DIVERSITY AND COMMUNITY DYNAMICS OF PROTISTS AND THEIR VIRUSES IN THE SKAGERRAK

Sandra Gran Stadniczeñko

Dissertation presented for the degree of Philosophiae Doctor (PhD)

2019

Section for Aquatic Biology and Toxicology
Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

© Sandra Gran Stadniczeñko, 2019

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2097

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

“With every drop of water you drink,
every breath you take,
you're connected to the sea.
No matter where on Earth you live.”

Sylvia A. Earle

To my Mom

To my family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Prof. **Bente Edvardsen.** Thank you, Bente, for giving me the opportunity to do this PhD, for your mentorship and all your help in the development of this thesis.

Also, to my co-supervisors **Wenche Eikrem**, **Tom Andersen** and **Josefin Titelman**, for your additional inputs, especially Josefin for your comments and encouragements when it got really difficult.

This thesis was possible due to the financial support from the Research Council from Norway through the HAPTODIV and MICROPOLAR projects. I would also thank the BioMarKs project that provided for some of the samples, and ASSEMBLE for funding for the research stay.

A special thanks to **Elianne D. Egge**, my predecessor in the PhD. If you did not take all those samples, I wouldn't have had anything to work from. I also want to thank you for all your patience and knowledge during our time together at the office when I needed it.

Sissel Brubak, thank you for your help in the laboratory when I was not able to process the samples and to **Rita Amundsen** for assistance during samplings. Thank you both, and **Berit Kaasa**, for all the morning talks in the lunch room that made easy the start of each day.

Jan Heuschele… I do not have enough words to thank you for all your support both in statistics and personally. You did not only save my data when I got stuck with R a million times, but you were always there when I had a bad day, and I really appreciate it. I wish you all the best for the future; you deserve it.

To my AQUA colleagues **Luka, Elisabeth, Marcin**, **Raoul**, **Torben, Fran**, **Roul**, **Daniel, Lina, Maryia**, **Zhanna** and especially **Andrea**, some of you already away from UiO, but I loved our chats at the lunch room which made the days better. Andrea, I made it! And you will too!!!

Luka, Ramiro and **Anders** for your collaborations in my thesis work. I learned a lot form all and each one of you, microscopy… bioinformatics…phylogenies… which are of primary importance in this thesis.

The biggest thanks to my friends and family which were a great support during this PhD. And thus, I will address them in their languages.

Fran Lu, ni te imaginas lo importante que has sido en esta época de subidas y bajadas. Gracias por tu apoyo a diario cuando todo iba bien o mal. ¡Tú mismo pasaste por lo mismo y has sabido animarme cuando ya estaba al límite. Me siento muy afortunada de tenerte como amigo y, para qué mentir, ¡como un hermano! te luvo mucho!

Diana, **Isa**, **Tania**, **Paloma**, **Vane**, **Marta**, **Almu**… Tantos amigos que dejé al venirme a Oslo y que seguís ahí a pesar de la distancia. Muchas gracias a todos por vuestro apoyo. Me siento muy orgullosa de teneros! **Årul Jørgensen** and **Ana Martínez** muchísimas gracias por vuestra ayuda sobre la lengua inglesa de último momento!

Johann, mon amour, merci pour tout ton support et pour m'avoir donné nos fils Julia et Liam. Ils sont le meilleur cadeau que j'ai jamais eu. Vous êtes ce qui me fait sourire chaque jour. Merci beaucoup pour ta patience et pour tous les moments qu'on a vécu ensemble. Je t'aime.

Julia y **Liam,** mis dos amores. Gracias por haber aparecido en mi vida. Sois lo que me hace seguir adelante en cada momento, la alegría de mis días. Os quiero muchísimo.

A mi familia, **papa, Timur** y **Roci**. Por estar siempre allí, por apoyarme en mi decisión para estudiar Ciencias del Mar. Por empujarme a perseguir mis sueños y dejar el nido. Por seguir ahí en los buenos y malos momentos, sobre todo estos últimos meses tan duros para todos nosotros, os quiero.

Babciu i **ciociu,** bardzo wam dziękuję za wsparcie w ciągu tych ostatnich lat, w szczególnie w tych ostatnich trudnych miesiącach.

Mamo… Dziekuje ci za wszystko. Nie masz pojęcia, jak bardzo za tobą tęsknię. Odszedłaś w momencie kiedy najbardziej cię potrzebowałam. Mam nadzieję że jesteś ze mnie dumna, gdziekolwiek jesteś. Kocham cię bardzo.

Oslo, November 2018

Sandra Gran Stadniczeñko

TABLE OF CONTENTS

SUMMARY

Marine coastal areas are among the most productive ecosystems in the world. Protists play key ecological roles in these systems, and it is thus essential to characterise who the main players are. Yet, eukaryotic microbial diversity in marine habitats, especially the smaller nano- and picoplankton, is still poorly described and under-sampled.

The main theme of my thesis is to reveal the diversity and dynamics of protists and their viruses by molecular methods in the Outer Oslofjorden, Skagerrak. This work was also the first to perform studies on both the protist diversity and seasonal patterns and their co-occurrent viruses in Norwegian coastal waters by high-throughput sequencing (HTS, 454-pyrosequencing) throughout two-years (2009-2011). We also aimed to test the 18S and 28S rRNA marker genes to study the haptophyte plankton community and to compare the application of high-throughput sequencing HTS with the scanning electron microscopy method for analysis of the coccolithophore community. Further, we wanted to improve the knowledge about the ichthyotoxic species in the Skagerrak.

Our results have revealed a higher richness of protists and algae-infecting viruses compared to previous surveys carried out for more than a decade in the Skagerrak. This demonstrates that a vast diversity remains to be described both morphologically and genetically. We also revealed which species dominate in the HTS and found taxa not previously recorded in the area by microscopy or taxa novel to science. Further, we provided a curated 28S rRNA gene reference database based on cultures, contributed to linking molecular and morphological data and 28S to 18S rRNA gene sequences without cultured representatives and also improved the metabarcoding methodology. Our studies may serve as a baseline for future surveys and monitoring of planktonic communities to understand the effects of environmental and climate changes.

LIST OF PAPERS

This thesis is built upon the following four original research papers, which will be referred to in the text by their Roman numerals:

I. **Gran-Stadniczeñko, S.**, Egge, E. D., Hostyeva, V., Logares, R., Eikrem, W. & Edvardsen, B. (2018). Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities. *Journal of Eukaryotic Microbiology.* https://doi.org/10.1111/jeu.12700

II. **Gran-Stadniczeñko, S.***, Šupraha, L.*, Egge, E. D., & Edvardsen, B. (2017). Haptophyte diversity and vertical distribution explored by 18S and 28S ribosomal RNA gene metabarcoding and scanning electron microscopy. *Journal of Eukaryotic Microbiology*, *64*(4), 514-532.

* These authors contributed equally

III. Engesmo, A., Strand, D., **Gran-Stadniczeñko, S.**, Edvardsen, B., Medlin, L. K., & Eikrem, W. (2018). Development of a qPCR assay to detect and quantify ichthyotoxic flagellates along the Norwegian coast, and the first Norwegian record of *Fibrocapsa japonica* (Raphidophyceae). *Harmful Algae*, *75*, 105-117.

IV. **Gran-Stadniczeñko, S.,** Krabberød A. K., Egge, E. D., Yau, S., Sandaa R. A. & Edvardsen, B. Seasonal dynamics of algal infecting viruses in the Skagerrak, North Atlantic and inferred interactions with protists. Manuscript.

Papers I-III and Figure 2 are re-printed with the permission of the publishers.

INTRODUCTION

Marine protists and their ecological importance

Marine coastal areas are among the most productive ecosystems in the world. About 33% of the total marine biomass is formed by protists (Bar-On et al. 2018). Protista is an extremely diverse and paraphyletic group of mostly unicellular microeukaryotes, but there are numerous colonial and multicellular taxa. Highly abundant in the euphotic zone, protists play an essential role with diverse ecological functions (Figure 1), and it is thus important to characterise who the main players are (del Campo et al. 2016). Protists: i) serve as the base of marine food webs (Cushing 1989; Massana et al. 2006), ii) influence the productivity and stability of ecosystems across trophic levels (Ptacnik et al. 2008) and, iii) affect the ability of the ocean to function as a natural carbon sink by fixing CO_2 (e.g. De La Rocha and Passow 2007). The grazing activity of heterotrophic protists, together with marine viruses, makes them important control agents of nano- and picoplankton, maintaining populations at relatively stable concentrations in oceanic systems (Suttle 2005; Logares et al. 2012; Prowe et al. 2012). For example, Maar (2003) showed that the grazing impact of ciliates and heterotrophic dinoflagellates in the spring blooms in Skagerrak exceeded that of copepods with a factor of 3 to 4.

The species composition and the relative abundance of the different taxa present in plankton communities undergo seasonal changes, which are of major importance in the coupling between primary producers and higher trophic levels. Identifying the mechanisms that shape marine protist communities has been a centre of interest for researches in the past years. Several mechanisms have been described as causatives: i) external forces including temperature and nutrient supply, transport, physical mixing and inland runoffs (Fuhrman et al. 2015; **Paper I**) ii) intra- and inter-specific competition (Barton et al. 2010), iii) selective grazing from predators (Naselli-Flores et al. 2007), and iv) viral infection (Bratbak et al. 1993; Nagasaki and Yamaguchi 1998; **Paper IV)**.

The oceans are the largest habitat on our planet. Environmental quality, sustainability and functioning of marine ecosystems are closely linked to the species richness and diversity (Loreau et al. 2001). Given the massive and irreversible loss of biodiversity due to human activities and climate change, it is essential to investigate the role and the consequences of biodiversity changes in marine ecosystems (Loreau et al. 2002). Also, the diversity of the protist community impacts it's functioning as different species have different physiology and modes of nutrition. Some researchers have suggested that greater efficiency is reached in ecosystems with higher

diversity (e.g. Barton et al. 2010). Yet, eukaryotic microbial diversity in marine habitats, especially the smaller nano- and picoplankton, is still poorly described and under-sampled.

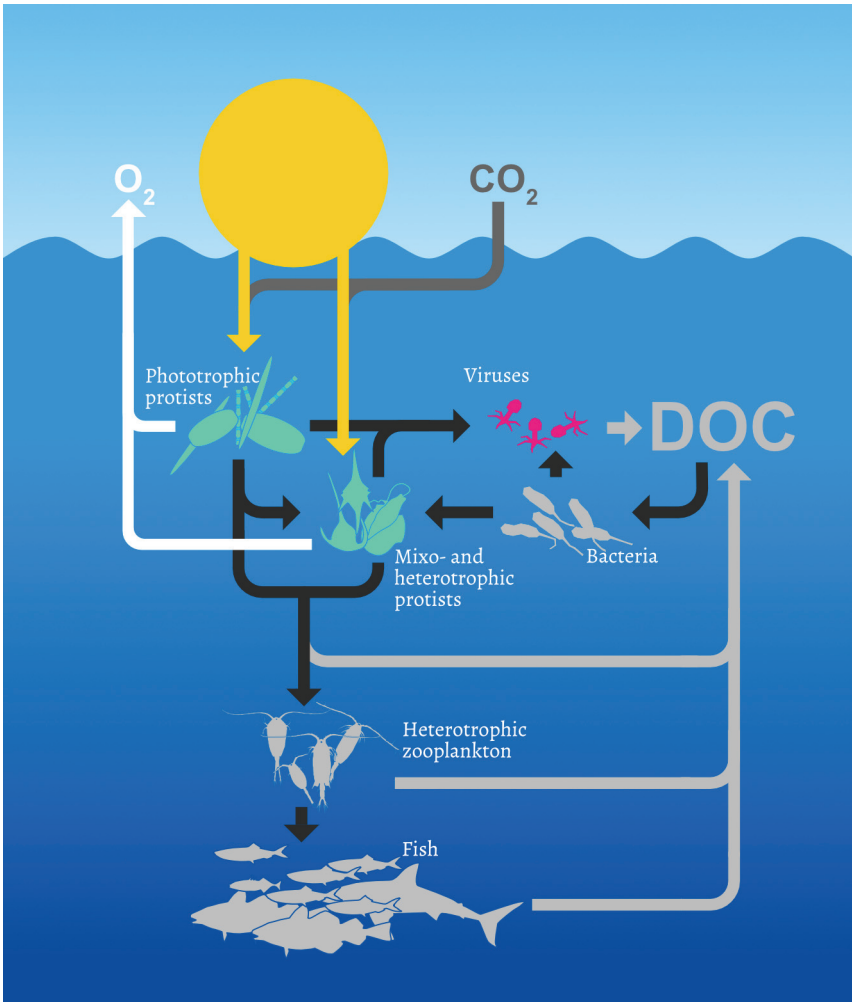


Figure 1. A simplified illustration representing the microbial and viral loops and how they fit in the pelagic carbon cycle. Emphasised are the target groups: protists (green**, Paper I-IV**) and viruses (pink, **Paper IV**). DOC: dissolved organic carbon. Image by Jan Heuschele.

Apart from the intrinsic ecological value that protists possess, their economic value is reflected in the industry. Protists are used in wastewater treatment (Madoni 2003), as biofuel (Hannon et al. 2010), as food in aquaculture (Wikfors et al. 1996) and also in the development of medicines (Jha and Zi-rong 2004). Protist can also have negative impacts on the economy and health by affecting fisheries and tourism (e.g. red tides). Some also act as infectious agents, such as the dinoflagellate *Alexandrium tamarense* and the diatom *Nitzschia pseudodelicatissima* (Enevoldsen 2003).

Protist diversity

Protista is a paraphyletic group that comprises the classically known protozoa, algae and lower-fungi, with more than 200000 described species (Corliss 2002; Adl et al. 2012; Pawlowski 2014). They encompass a wide range of size fractions from 0.2 µm (unicellular protists, Sieburth et al. 1978) up to several meters (multicellular kelp, Schiel and Foster 2015). Different trophic modes, photo-, hetero- and mixotrophy, in addition to decomposer, mutualistic symbiotic and parasitic organisms are found within the protist diversity. Phototrophs encompass the primary producers from different taxonomic groups such as e.g. diatoms, haptophytes, chlorophytes and some dinoflagellates. Despite that larger phytoplankton (micro and nanophytoplankton) may dominate in biomass, small picoeukaryotes (eukaryotes with cell size <2-3µm) have a high turnover rate and productivity representing an important carbon source for the heterotrophic protist community. Protist communities on continental shelves are dominated in biomass by diatoms, dinoflagellates and haptophytes (Simon et al. 2009). Heterotrophic protists include consumers, decomposers and parasites from different taxa such as ciliates, dinoflagellates, cryptophytes, heterokonts and certain euglenoids. Some dinoflagellates and ciliates are mixotrophs being able to both ingest prey and carry out photosynthesis (Maar 2003) allowing them greater flexibility in terms of resource acquisition (Ward et al. 2011).

The taxonomic classification of the eukaryotic kingdom has repeatedly changed (see, e.g., Adl et al., 2012, 2018). Resolving the phylogenetic relationship between the different taxonomic groups is one of the primary topics in evolutionary biology. Protista contains all eukaryotic microorganisms that cannot be phylogenetically placed in other eukaryotic kingdoms (Animalia, Plantae or the Higher-Fungi) but may be closely related to some of them. Pawlowski (2014) classifies members of Protista in seven separate monophyletic groups: Amoebozoa, Opisthokonta, Archaeplastida, Stramenopila, Alveolata, Rhizaria and Excavata, which will be shortly addressed here in addition to the "Hacrobia" supergroup (Figure 2).

Amoebozoa consists of about 2400 species (Pawlowski et al. 2012) classified in diverse groups of lobose naked, testate (cells partially enclosed by a hard shell) and lobose flat-shaped amoeba, but also of cellular and acellular slime molds, in addition to diverse reticulate, filose and amoeboflagellate forms (Tekle et al. 2017). These organisms are mainly free-living in both aquatic and soil environments, but can also be symbionts or parasites (Adl et al. 2012) and thus, be responsible for diseases in different organisms such as the amoebiasis in humans (Entamoeba species, Wilmot 1962).

Opisthokonta is an extensive group comprising species of mostly multicellular eukaryotes, Metazoa and Fungi, as well as the less diverse protistan classes Choanoflagellata and Mesomycetozoa. There are more than 1.5 million species within the Opisthokonta of which only about 700 are protists (Pawlowski 2014). The choanoflagellates get their name "collar-flagellates" thanks to a collar-like structure of microvilli that surrounds a single flagellum, used as a mechanism for filter feeding (Pettitt et al. 2002). These unicellular free-living heterotrophic nanoflagellates have critical roles as bacteriovors in the microbial food web and carbon cycle (Boenigk and Arndt 2002). Their sister group Mesomycetozoa is, however, parasitic and saprophytic, toxic to both aquatic and terrestrial animals (Mendoza et al. 2002).

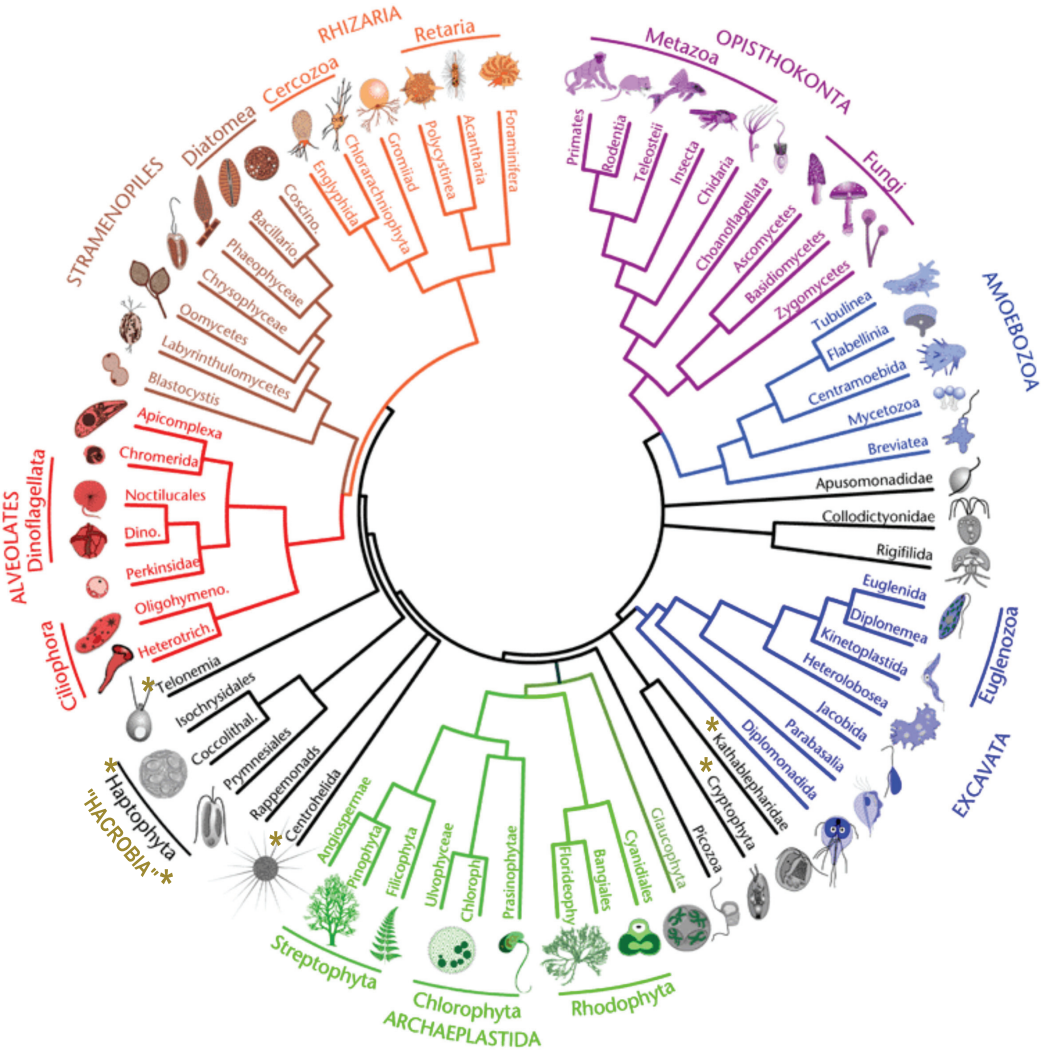


Figure 2. A modified version of the phylogenomic tree of eukaryotes from Pawlowski (2014). It presents the seven different described supergroups where the majority of protists are placed. The modification marks five of the 12 independent lineages as members of the supergroup "Hacrobia" proposed by Okamoto et al. (2009). This image is reprinted with permission from the publisher.

Archaeplastida encompasses the photosynthetic Rhodophyta, Chloroplastida and Glaucophyta. Even though its phylogenetic support is weak (Deschamps and Moreira 2009), this clade is strongly supported based on their plastids acquired by cyanobacterial primary endosymbiosis (Rockwell et al. 2014). Rhodophyta (red algae) comprises more than 5000 species of mostly multicellular marine algae. Chloroplastida includes both the chlorophytes (green algae, primarily aquatic) and the plants (Streptophyta). Lastly, Glaucophyta comprises few freshwater unicellular species.

Stramenopila, also known as Heterokonta, comprise ~30000 species. Most of them belong to Ochrophyta (photosynthetic heterokonts) ranging from unicellular groups such as the Bacillariophyta (diatoms) and Chrysophyceae (golden algae) to the multicellular Phaeophyceae (brown algae; i.e. kelps). Ochrophytes are characterised by bearing two different flagella, one forwardly directed with tripartite tubular hairs (mastigonemes) and one smooth and backwardly directed (Graham et al. 2009). Heterotrophic taxa are also found within Stramenopila, including free-living groups (i.e. Bicosoecida), endocommensal (i.e. Opalinea) and parasitic groups (i.e. Oomycota) (Pawlowski 2014). Also, molecular phylogenetic studies of stramenopiles have revealed several clades of heterotrophic marine flagellate without cultured representatives (MAST, Massana et al. 2004).

Alveolata owe their name to the presence of cortical alveoli, flat vesicles arranged in a continuous layer that supports the membrane (Cavalier-Smith 1991). Including more than 16000 species, Alveolata comprises three major phyla (Pawlowski 2014): i) Ciliophora, which include mostly free-living heterotrophs with abundant cilia covering their surface used for locomotion; ii) Dinoflagellata, mainly motile cells with flagella, have a wide range of trophy (photo-, mixo-, heterotrophic and parasitic or endosymbiont) and can be naked or armoured. Some can cause toxic algal blooms, and some are bioluminescent and; iii) Apicomplexa, a large group of non-motile cells, are obligate parasites which form infectious spores causing severe diseases as the Malaria (*Plasmodium vivax*) or Toxoplasmosis (*Toxoplasma gondii*). Alveolata also comprises the parasitic groups Syndiniales and Perkinsidae and the phototrophic Chromeridae (Pawlowski 2014). Further, phylogenetic analyses have shown uncultured marine alveolate (MALV) lineages in marine planktonic communities (Díez et al. 2001; López-García et al. 2001).

The **Rhizaria** supergroup was exclusively established on molecular data (Cavalier-Smith 2002). It includes phototrophs, heterotrophs, single cells and colonial organisms (Pawlowski and Burki 2009) that are key players in the biological carbon pump at the upper ocean (Guidi et al. 2016). Most rhizarians feed by capturing and engulfing prey with a filose or reticulose

pseudopodia (extensions of their cytoplasm) that can reach sizes of more than a centimetre (Caron 2016). Two major groups compose this supergroup, the Cercozoa and the Retaria, among other minor lineages. The Cercozoa are unicellular, mostly heterotrophic eukaryotes (amoeboids and flagellates), but also include the phototrophic Chlorarachniophyta (Pawlowski 2014). The Retaria clade comprises the Foraminifera and Polycystinea groups, characterised by presenting intricate skeletons (usually of silica) which are ultimately deposited on the sea-bed forming microfossils (Saraswati and Srinivasan 2016). The Acantharea group which forms non-fossilising skeletons is also included in this clade.

The **Excavata** phylogenetic classification is not well supported (Hampl et al. 2009). Nevertheless, most excavates possess a characteristic "excavated" longitudinal feeding groove. They comprise parasitic groups such as the Retortamonadida, Diplomonadida and Parabasalia (Simpson 2003). Euglenozoa includes photosynthetic, heterotrophic and mixotrophic flagellates common in shallow planktonic and benthic ecosystems (Walne and Kivic 1990). Further, the free-living heterotrophic Jakobida are important planktonic and benthic bacterivorous excavates (O'Kelly 1993), able to survive in anoxic-sulphurous waters (Stock et al. 2009).

Finally, the **"Hacrobia**" was first introduced as a supergroup by Okamoto et al. (2009) including about 580 species (Pawlowski 2014). Its classification has been subject to constant changes and is kept here as a practical group, with the name placed in quotes. It includes the well-studied and abundant Haptophyta and Cryptophyta together with the heterotrophs Telonemia, Katablepharida and Centroheliozoa (Okamoto et al. 2009). In the recent taxonomic overview by Adl et al. (2018), Haptophyta and Centroplasthelida were placed in the high-ranked group Haptista and Cryptophyta in Cryptista. Haptophytes are mostly photosynthetic organisms with a wide morphology range that appear as single-celled or forming colonies in a mucilaginous matrix (Edvardsen and Imai 2006). Their distinguishing feature is a unique third appendage used in food handling called haptonema. They play a significant role in the global biogeochemical cycle as primary producers, bacterivores, and formers of extensive blooms in marine ecosystems (Iglesias-Rodríguez et al. 2002; Granéli et al. 2012). Cryptophyta (or Cryptomonads) are biflagellate, mainly phototroph and non-toxic organisms, and thus they are ecologically important as a food source for some aquatic animals and protists (Pedrós-Alió et al. 1995).

Study methods on protists communities

Due to the importance of microorganisms in aquatic environments, interest in the ecology of protist communities has significantly increased in the last years (see, e.g., Massana et al. 2006; Logares et al. 2012; de Vargas et al. 2015; Simon et al. 2015). Such studies usually incorporate multiple techniques to asses protist community structure, function and evolution.

Microscopy

For more than two centuries the study of aquatic protists diversity has been based on light microscopy (LM), which allows for morphological description, size range and motility (e.g. Ikävalko and Gradinger 1997). In the 1950-ies the introduction of the scanning electron microscopy (SEM), allowed the discovery and description of species in the pico- (0.2-2 μm) and nanofraction (2-20 μm) with higher resolution (e.g. Braarud et al. 1952; Deflandre and Fert 1953; **Paper II**). In addition, the introduction of transmission electron microscopy (TEM), allowed the characterisation of inner structures of cells (e.g. Yabuki et al. 2013). Today, both transmission and scanning EM allows for high-resolution images with 100000 times magnification.

Microscopical methods also present important drawbacks that may have consequences for biodiversity estimation. Much smaller sample volumes are used for microscopy (10 ml for LM, **Paper I** and **III**; 300 ml for SEM, **Paper II**) compared to molecular approaches, where several litres of water can be filtered and analysed (20 L, **Paper I-IV**). This only allows observation of a limited number of species overlooking rare taxa in ecological surveys. In addition, some species have fragile structures that may get damaged during the sample preparation, preventing thus their identification. Microscopy may also overestimate protist richness, as different phenotypes of a certain species could be identified as separate species. Besides, the expertise of the taxonomist will influence the taxonomic identification. Regarding the different types of microscopy, other limitations must be addressed. Light microscopy only allows direct identification of organisms in the micro size fraction as the resolution is insufficient for identification of smaller cells (<10 μm, **Paper I**). Finally, while SEM allows detection in all size fractions, this method is time-consuming and thus not suitable for routine monitoring or surveys with a high number of samples (**Paper III**).

Molecular techniques

The advent of molecular surveys on protist communities represents a step forward in community studies. During the last years, new molecular techniques such as DNA sequencing have proven to be indispensable tools for examining the marine microbial diversity (Medlin and Kooistra 2010) and inferring their phylogeny and taxonomic placement. Molecular techniques have revealed the existence of an immense variety of novel marine and uncultured protists (e.g. López-García et al. 2001; Guillou et al. 2004; Massana et al. 2004). Such techniques have the potential to be more specific than traditional methods as it is possible to work with larger sample volumes and allow for obtaining species-specific data without isolation or culture of species (Medlin and Kooistra 2010, **Papers I** and **II**).

The small subunit ribosomal 18S rRNA gene is the most widely used marker as it allows the detection of metabolically active (living) cells, to classify known species present in marine eukaryotic microbial communities and to assess the phylogenetic affiliations of unknown sequences (see, e.g., López-García et al. 2001; **Papers I**, **II** and **III**). It offers diverse advantages: i) its variability is enough to allow for distinction between the different taxa from kingdoms to species, ii) it has conservative regions which are identical in most eukaryotes which allow amplification by PCR using universal primers, iii) it has the same function in all organisms, iv) there is no evidence for lateral gene transfer and v) it is present in multiple copies within each cell making PCR easy. Sequencing of the 18S rRNA gene has allowed scientists to create large databases with reference sequences that are available to compare environmental sequences to those of known species such as the PR2 (Guillou et al. 2013). However, the 18S rRNA gene is not sufficiently variable to resolve interspecies relationships in several taxa (Pawlowski et al. 2012). Some alternative markers are the large subunit 28S rRNA gene, valuable for certain groups such as ciliates (Gentekaki and Lynn 2009), diatoms (Hamsher et al. 2011) and haptophytes (**Paper II**), and the internal transcribed spacer (ITS) rDNA which is the main fungal barcode marker (Schoch et al. 2012).

High throughput sequencing

High throughput sequencing (HTS) of rDNA marker regions is also referred to as **metabarcoding**. HTS is a powerful technology that enables massive parallel sequencing of clonally amplified DNA templates or single DNA molecules (Margulies et al. 2005). The result is hundreds of megabases to gigabases sequences (reads) in a single run, which are consecutively assembled into genotypes by powerful computer tools. This technology grants faster and

cheaper sequencing than traditional Sanger DNA sequencing and has allowed for a broad description of the environmental microbial diversity. HTS has also revealed uncultured and rare taxa in different habitats (e.g. Bates et al. 2013; Debroas et al. 2015; de Vargas et al. 2015). HTS provides qualitative genomic information and sequence read abundances can be used to infer quantitative information on microbial taxa (Egge et al. 2013; **Papers I-IV**). However, PCR and sequencing processes involved in HTS are known to incorporate errors, such as chimeras, that can create spurious phylotypes (Quince et al. 2009; Huse et al. 2010). It is thus important to identify and remove these errors through rigorous bioinformatic filtering of the data (Quince et al. 2011; Schloss et al. 2011).

Many of the reads acquired in aquatic protist studies represent known taxa that are already described from microscopy studies, but for which no DNA sequences or cultures are available. Combining HTS with microscopy allows for linking genotype to a morphologically and genetically characterised organism, culture or isolated cells from a natural sample. However, classification of genotypes based on HTS can often only be done to higher taxonomic levels (e.g. family, genus or a phylogenetic clade), and not to species level (e.g. Bachy et al. 2011) due to lacking reference DNA sequences (**Paper I** and **II**). A combination of high-resolution microscopy with environmental sequencing or metabarcoding, phylogenetic analysis and a curated DNA reference sequence database improves the resolution of microbial taxon identification and distribution patterns (**Papers I** and **II**).

Quantitative polymerase chain reaction (qPCR)

The qPCR is a variant of the polymerase chain reaction (PCR) used to amplify and simultaneously quantify DNA and RNA molecules (Heid et al. 1996). It is often used for quantification of toxic algae in monitoring programs as it allows detecting minimal amounts of DNA/RNA present in the samples (Zamor et al. 2012, **Paper III**). This quantitative method is fast, extremely accurate and cost-effective and can be applied in preserved environmental samples (**Paper III**). A major drawback is that it only allows for detection of one or few species at a time.

Fluorescence in situ hybridisation (FISH)

FISH is a combination of molecular and microscopical techniques that allows visualisation of the cells as well as locating sequences of nucleic acids in morphologically preserved cells using a target labelled DNA or RNA probe (Hosoi-Tanabe and Sako 2006). This technique has been widely used in microbial ecology studies for easy microalgae identification (Rublee et al. 1999;

Hosoi-Tanabe and Sako 2005; Dittami et al. 2013). As qPCR, FISH is highly specific and allows a rapid identification but is an expensive and time-consuming technique (Hosoi-Tanabe and Sako 2006).

Methodology workflow used in this thesis

A schematic representation of the different steps in community studies is presented in Figure 3, and detailed protocols are described in **Papers I-IV**. Briefly, monthly water samples and hydrographical data were taken between September 2009 and June 2011 at two different depths (1m and deep chlorophyll maximum, DC) at the OF2 station, Outer Oslofjorden, Northern Skagerrak. For molecular analyses, 20 L of water were collected with Niskin bottles, prefiltered with a 45µm nylon mesh and filtrated with a peristaltic pump to obtain the size fractions 45-3 µm (nanoplankton), 3-0.8 µm (picoplankton) and 0.45-0 µm (virus). The 18S and 28S rRNA genes for protists and MCP gene for virus were amplified by PCR and 454-pyrosequenced. A posterior bioinformatic and statistical pipeline was performed in QIIME and R. For light microscopy, 100 ml were collected from the Niskin bottles and examined using Utermöhl's sedimentation technique (Utermöhl 1958). Phytoplankton taxa were identified according to Throndsen et al. (2007), and biovolumes were estimated using the HELCOM 2006 protocol (Olenina et al. 2006). Three hundred ml of water sample were collected at eight depths (1, 2, 4, 8, 12, 16, 20, and 40 m) for analysis under a Zeiss Supra35-VP scanning electron microscope. Quantitative analysis of the coccolithophore community was conducted following Bollmann et al. (2002).

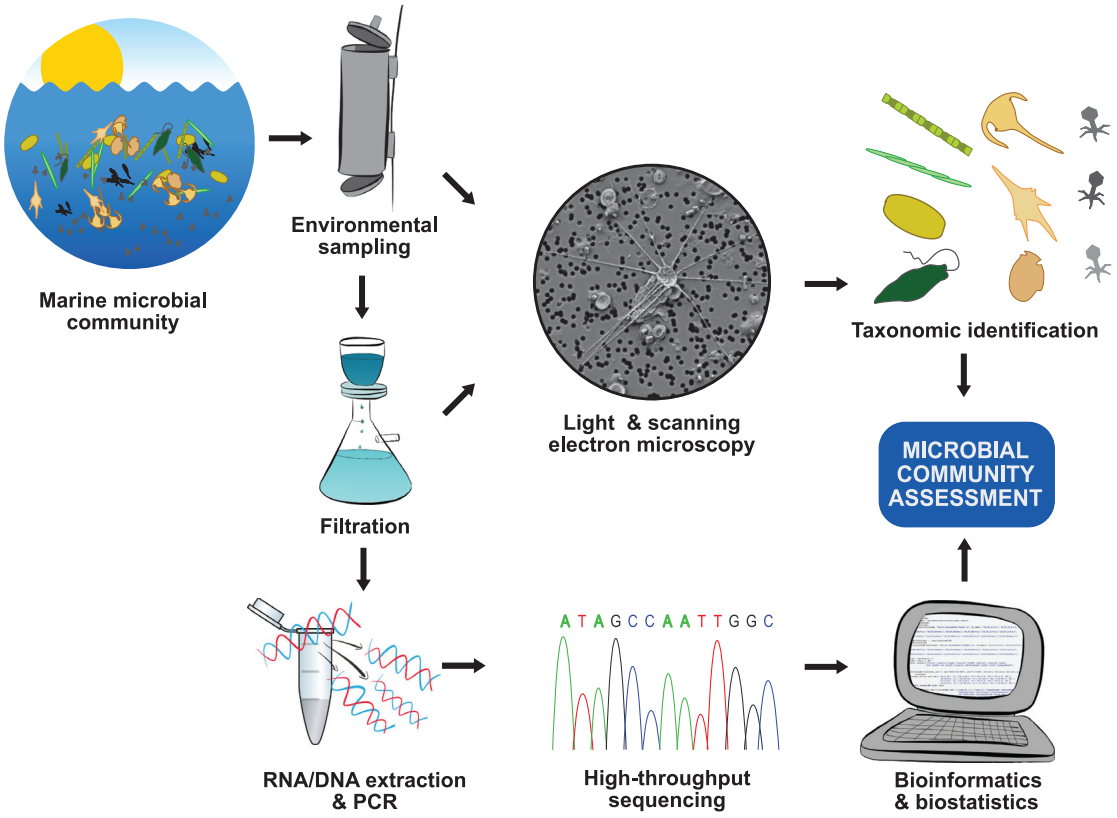


Figure 3. Illustration of the study workflow that was applied for the protist and viral communities in **Papers I-IV**.

Diversity and dynamics of protists in the Skagerrak.

The rise and improvement of molecular methodologies during recent years have allowed a more profound characterisation of new microeukaryote species in marine habitats (Massana 2015). While studies targeting marine protistan diversity have increased in different areas (e.g. Countway et al. 2005; Stock et al. 2009; Bachy et al. 2011), only the haptophyte community has been inferred by these methods in Norwegian coastal waters (Egge et al. 2013, 2015a, b). Previous studies on protist succession and species composition in the Skagerrak area have been carried out in several taxonomic surveys for over a century (e.g. Hjort and Gran 1900; Braarud and Bursa 1939; Braarud et al. 1953; Kuylenstierna and Karlson 1994) based on light, electron and epifluorescence microscopy and flow cytometry, and mostly focusing on photosynthetic taxa (Backe-Hansen and Throndsen 2002; Bratbak et al. 2011). Such methods have, however, many limitations when detecting the full diversity.

The supply of water from the Baltic and Atlantic brings in allochthonous plankton that contributes to a species-rich protist community in the Skagerrak area (Andersen et al. 2001). The research in **Papers I–IV** is all based on work in the Outer Oslofjorden (Box 1). Furthermore, in the Outer Oslofjorden changes in hydrodynamical and meteorological processes together with nutrient availability, grazing pressure, competition and viral infection (**Paper IV**), cause inter-annual variations in the protist species composition (Braarud et al. 1953) with different environmental preferences.

More than 700 species from diverse phytoplanktonic groups are present in the Norwegian coastal waters (Throndsen et al. 2007). The "Checklist of phytoplankton in the Skagerrak-Kattegat" (including heterotrophic protists) has registered a total amount of 178 species of dinoflagellates, 177 species of diatoms, and 291 species from other algal/protist groups (such as euglenoids, ciliates and haptophytes among others, Kuylenstierna and Karlson 2006). However, due to the small size and fragility of pico-and nano-plankton, some organisms are sometimes difficult to identify to the species level. Recently, molecular identification of these tiny eukaryotic organisms has elucidated in greater detail a vast diversity with less biased qualitative approaches of the microbial communities even at low relative abundances (Egge et al. 2013; Nitsche et al. 2017; **Paper I** and **II**). Seasonal studies based on light and electron microscopy of the Skagerrak plankton community have shown important protists genera to be common in the area, such as *Micromonas, Chaetoceros*, *Prymnesium*, *Chrysochromulina, Gyrodinium* and *Heterocapsa* (Kuylenstierna and Karlson 1994). Recent metabarcoding studies at the Skagerrak-Kattegat targeting biodiversity of small eukaryotes of specific groups have found cryptophytes and choanoflagellates (Nitsche et al. 2017; **Paper I**) to also be abundant in

this area. As new taxa are discovered and new sequences are added to reference libraries, taxonomical re-assignments of OTUs may be needed to get an improved overview of the protistan community (**Paper I** and **II)**. For example, the picoflagellate *Micromonas commoda,* belonging to class Mamiellophyceae (Chlorophyta)*,* was recently separated from *Micromonas pusilla* (Simon et al. 2017) and was found to be an abundant species in the Skagerrak area (**Paper I**). Still, the protist community at the Skagerrak encompasses a large unknown diversity, highlighting the need for more DNA-reference sequences to be able to link a DNA sequence or genotype to a protist species (**Papers I**, **II**, **III**).

Box 1. The Outer Oslofjorden -Study site

The Outer Oslofjorden is situated in the northern part of the Skagerrak, with the Drøbaksundet, a sill with a threshold depth of 19 m, as the upper limit, and stretching down to Færder fyr (Figure 4). The Skagerrak, off the coasts of Norway, Sweden and Denmark, undergoes strong seasonal changes in meteorological and hydrological conditions, and irradiance resulting in a heterogeneous environment. Hydrological conditions such as brackish and saline water currents (e.g. the Baltic current and the North Atlantic current), inland runoffs and tidal forces, together with differences in wind conditions lead to changes in circulation patterns and physicochemical properties at the Skagerrak coastal water masses (Sætre 2007).

The Norwegian coast of the Skagerrak area holds two-thirds of the Norwegian population. It is thus a resource for recreation and is holding important harbours. Runoffs from industry, forestry, agriculture and wastewater from the population cause major environmental impacts (Walday et al. 2017). The pollution derived from human activities causes stress to marine ecosystems of the Skagerrak such as eutrophication (Boesch et al. 2006) and reduction of fish populations (Kålås et al. 2006). In addition, climate change is affecting the environment by increasing freshwater runoff and eutrophication, extending species distribution further north (e.g. Lindley & Batten 2002) and modifying the general biodiversity (e.g. Norderhaug et al. 2015). Finally, the presence and blooms of toxic algae in the Norwegian coastal waters can affect the aquaculture and recreation in this area. Monitoring programs are therefore necessary to identify the state of the marine environment, and detect and prevent environmental problems in the Skagerrak (Walday et al. 2017).

The phytoplankton composition at Outer Oslofjorden monitoring location (OF2) in the Skagerrak is considered to represent that in the Norwegian coastal waters along the northern Skagerrak coast (Braarud and Bursa 1939; Dragsund et al. 2006). The OF2 is thus a monitoring station of water quality in Outer Oslofjorden (Walday et al. 2017).

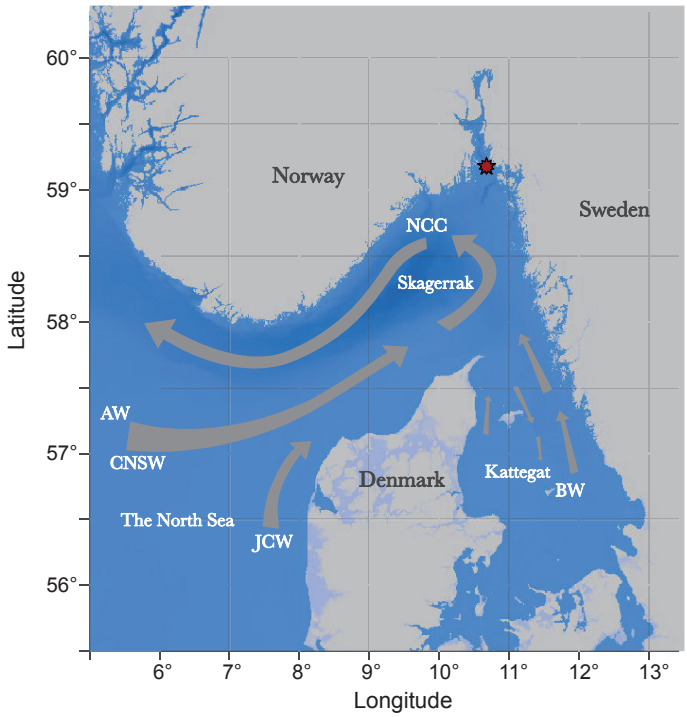


Figure 4. Surface circulation pattern in the Skagerrak and Kattegat off the coast of Norway, Sweden and Denmark. The OF2 station (59.17 N, 10.69 E) is indicated with a star. NCC: Norwegian Coastal Current; AW: Atlantic Water; CNSW: Central North Sea Water; BW: Baltic Water; JCW: Jutland Coastal Water. The map was obtained from https://www.gebco.net/ using the ggmap and marmap packages in R (R Development Core Team 2017).

Protist-infecting viruses

Marine viruses are a highly abundant, dynamic and diverse component of the planktonic community (Bergh et al. 1989). The total viral abundance in the marine environment is highest in productive coastal waters (~10^8 viruses ml-1) decreasing with depth and distance from the shore (Suttle 2005). Viruses play key roles in marine ecosystems. They cause significant mortality in microbial communities and thus alter hosts abundance and distribution, and can sustain the coexistence of competing species (Thingstad 2000). Viruses also affect the nutrient and carbon cycling (Figure 1) by converting microbial biomass into dissolved and particulate organic matter (Bratbak et al. 1994), that is used by heterotrophic prokaryotes and other degraders (Suttle 2005). Further, viruses can prevent and terminate algal blooms (Castberg et al. 2001). However, by lysis of certain phytoplankton, such as the coccolithophores, viruses increase dimethyl sulfide (DMS) emission. This volatile compound induces atmospheric cloud formation, modifying the planetary albedo and indirectly affecting the climate (Charlson et al. 1987).

Novel molecular techniques allow for investigating the diversity of viruses and their hosts in natural habitats. Most viruses have shown to be host-specific (Short 2012), but viruses able to infect and lyse different protist species have also been described (Johannessen et al. 2015). Most of the isolated viruses infecting photosynthetic protists (algal infecting viruses) are large double-stranded DNA (dsDNA) with genome sizes of >200 kb, hence referred to as giant viruses. They are taxonomically assigned to the *Phycodnaviridae* (Van Etten 2000) and the *Mimiviridae* (La Scola 2003) families. These two families are ubiquitous and have been isolated in freshwater environments (Short et al. 2011). Whereas all the *Phycodnaviridae* viruses infect algae, the *Mimiviridae* have both photosynthetic (i. e. *Chrysochromulina ericina* virus and *Pyramimonas orientalis* virus, (Sandaa et al. 2001) and non-photosynthetic protists as hosts (i.e. *Acanthamoeba polyphaga*, Suzan-Monti et al. 2007). However, a large number of virus taxa within these two families are still not genetically characterised.

Host-viruses interactions have been described as: i) acute boom-bust infections, where a specific virus terminates a dense host bloom within hours, and ii) persistent infections that allow the coexistence between the host and its viruses (Sandaa and Bratbak 2018) due to the host's viral resistance, immunity and/or strain specificity, or the virus becoming less virulent (Dimmock et al. 2016). These dynamics can be measured by correlating viral diversity and abundance with that of the hosts. Several studies have focused on haptophyte-virus diversity and dynamics in West Norwegian Coastal waters, especially in the Raunefjorden (Larsen et al. 2004; Johannessen et al. 2017). Nonetheless, **Paper IV** is the first study aiming the complete protist-viral community diversity and seasonal dynamics in Norwegian waters.

AIMS OF THIS THESIS

The main objective of this thesis was to study the diversity and dynamics of the protist community and their viruses in the Outer Oslofjorden (Skagerrak) by the use of molecular methods such as high-throughput sequencing. The specific aims for the different papers were:

- The aim of **Paper I** was to study the changes of the protist community in the Skagerrak through the seasons by combining high-throughput sequencing and microscopy, and which are the main abiotic drivers for these changes. We also aimed to reveal which species dominate in the HTS and if we find taxa not previously recorded and reported in the area by microscopy, or taxa novel to science.

- In **Paper II** we wanted to explore the haptophyte community at the Outer Oslofjorden using high-throughput sequencing (HTS), comparing the 18S and 28S rRNA marker genes. We also wanted to compare the application of HTS with the scanning electron microscopy method for analysis of the coccolithophore community.

- **Paper III** aimed to develop a rapid detection and enumeration method for ichthyotoxic species, which can be used in algal monitoring as a complement to LM. We further wanted to improve our knowledge about the seasonal distribution of ichthyotoxic species present in the Skagerrak.

- The aim of **Paper IV** was to reveal the diversity and community dynamics of algae-infecting viruses in the Outer Oslofjorden by metabarcoding of the major capsid protein (MCP) gene. We also aimed to detect co-occurrences between the viruses with various protists and if co-occurrences give information about potential virus-algal hosts relationships.

RESULTS AND DISCUSSION

The planktonic communities at the Skagerrak are highly diverse and dynamic.

One of the major obstacles microbial ecologists nowadays face is the lack of consensus when studying natural communities (Goodwin et al. 2017). Differences in the choice of gene markers, sequencing platforms and bioinformatic pipelines hamper comparisons between studies. Several metabarcoding studies have compared DNA vs RNA templates with the intent of indicating which molecule best represents the real microbial community. Some of these studies did not find any significant differences in the community structure targeting special groups (e.g. haptophytes, Bittner et al. 2013) or the entire protist communities (Massana et al. 2015). Egge et al. (2013), however, did capture more diversity in the haptophyte community with RNA than with DNA as a template. This may be due to the bias of the variability in rDNA copy numbers among taxonomic groups (Not et al. 2009), where protists with large copy numbers overshadow those with small genomes and few copies. Also, DNA from dead organisms can be detected. This is not the case of RNA which is metabolically active at the time of collection (Stoeck et al. 2007). RNA has thus been the template chosen for the present thesis (**Papers I-IV**). However, RNA requires reverse transcription into cDNA which may introduce additional chimeras (Egge et al. 2013) that need to be identified and removed.

