Protist diversity and biogeography in the Pacific sector of the
Southern Ocean

by

Christian Wolf

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

Doctor of Philosophy
in Biology

Approved Dissertation Committee:
Dr. Katja Metfies
Jacobs University Bremen
Alfred Wegener Institute for Polar and Marine Research
Prof. Dr. Matthias Ullrich
Jacobs University Bremen
Dr. Gunnar Gerdts
Alfred Wegener Institute for Polar and Marine Research
Prof. Dr. Jelle Bijma
Jacobs University Bremen
Alfred Wegener Institute for Polar and Marine Research

Date of Defense: December 13, 2012

School of Engineering and Science
“Anyone who has never made a mistake has never tried anything new”
Albert Einstein
# Table of contents

Acknowledgements                                                                 5  
1. Summary                                                                      6  
2. General introduction                                                        8  
  2.1 The Southern Ocean and climate change                                      8  
  2.2 Marine eukaryotic protists in the Southern Ocean                          13  
  2.3 Molecular methods                                                         16  
    2.3.1 Automated ribosomal intergenic spacer analysis (ARISA)                17  
    2.3.2 Sequencing of 18S rRNA gene clone libraries                         19  
    2.3.3 Next-generation sequencing: 454-pyrosequencing                       21  
3. Aims and outline of the dissertation                                           23  
  3.1 Aims                                                                       23  
  3.2 Outline                                                                   23  
4. Manuscripts                                                                25  
  4.1 List of manuscripts                                                     25  
  4.2 Statement of contribution to the manuscripts                        26  
  4.3 Manuscript I: Investigating eukaryotic protist diversity – a comparison of clone library and 454-pyrosequencing data 27  
  4.4 Manuscript II: Oceanographic fronts in the Southern Ocean determine biogeographic differences in eukaryotic protist communities – new insights based on 454-pyrosequencing 53  
  4.5 Manuscript III: Regional variability in eukaryotic protist communities in the Amundsen Sea 92  
5. Synthesis and future perspectives                                             125  
  5.1 Investigation of eukaryotic protist diversity with molecular approaches 125  
  5.2 Protist assemblages in the Pacific sector of the Southern Ocean          127  
  5.3 Future perspectives                                                      129  
6. References                                                                  131  
7. Statutory Declaration                                                        138
Acknowledgements

First, I want to thank Dr. Katja Metfies for her supervision and scientific advice. Thanks for having your door always open to discuss and to resolve problems. I am grateful for the opportunity to get my PhD in the warm atmosphere of your working group.

I also would like to thank my other PhD committee members, namely Prof. Dr. Matthias Ullrich, Dr. Gunnar Gerdts and Prof. Dr. Jelle Bijma for examining my thesis and for their scientific advice.

I would like to thank the Initiative and Networking Fund of the Helmholtz Association for funding. I also want to thank the Alfred Wegener Institute and the Helmholtz Graduate School for Polar and Marine Research (POLMAR) for scientific and financial support.

Special thanks go to Estelle Kilias for going the road to the PhD with me. It was always good to have someone in the same boat. Thank you!

I want to thank Stephan Frickenhaus, Ilka Peeken and Mirko Lunau for their contributions to the manuscripts and for their scientific advice.

I am grateful to Katharina Kohls, Eva-Maria Nöthig and Steffi Gäbler-Schwarz for their time, advice and for the fruitful discussions (in particular the non-scientific ones).

I also want to thank Anja Nicolaus, Kerstin Oetjen and Annika Schröer for technical support in the laboratory and Fabian Kilpert for bioinformatical support.

I am also grateful to all my friends, who were always present when I needed them, especially Jennifer and “Ella”, who gave me the strength to go ahead.

Finally yet importantly, I want to thank my family and especially my parents, Ritva und Gerhard Wolf, who made it possible to find my way. It was not always easy, but you always believed in me. Thank you!
1. Summary

The objectives of this thesis were the establishment of molecular approaches in the diversity investigation of eukaryotic protists in the Southern Ocean, the comparison of different approaches and the delivery of a comprehensive and taxon detailed overview of protist assemblages in the Pacific sector of the Southern Ocean, especially in the Amundsen Sea. The molecular approaches used to achieve these goals were automated ribosomal intergenic spacer analysis (ARISA), sequencing of 18S rRNA gene clone libraries and 454-pyrosequencing.

The comparison of 18S rRNA clone library sequences with the results of 454-pyrosequencing was conducted with four Arctic water samples, focusing on picoplankton (0.4-3 µm), and with one Antarctic water sample, covering the whole size spectrum (>0.2 µm). It turned out that the two methods delivered different results. Both approaches discovered phylotypes that were not found with the other approach. The abundant biosphere, defined by the 454-pyrosequencing approach, was not fully recovered by the clone library approach. The cloning approach was biased against several groups, e.g. haptophytes in the Arctic samples and diatoms in the Antarctic sample. In summary, prior cloning data have to be handled with care, when compared with 454-pyrosequencing data. Additionally, cloning data are only of limited suitability to serve as a backbone for phylogenetic analysis of 454-pyrosequencing data.

The results of this comparison led to the decision to use ARISA and 454-pyrosequencing for the further protist diversity investigations. First, the hypothesis that distinct protist community assemblages characterize large-scale water masses was tested. The composition and biogeography of late summer eukaryotic protist assemblages along a transect from the coast of New Zealand to the eastern Ross Sea was determined. ARISA and 454-pyrosequencing were used in combination with flow cytometry and pigment measurements via high performance liquid chromatography (HPLC) to study the protist assemblage. Distinct biogeographic patterns defined by the different oceanic regions were revealed. Different water masses harboured different microbial communities, and environmental gradients limited their dispersal. Picoeukaryotes were of minor importance throughout the investigated transect and were nearly absent south of the Polar Front. Dinoflagellates, Syndiniales, and small stramenopiles dominated the Subantarctic Zone, whereas the importance of diatoms increased southwards, in the Polar Frontal Zone, the Antarctic Zone and the Subpolar Region. South of the Polar Front, haptophytes were the dominating group.
Second, the investigation focused on the Amundsen Sea to see if the protist community assemblages vary in different areas of a single large-scale water mass. The composition and structure of late summer eukaryotic protist assemblages along a west-east transect in the Amundsen Sea were analysed. ARISA and 454-pyrosequencing were combined with HPLC. Characteristic communities offshore and inshore were revealed, but the differences were weaker, compared to those found along the north-south transect. In general, total chlorophyll $a$ and microeukaryotic contribution were higher in inshore samples. Picoeukaryotes were also of minor importance. Diatoms were the dominating group across the entire area, at which *Eucampia* sp. and *Pseudo-nitzschia* sp. were dominating inshore and *Chaetoceros* sp. was dominating offshore. At the eastern most station, the assemblage was dominated by *Phaeocystis* sp. Under the ice, ciliates showed their highest and haptophytes their lowest abundance.

This thesis sheds light on the use and applicability of several molecular methods for the investigation of protist assemblages in polar waters. It delivers a comprehensive and taxon detailed overview of the eukaryotic protist composition during the austral summer in the Pacific sector of the Southern Ocean, especially in the Amundsen Sea. This thesis constitutes as groundwork for future investigations of protist assemblage changes in this area.
2. General introduction

2.1 The Southern Ocean and climate change

The Southern Ocean, defined as waters south of 50°S, is the most important region in the regulation of global climate and global marine carbon cycle (Sarmiento & LeQuere 1996, Sarmiento et al. 1998). It accounts for approx. 25% of the oceanic uptake of atmospheric CO₂, although it represents only 10% of total ocean surface area (Takahashi et al. 2002). In the Southern Ocean, some of the highest concentrations and deepest penetration of anthropogenic carbon are found (Lo Monaco et al. 2005, Waugh et al. 2006). Especially the northward flowing Antarctic Bottom Water, recently in contact with the atmosphere, is an important water mass for the storage of anthropogenic carbon in the deep ocean (McNeil et al. 2001). The cooling of the southward flowing subtropical surface waters accompanied with an increase in CO₂ solubility is one part of the carbon sink (solubility pump). The biological pump is another part, at which phytoplankton photosynthesis creates a gradient between the ocean and the atmosphere by reducing the surface water partial pressure of CO₂ (Arrigo et al. 2008). The fixed organic carbon may be stored for hundreds of years within the deep ocean, when it sinks out of the upper mixed layer (Broecker 1991). Hence, Southern Ocean regions with high rates of primary production and an active biological pump are important for the sequestration of CO₂ from the atmosphere (Arrigo et al. 2008).

The Southern Ocean is characterized by different oceanic currents and fronts (Fig. 1a). The most important current, the Antarctic Circumpolar Current (ACC), is of circumpolar character and plays a central role in the global thermohaline circulation, as it connects the Atlantic, Pacific and Indian Oceans. It flows clockwise from west to east around Antarctica (West Wind Drift), varies in depth and width, and transports a water volume of $1 \times 10^6$ m³ s⁻¹ on a 24,000 km long path (Rintoul et al. 2001, Boning et al. 2008, Thompson 2008). A counter-current to the ACC is the Antarctic Coastal Current (East Wind Drift), which flows westwards parallel to the coast of Antarctica. Several well-defined gyres are formed at the interference of the ACC and the Antarctic continental shelf, the two largest being the Weddell Gyre and the Ross Gyre (Fig. 1a).
Fig. 1 The Southern Ocean. a) Overview of the Southern Ocean and its main currents (after Rintoul (2011)). b) Oceanic fronts and regions of the Southern Ocean (locations of fronts after Orsi et al. (1995)).
The Southern Ocean is subdivided into four major regions (Fig. 1b), i.e. from north to south, the Subantarctic Zone (SAZ), the Polar Frontal Zone (PFZ), the Antarctic Zone (AZ) and the Subpolar Region (SR). The Subtropical Front (STF) in the north and the Subantarctic Front (SAF) in the south determine the SAZ. At the STF, the average Sea Surface Temperature (SST) changes from about 12°C to 7°C and stays above 4°C in the SAZ. The PFZ expands between the SAF and the Polar Front (PF). The AZ (average SST lower than 2°C) lies south of the PFZ and is separated by the ACC Southern Boundary (ACCSB) from the SR, where the average SST is around -1°C. The SST values are given in (Smith et al. 2005).

Sea surface salinity in the open Southern Ocean is ranging on average between 33.5 and 34.9 psu (Smith et al. 2005). In sea ice covered regions the salinity fluctuations are much higher, because of ice formation and melting. In brine systems, salinity can reach up to 200 psu (Thomas & Dieckmann 2003).

Sea ice plays a pivotal role in the Southern Ocean. The sea ice extent in winter can reach up to 20 million square kilometres. However, in summer, it decreases to about 4 million square kilometres (Fig. 2). In the far south of the Southern Ocean, light availability is extremely variable during the annual cycle. In winter, there is no or only little light, whereas in summer, the sun is shining during the whole day (polar day). Furthermore, the variable sea ice cover influences the light availability in the water column.

**Fig. 2** Southern Hemisphere 28-year average ice concentration maps for February and September, the months of average minimum and maximum sea ice extents, respectively (after Cavalieri and Parkinson (2008)).
The Southern Ocean is commonly characterized by its abundant macronutrients combined with low rates of annual average net primary production (Arrigo et al. 1998, Moore & Abbott 2000). In general, the lowest production rates are implicated with pelagic waters north of the sea ice zone (SIZ), with rates ranging from 0.08 to 0.22 g C m\(^{-2}\) d\(^{-1}\) in June to 0.5 to 1.0 g C m\(^{-2}\) d\(^{-1}\) in December (Arrigo et al. 1998). These areas of the Southern Ocean are regarded as high-nutrient, low-chlorophyll (HNLC) regions and low production rates are the result of several factors, including trace metal limitation, deep mixing of the upper water column, and low sun angles (Martin 1990, Mitchell & Holm-Hansen 1991, Boyd et al. 2000). Exceptions to the usually low production rates north of the SIZ are found along oceanic fronts, such as the Antarctic Polar Front, where divergence of surface waters brings waters with high nutrient concentrations to the surface, enforcing phytoplankton growth (Hense et al. 2000, Moore & Abbott 2000). Offshore islands, like South Georgia Island, are another exception, where current flow past rough or shallow topography can increase the flux of nutrients into surface waters (Korb & Whitehouse 2004). In general, the highest primary production rates in the Southern Ocean can be found in continental shelf areas (Smith & Gordon 1997, Sweeney 2003, Arrigo & van Dijken 2004), coastal polynias (Arrigo & Van Dijken 2007), and in the marginal ice zone (MIZ) (Smith & Nelson 1986). In these areas, production rates often exceed 2 g C m\(^{-2}\) d\(^{-1}\) (Louanchi et al. 1999, Sweeney 2003).

In the Southern Ocean, changes due to climate change are already apparent. Over the second half of the 20\(^{th}\) century, the sea surface temperature in the Southern Ocean has increased by \(~0.2\)°C, while the warming has concentrated within the Antarctic Circumpolar Current (ACC) (Gille 2008). During the past 50 years, an increase of atmospheric temperatures over the Antarctic Peninsula and West Antarctica has been observed, as well as a warming of the bottom water in the Weddell Sea (Mayewski et al. 2009). Environmental change in some areas of Antarctica and the Southern Ocean is more pronounced than in other areas of the world. The Western Antarctic Peninsula has warmed as much as 4.8 times of the global average (Montes-Hugo et al. 2009). For the future, model projections indicate a warming of the Antarctic interior by 3.4°C ± 1°C and a decreasing of sea ice extent by \(~30\)% over the current century (Mayewski et al. 2009).

The Pacific Sector of the Southern Ocean and especially the Amundsen Sea (Fig. 3) are the least studied oceanic regions in the world (Griffiths 2010, Griffiths et al. 2011). Severe ice conditions year-round and the geographic remoteness make sampling in this area difficult. The biodiversity of the Amundsen Sea, especially of the coastal and shelf areas, is almost unknown (Kaiser et al. 2009).
Fig. 3 Oceans and Seas around Antarctica.
2.2 Marine eukaryotic protists in the Southern Ocean

Eukaryotic organisms able to exist as single cells are stated as marine eukaryotic protists. They show an enormous range in size. The smallest known protist with a size less than 1 µm is a species from the genus *Ostreococcus*, whereas the largest protists can reach more than 1 cm in diameter (Caron et al. 2012). Plenty of protistan species form colonies (e.g. *Phaeocystis* spp.), including few to many cells.

Within the domain Eukarya, protists have representatives in every supergroup (Fig. 4). The opisthokonts contain the choanoflagellates, which are small, unicellular, heterotrophic flagellates. They are common in the marine plankton and can constitute important bacterivorous species (Throndsen 1970, Marchant 1985). Amoebozoans include many of the amoeboid taxa that are more important in marine benthic than in pelagic communities (Moran et al. 2007, Caron et al. 2012). The supergroup of Archaeplastida contains ecologically important taxa of chlorophytes and prasinophytes with many single-celled and colony forming taxa. They constitute a diverse and important part in oceanic ecosystems (Guillou et al. 2004, Worden 2006). Alveolates and stramenopiles together form the overarching supergroup of chromalveolates. The alveolates include the ciliates and dinoflagellates that are extremely widespread in marine ecosystems (Caron et al. 2012). Recent molecular investigations highlight the importance of the order Syndiniales within the dinoflagellates, which contains many parasitic forms (Guillou et al. 2008). The Rhizaria supergroup contains taxa ranging from small heterotrophic and phototrophic forms to some of the largest solitary and colonial unicellular organisms (Moreira et al. 2007, Pawlowski & Burki 2009). The most important clades are radiolarians, foraminifers, acantharias, and cercomonads. The stramenopiles include several very successful marine protists, most notably the diatoms. Recent gene surveys have revealed new lineages of small, heterotrophic marine stramenopiles (MAST), which have not been detected until recently with traditional approaches (Massana et al. 2004). Haptophytes, telonemids, and cryptophytes also contain many single-celled taxa. Among the haptophytes, coccolithophorids, and the genus *Phaeocystis* are the most important. The latter can form huge blooms in the Southern Ocean (Bodungen et al. 1986, Arrigo et al. 1999). Excavates and discicristates contain only few free-living marine taxa and are mainly restricted to shallow planktonic and benthic ecosystems, but also appear in deep-sea sediments (Buck et al. 2000).
Protists are the largest contributors to organic matter in the ocean. They constitute an important determinant of the structure and efficiency of Antarctic marine food webs and the flux of particles to the deep ocean (Smith Jr. & Sakshaug 1990, Priddle et al. 1992). The composition of the protistan community determines the fate of primary production due to interspecific differences in rates of growth, grazing mortality, and vertical flux (Moline & Prezelin 1996, Mengesha et al. 1998, Smetacek et al. 2002). Little is known about the factors affecting the diversity, distribution, and abundance of protists in the Southern Ocean, because most studies do not report detailed species information (Smetacek et al. 2002, Garibotti et al. 2005). The Southern Ocean is predicted to experience changes in temperature, stratification, mixed-layer depth and acidity, which will affect light climate and nutrient availability (Davidson et al. 2010). These effects will influence the composition and trophodynamics of the Antarctic marine ecosystem (Boyd 2002, Orr et al. 2005, Tortell et al. 2008). Detecting changes in microbial communities will not be possible without information about the current distributions and abundances of protists (Davidson et al. 2010).

In silica rich regions of the Southern Ocean, diatoms are generally dominating the protist biomass (Smetacek et al. 2004, de Salas et al. 2011). Long chain-forming
(Fragilariopsis kerguelensis, Pseudonitzschia spp. and Chaetoceros spp.) and large-celled (Corethron pennatum and Thalassiothrix antarctica) diatoms are the major species in the open, iron-limited ACC. In contrast, the dominant diatoms in near-shore, iron-rich environments are more similar to diatoms from the continental shelves of the world ocean, i.e. weakly silicified species of the genera Thalassiosira and Chaetoceros (Smetacek et al. 2004). When silica is depleted, dinoflagellates and other small flagellates are more favoured (de Salas et al. 2011). In many regions, colonies of the haptophyte Phaeocystis sp. dominate the biomass (Smetacek et al. 2004).
2.3 Molecular methods

Traditional morphology-based taxonomies suffer from limitations and difficulties for rapid and easy identification, especially for species with small cell size (<10 µm) and/or few morphological features (Caron et al. 2012). In recent years, molecular approaches have emerged and established as powerful tools for broad and relatively rapid diversity assessments of protistan assemblages (Caron et al. 2009). Within morphologically defined species, genetic information has been useful for the identification of cryptic taxa (Pfandl et al. 2009). Genetic approaches have also provided additional information for distinguishing species with amorphous morphologies (Nassonova et al. 2010).

One of the most important gene marker used for these molecular approaches is the 18S rRNA gene. It is a widely employed and valuable ‘bar-code’ to assess eukaryotic diversity, as it is universally present in living organisms and there is significant sequence data for comparison in public databases, such as GenBank (Ebenezer et al. 2012). In addition, the 18S rRNA gene contains regions that are well conserved within a species and generally different between species (Ki 2012). The gene marker is commonly used for different molecular approaches, such as fluorescent in situ hybridization (FISH) (Hosoi-Tanabe & Sako 2005), microarrays (Metfies & Medlin 2008), sequencing of clone libraries (Diez et al. 2001) and next-generation sequencing (Cheung et al. 2010, Stoeck et al. 2010).

The basis of these molecular approaches forms the collection, extraction, and purification of nucleic acids from environmental samples. The second step is the application of one or more methods for assessing sequence diversity within the samples. Finally, the interpretation of this genetic diversity in accordance with databases containing taxonomic information is completing the survey. To date, two basic genetic approaches have been established, those that use DNA sequences themselves, and those that apply DNA fragment analysis to assess the composition of the most abundant taxa within a community (Caron et al. 2012). In the following, the three molecular approaches used in this work will be briefly described.
2.3.1 Automated ribosomal intergenic spacer analysis (ARISA)

The most common DNA fragment analysis methods are denaturing gradient gel electrophoresis (DGGE) (Hamilton et al. 2008, Not et al. 2008), terminal restriction fragment length polymorphism (T-RFLP) (Vigil et al. 2009), and ARISA. They all use species-specific fragment lengths as genetic signatures of eukaryotic taxa and are relatively fast and inexpensive. These approaches are commonly referred as fingerprint methods, because they deliver a snapshot of a subset of the dominant taxa within a community (Caron et al. 2012).

