405**, 1049–1052.

Halvorsen MH, Smedsrud LH, Zhang R, & Kloster K (2015). Fram Strait spring ice export and September Arctic sea ice. *The Cryosphere Discussions*, **9**, 4205–4235.

Hartmann DL, Klein Tank AMG, Rusticucci M, Alexander LV, Brönnimann S, Charabi Y, Dentener FJ, Dlugokencky EJ, Easterling DR, Kaplan A, Soden BJ, Thorne BW, Wild M, Zhai PM (2013) Observations: Atmosphere and SurfaceIn: Climate Change 2013: The Physical Science BasisContribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.)]Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hegseth EN, Sundfjord A (2008) Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. *Journal of Marine Systems*, **74**, 108–119.

Hilligsøe KM, Richardson K, Bendtsen J, Sørensen LL, Nielsen TG, Lyngsgaard MM (2011) Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux. *Deep–Sea Research I*, **58**, 826–838.

Horner R (1985) Ecology of sea ice microalgae. *In*: Horner, R. (ed) Sea ice biota. CRC, Fla. p. 83–103.

Hoppenrath M, Beszteri B, Drebes G, Halliger H, Beusekom JEE, Janisch S, Wiltshire K (2007) Thalassiosira species (Bacillariophyceae, Thalassiosirales) in the North Sea at Helgoland (German Bight) and Sylt (North Frisian Wadden Sea) a first approach to assessing diversity. *European Journal of Phycology*, **42**, 271–288. doi: 10.1080/09670260701352288.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, 143–152.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12** (7), 1889–1898.

Iida T, Mizobata K, Saitoh SI (2012) Interannual variability of coccolithophore *Emiliania huxleyi* blooms in response to changes of water column stability in the eastern Bering Sea. *Continental Shelf Research*, **34**, 7–17.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945

Jensen F, Hansen WB (2000) Ciliates and heterotrophic dinoflagellates in the marginal ice zone of the central Barents Sea during spring. *Journal of the Marine Biological Association*, **80**, 45–54.

Kaartokallio H, Kuosa H, Thomas DN, Granskog MA, Kivi K (2007) Biomass, composition and activity of organism assemblages along a salinity gradient in sea ice subjected to river discharge in the Baltic Sea. *Polar Biology*, **30**, 183–197.

Kermarrec L, Bouchez A, Rimet F, Humbert JF (2013) First evidence of the existence of semi–cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kützing) Kützing complex (Bacillariophyta). *Protist*, **164** (5), 686–705.

Kermarrec L, Franc A, Rimet F, Chaumeil P, Frigero JM, Humbert JF, Bouchez A (2014) A next–generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, **33**, 349–363.

Kilias E, Kattner G, Wolf C, Frickenhaus S, Metfies K (2014a) A molecular survey of protist diversity through the central Arctic Ocean. *Polar Biology*, **37**, 1271–1287. doi:10.1007/s00300–014–1519–5.

Kilias ES, Peeken I, and Metfies K (2014b) Insight into protist diversity in Arctic sea ice and meltpond aggregate obtained by pyrosequencing. *Polar Research*, **33**. doi:http://dx.doi.org/10.3402/polar.v33.23466.

Kilias E, Wolf C, Metfies K (2015) Characterizing variability in marine protist communities via ARISA fingerprints – a method evaluation. *ASLO limnology and oceanography, methods*, **13**, 74–80. doi:10.1002/lom3.10008.

Kirtman B, Power SB, Adedoyin JA., Boer GJ, Bojariu R, Camilloni I, Doblas-Reyes FJ, Fiore AM, Kimoto M, Meehl GA, Prather M, Sarr A, Schär C, Sutton R, van Oldenborgh GJ, Vecchi G, Wang HJ (2013) Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J., Nauels A, Xia Y, Bex V, Midgley PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual–index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, **79**, 5112–5120.

Krembs C, Eicken H, Deming JW (2011) Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3653–3658.

Krumpen T, Gerdes R, Haas C, Hendricks S, Herber A, Selyuzhenok V, Smedsrud L, Spreen, G (2015) Recent summer sea ice thickness surveys in the Fram Strait and associated volume fluxes. *The Cryosphere Discussions*, **9**, 5171–5202.

Kwok R, Rothrock DA (2009) Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. *Geophysical Research Letters*, **36**, L15501. doi:10.1029/2009GL039035.

Laney SR, Sosik HM (2014) Phytoplankton assemblage structure in and around a massive under–ice bloom in the Chukchi Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, **105**, 30–41.