The most widely used target region is the 18S rRNA gene (encoding the small subunit ribosomal RNA), which consists of ~1800 bp in protists. Since almost no HTS sequencing technology is able to analyse the complete 18S rDNA gene, several studies on protist communities have performed comparisons between the two most used marker gene regions of 18S rRNA gene, the V4 and V9. Compared to the V9 (ca 150 bp), the V4 region is longer (ca 400 bp) and gives more phylogenetic information. Results demonstrated that estimates depended on the region targeted and varies with taxa. Giner et al. (2016) found that the V4 region gave a better estimation of the relative cell abundances of chlorophytes whereas the V9 was better for some stramenopile groups (MAST-4, MAST-7 and Pelagophyceae). Dunthorn et al. (2012) recommend the V4 region in ciliate studies. Furthermore, as most of the 18S rRNA gene reference sequences in public databases do not cover the entire V9 region (Tragin et al. 2018), the difficulty of identifying rare or uncultured taxa is more considerable with this region.

Another molecular marker used in protist surveys is the 28S rRNA gene (encoding the large subunit ribosomal RNA). The 28S rRNA gene is longer (> 3000 bp in protists) than the 18S and its mutation rate in some hypervariable regions (D1-D2) is higher than the 18S, making

parts of this gene a useful molecular marker for monitoring programs. However, more reference sequences are currently available for the 18S than for the 28S rRNA gene.

In **Paper I** we used high-throughput sequencing to study the diversity and temporal and spatial dynamics of the total protists assemblage at the Outer Oslofjorden (Skagerrak). We analysed the V4 region of the 18S rRNA gene on 21 sampling dates (2009-2011) at two depths, using the primers described by Stoeck et al. (2010). Due to the low cell abundances in aquatic habitats, large sampling volumes are necessary to resolve the full planktonic protist community richness (Rodríguez-Ramos et al. 2014). To do so, and to include species that are overlooked by microscopy, we sampled 20 L of seawater for each date and depth (**Papers I-IV**). We here focused on the smaller protists that are least known from previous microscopical studies, the pico-, nano- and smallest micro-plankton communities (about 0.8 - 45 µm in size) (**Papers I** and **II**). Important species of the dinoflagellate, diatom and ciliate groups larger than 45 µm were mostly excluded in this study. However, DNA from broken cells during filtration may have been included.

A similarity value of 98% might be appropriate to distinguish protists at species-level (Caron et al. 2009). Therefore, we defined an operational taxonomic unit (OTU) as the taxonomic entity represented by 18S rRNA gene sequences sharing >98% identity, which was used as a proxy for species (**Paper I**). However, in some taxonomic groups, such as haptophytes (e.g. Egge et al. 2015a), and dinoflagellates (Edvardsen et al. unpubl. data) closely related species may differ by less than 1% at this marker. Further, due to the possibility of obtaining spurious data in our data samples (e.g. dead DNA, leaks during filtration, sequencing errors or chimeras), we removed all the OTUs that contained less than ten reads in the total dataset.

We found a very diverse and dynamic protistan community at the Outer Oslofjorden (Figure 5), with a total of 2032 OTUs, compared to the more than 700 species that Throndsen et al. (2007) estimated with morphological observations for the Norwegian coastal waters. Our results almost doubled the ca 1200 species of pelagic and benthic protist recorded based on microscopy in Norwegian marine waters, according to the Norwegian Species Information Centre (Artsdatabanken 2018, Antall arter i norsk natur 2016), of which 1020 belong to a phylum with microalgal representatives.

However, taxonomical assignation revealed that several OTUs represented the same species, reducing thus the real number of taxa in the studied area. The detailed diversity and seasonality of the Outer Oslofjorden protistan community are presented in **Paper I.** In the following I focus on the most abundant and newly recorded taxa.

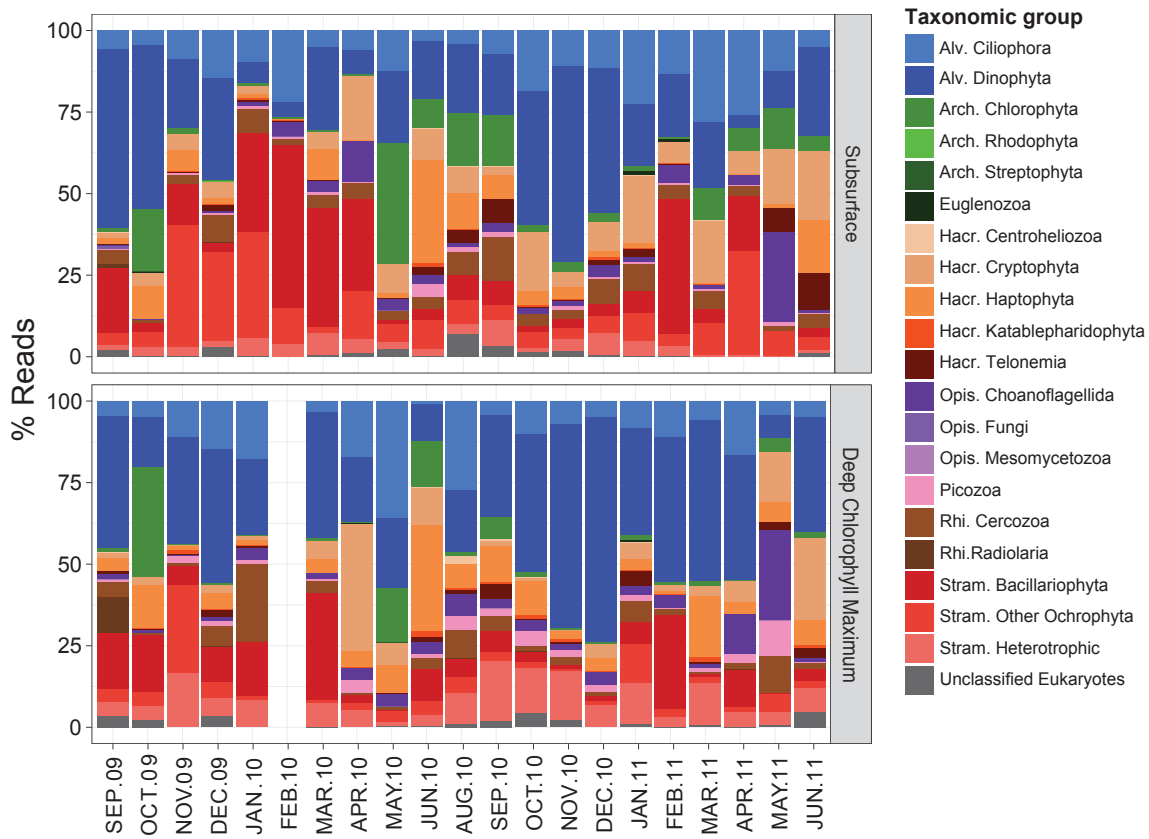


Figure 5. Succession of proportions of reads of the 18 major taxonomic groups across the 21 temporal samples at the Outer Oslofjorden. Alv.: Alveolata; Arch.: Archaeplastida; Hacr.: Hacrobia; Opis.: Opisthokonta; Rhi.: Rhizaria and Stram.: Stramenopila.

Protist and viral diversity

The Outer Oslofjorden protist community was diverse, with representatives of the supergroups Alveolata, Archaeplastida, Excavata, Opisthokonta, Rhizaria and Stramenopila, together with the "Hacrobia". The latter was proposed by Okamoto et al. (2009), but its classification is not well established. We reported 69 potentially new species and 40 potentially new genera for the Skagerrak area that are not registered in the Nordic Microalgae and Aquatic Protozoa (NOD) database (Karlson 2015) nor in the Norwegian Biodiversity Information Centre (Artsdatabanken 2018). Some of those newly recovered species might have been misidentified or overlooked in past microscopical surveys due to their small size or fragility. From our total dataset, only 4% of OTUs could not be classified further than to the eukaryotic kingdom. Also, comparing to the PR2 database, we could only taxonomically assign the 28% of our OTUs to genus level and 19% to species level (representing 155 genera and 144 species of cultured taxa). This indicates a lack of gene reference sequences from isolated species.

Dinoflagellates were the most diverse and abundant group, followed by diatoms and haptophytes. This accords with Not et al. (2012), who described that dinoflagellates reach their highest abundances in estuaries and coastal waters. The dinoflagellates that accounted most to our dataset were of the genera *Karenia, Karlodinium* and *Akashiwo*. Species of these genera have previously formed harmful algal blooms (HABs) in the Oslofjorden (Throndsen et al. 2007, **Paper III**). The most abundant OTU was identical to the sequence of *K. papilionaceae* (HM067005), a species that may cause neurotoxic shellfish poisoning (NSP, Enevoldsen 2003). This species has not previously been recorded in the Skagerrak or Norwegian waters. The recorded species *Karenia brevis* has however been lately separated into four different *Karenia* spp. including *K. papilionaceae*. Also, cultures of this species have shown a second stage of small cells not identifiable under the light microscopy (Carmelo Tomas, pers. commun.). These aspects suggest that *K. papilionaceae* may have been overlooked in the past. In **Paper III** we also recorder high abundances of *Karenia mikimotoi* and *Karlodinium veneficum* by both, LM and qPCR methods.

Similar was the case of the newly described Chlorophyta species *Micromonas commoda,* recently separated from *Micromonas pusilla* (Simon et al. 2017). For the first, time we recorded *M. commoda* from the Skagerrak and found it among the most abundant taxa in our dataset (**Paper I**). *Micromonas pusilla* was found to dominate the eukaryotic picoplankton community in the North Atlantic coastal and Arctic waters (Not et al. 2004, 2005). Therefore, we here suggest that it is *M. commoda* and not *M. pusilla* that dominates in the Oslofjorden.

Although not abundant, we recorded for the first time in Norwegian coastal waters the Raphidophyceae *Fibrocapsa japonica*, both with qPCR and 454 pyrosequencing (**Paper III**) a species not previously found in Norwegian coastal waters. This species has however been previously recorded off the German Bight (Rademaker et al. 1998) and Swedish west coast (www.smhi.se/klimatdata/oceanografi/havsmiljodata), which may suggest a northward dispersal. Our finding is thus of interest for future monitoring programs of HABs in the Norwegian coastal waters as *F. japonica* may pose future challenges to fish-farmers, wildlife and tourism.

We also found several clades of uncultured marine alveolates (MALV I-V), taxonomically placed in the clade Syndiniales (Guillou et al. 2008). They are parasitic organisms and may have a key role in regulating high-biomass populations of HABs forming species (Chambouvet et al. 2008). López-García et al. (2001) were the first to describe the MALV group in marine 18S rRNA gene molecular surveys by environmental clone libraries. Members of the uncultured marine stramenopiles (MAST, Massana et al. 2004) were also found at our sampling site. MAST are generally abundant in clone libraries, which suggests that they may strongly

influence the food web and the biogeochemical cycles (del Campo et al. 2016). Different clone library studies show that MALV and MAST clades are dominant in marine surveys (e.g. Koid et al. 2012; Massana et al. 2014).

With the use of haptophyte specific primers and SEM observations, we were able to determine new Haptophyta taxa (**Paper II**). More than half of our 18S rRNA OTUs were recorded for the first time for the Outer Oslofjorden and Skagerrak. We wanted to address special attention to the rare coccolithophore *Tergestiella adriatica* considered to be extinct after the Cretaceous–Paleogene (K/Pg) extinction event (66 million years ago). However, *T. adriatica* was recently re-discovered at the coasts of Japan and Croatia (Hagino et al. 2015), and now in the Skagerrak (**Paper II**). With our SEM observations, we were able to link morphological and molecular data without cultured representatives such as *Braarudosphaera bigelowii* (Figure 6a) and recorded, for the first time in this area, six coccolithophore species such as the *Calciopappus caudatus* (Figure 6b).

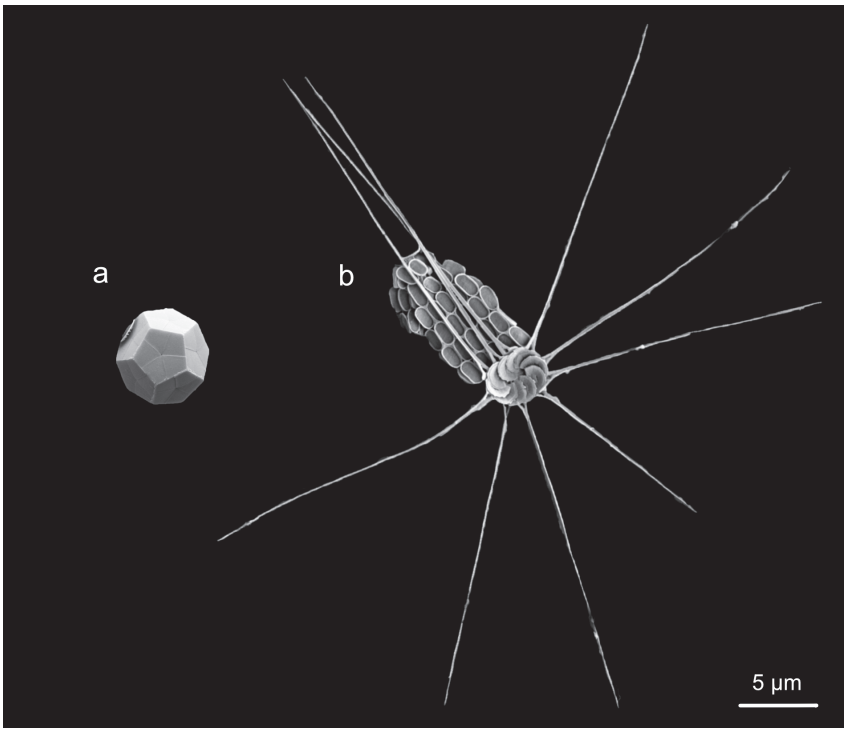


Figure 6. Scanning electron micrographs of two coccolithophore morphotypes detected at the Outer Oslofjorden: a) *Braarudosphaera bigelowii* and b) *Calciopappus caudatus*. While *B. bigelowii* was also detected with SEM and the 18S rRNA marker gene, *C. caudatus* was only found by SEM.

We have also provided new knowledge about the virus community infecting photosynthetic protists (algae) in the Skagerrak by targeting the major capsid protein (MCP) marker gene and defining OTUs at 97% sequence similarity (**Paper IV**). This is a broadly used marker when

targeting *Phycodnaviridae* and *Mimiviridae* (e.g. Johannessen et al. 2017). Within our most abundant OTUs, we detected viruses infecting different hosts previously described in the Norwegian coastal waters such as *Micromonas pusilla* virus (MpV1), *Haptolina hirta* virus (HhV-Of01), *Ostreococcus* spp. viruses (OtV, OlV and OmV viruses) and *Chrysochromulina ericina* viruses (CeV*)*. We also obtained many clades of OTUs without any cultured or environmental reference sequences, which also was the case in studies from other areas (see, e.g., Clerissi et al. 2015).

Our findings are in accordance with previous surveys pinpointing the lack of characterised organisms and the need for more cultured and characterised reference strains of protists (Koid et al. 2012; Massana et al. 2014) and viruses (Clerissi et al. 2015), and can set a baseline for future studies to broaden the community knowledge. **Papers I**, **II** and **IV** revealed a higher protist and viral richness than previous studies in the Skagerrak.

As sequencing techniques are evolving and new primers targeting specific groups are continuously being developed, we may even obtain a better estimation of the real microbial diversity in future surveys. The choice of primers is essential to address since differences in targeted groups have been observed. In **Paper I** and **III** we chose the universal eukaryotic primers by Stoeck et al. (2010) to target the V4 region. These primers allowed us to obtain good phylogenetic information of our protist community, but also some drawbacks were found. Due to mismatches in those primers, they failed to target some taxonomic groups such as the Foraminifera and to some degree the Haptophyta, leading to an underrepresentation of the total protist diversity. The same was observed in **Paper II** where we compared 18S and 28S Haptophyte specific primers. Here, the 28S primers failed to pick up the Noelaerhabdaceae, which is the most abundant coccolithophore family in our dataset according to the 18S primers. Besides, when studying the marine viral community, the major capsid protein (MCP) primers (used in **Paper IV**) are better suited for capturing the *Mimiviridae* family, whereas the *Phycodnaviridae* family is better captured by the DNA polymerase B (polB) primers (Wang et al. 2015). The primer choice is thus a critical aspect since certain important groups may be overlooked. A final issue to consider in the future is the use of OTUs which can group different species as one. This problem has been addressed by shifting from OTUs to amplicon sequence variants (ASVs) methods. These methods use all unique biological sequences distinguishing sequence variants differing by as little as one nucleotide (Callahan et al. 2017).

Seasonality and driving forces

To understand the factors that drive changes in the protist community, it is essential to study their temporal trends (**Paper I** and **IV**). The review by Fuhrman et al. (2015) indicates that different biotic and abiotic environmental factors drive changes in microbial communities over multiple timescales ranging from hours to years (**Paper I**). We addressed changes in protistan community composition during a two-year period and in two different depths within the well-lit euphotic zone (subsurface: SS and bottom of the deep chlorophyll maximum: DC). We observed marked seasonal and spatial variations in protist composition and relative abundances with differences observed between the two depths. The community presented highest diversity in the summer-early autumn season, probably due to allochthonous plankton brought by the North Atlantic current into the Oslofjorden (Andersen et al. 2001). Further, all our most abundant OTUs did peak in abundances two or more times during the study period except for one, indicating that they present a seasonal dynamic. Diatoms dominate during the spring bloom whereas dinoflagellates have their highest proportion in autumn-winter and haptophytes peaked in June. This seasonal succession in composition and abundance is well established in previous microscopy-based surveys in temperate coastal ecosystems (e.g. Hasle and Smayda 1960; Thomsen et al. 1992).

We also assessed the temporal distribution of the different trophy modes. Our results are in accordance with the patterns found in the TARA Oceans expedition (de Vargas et al. 2015), as we observed that the trophic groups showed a clear seasonality, and the heterotrophs were more diverse than the autotrophs. The latter presented higher abundances at the SS than at the DC. Permutational multivariate analysis of variance (PERMANOVA) analyses indicated that a low percentage (28%) of the temporal variation in the protist community is driven by environmental factors, mainly temperature and salinity. Similar findings were recently found in other temperate areas where a large part of the community variance could not be explained by environmental factors (Giner et al. 2018). Both temperature and salinity were negatively correlated with nutrient concentrations in the Outer Oslofjorden. Diatoms and dinoflagellates dominated and drove the general variation of the protist community (**Paper I**). Simon et al. (2015) proposed that the lack of including biotic factors in the analyses (e.g. mutualism, predation, parasitism and virus infection), may be the reason of the low correlations detected. Such factors were, however, not included in our analysis. Future studies need thus to incorporate more biotic and abiotic environmental factors to describe what drives most of the variability of the protist communities at the Skagerrak.

Comparison between HTS relative read abundance and light microscopical cell counts method (Utermöhl 1958) for the phytoplanktonic community showed over and underrepresentation of the target groups. Microscopy methods allowed good resolution for larger protists but are insufficient for the smallest ones (Massana 2015, **Papers I** and **III**). This, together with the choice of primers for HTS, the fragility of some cell structures, prefiltration and differences in sampled volumes for both methods are some reasons for such over- and underrepresentations (**Paper I** and **III**). However, we observed a similar phytoplanktonic seasonal pattern for both HTS and light microscopy (**Paper I**). These findings are important in the ongoing transition in monitoring methods from the Utermöhl method to molecular biological methods.

Finally, we studied the seasonal changes in the viral community, together with possible hosts (**Paper IV**). The viral community presented a clear temporal variation, but not a recurring seasonal pattern as in some other studies (e.g. Pagarete et al. 2013). This lack of seasonality could be explained by the primer choice, PCR-biases towards amplification of specific genotypes or the interannual variation of the host community structure (Short 2012). The most observed pattern in **Paper IV** was a coexistence between virus and host during long periods. A possible explanation can be long virulence periods were the virus coexists with its hosts, only infecting a part of the population to ensure virus proliferation (Sandaa and Bratbak 2018). Johannessen et al. (2015) also pointed out the possibility of viruses to infect several similar or even different hosts, allowing them to proliferate on different host species.

We also observed positive and negative correlations of viruses to very diverse protist groups (Figure 7). Such correlations do not necessarily imply an infection of these viruses to such diverse hosts, but that they may influence one host group, consequently changing the growth conditions for another group. Another explanation can be that these protist groups may show similar or opposite responses to other environmental factors than infection by viruses. Our comparison between the relative abundance of viruses and their potential host may give new insight into the virus-algal host dynamics and the ecological role of algal viruses. We here suggest some relationships that can be investigated in future studies. However, in our study, we have only studied viruses in seawater passing a 0.45 μm pore-size filter, which include both viruses in the water mass and within host cells (<200 μm) that might have been disrupted during filtration (**Paper IV**). Viruses within intact eukaryote cells were not studied and thus may be overlooked, which can lead to problems in linking viruses with potential hosts.

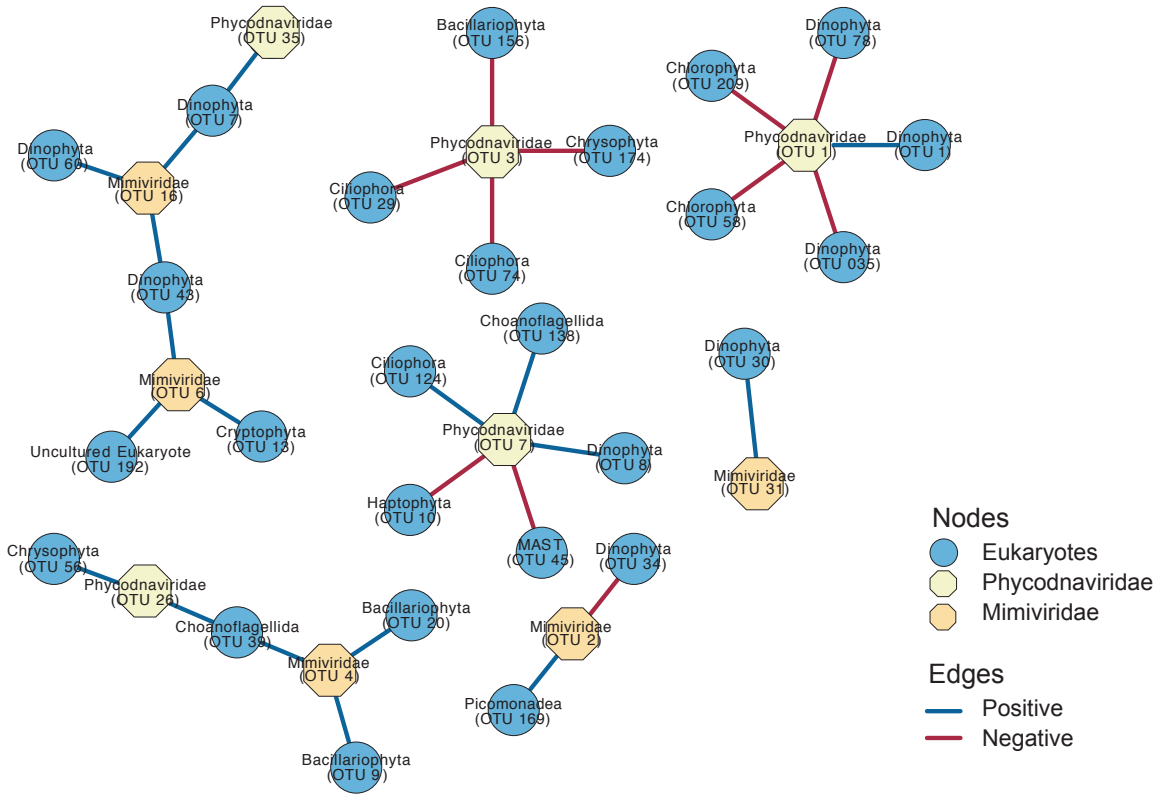


Figure 7. Network analysis revealing the co-occurrence between virus and protist taxa (represented as beige and blue nodes respectively. Lines between nodes indicate positive (blue) and negative (red) correlations ($p < 0.05$) between the abundances of linked taxa. The network was visualised by Cytoscape V3.3.0.

The proportional abundance of the major protistan and viral taxa varied between samplings. These changes could be explained by the long (monthly) sampling intervals (Countway et al. 2005). The sampling frequency is one of the important aspects when trying to detect seasonal patterns in microbial communities. We sampled once a month during two years to assess the protistan and viral community composition and seasonality with a main focus on the most abundant taxa. Several studies have demonstrated that for the protistan community once a month is a suitable frequency to detect the seasonality (**Paper I**). An exception may be when the target is part of the rare protistan biosphere. Lynch and Neufeld (2015) address the rare biosphere (the low-abundance taxa) as important contributors to assessments of α-diversity (species diversity in sites or habitats at a local scale) and β-diversity (changes in species diversity between different environments). Kim et al. (2011) reported rapid changes in the relative proportion of some rare taxa during a three-day incubation period, suggesting that these rare organisms are ecologically important under changing environmental conditions. We did not focus on the rare taxa, but their inclusion and closer sampling frequencies are aspects to consider in future studies to get a better knowledge on the total protistan biosphere in the

Skagerrak. Contrarily, when dealing with viral populations, these sampling intervals are not optimal. Due to the fast changes in viral communities, (e.g. high virus decay rate and boom-bust episodes), we need closer sampling intervals (e.g. weekly or even daily) to catch direct link between host-virus (**Paper IV**).

18S rRNA gives a better estimation of the haptophyte diversity.

The division Haptophyta was targeted in **Paper II.** Haptophytes are ecologically highly important and major primary producers in open oceans (Andersen et al. 1996; Liu et al. 2009). Also, calcifying haptophytes (e.g. *Emiliania huxleyi* and *Gephyrocapsa oceanica*) play a key role in the biogeochemical carbon cycle (Iglesias-Rodríguez et al. 2002), and blooms of non-calcifying haptophytes (e.g. *Prymnesium* spp.) may substantially impact coastal ecosystems through toxin production (Granéli et al. 2012).

In **Paper II** we performed a rigorous comparison between two different molecular markers (18S and 28S rRNA) on the Haptophyta community in the Outer Oslofjorden. We used haptophyte specific primers targeting the V4 region of the 18S, and the D1-D2 region of the 28S rRNA gene. We also used curated Haptophyte 18S and 28S reference databases for taxonomic assignment (**Paper II)**. As these two markers offer complementary views of environmental haptophyte communities, it is essential to compare richness and taxonomic composition of 18S and 28S metabarcoding datasets obtained from the same samples. Finally, we compared our high-throughput sequencing (HTS) results with those of counts of coccolithophores by scanning electron microscopy (SEM) to asses which of the two markers gave a more accurate image of the coccolithophore and haptophyte community.

Using the 18S rRNA gene we were able to resolve more defined clades compared to with the 28S rRNA gene (Figure 8). With the 28S rRNA gene, we could not taxonomically assign some haptophyte taxa to a major clade consisting of sequences from cultures or only environmental samples. Also, two families were only detected by the 18S rRNA gene and only one by the 28S rRNA gene. However, the number of described haptophyte species for which there are reference sequences available is higher for the 18S than 28S rRNA gene (96 vs 76; Edvardsen et al. 2016). This difference in numbers of references currently available makes the 18S rRNA gene a more useful marker for identifying haptophyte species in an environmental sample.

For the most diverse and abundant families, we observed similar OTUs and read percentage with both markers. However, we obtained a considerably higher richness using the 28S rRNA than with 18S. The 18S rRNA gene in haptophytes may differ in only a few base pairs between closely related species, and some short variable regions used for HTS metabarcoding may be identical (Bittner et al. 2013; Edvardsen et al. 2016). This may lead to an apparently reduced richness in our 18S dataset. Further, the 28S rRNA gene is more variable than 18S. Therefore, regions of the 28S might be more appropriate barcodes for distinguishing recently diverged species (Liu et al. 2009; Bittner et al. 2013). Some intra-species variation may still

occur, and this will thus lead to overestimated species richness (Liu et al. 2009). Such overestimation could be the case in our 28S dataset.

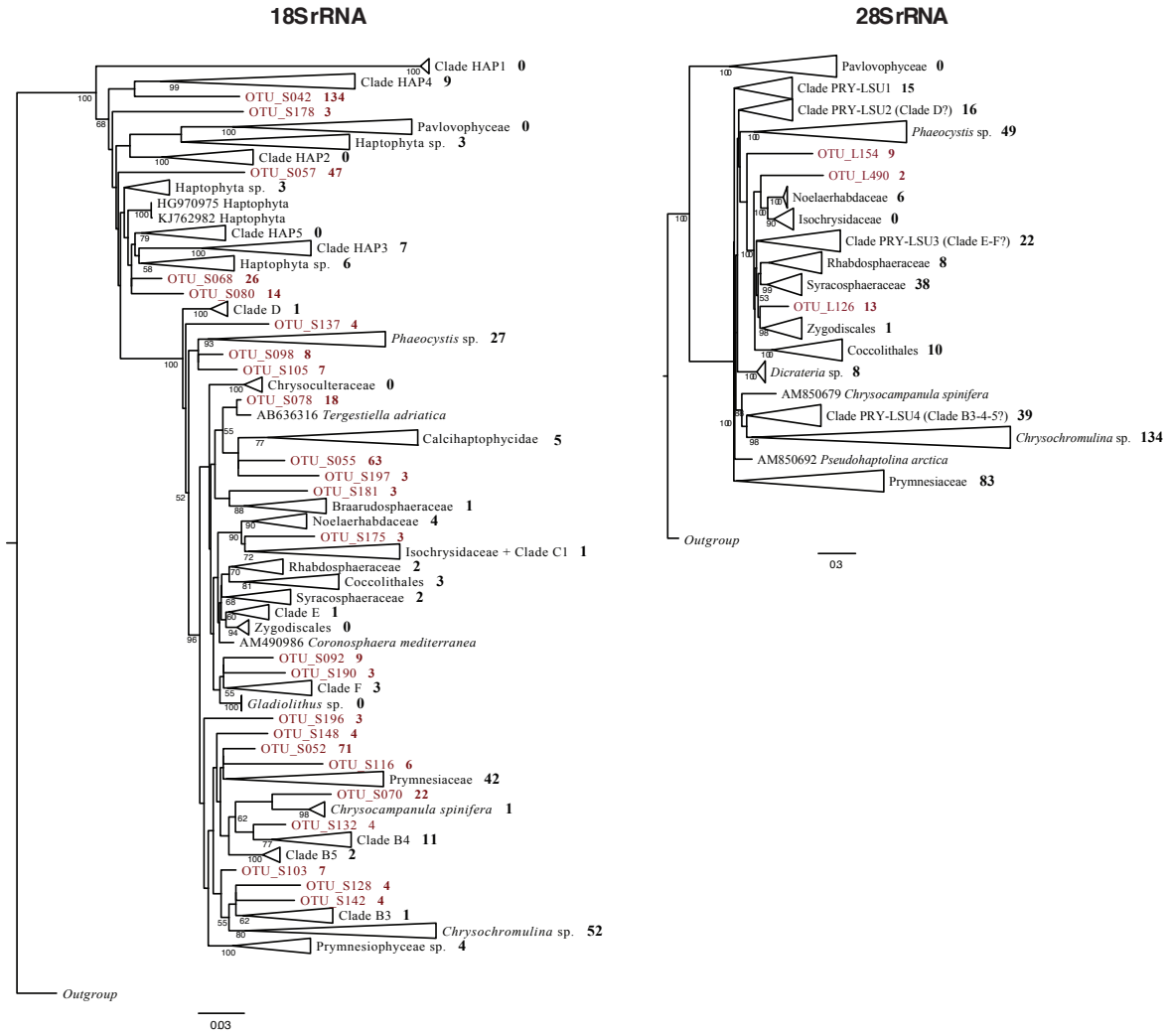


Figure 8. Maximum-likelihood (RAxML) haptophyte phylogeny of 215 18S rRNA OTUs and 432 28S rRNA OTUs. Support values are inferred using the GTRCAT model with 100 bootstraps. Bold numbers indicate the number of OTUs presented in each clade. Numbers next to OTU names indicate the number of reads in each OTU.

We also performed a taxonomical identification of coccolithophore morphospecies with SEM (**Paper II**). Coccolithophores are well preserved in samples prepared for SEM. This makes them suitable for qualitative and quantitative surveys (Young et al. 2003). Besides, the coccolithophore morphospecies taxonomy is well established (Cros and Fortuno 2002; Young et al. 2003; Jordan et al. 2004), as is the methodology for quantitative analyses (Bollmann et al. 2002). Therefore, we could directly compare our coccolithophore cell counts by SEM to those obtained by HTS. Comparison between methods showed that the 18S rRNA gene provides

good correspondence of relative abundance and taxonomic diversity with SEM observations. Finally, the 28S rRNA gene provided poor estimates in coccolithophore relative abundance and overestimated the diversity. This might be due to a mismatch in the used 28S rRNA primer, which failed to pick up the most abundant family (Noelaerhabdaceae).

In **Paper II** we suggest that SEM resolves both diversity and abundance for specific families but comes with known drawbacks. It uses small sample volumes and is a time-consuming method which requires taxonomic expertise. Thus, it is not suitable for routine monitoring of coccolithophore communities or surveys with a large number of samples. Also, if used on the full protist community it may overlook cryptic and/or fragile species. Overall, our results and curated 28S rRNA gene reference database based on cultures in **Paper II** may serve as a contribution to link microscopical (morphological) and molecular data, in addition to link the 28S to the 18S rRNA gene sequences of haptophytes without a cultured representative, improving thus the metabarcoding methodology.

FINAL CONCLUSIONS

Papers I and **IV** were the first metabarcoding studies targeting the protist and viral community through the seasons at the Skagerrak. Metabarcoding can unveil a detailed community composition and allows large sample volumes. The data collected on diversity, abundance and temporal distribution allowed us to draw a picture of the planktonic community in this area. Our results have revealed a high richness of protist and algae-infecting viruses compared to previous surveys through a decade. We also demonstrated that a vast diversity remains to be described both morphologically and genetically. At present molecular analyses of protist and virus diversity are restricted by a limited number of reference DNA-sequences (**Paper II**; Johannessen et al. 2017). New reference sequences need thus to be generated from cultures and isolated cells to improve the resolution and specificity of metabarcoding communities. We propose that 28S can be a useful complement to 18S in metabarcoding studies. Microscopy methods presented some drawbacks, such as that they are time-consuming techniques and require taxonomic expertise, and thus are not optimal for routine monitoring or surveys with a large number of samples. As neither metabarcoding nor microscopical method shows the full picture alone, they should be used complementarily.

In this thesis, we have increased the knowledge of protist and algae-infecting viral diversity and temporal distribution in the Norwegian coastal waters. We provided a curated 28S rRNA gene reference database based on cultures, contributed to linking molecular and morphological data and 28S to 18S rRNA gene sequences without cultured representatives and also improved the metabarcoding methodology.

FUTURE PERSPECTIVES

Our studies may serve as a baseline for future surveys and monitoring of planktonic communities to understand the effects of environmental and climate changes. Future studies may consider including more environmental factors to detect how climate change and eutrophication affects the protist community in the Skagerrak, such as light and day-length. We also suggest the use of newer sequencing technologies that allow for better diversity estimates, such as Illumina and exact sequence variants. PacBio could also be useful as a complement in metabarcoding to sequence the complete ribosomal DNA transcribed unit. Other molecular technologies such as FISH, which allows for visualisation and mapping the genetic material of individual cells, could also be of interest to give a better quantitative estimation of selected taxa in the planktonic community. We also suggest the development of a curated concatenated 18S and 28S rRNA reference database, that too will confirm the link between 18S and 28S rRNA clades without cultured representatives. As for the virus, we recommend that future studies also include the viruses present in eukaryotic cells to better link viruses with potential hosts.

REFERENCES

Adl S. M., Bass D., Lane C. E., Lukeš J., Schoch C. L., Smirnov A., Agatha S., Berney C., et. al. 2018. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*.

Adl S. M., Simpson A. G., Lane C. E., Lukes J., Bass D., Bowser S. S., Brown M. W., Burki F., et. al. 2012. The revised classification of eukaryotes. *J Eukaryot Microbiol*, **59**:429–493.

Andersen I., Berge J. A., Andersen J. H., Berntsson I., Danielssen D., Foverskov S., Fyrberg L., Gjøsæter J., et. al. 2001. The Skagerrak – environmental state and monitoring prospects. Forum Skagerrak.

Andersen R. A., Bidigare R. R., Keller M. D. & Latasa M. 1996. A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, **43**:517–537.

Artsdatabanken. 2018. The Norwegian Biodiversity Information Centre. Norway. Available from: https://www.artsdatabanken.no/

Bachy C., López-García P., Vereshchaka A. & Moreira D. 2011. Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night. *Frontiers in Microbiology*, **2**:106.

Backe-Hansen P. & Throndsen J. 2002. Pico- and nanoplankton from the inner Oslofjord, eastern Norway, including description of two new species of *Luffisphaera* (incerta sedis). *Sarsia: North Atlantic Marine Science*, **87**:55–64.

Bar-On Y. M., Phillips R. & Milo R. 2018. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, **115**:6506–6511.

Barton A. D., Dutkiewicz S., Flierl G., Bragg J. & Follows M. J. 2010. Patterns of Diversity in Marine Phytoplankton. *Science*, **327**:1509–1511.

Bates S. T., Clemente J. C., Flores G. E., Walters W. A., Parfrey L. W., Knight R. & Fierer N. 2013. Global biogeography of highly diverse protistan communities in soil. *The ISME Journal*, **7**:652–659.

Bergh Ø., Børsheim K. Y., Bratbak G. & Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature*, **340**:467–468.

Bittner L., Gobet A., Audic S., Romac S., Egge E. S., Santini S., Ogata H., Probert I., Edvardsen B. & de Vargas C. 2013. Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Molecular Ecology*, **22**:87–101.

Boenigk J. & Arndt H. 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. *Antonie van Leeuwenhoek*, **81**:465–480.

Boesch D., Sverige & Statens Naturvårdsverk. 2006. Eutrophication of Swedish seas: final report. Stockholm, Swedish Environmental Protection Agency.

Bollmann J., Cortés M. Y., Haidar A. T., Brabec B., Close A., Hofmann R., Palma S., Tupas L. & Thierstein H. R. 2002. Techniques for quantitative analyses of calcareous marine phytoplankton. *Marine Micropaleontology*, **44**:163–185.

Braarud T. & Bursa A. 1939. The phytoplankton of the Oslo fjord, 1933-1934. *J. Dybwad for Det Norske Videnskaps-akademi i Oslo*.

Braarud T., Gaarder K. R., Markali J. & Nordli E. 1952. Coccolithophorids studied in the electron microscope. Observations on *Coccolithus huxleyi* and *Syracosphaera carterae*. *Nytt Mag. Bot*, **1**:129–134.

Braarud T., Graarder K. R. & Grøntved J. 1953. The phytoplankton of the North Sea and adjacent waters in May 1948. Copenhague, Le Conseil.

Bratbak G., Egge J. & Heldal M. 1993. Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series*, **93**:39–48.

Bratbak G., Jacquet S., Larsen A., Pettersson L. H., Sazhin A. F. & Thyrhaug R. 2011. The plankton community in Norwegian coastal waters—abundance, composition, spatial distribution and diel variation. *Continental Shelf Research*, **31**:1500–1514.

Bratbak G., Thingstad F. & Heldal M. 1994. Viruses and the microbial loop. *Microbial Ecology*, **28**:209–221.

Callahan B. J., McMurdie P. J. & Holmes S. P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **11**:2639–2643.

del Campo J., Guillou L., Hehenberger E., Logares R., López-García P. & Massana R. 2016. Ecological and evolutionary significance of novel protist lineages. *European Journal of Protistology*, **55**:4–11.

Caron D. A. 2016. The rise of Rhizaria. *Nature*, **532**:444–445.

Caron D. A., Countway P. D., Savai P., Gast R. J., Schnetzer A., Moorthi S. D., Dennett M. R., Moran D. M. & Jones A. C. 2009. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Applied and Environmental Microbiology*, **75**:5797–5808.

Castberg T., Larsen A., Sandaa R., Brussaard C., Egge J., Heldal M., Thyrhaug R., van Hannen E. & Bratbak G. 2001. Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Marine Ecology Progress Series*, **221**:39–46.

Cavalier-Smith T. 1991. Cell diversification in heterotrophic flagellates. *In*: The Biology of Free-living Heterotrophic Flagellates. Clarendon Press. p. 113–131.

Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology*, **52**:297–354.

Chambouvet A., Morin P., Marie D. & Guillou L. 2008. Control of Toxic Marine Dinoflagellate Blooms by Serial Parasitic Killers. *Science*, **322**:1254–1257.

Charlson R. J., Lovelock J. E., Andreae M. O. & Warren S. G. 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, **326**:655–661.

Clerissi C., Desdevises Y., Romac S., Audic S., de Vargas C., Acinas S. G., Casotti R., Poulain J., Wincker P., Hingamp P., Ogata H. & Grimsley N. 2015. Deep sequencing of amplified *Prasinovirus* and host green algal genes from an Indian Ocean transect reveals interacting trophic dependencies and new genotypes: Marine algae and their viruses in the Indian Ocean. *Environmental Microbiology Reports*, **7**:979–989.

Corliss J. O. 2002. Biodiversity and biocomplexity of the protists and an overview of their significant roles in maintenance of our biosphere. *Acta Protozoologica*, **41**:199–219.

Countway P. D., Gast R. J., Savai P. & Caron D. A. 2005. Protistan Diversity Estimates Based on 18S rDNA from Seawater Incubations in the Western North Atlantic. *J. Eukaryot. Microbiol.*, **52**:95–106.

Cros L. & Fortuno J. M. 2002. Atlas of Northwestern Mediterranean coccolithophores. *Scientia Marina (Barcelona)*, **66**:7–182.

Cushing D. H. 1989. A difference in structure between ecosystems in strongly stratified waters and in those that are only weakly stratified. *Journal of Plankton Research*, **11**:1–13.

De La Rocha C. L. & Passow U. 2007. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep Sea Research II*, **54**:639–658.

Debroas D., Hugoni M. & Domaizon I. 2015. Evidence for an active rare biosphere within freshwater protists community. *Molecular Ecology*, **24**:1236–1247.

Deflandre G. & Fert C. 1953. Etude des coccolithophorides des vases actuelles au microscope electronique-orientation des particules elementaires de calcaire en rapport avec les notions dheliolithae et dortholithae. Comptes rendus hebdomadaires des seances de l´academie des sciences. **236**:328–330.

Deschamps P. & Moreira D. 2009. Signal Conflicts in the Phylogeny of the Primary Photosynthetic Eukaryotes. *Molecular Biology and Evolution*, **26**:2745–2753.

Díez B., Pedós-Alió C. & Massana R. 2001. Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. *Applied and Environmental Microbiology*, **67**:2932–2941.

Dimmock N. J., Easton A. J. & Leppard K. N. 2016. Introduction to Modern Virology. 7th ed. Wiley-Blackwell.

Dittami S. M., Hostyeva V., Egge E. S., Kegel J. U., Eikrem W. & Edvardsen B. 2013. Seasonal dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. *Environmental Science and Pollution Research*, **20**:6719–6732.

Dragsund E., Aspholm O., Tangen K., Bakke S. M., Heier L. & Jensen T. 2006. Overvåking av eutrofitilstanden i Ytre Oslofjord. Femårsrapport 2001-2005. nr. 2006-0831. Oslo, Det Norske Veritas AS.

Dunthorn M., Klier J., Bunge J. & Stoeck T. 2012. Comparing the Hyper-Variable V4 and V9 Regions of the Small Subunit rDNA for Assessment of Ciliate Environmental Diversity. *Journal of Eukaryotic Microbiology*, **59**:185–187.

Edvardsen B., Egge E. S. & Vaulot D. 2016. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review. *Perspectives in Phycology*, :77–91.

Edvardsen B. & Imai I. 2006. The Ecology of Harmful Flagellates Within Prymnesiophyceae and Raphidophyceae. *In*: Ecology of Harmful Algae. Vol. 189. Berlin Heidelberg, Springer-Verlag.

Egge E., Bittner L., Andersen T., Audic S., de Vargas C. & Edvardsen B. 2013. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS One*, **8**:e74371.

Egge E. S., Eikrem W. & Edvardsen B. 2015a. Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *Journal of Eukaryotic Microbiology*, **62**:121–140.

Egge E. S., Johannessen T. V., Andersen T., Eikrem W., Bittner L., Larsen A., Sandaa R. A. & Edvardsen B. 2015b. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol Ecol*, **24**:3026–3042.

Enevoldsen H. 2003. Manual on Harmful Marine Microalgae. Gustaaf M. Hallegraeff, Donald Mark Anderson, and Allan D. Cembella. Paris: Unesco.

Fuhrman J. A., Cram J. A. & Needham D. M. 2015. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol*, **13**:133–146.

Gentekaki E. & Lynn D. H. 2009. High-Level Genetic Diversity but No Population Structure Inferred from Nuclear and Mitochondrial Markers of the Peritrichous Ciliate Carchesium polypinum in the Grand River Basin (North America). *Applied and Environmental Microbiology*, **75**:3187–3195.

Giner C. R., Balagué V., Krabberød A. K., Ferrera I., Reñé A., Garcés E., Gasol J. M., Logares R. & Massana R. 2018. Quantifying long-term recurrence in planktonic microbial eukaryotes. *Molecular Ecology*.

Giner C. R., Forn I., Romac S., Logares R., de Vargas C. & Massana R. 2016. Environmental Sequencing Provides Reasonable Estimates of the Relative Abundance of Specific Picoeukaryotes. *Applied and Environmental Microbiology*, **82**:4757–4766.

Goodwin K. D., Thompson L. R., Duarte B., Kahlke T., Thompson A. R., Marques J. C. & Caçador I. 2017. DNA Sequencing as a Tool to Monitor Marine Ecological Status. *Frontiers in Marine Science*, **4**:107.

Graham J. E., Wilcox L. W. & Graham L. E. 2009. Algae. Second. Pearson.

Granéli E., Edvardsen B., Roelke D. L. & Hagström J. A. 2012. The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae*, **14**:260–270.

Guidi L., Chaffron S., Bittner L., Eveillard D., Larhlimi A., Roux S., Darzi Y., Audic S., Berline L., et. al. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, **532**:465–470.

Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., et. al. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res*, **41**:D597–D604.

Guillou L., Eikrem W., Chrétiennot-Dinet M.-J., Le Gall F., Massana R., Romari K., Pedrós-Alió C. & Vaulot D. 2004. Diversity of Picoplanktonic Prasinophytes Assessed by Direct Nuclear SSU rDNA Sequencing of Environmental Samples and Novel Isolates Retrieved from Oceanic and Coastal Marine Ecosystems. *Protist*, **155**:193–214.

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

Hagino K., Young J. R., Bown P. R., Godrijan J., Kulhanek D. K., Kogame K. & Horiguchi T. 2015. Re-discovery of a "living fossil " coccolithophore from the coastal waters of Japan and Croatia Funahama Shukutsu Tomari Rovinj stations of KT90-9. *Marine Micropaleontology*, **116**:28–37.

Hampl V., Hug L., Leigh J. W., Dacks J. B., Lang B. F., Simpson A. G. B. & Roger A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." *Proceedings of the National Academy of Sciences*, **106**:3859–3864.

Hamsher S. E., Evans K. M., Mann D. G., Poulíčková A. & Saunders G. W. 2011. Barcoding Diatoms: Exploring Alternatives to COI-5P. *Protist*, **162**:405–422.

Hannon M., Gimpel J., Tran M., Rasala B. & Mayfield S. 2010. Biofuels from algae: challenges and potential. *Biofuels*, **1**:763–784.

Hasle G. R. & Smayda T. J. 1960. The annual phytoplankton cycle at Drobak, Oslofjord. *Nytt. Mag. Bot.*, **8**:53–75.

Heid C. A., Stevens J., Livak K. J. & Williams P. M. 1996. Real Time Quantitative PCR. *Genome Research*, **6**:986–994.

Hjort J. & Gran H. H. 1900. Hidrographic-biological investigations of the Skagerrak and the Christiania Fiord. *O. Andersens*.