The ARISA approach is used to describe the diversity of communities and facilitates the comparison of different samples. It is well established for investigations of prokaryotic diversity (Danovaro et al. 2009, Smith et al. 2010). However, to our knowledge, this method has not yet been applied for diversity investigations of eukaryotic phytoplankton. The approach is based on the length heterogeneity of the intergenic regions between the 18S rRNA and 28S rRNA genes, i.e. different taxa have different internal transcribed spacer (ITS) fragment lengths (Baldwin 1992). The overview is only qualitative and approximate, because of possible overlapping of fragment sizes among different species. The number of fragments that can be detected also limits the sensitivity of DNA fragment based approaches. Thus, these approaches are not appropriate for assessing the total species richness (Bent et al. 2007).

A short overview of the procedure is shown in Figure 5. After the DNA isolation from the environmental samples, the ITS region is amplified with fluorescence labelled primers. Subsequently, the fragment lengths can be measured on a sequencer by separating them via capillary electrophoresis. The fragment composition of the different samples can be compared and delivers a diversity overview.
**Fig. 5** Overview of the automated ribosomal intergenic spacer analysis (ARISA) approach. After DNA isolation, the ITS region is amplified with fluorescence labelled primers and the fragment size is measured on a sequencing machine.
2.3.2 Sequencing of 18S rRNA gene clone libraries

The cloning and sequencing approach has been widely used to study the diversity and community composition of protists in the Southern Ocean (Diez et al. 2001, Lopez-Garcia et al. 2001). These studies contributed significantly to illuminate eukaryotic protist diversity and composition in the marine environment and revealed a huge hidden diversity. The cloning approach provides information of long sequence fragments and therefore, the phylogenetic characterization is highly reliable. An overview of the procedure is shown in Figure 6. The 18S RNA gene is amplified from environmental samples and transferred into carriers (plasmids), so that every carrier contains an 18S RNA gene of one organism. Subsequently, the carriers are transferred into E. coli cells for enrichment (one E. coli contains one plasmid). After a growth phase, the colonies contain enough carrier DNA material that can be isolated. Then, the 18S RNA gene is reamplified and sequenced.

There are some restrictions and facts to be considered when dealing with cloning results. The cloning approach suffers from potential cloning biases that may mask the real community diversity (Forns et al. 1997). Also the diversity may be underestimated by the limited throughput (Bent & Forney 2008). The examination of typically 100 clones per library only provide general information on the structure of protist communities, but are not sufficient for meaningful comparisons (Cheung et al. 2010). Nevertheless, it is a good approach to retrieve comparable diversity patterns (Marande et al. 2009) and can be done routinely to help estimating the diversity of planktonic communities (Potvin & Lovejoy 2009). The cloning approach also delivers robust sequence data (because of the length of fragments of >1000 bp), especially when using small- and large-subunit, which strengthen the phylogenetic information and reduce the risk of misclassification (Marande et al. 2009). Hence, many studies used the cloning approach as reliable, comparable and well established method (Diez et al. 2001, Lopez-Garcia et al. 2001, Moon-van der Staay et al. 2001, Potvin & Lovejoy 2009).
Fig. 6 Overview of the clone library approach. After DNA isolation and amplification of the 18S rRNA gene, the amplicons are transferred into plasmids. The plasmids are integrated into E. coli cells and after growth of colonies, the plasmid DNA is isolated and the 18S rRNA gene is reamplified. Subsequently, the fragments can be sequenced.
2.3.3 Next-generation sequencing: 454-pyrosequencing

During the past decade, 454-pyrosequencing (Margulies et al. 2005) has been established as a high throughput sequencing methodology and is replacing the Sanger sequencing for comparative metagenomics (Kunin et al. 2010). This approach allows assessing microbial communities with high resolution, based on sufficient deep taxon sampling (Margulies et al. 2005, Stoeck et al. 2010). The method has revolutionized surveys of microorganism diversity (Medinger et al. 2010). The principles of the approach are explained in Margulies et al. (2005) and a brief overview is shown in Figure 7. In this work, a ~670 bp fragment of the 18S rRNA gene, containing the highly variable V4 region, was used for the diversity survey. After amplification of the fragment, the samples are directly sequenced. Therefore, single strands are captured with capture beads (one molecule per bead) and enriched via amplification. This results in a library containing beads, each with sufficient fragments of one organism. The beads are loaded into picotiter plates, so that every well contains one bead. The nucleotides (A, C, G, and T) are alternately washed over the plates and the light signal, occurring when a nucleotide is incorporated, is measured. For the bioinformatic processing of the obtained sequences, there are several pipelines and procedures existing. Crucial points are the quality filtering (e.g. removal of too short and too long reads, non-target sequences, chimeras and sequences with too many ambiguous bases), the clustering into operational taxonomic units (OTUs) and the taxonomical affiliation of the resulting OTUs.

The advantage of this approach (compared to the cloning approach) is the recovery of several thousand sequences per sample in a four-hour run. It even allows the revealing of the rare biosphere (phylogenotypes with an abundance <1%) and has already led to the conclusion that it is more diverse and larger than formerly estimated (Sogin et al. 2006). However, the length of the fragments (~600 bp) is too short for robust phylogenetic analyses, at least for affiliations down to the species level.
Fig. 7 Overview of the 454-pyrosequencing approach. After DNA isolation, V4 region of the 18S rRNA gene is amplified. The single strands are captured with capture beads (one molecule per bead) and enriched via amplification. The beads are loaded into picotiter plates, so that every well contains one bead. The nucleotides (A, C, G and T) are alternately washed over the plates and the light signal, occurring when a nucleotide is incorporated, is measured.
3. Aims and outline of the dissertation

3.1 Aims

Molecular approaches have recently emerged as tools for broad and relatively rapid diversity assessments of protistan assemblages. However, so far, they are not implemented in all areas of investigation and the applicability and interpretation of molecular data is still under discussion. One aim of this thesis is the establishing of molecular approaches in the diversity investigation of eukaryotic protists in the Southern Ocean and the comparison of different approaches (described in 2.3).

The Pacific sector of the Southern Ocean is one of the least investigated regions of the oceans, because of severe ice conditions year-round and the geographic remoteness. Not much is known about the diversity and composition of protists in this area, especially in the Amundsen Sea. Hence, the second aim of the thesis is to deliver a comprehensive and taxon detailed overview of protist assemblages in the Pacific sector of the Southern Ocean, especially the Amundsen Sea. In doing so, the questions if different large-scale water masses are characterized by distinct protist community assemblages and if a single large-scale water mass shows regional distinct community patterns will be addressed. Additionally, the extent and role of the rare biosphere as background population, with regard to the “everything is everywhere” hypothesis, will be highlighted.

3.2 Outline

Comparison of molecular approaches

There are several molecular approaches for determining the diversity and composition of protist assemblages. The two, which deliver the most taxon detailed information, are sequencing of clone libraries and 454-pyrosequencing. The clone library approach suffers from low throughput, but it might be useful as backbone for refined phylogenetic analysis of 454-pyrosequencing data. To assess the applicability, comparability, advantages and disadvantages of these methods, Manuscript I addresses the comparison between these two approaches. The manuscript shows that our understanding of protist diversity and structure varies strongly depending on the molecular method used. Furthermore, comparisons of new 454-pyrosequencing data with previously published clone library data of protist diversity have
Aims and outline

to be handled with care and should not be over interpreted. Moreover, the cloning approach seems to be of very limited suitability as backbone for a refined phylogenetic analysis of phylotypes identified by 454-pyrosequencing data.

Diversity and composition of protist assemblages in the Pacific sector of the Southern Ocean

Changes in temperature, stratification, mixed-layer depth and acidity in the Southern Ocean, due to climate change, will affect light climate and nutrient availability. These effects will influence the entire Antarctic marine ecosystem, from small protists up to higher trophic levels. Detecting changes in microbial communities will not be possible without information about the current distributions and abundances of protists. Manuscript II and III describe the diversity and composition of eukaryotic protists in the scarce investigated Pacific sector of the Southern Ocean. Samples were collected by the author during the RV Polarstern cruise ANT XXVI/3 (January 2010 – April 2010) from Wellington to Punta Arenas. These samples were analyzed with molecular approaches, such as ARISA and 454-pyrosequencing, combined with flow cytometry (only Manuscript II) and high-performance liquid chromatography (HPLC). Both manuscripts show the applicability of ARISA for investigating the structure of protist assemblages in large numbers of samples.

Manuscript II focuses on a north-south transect, crossing all major oceanic fronts of the Southern Ocean. Each water mass in the studied area harboured characteristic protist assemblages. The rare biosphere did not appear as a background population that contains species, which are abundant under the other environmental conditions. However, the present study does not resolve possible seasonal effects. The most prominent separator in the investigated area was the PF. Dinoflagellates and smaller cells, like labyrinthulids and Syndiniales, were dominating north of the PF, whereas haptophytes and diatoms became more important south of the PF.

Manuscript III concentrates on a west-east transect along the coast of Antarctica from the eastern Ross Sea to the Amundsen Sea. The manuscript shows that within a single water mass, protist assemblages differed in dimensions and species composition, according to geographical and environmental conditions. There were two major groups, the offshore and the inshore group. Biomass and microeukaryotes contribution to total Chl a were highest in the inshore group, whereas in the offshore group the contribution of nanoeukaryotes was the highest across the entire transect. Diatoms were the most prominent protist class, and the diatom species appearing as most abundant differed among the locations.
4. Manuscripts

4.1 List of manuscripts

This thesis is based on the following manuscripts:

I. Christian Wolf, Estelle S. Kilias and Katja Metfies
   Investigating eukaryotic protist diversity – a comparison of clone library and
   454-pyrosequencing data

II. Christian Wolf, Ilka Peeken, Mirko Lunau, Stephan Frickenhaus, Estelle S. Kilias and Katja Metfies
   Oceanographic fronts in the Southern Ocean determine biogeographic
   differences in eukaryotic protist communities – new insights based on 454-
   pyrosequencing

III. Christian Wolf, Ilka Peeken, Stephan Frickenhaus, Estelle S. Kilias and Katja Metfies
    Regional variability in eukaryotic protist communities in the Amundsen Sea

Other manuscripts prepared with contribution of the candidate from the period:

Estelle S. Kilias, Eva-Maria Nöthig, Christian Wolf and Katja Metfies
   Genetic diversity of picoeukaryotes at the long-term observatory “Hausgarten”
   to be submitted

Estelle S. Kilias, Eva-Maria Nöthig, Ilka Peeken, Christian Wolf and Katja Metfies
   Protist distribution in the western Fram Strait in summer investigated by molecular
   techniques
   to be submitted
Estelle S. Kilias, Gerhard Kattner, Christian Wolf and Katja Metfies
Exploring picoeukaryotic diversity by the implementation of molecular tools – a survey through the Arctic Ocean
to be submitted

Stefan Thiele, Isabelle Schulz, Christian Wolf, Bernhard M. Fuchs, Philipp Assmy, Katja Metfies, Victor Smetacek, Estelle S. Kilias and Rudolf Amann
Comparison of classical methods with modern molecular approaches for the investigation of pico- and nanoplankton assemblages
in preparation

4.2 Statement of contribution to the manuscripts

Manuscript I
The experiments were planned together with Estelle S. Kilias and Katja Metfies and performed by Estelle S. Kilias and myself. Estelle S. Kilias and I analyzed the data. I wrote the manuscript together with Estelle S. Kilias. Both authors contributed equally to the manuscript.

Manuscript II
The experiments were planned together with Katja Metfies. I performed the sampling and the molecular experiments. Mirko Lunau conducted flow cytometry and Ilka Peeken performed HPLC. The pipeline for the molecular analysis was developed and evaluated in co-operation with Estelle S. Kilias. Stephan Frickenhaus advised me on bioinformatical questions. I conducted all lab work related to this manuscript, the data analysis and writing of the manuscript.

Manuscript III
The experiments were planned together with Katja Metfies. I performed the sampling and the molecular experiments. Ilka Peeken performed HPLC. Stephan Frickenhaus advised me on bioinformatical questions. Estelle S. Kilias contributed to the discussion of the results. I conducted the data analysis and wrote the manuscript.
4.3 Manuscript I

Investigating eukaryotic protist diversity – a comparison of clone library and 454-pyrosequencing data

Christian Wolf*, Estelle Silvia Kilias* and Katja Metfies

Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven, Germany

* These authors contributed equally to this work

Journal of Phycology, submitted
Abstract

We compared two molecular approaches, i.e. sequencing of 18S rRNA clone libraries and 454-pyrosequencing, which are commonly used for describing protist diversity. The comparison was conducted with four Arctic water samples, focusing on the picoplankton (0.4-3 µm), and with one Antarctic water sample, examining the whole size spectrum (>0.4 µm). We found different outcomes between the two different methods. Both approaches revealed phylotypes that were not found with the other approach. The abundant biosphere, defined by the 454-pyrosequencing approach, was not fully recovered by the clone library approach. We found a bias of the cloning method against several groups, e.g. haptophytes in the Arctic samples and diatoms in the Antarctic sample. In summary, prior cloning data have to be handled with care, when compared with 454-pyrosequencing data. Additionally, cloning data are only of limited suitability as backbone for phylogenetic analysis of 454-pyrosequencing data.

Key index words
18S rRNA gene, 454-pyrosequencing, Clone library, Eukaryotic protists, Genetic diversity, Phytoplankton
Introduction

Until recently, the majority of phylogenetic investigations of the eukaryotic protist diversity were based on the analysis of 18S rRNA gene clone libraries. Sequencing of 18S rRNA gene clone libraries provides sequence information of the complete gene and allows a reliable phylogenetic characterization (Diez et al., 2001, Lovejoy et al., 2006). These studies contributed significantly to elucidate eukaryotic phytoplankton diversity and community composition in the marine environment. They revealed a huge hidden diversity, especially originating from the picoeukaryotic size fraction. Representatives of all major phytoplankton taxa could be found in the sequence libraries (Diez et al., 2001, Lovejoy et al., 2006, Lovejoy & Potvin, 2011). However, it is expected that the real diversity is even higher than observed in the libraries (Diez et al., 2001, Lopez-Garcia et al., 2001, Moon-van der Staay et al., 2001).

In summary, sequencing of the 18S rRNA gene is a reliable approach (Cheung et al., 2010, Diez et al., 2001, Lovejoy et al., 2006, Lovejoy & Potvin, 2011). It is used as a gold standard in molecular assessments of phytoplankton diversity. The 18S rRNA gene is such a widely employed and valuable ‘bar-code’ to assess eukaryotic phytoplankton diversity, because of its slow evolutionary rate and its occurrence in all eukaryotic organisms (Amann & Kuhl, 1998, Vaulot et al., 2008). It is sufficiently slow to allow differentiation between organisms at different taxonomic levels (Vaulot et al., 2008).

However, quantitative interpretation of 18S rRNA clone libraries is challenged in various ways, e.g. by the presence of multiple gene copies that may not be identical in all species and thus falsify the diversity and community structure. Furthermore, sequencing of 18S rRNA gene clone libraries has a number of other drawbacks, such as its vulnerability towards a potential bias induced by primer specificity (Farris & Olson, 2007) and the production of chimerical sequences (Berney et al., 2004), a disparate incorporation rate within the cloning process and a limited throughput (Bent & Forney, 2008). Additionally, the analysis is time-consuming and cost-intensive.

The introduction of the 454-pyrosequencing technology for the assessment of microbial communities is an attractive alternative to the sequencing of clone libraries. It is independent of the cloning step and allows high-resolution sequencing of microbial sequences (Margulies et al., 2005). Furthermore, compared to the analysis of clones, the massively parallel pyrosequencing provides more sequences and uncovers more organisms by less chimera formation and less costs (Huse et al., 2008). In respect of the vast microbial diversity, the greater sampling depth is advantageous and even allows elucidating the diversity of the
rare biosphere (Sogin et al., 2006). However, one caveat of the pyrosequencing approach is the tendency to overestimate the number of rare phylotypes, because of sequencing errors. Such errors will run the risk of inflating the diversity estimates, due to the fact that every single read is considered to represent a community member (Kunin et al., 2010). An additional caveat is the short sequence length of approximately 500 bp, which results in a less robust phylogenetic affiliation. Here, the clone library approach is advantageous over the 454-pyrosequencing approach because it delivers longer sequences that cover the whole 18S rRNA gene and are better suited for phylogenetic analyses (Diez et al., 2001, Lovejoy et al., 2006).

Regarding the advantages and drawbacks of 18S rRNA gene clone libraries and 454-pyrosequencing for the assessment of protist communities, we would like to address three main questions in this study: (i) How do the choice of a molecular method influence our understanding of protist diversity and community structure? (ii) Do clone library data reflect the abundant biosphere? (iii) Are clone library data suitable to serve as backbone for phylogenetic analysis of 454-pyrosequencing data?

To answer these questions, we analyzed four samples from the Arctic Ocean, comprising the picoeukaryotic fraction (0.4-3 µm) and one sample from the Southern Ocean, comprising the whole size fraction (>0.4 µm). We choose the sampling setup to exclude possible bias induced by cell size or geographical location.
Material and Methods

Location and sampling
The study area comprises four stations, located in the Fram Strait (Arctic Ocean), as well as one station from the Southern Ocean (Fig. 1). All samples are part of other larger studies (Kilias et al. in prep., Wolf et al. submitted). The four Arctic stations extended between 2-6°E longitudes and 78-80°N latitudes and were sampled during the ARK XXIV/2 cruise onboard the RV Polarstern in July 2009. The Antarctic station, located at 164.9°W longitude and 69°S latitude, was sampled during the RV Polarstern cruise ANT XXVI/3 in February 2010. The Arctic samples were collected at the subsurface maximum chlorophyll layer with a rosette system, fitted with Niskin bottles and appointed with depth, temperature, salinity, and fluorescence profilers. The Antarctic sample was collected using the ship pumping system (membrane pump), located at the bow at 8 m depth below the surface. In both cases, 1.5 l of sea water were successively filtered with a pressure of 200 mbar onto Isopore Membrane Filters (Millipore, USA) with a pore size of 10 µm, 3 µm and finally 0.4 µm. The filters were stored at -80°C until further treatment in the laboratory.

DNA extraction
DNA extraction was carried out with the E.Z.N.A TM SP Plant DNA Kit (Omega Bio-Tek, USA) following the manufacturer’s instructions. DNA concentration was determined with a NanoDrop 1000 system (Thermo Fisher Scientific, USA).

Clone library construction
The 18S rRNA gene of the Arctic samples was amplified using the specific primers 82F (5’-GTA AAA CTG CGA ATG GCT CAT-3’) and 1528R (5’-TGA TCC TTC TGC AGG TTC ACC TAC-3’) and genomic DNA isolated from the 0.4 µm filter as template. The amplification of the Antarctic sample was conducted using the primer combination 300F (5’-AGG GTT CGA TTC CGG AG-3’) and 1200R (5’-CAG GTC TGT GAT GCC C-3’), because the former combination resulted in a poor PCR product. Furthermore, the whole protist assemblage was used for the methodological comparison of the Antarctic sample. In this respect, the 18S rRNA gene of each fraction was amplified and equal volumes of each PCR product were pooled before the purification. The PCR reaction mixture contained 1 x HotMaster Taq Buffer containing 2.5 mM Mg\(^{2+}\) (5 Prime, USA), 0.4 U of HotMaster Taq polymerase (5 Prime, USA), 10 mg/ml BSA, 10 mM dNTP-mix (Eppendorf, Germany), 10
µM of each Primer and 1 µl of template DNA in a final volume of 20 µl. PCR reactions were carried out in a Mastercycler (Eppendorf, Germany) under the following conditions: an initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 45 s, 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. The purification of the resulting PCR fragment was carried out with the Gel Purification Kit (Invitrogen, USA), following the manufacturers protocol. Subsequently, the fragment was cloned into the pDrive Cloning Vector (QIAGEN, Germany) taking advantage of the PCR Cloning Kit (QIAGEN, Germany) and transformed into TOP10 chemo-competent E.coli cells (Invitrogen, USA). Positive colonies were screened for similar inserts by performing a restriction fragment length polymorphism (RFLP) analysis using the multiclit enzyme Hae III (New England Biolabs, USA). Clones with a similar RFLP pattern were considered to display the same phylotype and grouped into an OTU (operational taxonomic unit). One to two representatives of each OTU were sequenced using the 300F (see above) and 528F (5’-GCG GTA ATT CCA GCT CCA A-3’) primer under the following conditions: an initial denaturation step at 96°C for 1 min, 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s and extension at 60°C for 4 min. The terminal sequencing was carried out on an ABI Prism 310 Genetic Analyzer (Applied Biosystems, USA).