Laxon SW, Giles KA, Ridout AL, Wingham DJ, Willatt R, Cullen R, Kwok R, Schweiger A, Zhang J, Haas C, Hendricks S, Krish R, Kurtz N, Farrell S, Davidson M (2013) CryoSat–2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, **40**, 1–6. doi:10.1002/GRL.50193.

Lee KH, Jeong HJ, Jang TY, Lim AS, Kang NS, Kim JH, Park KT, Lee K (2014) Feeding by the newly described mixotrophic dinoflagellate *Gymnodinium smaydae*: Feeding mechanism, prey species, and effect of prey concentration. *Journal of Experimental Marine Biology and Ecology*, **459**, 114–125.

Leu E, Soreide JED, Hessen O, Falk–Petersen S, Berge J (2011) Consequences of changing sea–ice cover for primary and secondary producers in the European Arctic shelf seas: Timing, quantity, and quality. *Progress in oceanography*, **90,** 18–32.

Li WKW, McLaughlin, FA, Lovejoy C Carmack EC (2009) Smallest algae thrive as the Arctic Ocean freshens. *Science*, **326**, 539.

Lovejoy C, Massana R, Pedrós–Alió C (2006) Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Applied and environmental microbiology*, **72**, 3085–3095.

Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, Potvin M, Massana R, Pedros–Alio C (2007) Distribution, phylogeny, and growth of cold–adapted picoprasinophytes in arctic seas. *Journal of Phycology*, **43**, 78–89. doi:10.1111/j.1529–8817.2006.00310.

Luddington IA, Kaczmarska I, Lovejoy C (2012) Distance and character based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PLoS ONE*, **7**, e45664.

Mahé F, Mayor J, Bunge J, Chi J, Siemensmeyer T, Stoeck T, Wahl B, Paprotka T, Filker S Dunthorn M (2014) Comparing high–throughput platforms for sequencing the V4 region of SSU–rDNA in environmental microbial eukaryotic diversity surveys. *The Journal of eukaryotic microbiology*, **62**, 338–345.

Majaneva M, Hyytiäinen K, Varvio SL, Nagai S, Blomster J (2015) Bioinformatic Amplicon Read Processing Strategies Strongly Affect Eukaryotic Diversity and the Taxonomic Composition of Communities. *PLoS ONE*, **10** (6), e0130035. doi:10.1371/journal.pone.0130035.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Baverman MS, Chen YJ, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S,

Ho CE, Irzyk GP, Jando SC, Alenquer MLI et al. (2005) Genome sequencing in microfabricated highdensity picolitre reactors. *Nature*, **437**, 376–380.

Matantseva OV, Skarlato SO (2013) Mixotrophy in microorganisms: Ecological and cytophysiological aspects. *Journal of Evolutionary Biochemistry and Physiology*, **49** (4), 377–388.

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum–likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, **11**, 538.

Maslanik J, Stroeve J, Fowler C, Emery W (2011) Distribution and trends in Arctic sea ice age through spring 2011. *Geophysical Research Letters*, **38**, L13502. doi:10.1029/2011GL047735.

Masson–Delmotte V, Schulz M, Abe–Ouchi A, Beer J, Ganopolski A, González Rouco JF, Jansen E, Lambeck K, Luterbacher J, Naish T, Osborn T, Otto–Bliesner B, Quinn T, Ramesh R, Rojas M, Shao X Timmermann A (2013) Information from Paleoclimate ArchivesIn: Climate Change 2013: The Physical Science BasisContribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley M (eds.)].Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Melnikov IA (1997) The Arctic Sea Ice Ecosystem. Gordon and Breach Science Publishers, Amsterdam, 204 pp.

Melnikov IA, Korneeva GA, Zhitina LS, Shanin SS (2003) Dynamics of ecological–biochemical characteristics of sea ice in coastal waters of the White Sea. *The Biological bulletin*, **30**, 164–170.

Melnikov IA (2009) Recent sea ice ecosytem in the Arctic Ocean: a review, p. 57–72. In J. C. J. Nihoul and K. A. G. [eds.], Influence of Climate Change on the Changing Arctic and Sub–Arctic Conditions. Springer.

Metfies K, von Appen W–J, Kilias E, Nicolaus A, Nöthig E–M (2016) Biogeography and Photosynthetic Biomass of Arctic Marine Pico– Eukaroytes during Summer of the Record Sea Ice Minimum 2012. *PLoS ONE*, **11** (2), e0148512. doi:10.1371/journal.pone.0148512.