Hosoi-Tanabe S. & Sako Y. 2005. Rapid detection of natural cells of *Alexandrium tamarense* and *A. catenella* (Dinophyceae) by fluorescence in situ hybridization. *Harmful Algae*, **4**:319–328.

Hosoi-Tanabe S. & Sako Y. 2006. Development and application of fluorescence in situ hybridization (FISH) method for simple and rapid identification of the toxic dinoflagellates *Alexandrium tamarense* and *Alexandrium catenella* in cultured and natural seawater. *Fisheries Science*, **72**:77–82.

Huse S. M., Welch D. M., Morrison H. G. & Sogin M. L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering: Ironing out the wrinkles in the rare biosphere. *Environmental Microbiology*, **12**:1889–1898.

Iglesias-Rodríguez M. D., Brown C. W., Doney S. C., Kleypas J., Kolber D., Kolber Z., Hayes P. K. & Falkowski P. G. 2002. Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. *Global Biogeochemical Cycles*, **16**:20–47.

Ikävalko J. & Gradinger R. 1997. Flagellates and heliozoans in the Greenland Sea ice studied alive using light microscopy. *Polar Biology*, **17**:473–481.

Jha R. & Zi-rong X. 2004. Biomedical Compounds from Marine organisms. *Marine Drugs*, **2**:123–146.

Johannessen T., Larsen A., Bratbak G., Pagarete A., Edvardsen B., Egge E. & Sandaa R. A. 2017. Seasonal Dynamics of Haptophytes and dsDNA Algal Viruses Suggest Complex Virus-Host Relationship. *Viruses*, **9**:84.

Johannessen T. V., Bratbak G., Larsen A., Ogata H., Egge E. S., Edvardsen B., Eikrem W. & Sandaa R. A. 2015. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology*, **476**:180–188.

Jordan R. W., Cros L. & Young J. R. 2004. A revised classification scheme for living haptophytes. *Micropaleontology*, **50**:55–79.

Kålås J., Viken Å. & Bakken T. 2006. Norwegian red list. Norwegian Species Information Centre, Trondheim.

Karlson B. 2015. Nordic Microalgae. http://www.nordicmicroalgae.org., World-wide electronic publication.

Kim D. Y., Countway P. D., Gast R. J. & Caron D. A. 2011. Rapid Shifts in the Structure and Composition of a Protistan Assemblage During Bottle Incubations Affect Estimates of Total Protistan Species Richness. *Microbial Ecology*, **62**:383–398.

Koid A., Nelson W. C., Mraz A. & Heidelberg K. B. 2012. Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. *Appl Environ Microbiol*, **78**:3958–3965.

Kuylenstierna M. & Karlson B. 1994. Seasonality and composition of pico- and nanoplanktonic cyanobacteria and protists in the Skagerrak. *Botanica Marina*, **37**:17–33.

Kuylenstierna M. & Karlson B. 2006. Checklist of phytoplankton in the Skagerrak-Kattegat.

La Scola B. L. 2003. A Giant Virus in Amoebae. *Science*, **299**:2033–2033.

Larsen A., Flaten G. A. F., Sandaa R.-A., Castberg T., Thyrhaug R., Erga S. R., Jacquet S. & Bratbak G. 2004. Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. *Limnology and Oceanography*, **49**:180–190.

Lindley J. A. & Batten S. D. 2002. Long-term variability in the diversity of North Sea zooplankton. *Journal of the Marine Biological Association of the United Kingdom*, **82**: 31-40.

Liu H., Probert I., Uitz J., Claustre H., Aris-Brosou S., Frada M., Not F. & de Vargas C. 2009. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *PNAS*, **106**:12803–12808.

Logares R., Audic S., Santini S., Pernice M. C., de Vargas C. & Massana R. 2012. Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J*, **6**:1823.

López-García P., Rodríguez-Valera F., Pedrós-Alió C. & Moreira D. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**:603–607.

Loreau M., Downing A., Emmerson M., Gonzalez A., Hughes J., Inchausti P., Joshi J., Norberg J. & Sala O. 2002. A new look at the relationship between diversity and stability. *In*: Michel Loreau Pablo Inchausti S. N. (ed.), Biodiversity and Ecosystem Functioning Synthesis and Perspectives. New York, Oxford University Press. p. 79–91.

Loreau M., Naeem S., Inchausti P., Bengtsson J., Grime J. P., Hector A., Hooper D. U., Huston M. A., Raffaelli D., Schmid B., Tilman D. & Wardle D. A. 2001. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science*, **294**:804–808.

Lynch M. D. & Neufeld J. D. 2015. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol*, **13**:217–229.

Maar M. 2003. Distributions of zooplankton in relation to biological-physical factors. Roskilde, Denmark, National Environmental Research Institute. 142 p.

Madoni P. 2003. Protozoa as indicators of wastewater treatment efficiency. *In*: Handbook of Water and Wastewater Microbiology. London ; San Diego, Academic Press. p. 361–371.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., et. al. 2005. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature*, **437**:376–380.

Massana R. 2015. Protistan Diversity in Environmental Molecular Surveys. *In*: Marine Protists: Diversity and Dynamics. Tokyo, *Springer*. p. 3–21.

Massana R., del Campo J., Sieracki M. E., Audic S. & Logares R. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J*, **8**:854–866.

Massana R., Castresana J., Balague V., Guillou L., Romari K., Groisillier A., Valentin K. & Pedros-Alio C. 2004. Phylogenetic and Ecological Analysis of Novel Marine Stramenopiles. *Applied and Environmental Microbiology*, **70**:3528–3534.

Massana R., Gobet A., Audic S., Bass D., Bittner L., Boutte C., Chambouvet A., Christen R., et. al. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ Microbiol*, **17**:4035–4049.

Massana R., Terrado R., Forn I., Lovejoy C. & Pedros-Alio C. 2006. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol*, **8**:1515–1522.

Medlin L. K. & Kooistra W. H. C. F. 2010. Methods to Estimate the Diversity in the Marine Photosynthetic Protist Community with Illustrations from Case Studies: A Review. *Diversity*, **2**:973–1014.

Mendoza L., Taylor J. W. & Ajello L. 2002. The Class Mesomycetozoea: A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary. *Annual Review of Microbiology*, **56**:315–344.

Nagasaki K. & Yamaguchi M. 1998. Intra-species host specificity of HaV (*Heterosigma akashiwo* virus) clones. *Aquatic Microbial Ecology*, **14**:109–112.

Naselli-Flores L., Padisák J. & Albay M. 2007. Shape and size in phytoplankton ecology: do they matter? *Hydrobiologia*, **578**:157–161.

Nitsche F., Thomsen H. A. & Richter D. J. 2017. Bridging the gap between morphological species and molecular barcodes - Exemplified by loricate choanoflagellates. *European Journal of Protistology*, **57**:26–37.

Norderhaug K., Gundersen H., Pedersen A., Moy F., Green N., Walday M., Gitmark J., Ledang A., Bjerkeng B., Hjermann D. & Trannum H. 2015. Effects of climate and eutrophication on the diversity of hard bottom communities on the Skagerrak coast 1990-2010. *Marine Ecology Progress Series*, **530**:29–46.

Not F., del Campo J., Balagué V., de Vargas C. & Massana R. 2009. New Insights into the Diversity of Marine Picoeukaryotes. *PLoS One*, **4**:e7143.

Not F., Latasa M., Marie D., Cariou T., Vaulot D. & Simon N. 2004. A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Applied and Environmental Microbiology*, **70**:4064–4072.

Not F., Massana R., Latasa M., Marie D., Colson C., Eikrem W., Pedrós-Alió C., Vaulot D. & Simon N. 2005. Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnology and Oceanography*, **50**:1677–1686.

Not F., Siano R., Kooistra W. H. C. F., Simon N., Vaulot D. & Probert I. 2012. Diversity and Ecology of Eukaryotic Marine Phytoplankton. **64**:1–53.

Okamoto N., Chantangsi C., Horák A., Leander B. S. & Keeling P. J. 2009. Molecular Phylogeny and Description of the Novel Katablepharid *Roombia truncata* gen. et sp. nov., and Establishment of the Hacrobia Taxon nov.Stajich J. E. (ed.). *PLoS ONE*, **4**:e7080.

O'Kelly C. J. 1993. The Jakobid Flagellates: Structural Features of Jakoba, Reclinomonas and Histiona and Implications for the Early Diversification of Eukaryotes. *The Journal of Eukaryotic Microbiology*, **40**:627–636.

Olenina I., Hajdu S., Edler L., Wasmund N., Busch S., Göbel J., Gromisz S., Huseby S., Huttunen M., Jaanus A., Kokkonen P., Ledaine I. & Niemkiewicz E. 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. *HELCOM Balt.Sea Environ. Proc.*, **106**:144 pp.

Pagarete A., Chow C.-E. T., Johannessen T., Fuhrman J. A., Thingstad T. F. & Sandaa R. A. 2013. Strong Seasonality and Interannual Recurrence in Marine Myovirus Communities. *Applied and Environmental Microbiology*, **20**:6253–6259.

Pawlowski J. 2014. Protist Evolution and Phylogeny. *In*: John Wiley & Sons Ltd (ed.), eLS. Chichester, John Wiley & Sons, Ltd.

Pawlowski J., Audic S., Adl S., Bass D., Belbahri L., Berney C., Bowser S. S., Cepicka I., et. al. 2012. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biology*, **10**:e1001419.

Pawlowski J. & Burki F. 2009. Untangling the Phylogeny of Amoeboid Protists. *Journal of Eukaryotic Microbiology*, **56**:16–25.

Pedrós-Alió C., Massana R., Latasa M., Garcí-Cantizano J. & Gasol J. M. 1995. Predation by ciliates on a metalimnetic *Cryptomonas* population: feeding rates, impact and effects of vertical migration. *Journal of Plankton Research*, **17**:2131–2154.

Pettitt M. E., Orme B. A. A., Blake J. R. & Leadbeater B. S. C. 2002. The hydrodynamics of filter feeding in choanoflagellates. *European Journal of Protistology*, **38**:313–332.

Prowe A. E. F., Pahlow M., Dutkiewicz S., Follows M. & Oschlies A. 2012. Top-down control of marine phytoplankton diversity in a global ecosystem model. *Progress in Oceanography*, **101**:1–13.

Ptacnik R., Solimini A. G., Andersen T., Tamminen T., Brettum P., Lepisto L., Willen E. & Rekolainen S. 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *PNAS*, **105**:5134–5138.

Quince C., Lanzén A., Curtis T. P., Davenport R. J., Hall N., Head I. M., Read L. F. & Sloan W. T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, **6**:6.

Quince C., Lanzen A., Davenport R. J. & Turnbaugh P. J. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**:38.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. (Team R. D. C. (ed.)). R Foundation for Statistical Computing. Available from: http://www.r-project.org

Rademaker M., Reckermann M., Tillmann U., Tillmann A., Colijn F., Zevenboom W. & Houpt P. 1998. *Fibrocapsa japonica* and *Heterosigma akashiwo*: new observations. *Harmful Algae News*, **17**:8–10.

Rockwell N. C., Lagarias J. C. & Bhattacharya D. 2014. Primary endosymbiosis and the evolution of light and oxygen sensing in photosynthetic eukaryotes. *Frontiers in Ecology and Evolution*, **2**.

Rodríguez-Ramos T., Dornelas M., Marañón E. & Cermeño P. 2014. Conventional sampling methods severely underestimate phytoplankton species richness. *Journal of Plankton Research*, **36**:334–343.

Rublee P. A., Kempton J., Schaefer E., Burkholder J. M., Glasgow, Jr H. B. & Oldach D. 1999. PCR and FISH detection extends the range of *Pfiesteria piscicida* in estuarine waters. *Virginia journal of science*, **50**:325–335.

Sætre R. 2007. The Norwegian coastal current: oceanography and climate. Akademika Pub.

Sandaa R. A. & Bratbak G. 2018. Is the Virus Important? And Some Other Questions. *Viruses*, **10**:442.

Sandaa R.-A., Heldal M., Castberg T., Thyrhaug R. & Bratbak G. 2001. Isolation and Characterization of Two Viruses with Large Genome Size Infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology*, **290**:272–280.

Saraswati P. K. & Srinivasan M. S. 2016. Siliceous-Walled Microfossils. *In*: Micropaleontology. Cham, Springer International Publishing. p. 121–132.

Schiel D. R. & Foster M. S. 2015. The biology and ecology of giant kelp forests. University of California Press.

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

Schoch C. L., Seifert K. A., Huhndorf S., Robert V., Spouge J. L., Levesque C. A., Chen W., et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, **109**:6241–6246.

Short S. 2012. The ecology of viruses that infect eukaryotic algae. *Environ Microbiol*, **14**:2253–2271.

Short S., Rusanova O. & Staniewski M. 2011. Novel phycodnavirus genes amplified from Canadian freshwater environments. *Aquatic Microbial Ecology*, **63**:61–67.

Sieburth J. M., Smetacek V. & Lentz J. 1978. Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnology and Oceanography*, **23**:1256–1263.

Simon M., López-García P., Deschamps P., Moreira D., Restoux G., Bertolino P. & Jardillier L. 2015. Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J*.

Simon N., Cras A. L., Foulon E. & Lemée R. 2009. Diversity and evolution of marine phytoplankton. *Comptes Rendus - Biologies*, **332**:159–170.

Simon N., Foulon E., Grulois D., Six C., Desdevises Y., Latimier M., Le Gall F., Tragin M., Houdan A., Derelle E., Jouenne F., Marie D., Le Panse S., Vaulot D. & Marin B. 2017. Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and. *Protist*, **168**:612–635.

Simpson A. G. B. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *International Journal of Systematic and Evolutionary Microbiology*, **53**:1759–1777.

Stock A., Jürgens K., Bunge J. & Stoeck T. 2009. Protistan diversity in suboxic and anoxic waters of the Gotland Deep (Baltic Sea) as revealed by 18S rRNA clone libraries. *Aquatic Microbial Ecology*, **55**:267–284.

Stoeck T., Bass D., Nebel M., Christen R. & Meredith D. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. **19**:21–31.

Stoeck T., Zuendorf A., Breiner H. W. & Behnke A. 2007. A molecular approach to identify active microbes in environmental eukaryote clone libraries. *Microbial Ecology*, **53**:328–339.

Suttle C. A. 2005. Viruses in the sea. *Nature*, **437**:356–361.

Suzan-Monti M., Scola B. L., Barrassi L., Espinosa L. & Raoult D. 2007. Ultrastructural Characterization of the Giant Volcano-like Virus Factory of *Acanthamoeba polyphaga Mimivirus*.Digard P. (ed.). *PLoS ONE*, **2**:e328.

Tekle Y. I., Wood F. C., Katz L. A., Cerón-Romero M. A. & Gorfu L. A. 2017. Amoebozoans are Secretly but Ancestrally Sexual: Evidence for Sex Genes and Potential Novel Crossover Pathways in Diverse Groups of Amoebae. *Genome Biology and Evolution*, **9**:375–387.

Thingstad T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, **45**:1320–1328.

Thomsen H. A., Hansen G., Larsen J., Moestrup Ø. & Vørs N. 1992. Phytoplankton and heterotrophic nanoplankton. *In*: Plankton dynamics and material fluxes in the Kattegat. T. Fenchel. Miljostyrelsen. p. 31–61.

Throndsen J., Hasle G. R. & Tangen K. 2007. Phytoplankton of Norwegian Coastal Waters. Oslo, Almater Forlag AS.

Tragin M., Zingone A. & Vaulot D. 2018. Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta: Comparison of V4 and V9 18S metabarcodes. *Environmental Microbiology*, **20**:506–520.

Utermöhl H. 1958. Zur Vervollkommung der quantitativen Phytoplankton-Methodik. *Internationale Vereinigung fuer Theoretische und Angewandte Limnologie Verhandlungen*, **9**:38.

Van Etten J. 2000. Phycodnaviridae. *In*: Virus taxonomy, classification and nomenclature of viruses, seventh report. San Diego, Academic Press. p. 183–193.

de Vargas C., Audic S., Henry N., Decelle J., Mahé F., Logares R., Lara E., Berney C., Le Bescot N., Carmichael I. P. M., Poulain J., Romac S., Colin S., Aury J.-M., Bittne L. & Karsenti E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**:1261605–1–11.

Walday M. G., Gitmark J. K., Naustvoll L. J. & Selvik J. R. 2017. Overvåking av Ytre Oslofjord 2014-2018. Årsrapport for 2016.

Walne P. & Kivic P. 1990. Euglenida. *In*: Handbook of Protoctista. Jones and Barlett. p. 270–287.

Wang M.-N., Ge X.-Y., Wu Y.-Q., Yang X.-L., Tan B., Zhang Y.-J. & Shi Z.-L. 2015. Genetic diversity and temporal dynamics of phytoplankton viruses in East Lake, China. *Virologica Sinica*, **30**:290–300.

Ward B. A., Dutkiewicz S., Barton A. D. & Follows M. J. 2011. Biophysical aspects of resource acquisition and competition in algal mixotrophs. *Am Nat*, **178**:98–112.

Wikfors G. H., Patterson G. W., Ghosh P., Lewin R. A., Smith B. C. & Alix J. H. 1996. Growth of post-set oysters, *Crassostrea virginica*, on high-lipid strains of algal flagellates *Tetraselmis* spp. *Aquaculture*, **143**:411–419.

Wilmot A. J. 1962. Clinical Amoebiasis. *Clinical Amoebiasis*.

Yabuki A., Eikrem W., Takishita K. & Patterson D. J. 2013. Fine Structure of *Telonema subtilis* Griessmann, 1913: A Flagellate with a Unique Cytoskeletal Structure Among Eukaryotes. *Protist*, **164**:556–569.

Young J. R., Geisen M., Cros L., Kleijne A., Sprengel C., Probert I. & Ostergaard J. 2003. A guide to extant coccolithophore taxonomy. *Journal of Nannoplankton Research, Special Issue*, **1**:1–132.

Zamor R. M., Glenn K. L. & Hambright K. D. 2012. Incorporating molecular tools into routine HAB monitoring programs: Using qPCR to track invasive Prymnesium. *Harmful Algae*, **15**:1–7.

PAPER I

The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as Revealed by Metabarcoding and Microscopy

Sandra Gran-Stadniczeñko[a], Elianne Egge[a], Vladyslava Hostyeva[b], Ramiro Logares[c], Wenche Eikrem[a,b] & Bente Edvardsen[a]

a Department of Biosciences, University of Oslo, P. O. Box 1066 Blindern, 0316, Oslo, Norway
b Norwegian Institute for Water Research, Gaustadalléen 21, 0349, Oslo, Norway
c Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), 08003, Barcelona, Catalonia, Spain

Keywords
Biovolume; high-throughput sequencing; Oslofjorden; richness.

Correspondence
B. Edvardsen, Department of Biosciences, University of Oslo, P. O. Box 1066 Blindern, 0316 Oslo, Norway
Telephone number: +47-22-85-70-38;
FAX number: +47-22-85-47-26;
e-mail: bente.edvardsen@ibv.uio.no

Received: 30 April 2018; revised 15 October 2018; accepted October 18, 2018.

doi:10.1111/jeu.12700

ABSTRACT

Protist community composition and seasonal dynamics are of major importance for the production of higher trophic levels, such as zooplankton and fish. Our aim was to reveal how the protist community in the Skagerrak changes through the seasons by combining high-throughput sequencing and microscopy of plankton collected monthly over two years. The V4 region of the 18S rRNA gene was amplified by eukaryote universal primers from the total RNA/cDNA. We found a strong seasonal variation in protist composition and proportional abundances, and a difference between two depths within the euphotic zone. Highest protist richness was found in late summer-early autumn, and lowest in winter. Temperature was the abiotic factor explaining most of the variation in diversity. Dinoflagellates was the most abundant and diverse group followed by ciliates and diatoms. We found about 70 new taxa recorded for the first time in the Skagerrak. The seasonal pattern in relative read abundance of major phytoplankton groups was well in accordance with microscopical biovolumes. This is the first metabarcoding study of the protist plankton community of all taxonomic groups and through seasons in the Skagerrak, which may serve as a baseline for future surveys to reveal effects of climate and environmental changes.

PROTISTS are unicellular and multicellular algae and protozoans with a wide range of ecological functions (Massana 2015). Microalgae play key roles in coastal ecosystems contributing significantly to carbon flux through the microbial loop (Not et al. 2012), and are the main suppliers of photosynthetic products that higher trophic levels of the marine food web depend upon. Protists are morphologically and genetically diverse, and are present in all types of marine habitats (Massana 2015). Phytoplankton communities on continental shelves are dominated in biomass by diatoms, dinoflagellates, and haptophytes (Simon et al. 2009). In temperate seas, community composition and abundance undergo strong seasonal changes as a result of alterations in abiotic factors, such as irradiance, temperature and nutrient levels, and biotic factors, such as grazing, pathogens, and competition.

The Skagerrak, off the coasts of Norway, Sweden and Denmark, undergoes strong seasonal environmental variations due to changes in meteorological and hydrological conditions, and irradiance. The balance of hydrological forces from brackish Baltic currents, saline North Atlantic currents, and land runoff lead to considerable salinity and temperature fluctuations and seasonal water column stratification. The water currents also bring in allochtonous plankton, which further contribute to a species rich phytoplankton community in this area (Andersen et al. 2001). In addition, variation in nutrient availability and grazing pressure cause inter-annual variations in the protist species composition (Braarud et al. 1953) with different environmental preferences. The Outer Oslofjorden monitoring location in the Skagerrak is considered to represent the Southern Norwegian

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

coastal waters (Braarud and Bursa 1939; Dragsund et al. 2006).

Studies on protist taxonomic composition in the Skagerrak area have been carried out for over a century with a focus on diversity and dynamics based on light-, electron-, and epifluorescence microscopy and flow cytometry (Backe-Hansen and Throndsen 2002; Braarud et al. 1953; Bratbak et al. 2011; Dittami et al. 2013; Hasle and Smayda 1960; Hjort and Gran 1900; Kuylenstierna and Karlson 1994). These studies have revealed the dynamics and distribution of organisms belonging to different trophic (auto-, mixo- and heterotrophs) and taxonomic groups such as dinoflagellates, diatoms, haptophytes, cryptophytes, prasinophytes, dictyochophytes, and euglenoids. The overall seasonal pattern that has emerged can be described as follows: Low protist abundances are found in the Outer Oslofjorden during winter due to constant mixing of water masses combined with low solar irradiance (Dittami et al. 2013). An increase in irradiance and heat, together with brackish water inputs from the Baltic Current and river run-offs lead to water stratification in early spring, in February–March. Stratification leads to improved light conditions in the upper mixed layer which triggers the first spring bloom dominated by diatoms (mainly *Skeletonema*, *Thalassiosira*, *Chaetoceros, Pseudo-nitzschia* spp.), where nutrients are supplied from bottom waters (Paasche and Østergren 1980). A second bloom dominated by diatoms may occur in May–June with river run-offs as nutrient source. Strong summer stratification in July–August limits the transport of nutrients from deep waters to the upper water column, resulting in relatively low phytoplankton biomass and a dominance of mixotrophic and heterotrophic flagellates, including dinoflagellates and haptophytes. A third, smaller bloom may occur in August-September, when decreased stratification and wind mixing bring up nutrients to the upper, photic zone. Finally, heavy storms and a decrease in irradiance and temperature occur in late autumn forcing a decline in the general protist community (Braarud and Bursa 1939). The aforementioned microscopy studies were, however, limited to small water volumes (up to 50 ml) and identification to species level of cells larger than ca. 20 μm. Thus, little is yet known about seasonal dynamics of smaller, fragile, or less abundant protists.

New molecular techniques have proven to be an indispensable tool to examine the marine microbial diversity (Medlin and Kooistra 2010) to overcome the limitations of traditional methods, for example, microscopy. They have revealed the existence of an immense variety of novel protists (Epstein and López-García 2008) without the need for isolation or culturing (Medlin and Kooistra 2010). The small subunit (SSU) 18S rRNA gene is the most widely used marker to detect and classify known species present in marine eukaryotic microbial communities and to assess the phylogenetic affiliations of unknown sequences (see López-García et al. 2001). Recently, studies targeting the haptophytes in the Outer Oslofjorden with high-throughput sequencing have elucidated a vast diversity in a greater detail than has previously been obtained by microscopy (Egge et al. 2015a,b; Gran-Stadniczeñko et al. 2017).

Here, we investigate how the protist plankton community in the Skagerrak changes through the seasons by combining high-throughput sequencing (HTS) of the V4 region of the 18S rRNA gene and microscopy analyses of samples taken monthly over two years. We addressed the following questions: (i) How do HTS-inferred community composition and relative abundance change with season and depth? (ii) Which are the main abiotic drivers for these changes? (iii) What is the proportion of heterotrophic and autotrophic protists through the seasons? (iv) Which species dominate in the HTS dataset and what are their seasonal distributions? (v) Does HTS reveal taxa not previously recorded in the area, or taxa novel to science? (vi) How do HTS results compare to microscopy observations? Here, we reveal novel diversity not previously recorded in the Skagerrak, and how major protist components occur through the year. This study also contributes to a better understanding of protist plankton community structure and dynamics.

MATERIALS AND METHODS

Sampling

The sampling was performed as previously described in Egge et al. (2015a,b). Twenty-one coastal sampling campaigns were performed at the OF2 monitoring station (59.17 N, 10.69 E) located in the Outer Oslofjorden, Northern Skagerrak on board R/V Trygve Braarud. Samplings were conducted monthly for 2 yr, between September 2009 and June 2011 (with exception of February 2010 when samples and measurements were collected by the Ferrybox ships of opportunity, due to ice coverage) within the HAPTODIV project. Samples from September 2009 and June 2010 were also parts of the EU project BioMarKs (www.biomarks.org).

A conductivity-temperature-depth sensor (CTD, Falmouth Scientific Inc., Cataumet, MA) attached to a Niskin bottle rosette was used to obtain physicochemical water column profiles (temperature, conductivity/salinity, depth and fluorescence) from 1 to 100 m depth. Niskin bottles were used to collect water samples for nutrients (N, P, Si and Tot-P) and Chlorophyll-a (Chl-a) analysis at eight different depths (1, 2, 4, 8, 12, 16, 20, and 40 m). Water samples for nutrient analysis were frozen and stored in 20-ml scintillation vials until analysed in an autoanalyzer (Bran Luebbe Autoanalyzer 3). For Chl-a analyses, 100–500 ml water from each depth were filtered onto glass-fibre filters (Whatman GF/F, 25 mm, c. 0.8 mm mesh size), transferred to 2-ml cryotubes and frozen in liquid N_2 at −196 °C. Filters were incubated in 10 ml 90% acetone for 30–60 min and Chl-a was fluorometrically quantified with a Turner Designs fluorometer TD-700 (Turner Designs, Sunnyvale, CA) as described by Strickland and Parsons (1972).

Protist communities were collected by filtration onboard ship. At each sampling occasion, 20 liters of sea water was collected with 5 liters Niskin bottles at two different

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

depths: subsurface (1 m) and the depth for the bottom of the deep chlorophyll maximum (DC) when present, which was determined by visual inspection of the fluorescence on the CTD plots and tables. When no chlorophyll peak was observed, the depth for DC samples was 8 m. To remove large plankton, a prefiltration step was performed through a 45 μm nylon mesh into hydrochloric acid-washed plastic carboys. Protist cells were then collected by fractionated filtration with a peristaltic pump (Masterflex 07523-80; Cole-Parmer, Vernon Hills, IL) at a rate of 0.5–1 l/min, through 142 mm diameter polycarbonate filters (Millipore, Billerica, MA) with pore sizes of 3 and 0.8 μm, in a line giving the size fractions 45–3 μm (nano) and 3–0.8 μm (pico) plankton. To minimise RNA degradation, filtration was conducted for maximum 40 min. Finally, filters were cut in four and each piece was transferred into a 5-ml cryotube, which was frozen in liquid N_2 onboard ship, and stored at −80 °C. During the BioMarKs sampling in September 2009 and June 2010, prefiltration was performed at 20 μm giving a nano size fraction of 3–20 μm.

High-throughput sequencing

Total RNA was extracted and amplified as described in Egge et al. (2015a) using RNA NucleoSpin II (Macherey-Nagel, Düren, Germany). From each sample, ½ of the filter was extracted (representing a 10 liters water sample). Sixty microliters of RNA eluate was obtained and concentration was checked with a NanoDrop spectrophotometer (Wilmington, DE). Standard PCR with universal eukaryote partial 18S rRNA gene primers 1F and 300R (see Edvardsen et al. 2003) was performed to check for residual DNA in the RNA eluates. DNase (TURBO DNA-freeTM kit, Ambion, Austin, TX) treatment was performed with the samples where PCR products were observed by gel electrophoresis, as described in the manufacturer's protocol. To reverse-transcribe the RNA to cDNA, the High-Fidelity first Strand cDNA Synthesis Kit (Agilent, Santa Clara, CA) with random primers was used according to the manufacturer's protocol. In the synthesis reaction, approx. 100 ng of RNA per sample was used. Samples from the BioMarKs project (September 2009 and June 2010) were prepared as specified in Logares et al. (2012). PCR amplification of cDNA was done using the eukaryote specific primers by Stoeck et al. (2010) TAReuk454FWD1 (5′-CCAGCA(G/C)C(C/T)GCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGAT(C/T)(A/G)A-3′), adapted for 454-pyrosequencing. The forward primer contained a sample specific tag (MIDs). PCR was conducted in four separate reactions per sample on an Eppendorf thermocycler (Mastercycler, ep gradient S, Eppendorf). The PCR mixtures (25 μl) contained 5 μl 5× Phusion GC buffer, 0.5 μl of dNTP at a concentration of 10 μM, 0.75 μl of DMSO, 1 μl of each primer at a concentration of 10 μM, 0.25 μl of polymerase (Phusion, Finnzymes, Vantaa, Finland), 1 μl of template cDNA (10–60 ng/μl) and 15.5 μl sterilised PCR water. The PCR-programme included an initial denaturation at 98 °C for 30 s, followed by 30 cycles (denaturation at 98 °C for 10 s, annealing at 55 °C for 30 s, extension at 72 °C for 30 s) and a final extension at 72 °C for 10 min. Pooled PCR reactions were then purified with AMPure beads (BeckmanCoulter, Brea, CA), quantified with NanoDrop and pooled to obtain equal concentrations for sequencing. The samples were prepared for sequencing with Lib-L chemistry and sequenced unidirectionally from the forward primer on ½ of a 454 life sciences GS-FLX Titanum sequencing plate (454 Life Sciences, Branford, CT) at the Norwegian Sequencing Centre at the Department of Biosciences, University of Oslo (http://www.sequencing.uio.no). Raw SFF sequence files were deposited to GenBank under the project number PRJNA497792.

Bioinformatic pipeline

AmpliconNoise v.1.6.0 (Quince et al. 2011) was used to denoise the 454 reads, which were truncated at 400 bp. Reads with > 8 bp homopolymers and/or presenting mismatches in barcode or primers were removed. Perseus (incorporated in AmpliconNoise) was used to identify and remove putative chimeras. Some chimeras were also found by manual inspection by BLASTn against the NCBI nucleotide database and excluded. Clustering and taxonomic assignation of reads were performed with the "pick_open_reference_otus.py" command implemented in QIIME v.1.9.1 (Caporaso et al. 2010). UCLUST v. 1.2.22 (Edgar 2010) was used to cluster the reads into OTUs with 98% sequence identity. An initial taxonomical assignment was performed against the Protist Ribosomal Reference Database (PR2 v.1.0.0, including only sequences from cultures and longer than 800 bp; Guillou et al. 2013, https://github.com/vaulot/pr2database) at > 90% similarity, using the parameter "pick_open_reference_otus.py". Subsequent taxonomic assignments were done with all OTUs that did not initially match any phylum, using BLASTn within the software Geneious (v10.2.2) against the PR2 and then the NCBI databases. By manual BLAST, some of them were found to be chimeras and thus removed. All OTUs assigned to metazoans were removed from the data set. OTUs with less than 10 reads were excluded from the dataset, to remove possible spurious diversity. Scripts for the bioinformatics pipeline in Qiime and statistical analyses in R are found in File S1.

Phylogenetic analyses

The 16 most abundant OTUs (> 1% of total reads) were taxonomically placed by the EUKREF RAxML-EPA (Evolutionary Placement Algorithm) pipeline (del Campo, pers. commun., Berger et al. 2011; Stamatakis 2014) for a more reliable taxonomic assignation. Reference sequences of Gymnodiniales (Dinophyta), Geminigeraceae (Cryptophyta), Mamiellaceae (Chlorophyta), Mediophyceae (Bacillariophyceae), Chrysochromulinaceae and Noelaerhabdaceae (Haptophyta) were selected from the PR2 database, and Stephanoencidae (Choanoflagellata) from NCBI, and then aligned by MAFFT- E-INS-I v.7.300 (Katoh et al. 2009). Phylogenetic analyses using RAxML v.8.0.26 (Stamatakis

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

2014) based on reference sequences were performed implementing GTRGAMMA model with 100 bootstrap runs. OTUs were aligned to reference sequences by MAFFT -addfragments and added to the reference RAxML best tree with raxmlHPC-PTHREADS-SSE3 using GTRCATI. The analyses were conducted on the Abel cluster at the University of Oslo. Scripts for the phylogenetic analyses are found in File S1.

Identification of novel taxa

To assess new records for the area we compared the taxonomic assignments of the OTUs (to $\geq 90\%$ similarity in QIIME) in this study with species lists in the Norwegian Species Information Centre (Artsnavnebasen at Artsdatabanken 2018, http://www2.artsdatabanken.no/artsnavn/Contentpages/Sok.aspx) and the Nordic Microalgae and Aquatic Protozoa Checklist in Sweden and Norway (http://nordicmicroalgae.org). In addition, we checked all novel taxa by manual Blast against NCBI to verify the taxonomic assignation.

Statistical analyses

All statistical analyses and figures were performed in R software v.3.4.1 (R Development Core Team 2017). Treemap plots representing the complete protist community composition at the OF2 station during the 2 years were created based on read abundance and OTU richness, with the treemap package (Tennekes 2017). The Vegan package (Oksanen et al. 2017) was used in all diversity analyses. To compare the communities in the different samples, the dataset was normalised to equal sample sizes by rarefying (i.e. subsampling) using the "rrarefy" function, each of the 82 samples to the lowest number of reads found in a single sample (998 reads). As some OTUs occur in both nano- and pico-size fraction samples, the data from the two size fractions within a sample were pooled after subsampling to give 41 samples. Richness (number of OTUs per sample), proportional abundances and the Shannon diversity index H′ (Shannon 1948) determined by the "diversity" function in R, were used to investigate the seasonal variation in the community structure at the two studied depths. Nonparametric generalised additive model (GAM) was used to fit monthly linear diversity time trends with the "gam" function from the "mgcv" package. To test if the two studied depths were significantly different with respect to richness and diversity, Welch Two Sample *t*-test was applied. Bray–Curtis distances (Bray and Curtis 1957) were generated and used to produce a dissimilarity matrix based on OTU presence-absence data. Non-Metric Multidimensional Scaling (NMDS) analyses based on the dissimilarity matrix were performed to explore community patterns applying the monoMDS function. ANOSIM (Analysis of similarity) were used to test differences in composition between seasons. To analyse the correlations between environmental factors and community changes, canonical correspondence analysis (CCA), Mantel test and PERMANOVA (Permutational multivariate analyses of variance) were conducted. Similarity percentage analysis (SIMPER) was performed with the "simper" function to identify the OTUs that drove most of the differences in seasonal assemblages. To compare relative read abundance obtained by HTS with relative biovolume measured by microscopy of specific taxonomic groups (Bacillariophyceae, Chrysophyceae, Dictyochophyceae, Dinophyceae and Euglenophyceae), Welch Two Sample *t*-test was applied.

Microscopy

Total water samples (100 ml) were dispensed into flasks directly from the Niskin bottles and preserved in Lugol's solution (1% final concentration, Throndsen et al. 2007). The phytoplankton cell concentrations were determined in a 10-ml sub-sample that was allowed to settle overnight and subsequently counted in an inverted microscope (Nikon Diaphot 300; Nikon Corporation, Tokyo, Japan) in accordance with the method of Utermöhl (1958). Qualitative inspections were also made on vertical (0–20 m depth) and horizontal phytoplankton net samples (mesh size 10 μm) preserved with Lugol's solution (1% final conc.). Phytoplankton taxa were identified to the lowest level possible with light microscopy (LM) according to Throndsen et al. (2007). Biovolumes were estimated from cell counts using the HELCOM 2006 protocol (Olenina et al. 2006).

RESULTS

The outer Oslofjorden is a dynamic locality with respect to hydrographical conditions and protist composition and abundance. Here, we examined the community structure of the eukaryotic pico- and nano-plankton (passing a nylon sieve with 45 μm mesh size) at a monitoring station (OF2) during 2 years (2009–2011) with monthly samplings, and at two depths. This is the first paper on seasonal dynamics of the total planktonic protist community in the Oslofjorden using metabarcoding and microscopy.

Seasonal variations of environmental factors

The physicochemical parameters temperature, salinity, density, and chlorophyll fluorescence at the OF2 station in the upper water column (0–40 m) showed seasonal variations as shown in Table 1, Fig. S1 and File S2, and previously described by Egge et al. (2015b) and Dittami et al. (2013). The chlorophyll *a* concentration was highest in the upper 4 m of the water column at all times and usually higher in 1 and 2 m than at 4 m. To compare whether there was a difference in the small protist community composition and structure within the well-lit eutrophic zone, a sample at 1 m depth (subsurface, SS) and bottom of the chlorophyll fluorescence peak, here called "deep chlorophyll" and shortened DC were sampled and analysed. The depth for the DC samples ranged between 5 and 22 m depth. The hydrographical conditions in the upper 40 m, including the upper mixed layer and the

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 1. Physical and chemical conditions at the OF2 station in the sampling depths subsurface (1 m) and bottom of the deep chlorophyll maximum (DC) in the sampling period September 2009 to June 2011

		Temperature (°C)		Salinity (PSU)		Density (σt)		Chl-a (μg/l)		N [μM]		Si [μM]		P [μM]		Tot-P [μM]	
Sample	Depth (m) DC	1 m	DC	1 m	DC	1 m	DC	1 m	DC	1 m	DC	1 m	DC	1 m	DC	1 m	DC
Sep. 09	20	15.47	16.05	25.25	28.68	18.38	20.97	2.76	0.9	0.66	0.94	2.08	2.83	0.06	0.03	0.69	0.66
Oct. 09	5	11.83	13.1	31.41	32.79	23.86	24.71	0.6	0.3	4.44	4.17	5.47	3.93	0.11	0.1	0.81	0.82
Nov. 09	14	6.94	7.5	19.11	21.92	13.03	15.57	0.13	2.5	5	1.53	3.91	4.15	0.48	0.42	0.98	1.39
Dec. 09	8	5.6	6.37	22.93	26.62	18.08	20.94	0.62	0.48	8.92	4.43	14.71	5.2	0.45	0.45	1.13	0.96
Jan. 10	12	−1.25	1.27	24.2	29.46	19.42	23.63	7.71	2.95	0.64	4.43	1.05	2.82	0.45	0.26	0.94	0.93
Feb. 10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mar. 10	12	5.25	5.78	32.75	33.11	25.87	26.14	2.49	1.95	5.43	6.28	5.21	4.59	0.62	0.7	0.96	0.88
Apr. 10	6	5.76	4.1	24.83	27.72	19.56	22.03	2.33	1.33	2.93	1.09	6.62	0.57	0.19	0.16	0.23	0.21
May. 10	10	7.65	6.36	26.3	29.05	20.5	22.86	1.63	0.98	1.13	0.75	2.72	1.48	0.06	0.13	0.39	0.44
Jun. 10	10	14.93	11.8	22.16	29.71	16.12	22.57	0.38	0.35	0.14	0.16	3.03	1.41	0.11	0.07	0.36	0.32
Aug. 10	20	18.44	14.12	23.31	31.74	16.25	23.73	1.85	0.19	0.14	1.68	1.17	2.42	0.06	0.16	0.26	0.39
Sep. 10	20	15.76	15.76	25.12	25.13	18.22	18.31	1.44	1.38	0.14	0.14	0.72	0.66	0.06	0.08	0.29	0.32
Oct. 10	8	10.35	11.14	25.1	26.52	19.2	20.19	1.58	0.65	2.75	2.19	2.8	2.99	0.15	0.13	0.22	0.24
Nov. 10	8	8.09	8.96	29.87	30.52	23.24	23.66	0.43	0.1	5.48	2.59	4.86	2.75	0.28	0.23	0.31	0.26
Dec. 10	8	0.31	3.52	28.37	30.96	22.74	24.66	0.59	0.57	3.64	2.57	5.18	3.83	0.81	0.48	9.52	4.04
Jan. 11	8	0.69	1.79	29.12	29.78	23.34	23.84	0.89	0.78	3.71	2.28	6.55	4.55	0.39	0.29	0.74	0.61
Feb. 11	15	0.56	2.9	29.01	31.84	23.22	25.45	5.42	1.32	0.36	5.43	0.32	4.46	0.26	0.45	0.17	0.71
Mar. 11	22	−0.26	2.12	24.97	31.06	20.01	24.93	0.8	0.19	0.43	4.93	1.17	3.08	0.48	0.74	0.48	0.81
Apr. 11	8	5.12	3.61	16.1	20.95	10.82	14.42	0.36	0.7	6.04	3.48	5.74	0.67	0.49	0.16	3.92	1.36
May. 11	10	12.32	11.86	17.84	20.58	13.25	15.51	0.49	0.46	1.19	0.87	6.17	5.23	0.25	0.12	0.47	0.26
Jun. 11	10	16.06	10.85	19.31	28.04	13.71	21.43	1.9	0.37	1.71	1.87	13.15	2.59	0.17	0.11	0.3	0.14

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

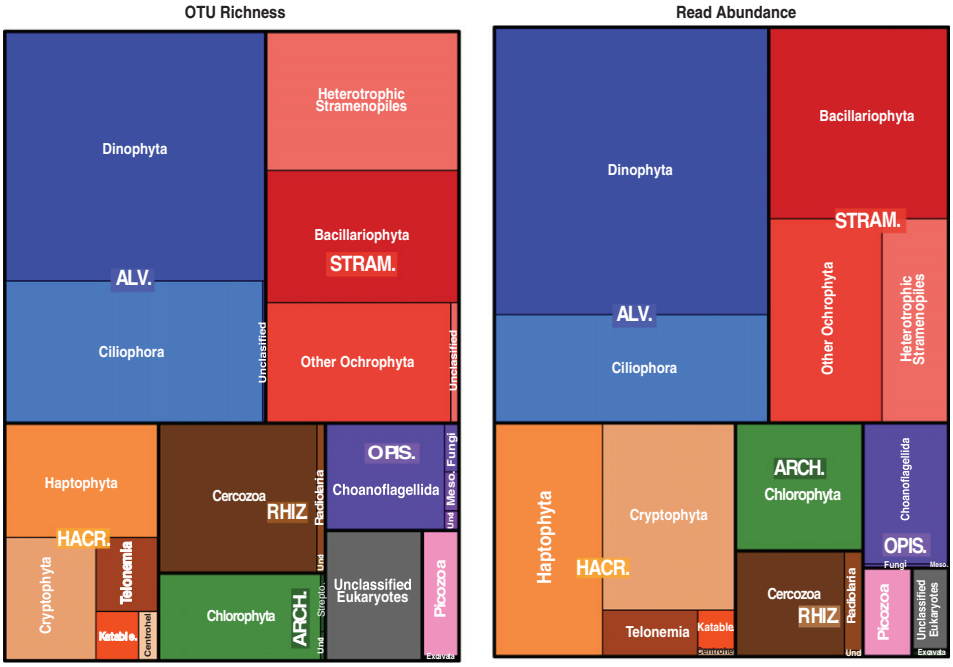


Figure 1 Tree maps displaying the taxonomic composition of the complete Outer Oslofjorden OF2 station protist HTS dataset at supergroup and phylum levels: OTU richness (left) and proportional read abundance (right) of supergroups Alveolata (ALV.), Stramenopila (STRAM.), Hacrobia (HACR.), Rhizaria (RHIZ.), Archeplastida (ARCH.), Opistokonta (OPIS.), Excavata and unclassified groups.

pycnocline, presented strong fluctuations. The seawater temperature increased during spring and summer up to 18.4 °C and decreased during autumn and winter to a minimum of −1.2 °C. An opposite pattern was observed for salinity and density. Highest values were found in winter or early spring with salinities up to 32.8, whereas lowest salinities were registered during late-spring and summer with minimum 16.1. Temporal patterns were also found in the Chl-*a* concentrations with highest values (up to 7.7 µg/l) during the main spring-bloom in late January 2010 and in February 2011. Concentrations of inorganic nutrient peaked during winter.

Taxonomic composition and relative abundance

After initial filtration of reads in the QIIME pipeline, including denoising by AmpliconNoise, we obtained 670,886 reads with average fragment length 375 bp, ranging between 184 and 400 bp. A second filtering step (removal of chimeras by Perseus, metazoan OTUs and OTUs with less than 10 reads in the whole dataset), resulted in 613,031 reads assigned to 2,032 OTUs (File S3). The taxonomic classification and absolute number of reads per OTU in each sample are presented in Table S1. Of the total OTUs, 1,791 were rare, with < 0.05% of the reads per OTU, while 44 OTUs were typically abundant, with > 0.5% of the reads per OTU, comprising 13.8% and 55.6% of the total reads, respectively. Most (95%) of the OTUs could be taxonomically assigned to one of 18 major micro-eukaryotic taxonomic taxa (superphylum to subphylum; Fig. 1). The remaining 5% were assigned as unclassified eukaryotes.

The infrakingdom Alveolata dominated the communities both in richness (36% of OTUs) and abundance (41% of reads). All alveolate OTUs except three were classified to a phylum: dinoflagellates or ciliates. Dinoflagellates were the most abundant phylum within alveolates, accounting for 67% of the total alveolate reads. Gymnodiniales was the most abundant order within dinoflagellates and the only order found in all samples (Fig. S2). Abundant taxa within the order Gymnodiniales were *Karenia* spp., *Karlodinium* sp., *Lepidodinium* sp., *Gyrodinium helveticum* and *Akashiwo sanguinea*. The second most abundant dinoflagellate group was Syndiniales, divided into the clades MALV I–V (marine alveolates without a cultured representative). MALV clades I–III were more abundant (> 0.5% of reads) than MALV IV and V (< 0.1%). Besides that, a few reads were assigned to Dinophysiales, Gonyaulacales, Noctilucales, Peridiniales, Prorocentrales, Suessiales, and Thoracosphaerales. Ciliates were both diverse (12.9% of total OTUs) and abundant, representing 33% of the alveolate reads (Fig. 1). Spirotrichea was the most represented class within ciliates (Fig. S2). The five most abundant ciliate OTUs had best match to the family Strombidiidae (Table S1).

Stramenopiles were the second most OTU-rich (26% of total OTUs) and abundant (20% reads) high rank group (subkingdom) after the alveolates (Fig. 1). Stramenopile OTUs were found in all samples, with highest abundances observed during spring. Two thirds were assigned to phylum Ochrophyta, and one-third was assigned to entirely

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

heterotrophic stramenopile phyla. Diatoms (Bacillariophyta) were the most diverse and abundant stramenopile group (36% of the total stramenopile reads) and were found during the entire sampling period (Fig. S2). The most important diatoms were the centric *Skeletonema marinoi*, *Thalassiosira nordenskioeldii* and *Chaetoceros neogracilis*. Other abundant diatoms were the centric *Chaetoceros debilis, Ch. calcitrans, Minidiscus trioculatus, Eucampia zoodiacus, Brockmanniella brockmannii, Ditylum brightwellii, Porosira pseudodenticulata, Leptocylindrus minimus, L. aporus, Proboscia alata,* and members of the pennate genus *Pseudo-nitzschia* (Table S1). Other ochrophyte groups present in all samples were Dictyochophyceae and Chrysophyceae. The class Dictyochophycea (silicoflagellates) was mainly represented by *Dictyocha speculum* (OTU 6). The next most abundant dictyochophyte was the picoflagellate *Florenciella parvula* present in 80% of the sampling dates, followed by the potentially ichthyotoxic species *Pseudochattonella verruculosa*, which was present in 50% of the sampling dates.

Within the heterotrophic stramenopiles, the most abundant groups were MAST-1, -3 and -7, which consist of marine stramenopiles without a cultured representative. Operational taxonomic units assigned to the heterotrophic stramenopile groups Bicoecea, Labyrinthulea, Oomyceta, Pirsonia, MAST-4, -6, -8, -9, -10, -12 and the phototrophic MOCH (marine ochrophyte without cultured representative) were also present in our dataset (Fig. S2).