454-pyrosequencing

The hypervariable V4 region of the 18S rRNA gene was amplified taking advantage of the primer combination 528F (5’-GCG GTA ATT CCA GCT CCA A-3’) and 1055R (5’-ACG GCC ATG CAC CAC CAC CCA T-3’). The PCR mixtures were composed as described previously for the clone library construction. Reaction conditions were as following: an initial denaturation at 94°C for 3 min, 30 cycles of denaturation at 94°C for 45 sec, annealing at 59°C for 1 min and extension at 72°C for 3 min, and a final extension at 72°C for 10 min. Subsequently, the amplicons were purified with the Mini Elute PCR Purification Kit (QIAGEN, Germany). In case of the Antarctic sample, an equal volume of PCR reaction of each size fraction was pooled and purified with the MiniElute PCR purification kit (Qiagen, Germany) following the manufacturer's instructions. Pyrosequencing was performed on a Genome Sequencer FLX system (Roche, Germany) by GATC Biotech AG (Germany).
Data analysis

The two raw sequences of each sequenced clone were assembled with the software Lasergene 10 (DNASTAR, USA) and a consensus sequence was built. All sequences (clone consensus sequences and 454-pyrosequencing reads) were checked for errors (reads with many unresolved bases) implied by the sequencing process and sequences with more than one uncertain base (N) were removed. Remaining sequences were checked for possible chimera formation by applying the detecting software UCHIME 4.2.40 (Edgar et al., 2011) and all sequences considered being chimeric were excluded from further analysis. Residual sequences were added to the Lasergene 10 software (DNASTAR, USA) and clustered into OTUs at the 97% similarity level. Subsequently, singletons from the 454-pyrosequencing data were removed. Consensus sequences of OTUs were aligned using the software HMMER 2.3.2 (Eddy, 2011). Subsequently, taxonomical affiliation was determined by placing the consensus sequences into a reference tree, consisting of 1200 high quality 18S rRNA gene sequences of Eukarya from the SILVA reference database (SSU Ref 108), using the software pplacer 1.0 (Matsen et al., 2010). The compiled reference database is available on request in ARB-format. Detected non-phytoplankton sequences originating from metazoans and fungi were removed. Rarefaction curves were computed using the freeware program Analytic Rarefaction 1.3. In case of the clone library sequences, a phylogenetic tree was generated using MEGA version 4 (Tamura et al., 2007) on the basis of Maximum Parsimony principles by the application of the Juke Cantor model and 1000 bootstrap restarts. The clone library sequences generated in this study have been deposited at GenBank under Accession No. JX840877-JX840942. The 454-pyrosequencing reads were deposited at GenBank’s Short Read Archive (SRA) under Accession No. SRA058841 (Arctic samples) and SRA056811 (Antarctic sample).
Results

The five clone libraries resulted in a total of 698 high quality clones (Table 1). After the RFLP analysis of 182 (HG1 and HGS3) to 117 (ANT25) clones, a total of 134 Arctic and 64 Antarctic clones were sequenced. Non-target sequences (metazoan and fungi sequences) were only found in the pooled ANT25 clone library (6%), while chimeras were formed in all clone libraries (6-19%), except in library HG4 (0%). Final clustering of the residual sequences resulted in seven (HG4) to 24 (HGS3) different phylotypes.

In total, between 7539 (HG4) and 45772 (ANT25) 454-pyrosequencing reads were obtained. The analytical process revealed 2-6% of chimeric sequences. The quality filtering reduced the initial read number to a final range of 5220 (HGN4) to 30561 (ANT25) reads, that resulted in 709 (HG4) to 1153 (ANT25) different phylotypes, based on a clustering at the 97% similarity level.

The clone library and 454-pyrosequencing approach provided different numbers of phylotypes that differ by several levels of magnitude. The rarefaction curves present an estimation of the local species richness based on the respective approach (Fig. 2). Two Arctic clone libraries (HG1 and HGN4) present a rarefaction curve that is saturated or almost saturated, while the species richness of the residual Arctic and the Antarctic clone library is not totally covered. The 454-pyrosequencing rarefaction curve for sample ANT25 reached the plateau, whereas the curves of the four Arctic samples (HG1, HG4, HGN4 and HGS3) ended in the slope phase.

Comparison of clone library and 454-pyrosequencing data set - Arctic

In total, 47 different phylotypes have been identified in the clone libraries from the Arctic samples (Fig. 3 and Table 2). The number of phylotypes in each sample ranged from seven (HG4) to 24 (HGS3). Most phylotypes (32) grouped within the alveolates, whereas the majority of 30 phylotypes affiliated with dinoflagellates and only two with ciliates. Picobiliphytes, cryptophytes, rhodophytes and stramenopiles were represented by two, three, one and two phylotypes, respectively. Seven phylotypes affiliated with chlorophytes, of which five were closely related to Micromonas pusilla. The majority of the phylotypes (37) were just present at one of the different stations. Only ten phylotypes could be found in more than one station. Three phylotypes (ARK_3, ARK_5 and ARK_15) were present in three samples. There was no phylotype, which was recovered from all samples. Samples HGN4 and HGS3 showed the highest similarity with five shared phylotypes. The most abundant phylotype in
the clone libraries was ARK_15, and affiliated with *Micromonas pusilla* in the phylogenetic tree. This phylotype contributed 77.1% of the clones in sample HG1, 65.9% in sample HG4, and 47.5% in sample HGN4. In sample HGS3, the phylotype was not found at all. In this sample, the phylotypes ARK_12 (Syndiniales clone) and ARK_14 (*Geminigera cryophila*) showed the highest clone abundance with 26.6% and 23.7%, respectively.

In total, 709 to 1014 phylotypes were obtained by 454-pyrosequencing. The data set was composed of 18.2-51.4% dinoflagellates, 1-3.4% ciliates, 16.3-33.1% haptophytes, 0-0.9% rhodophytes, 0.5-2.2% cryptophytes, 3.4-42.2% chlorophytes and 14.7-16.8% stramenopiles. In contrast, the clone libraries did neither contain haptophytes, nor rhodophytes, 0.6-28.1% cryptophytes, 4.3-82.9% chlorophytes, 1-2.2% stramenopiles, 3.5-64.3% dinoflagellates and 0.7-9.7% ciliates.

Thirteen out of the 47 clone library phylotypes (27.7%) could not be recovered in the 454-pyrosequencing data set. The clone libraries of samples HG1, HG4 and HGS3 each covered 20% of the abundant phylotypes of the 454 data set (data not shown). The clone library of sample HGN4 covered none of the abundant 454 phylotypes.

**Comparison of clone library and 454-pyrosequencing data set - Antarctic**

In the clone library generated from the Antarctic sample (ANT25), 19 different phylotypes have been found (Fig. 4 and Table 3). The majority of these phylotypes belonged to alveolates (10). Among the alveolates, eight phylotypes affiliated with dinoflagellates, whereas one belonged to the ciliates and Syndiniales, respectively. There were two phylotypes belonging to the haptophytes, whereas both were close to the genus *Phaeocystis*. Five phylotypes of the library belonged to the stramenopiles, at which two were representatives of diatoms. One cryptophyte, belonging to the genus *Geminigera*, and one picobiliphyte were found.

The 454-pyrosequencing revealed 1153 different phylotypes. The relative abundance of the phylotypes retrieved from the library and the respective relative abundance in the 454-pyrosequencing data set are shown in Table 3. Four out of the 19 clone library phylotypes (21.1%) were not found in the 454 data set. From the 12 abundant phylotypes in the 454 data set (data not shown), only four (33.3%) were found in the clone library. The phylotype with the highest relative abundance was the same in both data sets (clone ANT_13). The 454-pyrosequencing data set was composed of 30.8% haptophytes (39.5% in the clone library), 1.2% chlorophytes (0%), 1.6% cryptophytes (1%), 1.1% rhodophytes (0%), 22.9% diatoms (2%), 9.5% other stramenopiles (5.8%), 23.6% dinoflagellates (42.5%), 2.5% Syndiniales (1.9%), and 6.8% ciliates (4.8%).
Both the Arctic and Southern Ocean samples showed that ~20% of the phylotypes retrieved via the cloning approach were not found in the 454-pyrosequencing data, and only 20-30% of the abundant phylotypes of the 454-pyrosequencing data were found via the cloning approach.
Discussion

Although, culture-independent methods as traditional clone library and the novel 454-pyrosequencing are often used for screening microbial community structures (Cheung et al., 2010, Diez et al., 2001, Lovejoy et al., 2006), studies that directly compare both approaches are scarce. To our knowledge, those studies primarily focused on the genetic diversity of prokaryotes (Zhang et al., 2011). Here, we present the first comparison of both molecular methods for eukaryotic protists. The comparison has been carried out on picoeukaryotic protists in the Arctic and on the whole protist assemblage (covering all size classes) in the Antarctic. Independent of the size fractionation or geographical location, the two methods showed high discrepancies in distribution as well as in relative abundance.

(i) How does the choice of a molecular method influence our understanding of protist diversity and community structure?

Our data suggest that our understanding of protist diversity and community structure is strongly dependent on the molecular method used. According to the Arctic clone library data, the picoplankton communities were mainly dominated by chlorophytes, while haptophytes were absent. In contrast, the 454-data suggest that haptophytes are dominating in these samples and chlorophytes are less abundant. It can be excluded that it is hard or impossible to clone haptophytes, because numerous Phaeocystis sp. clones have been found in the Antarctic library. However, phylotypes that affiliated particularly to small non-colonial haptophytes have been already reported to occur in merely small numbers in clone library data (Lovejoy et al., 2006, Potvin & Lovejoy, 2009). Amacher et al. (2011) also proved that it is possible to retrieve Emiliania huxleyi (belonging to haptophytes) sequences via cloning. Additionally, we have checked the suitability of the used primers. All primers matched with haptophytes (including Phaeocystis and Emiliania) in the reference database. Nevertheless, it cannot be excluded, that primer efficiency could be an explanation for the observed discrepancy. There are several studies showing that different primer sets applied on the same sample resulted in different diversity and abundance patterns (Potvin & Lovejoy, 2009, Stoeck et al., 2010, Jeon et al., 2008). Mismatches between the primers and the 18S rRNA genes sequences (Liu et al., 2009), and primer competition (Potvin & Lovejoy, 2009) are considered to be responsible for it. However, the primer sets used in this study covered all major taxonomic groups. Nevertheless, we observed that the primer set used for the Arctic samples might be biased against haptophytes and the primer set used for the Antarctic sample against diatoms.
Another explanation for the discrepancies between the clone libraries and the 454 data sets could be the mutual impact of organisms during the PCR or cloning step. Amacher et al. (2011) showed that certain co-occurring organisms bias the recovery of organisms in clone libraries. Our data suggest that *Micromonas* sp. might influence the cloning efficiency of other organisms, especially on *Phaeocystis* sp. *Micromonas* sp. sequences were overrepresented in the Arctic clone libraries and no *Phaeocystis* sequences were retrieved. However, in three of the four 454-pyrosequencing samples, *Phaeocystis* sp. was more abundant than *Micromonas* sp. Thus, if *Micromonas* sp. occurs in sufficient abundance, it might have been favored during the cloning step. In contrast, in the Antarctic clone library, we observed a bias against diatoms. The 454-pyrosequencing revealed a tenfold higher diatom abundance than the clone library. Here, our data suggest that dinoflagellates or *Phaeocystis* had a repressing effect on diatoms during the cloning step.

Although, the picoeukaryotic and the whole protist approach both showed high differences between the clone libraries and the 454-pyrosequencing data, the aberration was higher for picoeukaryotes. Beside the previously discussed bias induced by the use of different primer sets, another possible bias may rely on the interpretation of the RFLP patterns. The interpretation of RFLP patterns may be biased because of incomplete DNA digestion, variable DNA load of the gels, similar cutting sites among species, or limited sensitivity of the detection technique (Wu et al., 2000). In this regard, the whole size community may be represented by more distinct patterns, while patterns of picoeukaryotes could be more similar. This would lead to a reduced picture of picoeukaryotic diversity, because a limited number (1-2) of representatives were sequenced for each interpreted group. Altogether, we have shown that the two different methods revealed different community compositions and the clone library approach even defalcated whole taxonomical groups.

(ii) Do clone library data reflect exclusively the abundant biosphere?
A recent assumption of previous studies (Pedros-Alio, 2006) is that clone libraries cover at least the abundant biosphere of protist communities. However, our results suggest that the clone library data did not cover the whole abundant biosphere of the 454-pyrosequencing data set. All clone libraries missed over 50% of the abundant biosphere, identified by 454-pyrosequencing. It should be mentioned that of the missing 50% most phylotypes were affiliated to *Phaeocystis* (Arctic samples) and to diatoms (Antarctic sample). In addition, we found phylotypes in the clone libraries that were not recovered in the 454-pyrosequencing data. This is most pronounced in the Arctic samples and supported by the 454-pyrosequencing
rarefaction curves, which showed that the samples were not exhaustively sequenced. Thus, with a higher sequencing effort the additional sequences in the clone libraries might have been recovered by the 454-pyrosequencing. In contrast, the rarefaction curves of the clone libraries simulate that the total diversity was sufficiently recovered, although the 454 approach revealed far more phylotypes. Our data support the common sense that the concept of rarefaction curves is questionable. The calculation is biased by the presence of multiple 18S rRNA gene copies. Additionally, the removal of singletons in the 454 approach during the analytical process consequently leads to a saturated curve (e.g. in the Antarctic sample). The calculation for clone library data is biased by the limited throughput, the low number of phylotypes retrieved and the overrepresentation of single phylotypes (e.g. *Micromonas* sp. in the Arctic samples).

We have observed that groups, which showed a low abundance in the 454 data (<2% in total), can also be retrieved via the cloning approach. However, this observation does not apply to all of the clone libraries, suggesting that the recovery of OTUs in clone libraries is random.

(iii) Are clone library data suitable to serve as a backbone for phylogenetic interpretation of 454-pyrosequencing data?

Based on the findings and remarks discussed above, the suitability of clone library data to serve as backbone for 454-pyrosequencing data is only limited, because the recovery of phylotypes extremely differs between the two methods, especially when focusing on picoplankton.

In conclusion, we have shown that our understanding of protist diversity and structure assessed with molecular methods varies strongly depending on the molecular method used. Furthermore, we suggest that comparisons of new 454-pyrosequencing data with previously published clone library data of protist diversity have to be handled with care and should not be over interpreted. Moreover, the cloning approach seems not to be adequate in general to resolve the abundant biosphere and appears to be of very limited suitability as backbone for a refined phylogenetic analysis of OTUs identified by 454-pyrosequencing data.
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 cruises. We are very grateful to S. Frickenhaus, F. Kilpert and B. Beszteri for their bioinformatical support. We also want to thank A. Schroer, A. Nicolaus and K. Oetjen for excellent technical support in the laboratory and Steven Holland for providing access to the program Analytic Rarefaction 1.3.
References


### Table 1 Summary of recovered clones and 454-pyrosequencing reads.

<table>
<thead>
<tr>
<th></th>
<th>HG1</th>
<th>HG4</th>
<th>HGN4</th>
<th>HGS3</th>
<th>ANT25</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clone library:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High quality clones</td>
<td>175</td>
<td>179</td>
<td>101</td>
<td>139</td>
<td>104</td>
</tr>
<tr>
<td>Phylotypes</td>
<td>16</td>
<td>7</td>
<td>13</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td><strong>454-pyrosequencing:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total reads</td>
<td>9830</td>
<td>7539</td>
<td>7938</td>
<td>8786</td>
<td>45772</td>
</tr>
<tr>
<td>High quality reads</td>
<td>8154</td>
<td>5434</td>
<td>5220</td>
<td>7020</td>
<td>30561</td>
</tr>
<tr>
<td>OTUs (97 %)</td>
<td>754</td>
<td>709</td>
<td>829</td>
<td>1014</td>
<td>1153</td>
</tr>
</tbody>
</table>
Table 2 Phylogenetic affiliations of the Arctic clone phylotypes and their relative abundance in the libraries and the 454-pyrosequencing data set.

<table>
<thead>
<tr>
<th>Phylotype</th>
<th>Closest match (Maximum identity %)</th>
<th>Taxonomic group</th>
<th>HG1 (%)</th>
<th>HG4 (%)</th>
<th>HGN4 (%)</th>
<th>HGS3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARK_1</td>
<td>Bolidomonas pacifica (92)</td>
<td>Stramenopiles</td>
<td>1.1 / 2.8</td>
<td>- / 0.7</td>
<td>1.0 / 0.2</td>
<td>- / 0.5</td>
</tr>
<tr>
<td>ARK_2</td>
<td>Clone EU793918.1</td>
<td>Dinoflagellates</td>
<td>0.6 / 0.8</td>
<td>- / 0.5</td>
<td>- / 0.5</td>
<td>- / 2.6</td>
</tr>
<tr>
<td>ARK_3</td>
<td>Clone HM135092.1 (98)</td>
<td>Dinoflagellates</td>
<td>0.6 / 0.4</td>
<td>- / 0.4</td>
<td>4.0 / 0.9</td>
<td>7.2 / 0.2</td>
</tr>
<tr>
<td>ARK_4</td>
<td>Clone JF791003.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>5.0 / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_5</td>
<td>Clone EU819790.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / 0.24</td>
<td>1.1 / 0.2</td>
<td>3.0 / 0.5</td>
<td>1.4 / 1.7</td>
</tr>
<tr>
<td>ARK_6</td>
<td>Micromonas pusilla (99)</td>
<td>Chlorophytes</td>
<td>2.9 / 1.4</td>
<td>- / 1.9</td>
<td>- / 0.3</td>
<td>- / 1.3</td>
</tr>
<tr>
<td>ARK_11</td>
<td>Clone HQ438132.1 (94)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>8.9 / 0.7</td>
<td>- / 0.1</td>
</tr>
<tr>
<td>ARK_12</td>
<td>Syndiniales EU793925.1 (95)</td>
<td>Dinoflagellates</td>
<td>- / 0.2</td>
<td>- / 0.1</td>
<td>- / 0.8</td>
<td>26.6 / 0.6</td>
</tr>
<tr>
<td>ARK_13</td>
<td>Gyrodinium AB120001.1</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>4.0 / -</td>
<td>5.0 / -</td>
</tr>
<tr>
<td>ARK_14</td>
<td>Geminigera cryophila (99)</td>
<td>Cryptophytes</td>
<td>0.6 / 0.4</td>
<td>- / 0.1</td>
<td>- / -</td>
<td>23.7 / 0.1</td>
</tr>
<tr>
<td>ARK_15</td>
<td>Micromonas pusilla (99)</td>
<td>Chlorophytes</td>
<td>77.1 / 14.4</td>
<td>65.9 / 1.1</td>
<td>47.5 / 0.4</td>
<td>- / 1.1</td>
</tr>
<tr>
<td>ARK_16</td>
<td>Clone AY295399.1 (91)</td>
<td>Ciliates</td>
<td>8.0 / 1.2</td>
<td>- / 0.8</td>
<td>- / 0.4</td>
<td>0.7 / 0.2</td>
</tr>
<tr>
<td>ARK_17</td>
<td>Clone EU682572.1 (97)</td>
<td>Ciliates</td>
<td>1.7 / 0.3</td>
<td>- / 0.3</td>
<td>- / 0.2</td>
<td>- / 0.1</td>
</tr>
<tr>
<td>ARK_20</td>
<td>Clone HQ43812.9 (98)</td>
<td>Dinoflagellates</td>
<td>1.1 / -</td>
<td>- / 0.0</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_21</td>
<td>Clone JN934892.1 (95)</td>
<td>Picobiliphytes</td>
<td>1.1 / 0.0</td>
<td>- / 0.0</td>
<td>- / 0.0</td>
<td>- / 0.0</td>
</tr>
<tr>
<td>ARK_25</td>
<td>Gyrodinium sp. (98)</td>
<td>Dinoflagellates</td>
<td>0.6 / 0.4</td>
<td>- / -</td>
<td>- / 0.1</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_26</td>
<td>Woloszynskia sp. (99)</td>
<td>Dinoflagellates</td>
<td>0.6 / 0.4</td>
<td>- / 0.0</td>
<td>- / 0.1</td>
<td>- / 0.2</td>
</tr>
<tr>
<td>ARK_29</td>
<td>Micromonas pusilla (99)</td>
<td>Chlorophytes</td>
<td>0.6 / 3.0</td>
<td>0.6 / 1.1</td>
<td>- / 0.1</td>
<td>- / 0.6</td>
</tr>
<tr>
<td>ARK_30</td>
<td>Clone HQ222463.1 (98)</td>
<td>Picobiliphytes</td>
<td>0.6 / -</td>
<td>- / 0.0</td>
<td>- / -</td>
<td>- / 0.0</td>
</tr>
<tr>
<td>ARK_31</td>
<td>Micromonas pusilla (99)</td>
<td>Chlorophytes</td>
<td>2.3 / -</td>
<td>- / -</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_33</td>
<td>Clone AJ420693.1 (96)</td>
<td>Rhodophytes</td>
<td>0.6 / -</td>
<td>- / -</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_38</td>
<td>Clone EU682636.1 (97)</td>
<td>Chlorophytes</td>
<td>- / -</td>
<td>1.7 / -</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_46</td>
<td>Micromonas pusilla (91)</td>
<td>Chlorophytes</td>
<td>- / -</td>
<td>0.6 / -</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_47</td>
<td>Syndiniales EU793375.1 (90)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>29.6 / -</td>
<td>- / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_58</td>
<td>Clone EU793946.1 (88)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / 0.1</td>
<td>10.9 / -</td>
<td>0.7 / -</td>
</tr>
<tr>
<td>ARK_60</td>
<td>Clone EU793957.12 (92)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>2.0 / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_62</td>
<td>Clone EU682577.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>6.9 / -</td>
<td>- / -</td>
</tr>
<tr>
<td>ARK_68</td>
<td>Clone EF172940.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / 0.1</td>
<td>4.0 / 0.1</td>
<td>- / 0.0</td>
</tr>
<tr>
<td>ARK_69</td>
<td>Clone JF826365.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / 0.1</td>
<td>- / 0.3</td>
<td>2.0 / 0.1</td>
<td>- / 0.1</td>
</tr>
<tr>
<td>ARK_70</td>
<td>Clone HQ438143.1 (95)</td>
<td>Dinoflagellates</td>
<td>- / 0.0</td>
<td>- / 0.1</td>
<td>1.0 / 0.1</td>
<td>1.4 / 0.1</td>
</tr>
<tr>
<td>ARK_72</td>
<td>Clone EU793201.1 (98)</td>
<td>Dinoflagellates</td>
<td>- / -</td>
<td>- / -</td>
<td>- / -</td>
<td>0.7 / -</td>
</tr>
<tr>
<td>ARK_76</td>
<td>Clone EU793383.1 (90)</td>
<td>Dinoflagellates</td>
<td>- / 0.1</td>
<td>- / 0.1</td>
<td>- / -</td>
<td>2.2 / 0.2</td>
</tr>
<tr>
<td>ARK_78</td>
<td>Clone EF195735.1 (90)</td>
<td>Cryptophytes</td>
<td>- / 0.3</td>
<td>- / 0.3</td>
<td>- / 0.2</td>
<td>0.7 / 0.1</td>
</tr>
</tbody>
</table>
Table 2 continued