Michel C, Gissel Nielsen T, Nozais C, Gosselin M (2002) Significance of sedimentation and grazing by ice micro– and meiofauna for carbon cycling in annual sea ice (northern Baffin Bay). *Aquatic microbial ecology*, **30**, 57–68.

Mitra A, Flynn KJ, Tillmann U, Raven JA, Caron D, Stoecker DK, Not F, Hansen PK, Hallegraeff G, Sanders R, Wilken S, McManus G, Johnson M, Pitta P, Vage S, Berge T, Calbet A, Thingstad F, Jeong HJ, Burkholder J, et al. (2016) Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist*, **167** (2), 106–120.

Moran XAG, Lopez–Urrutia A, Calvo–Diaz A, Li WKW (2010) Increasing importance of small phytoplankton in a warmer ocean. *Global change biology*, **16**, 1137–1144.

Nicolaus M, Katlein C, Maslanik J, Hendricks S (2012) Changes in Arctic sea ice result in increasing light transmittance and absorption. *Geophysical Research Letters*, **39**, L24501. doi:10.1029/2012GL053738.

Niemi A, Michel C, Hille K, Poulin M (2011) Protist assemblages in winter sea ice: Setting the stage for the spring ice algal bloom. *Polar Biology*, **34**, 1803–1817.

Overland JE, Wang M (2013) When will the summer Arctic be nearly sea ice free? *Geophysical Research Letters*, **40** (10), 2097–2101.

Passow U (1991) Species–specific sedimentation and sinking velocities of diatoms. *Marine Biology*, **108**, 449–455. http://dx.doi.org/10.1007/BF01313655.

Peinert R, Antia A, Bauerfeind E, von Bodungen B, Haupt O, Krumbholz M, Ramseier RO, Voss M, Zeitzschel B (2001) Particle flux variability in the polar and Atlantic biogeochemical provinces of the Nordic Seas. In: Schäfer P, Ritzrau W, Schlüter M, Thiede J. (Eds.), The northern North Atlantic: A changing environment. Springer, Berlin, pp. 45–66.

Perovich DK, Polashenski C (2012) Albedo evolution of seasonal Arctic sea ice. *Geophysical Research Letters*, **39**, L08501. doi:10.1029/2012GL051432.

Perrette M, Yool A, Quartly GD, Popova EE (2011) Near–ubiquity of ice–edge blooms in the Arctic. *Biogeosciences*, **8**, 515–524.

Pfandl K, Chatzinotas A, Dyal P, Boenigk J (2009) SSU rRNA gene variation resolves population heterogeneity and ecophysiological differentiation within a morphospecies (Stramenopiles, Chrysophyceae). *Limnology and Oceanography*, **54** (1), 171–181.

Polyakov IV, Alekseev GV, Timokhov LA, Bhatt US, Colony RL, Simmons HL, Walsh D, Walsh JE, Zakharov VF (2005) One more step toward a warmer Arctic. *Geophysical Research Letters*, **32** (17), L17605.

Popova EE, Yool A, Coward AC, Dupont F, Deal C, Elliott S, Hunke E, Jin MB, Steele M, Zhang JL (2012) What controls primary production in the Arctic Ocean? Results from an

intercomparison of five general circulation models with biogeochemistry. *Journal of geophysical research*, **117**, C00D12. doi:10.1029/2011JC007112.

Posch T, Eugster B, Pomati F, Pernthaler J, Pitsch G and Eckert EM (2015) Network of Interactions Between Ciliates and Phytoplankton During Spring. *Frontiers in Microbiology*, **6**, 1289. doi: 10.3389/fmicb.2015.01289.

Post E, Bhatt US, Bitz CM, Brodie JF, Fulton TL, Hebblewhite M, Kerby J, Kutz SJ, Stirling I, Walker DA (2013) Ecological consequences of sea–ice decline. *Science*, **341** (6145), 519–524.

Poulíčková A, Veselá J, Neustupa J, Skaloud P (2010) Pseudocryptic diversity versus cosmopolitanism in diatoms: a casestudy on *Navicula cryptocephala* Kütz. (Bacillariophyceae) andmorphologically similar taxa. *Protist*, **161**, 353–369.