The subkingdom "Hacrobia" was also considerably rich and abundant, with haptophytes, cryptophytes and telonemians contributing to 91% of the total Hacrobia reads (Fig. 1). Prymnesiales was the most abundant, frequently detected and diverse haptophyte order. Within this order, OTUs with best match to *Emiliania huxleyi* and *Chrysochromulina simplex* were the most abundant, followed by *Chrysochromulina acantha*. Other abundant haptophytes were assigned to *Prymnesium faveolatum, Imantonia rotunda*, and the bloom-forming species *Phaeocystis pouchetii* (Table S1). Two cryptophytes, *Teleaulax amphioxeia* and *Teleaulax gracilis,* were among the most abundant protists. Of the 29 telonemian OTUs found, only one had match with *Telonema antarcticum*, the rest belonged to unclassified *Telonemia* Group 1 and 2 (Table S1). Members of the heterotrophic phyla Katablepharida and Centroheliozoa were found in low proportions (Fig. S2).

Archaeplastida was primarily represented by chlorophytes (~7.6% total reads). Within Chlorophyta the most abundant and diverse group was Mamiellophyceae represented by 4.6% of all reads. *Micromonas commoda,* a picoflagellate belonging to this class, was among the most abundant OTUs. OTUs assigned to *Micromonas* spp. represented 6.7% of the reads in the pico size fraction. Other major components of the Chlorophyta community belonged to Pycnococcaceae (*Pycnococcus provasolii*), Trebouxiophyceae (*Amphikrikos nanus*), Pyramimonadales (*Pyramimonas* spp., *Pterosperma cristatum*), Nephroselmidophyceae (Fig. S2; *Nephroselmis pyriformis*), which are all pico- or small nanoplankton.

A total of 101 Opisthokonta OTUs were detected. Most of them were identified as choanoflagellates, mainly represented by the order Acanthoecida, which was present in all our samples (Fig. S2). The most abundant opisthokont OTU was placed close to *Calliacantha* spp. in our RAxML phylogeny. An additional BLAST against NCBI showed it to be nearly identical to the sequence KU587842 of *Calliacantha natans,* differing in only two bases, one being in a homopolymer. Five Fungi and four Mesomycetozoa OTUs were also found in low abundances (< 0.2% of total reads).

Rhizarians, mainly cercozoans, were diverse (178 OTUs) and detected in relative high abundances (6% of total reads; Fig. 1). Picozoa was found to be rather diverse (33 OTUs), and quite abundant (~2% reads). They are known as heterotrophic picoplankton with only one described species (*Picomonas judraskeda*). We found an unknown picozoa (OTU 21), differing from the only described species in seven positions, to be among the 25 most abundant. It was however identical to the sequence JN934893 of an uncultured picozoa isolated from Maine, USA. Finally, Excavata was represented by one abundant OTU, with 100% match to *Eutreptiella gymnastica* (accession number KF559331).

Most abundant OTUs

The 16 most abundant OTUs with > 1% of total reads per OTU, were more accurately taxonomically placed by RAxML-EPA. Separate trees for each taxon are presented in Fig. S3. Five of these OTUs were assigned to the dinoflagellate order Gymnodiniales, placed close to *Karenia papillonaceae* (OTU 1), *Karenia/Karlodinium* sp. (OTU 3), *Lepidodinium chlorophorum/L. viride*. (OTU 7), *G. helveticum* (OTU 8) and *A. sanguinea* (OTU 14) respectively. OTU 1 had identical sequence to *K. papillonaceae* whereas OTU 3 differed in one base pair from these, and in one other base pair from *Karlodinium micrum*. Two cryptophyte OTUs were identical to reference sequences of *T. amphioxeia* (OTU 2) and *T. gracilis* (OTU 13). The four most represented stramenopile OTUs had identical sequences to the diatoms *S. marinoi*, *T. nordenskioeldii* and *Ch. neogracilis* (OTUs 4, 9, and 16, respectively) and the dictyochophyte *D. speculum* (OTU 6). The fifth most abundant (OTU 5) was phylogenetically placed closest to *M. commoda*, differing in three base pairs. Three haptophyte OTUs were among the 16 most abundant. OTUs 10 and 15 were identical to *E. huxleyi* and *C. acantha* reference sequences respectively, whereas OTU 12 differed in two base pairs to that of *C. simplex*. The most abundant opisthokont OTU (OTU 11) was phylogenetically placed closest to *C. natans/C. longicaudata,* differing in one base pair from *C. natans*.

Seasonal variation in taxonomic groups at two depths as revealed by HTS

Succession of the 18 major taxonomic groups (from superphylum to subphylum) at the two studied depths through

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

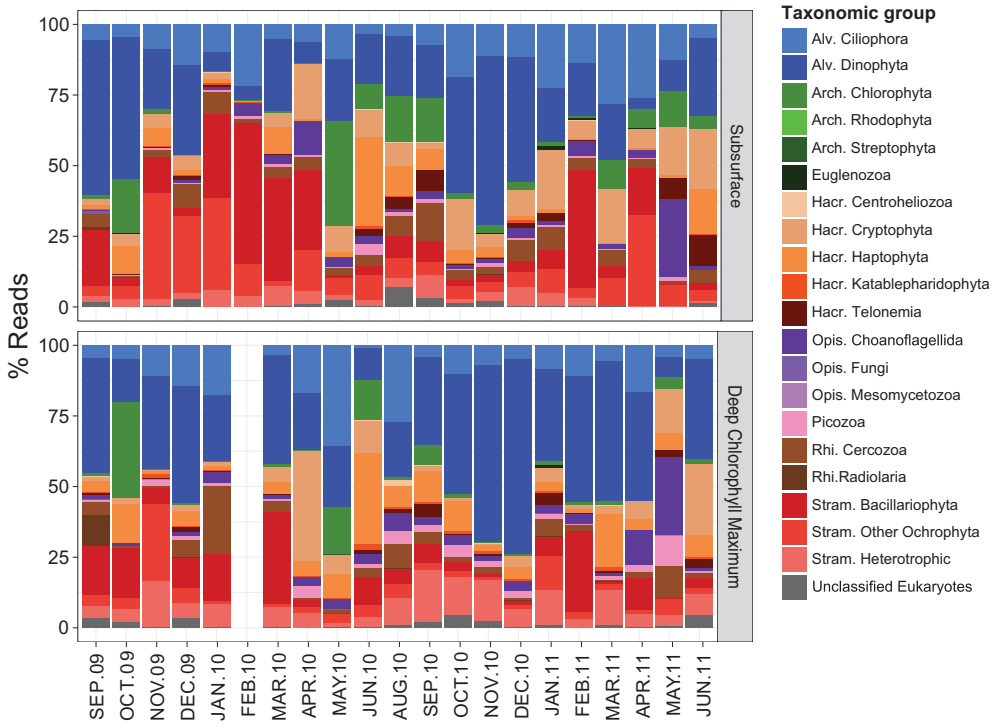


Figure 2 Succession of proportions of reads of the 18 major taxonomic groups across the 21 temporal samples at OF2 station.

the sampling period is shown as proportion of reads in Fig. 2 and OTUs in Fig. S4. All groups were found at both depths, but relative read abundance varied through the year and between the two depths (Fig. 2). There was no clear seasonal trend in proportion of OTU richness of the different groups, nor was there a consistent difference between the depths (Fig. S4). At both depths, alveolates were as a rule the most abundant group during autumn and early winter. Dinoflagellates were usually the most important alveolate group through the year, except for five samples where ciliates showed highest relative abundance (Fig. 2). The proportion of dinoflagellate reads was higher at the DC than at the SS from September 2010 to April 2011. This pattern was not clearly observed the year before. At lower taxonomic levels, reads representing *Lepidodinium* sp. and *A. sanguinea* were more abundant at the SS than at DC. Contrarily, *G. helveticum* and *K. papillonaceae* were more dominant at DC (Fig. 3), which partly explains the relatively high dinoflagellate abundances in March 10 and September 10 to April 11, respectively. The heterotrophic MALV clades I–III were present during the entire study period with few exceptions. MALV IV and V, however, appeared only in a few samples during summer-autumn and in very low abundances (< 0.3%). There were no clear differences in MALV distributions between the two depths.

Stramenopiles varied through the year and had highest proportional abundance during the winter–spring 2010. Their dominance was less pronounced during winter–spring 2011 (Fig. 2). This seasonal trend was more marked at the SS than at the DC. Diatoms and other ochrophytes showed higher proportions at the SS than DC. The phototrophs *D. speculum*, *T. nordenskioeldii* and *Ch. neogracilis* were amongst the most dominant taxa during the winter-spring 2010, and *S. marinoi* during the spring bloom in 2011 (Fig. 3), indicating the importance of these species during the spring blooms. Heterotrophic stramenopiles (mainly MAST OTUs; Fig. S2) and Picozoa showed higher proportions at the DC compared to SS. Members of Picozoa were present on all sampling occasions and depths.

Chlorophytes, haptophytes, and cryptophytes showed highest relative abundance during spring and summer, and similar at both depths (Fig. 2). The heterotrophic groups katablepharids, telonemians, choanoflagellates, and cercozoans were also present during all seasons but in low relative abundances (max. ~5% of reads per group in each sample). Exceptions were the cercozoans and choanoflagellates that showed a peak in January 2010 and May 2011, respectively. The choanoflagellate peak was due to the high proportion of *C. natans/C. longicaudata* reads (OTU 11) found in May 2011 at both depths (Fig. 3).

Seasonal dynamics of functional groups

We classified the OTUs into three trophic groups based on their taxonomic assignation: heterotrophs (choanoflagellates, picozoa, heterotrophic non-ochrophyte stramenopiles, ciliates, telonemia, radiolarians, katablepharids, cercozoans, fungi, centroheliozoans, mesomycetozoans, and members of class Syndiniophyceae within Dinophyta), autotrophs (cryptophytes, ochrophytes, haptophytes, rhodophytes, euglenoids,

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

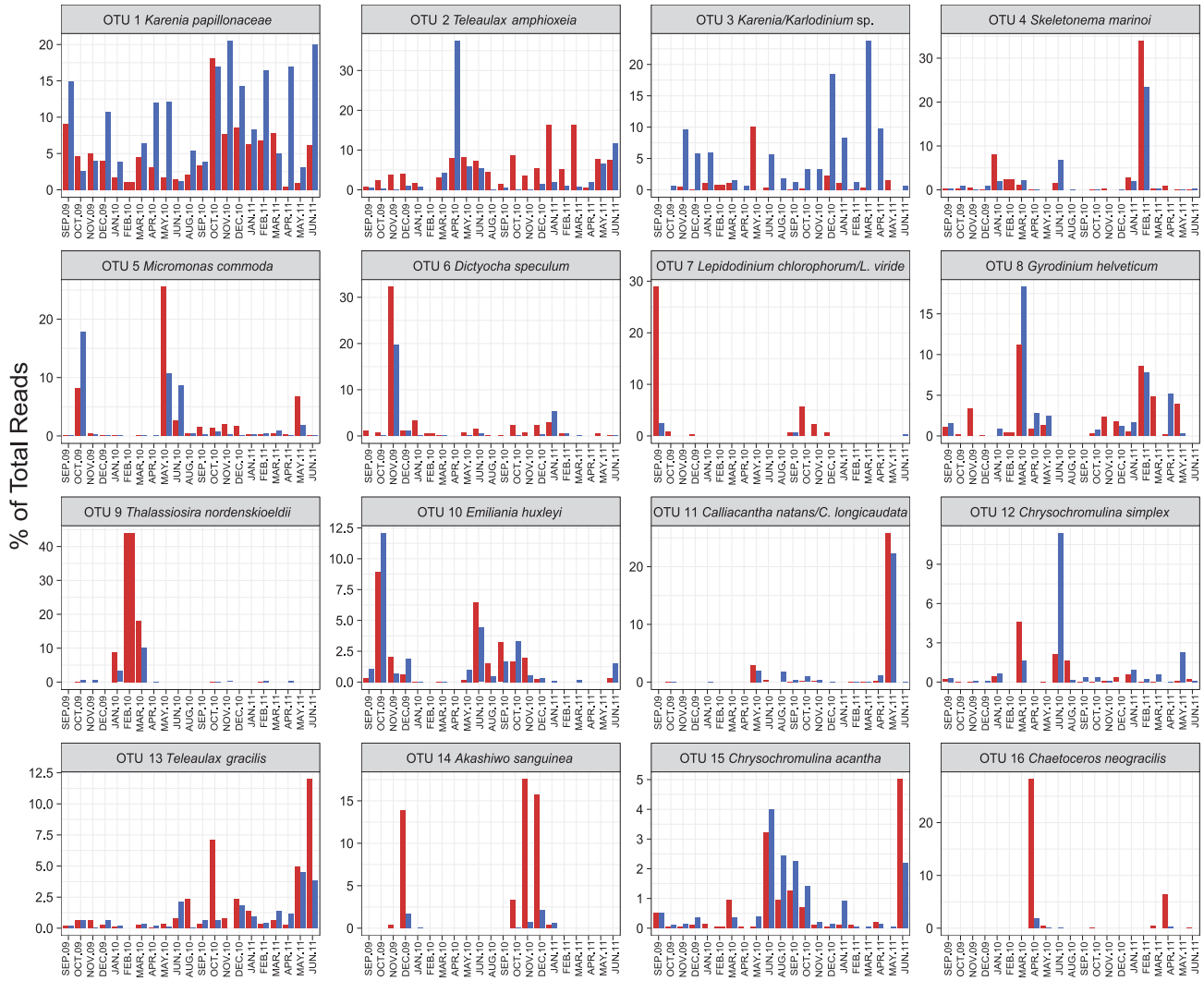


Figure 3 Temporal dynamics of the 16 most abundant OTUs (> 1% of total reads per OTU) at subsurface (SS, red) and bottom of deep chlorophyll maximum (DC, blue).

chlorophytes and streptophytes), and dinoflagellates consisting of autotrophic, mixotrophic and heterotrophic taxa, except the heterotrophic class Syndiniophyceae (Fig. 4). Within some of the autotrophic groups, some members have, however, lost their photosynthetic ability or may be mixotrophic.

Heterotrophs generally contributed more to the richness (per cent of total OTUs) than autotrophs through the sampling period at both depths, except in June 2010 and 2011 (SS), and October 2009 and June 2010 (DC), when the autotrophic community was more diverse (Fig. S5). The ratio of autotrophic- to heterotrophic OTU richness was rather similar through the entire study period. The dinoflagellate contribution of the OTUs varied between 6% and 18% and showed no clear pattern over the year and was similar at the two depths.

When comparing proportional abundances among trophic modes, a seasonal pattern was observed (Fig. 4). Reads corresponding to autotrophic groups dominated at the SS during the winter to summer period (January–August 2010, ~50% of reads) and dropped considerably during autumn (September–December 2011, ~36% of reads). In autumn, dinoflagellates reached their highest proportional abundance, especially the phototrophic *A. sanguinea* (Fig. 3). At the DC, fluctuations in read proportions were observed for all trophic groups. At DC autotrophs dominated in the autumn 2009 and spring and summer 2010, coinciding with high proportions of reads observed for *M. commoda*, *D. speculum*, *E. huxleyi*, and *Chrysochromulina* spp. (Fig. 3, 4). Heterotrophs were proportionally more abundant during winter 2009, autumn 2010 and spring 2011, and dinoflagellates took over during the autumn 2010 and winter 2011. Welch two-sample *t*-test showed significant differences in autotrophic proportional read abundances between the two depths ($P = 0.007$). However, no differences were found for proportional read abundances of heterotrophs and dinoflagellates between the two depths ($P = 0.07$ and $P = 0.12$, respectively). Autotrophic relative abundance was also higher in the subsurface than in

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

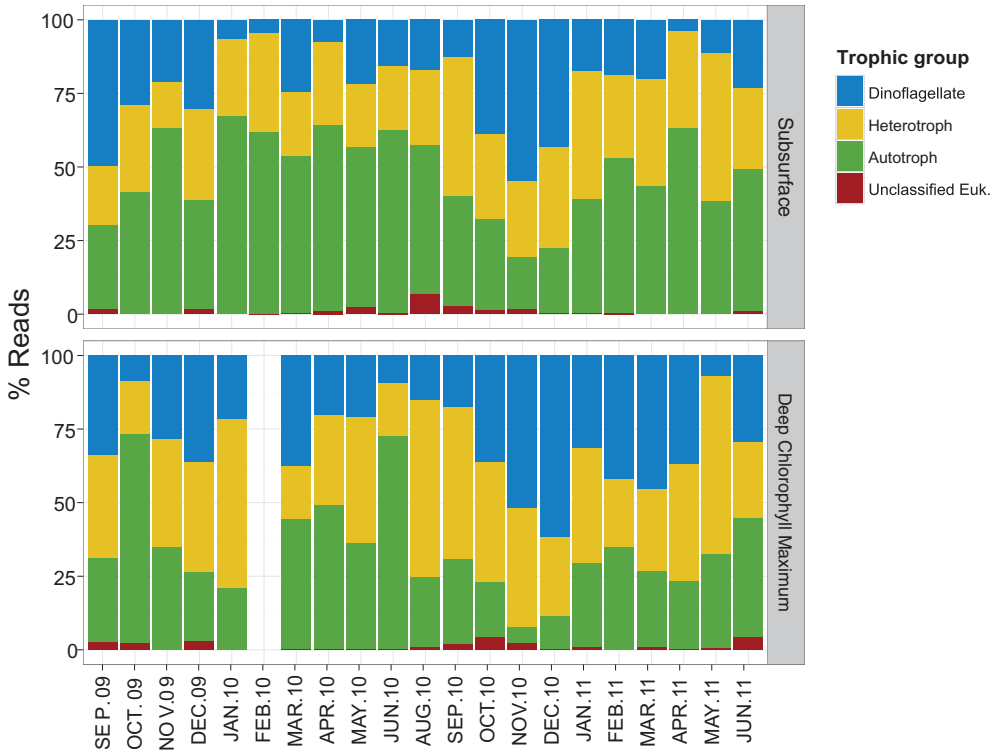


Figure 4 Succession of proportions of reads of the different trophic groups across the 21 temporal samples at OF2 station.

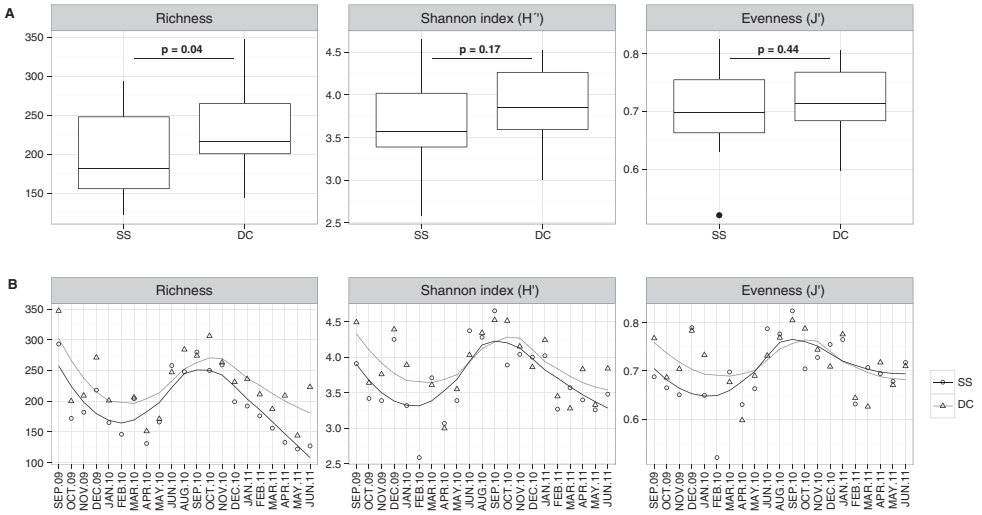


Figure 5 OTU richness, Shannon (H′) index, and Evenness (J′) index at both depths for the study site. (**A**) Boxplots. (**B**) Generalised additive model (GAM) smoothing curves fitted to the diversity indices temporal trends.

the DC, with peaks coinciding with the peaks of Chl-a observed in the isopleths (Fig. S1b).

Community structure in relation to environmental factors

Protist communities at OF2 showed a seasonal pattern. The richness median at the SS (182 OTUs, range 122–293) was significantly different than at the DC (217 OTUs, range 144–347; *t*-test $P = 0.04$; Fig. 5A and Table S2). In contrast, the SS and DC presented similar mean values of Shannon diversity index (3.57 and 3.85 respectively [*t*-test $P = 0.17$]) as well as Pielou's evenness index (0.70 and 0.71 respectively [*t*-test $P = 0.44$]). Richness and the Shannon index strongly differed between samples. They displayed a similar seasonal pattern at both depths

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

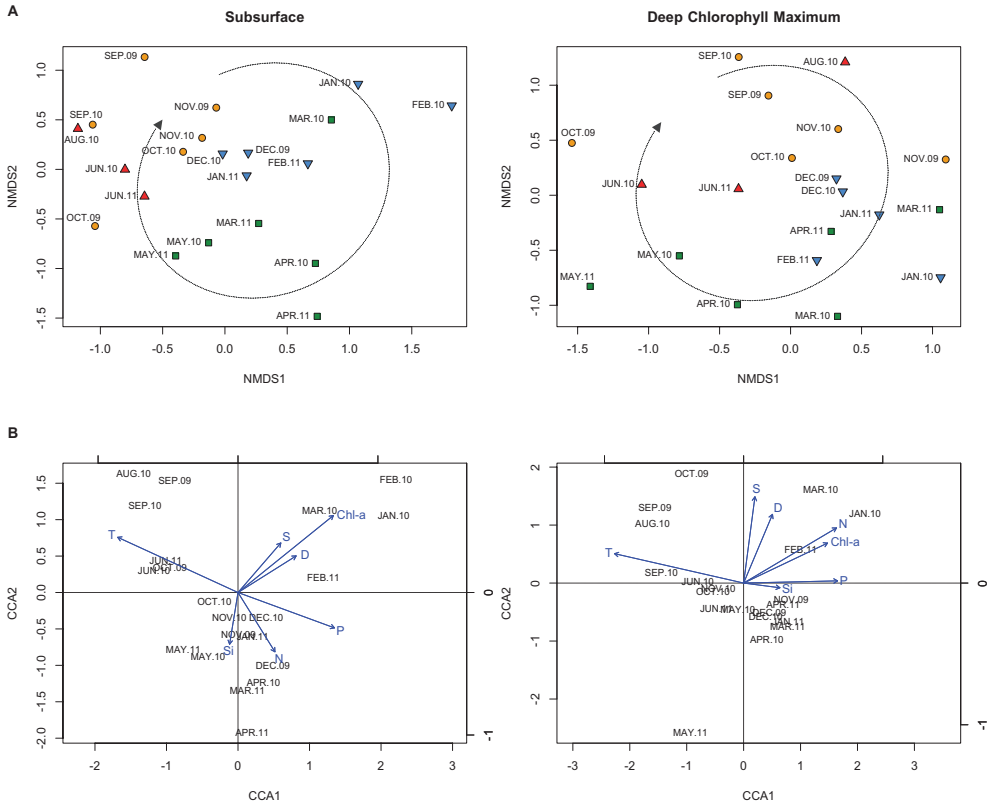


Figure 6 Ordination plots for both studied depth. (**A**) The diagram shows seasonal changes in the protist community composition. Seasons are indicated by different colours: spring (green), summer (red), autumn (yellow) and winter (blue). (**B**) Canonical Correspondence Analyses (CCA) plots showing correlations between seasonal communities and environmental factors.

reaching highest values during the summer-autumn seasons (Fig. 5B). Evenness also showed a seasonal pattern but the range was not large (0.52–0.82 for SS, 0.6–0.81 for DC).

Seasonality at both depths was also inferred by the ordination analyses based on Bray–Curtis dissimilarities (Fig. 6A), where protist communities presented four distinct seasonal clusters placed in a circular pattern. Summer and winter communities were more different at SS than at DC (ANOSIM: $R = 0.4558$, $P = 0.001$ and $R = 0.1976$, $P = 0.017$ for SS and DC respectively). The CCA (Fig. 6B) analyses were run to detect possible correlations between the environmental factors and the variations in protist communities. Temperature was found to be the most significant factor at both depths. In agreement with CCA, PERMANOVA results indicated that 20% of the seasonal variation in the SS protist community could be explained by the temperature ($P = 0.001$) and phytoplankton biomass (Chl-*a*; $P = 0.002$). In addition, salinity accounted for 8% of the variability, but this effect was not significant ($P = 0.08$; Table S3). In contrast, only temperature was a significant factor ($P = 0.001$) at the DC community explaining 10% of its variability.

The Venn diagram shows the percentage OTUs that are unique or shared between seasons (Fig. S6). The proportions were similar at both depths. Only ~7.5% of the total OTUs recorded during the 2-yr sampling were shared among the four seasons. Those consisted mostly of dinoflagellates, ochrophytes, heterotrophic stramenopiles, and ciliates OTUs.

SIMPER analyses results showed that eight OTUs (representing *Ch. neogracilis*, *S. marinoi*, *T. nordenskioeldii*, *D. speculum,* Pelagophyceae sp., *M. commoda*, *T. amphioxeia*, *A. sanguinea*, *K. papillonaceae*, and *Lepidodinium* sp.) contributed the most to the separation between seasons in the SS protist community (> 3% contribution per OTU; Table 2). Members of the infrakingdom Stramenopila contributed most to the community composition variation by season. At the DC, seven OTUs from five different phyla were the main responsible for the seasonal differences in community composition (*S. marinoi*, *M. commoda*, Colpodea sp., *T. amphioxeia*, *G. helveticum*, *Karenia* spp.), with the alveolates being the most important.

Novel records for Scandinavian waters

We detected 69 potentially new species and 40 potentially new genera of protists for Scandinavian coastal waters (see Table S4a) as compared to existing species lists for Norwegian and Scandinavian waters. For diatoms, 10 new records were found based on HTS, not yet observed by

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 2. Contribution of variance of top OTUs between seasons by SIMPER analysis

Depth	Season	OTU ID	Av. Diss	SD	Contr. %	Cum contr. %	Taxonomy
Subsurface	Spring–Summer	OTU 16	0.03	0.05	3.44	3.44	Bacillariophyta; Bacillariophyceae; *Chaetoceros neogracilis*
		OTU 5	0.03	0.04	3.29	6.72	Chlorophyta; Mamiellales; *Micromonas commoda*
	Summer–Autumn	OTU 7	0.03	0.05	3.991	3.991	Dinophyta; Gymnodiniales; *Lepidodinium chlorophorum/L. viride*
		OTU 6	0.03	0.06	3.755	7.746	Dictyochophyceae; Dictyochales; *Dictyocha speculum*
		OTU 1	0.03	0.02	3.332	11.079	Dinophyta; Gymnodiniales; *Karenia papillonaceae*
	Autumn–Winter	OTU 9	0.04	0.08	5.38	5.38	Bacillariophyta; Thalassiosirales; *Thalassiosira nordenskioeldii*
		OTU 4	0.04	0.06	4.74	10.11	Bacillariophyta; Thalassiosirales; *Skeletonema marinoi*
		OTU 14	0.03	0.04	3.91	14.03	Dinophyta; Gymnodiniales; *Akashiwo sanguinea*
		OTU 7	0.03	0.05	3.84	17.87	Dinophyta; Gymnodiniales; *Lepidodinium chlorophorum/L. viride*
		OTU 6	0.03	0.06	3.74	21.61	Dictyochophyceae; Dictyochales; *Dictyocha speculum*
		OTU 17	0.03	0.05	3.12	24.72	Pelagophyceae; Pelagophycea sp.
	Winter–Spring	OTU 9	0.05	0.08	6.24	6.24	Bacillariophyta; Thalassiosirales; *Thalassiosira nordenskioeldii*
		OTU 4	0.04	0.06	4.66	10.90	Bacillariophyta; Thalassiosirales; *Skeletonema marinoi*
		OTU 2	0.03	0.02	3.61	14.52	Cryptophyta; Pyrenomonadales; *Teleaulax amphioxeia*
		OTU 16	0.03	0.05	3.56	18.08	Bacillariophyta; Bacillariophyceae; *Chaetoceros neogracilis*
		OTU 5	0.03	0.05	3.32	21.40	Chlorophyta; Mamiellales; *Micromonas commoda*
		OTU 17	0.03	0.05	3.17	24.56	Pelagophyceae; *Pelagophycea* sp.
		OTU 14	0.02	0.04	3.01	27.58	Dinophyta; Gymnodiniales; *Akashiwo sanguinea*
Deep chlorophyll maximum	Spring–Summer	OTU 2	0.05	0.05	5.66	5.66	Cryptophyta; Pyrenomonadales; *Teleaulax amphioxeia*
		OTU 1	0.04	0.03	4.87	10.53	Dinophyta; Gymnodiniales; *Karenia papillonaceae*
		OTU 3	0.03	0.04	3.87	14.40	Dinophyta; Gymnodiniales; *Karenia/Karlodinium* sp.
		OTU 64	0.03	0.04	3.65	18.06	Ciliophora; Colpodea sp.
	Summer–Autumn	OTU 1	0.04	0.03	5.54	5.54	Dinophyta; Gymnodiniales; *Karenia papillonaceae*
		OTU 64	0.03	0.04	3.77	9.31	Ciliophora; Colpodea sp.
		OTU 2	0.03	0.02	3.57	12.88	Cryptophyta; Pyrenomonadales; *Teleaulax amphioxeia*
		OTU 5	0.02	0.03	3.12	16.00	Chlorophyta; Mamiellales; *Micromonas commoda*
	Autumn–Winter	OTU 1	0.04	0.02	4.63	4.63	Dinophyta; Gymnodiniales; *Karenia papillonaceae*
		OTU 3	0.03	0.03	4.15	8.78	Dinophyta; Gymnodiniales; *Karenia/Karlodinium* sp.
		OTU 4	0.03	0.05	3.60	12.38	Bacillariophyta; Thalassiosirales; *Skeletonema marinoi*
	Winter–Spring	OTU 3	0.04	0.03	5.54	5.54	Dinophyta; Gymnodiniales; *Karenia/Karlodinium* sp.
		OTU 2	0.04	0.06	5.47	11.01	Cryptophyta; Pyrenomonadales; *Teleaulax amphioxeia*
		OTU 1	0.03	0.02	3.63	14.64	Dinophyta; Gymnodiniales; *Karenia papillonaceae*
		OTU 4	0.03	0.04	3.61	18.25	Bacillariophyta; Thalassiosirales; *Skeletonema marinoi*
		OTU 8	0.02	0.03	3.16	21.42	Dinophyta; Gymnodiniales; *Gyrodinium helveticum*

Av. Diss = average dissimilarity between seasons; SD = standard deviation; Contr. % = percentage contribution of variance OTU; Cum. Contr. % = cumulative contribution of OTU in per cent. The taxonomy is based on EPA phylogenetic placement (see text). Method for taxonomic classification was EPA phylogeny for the 16 most abundant OTUs (OUT 1 – OTU 16), for the rest UCLUST against the PR2 database was used.

microscopy, for example, the genera *Eunotogramma* and *Tenuicylindrus*. Of the dinoflagellate species, eight new records were registered based on HTS; for example, the genera *Luciella* and *Adenoides*. Within the alveolates there were also 23 new ciliate records (Table S4b). Of other phytoplankton groups can be mentioned three new records of Chlorophyta, one new each of Cryptophyta, Pelagophyceae (genus *Ankylochrysis*) and Raphidophyceae (genus *Haramonas*), and Bolidophyceae (*Bolidomonas pacifica*). No new records of chrysophyceae, dictyochophyce, or euglenophytes were identified.

Community structure revealed by metabarcoding versus light microscopy

With metabarcoding, we targeted the protist-plankton community in the size range ca. 0.8–45 μm. The light microscopy (LM) cell counts were done on the total water sample, including all size groups. Cells smaller than c. 15–20 μm could, however, not be taxonomically identified to species under the light microscope. The comparison between methods was therefore done at the class and not species level. The main phytoplankton groups identified and counted by microscopy were the diatoms, with 51 recorded taxa or categories, and dinoflagellates (Dinophyta), with 59. In addition, two dictyochophyte, two euglenophyte and one chrysophyte taxa were observed and counted. Microscopy cell counts were transformed into biovolume to allow the comparison to relative read abundance (Table S5 and Fig. S7).

Comparisons of HTS reads and light microscopy biovolumes were performed on five major phytoplankton groups possible to identify by both methods. Ratios of the taxonomic groups (Fig. 7A) showed that Chrysophyceae and

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

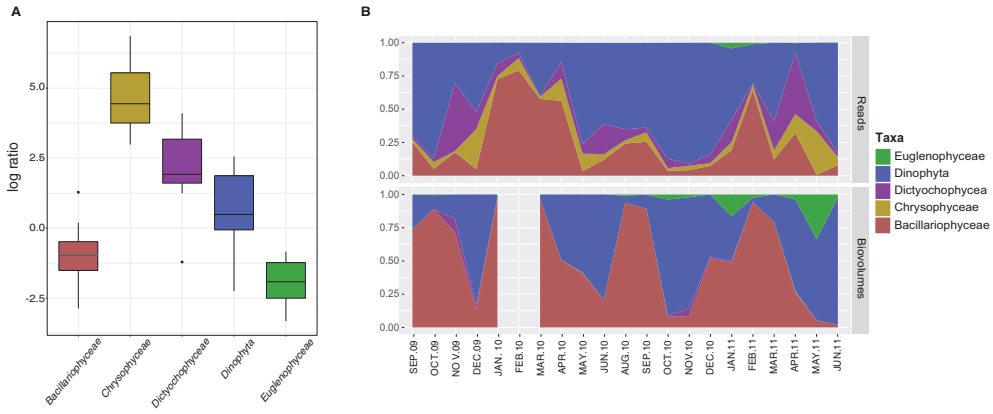


Figure 7 Comparison of metabarcoding reads versus microscopy data (biovolumes) for five phytoplankton groups. (**A**) Boxplots representing the log-ratio between reads and biovolumes. (**B**) Temporal variation in read and biovolume proportions.

Dictyochophyceae were overrepresented in HTS compared to LM, whereas Bacillariophyceae and Euglenophyceae were underrepresented. Welch *t*-test, however, showed that significant differences between the methods were only found for Bacillariophyceae ($P = 0.007$), Dictyochophyceae ($P = 0.002$), and Chrysophyceae ($P < 0.001$; Table S6). Dinoflagellates were not significantly different between methods ($P = 0.222$). The proportions of the five phytoplankton classes through the study period showed both similarities and differences between the two methods (Fig. 7B). Bacillariophyceae was the most abundant group during the winter and early spring in both years using both LM and HTS. There was also a diatom peak during summer (August–September) 2010 observed with both methods, but more pronounced by microscopy, where the diatoms were assessed to constitute 92% of the biovolume compared to 25% the reads. Microscopical counts showed that the dominating diatoms in September–November 2009 and August–September 2010 belonged to *Chaetoceros* species forming large chains that were probably removed to some extent by prefiltration prior to HTS.

Dinoflagellates showed a similar pattern, with peaks with both methods during spring to early summer (May–June) and late autumn (October–November) 2010, and May-June 2011. Dinophyta was the dominating group (> 60% of reads) in most HTS samples. In the microscopy counts this group dominated during late-autumn and late-spring, following the diatom blooms. Chrysophyceae was detected with HTS in all samples at low proportions, except for December 2009 and May 2011. With LM, Chrysophyceae, represented by the colony-forming *Dinobryon* sp., was barely detected in six samples. Dictyochophyceae reads were recovered in all HTS samples and was the group with highest proportions in November 2009 and April 2011. They were assigned to the genera *Florenciella*, *Pseudochattonella* and *Apedinella,* as determined by classification against PR2. In addition, 14 OTUs classified to unknown Dictyochophyceae were more than 90% similar to *Dictyocha* spp. as revealed by BLAST against NCBI. With microscopy only one dictyochophyte genus was detected: *Dictyocha,* found in nine samples at low biovolumes. Euglenophyceae was only detected in a few samples by both methods but seemed to be better detected by LM (January 2011 and May 2011).

DISCUSSION

This is the first long-term study of the protist community of the Oslofjorden and the Skagerrak by metabarcoding. A total of 2,026 OTUs from different trophic groups were revealed. This amount of OTUs is almost three times the number of taxa previously recorded for the Norwegian coastal waters through morphological observation (about 700 phytoplankton species according to Throndsen et al. (2007)). Our OTUs were defined at a 98% similarity level, as this has been found to be suitable for species-level distinctions of most protist groups (Caron et al. 2009). We found, however, many OTUs matching reference sequences from the same species, thus, the number of OTUs probably represents an inflated estimate of the true species diversity. Such a result indicates that a lower similarity level is needed to estimate the true diversity for some microeukaryotic groups. However, in some taxonomic groups, such as diatoms and haptophytes, different species may have identical V4 18S rRNA gene sequence, and a higher clustering level than 98% is needed to separate to species level (Egge et al. 2013). Furthermore, some microeukaryotic taxa are difficult to cultivate and/or identify through microscopy, and therefore, no molecular references are available. This may explain the number of unclassified taxa (4% of OTUs) obtained in our study.

Community composition

Alveolates, Stramenopiles and "Hacrobia" were the dominating supergroups in this study. Dinoflagellates were the most abundant phylum. They are, after diatoms, considered the most important primary producers in the ocean, reaching their highest abundances in estuaries and coastal

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

marine waters (Not et al. 2012). The high dinoflagellate contribution at the outer Oslofjorden is thus in accordance with Not and co-authors. Members of the genera *Karenia, Karlodinium* and *Akashiwo,* three of the most abundant genera in our waters, may form blooms associated to mortalities of fish or birds in marine coastal waters (Jones et al. 2017; Tangen 1977). *Karenia papillonaceae* was the most abundant OTU but has not previously been recorded in the Skagerrak or Norwegian waters by microscopy. This species has a second stage of small cells in culture (Carmelo Tomas, pers. commun.) that cannot be identified to species under the light microscope. It may thus have been misidentified or overlooked in past microscopical surveys. The uncultured marine alveolates group named MALV was first described by López-García et al. (2001) in 18S rRNA gene marine molecular surveys by environmental clone libraries. Members of MALV I-V have been phylogenetically placed in the dinoflagellate order Syndiniales and renamed Syndiniales group I–V (Guillou et al. 2008). Syndiniales group I and II are all assumed to be parasitic, and within group II we find the genus *Amoebophrya*. Notably, members of the MALV clades I–V were detected in all occasions at OF2, and mainly clades MALV-II and III. MALV-II has been described as a predominant group in marine metabarcoding surveys (Koid et al. 2012; Massana et al. 2011) and as a potential parasite of the class Dinophyceae (Park et al. 2004), which is similar to our results.

The most abundant diatoms in the HTS dataset, *S. marinoi*, *T. nordenskioeldii*, and *Ch. neogracilis,* are described as common diatom species in temperate coastal waters (Throndsen et al. 2007). The TARA Oceans survey included samples across the global ocean euphotic zone south of the 44°N latitude. In that survey, *Thalassiosira* and *Chaetoceros* were also two of the most diverse and cosmopolitan diatom genera, whereas *Skeletonema* was underrepresented compared to microscopy (Malviya et al. 2016). All the abundant diatoms found by HTS in this study are well-known species from Norwegian coastal waters (Throndsen et al. 2007; see Table S1). *Leptocylindrus aporus* was previously named *L. danicus* var. *aporus* but renamed by Nanjappa et al. (2013) and was found in our HTS-dataset.

Dictyochophyceae was the second most represented stramenopile class, represented mainly by *D. speculum*. This is a cold-water species with cosmopolitan distribution (Chang et al. 2003; Glezer 1970; Rigual-Hernández et al. 2010). It can be a major component in coastal and estuarine waters and has previously been linked to fish mortalities (Henriksen et al. 1993). *Dictyocha speculum* is a common species in the Oslofjorden (Throndsen et al. 2007). Another dictyochophyte recorded here, the picoflagellate *F. parvula,* was first described from the English Channel in 2004 (Eikrem et al. 2004). The ichthyotoxic dictyochophyte *P. verruculosa,* found in this study, was recorded for the first time in northern Europe (Germany) in 2000 (Riisberg and Edvardsen 2008). The cold-water species *Pseudochattonella farcimen*, that has formed ichthyotoxic blooms in the Skagerrak since 1998 (Edvardsen et al. 2007) was not recorded in our dataset. These two *Pseudochattonella* species differ in only one position within the V4 18S rRNA gene region (Riisberg and Edvardsen 2008) and may have been clustered together as *P. verruculosa*.

Marine Stramenopiles (MAST) include a large number of predominantly heterotrophic groups and are well represented both in the plankton and in sediments, playing a key role in marine ecosystems (Logares et al. 2012). MAST-1, -3, -4, and -7 have previously been found in open ocean and coastal systems. Although MAST-1 and 3 were the most abundant in our dataset, MAST-4 was present in small abundances. MAST-4 is a dominant group in most oceans but is absent in waters < 4 °C (Massana et al. 2006). Although the Oslofjorden waters are below such temperatures during half of the year, we detected MAST-4-related OTUs in our dataset. This may be due to the presence of the cyanobacteria *Synechoccocus,* which seems to be a prey for MAST-4 (Lin et al. 2012). The MAST-1, -6, -9, and -12, recorded in this study, are important clades in both planktonic and sediment samples (Massana et al. 2014).

In a previous study using metabarcoding with haptophyte-specific primers and with the same samples (Egge et al. 2015a,b), Prymnesiales was the most abundant, frequent, and diverse haptophyte order. As also found by Egge et al. (2015b), the most abundant haptophyte OTUs had best match to *E. huxleyi* and *C. simplex*. Members of Prymnesiales are abundant in the Skagerrak coastal waters and usually have densities over one million cells per litre during summer (Lekve et al. 2006) corresponding to our findings. *Prymnesium faveolatum* was among the most abundant haptophytes in this study, a species not previously recorded by microscopy in the Skagerrak (Artsdatabanken 2018).

The two cryptophytes *T. amphioxeia* and *T. gracilis* were among the 16 most abundant OTUs. *Teleaulax amphioxeia* is well know from brackish waters in Europe (Throndsen et al. 2007), whereas *T. gracilis* was first described in 2012 from the Atlantic coast of Spain (Laza-Martínez et al. 2012) and has not been recorded by microscopy in Norwegian waters (Artsdatabanken 2018).

Within Chlorophyta the species *M. commoda,* a picoflagellate belonging to class Mamiellophyceae, was among the 16 most abundant OTUs. This newly described species was recently separated from *Micromonas pusilla* (Simon et al. 2017). This is the first time *M. commoda* has been recorded from the Skagerrak. *Micromonas pusilla* has been shown to dominate the eukaryotic picoplankton in North Atlantic coastal and Arctic waters (Not et al. 2004, 2005). Our findings, however, suggest that it is *M. commoda,* and not *M. pusilla* that dominates in Oslofjorden (Fig. S3). OTUs assigned to *Micromonas* spp. represented almost 7% of the read abundance in the pico-size fraction in our study, and thus was less dominant compared to the findings by Not et al. (2004), where they used fluorescence in situ hybridisation (FISH) for quantification. The remaining major components of the Chlorophyta community, all pico- or small nanoplankton, have previously been recorded from the North Atlantic, except

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

for *A. nanus,* which is usually found in freshwater (John et al. 2002).

Within Opisthokonta we found an abundant choanoflagellate OTU differing in only two bases (one being in a homopolymer) from *C. natans,* recently genetically characterized by (Nitsche et al. 2017). This difference could be explained as a sequencing error. This species has not previously been recorded from the Oslofjorden, but was found to be a dominant choanoflagellate species in the winter–spring community in the southern Kattegat (Thomsen et al. 2016). In addition, it was the second-most globally abundant choanoflagellate in the Tara Oceans data, exhibiting highest relative abundances at cold-water stations (Nitsche et al. 2017).

Temporal variation

The richness index and Shannon diversity index (H) showed seasonal fluctuations at both depths. They reached maximum values during late summer-early autumn (June–September) and were generally higher at the deep chlorophyll than at the subsurface. Evenness varied slightly through the seasons, with generally high values (> 0.6), which allows the detection also of rare taxa (Caporaso et al. 2012). The high diversity in late summer-autumn has been proposed to be due to the influence of the North Atlantic current that brings in allochtonous plankton taxa (Andersen et al. 2001).

Marked seasonal variations in the protist community were observed at both depths, with a distinct separation between summer and winter. This seasonality revealed by HTS coincides with microscopy cell counts in this study, as well as previous microscopy-based studies in the Oslofjorden (Hasle and Smayda 1960). Taxonomic groups with marked increase in richness in June-September are Haptophyta, Chlorophyta and Cercozoa (Fig. S4). This agrees with Hasle and Smayda (1960) showing coccolithophore haptophytes to be present mainly during June–November and with Egge et al. (2015b) showing highest haptophyte diversity in this period by metabarcoding using haptophyte specific primers. Pico- and nanoplanktonic chlorophytes requires electron microscopy or molecular methods for identification, and there are no previous seasonal surveys on richness or number of taxa of this group from the Skagerrak. Diatoms were found in this study to have highest richness during autumn-winter. Lange et al. (1992) similarly found the highest diversity of diatoms during autumn–winter, which was explained by the period of major advection of foreign species introduced into the Skagerrak by the Jutland and Dooley currents from the North Atlantic. Highest diversity of dinoflagellates was in the autumn (September–December, Fig. S4 and Table S5), which has also previously been found by microscopy (Throndsen et al. 2007).

Large changes in proportional abundance of the major taxonomic groups were observed between samplings. This could be explained by the long (monthly) sampling intervals (Countway et al. 2005). The seasonal dynamics here are consistent with previous observations such that diatoms dominate during the spring bloom, whereas dinoflagellates have their highest proportion in autumn–winter (Hasle and Smayda 1960). Haptophyte proportional abundance peaked in June (Fig. 2). Members of *Chrysochromulina* were most abundant during the summer (Fig. 3), which corresponds with previous findings (Lekve et al. 2006) suggesting that they are favoured by low nutrient concentrations and high freshwater influence (Edvardsen and Paasche 1998).

Profound differences in community composition by season were also indicated at both depths by ordination analyses (NMDS), where four clear clusters were observed (Fig. 6B). According to SIMPER analyses 10 OTUs generated most of the differences between seasons, corresponding to the most abundant OTUs. CCA ordination and PERMANOVA analyses showed that temperature and salinity influenced the community structure. Temperature and salinity displayed negative correlations with nutrient concentrations which indicate that terrestrial and riverine inputs bring nutrients to the Outer Oslofjorden. As proposed by Simon et al. (2015), the detection of few correlations may result from biotic factors (e.g. predation, mutualism, parasitism and virus) not being included in this study.

Trophic status through the season and by depth

Ratios between percentage of autotrophic and heterotrophic OTUs was similar through the two years (Fig. S5). Heterotrophs were more diverse than autotrophs through most of the sampling period except in June 2010 and 2011. Similar results were found in the TARA Oceans expedition where heterotrophic groups contributed more to the richness than autotrophic (de Vargas et al. 2015). The relative abundance of trophic groups showed a clear seasonal pattern, especially at the subsurface. The proportion of reads assigned to autotrophic groups was highest during winter to spring, the period with highest chl-*a* concentrations (> 2 μg/l), and lowest surface water temperatures (−1 to + 5 °C), and lowest in the late autumn to early winter when water temperature was gradually decreasing from 12 to 0 °C. The opposite pattern was found for dinoflagellates. This pattern is similar to that found by Piredda et al. (2017) studying protist plankton communities in Gulf of Naples, Italy.

Novel records for Scandinavian waters

We recorded 69 potentially new species and 40 potentially new genera for the Skagerrak area that are not registered in the Norwegian Biodiversity Information Centre (Artsdatabanken 2018) nor the Nordic Microalgae and Aquatic Protozoa (NOD) database (Karlson et al. 2015). The number of pelagic and benthic protist species recorded in Norwegian marine waters based on microscopy are estimated to ca. 1,200, according to the Norwegian Species Information Centre (Artsdatabanken 2018, Antall arter i norsk natur 2016). About 1,020 of these species belong to a phylum with microalgal representatives. Throndsen et al.