<table>
<thead>
<tr>
<th>Phylotype</th>
<th>Closest match</th>
<th>Taxonomic group</th>
<th>Clones (%) / 454 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Maximum identity %)</td>
<td>HG1</td>
<td>HG4</td>
</tr>
<tr>
<td>ARK_38</td>
<td>Clone EU682636.1 (97)</td>
<td>Chlorophytes</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_46</td>
<td>Micromonas pusilla (91)</td>
<td>Chlorophytes</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_47</td>
<td>Syndiniales EU793375.1 (90)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_58</td>
<td>Clone EU793946.1 (88)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_58</td>
<td>Clone EU682577.1 (98)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_60</td>
<td>Clone EU793957.12 (92)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_69</td>
<td>Clone EU793201.1 (91)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_70</td>
<td>Clone HQ438143.1 (95)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_72</td>
<td>Clone EU793201.1 (98)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_76</td>
<td>Clone EU793383.1 (90)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_78</td>
<td>Clone EF195735.1 (90)</td>
<td>Cryptophytes</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_82</td>
<td>Clone EU793221.1 (94)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_83</td>
<td>Clone EU793700.1 (94)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_86</td>
<td>Clone EU793708.1 (92)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_87</td>
<td>Clone HM561117.1 (95)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_90</td>
<td>Clone HQ222399.1 (95)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_91</td>
<td>Clone FJ537539.1 (92)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_92</td>
<td>Bolidomonas pacifica (95)</td>
<td>Stramenopiles</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_93</td>
<td>Bathycoccus prasinos (98)</td>
<td>Chlorophytes</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_97</td>
<td>Clone JF826365.1 (91)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_100</td>
<td>Clone AF290050.2 (95)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_102</td>
<td>Clone GU819971.1 (95)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_103</td>
<td>Clone EU818505.2 (97)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
<tr>
<td>ARK_104</td>
<td>Clone EU793381.1 (96)</td>
<td>Dinoflagellates</td>
<td>-/-</td>
</tr>
</tbody>
</table>
Table 3 Phylogenetic affiliations of the Antarctic clone phylotypes and their relative abundance in the library and the 454-pyrosequencing data set.

<table>
<thead>
<tr>
<th>Phylotype</th>
<th>Closest match (Maximum identity %)</th>
<th>Taxonomic group</th>
<th>Clones (%) / 454 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANT_1</td>
<td>clone SGPX577 (98)</td>
<td>Dinoflagellates</td>
<td>2.9 / -</td>
</tr>
<tr>
<td>ANT_2</td>
<td><em>Gyrodinium fusiforme</em> (99)</td>
<td>Dinoflagellates</td>
<td>2.9 / -</td>
</tr>
<tr>
<td>ANT_3</td>
<td>clone SIF_2C7 (99)</td>
<td>Dinoflagellates</td>
<td>3.9 / &lt;1</td>
</tr>
<tr>
<td>ANT_4</td>
<td>clone B16 (98)</td>
<td>Dinoflagellates</td>
<td>1.0 / &lt;1</td>
</tr>
<tr>
<td>ANT_5</td>
<td>clone SHAX878 (95)</td>
<td>Dinoflagellates</td>
<td>2.9 / &lt;1</td>
</tr>
<tr>
<td>ANT_6</td>
<td>clone CNCIII51_20 (99)</td>
<td>Dinoflagellates</td>
<td>20.2 / 8.7</td>
</tr>
<tr>
<td>ANT_7</td>
<td><em>Azadinium spinosum</em> (99)</td>
<td>Dinoflagellates</td>
<td>7.7 / 1.0</td>
</tr>
<tr>
<td>ANT_8</td>
<td><em>Gyrodinium rubrum</em> (96)</td>
<td>Dinoflagellates</td>
<td>1.0 / &lt;1</td>
</tr>
<tr>
<td>ANT_9</td>
<td>DH147-EKD20 (94)</td>
<td>Syndiniales</td>
<td>1.9 / &lt;1</td>
</tr>
<tr>
<td>ANT_10</td>
<td><em>Salpingella acuminata</em> (99)</td>
<td>Ciliates</td>
<td>4.8 / 3.9</td>
</tr>
<tr>
<td>ANT_11</td>
<td>clone KRL01E30 (87)</td>
<td>Picobiliphytes</td>
<td>2.9 / &lt;1</td>
</tr>
<tr>
<td>ANT_12</td>
<td><em>Geminigera cryophila</em> (99)</td>
<td>Cryptophytes</td>
<td>1.0 / &lt;1</td>
</tr>
<tr>
<td>ANT_13</td>
<td>clone B1 (99)</td>
<td>Haptophytes</td>
<td>21.2 / 19.9</td>
</tr>
<tr>
<td>ANT_14</td>
<td>clone B1 (99)</td>
<td>Haptophytes</td>
<td>18.3 / 4.3</td>
</tr>
<tr>
<td>ANT_15</td>
<td>clone F11N10 (91)</td>
<td>Diatoms</td>
<td>1.0 / &lt;1</td>
</tr>
<tr>
<td>ANT_16</td>
<td><em>Hemiaulus sinensis</em> (96)</td>
<td>Diatoms</td>
<td>1.0 / &lt;1</td>
</tr>
<tr>
<td>ANT_17</td>
<td>clone RA070625T.073 (96)</td>
<td>Stramenopiles</td>
<td>2.9 / -</td>
</tr>
<tr>
<td>ANT_18</td>
<td>clone CNCIII05_73 (93)</td>
<td>Stramenopiles</td>
<td>1.9 / -</td>
</tr>
<tr>
<td>ANT_19</td>
<td>clone 14H3Te6QW (95)</td>
<td>Stramenopiles</td>
<td>1.0 / &lt;1</td>
</tr>
</tbody>
</table>
Figure captions

Fig. 1 Map of the sampling stations located within A) the long-term observatory “Hausgarten” (Fram Strait, Arctic) and B) the Southern Ocean.

Fig. 2 Rarefaction curves of A) clone libraries and B) 454-pyrosequencing.

Fig. 3 ARK24: Phylogenetic tree based on the 18S rRNA gene sequences retrieved from the clone libraries and rooted with *Staurosira* sp. Calculation of the tree has been performed with maximum likelihood under the implementation of the Juke-Cantor model and 1000 bootstraps replications. The symbols are standing for the respective station, where the clone had been found. The triangle stands for HG1, the circle for HGS3, the rhombus for HGN4 and the square for HG4.

Fig. 4 ANT25: Phylogenetic tree based on the 18S rRNA gene sequences retrieved from the clone libraries and rooted with *Micromonas pusilla*. Calculation of the tree has been performed with maximum likelihood under the implementation of the Juke-Cantor model and 1000 bootstraps replications.
Figures

Fig. 1
Fig. 2

A

OTUs

Number of clones sampled

HG1
HG4
HGN4
HGS3
ANT25

B

OTUs

Number of reads sampled
Fig. 3
Fig. 4
4.4 Manuscript II

Oceanographic fronts in the Southern Ocean determine biogeographic differences in eukaryotic protist communities – new insights based on 454-pyrosequencing

Christian Wolf\textsuperscript{1}, Stephan Frickenhaus\textsuperscript{1}, Estelle S. Kilias\textsuperscript{1}, Mirko Lunau\textsuperscript{1,2}, Ilka Peeken\textsuperscript{1,3} and Katja Metfies\textsuperscript{1}

\textsuperscript{1} Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
\textsuperscript{2} Leibniz Institute for Baltic Sea Research, Seestraße 15, 18119 Rostock, Germany
\textsuperscript{3} MARUM – Center for Marine Environmental Sciences, University of Bremen, Leobener Straße, 28359 Bremen, Germany

Polar Biology, submitted
Abstract

We determined the composition and biogeography of late summer eukaryotic protist assemblages along a transect from the coast of New Zealand to the eastern Ross Sea. We used state of the art molecular approaches, such as automated ribosomal intergenic spacer analysis (ARISA) and 454-pyrosequencing, combined with flow cytometry and pigment measurements via high performance liquid chromatography (HPLC) to study the protist assemblage. We found distinct biogeographic patterns defined by the oceanographic fronts in the particular region. Different water masses harboured different microbial communities, and environmental gradients, especially temperature, limited their dispersal. In contrast to the Arctic Ocean, picoeukaryotes had minor importance throughout the investigated transect and were nearly absent south of the Polar Front. Dinoflagellates, Syndiniales, and small stramenopiles were dominating the Subantarctic Zone, whereas the importance of diatoms increased southwards, in the Polar Frontal Zone, Antarctic Zone and Subpolar Region. South of the Polar Front, haptophytes were the dominating group. This study delivers a comprehensive and taxon detailed overview of the eukaryotic protist composition in the investigated area during the austral summer.

Keywords
Phytoplankton, Microbial diversity, ARISA, Next-generation-sequencing, HPLC, Flow cytometry
**Introduction**

It is generally accepted that microorganisms are essential for the functioning of marine ecosystems (Azam and Malfatti 2007). Protists are the largest contributors to organic matter in the ocean and they constitute an important determinant of the structure and efficiency of Antarctic marine food webs and the flux of particles to the deep ocean (Priddle et al. 1992). However, the extent of their diversity, in particular in polar oceans, remains unclear (Pedros-Alio 2006). The Southern Ocean is predicted to experience changes in temperature, stratification, mixed-layer depth, and acidity, which will affect light climate and nutrient availability (Davidson et al. 2010). These effects will affect the composition and trophodynamics of the Antarctic marine ecosystem (Boyd 2002; Orr et al. 2005; Tortell et al. 2008). Detecting changes in microbial communities will not be possible without information about the current distributions and abundances of protists (Davidson et al. 2010).

The influence of hydrography on microbial communities in polar oceans recently attracted more interest. There are several studies from the Arctic Ocean (Galand et al. 2009; Hamilton et al. 2008; Lovejoy and Potvin 2011) and the Southern Ocean (Diez et al. 2004), addressing the question how hydrography is shaping microbial distribution. Diez et al. (2004) e.g. showed in the Atlantic sector of the Southern Ocean that picoeukaryotic assemblages were characteristic of each water mass.

Recent investigations of eukaryotic protist composition and distribution in the Southern Ocean mainly used traditional microscopic (Kopczynska et al. 2007) or pigment extraction based (Peeken 1997) methods or a combination of both (Ishikawa et al. 2002; Wright et al. 2009; Ning et al. 1996). However, microscopic approaches have difficulties in identifying small cells (<10 µm) and thus, this fraction is understudied. Nevertheless, latest studies in the North Sea and the Arctic Ocean report that the smallest size fraction is expected to become more important in marine ecosystems in the future, eventually due to climate change (Daufresne et al. 2009; Li et al. 2009). The introduction of high performance liquid chromatography (HPLC) and flow cytometry into marine biology complemented microscopic and simple biomass investigations of marine phytoplankton. Analyzing group-specific marker pigments, using HPLC analysis, allows determining various size classes within the phytoplankton, including pico- and nanoplanckton (Peeken 1997; Wright et al. 2009; Wulff and Wangberg 2004). Furthermore, the optical analysis of plankton using flow cytometry provides a number of unique features, like e.g. potentially near real time simultaneous determination of size, pigment type (chlorophyll a, chlorophyll b, phycobillines) and the
amount of autofluorescence on a single cell level (Czechowska et al. 2008; Wang et al. 2010) in parallel to the quantitative protist assessment. However, flow cytometry and the extraction of pigments from filtered samples followed by HPLC have a low taxonomical resolution, the results critically depend on the algorithm used for the analysis and the target organisms are limited to autotrophs. Here, molecular tools are advantageous to get high-resolution deep taxonomical information. Molecular techniques deliver detailed information on protist diversity regardless of cell sizes.

So far, there are only a few investigations in the Southern Ocean using molecular methods, such as denaturant gradient gel electrophoresis (Diez et al. 2004) and 18S ribosomal ribonucleic acid (rRNA) gene cloning and sequencing (Diez et al. 2001; Lopez-Garcia et al. 2001). The ARISA (automated ribosomal intergenic spacer analysis) approach, providing a quick overview of the diversity and facilitates the comparison of different samples, is well established for investigations of prokaryotic diversity (Danovaro et al. 2009; Smith et al. 2010). However, to our knowledge, this approach has not yet been applied for diversity studies of eukaryotic phytoplankton. The majority of published molecular studies concentrated on cloning and sequencing of the 18S rRNA gene to elucidate the diversity of protists in the Southern Ocean (Diez et al. 2001; Lopez-Garcia et al. 2001). Bent and Forney (2008) noted that the community diversity may be underestimated due to the limited throughput of this method. During the past decade 454-pyrosequencing (Margulies et al. 2005) has been established as a high throughput sequencing methodology and is replacing the Sanger sequencing for comparative metagenomics (Kunin et al. 2010). This next generation high throughput sequencing approach allows assessing microbial communities with high resolution, based on sufficient deep taxon sampling (Margulies et al. 2005; Stoeck et al. 2010). The 454-pyrosequencing method has led to the conclusion that the rare biosphere (phylotypes with an abundance <1% of total sequences) is more diverse and larger than formerly estimated (Sogin et al. 2006).

The objective of this study is to determine the composition and biogeography of late austral summer eukaryotic protist assemblages, including the picoplankton fraction (cells <2 μm), on a transect starting in the north at the coast of New Zealand and ending in the south in the eastern Ross Sea. This area consists of very distinct oceanic regions, separated by oceanic fronts. We combined flow cytometry and pigment measurements (HPLC) with state of the art molecular approaches, such as ARISA and 454-pyrosequencing. The flow cytometry, pigment, and ARISA approach provide an overview of differences in structure and diversity of the entire investigated area. The 454-pyrosequencing of selected representative samples
gives detailed information about the species composition, dominant representatives and the distribution of phylotypes in the observed area.
Materials & Methods

Location and sampling

A total of 31 surface water samples were taken on a regular basis (approx. every 100 km) during the RV Polarstern cruise ANT XXVI/3 between 1st and 12th February 2010 in the Pacific sector of the Southern Ocean (Fig. 1) along a north-south transect from the coast of New Zealand to the eastern Ross Sea (within 45.57°S to 72.38°S latitude and 179.85°E to 164.15°W longitude). Precise locations and physical conditions are shown in table 1. The cruise crossed the dominant oceanographic fronts and involved the major oceanic regions of the Southern Ocean (Fig. 1). The locations of the fronts, including the Subantarctic Front (SAF), the Polar Front (PF), the Southern Antarctic Circumpolar Current Front (SACCF) and the Southern Boundary of the ACC (SB), are taken from Orsi et al. (1995). The zone between the Subtropical Front (STF) and the SAF is stated as the Subantarctic Zone (SAZ). The zone between the SAF and the PF is termed as Polar Frontal Zone (PFZ). The Antarctic Zone (AZ) is arranged between the PF and the SACCF. The zone between the SACCF and the SB is referred to as the Southern Zone (SZ). South of the SB, the Subpolar Region (SR) is finishing the oceanic regions (Orsi et al. 1995; Pollard et al. 2002). Due to the coincidence of the SACCF and the SB in the investigated area, the SACCF and the Southern Zone (SZ) will be neglected in the following.

All water samples were collected using the ship’s pumping system (membrane pump) located at the bow at 8 m below the surface. For the flow cytometry analysis the samples were filled into 5 ml cryovials, preserved with glutaraldehyde (Sigma, 0.1% final), and stored at -80°C until analysis in the laboratory. For the determination of pigments, four-litre-samples were immediately filtered onto 25 mm Whatman GF/F filters and stored at -80°C until further analysis in the laboratory. For molecular analysis, two-litre-samples were immediately fractionated by filtering them on Isopore Membrane Filters (Millipore, USA) with a pore size of 10 µm, 3 µm and 0.2 µm. Filters were stored at -80°C until analysis in the laboratory.

Flow cytometry

An Accuri C6 flow cytometer (BD Biosciences, USA) was used to measure abundance, size (forward scatter signal was calibrated with the Invitrogen size calibration kit F-13838), as well as red (FL3) and orange (FL2) autofluorescence after excitation with a blue (488 nm) and
a red (620 nm) laser. Each sample was analyzed for 3 min at fast flow rate (66 µl min⁻¹). Phytoplankton cells were separated from detritus using scatter plots of FL2 vs. FL3, followed by group identification based on their specific signature in forward scatter (FSC) vs. FL4 (far red). Polychromatic latex beads (1 µm, Polysciences) were added to every sample to monitor the performance of the optical system and to double-check the internal volume calibration of the instrument.

The amount of total FL3 (Chl a) fluorescence was calculated from the raw data (FL3 x abundance) and used as biomass proxy, while the ratio of FL3/FL2 served as an indicator for shifts in the plankton community and physiology, respectively (Shalapyonok et al. 2001).

Pigment analysis (HPLC)

Samples were measured using a Waters HPLC-system, equipped with an auto sampler (717 plus), pump (600), photo diode array (2996), a fluorescence detector (2475) and EMPOWER software. For analytical preparation, 50 µl internal standard (canthaxanthin) and 1.5 ml acetone were added to each filter sample and then homogenized for 20 sec in a Precellys® tissue homogenizer. After centrifugation, the supernatant liquid was filtered through a 0.2 µm polytetrafluoroethylene (PTFE) filter (Rotilabo) and placed in Eppendorf cups. An aliquot (100 µl) was transferred to the auto sampler (4°C). Just prior to analysis, the sample was premixed with 1 M ammonium acetate solution in the ratio 1:1 (v/v) in the auto sampler and injected onto the HPLC-system. The pigments were analyzed by reverse-phase HPLC, using a VARIAN Microsorb-MV3 C8 column (4.6x100 mm) and HPLC-grade solvents (Merck). Solvent A consisted of 70% methanol and 30% 1 mol L⁻¹ ammonium acetate and solvent B contained 100% methanol. The gradient was modified after Barlow et al. (1997). Eluting pigments were detected by absorbance (440 nm) and fluorescence (Ex: 410 nm, Em: >600 nm). Pigments were identified by comparing their retention times with those of pure standards and algal extracts. Additional confirmation for each pigment was done with representative samples using on-line diode array absorbance spectra between 390-750 nm. Pigment concentrations were quantified based on peak areas of external standards, which were spectrophotometrically calibrated using extinction coefficients published by Bidigare (1991) and Jeffrey et al. (1997). For correction of experimental losses and volume changes, the concentrations of the pigments were normalized to the internal standard canthaxanthin. Phytoplankton group composition was calculated applying the Chemtax program and input ratios of Mackey et al. (1996). For the comparison with the flow cytometry data, the
following groups were combined: prasinophytes and pelagophytes for picoplankton (<2 µm), haptophytes and cryptophytes for nanoplankton (2-20 µm), and dinoflagellates and diatoms for microplankton (20-200 µm). Their respective contribution to total biomass is based on their Chemtax derived chlorophyll $a$ concentration.