Poulin M, Daugbjerg N, Gradinger R, Ilyash L, Ratkova T, von Quillfeldt C (2011) The pan Arctic biodiversity of marine pelagic and sea–ice unicellular eukaryotes: A first–attempt assessment. *Marine Biodiversity*, **41**, 13–28.

Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, **46** (1), 48–50.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, **35**, 7188–7196.

Przytulska A, Comte J, Crevecoeur S, Lovejoy C, Laurion I, Vincent WF (2016) Phototrophic pigment diversity and picophytoplankton in permafrost thaw lakes, *Biogeosciences*, **13**, 13–26. doi:10.5194/bg–13–13–2016.

Quillfeldt CH von (2000) Common diatom species in Arctic spring blooms: their distribution and abundance. *Botanica Marina*, **43**, 499–516.

Quillfeldt CH von (2001) Identification of some easily confused common diatom species in Arctic spring blooms. *Botanica Marina*, **44** (4), 375–389.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.

Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank–abundance distributions. *Nature Methods*, **7**, 668–669.

Reigstad M, Wassmann P (2007) Does Phaeocystis spp. contribute significantly to export of organic carbon? *Biogeochemisty*, **83**, 217–234. DOI 10.1007/s10533–007–9093–3.

Rhein M, Rintoul SR, Aoki S, Campos E, Chambers D, Feely RA, Gulev S, Johnson GC, Josey SA, Kostianoy A, Mauritzen C, Roemmich D, Talley LD, Wang F (2013)

Observations: OceanIn: Climate Change 2013: The Physical Science Basis Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley M (eds.)]Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Rousseau V, Becquevort S, Parent JY, Gasparini S, Daro MH, Tackx M, Lancelot C (2000) Trophic efficiency of the planktonic food web in a coastal ecosystem dominated by Phaeocystis colonies. *Journal of Sea Research*, **43** (3), 357–372.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain–terminating inhibitors. *Proceedings of the National Academy of Sciences*, **74**, 5463–5467.

Sakshaug E (1989) The physiological ecology of polar phytoplankton. In: Rey L, Alexander V. (Eds.), Proceedings 01 the 6th Conference of the Comite Arctique International. Leiden, pp. 61–89.

Schauer U, Fahrbach E, Osterhus S, Rohardt G (2004) Arctic warming through the Fram Strait: Oceanic heat transport from 3 years of measurements. *Journal of geophysical research*, **109**, C06026. doi:10.1029/ 2003JC001823.

Schauer U, Beszczynska–Möller A, Walczowski W, Fahrbach E, Piechura J, Hansen E (2008) Variation of measured heat flow through the Fram Strait between 1997 and 2006. In: Dickson, R.R, Meincke, J, Rhines, P. (Eds.), Arctic–Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate. Springer Science and Business Media B.V, Dordrecht, pp. 65–85.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJ van, Weber CF (2009) Introducing mothur: open–source, platform–independent, community–supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, **75**, 7537–7741.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA–based studies. *PLoS One*, **6**, e27310.

Schloss PD (2013) Secondary structure improves OTU assignments of 16S rRNA gene sequences. *The ISME Journal*, **7**, 457–460.

Serreze MC, Barrett AP, Slater AG, Woodgate RA, Aagaard K, Lammers RB, Steele M, Moritz R, Meredith M, Lee CM (2006) The large-scale freshwater cycle of the Arctic. *Journal of Geophysical Research*, **111**, C11010. doi:10.1029/2005JC003424.

Smedsrud LH, Sirevaag A, Kloster K, Sorteberg A, Sandven S (2011) Recent wind driven high sea ice area export in the Fram Strait contributes to Arctic sea ice decline. *The Cryosphere*, **5**, 821–829.

Smith ND, Turner AH (2005) Morphology's role in phylogeny reconstruction: perspectives from Paleontology. *Systematic Biology*, **54**, 166–173.

Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, Wang X, Mai V (2012) A large–scale benchmark study of existing algorithms for taxonomy–independent microbial community analysis. *Briefings in bioinformatics*, **13** (1), 107–121

Stroeve J, Holland MM, Meier W, Scambos T, Serreze M (2007) Arctic sea ice decline: faster than forecast. *Geophysical Research Letters*, **34**. doi:10.1029/2007GL029703.

Stroeve J, Serreze M, Drobot S, Gearheard S, Holland M, Maslanik J, Meier W, Scambos T (2008) Arctic sea ice extent plummets in 2007. *EOS Trans, AGU*, **89** (2), 13–14.