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

(2007) estimated that more than 700 phytoplankton species may be present in Norwegian coastal waters. The approximately 2,000 planktonic protist OTUs reported here in the Skagerrak, passing a 45 μm sieve and collected on 0.8 μm pore size filters, and after a strict read filtering, are considerably more than the number of protists observed in the microscopy through all times, but still of the same magnitude. Furthermore, some closely related species have identical V4 18S rRNA gene region (e.g. among haptophytes Egge et al. 2015a, and dinoflagellates, Edvardsen et al., unpubl. data), and one OTU may thus represent more than one species. In our study, however, several OTUs represent the same species, which reduces the number of taxa. This study focuses on the smallest protists taxa that are difficult to identify in the light microscope. Many species of the important groups, dinoflagellates, diatoms and ciliates may be larger than 45 μm, and are not included in metabarcoding data. Since previous studies are mainly based on morphological techniques, some parasitic groups (e.g. Syndiniales) have been overlooked and recorded for the first time with metabarcoding.

High-throughput sequencing versus microscopy comparison

A few previous studies have compared metabarcoding to microscopy quantitative surveys, focusing on Arctic lakes (Majaneva et al. 2012) or targeting a specific group (Bachy et al. 2013; Young et al. 2014). Our study is one of the first comparing microscopy cell counts and HTS data of several protist classes from marine waters during a long-term time series. Our results showed some clear differences between the two approaches. Bacillariophyceae and Euglenophyceae proportional abundances were underrepresented by HTS of the nano-picoplankton, and the latter was almost overlooked by this approach. The only euglenoid genus found with both approaches was *Eutreptiella*, which ranges between 12 and 115 μm in cell size (Throndsen et al. 2007). The underestimation by metabarcoding can be explained by the < 45 μm prefiltration of the water samples for RNA extraction. It can also partly be explained by the V4 18S rRNA gene PCR primers used by Stoeck et al. 2010; that seem to be poor in amplifying members of Euglenophyta compared to amplification using chloroplast gene targeting primers (Amaral-Zettler et al. 2011). Bacillariophyceae was found abundant by both approaches, but significant differences through the year were found between the methods. Such discrepancies were also found for diatoms in freshwater studies (Xiao et al. 2014). Seven large, chain-forming *Chaetoceros* species were only detected by microscopy in our study. Prefiltration can explain this underestimation by metabarcoding also in this case.

In contrast, the classes Chrysophyceae and Dictyochophycea were favoured in the metabarcoding compared to microscopy. Of Chrysophyceae, only the genus *Dinobryon*, which forms large colonies, was observed by microscopy. *Dinobryon* sequences were included in our reference sequence database, but it was not detected by HTS. Many other OTUs were however assigned to Chrysophyceae.

The class Dictyophyceae was also detected by both methods but favoured by HTS in both relative read abundance and richness. Only *D. speculum* was observed by microscopy, while by metabarcoding, *Florenciella*, *Pseudochattonella, Apedinella*, and several unclassified dictyocophytes were also detected.

Dinoflagellates have a wide size distribution. In LM, we included all size groups that could be identified under a light microscope (larger than c. 15 μm), whereas in HTS we analysed the 45–0.8 μm size fraction. Compared to taxa with similar cell size, dinoflagellates have large genomes (Hackett et al. 2004) and putatively high rRNA gene copy number (Prokopowich et al. 2003) thus, an overrepresentation in metabarcoding surveys based on DNA may be expected. However, in this study, rRNA was isolated from the plankton, converted to cDNA by RT-PCR and then cDNA was amplified in the PCR, which is expected to reduce the bias for organisms with large genome size (Not et al. 2009). Indeed, the log-ratio of proportion of reads to cell count-based biovolume of Dinophyceae was 0.99, and the dinoflagellates were less overrepresented in the HTS dataset compared to microscopy, than Cryptophyceae and Dictyocophyceae.

Another important aspect in this comparison is that in LM only a small volume was analysed (10 ml) allowing very few species to be observed compared to HTS, where 20 liters were filtrated and RNA from ca. 10 liters were used in the further processing (RT-PCR to cDNA).

CONCLUDING REMARKS

A high diversity of protists was revealed by metabarcoding compared to previous surveys by microscopy through a decade. About 70 protist taxa was recorded for the first time. Metabarcoding can reveal a detailed protist composition and allows large sample sizes. The protist community composition and relative abundance in the Oslofjorden show large variation though the year. There was a difference in protist community structure between the two sampled depths, with higher proportional abundance of autotrophs found in the subsurface than at the deep chlorophyll maximum. The seasonal pattern in relative read abundance of major phytoplankton groups was well in accordance with microscopy biovolumes of the same groups. However, when comparing proportion of reads with biovolumes for major phytoplankton groups, some are overrepresented and other underrepresented in HTS versus microscopy. As neither method shows the full picture, they should be used complementary to each other. More reference sequences, connecting a genotype to a morphotype, are needed to enable a more precise taxonomic identification and reducing the number of OTUs with best match to an "uncultured marine eukaryote". This may also improve the assessment of the actual taxon richness instead of OTU richness. This study may serve as a baseline for future studies and

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

monitoring to reveal effects of environmental and climate change.

ACKNOWLEDGMENTS

Financial support was given by the University of Oslo, MN Faculty to SGS and BE, by the Research Council of Norway through grant 190307 HAPTODIV to BE, EDE, and WE, by grant 225956/E10 MIKROPOLAR to EDE and BE, and by the EU project BioMarKs (2008-6530, ERA-net Biodiversa, EU) to BE and RL. We thank Rita Amundsen, captain Sindre Holm and the crew onboard R/V Trygve Braarud for assistance during sampling, Sissel Brubak and Berit Kaasa for analysing Chlorophyll a and nutrient samples, Colomban de Vargas for leading the BioMArks project, Anette Engesmo for calculating phytoplankton biovolumes from cell numbers and Tom Andersen for statistical advice. High-throughput sequencing was performed at the Norwegian Sequencing Centre and bioinformatics analyses on the University of Oslo Abel cluster.

LITERATURE CITED

Amaral-Zettler, L. A., Zettler, E. R., Theroux, S. M., Palacios, C., Aguilera, A. & Amils, R. 2011. Microbial community structure across the tree of life in the extreme Río Tinto. *ISME J.*, 5:42–50.

Andersen, I., Berge, J. A., Andersen, J. H., Berntsson, I., Danielssen, D., Foverskov, S., Fyrberg, L., Gjøsæter, J., Granmo, A., Green, N., Hansen, O. S., Hylland, K., Håkansson, B., Johannessen, T., Karlson, B., Knutsen, J. A., Knutzen, J., Magnusson, J., Molvaer, J., Pedersen, B., Sjöberg, B., Szaron, J., Torstensen, E., Tveite, S. & Ærtebjerg, G. 2001. The Skagerrak – environmental state and monitoring prospects. Göteborg. 118 p.

Artsdatabanken. 2018. The Norwegian Biodiversity Information Centre, Norway. https://www.artsdatabanken.no/.

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

Backe-Hansen, P. & Throndsen, J. 2002. Pico- and nanoplankton from the inner Oslofjord, eastern Norway, including description of two new species of *Luffisphaera* (incerta sedis). *Sarsia North Atl. Mar. Sci.*, 87:55–64.

Berger, S. A., Krompass, D. & Stamatakis, A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.*, 60:291–302.

Braarud, T. & Bursa, A. 1939. The phytoplankton of the Oslo Fjord 1933–1934. *Hvalrådets Skr.*, 19:6–63.

Braarud, T., Gaarder, K. R. & Grøntved, J. 1953. The phytoplankton of the North Sea and adjacent waters in May 1948. Conseil permanent international pour l'exploration de la mer, Copenhague. 93 pp.

Bratbak, G., Jacquet, S., Larsen, A., Pettersson, L. H., Sazhin, A. F. & Thyrhaug, R. 2011. The plankton community in Norwegian coastal waters-abundance, composition, spatial distribution and diel variation. *Cont. Shelf Res.*, 31:1500–1514.

Bray, R. & Curtis, T. 1957. An Ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.*, 27:325–349.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Gonzalez Peña, A., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7:335–336.

Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R. & Gilbert, J. A. 2012. The Western English Channel contains a persistent microbial seed bank. *ISME J.*, 6:1089–1093.

Caron, D. A., Countway, P. D., Savai, P., Gast, R. J., Schnetzer, A., Moorthi, S. D., Dennett, M. R., Moran, D. M. & Jones, A. C. 2009. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl. Environ. Microbiol.*, 75:5797–5808.

Chang, F. H., Zeldis, J., Gall, M. & Hall, J. 2003. Seasonal and spatial variation of phytoplankton assemblages, biomass and cell size from spring to summer across the north-eastern New Zealand continental shelf. *J. Plankton Res.*, 25:737–758.

Countway, P. D., Gast, R. J., Savai, P. & Caron, D. A. 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. *J. Eukaryot. Microbiol.*, 52:95–106.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P. & Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605-1-11.

Dittami, S. M., Hostyeva, V., Egge, E. S., Kegel, J. U., Eikrem, W. & Edvardsen, B. 2013. Seasonal dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. *Environ. Sci. Pollut. Res.*, 20:6719–6732.

Dragsund, E., Aspholm, O., Tangen, K., Bakke, S. M., Heier, L. & Jensen, T. 2006. Overvåking av eutrofitilstanden i Ytre Oslofjord. Femårsrapport 2001–2005, nr. 2006-0831. Oslo. 127 p.

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

Edvardsen, B., Eikrem, W., Shalchian-Tabrizi, K., Riisberg, I., Johnsen, G., Naustvoll, L. & Throndsen, J. 2007. *Verrucophora farcimen* gen. et sp. nov. (Dictyochophyceae, Heterokonta) - a bloom-forming ichthyotoxic flagellate from the Skagerrak, Norway. *J. Phycol.*, 43:1054–1070.

Edvardsen, B. & Paasche, E. 1998. Bloom dynamics and physiology of *Prymnesium* and *Chrysochromulina*. *NATO ASI Ser. G. Ecol. Sci.*, 41:193–208.

Edvardsen, B., Shalchian-Tabrizi, K., Jakobsen, K. S., Medlin, L. K., Dahl, E., Brubak, S. & Paasche, E. 2003. Genetic variability and molecular phylogeny of *Dinophysis* species (Dinophyceae) from norwegian waters inferred from single cell analyses of rDNA. *J. Phycol.*, 39:395–408.

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C. & Edvardsen, B. 2013. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS ONE*, 8:e74371.

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Egge, E. S., Eikrem, W. & Edvardsen, B. 2015a. Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *J. Eukaryot. Microbiol.*, 62:121–140.

Egge, E. S., Johannessen, T. V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R. A. & Edvardsen, B. 2015b. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol. Ecol.*, 24:3026–3042.

Eikrem, W., Romari, K., Latasa, M., Gall, F., Le Throndsen, J. & Vaulot, D. 2004. *Florenciella parvula* gen. et sp. nov. (Dictyochophyceae, Heterokontophyta), a small flagellate isolated from the English Channel. *Phycologia*, 43:658–668.

Epstein, S. & López-García, P. 2008. "Missing" protists: a molecular prospective. *Biodivers. Conserv.*, 17:261–276.

Glezer, Z. I. 1970. Silicoflagellatophyceae. *In:* Gollerbakh, M. M. (ed.), Cryptogamic plants of the U.S.S.R, Vol. 7. Israel Program for Scientific Translation, Jerusalem. p. 1–363.

Gran-Stadniczeñko, S., Šupraha, L., Egge, E. D. & Edvardsen, B. 2017. Haptophyte diversity and vertical distribution explored by 18S and 28S ribosomal RNA gene metabarcoding and scanning electron microscopy. *J. Eukaryot. Microbiol.*, 64:514–532.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A. L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41:D597–D604.

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

Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. 2004. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.*, 91:12.

Hasle, G. R. & Smayda, T. J. 1960. The annual phytoplankton cycle at Drøbak, Oslofjord. *Nytt Mag. Bot.*, 8:53–75.

Henriksen, P., Knipschildt, F., Moestrup, Ø. & Thomsen, H. A. 1993. Autecology, life history and toxicology of the silicoflagellate *Dictyocha speculum* (Silicoflagellata, Dictyochophyceae). *Phycologia*, 32:29–39.

Hjort, J. & Gran, H. H. 1900. Hydrographic-biological investigations of the Skagerrak and the Christiania Fiord. *Report on Norwegian Fishery- and Marine-Investigations*, Vol. I No 2. 49 pp.

John, D. M., Whitton, B. A. & Brook, A. J. (ed.) 2002. The freshwater algal flora of the British Isles: an identification guide to freshwater and terrestrial algae, Vol. 1. Cambridge University Press, Cambridge. 702 pp.

Jones, T., Parrish, J. K., Punt, A. E., Trainer, V. L., Kudela, R., Lang, J., Brancato, M. S., Odell, A. & Hickey, B. 2017. Mass mortality of marine birds in the Northeast Pacific caused by *Akashiwo sanguinea*. *Mar. Ecol. Prog. Ser.*, 579:111–127.

Karlson, B., Andreasson, A., Johansen, M., Karlberg, M., Loo, A. & Skjevik, A.-T. 2015. Nordic Microalgae. http://www.nordicmicroalgae.org

Katoh, K., Asimenos, G. & Toh, H. 2009. Multiple alignment of DNA sequences with MAFFT. *In:* Posada, D. (ed.), Bioinformatics for DNA Sequence Analysis. Humana Press, Totowa, NJ. p. 39–64.

Koid, A., Nelson, W. C., Mraz, A. & Heidelberg, K. B. 2012. Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. *Appl. Environ. Microbiol.*, 78:3958–3965.

Kuylenstierna, M. & Karlson, B. 1994. Seasonality and composition of pico- and nanoplanktonic cyanobacteria and protists in the Skagerrak. *Bot. Mar.*, 37:17–33.

Lange, C. B., Hasle, G. R. & Syvertsen, E. E. 1992. Seasonal cycle of diatoms in the Skagerrak, North Atlantic, with emphasis on the period 1980–1990. *Sarsia*, 77:173–187.

Laza-Martínez, A., Arluzea, J., Miguel, I. & Orive, E. 2012. Morphological and molecular characterization of *Teleaulax gracilis* sp. nov. and *T. minuta* sp. nov. (Cryptophyceae). *Phycologia*, 51:649–661.

Lekve, K., Bagøien, E., Dahl, E., Edvardsen, B., Skogen, M. D. & Stenseth, N. C. 2006. Environmental forcing as a main determinant of bloom dynamics of the *Chrysochromulina* algae. *Proc. Biol. Sci.*, 273:3047–3055.

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

Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C. & Massana, R. 2012. Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J.*, 6:1823–1833.

López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, 409:603–607.

Majaneva, M., Rintala, J. M., Piisilä, M., Fewer, D. P. & Blomster, J. 2012. Comparison of wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on sequencing of the 18S rRNA gene. *Polar Biol.*, 35:875–889.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A. & Bowler, C. 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl Acad. Sci.*, 113:E1516–E1525.

Massana, R. 2015. Protistan diversity in environmental molecular surveys. *In:* Ohtsuka, S., Suzaki, T., Horiguchi, T., Suzuki, N. & Not, F. (ed.), Marine Protists: Diversity and Dynamics. Springer, Tokyo. p. 3–21.

Massana, R., del Campo, J., Sieracki, M. E., Audic, S. & Logares, R. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J.*, 8:854–866.

Massana, R., Pernice, M., Bunge, J. A. & del Campo, J. 2011. Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. *ISME J.*, 5:184–195.

Massana, R., Terrado, R., Forn, I., Lovejoy, C. & Pedros-Alio, C. 2006. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ. Microbiol.*, 8:1515–1522.

Medlin, L. K. & Kooistra, W. H. C. F. 2010. Methods to estimate the diversity in the marine photosynthetic protist community with illustrations from case studies: a review. *Diversity*, 2:973–1014.

Nanjappa, D., Kooistra, W. H. C. F. & Zingone, A. 2013. A reappraisal of the genus *Leptocylindrus* (Bacillariophyta), with the addition of three species and the erection of *Tenuicylindrus* gen. nov. *J. Phycol.*, 49:917–936.

Nitsche, F., Thomsen, H. A. & Richter, D. J. 2017. Bridging the gap between morphological species and molecular barcodes - Exemplified by loricate choanoflagellates. *Eur. J. Protistol.*, 57:26–37.

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Not, F., del Campo, J., Balagué, V., de Vargas, C. & Massana, R. 2009. New insights into the diversity of marine picoeukaryotes. *PLoS ONE*, 4:e7143.
Not, F., Latasa, M., Marie, D., Cariou, T., Vaulot, D. & Simon, N. 2004. A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Appl. Environ. Microbiol.*, 70:4064–4072.
Not, F., Massana, R., Latasa, M., Marie, D., Colson, C., Eikrem, W., Pedrós-Alió, C., Vaulot, D. & Simon, N. 2005. Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol. Oceanogr.*, 50:1677–1686.
Not, F., Siano, R., Kooistra, W. H. C. F., Simon, N., Vaulot, D. & Probert, I. 2012. Diversity and ecology of eukaryotic marine phytoplankton. *In:* Piganeau, G. (ed.), Advances in Botanical Research, Vol. 64. Academic Press, Waltham, MA. p. 1–53.
Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H. & Wagner, H. 2017. vegan: Community Ecology Package. R package version 2.4-3.
Olenina, I., Hajdu, S., Edler, L., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M., Jaanus, A., Kokkonen, P., Ledaine, I. & Niemkiewicz, E. 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. *HELCOM Balt. Sea Environ. Proc.*, 106.
Paasche, E. & Østergren, I. 1980. The annual cycle of plankton diatom growth and silica production in the inner Oslofjord. *Limnol. Oceanogr.*, 25:481–494.
Park, M. G., Yih, W. & Coats, D. W. 2004. Parasites and phytoplankton, with special emphasis on dinoflagellate infections. *J. Eukaryot. Microbiol.*, 51:145–155.
Piredda, R., Tomasino, M. P., D'Erchia, A. M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W. H. C. F., Sarno, D. & Zingone, A. 2017. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiol. Ecol.*, 93:1–14.
Prokopowich, C. D., Gregory, T. R. & Crease, T. J. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46:48–50.
Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, 12:38.
R Development Core Team. 2017. R: A language and environment for statistical computing v.3.4.1. *R Found. Stat. Comput.*
Rigual-Hernández, A. S., Bárcena, M. A., Sierro, F. J., Flores, J. A., Hernández-Almeida, I., Sanchez-Vidal, A., Palanques, A. & Heussner, S. 2010. Seasonal to interannual variability and geographic distribution of the silicoflagellate fluxes in the Western Mediterranean. *Mar. Micropaleontol.*, 77:46–57.
Riisberg, I. & Edvardsen, B. 2008. Genetic variation in bloom-forming ichthyotoxic *Pseudochattonella* species (Dictyochophyceae, Heterokonta) using nuclear, mitochondrial and plastid DNA sequence data. *Eur. J. Phycol.*, 43:413–422.
Shannon, C. E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423.
Simon, N., Cras, A. L., Foulon, E. & Lemée, R. 2009. Diversity and evolution of marine phytoplankton. *C. R. Biol.*, 332:159–170.
Simon, N., Foulon, E., Grulois, D., Six, C., Desdevises, Y., Latimier, M., Le Gall, F., Tragin, M., Houdan, A., Derelle, E., Jouenne, F., Marie, D., Le Panse, S., Vaulot, D. & Marin, B. 2017. Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the type species *M. pusilla.* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and description of two new species based on the genetic and phenotypic characterization of cultured isolates. *Protist*, 168:612–635.
Simon, M., López-García, P., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P. & Jardillier, L. 2015. Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J.*, 9:1941–1953.
Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312–1313.
Stoeck, T., Bass, D., Nebel, M., Christen, R. & Meredith, D. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.*, 19:21–31.
Strickland, J. D. H. & Parsons, T. R. 1972. A Practical Handbook of Seawater Analysis. Fisheries Research Board of Canada, Ottawa. 328 pp.
Tangen, K. 1977. Blooms of *Gyrodinium aureolum* (Dinophygeae) in North European waters, accompanied by mortality in marine organisms. *Sarsia*, 63:123–133.
Tennekes, M. 2017. treemap: Treemap Visualization. R package version 2.4-2.
Thomsen, H. A., Nitsche, F. & Richter, D. J. 2016. Seasonal occurrence of loricate choanoflagellates in Danish inner waters. *Protist*, 167:622–638.
Throndsen, J., Hasle, G. R. & Tangen, K. 2007. Phytoplankton of Norwegian coastal waters. Almater Forlag AS, Oslo.
Utermöhl, H. 1958. Zur Vervollkommung der quantitativen Phytoplankton-Methodik. *Int. Ver. Theor. Angew. Limnol. Verh.*, 9:1–38.
Xiao, X., Sogge, H., Lagesen, K., Tooming-Klunderud, A., Jakobsen, K. S. & Rohrlack, T. 2014. Use of high throughput sequencing and light microscopy show contrasting results in a study of phytoplankton occurrence in a freshwater environment. *PLoS ONE*, 9:e106510.
Young, J. R., Liu, H., Probert, I., Aris-Brosou, S. & de Vargas, C. 2014. Morphospecies versus phylospecies concepts for evaluating phytoplankton diversity: the case of the Coccolithophores. *Cryptogam. Algol.*, 35:353–377.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Temporal variation in physicochemical parameters measured at the study site.
Figure S2. Heatmaps representing the temporal variations of all studied protist subgroups at the Outer Oslofjorden OF2 station.
Figure S3. Maximum likelihood RAxML-EPA (Evolutionary Placement Algorithm) trees of the 16 most abundant OTUs at the OF2 station.
Figure S4. Succession of proportions of OTUs of the 18 major taxonomic groups across the 21 temporal samples at station OF2.
Figure S5. Succession of proportions of OTUs of the different trophic groups across the 21 temporal samples at station OF2.
Figure S6. Venn diagram showing the unique and shared operational taxonomic units (OTUs) at OF2 station among the four different seasons during the study period.

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Figure S7. Tree maps displaying the taxonomic composition of the complete Outer Oslofjorden OF2 station protist microscopy dataset at class levels.
Table S1. Protist V4 18S rRNA gene OTUs recorded at station OF2 during the sampling period September 2009 to June 2011 with number of reads in each sample and taxonomic placement. Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.
Table S2. Richness, Shannon, and Evenness diversity results with Welch two-tailed test to check correspondence between two studied depths (subsurface and bottom of chlorophyll maximum).
Table S3. PERMANOVA results with permutation of environmental factors for two studied depths (subsurface and bottom of chlorophyll maximum).
Table S4. New taxa recorded for the Skagerrak. Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.
Table S5. Microscopical biovolumes and cell counts at station OF2. Biovolumes (in μm^3) were estimated from cell counts using the HELCOM 2006 protocol (Olenina et al. 2006). Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.
Table S6. Welch two sample *t*-test results, to test correspondence between light microscopy cell counts and metabarcoding proportional read abundance.
File S1. Scripts for the bioinformatics pipeline in Qiime, phylogenetic analyses and statistical analyses in R. Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.
File S2. Environmental data; salinity, temperature, density, fluorescence, nutrients by depth. Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.
File S3. OTU fasta file with DNA sequences of all OTUs with more than 10 reads. Available from FigShare at the site: https://figshare.com/s/31d2ea0e5a303fd8813e.

© 2018 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as revealed by Metabarcoding and Microscopy

SUPLEMENTARY INFORMATION

Sandra Gran-Stadniczeñko[a], Elianne Dunthorn Egge[a], Vladyslava Hostyeva[b], Ramiro Logares[c], Wenche Eikrem[a,b], Bente Edvardsen[a].

[a] Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway
[b] Norwegian Institute for Water Research, Gaustadalléen 21, 0349, Oslo, Norway.
[c] Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), E-08003 Barcelona, Catalonia, Spain.

Figure S1. Temporal variation of physicochemical parameters measured at the study site. Dotted line indicates the bottom of the deep chlorophyll maximum (DC) during the study period.

Figure S2. Heatmaps representing the temporal variations of all studied protist subgroups at the Outer Oslofjorden OF2 station. Numbers represent proportional read abundances. A: Ciliophora, B: Dinophyta, C: Chlorophyta, D: Streptophyta, E: Euglenozoa, F: Centroheliozoa, G: Cryptophyta, H: Haptophyta, I: Katablepharidophyta, J: Telonemia, K: Choanoflagellida, L: Fungi, M: Mesomycetozoa, N: Picozoa, O: Cercozoa, P: Radiolaria, Q: Ochrophyta, R: Heterotrophic Stramenopiles.

Figure S3. Maximum likelihood RAxML-EPA (Evolutionary Placement Algorithm) trees of the 16 most abundant OTUs at the OF2 station.

Figure S4. Succession of proportions of OTUs of the 18 major taxonomic groups across the 21 temporal samples at station OF2.

Figure S5. Succession of proportions of OTUs of the different trophic groups across the 21 temporal samples at station OF2.

Figure S6. Venn diagram showing the unique and shared operational taxonomic units (OTUs) at OF2 station among the four different seasons during the study period.

Figure S7. Tree maps displaying the taxonomic composition of the complete Outer Oslofjorden OF2 station protist microscopy dataset at class levels. A. Categories. B. Cell counts. C. Biovolumes.

Table S1. Protist V4 18S rRNA gene OTUs recorded at station OF2 during the sampling period September 2009 to June 2011 with number of reads in each sample and taxonomic placement. OTUs marked in red were removed after subsampling.

Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

Table S2. Richness, Shannon and Evenness diversity results with Welch two-tailed test to check correspondence between two studied depths (subsurface and bottom of chlorophyll maximum).

Table S3. PERMANOVA results with permutation of environmental factors for two studied depths (subsurface and bottom of chlorophyll maximum).

Table S4. New taxa recorded for the Skagerrak. A) genus and species names B) new records by groups. Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

Table S5. Microscopical biovolumes and cell counts at station OF2. Biovolumes (in µm3) were estimated from cell counts using the HELCOM 2006 protocol (Olenina et al. 2006). Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

Table S6. Welch two sample t-test results, to test correspondence between light microscopy cell counts and metabarcoding proportional read abundance.

File S1. Scripts for the bioinformatics pipeline in Qiime, phylogenetic analyses and statistical analyses in R. Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

File S2. Environmental data; salinity, temperature, density, fluorescence, nutrients by depth. Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

File S3. OTU fasta file with DNA sequences of all OTUs with more than 10 reads. Archived at https://figshare.com/s/31d2ea0e5a303fd8813e.

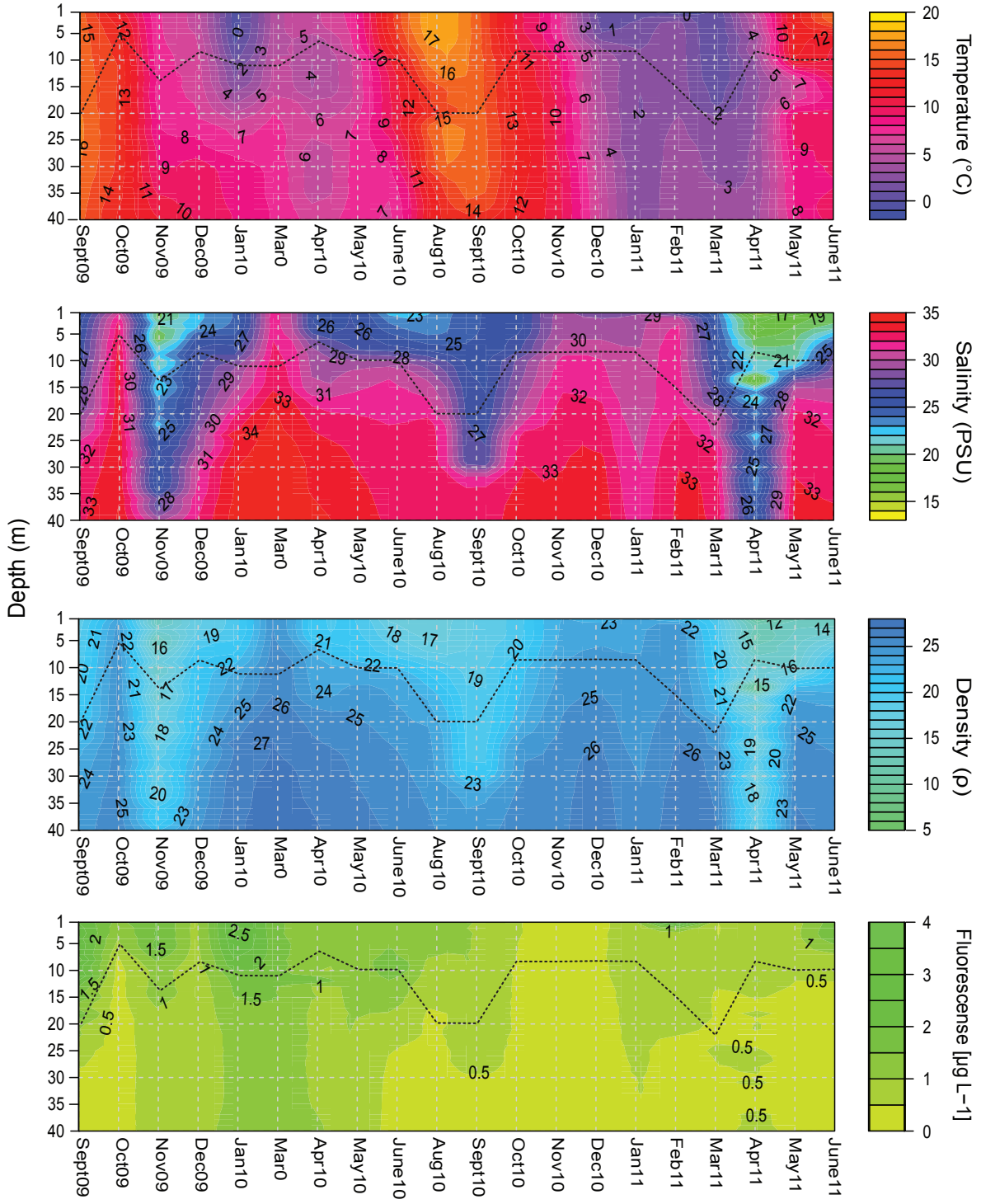


Figure S1. Temporal variation of physicochemical parameters measured at the study site. Dotted line indicates the bottom of the deep chlorophyll maximum (DC) during the study period.

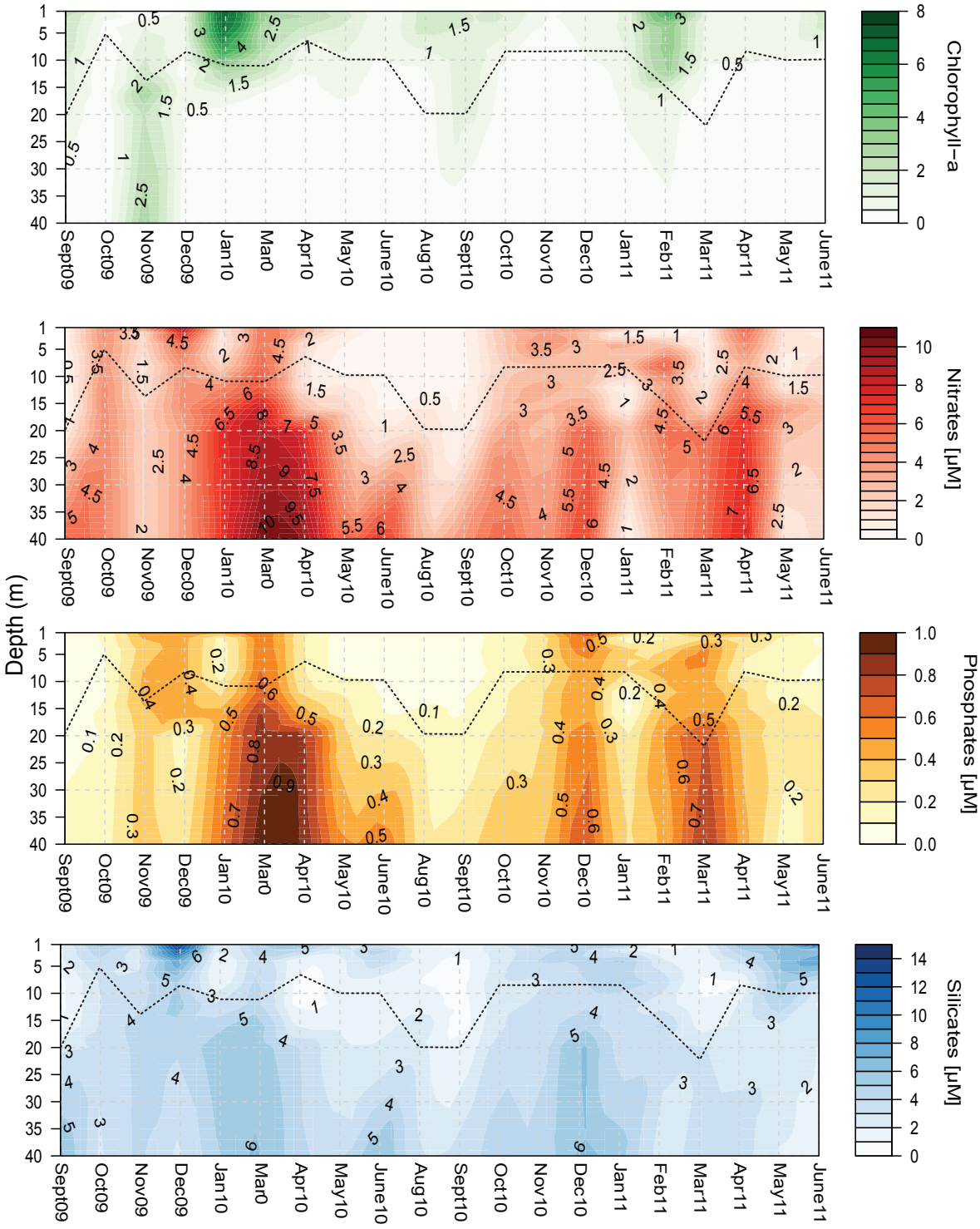


Figure S1 continued. Temporal variation of physicochemical parameters measured at the study site. Dotted line indicates the bottom of the deep chlorophyll maximum (DC) during the study period.

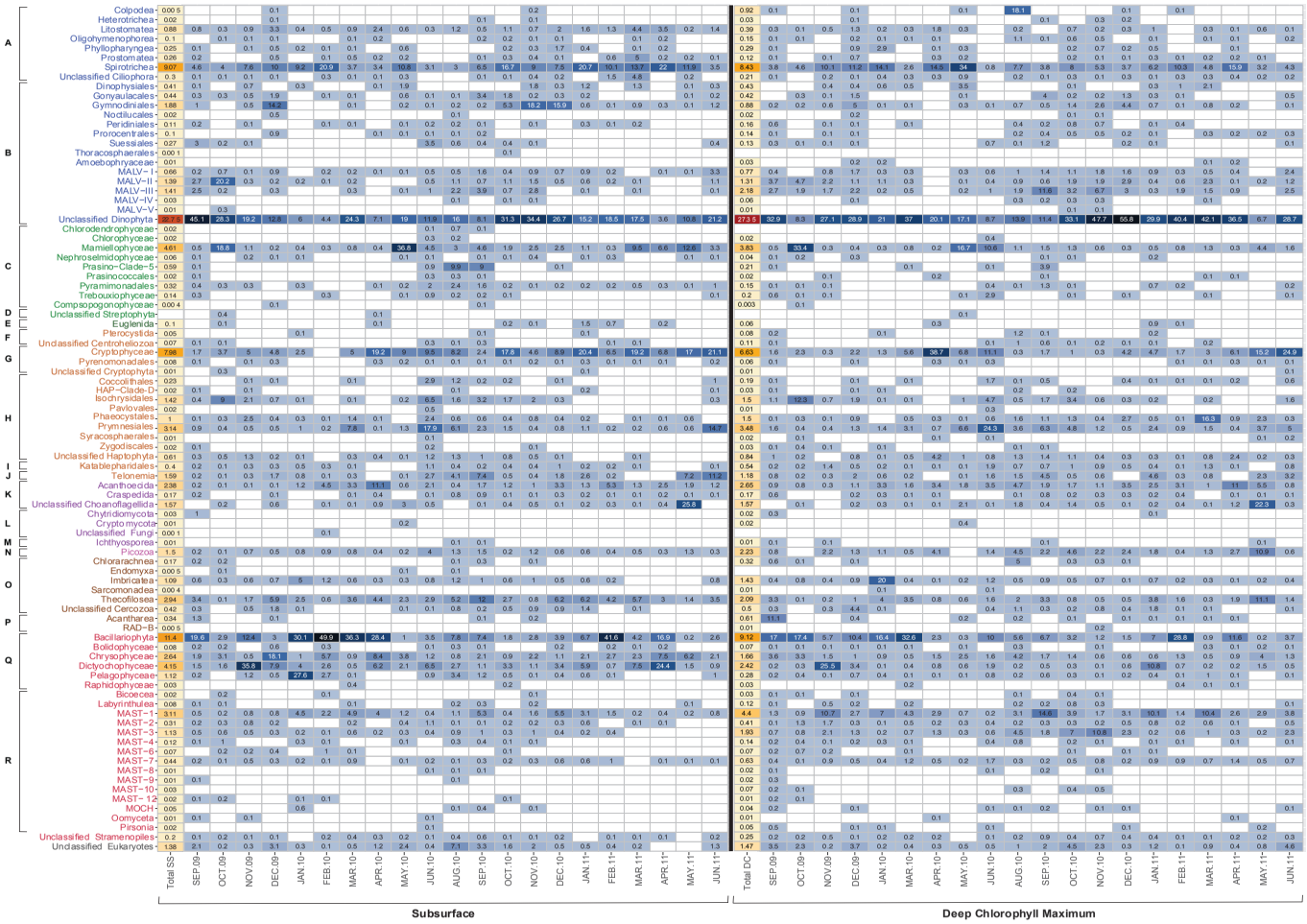


Figure S2. Heatmaps representing the temporal variations of all studied protist subgroups at the Outer Oslofjorden OF2 station. Numbers represent proportional read abundances. A: Ciliophora, B: Dinophyta, C: Chlorophyta, D: Streptophyta, E: Euglenozoa, F: Centroheliozoa, G: Cryptophyta, H: Haptophyta, I: Katablepharidophyta, J: Telonemia, K: Choanoflagellida, L: Fungi, M: Mesomycetozoa, N: Picozoa, O: Cercozoa, P: Radiolaria, Q: Ochrophyta, R: Heterotrophic Stramenopiles.

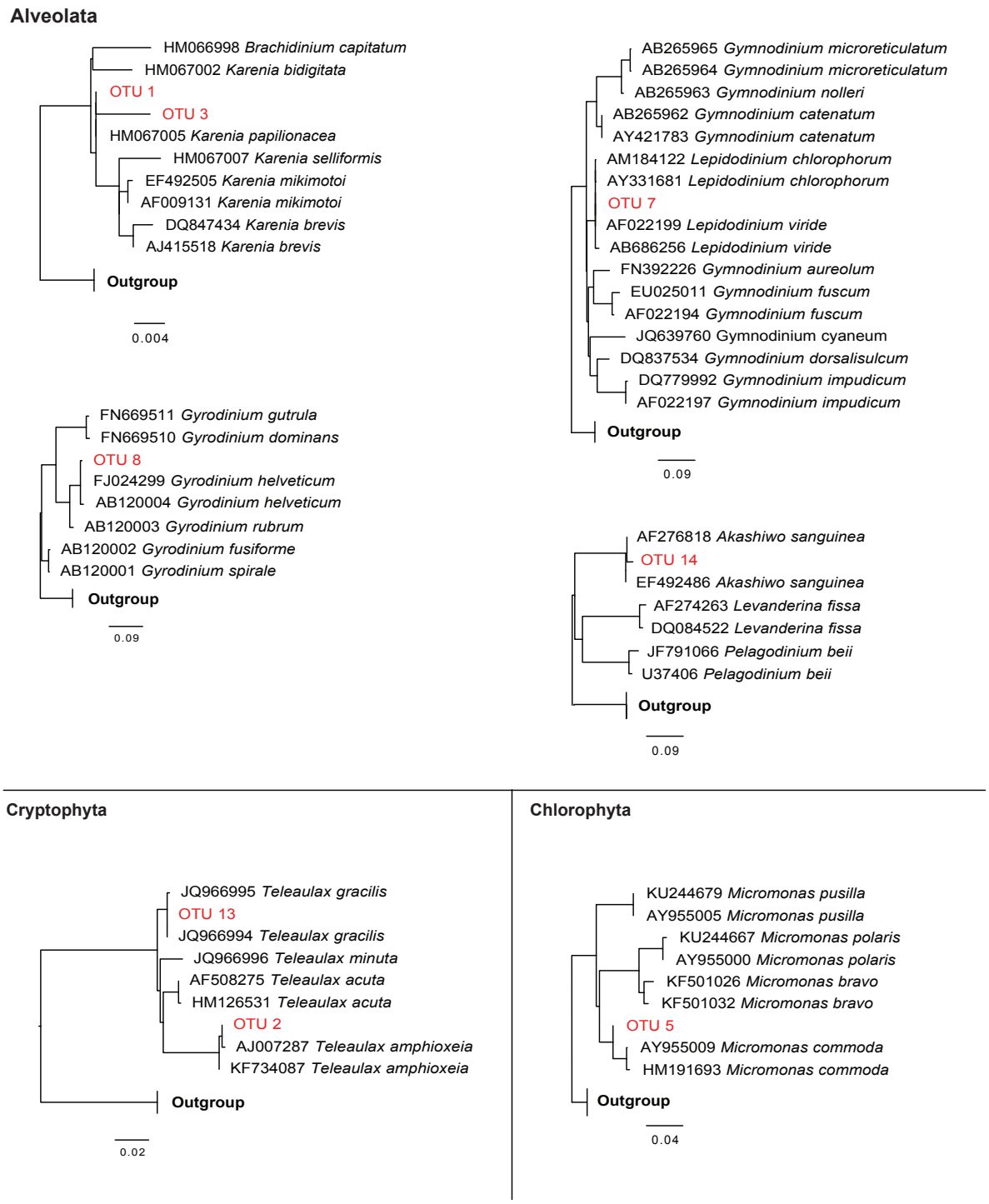


Figure S3. Maximum likelihood RAxML-EPA (Evolutionary Placement Algorithm) trees of the 16 most abundant OTUs at the OF2 station.

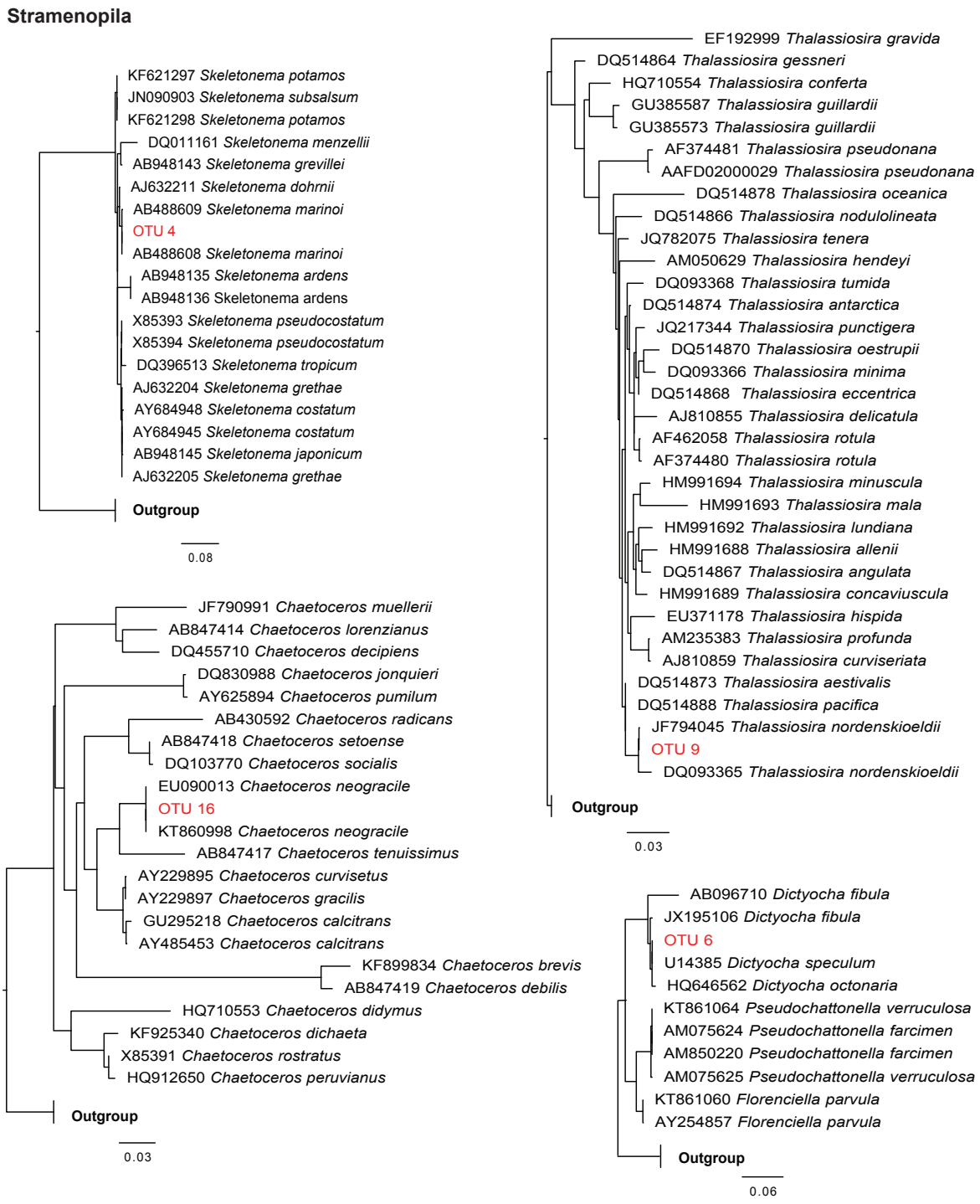


Figure S3 continued. Maximum likelihood RAxML-EPA (Evolutionary Placement Algorithm) trees of the 16 most abundant OTUs at the OF2 station.

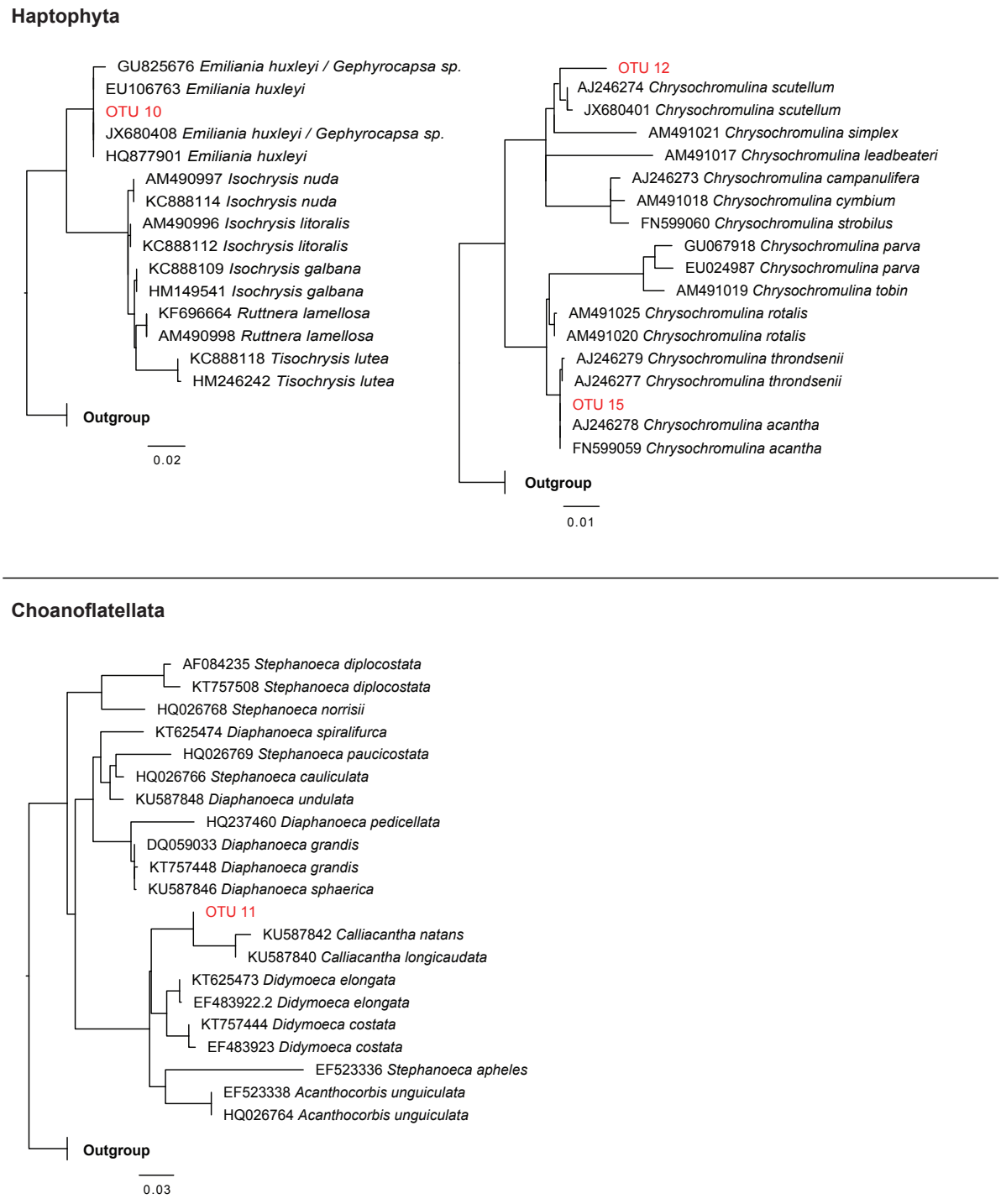


Figure S3 continued. Maximum likelihood RAxML-EPA (Evolutionary Placement Algorithm) trees of the 16 most abundant OTUs at the OF2 station.