DNA extraction

The DNA was extracted with the E.Z.N.A.TM SP Plant DNA Kit (Omega Bio-Tek, USA). At the beginning, the filters were incubated with lysis buffer. All further steps were performed as described in the manufacturer’s instructions. At the end, the DNA was eluted in 60 µl of elution buffer and the extracts were stored at -20°C until further analysis. DNA concentration was measured with a NanoDrop 1000 (Thermo Fisher Scientific, USA) (average DNA concentration: 24 ng µl$^{-1}$).

PCR amplification, ARISA

An equal volume of extracted DNA of each size fraction (>10 µm, 3-10 µm and 0.2-3 µm) from each sample was pooled. The ITS1 (internal transcribed spacer) region was amplified in triplicates with the primers 1528F (5’-GTA GGT GAA CCT GCA GAA GGA TCA-3’) (modified after Medlin et al. (1988)) and ITS2 (White et al. 1990). The 1528F primer was labelled with the dye 6-FAM (6-Carboxyfluorescein). The PCR mixtures contained 1 µl of DNA extract, 1 x HotMaster Taq Buffer containing 2.5 mmol L$^{-1}$ Mg$^{2+}$ (5 Prime, USA), 0.8 mmol L$^{-1}$ deoxynucleoside triphosphate (dNTP) mix (Eppendorf, Germany), 0.2 µmol L$^{-1}$ 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: an initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 45 sec, 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).

PCR amplification, 454-pyrosequencing

The ARISA approach served as a preliminary analysis for the 454-pyrosequencing and six samples were chosen to be sequenced. For each size fraction of a sample we amplified ~670
Manuscripts

bp fragments of the 18S rRNA gene, containing the highly variable V4-region, using the primer-set 528F (5’-GCG GTA ATT CCA GCT CCA A-3’) and 1055R (5’-ACG GCC ATG CAC CAC CAC CCA T-3’) (modified after Elwood et al. (1985)). The PCR mixtures were composed as described previously for ARISA. Reaction conditions were as following: an initial denaturation at 94°C for 3 min, 30 cycles of denaturation at 94°C for 45 sec, annealing at 59°C for 1 min and extension at 72°C for 3 min, and a final extension at 72°C for 10 min.

An equal volume of PCR reaction of each size fraction from each sample was pooled and purified with the MinElute PCR purification kit (Qiagen, Germany) following the manufacturer’s instructions. Pyrosequencing was performed on a Genome Sequencer FLX system (Roche, Germany) by GATC Biotech AG (Germany).

Data analysis, ARISA

Electropherograms were analyzed using the GeneMapper Software v4.0 (Applied Biosystems, USA). Peaks with a size smaller than 50 bp (corresponding to primer and primer-dimer peaks) were removed from the data set. To remove the background noise and to get sample-by-binned-operational-taxonomic-unit tables, the data were binned using the binning scripts, according to Ramette (2009), for R (R Development Core Team 2008). The resulting sample-by-binned-operational-taxonomic-unit tables were transformed into presence/absence matrices and the distances between the samples were calculated, using the Jaccard index implemented in the R package Vegan (Oksanen et al. 2011), which was also used in the following steps. The clustering of the samples was determined using the hclust function in R. The resulting clusters were tested for significant differences using the ANOSIM test. Multidimensional scaling (MDS) plots (maximum random starts of 1,000) were computed. The correlation between the ARISA distance matrix and the environmental distance matrix (temperature and salinity) was tested with a Mantel test (10,000 permutations), implemented in the R package ade4 (Dray and Dufour 2007). A PCA with the environmental parameters was performed (R package ade4).

Data analysis, 454-pyrosequencing

Raw sequence reads were processed to obtain high quality reads. The forward primer 528F, used for the sequencing, attaches approx. 25 bp upstream of the V4 region, which has in general a length of approx. 230 bp (Nickrent and Sargent 1991). Reads with a length under
300 bp were excluded from further analysis to assure the including of the complete hypervariable V4 region in the analysis and to get rid of the short reads. Unusually long reads, that were greater than the expected amplicon size (>670 bp) and reads with more than one uncertain base (N) were removed. Remaining reads were checked for chimeric sequences with the software UCHIME 4.2 (Edgar et al. 2011) and all reads considered being chimeric were excluded from further analysis. The high quality reads of all samples were clustered into operational taxonomic units (OTUs) at the 97% similarity level using the software Lasergene 10 (DNASTAR, USA). Subsequently, reads not starting with the forward primer were manually removed. Consensus sequences of each OTU were generated, which reduced the amount of sequences to operate with and attenuated the influence of sequencing errors and uncertain bases. The 97% similarity level has shown to be the most suitable to reproduce original eukaryotic diversity (Behnke et al. 2011) and has also the effect of bracing most of the sequencing errors (Kunin et al. 2010). Furthermore, known intragenomic SSU polymorphism levels can range to 2.9% in dinoflagellate species (Miranda et al. 2012). OTUs comprised of only one sequence (singletons) were removed. The consensus sequences were aligned using the software HMMER 2.3.2 (Eddy 2011). Subsequently, taxonomical affiliation was determined by placing the consensus sequences into a reference tree, containing about 1,200 high quality sequences of Eukarya from the SILVA reference database (SSU Ref 108), using the software pplacer 1.0 (Matsen et al. 2010). The compiled reference database is available on request in ARB-format. OTUs assigned to fungi and metazoans were excluded from further analysis. Rarefaction curves were computed using the freeware program Analytic Rarefaction 1.3. The data set generated in this study has been deposited at GenBank’s Short Read Archive (SRA) under Accession No. SRA056811.
Results

Environmental conditions

Along the investigated transect we crossed three fronts, the SAF, the PF and the SB. The four oceanic regions showed different environmental conditions, especially in terms of temperature (Table 1). Surface water temperatures of 8.8-16 °C and salinities of over 34 psu characterized the SAZ. The PFZ featured surface water temperatures of 3.1-6.9°C and salinities of 33.6-34 psu. The surface water temperatures in the AZ ranged between 0.7-2.4°C and the salinities between 33.4-33.5 psu. In the SR the surface water temperatures did not exceed 0.5°C and dropped to -1.3°C at the most southern sample station. Salinities in the SR ranged between 33.4-33.8 psu. No sea ice was present throughout the entire transect.

Structure/diversity overview

We used a combination of flow cytometry, HPLC and ARISA to identify the structure/diversity of the plankton assemblages in the investigated area. The observed patterns enabled us to select representative samples for subsequent 454-pyrosequencing.

Flow cytometry – The four investigated oceanic regions (SAZ, PFZ, AZ and SR) (Fig. 1) were well represented by the biological data (Fig. 2). The amplitude of the FL3/FL2 signal explicitly changed around the oceanic fronts (SAF, PF, SB), indicating a shift in the plankton community in these areas. Within the different oceanic regions the signal roughly ranged in the same order. The calculated biomass (0.2-50 μm fraction) roughly showed a pattern according to the oceanic regions. The highest biomass values occurred in the first two samples and in the AZ. Between the SAF and the PF, the biomass showed the lowest magnitude throughout the whole data set.

HPLC – Total Chl a concentrations showed a pattern according to the oceanographic fronts and oceanic regions (Fig. 3). The amplitude of the Chl a concentration explicitly changed in the area of the oceanic fronts (SAF, PF, and SB). The concentrations ranged between 0.1 and 3.8 μg L⁻¹ and were mostly greater than 0.5 μg L⁻¹. The average concentrations in the different regions were 1.05 ± 0.94 μg L⁻¹ (SAZ), 0.18 ± 0.06 μg L⁻¹ (PFZ), 1 ± 0.02 μg L⁻¹ (AZ) and 1.43 ± 0.87 μg L⁻¹ (SR). The highest concentrations were measured in sample 1 and 28 with 3.4 and 3.8 μg L⁻¹, respectively. The concentrations were lowest in the PFZ with values not exceeding 0.2 μg L⁻¹. The contribution of the different size
classes to total Chl \(a\) showed a pattern according to the oceanic regions. The biggest differences could be seen between the region north and south of the PF. North of the PF picoeukaryotes contributed 8.92\% ± 4.4\% on average to total Chl \(a\), whereas south of the PF they only accounted for 1.3\% ± 1.5\% on average of total Chl \(a\). Nanoeukaryotes were the dominating size class north of the PF. Here, they were contributing 57.7\% ± 10.5\% on average to total Chl \(a\). South of the PF nanoeukaryotes accounted for 38.7\% ± 13.4\% on average. The average contribution to total Chl \(a\) of microeukaryotes was 33.4\% ± 11.6\% north of the PF and 60\% ± 11.9\% south of the PF, where they were the dominant size fraction.

ARISA – The fragment length analysis of the ITS1 region of all 31 surface water samples resulted in 102 different fragments with a length of 50 to 432 bp, of which 15 only occurred in one sample (unique fragments). On average, the number of fragments of each sample was 39 and ranged between 27 (sample 31) and 52 (sample 2). The ordination analysis based on the ARISA profiles (Fig. 4) clustered the samples in three groups according to the two main fronts, the SAF and the PF (Fig. 1). On closer inspection, the three groups suit very well to the oceanic regions, the SAZ, the PFZ and a combination of AZ and SR (AZ/SR). The three groups show significantly different ARISA profiles (ANOSIM, \(R = 0.348, p = 0.001\)). The ARISA profiles distances are significantly correlated with the distances of the environmental condition profiles (Mantel test, \(r = 0.187, p = 0.008\)). Figure 5 shows the PCA of the environmental conditions with the three ARISA groups plotted. The two axes are explaining 96\% of the total variance. The three groups are characterized by different temperature and salinity regimes. Temperature and salinity values are decreasing from group A to C, as indicated by the corresponding axis.

Detailed community composition

Based on the pigments and ARISA analysis, one representative sample of each cluster/region was chosen for 454-pyrosequencing (sample 6 for the SAZ, sample 16 for the PFZ and sample 25 for the AZ/SR). Following the size of the clusters, one (sample 8) and two (sample 22 and 26) additional samples were sequenced for the SAZ and the AZ/SR, respectively, to strengthen the conclusions derived from the 454-pyrosequencing data (see below). Thus, in total six samples were sequenced.

454-pyrosequencing – The summary of recovered 454-pyrosequencing reads and the quality filtering can be seen in table 2. A total of 300,582 sequence reads were retrieved from 454-pyrosequencing, of which 76.4\% had an acceptable length (300-670 bp). After the quality
filtering, 58% of the total reads were left over for analysis. The number of analysed reads ranged between roughly 10,000 (sample 16) to over 50,000 (sample 22 and 26). Subsequent to the clustering, 3,841 different OTUs were generated. The number of OTUs (Table 2) for each sample ranged between 630 (sample 16) and 1,695 (sample 22), at which only 0.4% (sample 22) to 2.4% (sample 16) were abundant (number of reads ≥1% of total reads).

Rarefaction curves (Fig. 6) show that none of the samples has been sampled to saturation. However, the trend of the curves suggests that sample 16 harboured the lowest diversity.

An overview of the distribution of the OTUs across the three investigated regions can be seen in figure 7. Only 5% of all 3,841 OTUs were present in all three regions, 19% were present in two regions and the majority of 76% were present in only one region (unique). The proportion of unique OTUs in the SAZ, PFZ and AZ/SR was 53.8%, 13.5% and 71.1%, respectively.

The placement of all OTUs resulted in an assignment of all reads to major protist groups. The relative abundance of sequences assigned to these groups is shown in figure 8. The read abundance of haptophytes increased southwards from 12.3% (sample 6) to 46.8% (sample 22). In the AZ/SR, they were dominating the sequence assemblage. Chlorophytes were only occurring mentionable in the SAZ, where they constituted up to 6.4% of the reads. Pelagophytes showed their maximum read abundance in the PFZ with 4.1%. The read abundance of diatoms increased southwards from 2.7-3.8% in the SAZ to 12.9% in the PFZ and to 10.1-24.3% in the AZ/SR. Labyrinthulids only occurred in significant amounts in the SAZ, where they accounted for 5-13.8% of all reads. The group of marine stramenopiles (MAST) showed the highest read abundance in the PFZ and in sample 25 with 5.2% and 5.1%, respectively. Dinoflagellates were dominating the sequence assemblage in the SAZ, where they accounted for 31.6-58% of the reads. Their read abundance was decreasing southwards to 23.8% in the PFZ and to 30-37.3% in the AZ/SR. North of the PF Syndiniales accounted for 9.2-16.1% of the reads, whereas south of the PF they only accounted for 0.9-1.7%. Ciliates showed a southwards increase in sequence abundance from 0.4% to 4.7%.

Of the 3,841 different OTUs found in all samples, 36 were abundant in at least one sample. These 36 OTUs were present in at least two regions and even 23 were present in all three regions. Figure 9 shows the relative sequence read abundance of the abundant OTUs in all six samples in a colour-coded matrix. Emiliania sp. and Chrysochromulina sp. sequences only occurred north of the PF. Phaeocystis sp. 1 was detected in all samples, whereas the read abundance was increasing southwards and was most dominant in the AZ/SR with 19.15-
40.4% of all sequences. The OTU assigned to *Phaeocystis* sp. 2 was only occurring in the PFZ and AZ/SR. Chlorophytes (*Bathycoccus* sp. and unc. Prasinophyte) were only present in the first two regions. In general, the read abundance of diatoms (*Actinocyclus* sp., *Chaetoceros* sp., *Corethron* sp., *Eucampia* sp., *Pseudo-nitzschia* sp., *Thalassionema* sp. and *Thalassiosira* sp.) was increasing southwards. All diatom OTUs were rare or absent in the SAZ. Pelagophytes (*Pelagomonas* sp. and unc. Pelagophyte) and labyrinthulids (unc. Labyrinthulid 1-5) were only abundant in the SAZ. The unc. MAST OTU only occurred in the PFZ and the AZ/SR. Besides the haptophytes, the dinoflagellates were the numerically dominant group in our dataset across the entire transect. The most dominant OTUs were the *Perkinsus* sp. and the unc. Dinoflagellate 1, which accounted for up to 9.89% and 23.48% of the sequences, respectively. The dinoflagellate OTUs mostly showed the highest read abundance in the SAZ and the PFZ and some were absent in the AZ/SR. Syndiniales (unc. Syndiniales 1-2) only occurred north of the PF. The read abundance of the Ciliate 1 OTU was increasing southwards. The unc. Telonemia 1 OTU was rare, except in the PFZ. The proportion of the rare biosphere ranged between 25.85% (sample 22) and 48.36% (sample 8).
Discussion

Structure/diversity overview

With the combination of flow cytometry, HPLC and ARISA we gained an overview of the protist distribution and assemblage structure across the entire transect. The pattern, found here, is concatenated with the environmental gradients and the different communities are separated by the oceanic fronts in this area.

The biomass values were highly dynamic across the entire transect, but showed less variability within the particular oceanic regions. We found some differences in the amplitude between biomass, gained via flow cytometry, and Chl a, gained via HPLC. In contrast to the highest measured Chl a concentration in sample 28, the biomass amplitude for this sample was relatively low. This could be due to the fact, that the settings of the flow cytometry focus on the pico- and nanoeukaryotes and miss the larger microeukaryotes. The range of the Chl a values we have measured with HPLC in our study is commonly reported from Antarctic waters (Wright et al. 2009; Kozłowski et al. 2011; Ishikawa et al. 2002). In the southern Indian Ocean Chl a concentrations in late spring between 0.2 and 0.4 µg L⁻¹ were measured but in areas with Southern Ocean water influence, they can increase to 1.4 µg L⁻¹ (Schlüter et al. 2011). Higher Chl a concentrations were often connected with high abundances of large diatoms (Detmer and Bathmann 1997; Ishikawa et al. 2002). In our study, higher Chl a concentrations were also generally accompanied with a higher contribution of microeukaryotes (e.g. diatoms). The highest Chl a concentration of 3.8 µg L⁻¹ e.g. was obtained in sample 28 and was accompanied with a microeukaryotic proportion of 70%. The scattered Chl a peaks measured here, illustrate the patchiness of phytoplankton blooms in the Southern Ocean. Despite the generally high nutrient concentrations in waters of the Southern Ocean, micronutrient availability, especially the availability of iron, is considered responsible for the widespread low phytoplankton biomass and punctual occurrence of blooms. Hence, most areas of the Southern Ocean are regarded as high nutrient, low chlorophyll (HNLC) regions and presumably iron-limited (Banse 1996; Debaar et al. 1995; Martin et al. 1990). Areas with higher chlorophyll concentrations might be influenced by micronutrient supply via melting icebergs or dust from the atmosphere. Especially in the PFZ, Chl a concentrations are generally low and stable throughout the year (Banse 1996). This fits to our observation of the lowest biomass and Chl a concentration in that area. One explanation for the low biomass
could be that the surface layer was deeply mixed, which prevented positive photosynthesis and the build-up of biomass.

In temperate, subtropical and tropical open ocean waters of the Atlantic and Indian Ocean picoplankton can contribute 60-90% to total Chl $a$ (Maranon et al. 2001; Not et al. 2008). Even though the majority of picoeukaryotes in these areas are composed by cyanobacteria, picoeukaryotes can contribute up to 54% to total Chl $a$ (Not et al. 2008). In contrast to these data, in our study we found a maximum picoeukaryotic contribution to total Chl $a$ of only 16%. South of the PF picoeukaryotes were nearly absent. Diez et al. (2004) found out that south of the PF cells <5µm can contribute up to 80% to total Chl $a$. However, referring to cells <5µm when investigating picoplankton can be misleading, because Vanucci and Bruni (1998) showed great differences even between biomass estimates of cells <2µm and cells <3µm. The Subantarctic and polar open waters, investigated here, were dominated by nano- and microeukaryotes. These findings are in agreement with other observations of microeukaryotes dominating in ACC waters (Hewes 2009), especially south of the PF (Landry et al. 2001). The low water temperatures can be rejected as explanation for the low abundance of picoeukaryotes, because Tremblay et al. (2009) have described a picophytoplankton contribution to total phytoplankton abundance of 76% in the Canadian High Arctic. One difference between the Canadian High Arctic and the Southern Ocean is the more coastal character of the Arctic with much more freshwater input. In coastal waters high abundances of mamiellophytes, which are often dominating the picoplankton assemblage, have been described (Not et al. 2004). Mamiellophytes are always present and often very abundant in the Arctic (Lovejoy et al. 2006; Lovejoy et al. 2007), which can be an explanation for the higher proportion of picoeukaryotes in this region. Thus, there have to be other factors than temperature, like nutrient availability, competition or mixing of the surface layer, accounting for the low abundance of picoeukaryotes in the Southern Ocean.

We applied molecular tools to gain a deeper insight into the community structure and composition. In contrast to the flow cytometry and HPLC analysis, the ARISA analysis clustered the samples in only three groups. Either the genetic structure of the assemblages in the AZ and the SR has been very similar or the resolution of the fragment length analysis was not sufficient to detect differences. Nevertheless, in terms of genetic structure, the SAF and PF are separating three distinctive communities. Diez et al. (2004) also did a fingerprint analysis (DGGE) of eukaryotic protist assemblages across the main fronts in the Southern Ocean, but they focused only on the picoplankton. However, they determined characteristic
picoeukaryotic assemblages for each water mass in the Atlantic sector of the Southern Ocean and the SAF and PF constituted as separator, confirming our results.

The interpretation of molecular data requires the awareness of potential biases, which can lead to a tampering of the real situation. Primer specificity is one possible bias (Zhu et al. 2005). The amplification of the different species in an environmental sample can vary and the used primers might not capture some species. Zhu et al. (2005) also have described that the number of rRNA gene copies depends on the cell size and varies among eukaryotes from one to several hundreds. Especially the dinoflagellates seem to have more rRNA gene copies than the other taxonomical groups and thus might be overrepresented in molecular data. The ARISA approach can only serve as an approximate overview of the diversity structure, due to the qualitative character and the overlapping of fragment lengths among different species. The placement of sequences gained via 454-pyrosequencing has to be interpreted with care, because the length of only ~500-600 bp is decreasing the robustness of the phylogenetic annotation. Therefore, we generally did not classify the OTUs beyond the genus level.

Nevertheless, in summary, we have shown that the combination of established techniques (flow cytometry and HPLC) and new tools (ARISA) serves as a good approach to characterize the structure/diversity of protist assemblages in a large sample set spanning a huge geographical area and distinguishing between different regions separated by oceanic fronts.