Stroeve JC, Serrez, MC, Holland MM, Kay JE, Masklanik J, Barrett AP (2012) The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climate Change*, **110**, 1005–1027.

Stroeve JC, Markus T, Boisvert L, Miller J, Barrett A (2014) Changes in Arctic melt season and implications for sea ice loss. *Geophysical Research Letters*, **41**, 1216–1225. doi:10.1002/2013GL058951.

Wassmann P, Slagstad D (1993) Seasonal and annual dynamics of particulate carbon flux in the Barents Sea. A model approach. *Polar Biology*, **13**, 363–372.

Wassmann P, Reigstad M (2011) Future Arctic Ocean seasonal ice zones and implications for pelagic–benthic coupling. *Oceanography*, **24**, 220–231.

Weissenberger J, Grossmann S (1998) Experimental formation of sea ice: importance of water circulation and wave action for incorporation of phytoplankton and bacteria. *Polar Biology*, **20**, 178–188.

Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference–based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*, **3**, e1487.

Wiens JJ (2004) The role of morphological data in phylogeny reconstruction. *Systematic Biology*, **53**, 653–661.

Wolf C, Frickenhaus S, Kilias ES, Peeken I, Metfies K (2013) Regional variability in eukaryotic protist communities in the Amundsen Sea. *Antarctic Science*, **25**, 741–751.

Wolf C, Iversen M, Klaas C, Metfies K (2016) Limited sinking of *Phaeocystis* during a 12 days sediment trap study. *Molecular Ecology*, in press.

Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, von Dassow P et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, **324** (5924), 268–272.

Verity PG (1985) Grazing, respiration, excretion, and growth rates of tintinnids. *Limnology and Oceanography*, **30** (6), 1268–1282.

Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, Kilpert D, Worden AZ, Giovannoni SJ (2013) High–resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time–series Study site by phylogenetic placement of pyrosequences. *The ISME Journal*, **7**, 1322–1332.

Yamamoto–Kawai M, McLaughlin FA, Carmack EC (2011) Effects of ocean acidification, warming and melting of sea ice on aragonite saturation of the Canada Basin surface water. *Geophysical Research Letters*, **38**, L03601, doi:10.1029/2010GL045501.

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms, diversity & evolution*, **11**, 173–192.

11 Acknowledgment

First of all, I want to thank Dr. Katja Metfies for providing this interesting project and her supervision and guidance throughout my study. I want to express my deepest gratitude for all the pathbreaking and motivating words during tough and good times. Thank you for the big-hearted advice!

I am thankful to Dr. Ilka Peeken for the numerous, extensive and great discussions as well as motivating words during my thesis. A gratitude goes also to Prof. Frickenhaus for his great support and open ear for statistical questions.

I would like to express my sincere gratitude to my other PhD committee members: Prof. Dr. Matthias Ullrich, Dr. Mona Hoppenrath and Dr. Eva-Maria Nöthig for the fruitful discussions and warm words of support.

A special thanks goes to the Helmholtz Graduate School Polar and Marine Research (POLMAR) for the manifold and interesting courses. Thank you for having the door always open. In addition, I would like to thank the Initiative and Networking Fund of the Helmholtz Association for financial support.

My very sincere thanks to my colleagues and friends Johanna Hessel, Estelle Kilias, Angelina Kraft, Stefan Neuhaus, Kerstin Oetjen, Swantje Rogge, Pim Sprong, Henrieke Tonkes and, in particular Christian Wolf. Thank you for the productive discussions, constructive comments and laughs on my manuscripts. Thank you for always encouraging me, giving me big hugs and being around whenever I needed you. You made my day (especially the ‘mad Mondays’ and ‘freaky Fridays’).

I want to thank my beloved ‘Berliner Schnauzen’: Diana Bednarek, Chris Munke, Bianca Sandmann and Astrid Schories for always believing in me and for hours of profound, consoling talks. Thank you for the lovely way to bring more smile and color into my day.

A special thanks goes to my family ‘on the duck pond’ for always receiving me warmly and giving ear. I am especially grateful to Thomas Jendrossek and his “better half” for making me laugh so many times, for worrying about me and my vitamin balance and for banishing ghosts at day and night.

Finally yet importantly, I want to express my heartfelt thanks to my beloved grandparents in the “Sommerhaus”, my parents, Klaus and Cornelia, and my brother, Daniel, for their unconditional love and always believing in me. Thank you for the endless support and encouragement. Thank you for making me strong and always catching me when I fell.

Thank you!