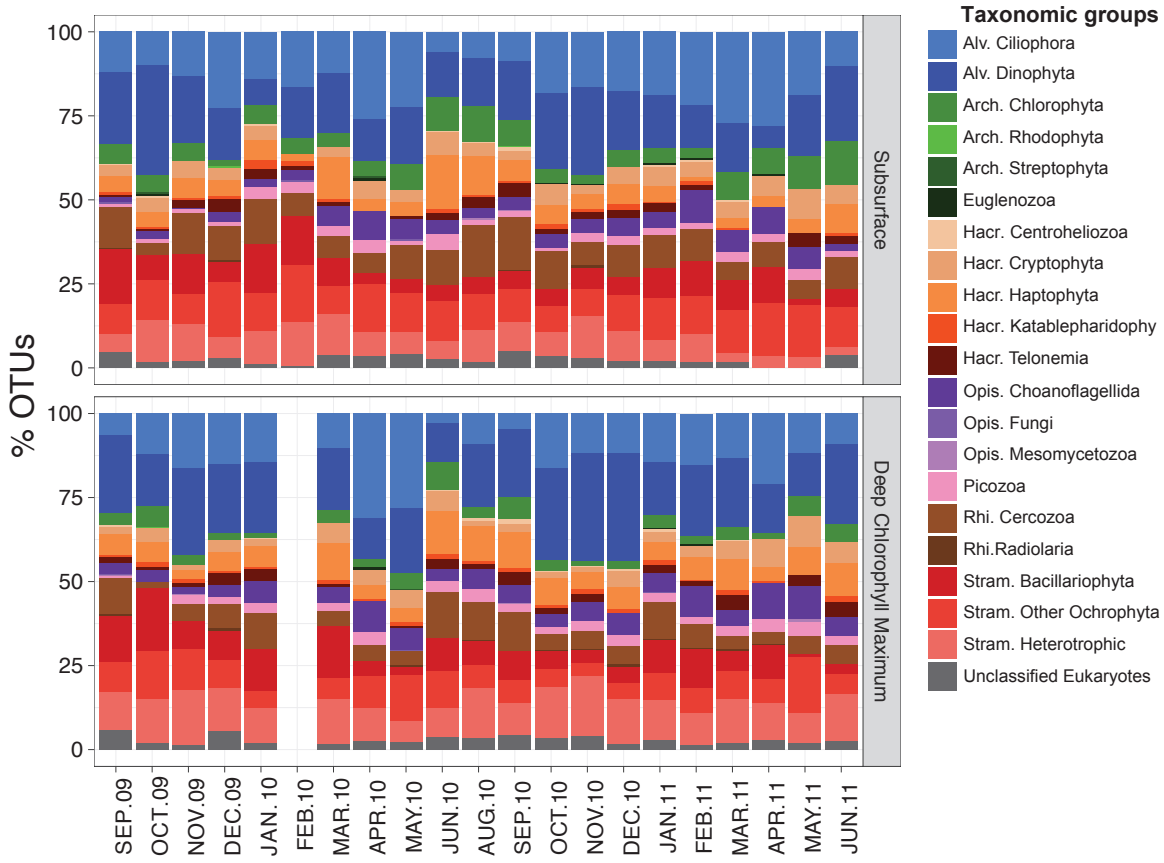


Figure S4. Succession of proportions of OTUs of the 18 major taxonomic groups across the 21 temporal samples at station OF2.

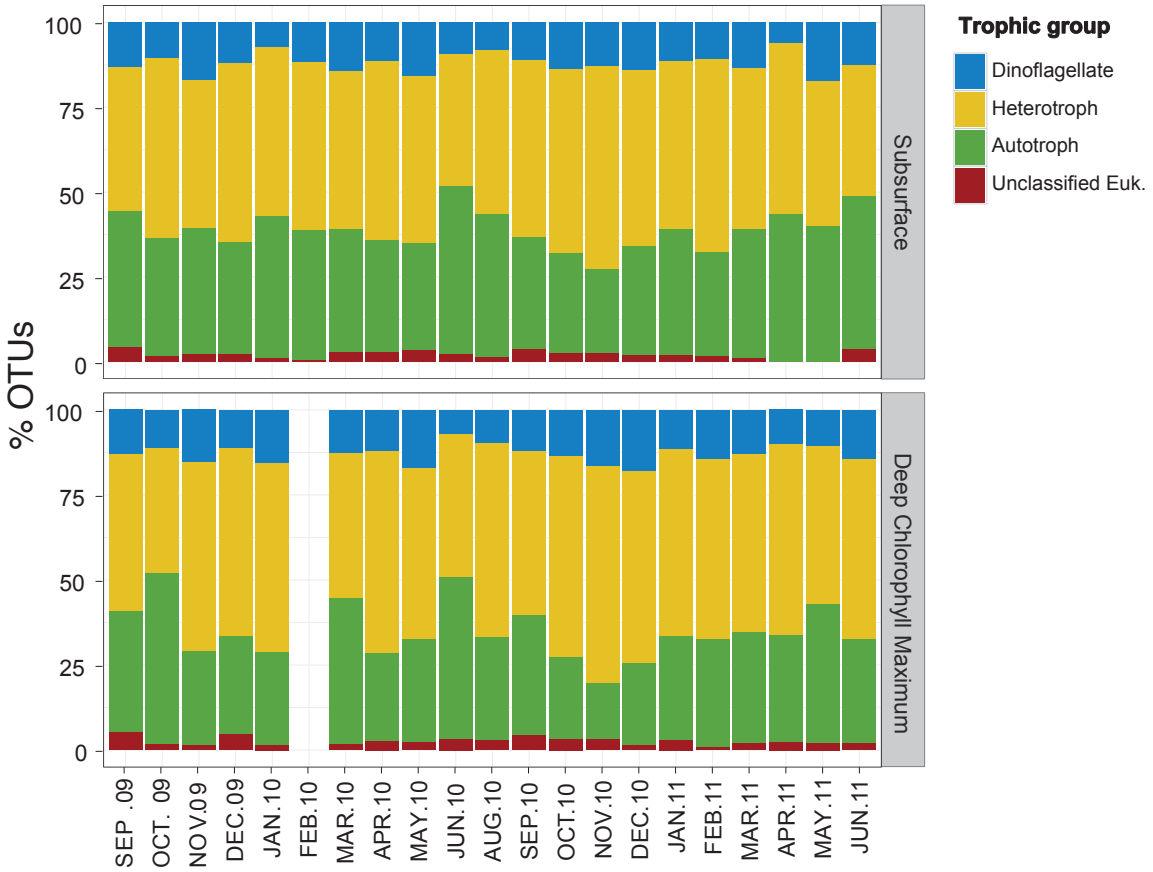


Figure S5. Succession of proportions of OTUs of the different trophic groups across the 21 temporal samples at station OF2.

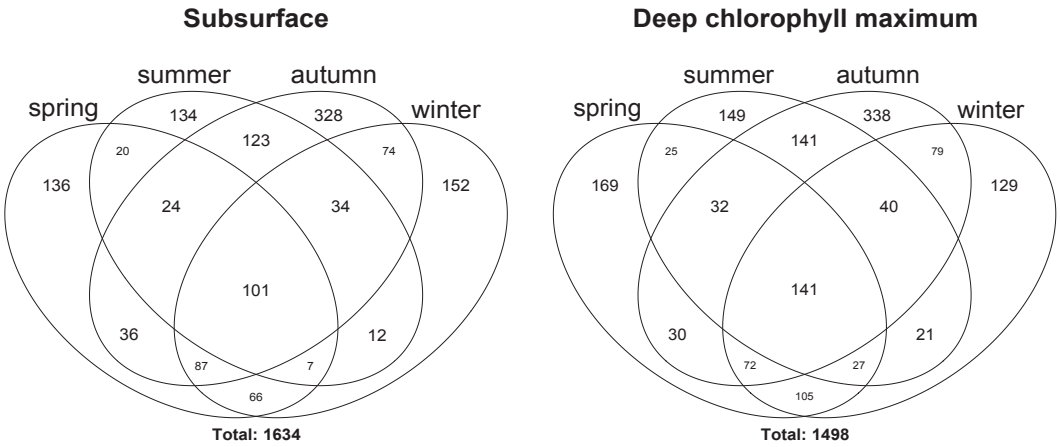


Figure S6. Venn diagram showing the unique and shared operational taxonomic units (OTUs) at OF2 station among the four different seasons during the study period.

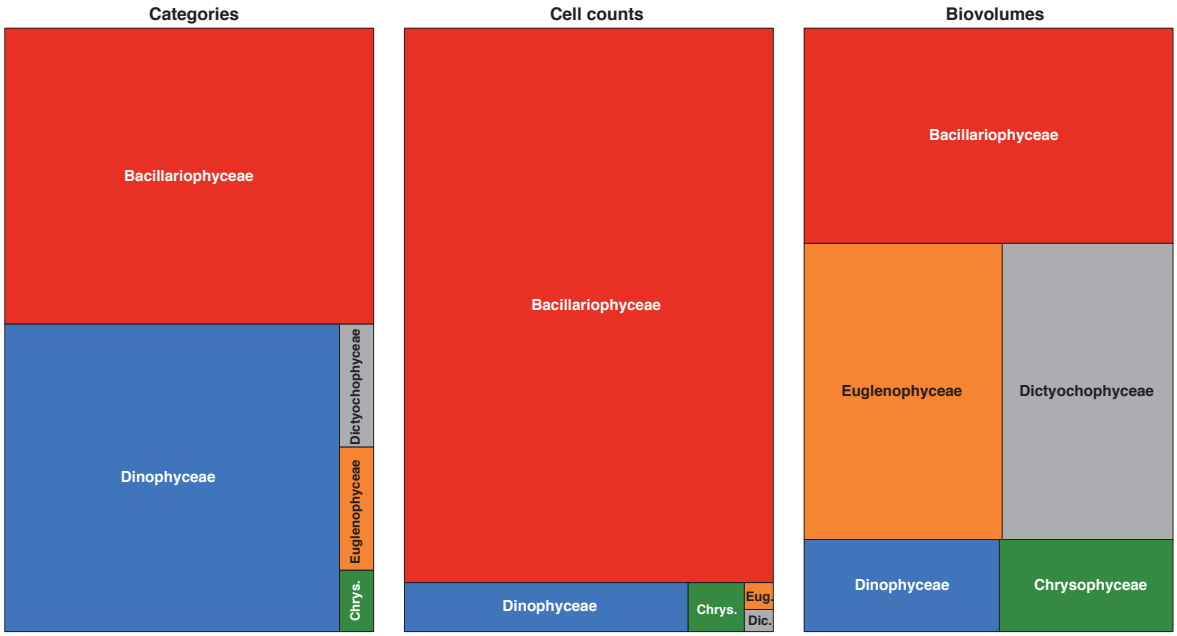


Figure S7. Tree maps displaying the taxonomic composition of the complete Outer Oslofjorden OF2 station protist microscopy dataset at class levels. A. Categories. B. Cell counts. C. Biovolumes.

Table S2. Richness, Shannon and Evenness diversity results with Welch two-tailed test to check correspondence between two studied depths (subsurface and bottom of chlorophyll maximum)

	Richness (N° OTUs)		Shannon (H´)		Evenness (J')	
Sample	**SS**	**DC**	**SS**	**DC**	**SS**	**DC**
SEP.09	293	347	3.91	4.49	0.69	0.77
OCT.09	172	200	3.42	3.64	0.67	0.69
NOV.09	182	209	3.39	3.76	0.65	0.70
DEC.09	218	271	4.25	4.39	0.79	0.78
JAN.10	165	201	3.32	3.89	0.65	0.73
FEB.10	146	NA	2.59	NA	0.52	NA
MAR.10	204	206	3.71	3.61	0.70	0.68
APR.10	131	151	3.07	3	0.63	0.60
MAY.10	166	171	3.39	3.55	0.66	0.69
JUN.10	258	247	4.37	4.03	0.79	0.73
AUG.10	248	284	4.28	4.34	0.78	0.77
SEP.10	280	274	4.65	4.52	0.82	0.81
OCT.10	250	306	3.89	4.51	0.70	0.79
NOV.10	259	263	4.04	4.15	0.73	0.74
DEC.10	199	231	4	3.86	0.75	0.71
JAN.11	192	236	4.02	4.24	0.77	0.78
FEB.11	176	211	3.27	3.45	0.63	0.64
MAR.11	156	187	3.57	3.28	0.71	0.63
APR.11	133	209	3.4	3.83	0.69	0.72
MAY.11	122	144	3.26	3.33	0.68	0.67
JUN.11	127	223	3.48	3.84	0.72	0.71
Welch T-test:						
t	-2.115		-1.396		-0.779	
df	38.986		38.850		38.143	
p-value	0.041		0.171		0.441	

Table S3. PERMANOVA results with permutation of environmental factors for two studied depths (subsurface and bottom of chlorophyll maximum)

	Subsurface						Deep chlorophyll maximum					
	Df	**SS**	**MS**	**F**	**R2**	**P**	**Df**	**SS**	**MS**	**F**	**R2**	**P**
Temperature	1	0.785	0.785	2.684	0.115	0.001*	1	0.641	0.641	2.210	0.109	0.001*
Salinity	1	0.389	0.389	1.330	0.057	0.083°	1	0.287	0.287	0.990	0.049	0.441
Density	1	0.343	0.343	1.171	0.050	0.203	1	0.262	0.262	0.901	0.044	0.647
Nitrates	1	0.353	0.353	1.206	0.052	0.176	1	0.273	0.273	0.941	0.046	0.589
Silicates	1	0.281	0.281	0.960	0.041	0.535	1	0.295	0.295	1.015	0.050	0.465
Phosphates	1	0.281	0.281	0.962	0.041	0.533	1	0.297	0.297	1.023	0.050	0.391
Chlorophyll-a	1	0.585	0.585	2.001	0.086	0.001*	1	0.350	0.349	1.204	0.059	0.192
Residuals	13	3.803	0.293		0.558		12	3.482	0.290		0.592	
Total	20	6.821			1		19	5.886			1	

P-values obtained by 999 permutations. * Significant P-values.

S6 Table. Welch two sample t-test results, to test correspondence between light microscopy cell counts and metabarcoding proportional read abundance

Taxa	mean proportional read Abundances	mean proportional Biovolumes	Difference	95% CI lower	95% CI upper	t	df	p-value
Bacillariophyceae	25.52	53.10	-27.58	-47.12	-8.04	2.87	34.15	0.007
Chrysophyceae	8.20	0.07	8.13	4.10	12.15	4.21	20.02	<0.001
Dictyochophyceae	11.84	1.16	10.68	4.31	17.05	3.48	21.42	0.002
Dinophyta	54.14	42.38	11.76	-7.44	30.96	1.24	37.11	0.222
Euglenophyceae	0.31	3.29	-2.99	-6.76	0.78	-1.66	19.50	0.114

PAPER II

The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Haptophyte Diversity and Vertical Distribution Explored by 18S and 28S Ribosomal RNA Gene Metabarcoding and Scanning Electron Microscopy

Sandra Gran-Stadniczeñko[a,1], Luka Šupraha[b,1], Elianne D. Egge[a] & Bente Edvardsen[a]

a Department of Biosciences, University of Oslo, PO Box 1066 Blindern, Oslo 0316, Norway
b Department of Earth Sciences, Uppsala University, Villavägen 16, Uppsala SE-75236, Sweden

Keywords
Abundance; coccolithophores; high-throughput sequencing; Oslofjorden; phylogeny; richness.

Correspondence
S. Gran-Stadniczeñko, Department of Biosciences, University of Oslo, P. O. Box 1066 Blindern, Oslo 0316, Norway
Telephone number: +47-22-85-73-65;
FAX number: +47-22-85-47-26;
e-mail: s.g.stadniczenko@ibv.uio.no

[1]Equal contribution.

Received: 15 September 2016; revised 6 December 2016; accepted December 7, 2016.
Early View publication January 29, 2017

doi:10.1111/jeu.12388

ABSTRACT

Haptophyta encompasses more than 300 species of mostly marine pico- and nanoplanktonic flagellates. Our aims were to investigate the Oslofjorden haptophyte diversity and vertical distribution by metabarcoding, and to improve the approach to study haptophyte community composition, richness and proportional abundance by comparing two rRNA markers and scanning electron microscopy (SEM). Samples were collected in August 2013 at the Outer Oslofjorden, Norway. Total RNA/cDNA was amplified by haptophyte-specific primers targeting the V4 region of the 18S, and the D1-D2 region of the 28S rRNA. Taxonomy was assigned using curated haptophyte reference databases and phylogenetic analyses. Both marker genes showed Chrysochromulinaceae and Prymnesiaceae to be the families with highest number of Operational Taxonomic Units (OTUs), as well as proportional abundance. The 18S rRNA data set also contained OTUs assigned to eight supported and defined clades consisting of environmental sequences only, possibly representing novel lineages from family to class. We also recorded new species for the area. Comparing coccolithophores by SEM with metabarcoding shows a good correspondence with the 18S rRNA gene proportional abundances. Our results contribute to link morphological and molecular data and 28S to 18S rRNA gene sequences of haptophytes without cultured representatives, and to improve metabarcoding methodology.

THE protist division Haptophyta encompasses more than 300 morphospecies of mostly pico- and nanoplanktonic flagellates (Edvardsen et al. 2016; Jordan et al. 2004; Thomsen et al. 1994). The group inhabits all seas and some also thrive in freshwater, exhibiting a high degree of morphological, physiological and functional diversity (Jordan and Chamberlain 1997). Haptophytes share common structural features, notably the production of unmineralized organic scales and possession of two flagella and a haptonema, although the latter was lost in a few members (e.g. Isochrysidales). Haptophytes are major primary producers in open oceans (Andersen et al. 1996; Liu et al. 2009) and the calcifying coccolithophores play a key role in the biogeochemical carbon cycle (Holligan et al. 1993; Iglesias-Rodríguez et al. 2002; Robertson et al. 1994). In addition, blooms of noncalcifying haptophytes can have a strong impact on coastal ecosystems through toxin production (Granéli et al. 2012; Moestrup 1994).

Being a large and diverse group, haptophytes commonly exhibit species- and even strain-specific physiological traits, which ultimately define their ecological and biogeochemical performance (Edvardsen and Paasche 1998; Langer et al. 2006; Ridgwell et al. 2009). Therefore, identifying haptophytes to a low taxonomic level is of great importance in ecological surveys. Morphological identification to the species level is particularly difficult in noncalcifying groups, which lack hard and mineralized body parts, and can only be accurately identified by the time-consuming examination of organic scales using electron microscopy (EM) (Edvardsen et al. 2011; Eikrem 1996; Eikrem and Edvardsen 1999). On the other hand, scanning electron microscopy (SEM)-based methods allow

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

for precise analysis of taxonomic composition and species abundance of coccolithophore communities (Bollmann et al. 2002; Cros and Fortuño 2002; Young et al. 2003).

With the advance of molecular techniques, notably environmental sequencing of clone-libraries and metabarcoding using high-throughput sequencing (HTS), it has become possible to overcome some of the limitations of microscopical analysis in investigating haptophyte communities (Edvardsen et al. 2016). Most importantly, molecular methods allow for detection of rare or fragile species that commonly remain unnoticed in ecological surveys. A number of studies specifically investigated haptophyte communities using molecular methods (Bittner et al. 2013; Egge et al. 2015a,b; Liu et al. 2009), each of them further improving the methodology and providing new insights about the diversity and distribution of the group. The major shift in understanding the diversity of haptophytes was provided by the early clone-library studies (Moon-van der Staay et al. 2000, 2001; Edvardsen et al. 2016 for comprehensive list of studies), as well as HTS-based works of Bittner et al. (2013) and Egge et al. (2015a,b). These studies identified an abundance of new haptophyte sequences (OTUs) and new haptophyte lineages, which could not be assigned to a cultured and genetically characterized taxon. This indicated that the haptophyte diversity in the modern oceans was largely underestimated in previous microscopic investigations, and that there are many morphospecies still to be described.

However, despite the rigorous processing of data during analysis of 454-reads, a significant portion of ribotypes may still represent chimeric sequences or amplification artefacts (Haas et al. 2011; Huse et al. 2010; Speksnijder et al. 2001). The lack of studies combining qualitative and quantitative morphological analysis of haptophyte communities with molecular approaches complicates the estimation of the actual species diversity and abundance. Conducting such investigations on haptophyte communities is difficult due to the mentioned methodological limitations in identifying noncalcifying species using a morphological approach. Therefore, coccolithophores are arguably the most appropriate group of haptophytes for such comparative investigation. A study by Young et al. (2014) aimed at comparing the molecular (environmental sequencing of clone-libraries) approach based on the 28S rRNA gene with the morphological (LM and SEM) analysis of coccolithophore communities. A good match between the morphological and molecular results was observed in terms of taxonomy, but the study found no strong correlation of the relative OTU and major morphotype abundances. Weak or no correlation between relative OTU abundance and biomass or cell number was also found by Egge et al. (2013) using 454 pyrosequencing in a haptophyte mock community experiment. Since HTS generates much more reads from environmental samples than sequences from clone-libraries, there is a need to compare it with the traditional SEM based approach to assess if this is an appropriate method for studying the diversity and proportional abundance of coccolithophores.

Previous metabarcoding studies of haptophytes (Bittner et al. 2013) and protists (Massana et al. 2015) comparing DNA vs. RNA as templates did not find any significant differences in community structure. However, Egge et al. (2013) found that RNA captured more of the diversity than with DNA where larger cells were favored. Using RNA may significantly reduce the bias due to variability in rDNA copy numbers among taxonomic groups (Not et al. 2009). Massana et al. (2015) found that Haptophyta were underrepresented in DNA compared to RNA surveys (average read ratio DNA/RNA was 7.4). In addition, compared to DNA, RNA is thought to more accurately picture which protists are metabolically active at the time of collection (Stoeck et al. 2007).

Further, comparison of different studies is complicated due to use of different marker genes. Of both 18S and 28S rRNA genes there are more than 600 different reference sequences available in gene databases from cultured material and environmental clone-libraries (Edvardsen et al. 2016). The number of described haptophyte species for which there are reference sequences available are higher for the 18S than 28S rRNA gene (96 vs. 76) (Edvardsen et al. 2016) which makes 18S a more useful marker for identifying species in an environmental sample. However, in haptophytes the 18S rRNA gene may differ in only a few base pairs between closely related species, and short variable regions used for HTS metabarcoding, such as the V4, may be identical (Bittner et al. 2013; Edvardsen et al. 2016). The 28S rRNA gene has more variable regions than 18S, and regions of the 28S such as the D1-D2 have therefore been suggested to be more appropriate barcodes for distinguishing recently diverged species (Bittner et al. 2013; Liu et al. 2009), although some intraspecies variation may occur, which will overestimate species richness (Liu et al. 2009). As these two markers offer complementary views of environmental haptophyte communities, it is important to compare richness and taxonomic composition of 18S and 28S metabarcoding data sets obtained from the same samples. To our knowledge, no study has so far done such a rigorous comparison for haptophytes.

The aim of our study was to investigate the Oslofjorden haptophyte community diversity and vertical distribution by using HTS with two RNA markers supplemented with scanning electron microscopy. We also aim at testing and improving the approach to study haptophyte diversity and proportional abundance. We addressed the following questions (i) Do we find novel taxa or species that have not previously been recorded in the Oslofjorden? (ii) Is there a difference in the community composition and proportional abundance by depth and size fraction? (iii) Do we obtain the same results with the 18 and 28S rRNA marker genes? (iv) Can we place 28S OTUs without a cultured representative in 18S rRNA gene-defined clades? (v) How does the qualitative and quantitative composition of the coccolithophore community compare between HTS and scanning electron microscopy?

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

MATERIALS AND METHODS

Sampling

Water samples from subsurface (1 m) and deep-chlorophyll maximum (DCM, 8 m) for high-throughput sequencing were collected at the OF2 station (59.19 N, 10.69 E) in the outer Oslofjorden, Skagerrak on August 21, 2013. Twenty litres of seawater from each sampling depth were collected with Niskin bottles and prefiltered through a 45-μm nylon mesh. Subsequently, an in-line filtration through 3-μm and 0.8-μm pore size polycarbonate (PC) filters (142-mm Millipore, Darmstads, Germany) was performed with a Millipore Tripod unit and a peristaltic pump for maximum 40 min. This procedure yielded two size fractions: nanoplankton (3–45 μm) and picoplankton (0.8–3 μm). Filters were fast frozen in liquid nitrogen and stored at −80 °C until RNA extraction. Water column profiling was carried out using a conductivity-temperature-depth device (CTD, Falmouth Scientific Inc., Cataumet, MA) and TD-700 fluorometer sensor (Turner Designs, Sunnyvale, CA) attached to a rosette. Nutrients measurements were performed on samples collected at eight different depths (1, 2, 4, 8, 12, 16, 20, and 40 m) as described in Egge et al. (2015b).

RNA extraction, PCR, and 454-pyrosequencing

RNA extraction, PCR, and 454 pyrosequencing of filtered samples were conducted following a modified protocol by Egge et al. (2015a). Briefly, RNA was isolated from 1 m and 8 m filter samples using RNA NucleoSpin II (Macherey-Nagel, Düren, Germany) and converted to cDNA by reverse transcription with the High-Fidelity 1st Strand cDNA Synthesis Kit (Agilent, Santa Clara, CA). For PCR amplification of the ribosomal 18S V4 RNA gene region, we used the forward 528Flong and reverse PRYM01+7 primers described in Egge et al. (2013), whereas for amplification of the 28S D1–D2 region we used the LSU 1 primer pair from Bittner et al. (2013). Amplification protocol for both markers was the same as in Egge 2013, with the exception of the annealing temperature (55 and 53 °C for 18S and 28S respectively). The amplicon library was processed as described by Roche (Basel, Switzerland) and sequenced with the 454 GS-FLX Titanium system at the Norwegian Sequencing Centre (NSC) at the Department of Biosciences, University of Oslo (www.sequencing.uio.no). Two technical replicates were analysed for each depth, size fraction and marker gene (total of 16 samples).

Processing and analyses of raw pyrosequencing data

Sequences were denoised using AmpliconNoise v.1.6.0 in QIIME (Quince et al. 2009). Putative chimeras were identified and removed using Perseus in AmpliconNoise. Further bioinformatic processing was done in Mothur v. 1.36.1 (Schloss et al. 2009), unless otherwise stated. Sequences shorter than 365 bp and with homopolymers > 8 bp were removed using the "trim.seqs"-command. Additional chimera check was done using the "chimera.uchime"-command, with default settings. Sequence clustering was done using the "cluster"-command with the average neighbour algorithm. To be able to compare our 18S rRNA results to previous studies (i.e. Egge et al. 2015a,b) we clustered our 18S OTUs at 99% similarity. Bittner et al. (2013) showed that the more variable 28S rRNA needed clustering at a lower similarity level to balance between species detection and spurious diversity and found 97% to be reasonable, which we used here too. To remove nonhaptophyte reads, the "classify.seqs"-command was used to perform a first OTU taxonomical assignation to phylum against the full Protist Ribosomal Data Base (PR2, Guillou et al. 2013) for the 18S OTUs, and against the SILVA LSU reference database (v. 123) for the 28S rRNA OTUs. The haptophyta OTUs were extracted with the "get.lineage"-command. Sequence similarity and a first taxonomic assignation of haptophyte reads was done by blast against the curated Haptophyta 18S rRNA gene database by Edvardsen et al. (2016) (Haptophyta-PiP) and a curated Haptophyta 28S rRNA gene database from this study, based on Bittner et al. (2013), consisting of 184 unique sequences (see description in Data S1). Both databases with alignments are available at figshare (https://figshare.com/account/home#/projects/11914). Single-tons and double-tons (OTUs with only one or two reads across all samples) were removed before further analysis. To compare our 18S OTUs to the OTUs obtained by HTS of samples from the same station taken in two earlier years (Egge et al. 2015a,b), we ran MegaBLAST (Morgulis et al. 2008) with the OTUs of the previous study as query sequences, and our OTUs as subject sequences. MegaBLAST was run on the University of Oslo Lifeportal (www.lifeportal.uio.no). Detection of an OTU from the previous study is defined as ≥ 99% sequence identity. To be able to compare between samples, they were rarefied (subsampled) to the smallest sample size.

Phylogenetic analyses

Phylogenies were performed following EUKREF RAxML-EPA (Evolutionary Placement Algorithm) pipeline (del Campo, pers. commun.) for a more reliable taxonomic assignation of reads than by BLAST. Curated haptophyte 18S and 28S rRNA gene reference alignments of the reference databases described above were created with MAFFT G-INS-i (http://mafft.cbrc.jp/alignment/server/). Reads were aligned against the reference alignments using "align_seqs.py". Gaps and hypervariable positions were removed using "filter_alignment.py" in QIIME. The alignment was checked manually in Geneious v.7.1.9, and positions that did not align well were edited. All known members of Prymnesiophyceae have a six A homopolymer (position 751–756 in reference sequence AJ004866 *Prymnesium polylepis* 18S rRNA gene) (Egge et al. 2015a), but in our 18S data set this homopolymer varied between 5 and 6 bp. Too short or long

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

homopolymers is a common error with 454 pyrosequencing (e.g. Gilles et al. 2011). To avoid inflated OTU richness, we truncated this A homopolymer to 5 bp in all the sequences. Maximum-likelihood analyses (RAxML v.8.0.26; Stamatakis 2006) was performed on the two reference alignments with substitution model GTR + CAT with 100 bootstraps run on the UiO Abel computer cluster. Finally, the program raxmlHPC-PTHREADS-SSE3 was run to place our 18S and 28S rRNA gene OTUs on our RAxML reference trees, where the alignments and the tree-files in newick format were the input files.

SEM analysis

Water samples were collected at 8 depths (1, 2, 4, 8, 12, 16, 20, and 40 m) using 5-l Niskin water samplers. A known volume from each sample, ranging between 250 (2, 4, 12, 16, 20, and 40 m samples) and 300 ml (1 m sample), was filtered under weak vacuum onto the polycarbonate filter (0.8-μm Cyclopore, 25-mm diameter, Whatman, Kent, UK) that was placed on cellulose nitrate membranes filter (0.8-μm Whatman) to ensure an even distribution of material. After the filtration, filters were dried in oven at 50 °C. Before the analysis under a Zeiss Supra35-VP scanning electron microscope, a piece of filter was mounted on an aluminium stub and sputter-coated with gold. Quantitative analysis of the coccolithophore community was conducted following Bollmann et al. (2002). The same number of fields of view (600 for 1 m and DCM samples, 300 for other samples) was analysed on each filter. Using 600 fields of view covered 6.89 mm^2 of the filter area, corresponding to 4.90 ml of analysed seawater at 1 m depth and 4.15 ml at DCM. Using 300 fields of view covered 3.45 mm^2, analysing 2.07–2.48 ml of seawater. Number of cells counted for each analysed sample ranged between 1,045 and 1,122 using 600 fields of view and 76–570 using 300 fields of view. Taxonomy of coccolithophores was determined to the lowest possible level using the standard taxonomic literature (Cros and Fortuño 2002; Jordan et al. 2004; Young et al. 2003).

Statistical analyses

For each 454-pyrosequencing sample, Shannon–Wiener's species diversity index ($H' = -\Sigma p_i(\ln\ p_i)$) and Pielou's evenness index ($J' = H'/\ln\ S$) were calculated, where p_i is the proportion of individuals of species i and S is the total number of species. Student's one sample t-test was then conducted to determine the significance between pseudo-replicates diversities for the two different marker data sets (Pielou 1966). OTU proportional abundances were normalized by performing a square root transformation, for each of the 16 samples. The Mean-difference (Bland–Altman) method was used on the normalized data to test replicate reproducibility. Differences between replicates were plotted and standard deviation of the differences were computed and added as dotted lines.

RESULTS AND DISCUSSION

Our aims were to investigate the Oslofjorden haptophyte community diversity and vertical distribution and to improve the approach to study haptophyte diversity by metabarcoding. We amplified 18S and 28S rRNA/cDNA genes with haptophyte-specific primers to identify the summer community in the Oslofjorden at two different depths. Most of the haptophyte species that have been morphologically described are found in the size fraction between 2 and 40 μm. In order to cover the entire range, we collected samples between 0.8 and 45 μm divided into two size fractions here called picoplankton (0.8–3 μm) and nanoplankton (3–45 μm). We used maximum-likelihood phylogenetic placement (RAxML-EPA) of OTUs and curated 18S and 28S rRNA gene sequence reference databases to determine their taxonomical assignation. The Oslofjorden haptophyte community has previously been described by high-throughput sequencing (Egge et al. 2015a,b). However, these studies used only 18S rRNA as a marker and only included one depth. Electron microscopy on coccolithophores was also performed on samples taken at eight different depths to assess the semi-quantitative capacity of metabarcoding data.

Hydrography

Figure 1 shows environmental variables at the OF2 sampling site at the day of sampling that may influence haptophyte species composition. At the surface (0–1 m) the

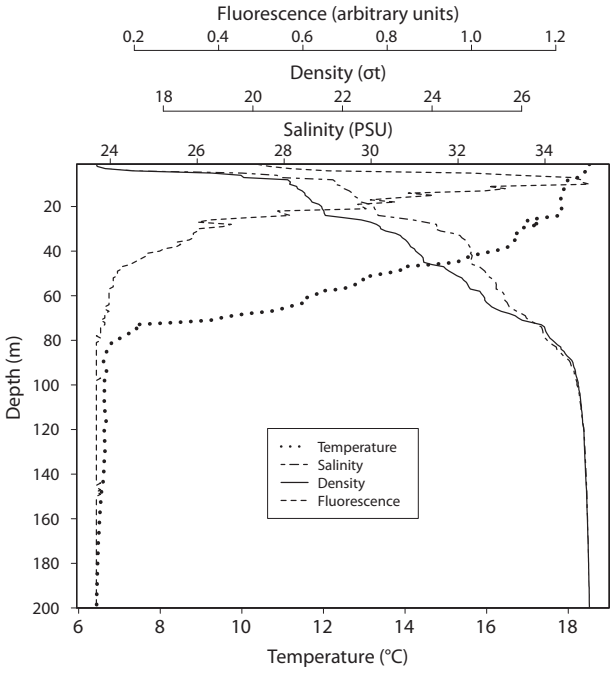


Figure 1 Depth profiles of temperature, salinity, density and fluorescence at outer Oslofjorden, OF2 station on 21 August 2013.

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

salinity was 23.7 PSU and increased by depth down to 90 m where it stabilized at 34.6 PSU. The temperature was 18.5 °C at the surface and gradually decreased down to 80 m stabilizing to 6.8 °C. The density plot indicates a shallow upper mixed-layer in 0–2 m with a pycnocline most pronounced at 2–8 m. The fluorescence (an estimate for relative phytoplankton biomass) increased with depth, reaching maximum values at 8–10 m, (deep-chlorophyll maximum, DCM). Nutrient concentrations by depths are shown in Fig. 2. Concentration of dissolved inorganic nitrogen ($[NO_3^-] + [NO_2^-]$) was near the detection limit (0–0.29 μM) in 1–20 m. Also phosphate (PO_4^{3-}) concentrations were low (0.26–0.29 μM) in the euphotic zone above 20 m. The silicate concentration peaked at 3 m (up to 3.34 μM), probably originated from a fresh water inflow from land.

Haptophyta richness

Denoising of initial reads and removal of putative chimaeras using AmpliconNoise generated a total of 120,282 amplicon reads of the V4 18S rRNA gene and 38,795 of the D1-D2 region of 28S rRNA, with similar numbers of reads among the replicates within each marker (Table S1). Removal of short (< 365 bp) reads and reads with homopolymers > 8 bp from the pooled within-marker data sets yielded 112,958 of reads of 18S rRNA and 30,981 reads of 28S rRNA. Of those, 112,399 18S rRNA reads were assigned to haptophytes (95.5%) compared to 30,892 of the 28S rRNA reads (99.7%). Subsequent removal of chimeric reads discarded 7,454 (6.6%) of the 18S rRNA and 288 (0.9%) of the 28S rRNA haptophyte reads. Finally, after clustering at 99% and 97% similarity and removal of singletons and double-singletons, the 18S rRNA data set contained 215 OTUs (104,345 reads) while the 28S rRNA data set contained 432 OTUs (29,751 reads), respectively.

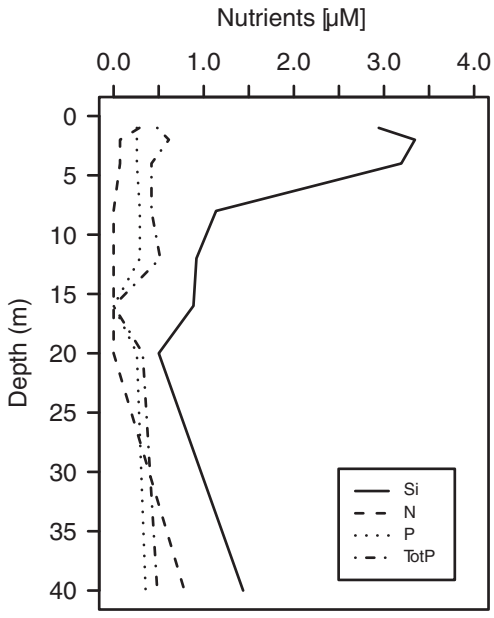


Figure 2 Concentrations (μM) of dissolved inorganic silicate (Si), nitrate + nitrite (N) and phosphate (P), and total phosphorus (dissolved and particulate, inorganic and organic, Tot-P) at outer Oslofjorden OF2 station on the sampling date (21 August 2013).

The haptophyte richness in the Skagerrak area is relatively high. To date, a total of 85 haptophyte species based on microscopy and 156 OTUs obtained by HTS, estimating species, have been recorded from the Skagerrak area (Egge et al. 2015a and references therein). Haptophyta currently comprises 312 morphologically described species (Edvardsen et al. 2016). We recovered considerably higher haptophyte diversity (215 OTUs) in the Oslofjorden than has previously been observed with microscopy in this area. We also recovered a higher 18S OTU richness on one single day than the only previous study applying HTS on monthly samples from 2 yrs (Egge et al. 2015a). In our study, both 1 m and deep-chlorophyll maximum were sampled, whereas in Egge et al. (2015a) only 1 m depth was analysed. In total 104 of the OTUs recovered from the DCM in this study were not detected (i.e. < 99% similar to any OTU) in Egge et al. (2015a), which contributed to a higher richness in this study (Table S2a).

The higher number of 18S OTUs in this study compared to Egge et al. (2015a) may also in part be due to differences in the bioinformatic filtering procedure. The RNA extraction, cDNA synthesis and PCR amplification protocols were identical; however, Egge et al. (2015a) included a very stringent manual OTU filtering step (manual editing of homopolymers and chimera check to GenBank-sequences by BLAST). This stringent filtering is not feasible with a higher number of OTUs, and also not fully reproducible, which is why we did not include it in this study. PCR and HTS techniques such as 454 are well known to introduce sequencing errors and chimeras, and it is near impossible to remove all such errors (e.g. Huse et al. 2010). On the other hand, too stringent filtering may theoretically remove genuine phylotypes.

Taxonomic composition and proportional abundance

Phylogenetic trees inferred from 18S and 28S reference sequences, with our OTUs added, are shown in Fig. 3. The taxonomic assignation is based on these trees. The number of OTUs within each clade are marked to the right of the collapsed and supported clades with bootstrap values > 50%. Some OTUs did not fall within a supported clade and are shown individually in the tree (number of reads per OTU are marked to the right).

18S rRNA

At 99% sequence similarity, we detected OTUs representing 18 supported clades at taxonomic levels from class to family. Within the class Prymnesiophyceae, the family Chrysochromulinaceae hosted the highest number of

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

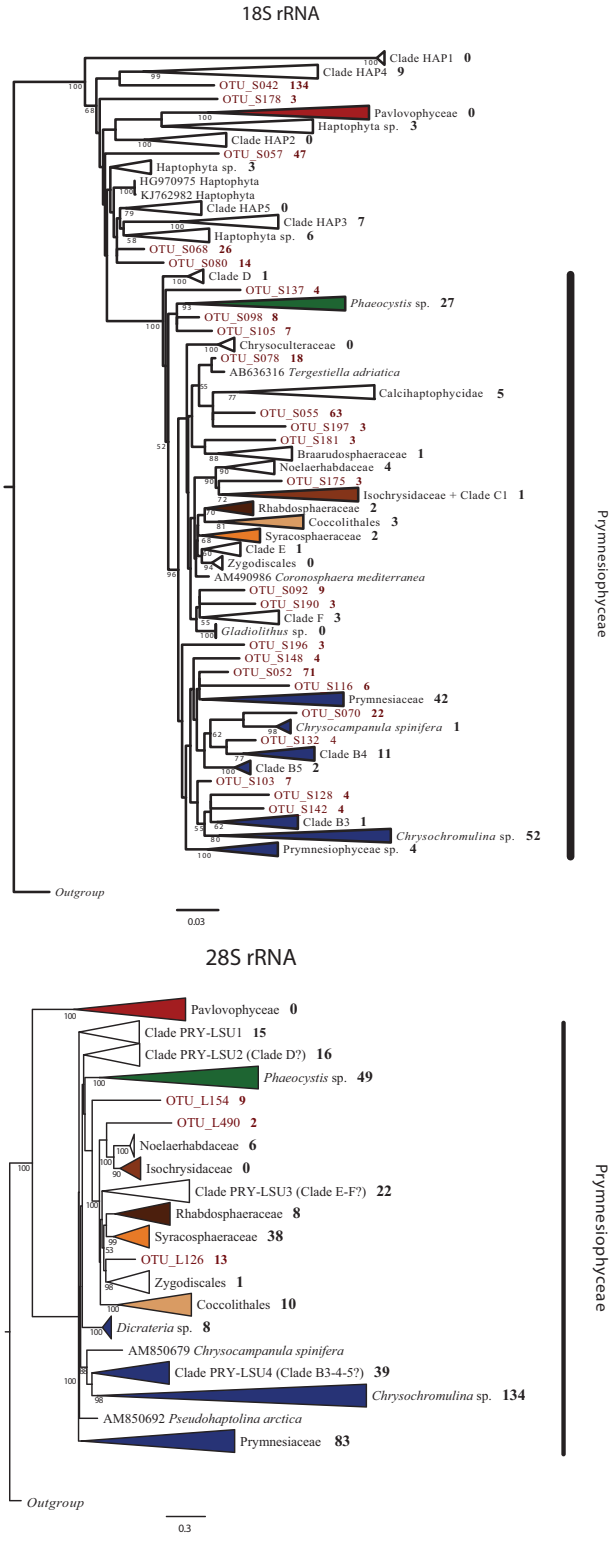


Figure 3 Maximum-likelihood (RAxML) haptophyte phylogeny for (**a**) 18S rRNA and (**b**) 28S rRNA. Support values are inferred using GTRCAT model with 100 bootstraps. Bold numbers indicate the number of OTUs presented in each clade. Numbers next to OTU names indicate the number of reads in each OTU.

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

OTUs (52, 24.2%), followed by Prymnesiaceae (42) and Phaeocystaceae (27) (Fig. 3a and Table S3). Noelaerhabdaceae and Syracosphaeraceae, were represented by only four and two OTUs, respectively. A total of 33 OTUs were placed outside the class Prymnesiophyceae, seven in Clade HAP3, nine in HAP4 and 12 were matching haptophyte environmental sequences without a clade name. The five remaining OTUs did not cluster with any reference sequence.

The proportional read abundance within each major clade is shown in Table S3, and of the 30 most abundant OTUs in Table 1. Chrysochromulinaceae was the family with the highest proportion of reads (33%) followed by Noelaerhabdaceae (18%), Prymnesiaceae (15%) and Syracosphaeraceae (12%). Phaeocystaceae represented 9% of the reads. A total of 3.1% of the reads belonged to OTUs that could not be assigned further than Prymnesiophyceae. OTUs that matched clades HAP3, HAP4, and sequences without phylogenetic placement to any supported clade accounted for 0.9% of the total reads.

The 10 most abundant OTUs were present in all samples (Fig. S1a). The first three 18S OTUs were nested within *Chrysochromulina* sp. (OTU_S001), *Emiliania huxleyi* (OTU_S002), and Syracosphaeraceae sp. (OTU_S003). They constituted 47% of the final reads (Tables 1, S2a). The 126 rarest OTUs (< 10 reads) represented 0.5% of the reads (Fig. S2).

28S rRNA

After clustering at 97% similarity we obtained OTUs within 12 supported clades. Chrysochromulinaceae was the most diverse family with 134 OTUs, followed by Prymnesiaceae (91), Phaeocystaceae (49), and Syracosphaeraceae (38) (Fig. 3b and Table S3). Ten or less OTUs were detected in each of the five remaining clades with cultured representatives. A major part of the OTUs (22%) formed clades without affiliation to any reference sequences from cultures (here called PRY-LSU1, PRY-LSU2, PRY-LSU3, and PRY-LSU4), and only three were not placed in any clade. These may possibly represent the clades without a cultured and sequenced representative as defined by the 18S rRNA gene (see Fig. 3b).

Highest proportion of reads (34%) was found in Chrysochromulinaceae, while Syracosphaeraceae represented the second most abundant family (17%) (Table S3). A similar proportion (17%) was also found for Prymnesiaceae followed by Phaeocystaceae (11.5%). A high proportion (15%) could not be assigned further than Prymnesiophyceae.

We found three 28S OTUs with more than 1,500 reads each, constituting 18% of the total. These clustered within the reference sequences of *Syracosphaera pulchra* (OTU_L001 and OTU_L002) and *Phaeocystis* sp. (OTU_L003) (Tables 2, S2b and Fig. S2). The 12 next most abundant OTUs contained between ~500 and 1,000 sequences. Thirty percentage of the 280 rarest OTUs (< 10 reads) belonged to the genus *Chrysochromulina* (Table S2b).

In samples from the Oslofjorden collected in August–September 2009 and 2010, Egge et al. (2015b) found that Prymnesiaceae and Chrysochromulinaceae were the most OTU-rich groups, followed by Phaeocystaceae and Calcihaptophycidae (de Vargas et al. 2007). The novel lineages HAP3 and HAP4 were also represented with 3–4 OTUs in the 2009 and 2010 late summer samples. Chrysochromulinaceae, Prymnesiaceae, *E. huxleyi*, Syracosphaerace, and Phaeocystaceae were proportionally the most abundant groups in these samples. Thus, both in terms of taxonomic distribution of 18S rRNA OTUs and proportional abundance of the different groups, our results are consistent with Egge et al. (2015b).

Previous microscopy surveys have also reported Prymnesiaceae and Chrysochromulinaceae to be very species rich in the Skagerrak and Kattegat. Jensen (1998) recorded c. 30 morphological species of *Chrysochromulina* (sensu lato), and scales of 20 undescribed forms, which morphologically resembled this group. Members of Chrysochromulinaceae and Prymnesiaceae have been reported as the most abundant noncalcifying haptophyte groups in June–September (Dahl and Johannessen 1998; Kuylenstierna and Karlson 1994; Lekve et al. 2006).