Detailed community composition

We performed 454-pyrosequencing with six representative samples, derived from the ARISA outcome, to offset the low taxonomical resolution of the methods described above. The distribution of OTUs (Fig. 7) indicates that at least the SAZ and AZ/SR represent distinct protist assemblages. The PFZ, sharing most of the OTUs with the other two regions, serves as a transition zone. Hence, we could argue that the “everything is everywhere” hypothesis, which has been debated controversially (Finlay and Fenchel 1999; Foissner 1999; Lachance 2004), does not apply here. However, it has to be kept in mind that the samples were not exhaustively explored (Fig. 6), and definitely conclusions cannot be made.

Nevertheless, we found several strong indications for biogeographic distribution. We observed sequences belonging to the genus *Emiliania* only in the SAZ and PFZ. South of the PF, we detected no sequences appertaining to coccolithophores. Gravalosa et al. (2008) found out, that the abundance of coccolithophores (mainly *Emiliania huxleyi*) in the eastern
Amundsen Sea decreases polewards. They detected no coccolithophores south of 65.50°S, confirming our findings. The *Phaeocystis* sp. 2 OTU, which was only occurring south of the SAF, is another example for biogeographic distribution. In the reference tree, this OTU was placed next to *Phaeocystis jahnii*, whereas the dominating *Phaeocystis* sp. 1 OTU, which also occurred north of the SAF, was placed near *Phaeocystis antarctica*. The occurrence of *Phaeocystis jahnii* south of the PF is stunning, because Zingone et al. (1999) were first describing this species in the Mediterranean Sea, where very different environmental settings can be found. Sequences most similar to *Phaeocystis jahnii* have also been found in the Ross Sea by Gast et al. (2004). It seems that the genus *Phaeocystis* is more diverse than presumed, which already Medlin and Zingone (2007) have illustrated. Furthermore, it is possible that the *Phaeocystis* species are able to cope with a wide range of environmental conditions.

In general, the sequence assemblages north of the SAF were dominated by dinoflagellates. South of the SAF diatoms and haptophytes appear more important. This matches to the findings of Landry et al. (2001), who described an enhanced importance of larger cells and a greater contribution of diatoms south of the PF. Ning et al. (1996) also observed a dominance of diatoms in cold Antarctic waters, and dinoflagellates dominating in the warmer waters north of the SAF. The most obvious reason for this shift in community composition is the nutrient limitation, especially silicon, north of the SAF (Ning et al. 1996). The diatom genera we identified to be among the abundant phylotypes (*Chaetoceros, Corethron, Pseudo-nitzschia, Thalassionema, and Thalassiosira*) were reported previously in open waters between Australia and Antarctica and in the Ross Sea (Gast et al. 2004; Ishikawa et al. 2002; Kopczynska et al. 2007). The genus *Actinocyclus* was also reported before in Antarctic waters, but in coastal regions like the Antarctic Peninsula (Olguin and Alder 2011). Finding comparative data for dinoflagellates and labyrinthulids is practically impossible, because most phylotypes we identified matched with unclassified representatives. The limited occurrence of chlorophytes, which were only among the abundant phylotypes north of the SAF, is congruent to the size class distribution derived from HPLC. The most dominant phylotype south of the PF belongs to the genus *Phaeocystis*. Our observation is supported by other studies (DiTullio et al. 2000; Elsayed et al. 1983), which revealed huge blooms of *Phaeocystis* in the Ross Sea. It seems that *Phaeocystis* cope best with the environmental conditions in the open ocean waters south of the PF, which are characterized by low temperatures and upwelling of nutrient rich deep water (Talley et al. 2011). Only four of the 36 abundant OTUs were abundant in all three regions, suggesting that only few species properly cope with the different environmental settings.
In conclusion, we have shown that each water mass in the studied area of the Southern Ocean harboured characteristic protist assemblages. The rare biosphere did not appear as a background population that contains species, which are abundant under the other environmental conditions. However, the present study does not resolve possible seasonal effects. The most prominent separator in the investigated area was the PF. Dinoflagellates and smaller cells, like labyrinthulids and Syndiniales, were dominating north of the PF, whereas haptophytes and diatoms became more important south of the PF. We also showed that the ARISA approach is very well suited for investigating the structure of eukaryotic phytoplankton assemblages in large numbers of samples, which are needed for assessments of biogeographical differences in microbial communities.
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. We are grateful to F. Kilpert and B. Beszteri for their bioinformatical support. We also want to thank A. Schroer, A. Nicolaus and K. Oetjen for technical support in the laboratory and Steven Holland for providing access to the program Analytic Rarefaction 1.3. We would like to acknowledge E. M. Nöthig and K. Kohls for their insightful discussions and reviews of this manuscript.
References


Manuscripts


Polar Front Zone at 170 degrees W. J Geophys Res-Oceans 106 (C7):13903-13915. doi:10.1029/1999jc000187


### Tables

**Table 1** Location and physical conditions of surface water samples (n.d. = not determined, SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ = Antarctic Zone, SR = Subpolar Region)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Longitude (°W)</th>
<th>Latitude (°S)</th>
<th>Temperature (°C)</th>
<th>Salinity (psu)</th>
<th>Oceanic region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>179.85</td>
<td>-45.57</td>
<td>n.d.</td>
<td>n.d.</td>
<td>SAZ</td>
</tr>
<tr>
<td>2</td>
<td>-179.08</td>
<td>-45.90</td>
<td>16.00</td>
<td>34.20</td>
<td>SAZ</td>
</tr>
<tr>
<td>3</td>
<td>-178.41</td>
<td>-46.56</td>
<td>16.00</td>
<td>34.39</td>
<td>SAZ</td>
</tr>
<tr>
<td>4</td>
<td>-178.41</td>
<td>-48.41</td>
<td>12.07</td>
<td>34.09</td>
<td>SAZ</td>
</tr>
<tr>
<td>5</td>
<td>-178.42</td>
<td>-49.35</td>
<td>12.91</td>
<td>34.09</td>
<td>SAZ</td>
</tr>
<tr>
<td>6</td>
<td>-178.41</td>
<td>-50.05</td>
<td>12.33</td>
<td>34.05</td>
<td>SAZ</td>
</tr>
<tr>
<td>7</td>
<td>-178.41</td>
<td>-51.52</td>
<td>10.88</td>
<td>34.00</td>
<td>SAZ</td>
</tr>
<tr>
<td>8</td>
<td>-178.41</td>
<td>-52.93</td>
<td>12.00</td>
<td>34.19</td>
<td>SAZ</td>
</tr>
<tr>
<td>9</td>
<td>-178.41</td>
<td>-54.14</td>
<td>10.96</td>
<td>34.17</td>
<td>SAZ</td>
</tr>
<tr>
<td>10</td>
<td>-177.01</td>
<td>-55.25</td>
<td>8.75</td>
<td>34.12</td>
<td>SAZ</td>
</tr>
<tr>
<td>11</td>
<td>-174.82</td>
<td>-56.76</td>
<td>6.86</td>
<td>33.98</td>
<td>PFZ</td>
</tr>
<tr>
<td>12</td>
<td>-174.26</td>
<td>-57.21</td>
<td>6.16</td>
<td>33.94</td>
<td>PFZ</td>
</tr>
<tr>
<td>13</td>
<td>-173.37</td>
<td>-58.22</td>
<td>6.12</td>
<td>33.95</td>
<td>PFZ</td>
</tr>
<tr>
<td>14</td>
<td>-172.06</td>
<td>-59.64</td>
<td>4.60</td>
<td>33.82</td>
<td>PFZ</td>
</tr>
<tr>
<td>15</td>
<td>-171.35</td>
<td>-60.38</td>
<td>3.12</td>
<td>33.58</td>
<td>PFZ</td>
</tr>
<tr>
<td>16</td>
<td>-170.93</td>
<td>-60.82</td>
<td>4.64</td>
<td>33.84</td>
<td>PFZ</td>
</tr>
<tr>
<td>17</td>
<td>-170.14</td>
<td>-61.62</td>
<td>2.38</td>
<td>33.41</td>
<td>AZ</td>
</tr>
<tr>
<td>18</td>
<td>-168.43</td>
<td>-63.29</td>
<td>1.27</td>
<td>33.52</td>
<td>AZ</td>
</tr>
<tr>
<td>19</td>
<td>-167.66</td>
<td>-64.05</td>
<td>0.70</td>
<td>33.54</td>
<td>AZ</td>
</tr>
<tr>
<td>20</td>
<td>-166.41</td>
<td>-65.18</td>
<td>0.46</td>
<td>33.82</td>
<td>SR</td>
</tr>
<tr>
<td>21</td>
<td>-166.37</td>
<td>-65.54</td>
<td>0.38</td>
<td>33.75</td>
<td>SR</td>
</tr>
<tr>
<td>22</td>
<td>-166.01</td>
<td>-66.39</td>
<td>0.25</td>
<td>33.81</td>
<td>SR</td>
</tr>
<tr>
<td>23</td>
<td>-165.35</td>
<td>-67.86</td>
<td>0.27</td>
<td>33.65</td>
<td>SR</td>
</tr>
<tr>
<td>24</td>
<td>-165.09</td>
<td>-68.63</td>
<td>-0.01</td>
<td>33.65</td>
<td>SR</td>
</tr>
<tr>
<td>25</td>
<td>-164.88</td>
<td>-69.09</td>
<td>-0.14</td>
<td>33.55</td>
<td>SR</td>
</tr>
<tr>
<td>26</td>
<td>-164.86</td>
<td>-69.38</td>
<td>-0.51</td>
<td>33.54</td>
<td>SR</td>
</tr>
<tr>
<td>27</td>
<td>-164.66</td>
<td>-70.17</td>
<td>-0.52</td>
<td>33.68</td>
<td>SR</td>
</tr>
<tr>
<td>28</td>
<td>-164.57</td>
<td>-70.67</td>
<td>-0.57</td>
<td>33.81</td>
<td>SR</td>
</tr>
<tr>
<td>29</td>
<td>-164.42</td>
<td>-71.32</td>
<td>-0.70</td>
<td>33.72</td>
<td>SR</td>
</tr>
<tr>
<td>30</td>
<td>-164.28</td>
<td>-72.12</td>
<td>-1.09</td>
<td>33.46</td>
<td>SR</td>
</tr>
<tr>
<td>31</td>
<td>-164.15</td>
<td>-72.38</td>
<td>-1.32</td>
<td>33.35</td>
<td>SR</td>
</tr>
</tbody>
</table>
Table 2 Summary of recovered 454-pyrosequencing reads, quality filtering and number of OTUs

<table>
<thead>
<tr>
<th></th>
<th>SAZ 6</th>
<th>SAZ 8</th>
<th>PFZ 16</th>
<th>PFZ 22</th>
<th>AZ/SR 25</th>
<th>AZ/SR 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total 454-reads</td>
<td>25,377</td>
<td>29,383</td>
<td>34,029</td>
<td>86,001</td>
<td>45,772</td>
<td>80,020</td>
</tr>
<tr>
<td>Average length (bp)</td>
<td>366</td>
<td>337</td>
<td>345</td>
<td>364</td>
<td>333</td>
<td>380</td>
</tr>
<tr>
<td>Acceptable length*</td>
<td>19,186</td>
<td>21,593</td>
<td>25,578</td>
<td>65,689</td>
<td>33,356</td>
<td>64,114</td>
</tr>
<tr>
<td>Quality filtering:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than one N</td>
<td>146</td>
<td>228</td>
<td>204</td>
<td>407</td>
<td>107</td>
<td>427</td>
</tr>
<tr>
<td>Chimeras</td>
<td>570</td>
<td>407</td>
<td>253</td>
<td>1,422</td>
<td>990</td>
<td>1,967</td>
</tr>
<tr>
<td>Incorrect forward primer</td>
<td>117</td>
<td>145</td>
<td>357</td>
<td>312</td>
<td>322</td>
<td>327</td>
</tr>
<tr>
<td>Singletons</td>
<td>2,078</td>
<td>1,309</td>
<td>653</td>
<td>5,361</td>
<td>3,498</td>
<td>3,723</td>
</tr>
<tr>
<td>Non-target organisms</td>
<td>1,170</td>
<td>3,039</td>
<td>14,246</td>
<td>4,444</td>
<td>4,023</td>
<td>2,863</td>
</tr>
<tr>
<td>Total filtered reads</td>
<td>15,105</td>
<td>16,465</td>
<td>9,865</td>
<td>53,743</td>
<td>24,416</td>
<td>54,807</td>
</tr>
<tr>
<td>OTU’s (97% similarity)</td>
<td>1,105</td>
<td>1,084</td>
<td>630</td>
<td>1,695</td>
<td>1,308</td>
<td>1,592</td>
</tr>
<tr>
<td>Abundant OTUs**</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>7</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Rare OTUs**</td>
<td>1,089</td>
<td>1,070</td>
<td>615</td>
<td>1,688</td>
<td>1,295</td>
<td>1,584</td>
</tr>
</tbody>
</table>

* reads with a minimum length of 300 bp and a maximum length of 670 bp; ** abundant OTU = number of reads ≥1% of total reads, otherwise it is rare
Figures

**Fig. 1** Study area, oceanographic fronts, oceanic regions and location of surface water samples. STF = Subtropical Front, SAF = Subantarctic Front, PF = Polar Front, SACCF = Southern ACC Front, SB = Southern ACC Boundary, SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ = Antarctic Zone, SZ = Southern Zone, SR = Subpolar Region. Locations of fronts after Orsi et al. (1995)
Fig. 2 Total biomass of Eukaryotes <50 µm in arbitrary units (AU) (bars) based on total Chl \( a \) fluorescence and the observed change in the FL3/FL2 signal (solid line), indicating shifts in the phytoplankton community composition/physiology as measured by flow cytometry.

Locations of oceanographic fronts (dotted lines) and regions are indicated. SAF = Subantarctic Front, PF = Polar Front, SB = Southern ACC Boundary. SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ = Antarctic Zone, SZ = Southern Zone, SR = Subpolar Region
Fig. 3 Total Chl a concentration (solid line) and size class distribution of total Chl a (bars) based on Chemtax identification of the various algae classes. Locations of oceanographic fronts (dotted lines) and regions are indicated. SAF = Subantarctic Front, PF = Polar Front, SB = Southern ACC Boundary. SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ = Antarctic Zone, SZ = Southern Zone, SR = Subpolar Region
Fig. 4 MDS plot based on Jaccard distances of all 31 samples gained via ARISA profiles. Ellipses are indicating three clusters, Subantarctic Zone (SAZ), Polar Frontal Zone (PFZ) and Antarctic Zone and Subpolar Region (AZ/SR)
Fig. 5 PCA of environmental conditions with plotted ARISA groups (A = SAZ, B = PFZ and C = AZ/SR). Both axes are explaining 96% of the variance (PC1: 78%, PC2: 18%). Group A shows the highest, cluster B intermediate and cluster C the lowest temperature and salinity values.
Fig. 6 Rarefaction analysis for each of the six sequenced samples based on clustering at the 97% similarity level
**Fig. 7** Venn diagram, illustrating the distribution of OTUs across the three regions. SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ/SR = Antarctic Zone and Subpolar Region.

The total OTU number is 3,841
**Fig. 8** Relative abundance of sequence reads, gained via 454-pyrosequencing, assigned to major taxonomic groups. Locations of oceanographic fronts are indicated (dashed lines). SAF = Subantarctic Front, PF = Polar Front, SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ/SR = Antarctic Zone and Subpolar Region
Fig. 9 Colour-coded matrix plot, illustrating the relative read abundance of abundant OTUs (abundance ≥1%, at least in one sample) in the six sequenced samples. White boxes indicate the absence of the respective OTU. SAZ = Subantarctic Zone, PFZ = Polar Frontal Zone, AZ/SR = Antarctic Zone and Subpolar Region.
4.5 Manuscript III

Regional variability in eukaryotic protist communities in the Amundsen Sea

Christian Wolf\(^1\), Stephan Frickenhaus\(^1\), Estelle S. Kilias\(^1\), Ilka Peeken\(^{1,2}\) and Katja Metfies\(^1\)

\(^1\) Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
\(^2\) MARUM – Center for Marine Environmental Sciences, University of Bremen, Leobener Straße, 28359 Bremen, Germany

Antarctic Science, submitted
Abstract

We determined the composition and structure of late summer eukaryotic protist assemblages along a west-east transect in the Amundsen Sea. We used state of the art molecular approaches, such as automated ribosomal intergenic spacer analysis (ARISA) and 454-pyrosequencing, combined with pigment measurements via high performance liquid chromatography (HPLC) to study the protist assemblage. We found characteristic communities offshore and inshore. In general, total chlorophyll \( a \) and microeukaryotic contribution were higher in inshore samples. In contrast to coastal regions of the Arctic Ocean, picoeukaryotes were of minor importance. Diatoms were the dominating group across the entire area, at which \textit{Eucampia} sp. and \textit{Pseudo-nitzschia} sp. were dominating inshore and \textit{Chaetoceros} sp. was dominating offshore. At the most eastern station, the assemblage was dominated by \textit{Phaeocystis} sp. Under the ice, ciliates showed their highest and haptophytes their lowest abundance. This study delivers a comprehensive and taxon detailed overview of the eukaryotic protist composition in the Amundsen Sea during the austral summer.

Keywords

Phytoplankton, Microbial diversity, ARISA, Next-generation-sequencing, HPLC
Introduction

The Pacific Sector of the Southern Ocean and especially the Amundsen Sea are the least studied oceanic regions in the world (Griffiths 2010; Griffiths et al. 2011). Severe ice conditions year-round and the geographic remoteness make sampling in this area very difficult. The biodiversity of the Amundsen Sea, especially of the coastal and shelf areas, is almost unknown (Kaiser et al. 2009). Recently, scientists began to highlight the diversity and distribution of isopods and phytoplankton in this isolated region (Fragoso and Smith 2012; Kaiser et al. 2009). Gravalosa et al. (2008) were concentrating on the distribution of coccolithophores and showed that their dispersion is restricted north of the Polar Front. Fragoso & Smith (2012) focused their study area near the coast and delivered an overview of the phytoplankton assemblage in this area. They revealed diatom dominated assemblages in offshore areas of the Amundsen Sea. However, they used pigment based and microscopic analysis and thereby, the taxonomical resolution was not much detailed. So far, there is no comprehensive survey of the whole eukaryotic protist spectrum in the Amundsen Sea existing.

In the course of the controversially conducted debate about the “everything is everywhere” hypothesis (Finlay and Fenchel 1999; Foissner 1999; Lachance 2004), many studies were focusing on the biogeography of protists (Finlay 2002; Finlay and Fenchel 2004; Galand et al. 2009). Our recent study, focusing on the distribution of eukaryotic protists along a transect from New Zealand to the coast of Antarctica, revealed distinct biogeographic patterns, defined by the oceanic fronts (Wolf et al. submitted). These patterns were driven by strong environmental gradients and included different large-scale water masses. To complement our knowledge about the biogeography of protists in the Pacific sector of the Southern Ocean, their distribution has to be highlighted on a smaller, more regional scale. Narrow environmental differences within a large-scale water mass have to be investigated.

Most investigations of eukaryotic protist composition and distribution in the Southern Ocean mainly used traditional microscopic and pigment extraction based methods (Ishikawa et al. 2002; Wright et al. 2009). However, microscopic surveys have difficulties in identifying small cells and pigment analysis only target autotrophic cells. Here, molecular tools are advantageous. The automated ribosomal intergenic spacer analysis (ARISA) approach provides a quick overview of the diversity and facilitates the comparison of different samples. It is well established for investigations of prokaryotic diversity (Danovaro et al. 2009; Smith et al. 2010) and recently, it was successfully implemented for the analysis of eukaryotic
phytoplankton diversity (Wolf et al. submitted). The newly emerging 454-pyrosequencing approach, e.g. of the V4 region of the 18S rRNA gene, allows assessing microbial communities with high resolution, based on sufficient deep taxon sampling (Margulies et al. 2005; Stoeck et al. 2010), regardless of cell size and nutrition.