Novel taxa or records for the area

We performed taxonomical assignation of OTUs by blast against the Haptophyta-PiP database. Detection of a cultured species or an environmental sequence is here defined as ≥ 99% or ≥ 97% sequence identity to one of the 18S or 28S rRNA OTUs, respectively. Of the 215 18S OTUs, only 20 had ≥ 99% match to a cultured species, whereas 47 had ≥ 99% match to an environmental sequence present in the Haptophyta-PiP database (Table S4). Thirty-six (16%) did not nest within any specific haptophyte clade in the phylogenetic tree (Fig. 3). Of the 432 28S OTUs, 47 had ≥ 97% match to a cultured species, (Table S4a, b). Comparing our 18S OTUs to Egge et al. (2015a,b) we found that 68 of our OTUs were ≥ 99% identical to any OTU recovered in Egge et al. 2015a (Tables S2a, S4a). The majority of these (62 OTUs) were present with ≥ 10 reads. Out of these 68 OTUs, 26 were ≥ 99% identical to OTUs from Egge et al. 2015a and at the same time < 99% identical to any sequence in the Haptophyta-PiP database. This suggests that these 26 OTUs have only been detected in the Oslofjorden. For instance, OTU_S072, whose closest match in the Haptophyta-PiP database was *Tergestiella adriatica* (95.8%), was 99.7% identical to an OTU nesting within a clade of environmental sequences classified as Calcihaptophycidae in Egge et al. (2015a) (cf. OTU113, fig. 4 in Egge et al. 2015a). Thus, it may represent a coccolithophore species that has not yet been sequenced. Of the most OTU-rich groups, the group with the highest proportion of OTUs matching environmental sequences was Phaeocystaceae and Chrysochromulinaceae (Table S4a). Within both of these groups, several morphological forms have been recognized that are not yet in culture and genetically characterized (Jensen 1998; Medlin and Zingone 2007). Our

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 1. List of the 30 most abundant haptophyte V4 18S rRNA OTUs detected

OTU ID	Total reads (*N*)	Total reads (%)	Total reads after subsampling (*N*)	Total reads after subsampling (%)	Depth	Size fraction	Group	Lowest taxonomic level possible to determine
OTU_S001	19,473	18.66	13,715	18.74	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S002	18,849	18.06	11,413	15.59	Both	Both	Noelaerhabdaceae	*Emiliania huxleyi*
OTU_S003	10,500	10.06	6,669	9.11	Both	Both	Syracosphaeraceae	Syracosphaeraceae
OTU_S004	5,186	4.97	3,893	5.32	Both	Both	Prymnesiaceae	*Prymnesium*
OTU_S005	4,234	4.06	3,364	4.60	Both	Both	Prymnesiaceae	*Haptolina*
OTU_S006	4,231	4.05	3,252	4.44	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU_S007	3,349	3.21	2,847	3.89	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S008	3,126	3.00	2,293	3.13	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S009	2,784	2.67	1,976	2.70	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S010	2,757	2.64	2,255	3.08	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S011	2,650	2.54	1,963	2.68	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU_S012	2,592	2.48	1,928	2.63	Both	Both	Prymnesiophyceae	Prymnesiophyceae
OTU_S013	2,334	2.24	1,614	2.21	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU_S014	2,148	2.06	1,258	1.72	Only DCM	Only nano	Syracosphaeraceae	Syracosphaeraceae
OTU_S015	2,132	2.04	1,690	2.31	Both	Both	Prymnesiales Clade B3–B4–B5	Prymnesiales Clade B4
OTU_S016	1,679	1.61	1,235	1.69	Both	Both	Prymnesiaceae	*Prymnesium polylepis*
OTU_S017	1,378	1.32	810	1.11	Both	Both	Calyptrosphaeraceae	*Calyptrosphaera sphaeroidea*
OTU_S018	1,284	1.23	927	1.27	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU_S019	1,037	0.99	775	1.06	Both	Both	Prymnesiales Clade B3–B4–B5	Prymnesiales Clade B4
OTU_S020	871	0.83	632	0.86	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S021	866	0.83	640	0.87	Both	Both	Phaeocystaceae	*Phaeocystis cordata*
OTU_S022	797	0.76	520	0.71	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU_S023	778	0.75	559	0.76	Both	Both	Prymnesiales Clade B3–B4–B5	Prymnesiales Clade B4
OTU_S024	765	0.73	557	0.76	Both	Both	Prymnesiales Clade B3–B4–B5	Prymnesiales Clade B4
OTU_S025	722	0.69	579	0.79	Both	Both	Prymnesiales Clade B3–B4–B5	Prymnesiales Clade B4
OTU_S026	633	0.61	577	0.79	Only 1 m	Both	Calcihaptophycidae	Calcihaptophycidae
OTU_S027	479	0.46	276	0.38	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU_S028	449	0.43	298	0.41	Both	Both	Prymnesiophyceae	Prymnesiophyceae
OTU_S029	443	0.42	287	0.39	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_S030	364	0.35	276	0.38	Both	Both	Phaeocystaceae	*Phaeocystis globosa*

The taxonomic assignment was based on phylogenetic placement.

result, that several OTUs had best match with DNA sequences or reads from environmental samples, supports previous studies showing that there is a large diversity of haptophytes that remains to be cultured and DNA sequence determined (e.g. Bittner et al. 2013; Liu et al. 2009; de Vargas et al. 2015).

Sample comparisons

After subsampling to the lowest number of reads (9,148 and 1,577 for 18S and 28S rRNA genes), four and 42 OTUs were removed, respectively. In the Venn diagram (Fig. 4) unique and shared OTUs of the 18S and 28S samples are presented separately. Only ~ 13% of the OTUs were present in all samples (28 for 18S and 53 for 28S). A considerable number of OTUs were found in only one sample (60% for 18S and 52% for 28S rRNA gene).

Reproducibility of PCR and 454 sequencing

High-throughput sequencing studies have often been criticized for lack of replication, as experimental errors can arise during sample and library preparation or sequencing and filtering (Robasky et al. 2014). In our study, we tested technical replicates (TR), dividing each filter in two and performing extraction, PCR, and pyrosequencing separately. The TR were highly similar for all pairs of samples ($R^2 > 0.9$, $p < 0.001$) presenting similar proportional abundances (Fig. 5b, S1). The Bland–Altman plot (Fig. S3)

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 2. List of the 30 most abundant haptophyte D1–D2 28S rRNA OTUs detected

OTU ID	Total reads (*N*)	Total reads (%)	Total reads after subsampling (*N*)	Total reads after subsampling (%)	Depth	Size fraction	Group	Lowest taxonomic level possible to determine
OTU_L001	1,980	6.66	740	5.87	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU_L002	1,830	6.15	685	5.43	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU_L003	1,555	5.23	679	5.38	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU_L004	957	3.22	494	3.92	Both	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU_L005	862	2.90	365	2.89	Both	Both	Prymnesiaceae	*Haptolina ericina/hirta/fragaria*
OTU_L006	782	2.63	339	2.69	Both	Both	Prymnesiaceae	*Haptolina*
OTU_L007	710	2.39	318	2.52	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_L008	676	2.27	299	2.37	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii/ C. campanulifera*
OTU_L009	659	2.22	319	2.53	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade-E–F?)
OTU_L010	648	2.18	269	2.13	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU_L011	645	2.17	224	1.78	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU_L012	597	2.01	232	1.84	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade-D?)
OTU_L013	576	1.94	257	2.04	Both	Both	Prymnesiaceae	*Prymnesium polylepis*
OTU_L014	560	1.88	257	2.04	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_L015	538	1.81	234	1.85	Both	Both	Prymnesiaceae	*Prymnesium kappa*
OTU_L016	516	1.73	234	1.85	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU_L017	491	1.65	198	1.57	Both	Both	Prymnesiaceae	*Prymnesium*
OTU_L018	468	1.57	160	1.27	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU_L019	467	1.57	188	1.49	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU_L020	426	1.43	180	1.43	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade-D?)
OTU_L021	412	1.38	186	1.47	Both	Both	Prymnesiaceae	*Haptolina*
OTU_L022	406	1.36	174	1.38	Both	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU_L023	367	1.23	137	1.09	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU_L024	355	1.19	178	1.41	Both	Both	Coccolithales	Coccolithaceae
OTU_L025	352	1.18	155	1.23	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU_L026	351	1.18	130	1.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU_L027	339	1.14	140	1.11	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU_L028	327	1.10	157	1.24	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU_L029	305	1.03	117	0.93	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU_L030	291	0.98	133	1.05	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*

The taxonomic assignment was based on phylogenetic placement.

performed for normalized OTU abundances clearly showed that larger disagreements between TR were found among the rare OTUs. With increasing abundances, the plots get consistently linear indicating high similarity. The TR had similar taxonomic composition and proportional abundance for the different major taxonomic groups (Fig. 5a, b).

We suggest that TR do not bring considerable extra information to studies focusing on the most abundant groups. However, many of the rare OTUs appeared in only one TR (Fig. S1), indicating the importance of TR in recovering low abundant OTUs. This agrees with findings by Massana et al. (2015).

Comparing markers

There are no previous studies comparing 18S and 28S rRNA gene as markers in metabarcoding studies of haptophytes. Here, we wanted to examine if the 28S marker gives the same resolution or higher than 18S. Compared to 18S rRNA, we found a higher haptophyte diversity using 28S haptophyte-specific primers (Fig. 6). However, it is not understood if this diversity represents intra- or interspecific variation. The number of OTUs detected was higher in the 28S than in the 18S samples (Table S1). In the 18S data set, members of Chrysochromulinaceae dominated in all samples followed by Prymnesiaceae, both in number of OTUs and read proportion. The next most abundant families in the 18S data set were Noelaerhabdaceae, Syracosphaeraceae, and Phaeocystaceae, whereas in the 28S data set members of Noelaerhabdaceae were almost missing (0.5%, Fig. 5b). This is likely due to mismatches between the 28S sequence of members of Isochrysidales and the LSU1 forward primer, one of which occurs at the 3′ end of the primer, which may prevent elongation. OTUs belonging to the families Isochrysidaceae and Braarudosphaeraceae were only detected in the 18S samples. In contrast, Helicosphaeraceae OTUs were only observed in the 28S data set. Respectively 4% and 15.5% of 18S and 28S OTUs could not be assigned to any known haptophyte family and was

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

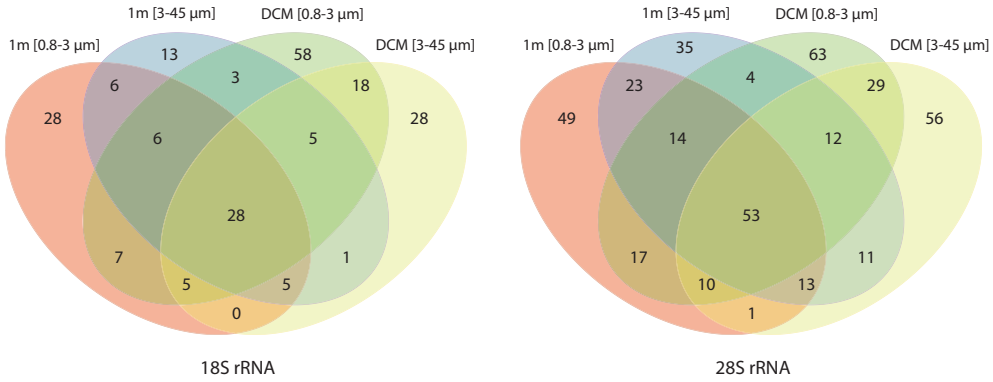


Figure 4 Four-way Venn diagram illustrating the number of unique and shared haptophyte 18S and 28S OTUs for the four set of samples studied at OF2 station.

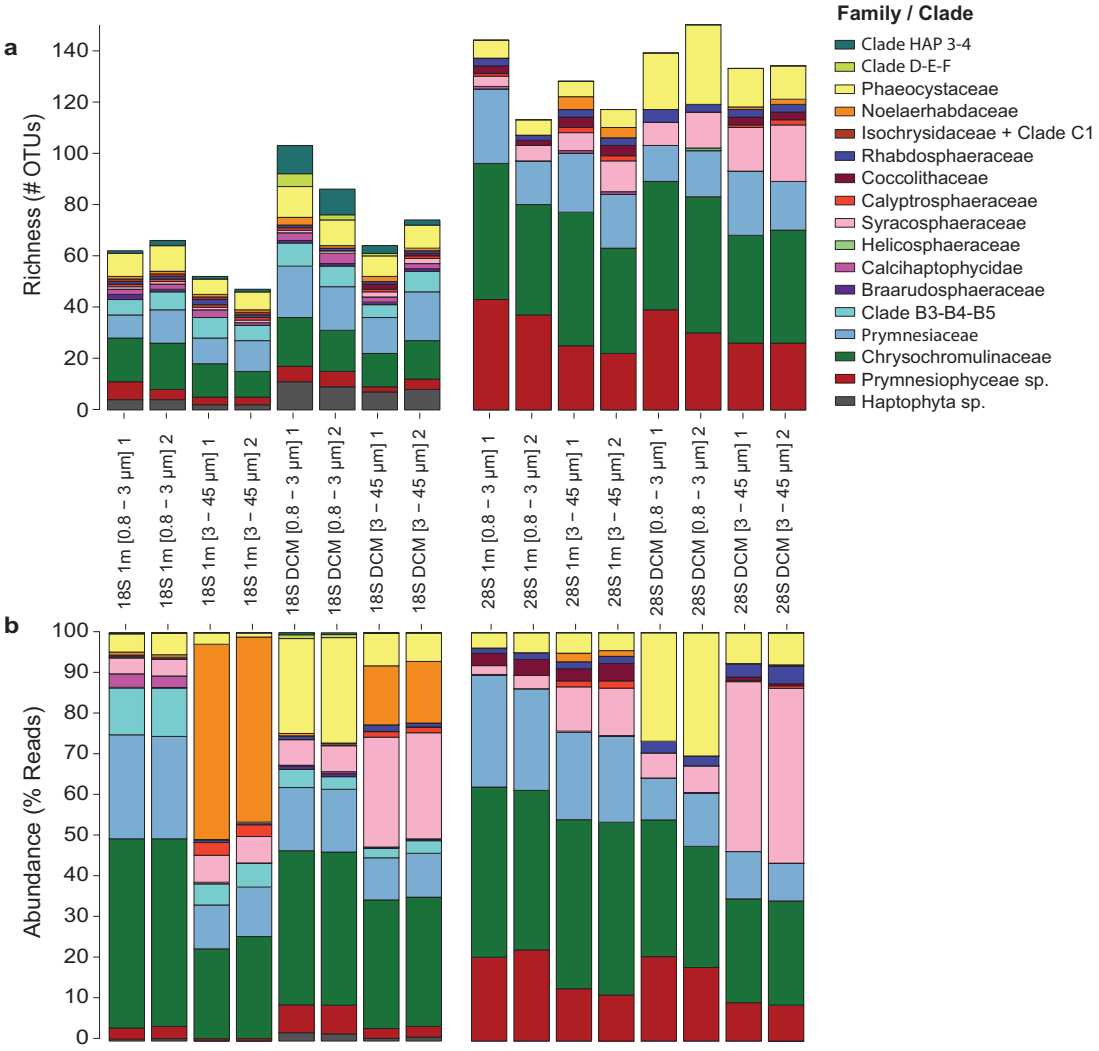


Figure 5 Haptophyte taxonomic distribution for all 18S rRNA (left) and 28S rRNA (right) samples and replicates 1 and 2. (**a**) OTU richness. (**b**) Proportional read abundances.

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

therefore assigned to either Prymnesiophyceae sp. or Haptophyta sp. The 18S marker provided a more accurate and extensive assessment of species identity than the 28S marker, due to the lower number of 28S reference sequences. In addition, for the 18S reference data set there are a number of defined clades (from class- to genus-level) without cultured representatives. We asked whether we could place 28S OTUs without a cultured representative in 18S-defined clades. We compared our 18S and 28S phylogenetic trees (Fig. 3) and identified three 28S clades (PRY-LSU2, -LSU3, and –LSU4) that may represent 18S defined clades without culture representatives (Clade-D, Clade-E + F, Clade B3–B5, respectively). Although differences in both number of OTUs and sequences abundances are found between the markers, both rRNA data sets revealed that the majority of haptophyte species can be assigned to a defined clade with either the 18S or the 28S rRNA gene.

Comparing depths and size fractions

Haptophyta communities may have different species composition according to depth: in the Mediterranean Sea (Bittner et al. 2013), in the South Pacific (Shi et al. 2009) and in the Red Sea (Man-Aharonovich et al. 2010). We aimed to examine if these differences also occur on the Norwegian coast. The highest number of OTUs in our data sets was found at the DCM in picoplankton (0.8–3 μm) samples for both markers, and pooled pico- and nanoplankton 1 m samples contained fewer OTUs than at DCM, corresponding with previous findings (Bittner et al. 2013).

Sixty-three 18S OTUs were present at both depths (Table S2a), whereas 46 were only found at 1 m and 102 only in the DCM. OTUs that clustered with *Isochrysis* sp. and *Tergestiella adriatica* were only found at 1 m (Table S2a). Contrarily, those belonging to Clades Prymnesiales B3, D, E, F, and HAP4, and *Coccolithus* sp. were only present in the DCM sample within 18S rRNA marker (Table S2a). The highest number of OTUs was found in DCM samples for both markers. The proportion of OTUs of each taxonomic group was similar among all samples (Fig. 5a). There is, however, a clear difference in proportional abundances by depth (Fig. 5b). As described in previous studies (Malinverno et al. 2003), differences in communities by depth can be explained by temperature, phosphorus and light availability. On our sampling date, temperature and phosphate at 1 m and DCM were equal and thus light availability and differences in salinity (23.7 PSU for 1 m and 29.1 PSU for DCM) could explain the differences found in community structure. Comparing size fractions, we observed that the picoplankton had higher number of OTUs than the nanoplankton, corresponding with previous findings by Bittner et al. (2013). OTUs nested within Clades B3, E, and F were exclusive in the pico-fraction and OTUs from *Chrysochromulina rotalis*, *Tergestiella adriatica*, and *Coccolithus* sp. were unique in the nano-fraction.

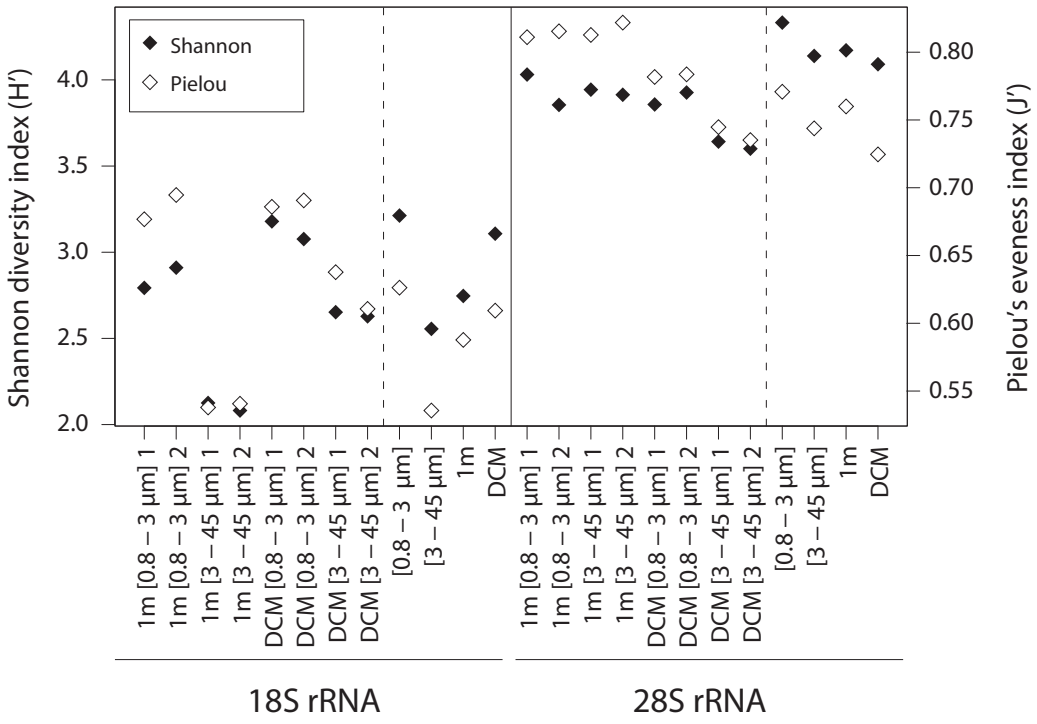


Figure 6 Shannon diversity index and Pielou's evenness index for all 18S and 28S samples studied at the OF2 station. The continued line splits the two different markers' data sets, whereas the dotted lines split the single samples from the pooled ones in size and depth.

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

The 28S data set contained 171 OTUs present at both depths (Table S2b). Of the remaining OTUs, 88 were only found at 1 m and 131 at the DCM. OTUs best matching *Chrysotila stipitata* and *Umbilicosphaera* sp. were uniquely found at 1 m, and *Coccolithus braarudii*, *Coronosphaera mediterranea*, whereas *Helicosphaera* sp. were unique for the DCM samples. As for the 18S data, the 28S contained larger number of OTUs in the picoplankton size fraction compared to the nanoplankton. The unique OTU for the picoplankton-fraction was clustered with *Helicosphaera* sp. OTUs assigned to *C. stipitata, C. braarudii, C. mediterranea, E. huxleyi*, and *Umbilicosphaera* sp. were only found in the nanoplankton samples. Coccolithophores were found in both size fractions, although almost all known species are larger than 3 μm. An explanation can be that during filtration cells break and ribosomes pass through the filter pores. However, some haploid stages of coccolithophores (i.e. *E. huxleyi*) and some members of Syracosphaeraceae are known to be < 3 μm in size.

Diversity and evenness

Tables with proportional abundances per sample and OTU (not shown, see Fig. S1) were used to calculate the Shannon's diversity and Pilou's evenness indexes. Shannon index (18S *t*-test: $t = 0.29$, df = 3, p-value = 0.79, and 28S *t*-test: $t = 0.88$, df = 3, p-value = 0.44) and Pilou's index (18S *t*-test: $t = 0.05$, df = 3, p-value = 0.97, and 28S *t*-test: $t = -0.37$, df = 3, p-value = 0.74) were similar among the technical replicates (Fig. 6). These results suggest that the metabarcoding methodology used in this study is adequate for obtaining robust beta-diversity and taxonomic descriptions. Similar findings were obtained by Massana et al. (2015). When looking at single samples in the 28S data sets, the nano-DCM samples appear less diverse than the rest. However, for 18S the 1 m nanoplankton seems to harbour the lowest diversity. When pooling the samples from the same depth, we see the same pattern, that the pico-fraction had both higher OTU richness and evenness than the nano-fraction, resulting in higher diversity in the picoplankton. However, regarding depth, the 18S DCM sample was more diverse than the 1 m one, whereas the opposite was found for the 28S samples. In the 18S data set, we observed few Noelaerhabdaceae OTUs (*E. huxleyi* as the most abundant) representing a high proportion of the total reads (Fig. 5). The Noelaerhabdaceae OTUs occurred in very low read proportions in the 28S data, due to a mismatch in the primers (discussed above). This difference is probably the reason why we observe an opposite diversity pattern by depth for the two markers.

Comparison of SEM and metabarcoding for analyses of coccolithophore communities

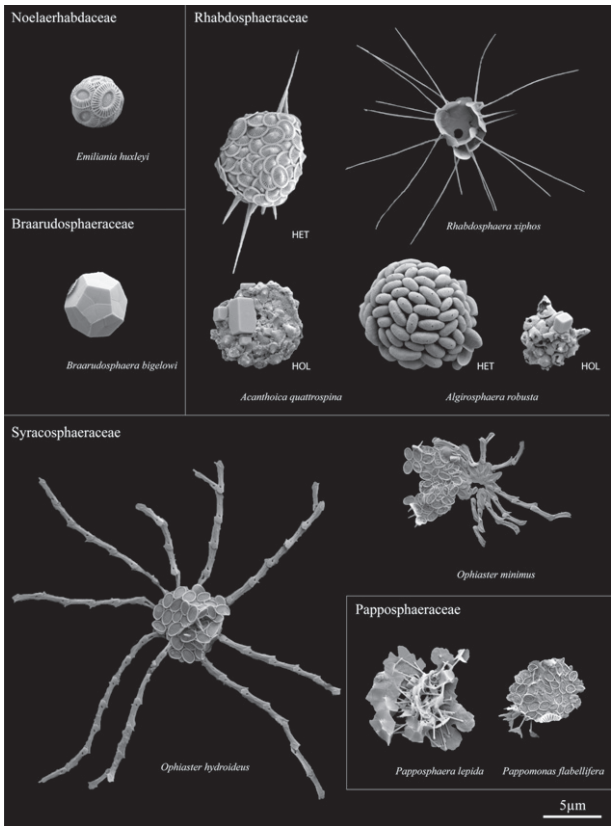
Species diversity estimation

Taxonomic analysis of coccolithophore community at OF2 using SEM detected 26 distinct coccolithophore morphotypes (from eight depths). When corrected for life-cycle phases and combination coccospheres (intermediate life-cycle forms), the complete taxonomic list numbered 22 coccolithophore species, two of which could not be precisely identified (Fig. 7 and Table 3). Our observations expand the checklist of species detected in the Scandinavian waters by 6, and the number of morphotypes by 12 (Egge et al. 2015a; Eikrem 1999). The total number of species detected at the two depths used for the method comparison (subsurface (1 m) and DCM (8 m)) was 14, with an increase in species number observed from 1 m (8 species) to DCM (14 species). The metabarcoding at these two depths generated a total of 29 coccolithophore OTUs based on the 18S rRNA, and 89 OTUs based on 28S rRNA. The increase in coccolithophore OTU diversity by depth was also detected with metabarcoding. The number of 18S rRNA OTUs increased from 13 to 23 and the number of 28S rRNA OTUs from 51 to 69 from 1 m to DCM. In the SEM analysis, Syracosphaeraceae was the most species-rich family at the two depths with eight species, followed by various holococcolithophore taxa (2 species) and Rhabdosphaeraceae (2 species). On the other hand, the 18S rRNA OTUs were mostly assigned to Noelaerhabdaceae (4), Rhabdosphaeraceae (2), and Syracosphaeraceae (3). The 28S rRNA OTUs showed highest richness of Syracosphaeraceae (38), Rhabdosphaeraceae (8) and Calyptrosphaeraceae (6) (Table 4).

The number of 18S rRNA OTUs obtained was similar to the number of observed species, but failed to account for the high level of richness within the Syracosphaeraceae and Rhabdosphaeraceae families. Also, the number of coccolithophore OTUs in our study corresponded to the number of OTUs assigned to the subclass Calcihaptophycidae (29) in the study by (Egge et al. 2015a) at the same location. On the other hand, the 28S rRNA marker seemed to overestimate the coccolithophore species richness, generating higher number of coccolithophore OTUs compared to the number of species identified in the SEM analysis. Similar results showing an overestimation of richness using the 28S rRNA gene as marker were obtained in the study of Young et al. (2014) using environmental sequencing of clone-libraries. This could be due to faster evolutionary rates in the D1-D2 region of the 28S than the V4 region of the 18S rRNA gene, giving a higher resolution for species-level identification, but also overestimating diversity (Liu et al. 2009). More 28S rRNA reference sequences are needed to obtain a better link between OTU and species based on this gene.

Due to the low number of reference sequences obtained from cultures of Syracosphaerales and cf. Syracospaherales *incertae sedis* (Edvardsen et al. 2016), only few direct matches between SEM and metabarcoding data were obtained at the species level. The most abundant species in our samples, *E. huxleyi*, was represented by the type A morphotype in SEM counts. The most abundant OTU assigned to *E. huxleyi* (OTU_S002) by phylogeny was 99.7% similar to sequences of cultures of this species (e.g. EU106795). One base differed, as the OTU_S002 had 5. As instead of 6 in a homopolymere region. Using 28S rRNA generated six OTUs matching the

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists



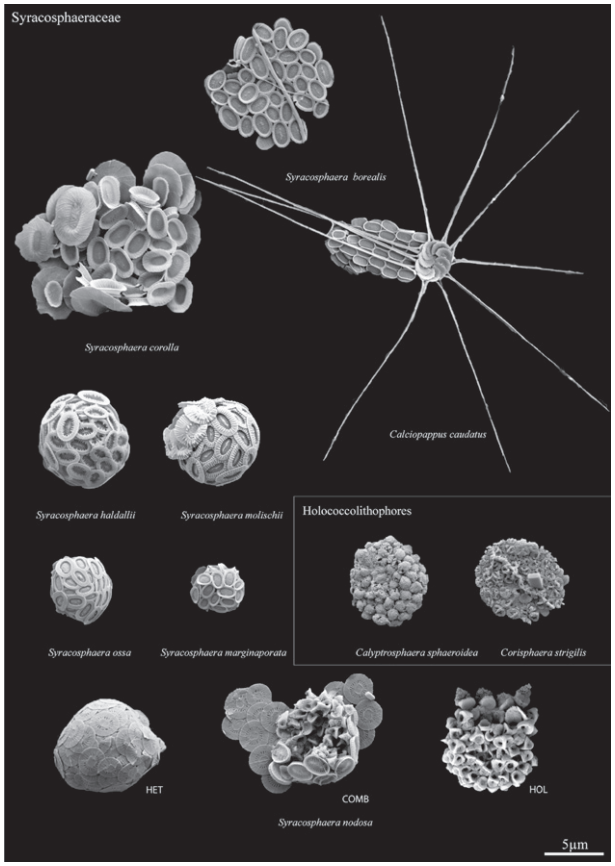


Figure 7 Scanning electron micrographs of coccolithophore morphotypes detected in this study. Abbreviations used for species detected in more than one life-cycle phase: HET, heterococcolith phase; COMB, combination coccosphere with both heterococcoliths and holococcoliths representing a transition phase; HOL, holococcolith phase.

cultured *E. huxleyi* reference. A good correspondence between SEM and metabarcoding was observed in *Braarudosphaeara bigelowii*, for which the OTU_S033 clustered with a clade including reference sequences from morphologically verified picked cells of *B. bigelowii*.

The Syracosphaeraceae family is represented with only two cultured species in the Haptophyta-PiP database (*S. pulchra* and *C. mediterranea*), neither of which were observed in the SEM analysis. Only two 18S rRNA OTUs clustered within Syracosphaeraceae, and none was placed close to the two species (Fig. 3a). However, 38 28S rRNA OTUs were assigned to this family, exhibiting different degrees of similarity with the two cultured references. The OTU_L098 was placed close to *C. mediterranea* 28S culture reference indicating that *C. mediterranea* was present in our samples. The OTUs clustering with the *S. pulchra* reference sequence likely represent different *Syracosphaera* species or members of Syracosphaeraceae. Similarly, a number of OTUs were placed within Rhabdosphaeraceae, represented by one cultured species, *Algirosphaera robusta* in both 18S and 28S rRNA gene reference databases. Both markers provided OTUs highly similar to this species' reference sequence, confirming the finding by SEM.

The Haptophyta-PiP database contained only two reference sequences for holococcolithophore taxa (both markers for *Calyptrosphaera sphaeroidea* and 28S rRNA for *Helladosphaera* sp.) obtained from cultured material. We detected OTUs assigned to *C. sphaeroidea* with both markers (OTU_S017, 18S and OTU_L053, L233, L350, 28S) and also detected this species by SEM. However, no OTUs clustered with *Helladosphaera* sp. Finally, a number of OTUs detected in our study matched the sequences of coccolithophore species that were not observed in the SEM survey. Those included *C. braarudi* (OTU_L137), *Coccolithus pelagicus/braarudi* (OTU_S075, 18S, which is identical for the two species), *Helicosphaera carteri/wallichii* (OTU_L377, 28S) and *Tergestiella adriatica* (OTU_S078, 18S). In addition, molecular analysis confirmed the presence of the members of Calcidiscaceae (OTU_L224) and Pleurochrysidaceae (OTU_L222), none of which were observed in the SEM analysis. Finally, a number of OTUs (pooled in category "Other" in Table 4 were placed within the Calcihaptophycidae clade but could not be placed to a clade with cultured representatives. Most of these were placed in defined clades consisting of environmental sequences only, such as Clades E and F for 18S and Clade PRY- LSU3.

Overall, the taxonomic assignment using metabarcoding is strongly constrained for some taxa, such as coccolithophores other than Coccolithales and Isochrysidales, by

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

the relatively low number of available reference sequences obtained from taxonomically verified material, as was shown in Young et al. (2014). This lack of reference sequences from morphologically characterized cells (from cultures or picked cells) resulted in overall good assignment of OTUs to the family level, but did not allow for detailed species-level identification for most of the coccolithophore OTUs in this study. The issue was especially pronounced in highly diverse families such as Syracosphaeraceae and Rhabdosphaeraceae, where 28S rRNA marker yielded high numbers of OTUs, but only two reference sequences were available for species-level taxonomic assignment. The power of metabarcoding combined with a curated reference sequence database was illustrated by our detection of a very rare species *Tergestiella adriatica*, that was considered to be extinct after K/Pg boundary (66 million years ago) until it was recently described in modern plankton (Hagino et al. 2015).

Abundance estimation

The quantitative analysis of coccolithophore community using SEM revealed a peak in abundance at 4 m (2.8×10^5 cells/l) decreasing gradually towards the 40 m depth (Fig. S4 and Table 3). The community was dominated by *E. huxleyi*, accounting for over 90% of the coccolithophore cell abundance in the top 4 m layer and decreasing in abundance and relative contribution (44–70%) in the deeper layers. Syracosphaeraceae showed an increase in contribution (up to 28%) in layers below 8 m depth, while other families, such as Rhabdosphaeraceae, Braarudosphaeraceae, and holococcolith taxa were present in lower numbers all along the vertical profile. It is important to note that the coccolithophore abundance was slightly lower at 1 m compared to the DCM.

The quantitative analysis using metabarcoding showed large variation in the proportional abundance of reads between the markers (Fig. 8). The 18S rRNA showed overall similar trends to SEM in describing proportional abundance of the coccolithophore families. The OTUs belonging to Noelaerhabdaceae accounted for 72% at the 1 m depth and decreased to 28% at the DCM. Syracosphaeraceae were the second most abundant group, increasing in proportional abundance from 16% at the 1 m depth to 61% at the DCM. On the other hand, the 28S rRNA highly underestimated the proportional abundance of Noelaerhabdaceae, likely due to mismatches with the LSU1 primer pair, and showed high relative contribution (34%) of taxonomically unidentified 28S rRNA OTUs. However, the trends in the proportional abundance of Syracosphaeraceae and Rhabdosphaeraceae 28S rRNA OTUs followed the same pattern as the SEM counts. Unlike the SEM counts, both 18S and 28S rRNA metabarcoding showed a significant contribution of Calyptrosphaeraceae, notably at 1 m depth.

Using HTS for quantitative analysis of haptophyte communities is a challenging task, as the amount of RNA extracted from each species varies with size, growth rate and is likely group-specific (Egge et al. 2013). Scanning electron microscopy has been a standard method for qualitative and quantitative analysis of coccolithophore communities for many years (Bollmann et al. 2002). This study was the first one using SEM quantitative analysis to test the usage of HTS for coccolithophore proportional abundance estimation. Environmental sequencing of clone-libraries has previously showed weak correspondence between SEM counts and proportional abundance of 28S rRNA gene OTUs (Young et al. 2014), largely owing to poor representation of Noelaerhabdaceae sequences. We observed the same marked underrepresentation compared to SEM counts. In this study, the low representation of Noelaerhabdaceae is likely related to primer mismatch, and the marker could still be suitable for proportional abundance estimation if primers without any mismatches are used. The main advantages of the SEM technique; a high degree of taxonomic precision, well-established, morphology-based taxonomy of the group (Cros and Fortuño 2002; Young et al. 2003) and availability of absolute abundance data, were all confirmed in this study.

Methodological considerations

In this study we used RNA as template because we were interested in living cells of haptophytes that seem to be more represented in the total RNA than DNA (Massana et al. 2015). However, RNA requires reverse transcription into cDNA which may introduce additional chimeras (Egge et al. 2013) that need to be identified and removed. In monitoring surveys of total protist community DNA could therefore be the preferred template.

Here, we contribute with a curated 28S rRNA gene reference database based on cultures with updated taxonomy verified by phylogeny. With this gene some haptophyte taxa could not be taxonomically assigned to a major clade consisting of sequences from cultures or only environmental samples (e.g. Clades HAP-3), due to a constrained reference database. We observed, however, a considerably higher richness using the 28S rRNA than 18S. Both markers showed, however, the same percentage of OTUs and reads for the most diverse and abundant families. We also tested the need for technical replicates in metabarcoding and found that it do not add considerable information in studies focusing on the most abundant groups, but is important in recovering low abundant OTUs.

Some of the possible drawbacks of the SEM method were also observed. Most importantly, the method is time consuming and requires taxonomic expertise, meaning that it is not suitable for routine monitoring of coccolithophore communities or surveys with a high number of samples. And the smaller sample volume that is practical to examine result in that some species are overlooked.

We suggest that 28S can be a useful complement to 18S in haptophyte metabarcoding studies. While the 28S rRNA gene seems to better distinguish between closely related coccolithophore species, the 18S rRNA gene has more reference sequences of defined clades and better assign to major clades. The observed mismatch to Isochrysidales with the LSU1 primer pair could account for

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 3. Species list of coccolithophores detected by SEM and their abundances (cells/l) at the sampled depths

Species	Author	1 m	2 m	4 m	8 m	12 m	16 m	20 m	40 m
Noelaerhabdaceae									
Emiliania huxleyi TYPE A	Young & Westbroek, 1991	214,077	237,131	253,518	170,860	140,227	99,769	80,972	16,387
Rhabdosphaeraceae									
Acanthoica quattrospina	Lohmann, 1903	602	2,410	964	2,410	2,009	1,928	964	482
Acanthoica quattrospina HOL[a]	Cros et al. 2000				241		482		
Algirosphaera robusta	(Lohmann 1902) Norris, 1984					402			
Algirosphaera robusta HOL[a]	(Schiller 1913) Deflandre, 1952		482						
Rhabdosphaera xiphos[a]	(Deflandre & Fert 1954) Norris, 1984				241	1,205	964	482	1,928
Syracosphaeraceae									
Calciopappus caudatus	Gaarder & Ramsfjell, 1954						482		
Ophiaster hydroideus	(Lohmann 1903) Lohmann, 1913				6,266	2,009	9,157	5,784	5,302
Ophiaster minimus[a]	Manton & Oates, 1983				1,446	402	964	482	964
Syracosphaera anthos	(Lohmann 1912) Janin, 1987								
Syracosphaera borealis	Okada & McIntyre, 1977	1,607	482	964	241	804		964	
Syracosphaera corolla[a]	Lecal, 1966			482		402	482	1,446	
Syracosphaera halldalii[a]	Gaarder in Gaarder & Hasle 1971 ex Jordan, 1994	803	1,446	1,928	5,784	4,822	9,639	10,121	1,928
Syracosphaera marginaporata	Knappertsbusch, 1993	7,230	7,712	13,977	46,992	23,706	15,423	18,797	7,712
Syracosphaera molischii	Young, 2003	201			7,953	4,420	4,338	2,892	482
Syracosphaera nodosa	Kamptner, 1941				1,446	1,607	964	1,928	
Syracosphaera nodosa COMB[a]	This study					402	482	482	
Syracosphaera nodosa HOL[a]	This study				1,687	1,607	4,820	3,374	
Syracosphaera ossa[a]	(Lecal 1966) Loeblich & Tappan, 1968	602	964	1,928	1,687	804		482	
Pappsphaeraceae									
Pappsphaera lepida	Tangen, 1972								964
Pappomonas flabellifera	Manton & Oates, 1975								482
Calyptrosphaeraceae (holococcoliths)									
Calyptrosphaera sphaeroidea	Schiller 1913			964	2,410	4,822	1,446		
Corisphaera strigilis[a]	Gaarder, 1962		482		723	804			
Undetermined taxa									
Undetermined HOL[a]							482		
Undetermined HET[a]							964		
Braarudosphaeraceae									
Braarudosphaera bigelowii	(Gran & Braarud 1935) Deflandre, 1947	201			1,446	8,840	5,784	3,856	

[a]Taxa previously not reported from the Oslofjorden area.
Columns in grey mark the stations used for the method comparison.

the difference in relative abundance estimates between markers. Therefore, future studies should use modified 28S primers without this mismatch. To improve the resolution and specificity of metabarcoding protist communities we recommend that more reference sequences of both the 28S and 18S rRNA genes are produced from cultures and isolated single cells. We also suggest the construction of a curated concatenated 18S and 28S rRNA reference database, that also will confirm the link between 18S and 28S rRNA clades without cultured representatives.

CONCLUSIONS

A tenth of the OTUs with both markers matched a cultured species. More than half of the 18S OTUs had not previously been recorded in the area, showing that the majority of the OTUs had a best match with an environmental sequence and were recorded for the first time for the Outer Oslofjorden and Skagerrak. Six coccolithophore species were recorded for the first time in the area by SEM. The species composition differed significantly at the two depths and the diversity revealed by 18S rRNA was significantly higher at

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Table 4. Number of OTUs and morphospecies within coccolithophore families obtained using the two molecular markers and the SEM

	18S rRNA (OTUs)			28S rRNA (OTUs)			SEM (morphospecies)			
	1 m	DCM	Total	1 m	DCM	Total	1 m	DCM	Total	Total (OF2)
Noelaerhabdaceae	1	4	**4**	6	2	**6**	1	1	1	**1**
Calcidiscaceae	0	0	**0**	1	0	**1**	0	0	0	**0**
Coccolithaceae	0	1	**1**	1	0	**1**	0	0	0	**0**
Helicosphaeraceae	0	0	**0**	0	1	**1**	0	0	0	**0**
Pontosphaeraceae	0	0	**0**	0	0	**1**	0	0	0	**0**
Rhabdosphaeraceae	2	1	**2**	5	8	**8**	1	2	2	**3**
Syracosphaeraceae	1	2	**2**	18	36	**38**	5	8	8	**11**
Pappospheraceae	0	0	**0**	0	0	**0**	0	0	0	**2**
Calyptrosphaeraceae	1	1	**1**	4	5	**6**	0	2	2	**2**
Pleurochrysidaceae	0	0	**0**	1	0	**1**	0	0	0	**0**
Braarudosphaeraceae	2	1	**2**	0	0	**0**	1	1	1	**1**
Other[a]	6	13	**17**	15	17	**25**	0	0	0	**2**
Sum:	13	23	29	51	69	87	8	14	14	22

The number of species per family obtained using SEM data includes species from 8 analysed depths. Values in bold represent the total number of coccolithophore OTUs detected by HTS at 1 m depth and at DCM as well as the total number of morphospecies detected by SEM at the 8 analysed depths.

[a]Category "Other" includes OTUs placed within the Watznaueriaceae and Isochrysidaceae families as well as OTUs placed within clades Calcihaptophycidae, Coccolithales, Zygodiscales, Clade-E, Clade-F and OTUs with an unclear tree placement. Taxa marked as "Other" in the SEM data represent two unidentified morphotypes with an unclear taxonomy.

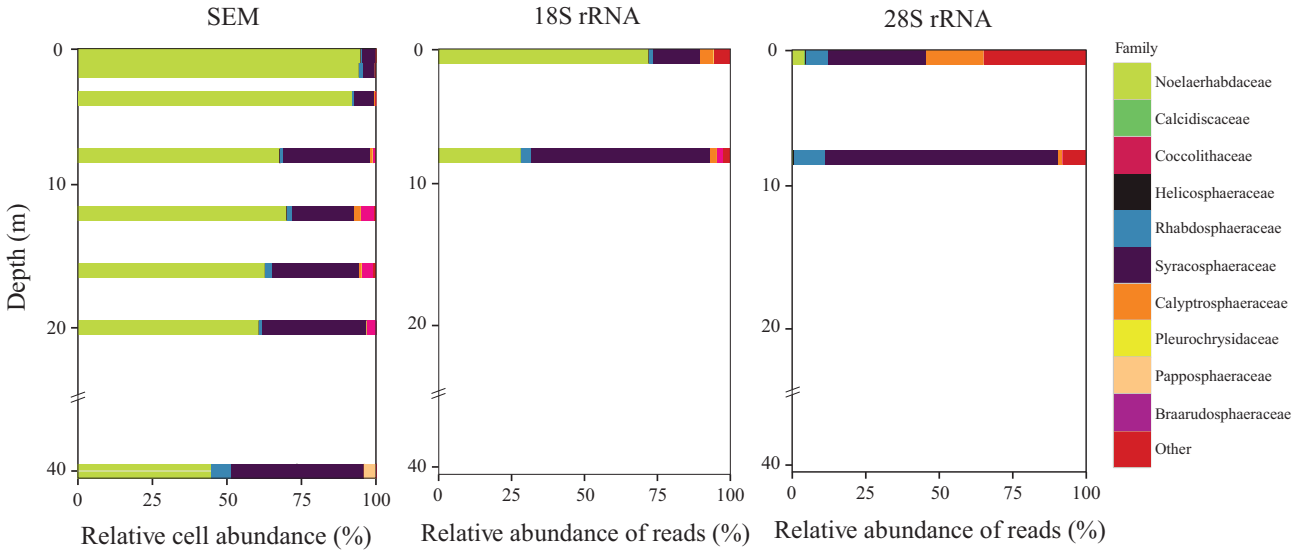


Figure 8 Proportional abundance of coccolithophore families inferred from SEM counts at eight depths and number of reads obtained using 18S and 28S rRNA markers at 1 and 8 m depths.

DCM than at subsurface. This shows that there is a need to sample more than one depth to reveal the full diversity. Further, the picoplankton size fraction contained more OTUs than the nanoplankton, even if only a few of the described haptophyte species are 3 μm or smaller (Edvardsen et al. 2016). We conclude from this that there is a large haptophyte diversity that remains to be described both morphologically and genetically, especially in the picoplankton, even in a relatively well-studied area for phytoplankton diversity, as the Skagerrak.

ACKNOWLEDGMENTS

This research was supported by the Research Council of Norway (FRIMEDBIO project 197823 to Jorijntje Henderiks, HAPTODIV project 190307 to BE and EE, and MICROPOLAR project 225956 to BE and EE), and the University of Oslo strategic funding (to SGS and BE). We thank captain Sindre Holm and crew on board R/V Trygve Braarud for assistance during field work, Ramiro Logares for advice on bioinformatics protocols and Tom Andersen

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

for statistical suggestions. AmpliconNoise was run within QIIME on the Biocluster at the Institut de Ciències del Mar (CSIC) in Barcelona, and Mothur on the Abel cluster at the University of Oslo.

LITERATURE CITED

Andersen, R. A., Bidigare, R. R., Keller, M. D. & Latasa, M. 1996. A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Res. II Top. Stud. Oceanogr.*, 43:517–537

Bittner, L., Gobet, A., Audic, S., Romac, S., Egge, E. S., Santini, S., Ogata, H., Probert, I., Edvardsen, B. & de Vargas, C. 2013. Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Mol. Ecol.*, 22:87–101.

Bollmann, J., Cortés, M. Y., Haidar, A. T., Brabec, B., Close, A., Hofmann, R., Palma, S., Tupas, L. & Thierstein, H. R. 2002. Techniques for quantitative analyses of calcareous marine phytoplankton. *Mar. Micropaleontol.*, 44:163–185.

Cros, L. & Fortuño, J. M. 2002. Atlas of Northwestern Mediterranean coccolithophores. *Sci. Mar.*, 66:7–182.

Dahl, E. & Johannessen, T. 1998. Temporal and spatial variability of phytoplankton and chlorophyll a: lessons from the south coast of Norway and the Skagerrak. *ICES J. Mar. Sci.*, 55:680–687.

Edvardsen, B., Egge, E. S. & Vaulot, D. 2016. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review. *Perspect. Phycol.*, 3:77–91.

Edvardsen, B., Eikrem, W., Throndsen, J., Sáez, A. G., Probert, I. & Medlin, L. K. 2011. Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). *Eur. J. Phycol.*, 46:202–228.

Edvardsen, B. & Paasche, E. 1998. Bloom dynamics and physiology of *Prymnesium* and *Chrysochromulina*. *NATO ASI Ser. G Ecol. Sci.*, 41:193–208.

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C. & Edvardsen, B. 2013. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS One*, 8: e74371.

Egge, E. S., Eikrem, W. & Edvardsen, B. 2015a. Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *J. Eukaryot. Microbiol.*, 62:121–140.

Egge, E. S., Johannessen, T. V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R.-A. & Edvardsen, B. 2015b. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol. Ecol.*, 24:3026–3042.

Eikrem, W. 1996. *Chrysochromulina throndsenii* sp. nov. (Prymnesiophyceae). Description of a new haptophyte flagellate from Norwegian waters. *Phycologia*, 35:377–380.