The objective of this study is to determine the composition of late summer eukaryotic protist assemblages in the Amundsen Sea, south of the southern boundary of the Antarctic Circumpolar Current. We used state of the art molecular approaches, such as ARISA and 454-pyrosequencing, and high-performance liquid chromatography (HPLC). Furthermore, we want to assess the impact of different environmental conditions on the biogeography of protists, within this oceanic region. The pigment and ARISA analysis provide an overview of differences in structure and diversity of the whole investigated area and the 454-pyrosequencing of selected samples gives more detailed information, about the species composition, dominant representatives and the distribution of the rare biosphere (phylotypes with an abundance < 1% of total sequences) in the observed area.
Materials & Methods

Location and sampling

A total of 34 surface-water samples were taken on a regular basis (approx. every 40 km) during the *RV Polarstern* cruise ANT XXVI/3 between 12 February and 22 March 2010 in the Amundsen Sea (Fig. 1 a) along a west-east transect from the eastern Ross Sea to the western Bellingshausen Sea (within 71.06°S to 74.39°S latitude and 160.27°W to 101.58°W longitude). All surface water samples were collected using the ship pumping system (membrane pump), located at the bow at 8 m depth below the surface. For the determination of pigments, four-litre-samples were immediately filtered onto 25 mm Whatman GF/F filters and stored at -80°C until further analysis in the laboratory. For molecular analysis, two-litre-samples were immediately fractionated, by filtering them on Isopore Membrane Filters (Millipore, USA) with a pore size of 10 µm, 3 µm and 0.2 µm. Filters were stored at -80°C until analysis in the laboratory.

Pigment analysis (HPLC)

Samples were measured using a Waters HPLC-system, equipped with an auto sampler (717 plus), pump (600), PDA (2996), a fluorescence detector (2475) and EMPOWER software. For analytical preparation, 50 µl internal standard (canthaxanthin) and 1.5 ml acetone were added to each filter sample and then homogenized for 20 sec in a Precellys® tissue homogenizer. After centrifugation, the supernatant liquid was filtered through a 0.2 µm PTFE filter (Rotilabo) and placed in Eppendorf cups. An aliquot (100 µl) was transferred to the auto sampler (4°C). Just prior to analysis, the sample was premixed with 1 M ammonium acetate solution in the ratio 1:1 (v/v) in the auto sampler and injected onto the HPLC-system. The pigments were analyzed by reverse-phase HPLC, using a VARIAN Microsorb-MV3 C8 column (4.6x100 mm) and HPLC-grade solvents (Merck). Solvent A consisted of 70% methanol and 30% 1 M ammonium acetate and solvent B contained 100% methanol. The gradient was modified after Barlow *et al.* (1997). Eluting pigments were detected by absorbance (440 nm) and fluorescence (Ex: 410 nm, Em: > 600 nm). Pigments were identified by comparing their retention times with those of pure standards and algal extracts. Additional confirmation for each pigment was done by comparing their absorbance spectra between 390-750 nm with the library of the standards. Pigment concentrations were quantified based on
peak areas of external standards, which were spectrophotometrically calibrated using extinction coefficients published by Bidigare (1991) and Jeffrey et al. (1997). For correction of experimental losses and volume changes, the concentrations of the pigments were normalized to the internal standard canthaxanthin. Phytoplankton group composition was calculated applying the Chemtax program and input ratios of Mackey et al. (1996). To estimate the various size classes of the phytoplankton, the following groups were combined: prasinophytes and pelagophytes for picoplankton (<2 µm), haptophytes and cryptophytes for nanoplankton (2-20 µm), and dinoflagellates and diatoms for microplankton (20-200 µm). Their respective contribution to total biomass is based on their Chemtax derived chlorophyll $a$ concentration.

DNA extraction

The DNA was extracted with the E.Z.N.A.TM SP Plant DNA Kit (Omega Bio-Tek, USA). At the beginning, the filters were incubated with lysis buffer. All further steps were performed as described in the manufacturer’s instructions. At the end, the DNA was eluted in 60 µl of elution buffer and the extracts were stored at -20°C until further analysis. DNA concentration was measured with a NanoDrop 1000 (Thermo Fisher Scientific, USA) (average DNA concentration: 23 ng µl$^{-1}$).

PCR amplification, ARISA

An equal volume of extracted DNA of each size fraction (>10 µm, 3-10 µm and 0.2-3 µm) from each sample was pooled. The ITS1 (internal transcribed spacer) region was amplified in triplicates 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 1528F primer was labelled at the 5’-end with the dye 6-FAM (6-Carboxyfluorescein). The PCR mixtures 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 µM 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: an initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 45 sec, 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).

PCR amplification, 454-pyrosequencing

Seven samples were sequenced (Table I). For each fraction of a sample, we amplified ~670 bp fragments of the 18S rRNA gene, containing the highly variable V4-region, using the primer-set 528F (5’-GCG GTA ATT CCA GCT CCA A-3’) and 1055R (5’-ACG GCC ATG CAC CAC CAC CCA T-3’) (modified after Elwood et al. (1985)). The PCR mixtures were composed as described previously for ARISA. Reaction conditions were as following: an initial denaturation at 94°C for 3 min, 30 cycles of denaturation at 94°C for 45 sec, annealing at 59°C for 1 min and extension at 72°C for 3 min, and a final extension at 72°C for 10 min. An equal volume of PCR reaction of each size fraction from each sample was pooled and purified with the MinElute PCR purification kit (Qiagen, Germany) following the manufacturer’s instructions. Pyrosequencing was performed on a Genome Sequencer FLX system (Roche, Germany) by GATC Biotech AG (Germany).

Data analysis, ARISA

Electropherograms were analyzed using the GeneMapper Software v4.0 (Applied Biosystems, USA). Peaks with a size smaller than 50 bp (corresponding to primer and primer dimer peaks) were removed from the data set. To remove the background noise and to get sample-by-binned-OTU (operational taxonomic unit) tables, the data were binned using the binning scripts, according to Ramette (2009), for R (R Development Core Team 2008). The resulting sample-by-binned-OTU tables were transformed into presence/absence matrices and the distances between the samples were calculated, using the Jaccard index implemented in the R package vegan (Oksanen et al. 2011), which was also used in the following steps. MetaMDS (maximum random starts of 300) plots were computed. Clusters were determined using the hclust function in R. To test, whether the resulting clusters differ significantly, an ANOSIM was performed. A Euclidean distance matrix with the normalized environmental parameters was calculated. The correlation between the ARISA distance matrix and the environmental distance matrix was tested with a Mantel test (10,000 permutations), implemented in the R package ade4 (Dray and Dufour 2007). A PCA with the environmental parameters was performed (R package ade4).
Data analysis, 454-pyrosequencing

Raw sequence reads were processed to obtain high quality reads. The forward primer 528F, used for the sequencing, attaches approx. 25 bp upstream of the V4 region, which has in general a length of approx. 230 bp (Nickrent and Sargent 1991). Reads with a length under 300 bp were excluded from further analysis to assure the including of the whole hyper variable V4 region in the analysis and to get rid of short reads. Unusually long reads, that were greater than the expected amplicon size (>670 bp) and reads with more than one uncertain base (N) were removed. Remaining reads were checked for chimeric sequences with the software UCHIME 4.2.40 (Edgar et al. 2011) and all reads considered being chimeric were excluded from further analysis. The high quality reads of all samples were clustered into operational taxonomic units (OTUs) at the 97% similarity level using the software Lasergene 10 (DNASTAR, USA). Subsequently, reads not starting with the forward primer were manually removed. Consensus sequences of each OTU were generated, which reduced the amount of sequences to operate with and attenuated the influence of sequencing errors and uncertain bases. The 97% similarity level has shown to be the most suitable to reproduce original eukaryotic diversity (Behnke et al. 2011) and has the effect of bracing most of the sequencing errors (Kunin et al. 2010). Furthermore, known intragenomic SSU polymorphism levels can range to 2.9% in dinoflagellate species (Miranda et al. 2012). OTUs comprised of only one sequence (singletons) were removed. The consensus sequences were aligned into a reference alignment obtained from SILVA (see below) using the software HMMER 2.3.2 (Eddy 2011). Subsequently, taxonomical affiliation was determined by placing the consensus sequences into a reference tree, containing about 1,200 high quality sequences of Eukarya from the SILVA reference database (SSU Ref 108), using the software pplacer 1.0 (Matsen et al. 2010). The compiled reference database is available on request in ARB-format. OTUs assigned to fungi and metazoans were excluded from further analysis. Rarefaction curves were computed using the freeware program Analytic Rarefaction 1.3. The data set generated in this study has been deposited at GenBank’s Short Read Archive (SRA) under Accession No. SRA057133.
**Results**

**Environmental conditions** – The investigated area showed a very heterogeneous setting, in terms of water depth, surface temperature, surface salinity and ice coverage (Fig. 1 a-d). Samples 33-46 and 49-57 were lying offshore, with water depths (Fig. 1 a) from 1,969 m (sample 57) to 4334 m (sample 34). Samples 47-48 (polynia) and 58-71 (Pine Island Bay) were inshore, with water depths from 398 m (sample 69) to 714 (sample 48). Sample 60 was lying over the continental slope and showed a greater depth (1,447 m).

The surface water temperature ranged between -1.63°C (sample 47) and -0.24°C (sample 60) (Fig. 1 b). Hence, the temperature only varied weakly, but showed significantly higher values at the most eastern sample sites (60-62), located at the transition to the Bellingshausen Sea.

Surface water salinity (Fig. 1 c) showed values between 32.35 PSU (sample 57) and 33.51 PSU (sample 34). In general, salinity was higher in the western part of the transect and declined eastwards. Among the eastern samples, sample 69 showed a very high salinity (33.32 PSU).

Most samples were located near the ice edge (Fig. 1 d) with no ice. At samples 45-48, we crossed an ice field to reach a polynia, with a high spatial variability of the ice cover (5-50%). Samples 57-58 were taken in an ice field and showed an ice coverage of 10-50%. Sample 69 was obtained in a region with 100% ice cover.

**Structure/diversity overview**

We used a combination of HPLC and ARISA to assess the impact of different environmental conditions on the structure/diversity of the plankton assemblages in the Subpolar Region.

**HPLC** – Total Chl a concentrations (Fig. 2 a) along the entire transect ranged between 0.11 µg L⁻¹ (sample 36) and 9.58 µg L⁻¹ (sample 70). In general, the highest Chl a concentrations occurred in samples lying inshore. The Chl a concentrations in these areas always exceeded 1 µg L⁻¹. However, the majority of samples (21) showed Chl a concentrations lower than 0.5 µg L⁻¹. All these samples, except of sample 68, were taken offshore.

The contribution of the three size classes (picoeukaryotes (0.2-2 µm), nanoeukaryotes (2-20 µm) and microeukaryotes (>20 µm)) to total Chl a showed, that picoeukaryotes (Fig. 2 b) did not significantly contribute to phytoplankton biomass throughout the entire transect.
The highest contribution of picoeukaryotes occurred in samples 47 (4%), 48 (3.2%), 54 (6.5%) and 62 (7.7%), mainly samples lying inshore. In all other samples, picoeukaryotes did not exceed a contribution of 1.9%. In general, nanoeukaryotes showed the highest contribution to total Chl a in offshore samples (Fig. 2 c). In these areas, they contributed up to 58.5% (sample 36). In offshore samples, they accounted for 33% ± 10% of Chl a on average, whereas in inshore samples they only accounted for 25% ± 15% on average. Sample 68, with a contribution of nanoeukaryotes of 63%, presented as an outlier, just like for total Chl a concentration. The lowest contribution of nanoeukaryotes (~14%) was shown by the two polynia samples (samples 47 and 48). Microeukaryotes were always the dominating size class (Fig. 2 d), except for samples 33, 35, 36 and 68, where nanoeukaryotes were dominating. Microeukaryotes contributed 35.8-84.5% to total Chl a, in which the highest values occurred generally in inshore samples (except of sample 68). In these areas, they contributed 73% ± 15% on average, whereas in offshore ocean samples they accounted for 66% ± 11% on average.

**ARISA** – The fragment length analysis of the ITS1 region of all 34 surface water samples resulted in 97 different fragments with a length of 50 to 432 bp, of which 16 only occurred in one sample (unique fragments). The number of fragments in each sample was 26 on average, ranging from nine (sample 50) to 49 (sample 68). The ordination analysis based on the ARISA profiles (Fig. 3) clustered the samples in three groups. Group A includes samples 33-42, group B contains samples 44-45 and 49-62 and group C includes samples 46-48 and 68-71. The three groups show significantly different ARISA profiles (ANOSIM, R = 0.637, p = 0.001). Groups A and B persist of offshore samples and are representing the western and eastern part of the transect, respectively. Group C consists of samples collected inshore. Samples 58-62 fall into group B, although they were located over the shelf.

The ARISA profiles distances are significantly correlated with the distances of environmental conditions profiles (Mantel test, r = 0.142, p = 0.023). Figure 4 shows the PCA of the environmental conditions with the three ARISA groups plotted in. The two axes are explaining 83% of the total variance. Group C is mainly separated from group A and B by lower water depths and a higher ice coverage. Group A is primarily separated from group B by higher salinities and lower temperatures.
Detailed community structure

To obtain taxon detailed information about the community, we sequenced seven samples (samples 41, 47, 51, 57, 62, 69 and 70), spanning the entire transect and including all three ARISA groups. Three samples (samples 41, 51 and 57) were taken in open ocean waters and four samples (samples 47, 62, 69 and 70) were taken above the shelf.

454-pyrosequencing – The summary of recovered 454-pyrosequencing reads is shown in table I. In total, 278 116 sequence reads were obtained from 454-pyrosequencing, of which 77.1% had an acceptable length (300-670 bp). After the quality filtering, 56.5% of the total reads were left for analysis. The number of analyzed reads ranged between 14219 (sample 47) and 29241 (sample 57). Subsequent to the clustering, 4044 different OTUs could be observed. The number of OTUs for each sample (Table 1) ranged between 893 (sample 47) and 1687 (sample 69), at which only 0.7% (sample 57 and 70) to 1.7% (sample 47) were abundant (number of reads ≥1% of total reads). The proportion of unique OTUs (i.e., OTUs occurring in one sample only) was 36%.

The rarefaction curves (Fig. 5) show that none of the samples demonstrates saturation. However, the stacking of the curves suggests that samples 41, 47 and 51 harboured the lowest diversity.

The relative abundance of sequences assigned to major protist groups is shown in figure 6. Haptophytes showed a read abundance of 9-17% in offshore samples and 14-37% in inshore samples, except in sample 69, where they accounted for only 3%. Chlorophytes occurred in significant amounts only inshore where they composed 1.7% to 6.3% of the reads. Sample 69 was again an exception, because here chlorophytes only accounted for 0.6% of the reads. Pelagophytes only occurred in great quantities in one offshore sample (sample 41) with 11.6% of the sequence reads. Diatoms were the dominating group in samples 41, 47, 57, 69 and 70 with a read abundance of 40%, 52%, 44.7%, 48.3% and 40.3%, respectively. In samples 51 and 62, they accounted for 28.2% and 11.3% of the reads, respectively. Labyrinthulids were occurring in significant amounts only inshore, in samples 62, 69 and 70, where their read abundance accounted for 2.5%, 2.5% and 6.3%, respectively. The read abundance of the marine stramenopiles (MAST) group was composing 1.4-6.6%, whereas the highest abundance occurred in sample 62. Dinoflagellates were dominating the sequence assemblage in sample 51 with 38%. In the other samples, they accounted for 9.5-21.2% of the reads. In general, Dinoflagellates showed a higher read abundance in offshore than in inshore
samples. The highest read abundance of Syndiniales occurred in sample 57 (12.6%). In the other samples, they accounted for 2.5-10.7% of the reads. Ciliates played a minor role in all sequence assemblages, except in sample 69, where they account for 17.6% of the sequences.

Of the 4,044 OTUs, 34 were abundant (i.e., abundance > 1%) in at least one sample. A detailed overview of the relative read abundances of the abundant phylotypes is shown in figure 7. Three phylotypes were abundant in all seven samples (Phaeocystis sp. 1, Eucampia sp. and unclassified (unc.) Dinoflagellate 1). They were also among the most abundant phylotypes across the entire transect. The Phaeocystis sp. 1 OTU showed the highest read abundance inshore, in samples 70 and 62, with 21% and 26.7%, respectively. However, in sample 69 it was almost rare, with the lowest read abundance of 1.5%. The chlorophytes, represented by Micromonas sp. and Pyramimonas sp., were only abundant inshore (sample 47 and 62), with 3.7% as highest sequence abundance. We found nine abundant phylotypes among the diatoms. The most abundant was Eucampia sp., with a read abundance up to 23.6% (sample 69). Only in sample 47 and 51, the most abundant diatom phylotype was not Eucampia sp., but Pseudo-nitzschia sp. (13.8%) and Chaetoceros sp. 1 (12.6%), respectively. Pelagomonas sp., belonging to the pelagophytes, showed a high read abundance offshore, in sample 41 (10.4%), whereas it was nearly rare in all other samples. Among the rest of the “other stramenopiles”, the unc. Labyrinthulid OTU showed the highest read abundance in sample 70 (4.8%). We found four abundant dinoflagellate phylotypes, at which the unc. Dinoflagellate 1 was the most abundant, with a read abundance ranging from 5.2% to 19.8%. The highest abundance appeared offshore in sample 51. The other dinoflagellate phylotypes did not exceed a read abundance of 2.1%. Among the abundant Syndiniales phylotypes, the unc. Syndiniales 2 showed the highest read abundance with 3.9% in sample 57. Ciliate phylotypes were only abundant in sample 69, whereat the unc. Ciliate 1 OTU showed the highest sequence abundance (3.4%). The rare biosphere accounted for 34.2% (sample 47) to 45.8% (sample 62) of all reads.
Discussion

Structure/diversity overview and biogeographic patterns

One aim of this study was to determine the structure and diversity of eukaryotic protist assemblages in the Amundsen Sea and to assess the impact of environmental conditions on their biogeographic patterns. We used a combination of pigment analysis (HPLC) and ARISA, to get an overview of the structure/diversity and the biogeographic patterns. The resulting ARISA profiles were linked with the environmental conditions.

Previous biogeographic classifications of surface waters are broad and of larger scale (Spalding et al. 2012). For shelf regions the existing classifications are more detailed (Spalding et al. 2007). In our previous study we confirmed characteristic protistan assemblages for each large-scale water mass in the Southern Ocean (Wolf et al. submitted). However, it is also important to study more regional patterns, to complement our knowledge about the diversity and biogeography of protists in the Pacific sector of the Southern Ocean.

In general, we observed clear differences of total Chl \(a\) concentrations between the samples taken offshore and inshore. Inshore the concentrations always exceeded 1 µg L\(^{-1}\). This is congruent with other studies, which observed higher Chl \(a\) concentrations in Antarctic shelf and coastal near waters than in open oceanic waters (Hashihama et al. 2008; Ishikawa et al. 2002; Olguin and Alder 2011; Garibotti et al. 2003). Along the entire transect, sample 70 showed the highest Chl \(a\) concentration with 9.58 µg L\(^{-1}\), indicating a large phytoplankton bloom in this area. The high Chl \(a\) value is not surprising, since recent investigations observed Chl \(a\) concentrations up to 8-14 µg L\(^{-1}\) in the shelf area of the Amundsen Sea (Alderkamp et al. 2012; Fragoso and Smith 2012; Mills et al. 2012).

The high Chl \(a\) concentrations we observed above the shelf were accompanied by higher proportions of microeukaryotes. Higher Chl \(a\) concentrations were often connected with high abundances of larger cells, like diatoms (Detmer and Bathmann 1997; Ishikawa et al. 2002). The geomorphology in the shelf areas promote upwelling and mixing and thus, the nutrient availability in this region is higher, which promotes the build-up of biomass and favours larger cells (Irwin et al. 2006). Picoeukaryotes were of minor importance throughout the entire transect, which is in contrast to the High Arctic, where picophytoplankton contribution to total phytoplankton abundance of up to 76% has been described (Tremblay et al. 2009). Nanoeukaryotes were the counterpart to microeukaryotes in the investigated area.
They showed their highest contribution in samples, where microeukaryotes were less abundant.

The ARISA profiles generally approve the existence of an offshore and an inshore group in the investigated area. The offshore group is split into a western and an eastern part, whereat the eastern part was characterized by lower salinities, due to melting ice in this area. Samples 58-62 belong to the second offshore group, although they were taken above the shelf. One explanation could be that these areas are more influenced by open oceanic water. In these areas, Circumpolar Deep Water (CDW) is flowing onto the continental shelf through troughs in the shelf as modified CDW (Alderka mp et al. 2012) and may influence the surface layer (upwelling). However, it appears more likely that wind is the major determining factor, influencing the direction of the surface currents.

Detailed community structure

The most prominent taxonomic group across the entire transect were the diatoms. This group was previously observed to dominate in the Amundsen Sea, especially in the sea ice zones (Alderka mp et al. 2012; Fragoso and Smith 2012; Mills et al. 2012). The most dominant diatom in the Pine Island Bay was \textit{Eucampia} sp. Garibotti \textit{et al.} (2003) found a large contribution to total diatom biomass of \textit{Eucampia antarctica} in the Marguerite Bay (Antarctic Peninsula). It seems that the conditions in bays may constitute an optimal environment for \textit{Eucampia} to grow. In the Amundsen Polynya, we found \textit{Pseudo-nitzschia} sp. as the most dominant diatom, whereas offshore, \textit{Chaetoceros} sp. was generally the dominating diatom. These two genera were previously reported to dominate in waters around Antarctica (Kopczynska \textit{et al.} 2001; Gomi \textit{et al.} 2005).

Sample 62, in contrast, showed a dominance of \textit{Phaeocystis} sp. A dominance of \textit{Phaeocystis antarctica} in several regions of the Amundsen Sea was previously reported (Alderka mp \textit{et al.} 2012; Mills \textit{et al.} 2012). Arrigo \textit{et al.} (1999) revealed that \textit{Phaeocystis antarctica} dominates where waters are deeply mixed, whereas diatoms dominate in highly stratified waters. Hence, the domination of \textit{Phaeocystis sp.} in sample 62 could be due to more deeply mixed water. Another explanation could be that the succession at the eastern edge of the transect was most advanced (post bloom), due to a longer time period free of ice, retraced via Advanced Microwave Scanning Radiometer (AMSR) satellite images (Spreen \textit{et al.} 2008). In polar waters, after a diatom dominated bloom, \textit{Phaeocystis} often dominated the post bloom situation (McMinn and Hodgson 1993; Poulsen and Reuss 2002).
Sample 69 showed the most extreme ice condition with 100%. Here, the read abundance of *Phaeocystis* was very low. The lack of wind stress, due to the ice coverage, could have caused the water to be highly stratified, and therefore led to a low *Phaeocystis* abundance (Arrigo et al. 1999; Goffart et al. 2000). Under the ice, ciliates showed their highest read abundance. This corresponds to other observations of the under-ice community structure, which revealed that heterotrophic biomass was dominated by ciliates (Ichinomiya et al. 2007).

In contrast to our previous study, focusing on the distribution of eukaryotic protists across the main oceanic fronts of the Southern Ocean (Wolf et al. submitted), the distribution of OTUs was more even. The proportion of unique OTUs was only half the amount (37%) as it was across the main fronts of the Southern Ocean (76%). This is distinctly visible in the distribution of the abundant biosphere (Fig. 7). There were only a few OTUs, which were not present in all samples, whereas in our previous study, there were much more gaps (20.4% opposite to 10.1%). Here, in the single large-scale water mass, the observed “rare biosphere” serves as a background population and several species may become abundant when the environmental conditions change. However, in both studies the rarefaction curves suggest that none of the samples have been exhaustively analyzed by sequencing. Nevertheless, the rarefaction curves indicate that the highest diversity was observed under the ice (sample 69).

There are some potential biases, which can lead to a tampering of the real situation, while interpreting molecular data. The amplification of the different species in an environmental sample can vary and some species might not be captured by the used primers (primer specificity) (Jeon et al. 2008; Liu et al. 2009; Stoeck et al. 2010). The ARISA approach can only serve as an approximate overview of the diversity structure, due to the qualitative character and the identical fragment lengths of some different species (Caron et al. 2012). Additionally, the number of rRNA gene copies depends on the cell size and varies among eukaryotes from one to several hundreds (Zhu et al. 2005). Especially the dinoflagellates seem to have more rRNA gene copies than the other taxonomical groups and thus might be overrepresented in molecular sequence data. The placement of sequences gained via 454-pyrosequencing has to be interpreted with care, because the length of only ~500-600 bp is affecting the robustness. Therefore, we generally did not classify the OTUs beyond the genus level.

In conclusion, we have shown that within a single water mass protist assemblages differed in dimensions and species composition, according to geographical and environmental conditions. There were two major groups, the offshore and the inshore group. Biomass and microeukaryotes contribution to total Chl *a* were highest in the inshore group, whereas in the
offshore group the contribution of nanoeukaryotes was the highest across the entire transect. Diatoms were the most prominent protist class, and the diatom species appearing as most abundant differed among the locations.
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. We are grateful to F. Kilpert and B. Beszteri for their bioinformatical support. We also want to thank A. Schroer, A. Nicolaus and K. Oetjen for technical support in the laboratory and Steven Holland for providing access to the program Analytic Rarefaction 1.3. We would like to acknowledge E. M. Nöthig and K. Kohls for their insightful discussions and reviews of this manuscript.
References


Davidson AT, Scott FJ, Nash GV, Wright SW, Raymond B (2010) Physical and biological control of protistan community composition, distribution and abundance in the


**Tables**

**Table I** Summary of recovered 454-pyrosequencing reads, quality filtering and number of OTUs. Samples are arranged from west to east.

<table>
<thead>
<tr>
<th></th>
<th>Sample 41</th>
<th>Sample 47</th>
<th>Sample 51</th>
<th>Sample 57</th>
<th>Sample 70</th>
<th>Sample 69</th>
<th>Sample 62</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total 454-reads</strong></td>
<td>29807</td>
<td>24109</td>
<td>34767</td>
<td>49355</td>
<td>33020</td>
<td>63836</td>
<td>43222</td>
</tr>
<tr>
<td><strong>Average length (bp)</strong></td>
<td>332</td>
<td>333</td>
<td>345</td>
<td>360</td>
<td>383</td>
<td>370</td>
<td>393</td>
</tr>
<tr>
<td><strong>Acceptable length</strong>*</td>
<td>21339</td>
<td>17332</td>
<td>26562</td>
<td>36695</td>
<td>26687</td>
<td>49221</td>
<td>36520</td>
</tr>
<tr>
<td><strong>Quality filtering:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>More than one N</strong></td>
<td>142</td>
<td>67</td>
<td>135</td>
<td>309</td>
<td>228</td>
<td>406</td>
<td>386</td>
</tr>
<tr>
<td><strong>Chimeras</strong></td>
<td>754</td>
<td>504</td>
<td>484</td>
<td>1464</td>
<td>1278</td>
<td>2096</td>
<td>1265</td>
</tr>
<tr>
<td><strong>Incorrect forward primer</strong></td>
<td>302</td>
<td>204</td>
<td>280</td>
<td>185</td>
<td>159</td>
<td>327</td>
<td>393</td>
</tr>
<tr>
<td><strong>Singletons</strong></td>
<td>1007</td>
<td>853</td>
<td>1377</td>
<td>3223</td>
<td>1730</td>
<td>3583</td>
<td>3025</td>
</tr>
<tr>
<td><strong>Non-target organisms</strong></td>
<td>1131</td>
<td>1485</td>
<td>2175</td>
<td>2273</td>
<td>5913</td>
<td>14619</td>
<td>3462</td>
</tr>
<tr>
<td><strong>Total filtered reads</strong></td>
<td>18003</td>
<td>14219</td>
<td>22111</td>
<td>29241</td>
<td>17379</td>
<td>28190</td>
<td>27989</td>
</tr>
<tr>
<td><strong>OTUs (97% identity)</strong></td>
<td>927</td>
<td>893</td>
<td>1161</td>
<td>1593</td>
<td>1219</td>
<td>1687</td>
<td>1554</td>
</tr>
<tr>
<td><strong>Abundant OTUs</strong></td>
<td>12</td>
<td>15</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td><strong>Rare OTUs</strong></td>
<td>915</td>
<td>878</td>
<td>1150</td>
<td>1582</td>
<td>1211</td>
<td>1674</td>
<td>1542</td>
</tr>
</tbody>
</table>

* reads with a minimum length of 300 bp and a maximum length of 670 bp;
** abundant OTU = number of reads ≥1% of total reads, otherwise it is rare
Fig. 1 Study area and environmental conditions. a Location of surface water samples and water depth. b Surface water temperature. c Surface water salinity. d Ice coverage.
Fig. 2 Total chlorophyll $a$ (Chl $a$) concentration and size class distribution of total Chl $a$ based on Chemtax identification of the various algae classes. a Total Chl $a$. b Proportion of picoeukaryotes. c Proportion of nanoeukaryotes. d Proportion of microeukaryotes.
Fig. 3 MDS plot based on Jaccard distances of all 34 samples, gained via ARISA profiles. Colours of the samples indicate the three groups (red = group A, blue = group B, green = group C).
Fig. 4 PCA of environmental conditions with plotted ARISA groups (A, B and C). Both axes are explaining 83% of the variance (PC1: 52%, PC2: 31%). Cluster A shows greater water depths and higher salinities. Cluster B is characterized by lower salinities. Cluster C shows lower water depths and a higher ice coverage. (d = axis scaling factor)
Fig. 5 Rarefaction analysis for each of the seven sequenced samples based on clustering at the 97% similarity level.
Fig. 6 Relative abundance of sequence reads, gained via 454-pyrosequencing, assigned to major taxonomic groups. Blue encircled samples = “offshore”, green encircled samples = “inshore”, * = 100% ice coverage.
Fig. 7 Colour-coded matrix plot, illustrating the relative read abundance of abundant OTUs (abundance ≥1%, at least in one sample) in the seven sequenced samples. White boxes indicate the absence of the respective OTU.
5. Synthesis and future perspectives

The Southern Ocean is important for the global climate and the global marine carbon cycle (Sarmiento & LeQuere 1996, Sarmiento et al. 1998). It comprises several different oceanic regions with high and low primary production. Protists are the largest contributors to organic matter in the oceans and are important for marine food webs and biogeochemical cycles (Smith Jr. & Sakshaug 1990, Priddle et al. 1992). Traditional methods to investigate protist assemblages are complemented more and more by molecular methods.

The main objectives of this thesis were the establishment of molecular approaches in the diversity investigation of eukaryotic protists in the Southern Ocean, the comparison of different molecular approaches and the delivery of a comprehensive and taxon detailed overview of protist assemblages in the Pacific sector of the Southern Ocean, especially in the scarce investigated Amundsen Sea. The hypotheses that distinct protist community assemblages characterize different large-scale water masses and that a single large-scale water mass shows weaker, but still detectable regional community variability was addressed. Furthermore, the extent and role of the rare biosphere as background population, with regard to the “everything is everywhere” hypothesis, was treated.

5.1 Investigation of eukaryotic protist diversity with molecular approaches

The recent development of molecular approaches has altered our knowledge and understanding of marine protist diversity and distribution (Diez et al. 2001, Sogin et al. 2006, Stoeck et al. 2010, Caron et al. 2012). The extent of protist diversity is much higher than previously assumed and the revealing of a vast rare biosphere fed the discussion about the “everything is everywhere” hypothesis (Finlay & Fenchel 1999, Lachance 2004, Sogin et al. 2006). However, so far, not all molecular approaches are implemented in all areas of investigation and there is still discussion about the applicability and interpretation of molecular data. The rapid evolution in the field of molecular biology and the consistently invention of new approaches impede keeping track of the most accepted, accurate and applicable method.

Manuscript I compared the two most popular approaches for obtaining taxon detailed diversity information of protists, i.e. sequencing of clone libraries and 454-pyrosequencing. It shows that our understanding of protist diversity and structure assessed with molecular methods varies strongly depending on the molecular method used. The clone library approach
may ignore entire taxonomical groups, e.g. in this study the haptophytes or diatoms, which leads to a falsified community picture. Furthermore, the study shows that comparisons of new 454-pyrosequencing data with previously published clone library data of protist diversity have to be handled with care and should not be over interpreted. The low throughput and the overrepresentation of single phylotypes in the clone library approach make a direct comparison difficult. Moreover, the cloning approach seems to be of very limited suitability as backbone for a refined phylogenetic analysis of OTUs identified by 454-pyrosequencing data, because over 50% of the abundant phylotypes in the 454-pyrosequencing data set were not recovered by this approach.

In **Manuscripts II** and **III**, the ARISA approach was established for investigations of the structure of eukaryotic phytoplankton. The different major water masses of the Southern Ocean, separated by the main oceanic fronts, were well resolved. Within a single large-scale water mass (Amundsen Sea), the ARISA approach even distinguished between offshore and inshore regions. Hence, the approach is very well suited for investigations of eukaryotic phytoplankton assemblage structure in large numbers of samples.

The combination of ARISA and 454-pyrosequencing seem to be an appropriate approach for investigating the diversity and structure of protist assemblages. This combination can deliver a broad overview of the community structure of large numbers of samples (ARISA) and it can give taxon detailed information (454-pyrosequencing). Examining large numbers of samples is crucial to reveal and evaluate biogeographical patterns and taxon detailed information is needed to reveal and assess future community changes, e.g. due to climate change.

A comparison or validation with data gained by non-molecular approaches (e.g. microscopy or pigment based methods) is always advisable, because molecular approaches suffer from several biases. Primer specificity is one of the drawbacks (Farris & Olson 2007). The primers used for amplification may attach with different efficiency to the different species or groups in one environmental sample and lead to an over amplification of several species or groups. Additionally, Zhu et al. (2005) have described that the number of rRNA gene copies depends on the cell size and varies among eukaryotes from one to several hundreds. Especially the dinoflagellates seem to have more rRNA gene copies than the other taxonomical groups and thus might be overrepresented in molecular data. The ARISA approach can only serve as an approximate overview of the diversity structure, due to the qualitative character and the overlapping of fragment lengths among different species. The
placement of sequences gained via 454-pyrosequencing has to be interpreted with care, as the length of only ~500-600 bp decreases the robustness of the phylogenetic annotation.

However, molecular approaches will be essential components for diversity investigations in the future, because they enable taxon detailed and cell size independent information. In combination with microscopy, pigment, and cultivation data, it will close many gaps of knowledge regarding the diversity and role of protists in the Southern Ocean.

5.2 Protist assemblages in the Pacific sector of the Southern Ocean

Many studies about protist diversity and distribution in the Southern Ocean focus on particular groups, e.g. diatoms (Gomi et al. 2005, Annett et al. 2010, Gomi et al. 2010) or coccolithophorids (Gravalosa et al. 2008), or give only an overview of the major groups (Peeken 1997, Kopczynska et al. 2007), or the biomass production (Arrigo et al. 1998, Arrigo et al. 2008). In particular, taxon detailed information about the smallest protists, the picoplankton (0.2-3 µm), is scarce (Diez et al. 2001, Lopez-Garcia et al. 2001). There is hardly any study combining taxon detailed investigation of all size classes (pico-, nano- and microplankton). Molecular studies are scarce in the Southern Ocean and concentrate on sequencing of clone libraries (Diez et al. 2001, Lopez-Garcia et al. 2001). There is a lack of taxon detailed investigations, especially in the least studied region of the world’s oceans, the Pacific sector of the Southern Ocean. Not much is known about the diversity and composition of protists in this area, especially in the Amundsen Sea. Information about the diversity and distribution of eukaryotic protists is crucial for assessing future changes.

Manuscripts II and III deal with the diversity and biogeography of eukaryotic protists in the Pacific sector of the Southern Ocean during austral summer.

We found distinct biogeographic patterns defined by the oceanographic fronts in the particular region (Manuscript II). In contrast to the Arctic Ocean, picoeukaryotes had minor importance throughout the investigated area and were nearly absent south of the Polar Front (PF). Dinoflagellates and smaller cells, like labyrinthulids and Syndiniales, were dominating north of the PF, whereas haptophytes and diatoms became more important south of the PF.

In the Amundsen Sea (Manuscript III), we found characteristic offshore and inshore communities. In general, microeukaryotic contribution was higher in inshore samples. In contrast to coastal regions of the Arctic Ocean, picoeukaryotes were of minor importance. Diatoms were the dominating group across the entire area, at which *Eucampia* sp. and *Pseudo-nitzschia* sp. were dominating inshore and *Chaetoceros* sp. was dominating offshore.
At the most eastern station, the assemblage was dominated by *Phaeocystis* sp. Under the ice, ciliates showed their highest and haptophytes their lowest abundance.

Both manuscripts in combination deliver a comprehensive and taxon detailed overview of the eukaryotic protist composition in the investigated area during the austral summer. They clearly show that different large-scale water masses are characterized by distinct protist community assemblages and that within a single large-scale water mass the regional differences are weaker, but still detectable. This overview serves as status quo for the assessment of future changes of protist assemblages due to climate change. Time will tell if smaller cells become more important in marine ecosystems in the future, due to warming surface water temperatures, as expected by several authors (Daufresne et al. 2009, Li et al. 2009). This would change food webs and biogeochemical cycles in the Southern Ocean, because so far, small cells (picoplankton) are of minor importance.

Some may argue that the main criticism of these studies is the descriptive character. A description of the protist diversity only delivers a picture about “who” is there and how abundant “they” are. These studies deliver no information about the importance of the revealed organisms for the ecosystem. An answer about how they are involved in food webs and biogeochemical cycles cannot be given satisfactory. Nevertheless, the description is the first step to get answers for the above-mentioned questions. Biodiversity is important at multiple levels for ecosystem functioning and stability and serves as an indicator for environmental conditions (Bagley et al. 2002). It represents a kind of biological insurance against the loss or poor performance of selected species (Folke et al. 1996, Naeem & Li 1997). Therefore, it plays a major role in the renewal and reorganization of ecosystems after disturbance (Folke et al. 2004). Hence, the description of protist biodiversity is crucial for the understanding of ecosystem functioning. First, we have to know “who” is there, before we can figure out what “they” do and how.

The molecular high throughput diversity data seem to be valuable for the discussion about the “everything is everywhere” hypothesis. Baas Becking (1934) was first to postulate “Everything is everywhere, but, the environment selects”. The idea behind this phrase was that all microbial species are cosmopolitan, but in a given environmental setting most of them are only latently present (de Wit & Bouvier 2006). The majority of microbial diversity is hidden from our observation, due to the detection limits. The “everything is everywhere” idea has been debated controversially in recent years (Finlay & Fenchel 1999, Foissner 1999, Lachance 2004). Studies focusing on the diversity and geographic distribution of protists found on the one hand biogeographic patterns (Cho & Tiedje 2000, Richards et al. 2005,
Synthesis and future perspectives

Galand et al. 2009) and on the other hand globally distributions (Finlay 2002, Finlay & Fenchel 2004, Richards et al. 2005). However, most of the previous studies concentrated on prokaryotes and on other regions than the Southern Ocean. The distribution of phylotypes found in this study indicates that different regions represent distinct protist assemblages. The rare biosphere did not appear as a background population that contains species, which are abundant under the other environmental conditions. However, the present study does not resolve possible seasonal effects. One could argue that the “everything is everywhere” hypothesis does not apply here. However, it has to be kept in mind that the samples were not exhaustively explored (454-pyrosequencing still not covers all sequences, drawbacks of methods), and definite conclusions cannot be made. Nevertheless, several strong indications for biogeographic distribution can be found with the applied methods.

5.3 Future perspectives

We are far away from exhaustively knowing and understanding the diversity and distribution of protists in the Southern Ocean. We just have started to drill deeper, beyond dominant and abundant species, to get the whole picture of protist biodiversity. To understand the processes, interactions and relationships of protists in food webs and biogeochemical cycles, there is a need to continue investigating their diversity and distribution in the Southern Ocean. We have to close the gaps of knowledge and to create a detailed and comprehensive diversity and distribution data set of the entire Southern Ocean, including all scarce investigated regions (e.g. the Indian sector). Detecting changes in composition and distribution of protist assemblages, due to climate change, is only possible with profound status quo information and continuous sampling and analysis. Therefore, yearly sampling of representative regions is needed. Additionally, information about the seasonal succession is needed. So far, most data was acquired during the austral summer and data from the austral winter is scarce. Seasonal information is crucial to understand all processes in marine ecosystems.

The molecular age provides us with several appropriate tools to create the needed knowledge relatively fast and inexpensive. Two appropriate ones, as shown in this thesis, are ARISA and next-generation sequencing. Further, the combination of these methods with more traditional methods, i.e. microscopy, cultivation, and studies of biogeochemical parameters, is needed to get morphological, ecophysiological and functional information. This information is just as crucial for the understanding of processes, interactions, and relationships of protists
in food webs and biogeochemical cycles, and for modelling future changes in a changing climate.

Furthermore, the gained information can be brought in a bipolar perspective by comparing them with similar data from the Arctic Ocean.
6. References


Baas Becking LGM (1934) Geobiologie of inleiding tot de milieukunde, W.P. Van Stockum & Zoon (in Dutch), The Hague, the Netherlands


References

Bloom in the polar Southern Ocean stimulated by iron fertilization. Nature 407:695-702


de Wit R, Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol 8:755-758


Lachance MA (2004) Here and there or everywhere? Bioscience 54:884-884

Li WKW, McLaughlin FA, Lovejoy C, Carmack EC (2009) Smallest Algae Thrive As the Arctic Ocean Freshens. Science 326:539-539


References


References


7. Statutory Declaration

I, CHRISTIAN WOLF, hereby declare that I have written this PhD thesis independently, unless where clearly stated otherwise. I have used only the sources, the data and the support that I have clearly mentioned. This PhD thesis has not been submitted for conferral of degree elsewhere.

Bremen, October 22, 2012

Signature: ____________________