Eikrem, W. 1999. The class Prymnesiophyceae (Haptophyta) in Scandinavian waters. Dissertation. University of Oslo, Oslo, Norway. 214 p.

Eikrem, W. & Edvardsen, B. 1999. *Chrysochromulina fragaria* sp. nov. (Prymnesiophyceae), a new haptophyte flagellate from Norwegian waters. *Phycologia*, 38:149–155.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T. & Martin, J.-F. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genom.*, 12:245.

Granéli, E., Edvardsen, B., Roelke, D. L. & Hagström, J. A. 2012. The ecophysiology and bloom dynamics of Prymnesium spp. *Harmful Algae*, 14:260–270.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41:D597–D604.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., The Human Microbiome Consortium, Petrosino, J. F., Knight, R. & Birren, B. W. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.*, 21:494–504.

Hagino, K., Young, J. R., Bown, P. R., Godrijan, J., Kulhanek, D. K., Kogame, K. & Horiguchi, T. 2015. Re-discovery of a "living fossil" coccolithophore from the coastal waters of Japan and Croatia. *Mar. Micropaleontol.*, 116:28–37.

Holligan, P. M., Fernández, E., Aiken, J., Balch, W. M., Boyd, P., Burkill, P. H., Finch, M., Groom, S. B., Malin, G., Muller, K., Purdie, D. A., Robinson, C., Trees, C. C., Turner, S. M. & van der Wal, P. 1993. A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North Atlantic. *Global Biogeochem. Cycles*, 7:879–900.

Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.*, 12:1889–1898.

Iglesias-Rodríguez, M. D., Brown, C. W., Doney, S. C., Kleypas, J., Kolber, D., Kolber, Z., Hayes, P. K. & Falkowski, P. G. 2002. Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. *Global Biogeochem. Cycles*, 16:1100.

Jensen, M. Ø. 1998. The genus *Chrysochromulina* (Prymnesiophyceae) in Scandinavian coastal waters – diversity, abundance and ecology. Dissertation. University of Copenhagen, Copenhagen, Denmark. 200 p.

Jordan, R. W. & Chamberlain, A. H. L. 1997. Biodiversity among haptophyte algae. *Biodivers. Conserv.*, 6:131–152.

Jordan, R. W., Cros, L. & Young, J. R. 2004. A revised classification scheme for living haptophytes. *Micropaleontology*, 50:55–79.

Kuylenstierna, M. & Karlson, B. 1994. Seasonality and composition of Pico- and Nanoplanktonic cyanobacteria and protists in the Skagerrak. *Bot. Mar.*, 37:17–33.

Langer, G., Geisen, M., Baumann, K.-H., Kläs, J., Riebesell, U., Thoms, S. & Young, J. R. 2006. Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochem. Geophys. Geosyst.*, 7:Q09006.

Lekve, K., Bagøien, E., Dahl, E., Edvardsen, B., Skogen, M. D. & Stenseth, N. C. 2006. Environmental forcing as a main determinant of bloom dynamics of the *Chrysochromulina* algae. *Proc. Biol. Sci.*, 273:3047–3055.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F. & de Vargas, C. 2009. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc. Natl Acad. Sci.*, 106:12803–12808.

Malinverno, E., Ziveri, P. & Corselli, C. 2003. Coccolithophorid distribution in the Ionian Sea and its relationship to eastern

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Mediterranean circulation during late fall to early winter 1997. *J. Geophys. Res.*, 108:8115.
Man-Aharonovich, D., Philosof, A., Kirkup, B. C., Le Gall, F., Yogev, T., Berman-Frank, I., Polz, M. F., Vaulot, D. & Béjà, O. 2010. Diversity of active marine picoeukaryotes in the Eastern Mediterranean Sea unveiled using photosystem-II psbA transcripts. *ISME J.*, 4:1044–1052.
Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Forn, I., Forster, D., Guillou, L., Jaillon, O., Kooistra, W. H., Logares, R., Mahe, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M. C., Probert, I., Romac, S., Richards, T., Santini, S., Shalchian-Tabrizi, K., Siano, R., Simon, N., Stoeck, T., Vaulot, D., Zingone, A. & de Vargas, C. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.*, 17:4035–4049.
Medlin, L. & Zingone, A. 2007. A taxonomic review of the genus *Phaeocystis*. *Biogeochemistry*, 83:3–18.
Moestrup, Ø. 1994. Economic aspects: 'blooms', nuisance species, and toxins. *In:* Green, J. E. & Leadbeater, B. S. C. (eds), The Haptophyte Algae. Clarendon Press, Oxford. 51:265–285.
Moon-van der Staay, S. Y., de Wachter, R. & Vaulot, D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 409:607–610.
Moon-van der Staay, S. Y., van der Staay, G. W. M., Guillou, L., Vaulot, D., Claustre, H. H. & Medlin, L. K. 2000. Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnol. Oceanogr.*, 45:98–109.
Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R. & Schäffer, A. A. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics*, 24:1757–1764.
Not, F., del Campo, J., Balagué, V., de Vargas, C. & Massana, R. 2009. New insights into the diversity of marine picoeukaryotes. *PLoS One*, 4(9):e7143.
Pielou, E. C. 1966. The measurement of diversity in different types of biological collections. *J. Theor. Biol.*, 13:131–144.
Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F. & Sloan, W. T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods*, 6:639–641.
Ridgwell, A., Schmidt, D. N., Turley, C., Brownlee, C., Maldonado, M. T., Tortell, P. & Young, J. R. 2009. From laboratory manipulations to Earth system models: scaling calcification impacts of ocean acidification. *Biogeosciences*, 6:2611–2623.
Robasky, K., Lewis, N. E. & Church, G. M. 2014. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.*, 15:56–62.
Robertson, J. E., Robinson, C., Turner, D. R., Holligan, P., Watson, A. J., Boyd, P., Fernandez, E. & Finch, M. 1994. The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991. *Deep Sea Res. I: Oceanogr. Res. Pap.*, 41:297–314.
Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75:7537–7541.
Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J. & Vaulot, D. 2009. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS One*, 4:e7657.
Speksnijder, A. G. C. L., Kowalchuk, G. A., De Jong, S., Kline, E., Stephen, J. R. & Laanbroek, H. J. 2001. Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl. Environ. Microbiol.*, 67:469–472.
Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22:2688–2690.
Stoeck, T., Zuendorf, A., Breiner, H. & Behnke, A. W. 2007. A molecular approach to identify active microbes in environmental eukaryote clone libraries. *Microb. Ecol.* 53:328–339.
Thomsen, H. A., Buck, K. R. & Chavez, F. P. 1994. Haptophytes as components of marine phytoplankton. *In:* Green, J. E. & Leadbeater, B. S. C. (eds), The Haptophyte Algae. Clarendon Press, Oxford. 51:187–208.
de Vargas, C., Aubry, M. P., Probert, I. & Young, J. 2007. Origin and evolution of coccolithophores: from coastal hunters to oceanic farmers. *In:* Falkowski, P. G. & Knoll, A. H. (eds.), Evolution of Primary Producers in the Sea. Elsevier, Boston, MA. p. 251–285.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P. & Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605-1-11.
Young, J. R., Geisen, M., Cros, L., Kleijne, A., Sprengel, C., Probert, I. & Østergaard, J. 2003. A guide to extant coccolithophore taxonomy. *J. Nannoplankton Res. Spec.*, 1:1–124.
Young, J. R., Liu, H., Probert, I., Aris-Brosou, S. & de Vargas, C. 2014. Morphospecies versus phylospecies concepts for evaluating phytoplankton diversity: the case of the Coccolithophores. *Cryptogamie Algol.*, 35:353–377.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1. (a) Heat map showing proportional abundances of all 18S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by colour (white indicates that no reads were recorded). (b) Heat map showing proportional abundances of all 28S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by colour (white indicates that no reads were recorded).
Figure S2. Rank-abundance curves for 18S (a) and 28S rRNA gene (b).
Figure S3. Mean-difference (Bland–Altman) plot showing level of agreement between technical replicates for OTU proportional abundances in the 18S and 28S rRNA gene

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

data sets.

Figure S4. Vertical distribution of coccolithophore families observed at the OF2 station.

Table S1. Total number of reads at the beginning and end of the bioinformatics analysis and changes in the number of unique sequences (OTUs, operational taxonomic units) along the analysis process.

Table S2. (a) Haptophyte V4 18S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement. (b) Haptophyte D1-D2 28S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement.

Table S3. Total and proportional read abundances and OTUs within each major clade for 18S and 28S rRNA genes.

Table S4. (a) Overview over matching of 18S OTUs to other databases. Total: Total number of OTUs in each group. ≥ 99% any sequence: Number of OTUs that have ≥ 99% BLAST match with either any sequence in the Haptophyta-PiP database, or an OTU from Oslofjorden from Egge et al. 2015a. ≥ 99% Hapto-PiP_ENV: Number of OTUs that have ≥ 99% BLAST match with an "environmental sequence" in the Haptophyta-PiP database. ≥ 99% Hapto-PiP_CULT: Number of OTUs that have ≥ 99% BLAST match with a sequence from a cultured species in the Haptophyta-PiP database. ≥ 99% OF OTUs: Number of OTUs that have ≥ 99% BLAST match with an OTU previously obtained by HTS of samples from Oslofjorden (these may represent either cultured species, environmental sequences obtained by Sanger sequencing, or novel sequences from the Egge et al. 2015a study). ≥ 99% OF OTU & < 99% with any Hapto-PiP sequence: Number of OTUs that have ≥ 99% match to an OTU from Egge et al. (2015 a,b), but at the same time is < 99% similar to any sequence present in Hapto-PiP. The numbers from ≥ 99% Hapto-PiP_ENV and ≥ 99% Hapto-PiP_CULT may not add up, because environmental sequences in the Haptophyta-PiP database may also come from species that exist in culture. (b) Overview over matching of 28S OTUs to the 28S haptophyta reference database, consisting of sequences from cultured strains.

Data S1. Description of how the haptophyte 28S reference database was created.

© 2016 The Authors Journal of Eukaryotic Microbiology published by Wiley Periodicals, Inc. on behalf of International Society of Protistologists

Haptophyte Diversity and Vertical Distribution Explored by 18S and 28S Ribosomal RNA Gene Metabarcoding and Scanning Electron Microscopy

SUPPLEMENTARY INFORMATION

Sandra Gran-Stadniczeñko[a], Luka Šupraha[b], Elianne D. Egge[a] and Bente Edvardsen[a]

[a] Department of Biosciences, University of Oslo, PO Box 1066 Blindern, Oslo 0316, Norway
[b] Department of Earth Sciences, Uppsala University, Villavägen 16, Uppsala 75236, Sweden

Figure S1a. Heat map showing proportional abundances of all 18S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by color (white indicates that no reads were recorded).

Figure S1b. Heat map showing proportional abundances of all 28S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by color (white indicates that no reads were recorded).

Figure S2. Rank-abundance curves for 18S (a) and 28S rRNA gene (b). The abundant OTUs are shown in blue (the three most abundant in dark blue). Red indicates rare taxa.

Figure S3. Mean-difference (Bland-Altman) plot showing level of agreement between technical replicates for OTU proportional abundances in the 18S and 28S rRNA gene datasets. Average differences (± 1.45 and ± 0.83 standard deviation of the difference for 18S and 28S respectively) are represented as dotted lines.

Figure S4. Vertical distribution of coccolithophore families observed at the OF2 station. Samples at 1 m and 8 m depths were used for the method comparison.

Table S1. Total number of reads at the beginning and end of the bioinformatics analysis and changes in the number of unique sequences (OTUs, operational taxonomic units) along the analysis process.

Table S2a. Haptophyte V4 18S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement.

Table S2b. Haptophyte D1-D2 28S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement.

Table S3. Total and proportional read abundances and OTUs within each major clade for 18S and 28S rRNA genes.

Table S4. a) Overview over matching of 18S OTUs to other databases. Total: Total number of OTUs in each group. ≥ 99% any sequence: Number of OTUs that have ≥ 99% BLAST match with either any sequence in the Haptophyta-PiP database, or an OTU from Oslofjorden from Egge et al. 2015a. ≥ 99% Hapto-PiP_ENV: Number of OTUs that have ≥ 99% BLAST match with an "environmental sequence" in the Haptophyta-PiP database. ≥ 99% Hapto-PiP_CULT: Number of OTUs that have ≥ 99% BLAST match with a sequence from a cultured species in the Haptophyta-PiP database. ≥ 99% OF-OTUs: Number of OTUs that have ≥ 99% BLAST match with an OTU previously obtained by HTS of samples from Oslofjorden (these may represent either cultured species, environmental sequences obtained by Sanger sequencing, or novel sequences from the Egge et al. 2015a study). ≥ 99% OF-OTU & < 99% with any Hapto-PiP sequence: Number of OTUs that have ≥ 99% match to an OTU from Egge et al. (2015 a,b), but at the same time is < 99% similar to any sequence present in Hapto-PiP. The numbers from ≥ 99% Hapto-PiP_ENV and ≥ 99% Hapto-PiP_CULT may not add up, because environmental sequences in the Haptophyta-PiP database may also come from species that exist in culture.
b) Overview over matching of 28S OTUs to the 28S haptophyta reference database, consisting of sequences from cultured strains.

File S1. Description of how the haptophyte 28S reference database was created.

IMPORTANT REMARK

The link to the 18S and 28S rRNA curated gene alignments presented in page 516 of **Paper I** is wrong. Please find them in the following links:

18S rRNA: https://figshare.com/articles/Diversity_and_distribution_of_haptophytes_reavealed_by_environmental_sequencing_a_review_Perspectives_in_Phycology_Supplementary_material_/2759983

28S rRNA: https://figshare.com/s/13fef76f5fa3d7eb9b56

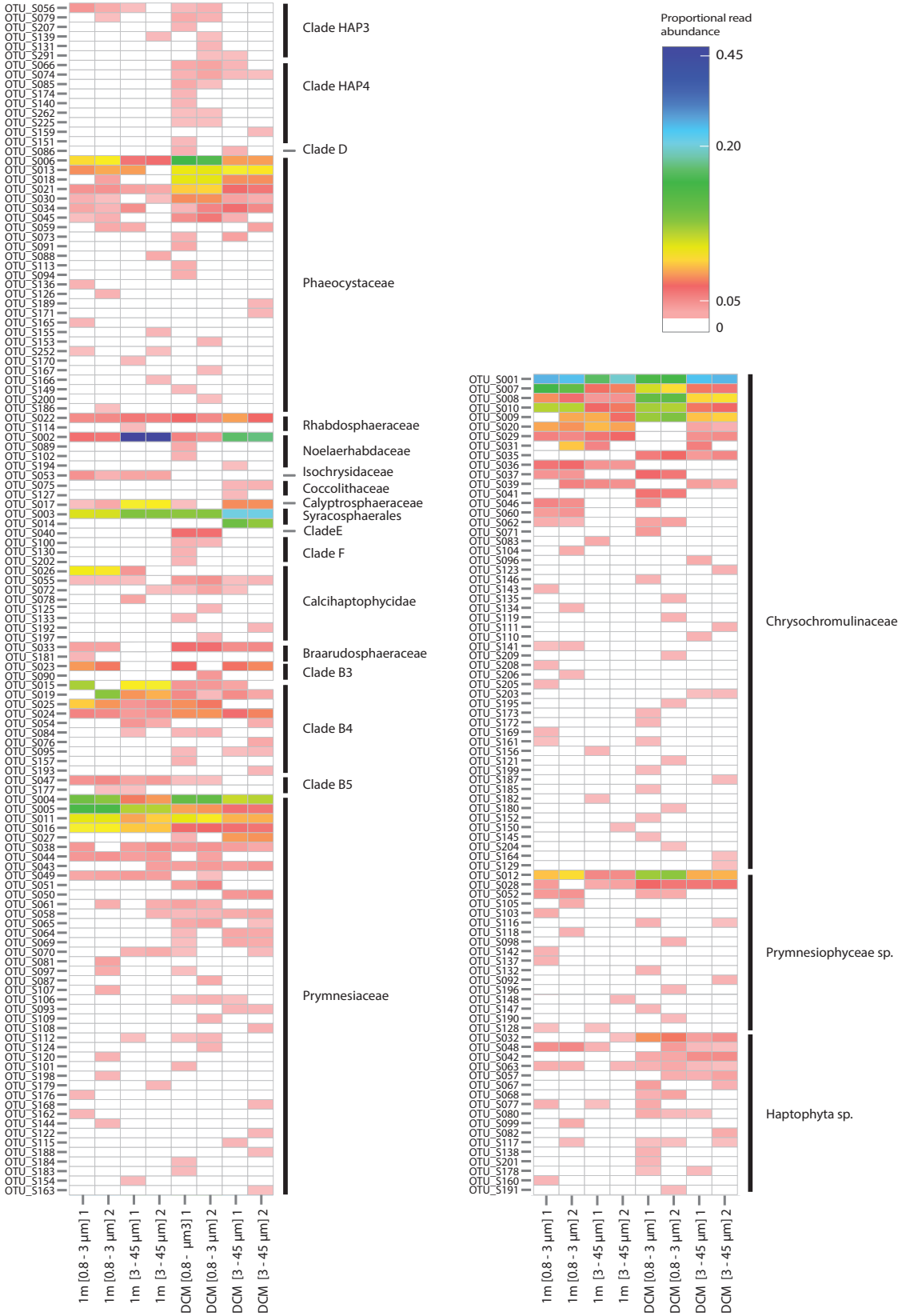


Figure S1a. Heat map showing proportional abundances of all 18S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by color (white indicates that no reads were recorded).

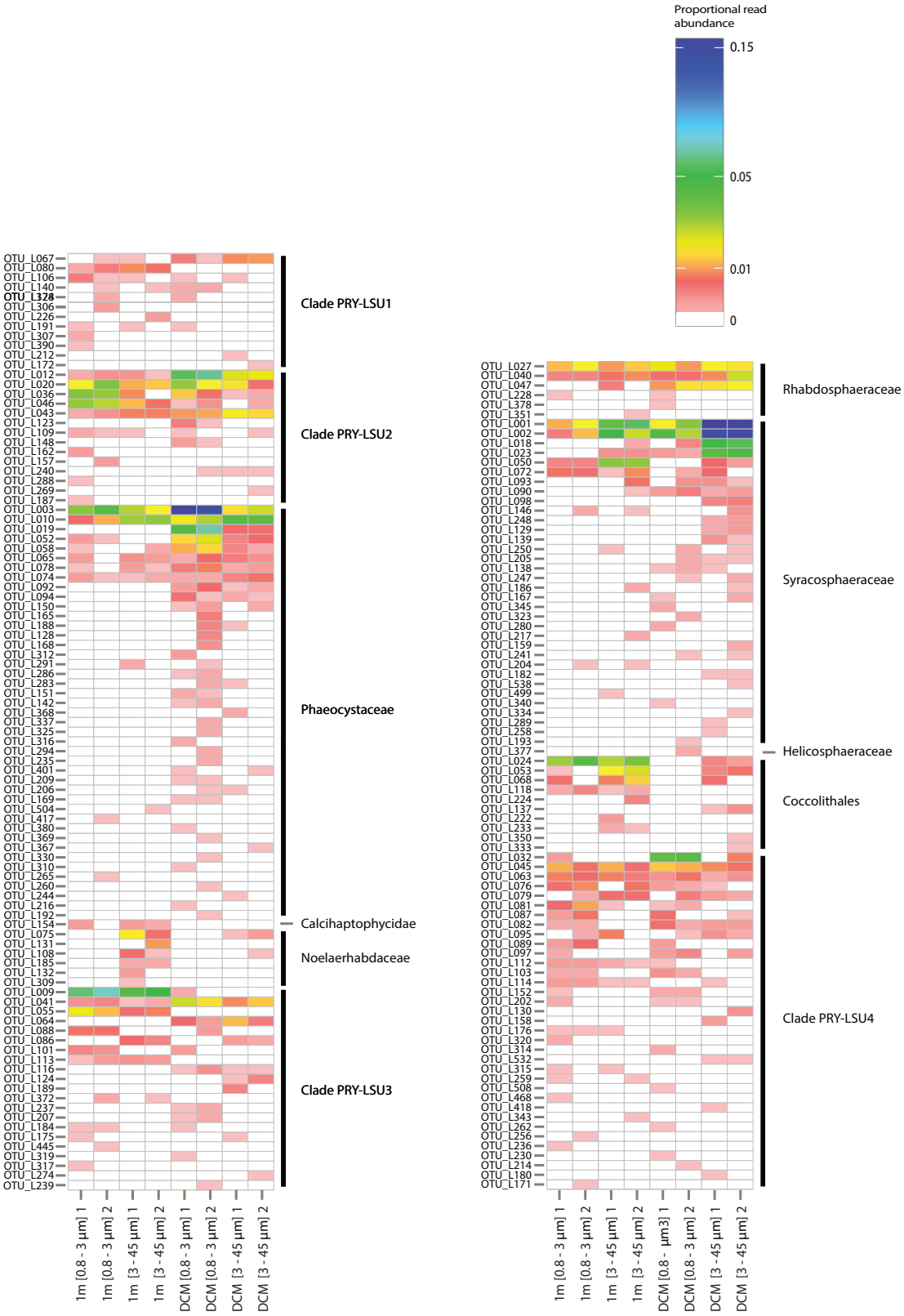


Figure S1b. Heat map showing proportional abundances of all 28S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by color (white indicates that no reads were recorded

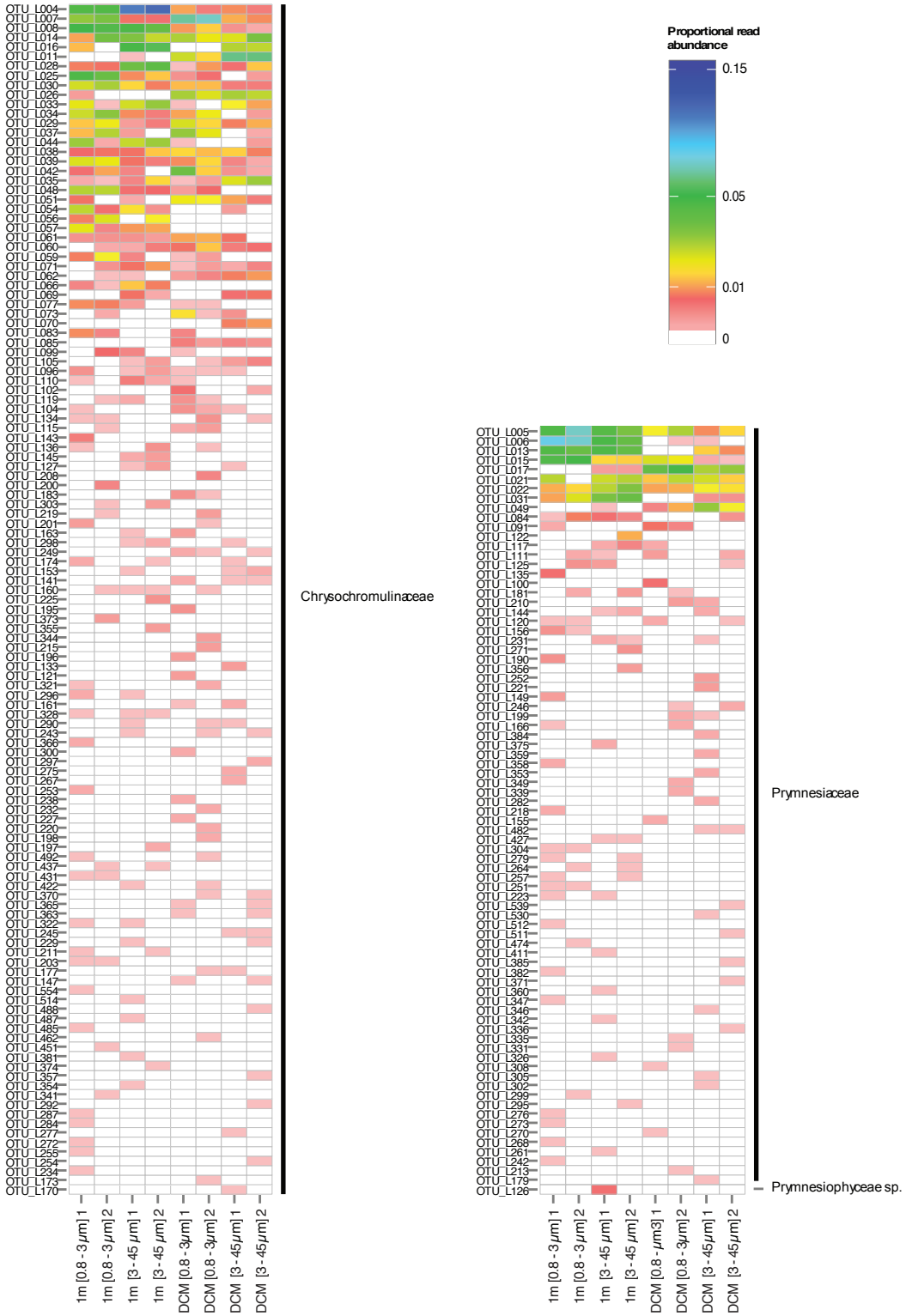


Figure S1b continued. Heat map showing proportional abundances of all 28S rRNA gene OTUs for the different samples and replicates. Proportional read abundance was scaled by color (white indicates that no reads were recorded).

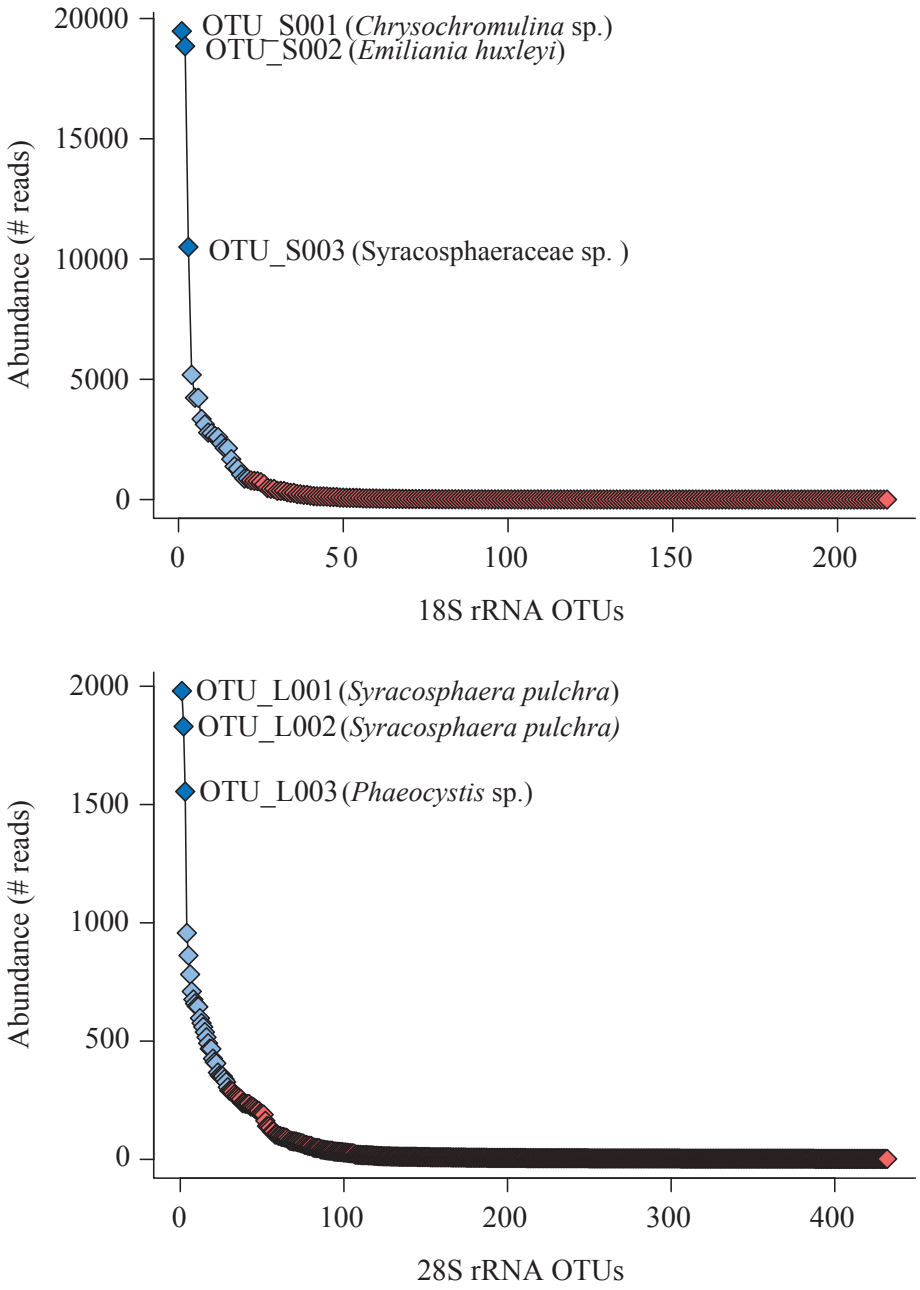


Figure S2. Rank-abundance curves for 18S (a) and 28S rRNA gene (b). The abundant OTUs are shown in blue (the three most abundant in dark blue). Red indicates rare taxa.

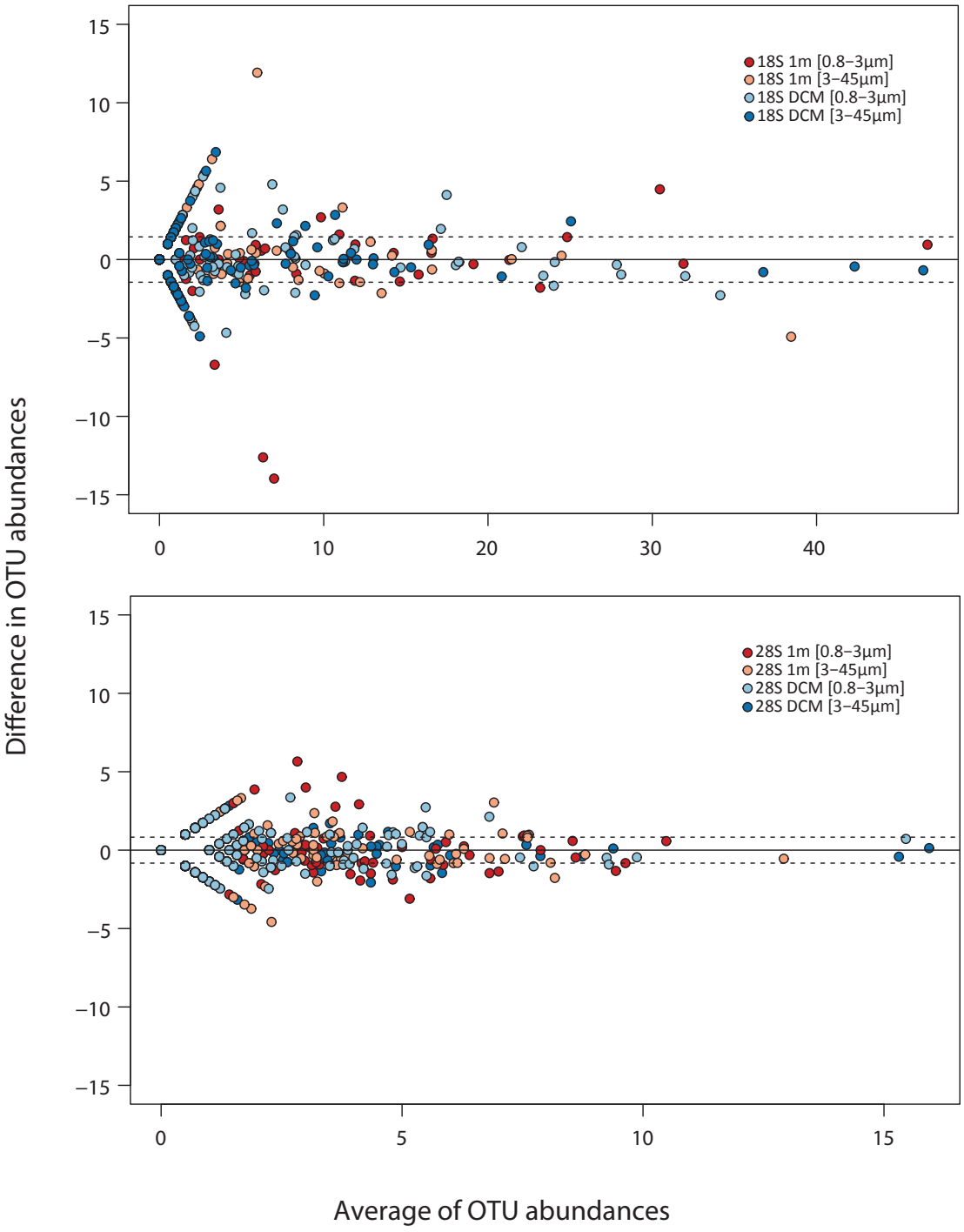


Figure S3. Mean-difference (Bland-Altman) plot showing level of agreement between technical replicates for OTU proportional abundances in the 18S and 28S rRNA gene datasets. Average differences (± 1.45 and ± 0.83 standard deviation of the difference for 18S and 28S respectively) are represented as dotted lines.

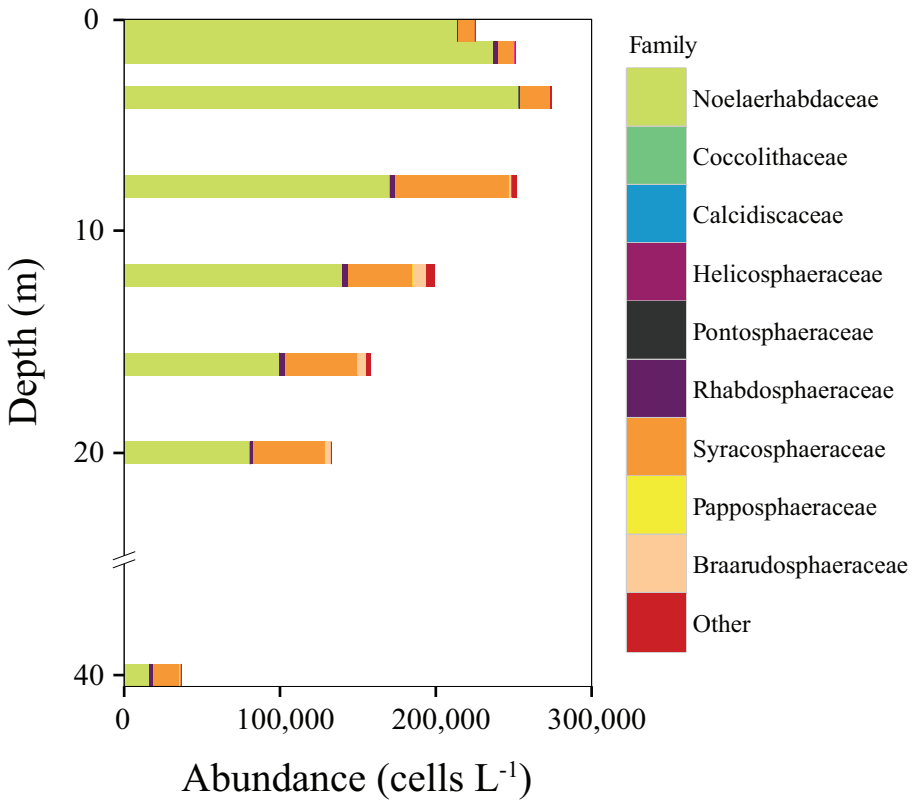


Figure S4. Vertical distribution of coccolithophore families observed at the OF2 station. Samples at 1 m and 8 m depths were used for the method comparison.

Table S1. Total number of reads at the beginning and end of the bioinformatics analysis and changes in the number of unique sequences (OTUs, operational taxonomic units) along the analysis process.

	18S rRNA		28S rRNA	
	Total reads	**Unique reads**	**Total reads**	**Unique reads**
AmpliconNoise output	120,282		38,795	
Reads after removal of <365 bp and homopolymers >8bp	112,958	2,096	30,981	2,594
Haptophyta reads	112,399	1,897	30,892	2,549
Chimeric reads removed	7,454		288	
Clustering output (number of OTUs)		722		1,135
(Double) Singletons removed		507		703
Final number of OTUs		**215**		**432**
Final number of reads	**104,345**		**29,751**	

Table S2a Haptophyte V4 18S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement.

OTU ID	Total reads (N)	Total reads (%)	Total reads after subsampling (N)	Total reads after subsampling (%)	Depth	Size fraction	Group	Lowest taxonomic level possible to determine
OTU S001	19473	18.66	13715	18.74	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S002	18849	18.06	11413	15.59	Both	Both	Noelaerhabdaceae	*Emiliania huxleyi*
OTU S003	10500	10.06	6669	9.11	Both	Both	Syracosphaeraceae	Syracosphaeraceae
OTU S004	5186	4.97	3893	5.32	Both	Both	Prymnesiaceae	*Prymnesium*
OTU S005	4234	4.06	3364	4.60	Both	Both	Prymnesiaceae	*Haptolina*
OTU S006	4231	4.05	3252	4.44	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S007	3349	3.21	2847	3.89	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S008	3126	3.00	2293	3.13	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S009	2784	2.67	1976	2.70	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S010	2757	2.64	2255	3.08	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S011	2650	2.54	1963	2.68	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU S012	2592	2.48	1928	2.63	Both	Both	Prymnesiophyceae	Prymnesiophyceae
OTU S013	2334	2.24	1614	2.21	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S014	2148	2.06	1258	1.72	Only DCM	Only nano	Syracosphaeraceae	Syracosphaeraceae
OTU S015	2132	2.04	1690	2.31	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S016	1679	1.61	1235	1.69	Both	Both	Prymnesiaceae	*Prymnesium polylepis*
OTU S017	1378	1.32	810	1.11	Both	Both	Calyptrosphaeraceae	*Calyptrosphaera sphaeroidea*
OTU S018	1284	1.23	927	1.27	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S019	1037	0.99	775	1.06	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S020	871	0.83	632	0.86	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S021	866	0.83	640	0.87	Both	Both	Phaeocystaceae	*Phaeocystis cordata*

OTU S022	797	0.76	520	0.71	Both	Both	Rhabdosphaeraceae	Algirosphaera robusta
OTU S023	778	0.75	559	0.76	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S024	765	0.73	557	0.76	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S025	722	0.69	579	0.79	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S026	633	0.61	577	0.79	Only 1m	Both	Calcihaptophycidae	Calcihaptophycidae
OTU S027	479	0.46	276	0.38	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU S028	449	0.43	298	0.41	Both	Both	Prymnesiophyceae	Prymnesiophyceae
OTU S029	443	0.42	287	0.39	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S030	364	0.35	276	0.38	Both	Both	Phaeocystaceae	*Phaeocystis globosa*
OTU S031	363	0.35	283	0.39	Both	Both	Chrysochromulinaceae	*Chrysochromulina scutellum*
OTU S032	355	0.34	273	0.37	Both	Both	Haptophyta	Haptophyta
OTU S033	307	0.29	229	0.31	Both	Both	Braarudosphaeraceae	Braarudosphaeraceae
OTU S034	280	0.27	192	0.26	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S035	278	0.27	196	0.27	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S036	229	0.22	187	0.26	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S037	220	0.21	183	0.25	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S038	196	0.19	140	0.19	Both	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU S039	192	0.18	130	0.18	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU S040	165	0.16	135	0.18	Only DCM	Only pico	Clade D-E-F	Clade E
OTU S041	143	0.14	118	0.16	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S042	134	0.13	82	0.11	Only DCM	Both	Haptophyta	Haptophyta
OTU S043	130	0.12	88	0.12	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU S044	129	0.12	96	0.13	Both	Both	Prymnesiaceae	*Prymnesium*
OTU S045	123	0.12	94	0.13	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S046	107	0.10	97	0.13	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S047	106	0.10	92	0.13	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B5
OTU S048	102	0.10	94	0.13	Both	Both	Haptophyta	Haptophyta

OTU S049	82	0.08	60	0.08	Both	Both	Prymnesiaceae	*Chrysocampanula spinifera*
OTU S050	79	0.08	49	0.07	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU S051	74	0.07	57	0.08	Only DCM	Only pico	Prymnesiaceae	Prymnesiaceae
OTU S052	71	0.07	60	0.08	Both	Only pico	Prymnesiophyceae	Prymnesiales
OTU S053	71	0.07	53	0.07	Only 1m	Both	Isochrysidaceae+Clade C1	*Isochrysis*
OTU S054	68	0.07	36	0.05	Both	Only nano	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S055	63	0.06	56	0.08	Both	Both	Calcihaptophycidae	Calcihaptophycidae
OTU S056	49	0.05	38	0.05	Both	Both	Clade HAP3-4	Clade HAP3
OTU S057	47	0.05	24	0.03	Only DCM	Both	Haptophyta	Haptophyta
OTU S058	47	0.05	26	0.04	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU S059	46	0.04	28	0.04	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU S060	44	0.04	41	0.06	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S061	38	0.04	32	0.04	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU S062	38	0.04	35	0.05	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S063	34	0.03	31	0.04	Both	Both	Haptophyta	Haptophyta
OTU S064	34	0.03	21	0.03	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU S065	34	0.03	24	0.03	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU S066	32	0.03	24	0.03	Only DCM	Both	Clade HAP3-4	Clade HAP4
OTU S067	31	0.03	21	0.03	Only DCM	Both	Haptophyta	Haptophyta
OTU S068	26	0.02	17	0.02	Only DCM	Only pico	Haptophyta	Haptophyta
OTU S069	26	0.02	17	0.02	Only DCM	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU S070	22	0.02	14	0.02	Both	Both	Prymnesiaceae	Prymnesiophyceae
OTU S071	22	0.02	19	0.03	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S072	22	0.02	16	0.02	Both	Both	Calcihaptophycidae	Calcihaptophycidae
OTU S073	22	0.02	18	0.02	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU S074	21	0.02	18	0.02	Only DCM	Both	Clade HAP3-4	Clade HAP4
OTU S075	20	0.02	7	0.01	Only DCM	Only nano	Coccolithales	*Coccolithus*

OTU S076	19	0.02	9	0.01	Only DCM	Only nano	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S077	19	0.02	14	0.02	Both	Both	Haptophyta	Haptophyta
OTU S078	18	0.02	11	0.02	Only 1m	Only nano	Calcihaptophycidae	*Tergestiella adriatica*
OTU S079	16	0.02	13	0.02	Both	Both	Clade HAP3-4	Clade HAP3
OTU S080	14	0.01	10	0.01	Only DCM	Both	Haptophyta	Haptophyta
OTU S081	14	0.01	13	0.02	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU S082	14	0.01	7	0.01	Both	Only nano	Haptophyta	Haptophyta
OTU S083	13	0.01	8	0.01	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S084	13	0.01	10	0.01	Both	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S085	12	0.01	10	0.01	Only DCM	Both	Clade HAP3-4	Clade HAP4
OTU S086	12	0.01	9	0.01	Only DCM	Both	Clade D-E-F	Clade D
OTU S087	11	0.01	8	0.01	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU S088	11	0.01	8	0.01	Only 1m	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S089	10	0.01	8	0.01	Only DCM	Only pico	Noelaerhabdaceae	*Emiliania huxleyi*
OTU S090	9	0.01	7	0.01	Only DCM	Only pico	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B3
OTU S091	9	0.01	8	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S092	9	0.01	4	0.01	Only DCM	Only nano	Prymnesiophyceae	Prymnesiophyceae
OTU S093	8	0.01	6	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU S094	8	0.01	6	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S095	8	0.01	5	0.01	Only DCM	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S096	8	0.01	7	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S097	8	0.01	8	0.01	Both	Only pico	Prymnesiaceae	*Prymnesium*
OTU S098	8	0.01	5	0.01	Only DCM	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S099	8	0.01	7	0.01	Only 1m	Only pico	Haptophyta	Haptophyta
OTU S100	8	0.01	7	0.01	Only DCM	Only pico	Clade D-E-F	Clade F
OTU S101	7	0.01	4	0.01	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU S102	7	0.01	4	0.01	Only DCM	Only pico	Noelaerhabdaceae	*Emiliania huxleyi*

OTU S103	7	0.01	7	0.01	Only 1m	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S104	7	0.01	7	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina scutellum*
OTU S105	7	0.01	7	0.01	Only 1m	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S106	7	0.01	6	0.01	Only DCM	Both	Prymnesiaceae	Prymnesiaceae
OTU S107	7	0.01	7	0.01	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU S108	7	0.01	5	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium polylepis*
OTU S109	7	0.01	5	0.01	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU S110	7	0.01	4	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S111	6	0.01	4	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S112	6	0.01	5	0.01	Both	Both	Prymnesiaceae	Prymnesiaceae
OTU S113	6	0.01	6	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S114	6	0.01	2	0.003	Only 1m	Only nano	Rhabdosphaeraceae	Rhabdosphaeraceae
OTU S115	6	0.01	3	0.004	Only DCM	Only nano	Prymnesiaceae	Prymnesiaceae
OTU S116	6	0.01	6	0.01	Only DCM	Both	Prymnesiophyceae	Prymnesiales
OTU S117	5	0.00	5	0.01	Both	Both	Haptophyta	Haptophyta
OTU S118	5	0.00	5	0.01	Only 1m	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S119	5	0.00	4	0.01	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S120	5	0.00	4	0.01	Only 1m	Only pico	Prymnesiaceae	Prymnesiaceae
OTU S121	5	0.00	3	0.004	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S122	5	0.00	3	0.004	Only DCM	Only nano	Prymnesiaceae	Prymnesiaceae
OTU S123	5	0.00	5	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina scutellum*
OTU S124	5	0.00	4	0.01	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU S125	4	0.00	3	0.004	Only DCM	Only pico	Calcihaptophycidae	Calcihaptophycidae
OTU S126	4	0.00	4	0.01	Only 1m	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S127	4	0.00	2	0.003	Only DCM	Only nano	Coccolithales	Coccolithales
OTU S128	4	0.00	2	0.003	Only 1m	Both	Prymnesiophyceae	Prymnesiophyceae
OTU S129	4	0.00	1	0.001	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*

OTU S130	4	0.00	4	0.01	Only DCM	Only pico	Clade D-E-F	Clade F
OTU S131	4	0.00	3	0.004	Only DCM	Only pico	Clade HAP3-4	Clade HAP3
OTU S132	4	0.00	4	0.01	Only DCM	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S133	4	0.00	3	0.004	Only DCM	Only pico	Calcihaptophycidae	Calcihaptophycidae
OTU S134	4	0.00	4	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S135	4	0.00	4	0.01	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S136	4	0.00	4	0.01	Only 1m	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S137	4	0.00	4	0.01	Only 1m	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S138	4	0.00	4	0.01	Only DCM	Only pico	Haptophyta	Haptophyta
OTU S139	4	0.00	3	0.004	Both	Both	Clade HAP3-4	Clade HAP3
OTU S140	4	0.00	3	0.004	Only DCM	Only pico	Clade HAP3-4	Clade HAP4
OTU S141	4	0.00	4	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S142	4	0.00	4	0.01	Only 1m	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S143	4	0.00	4	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S144	4	0.00	3	0.004	Only 1m	Only pico	Prymnesiaceae	Prymnesiaceae
OTU S145	4	0.00	2	0.003	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S146	4	0.00	4	0.01	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S147	4	0.00	3	0.004	Only DCM	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S148	4	0.00	3	0.004	Only 1m	Only nano	Prymnesiophyceae	Prymnesiophyceae
OTU S149	4	0.00	2	0.003	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S150	4	0.00	2	0.003	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S151	4	0.00	2	0.003	Only DCM	Only pico	Clade HAP3-4	Clade HAP4
OTU S152	4	0.00	2	0.003	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina scutellum*
OTU S153	4	0.00	3	0.004	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S154	4	0.00	2	0.003	Only 1m	Only nano	Prymnesiaceae	*Prymnesium polylepis*
OTU S155	4	0.00	3	0.004	Only 1m	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S156	4	0.00	3	0.004	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*

OTU S157	4	0.00	4	0.01	Only DCM	Only pico	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S158	3	0.00	0	0.000	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU S159	3	0.00	2	0.003	Only DCM	Both	Clade HAP3-4	Clade HAP4
OTU S160	3	0.00	3	0.004	Only 1m	Only pico	Haptophyta	Haptophyta
OTU S161	3	0.00	3	0.004	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S162	3	0.00	3	0.004	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU S163	3	0.00	1	0.001	Only DCM	Only nano	Prymnesiaceae	Prymnesiaceae
OTU S164	3	0.00	1	0.001	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S165	3	0.00	3	0.004	Only 1m	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S166	3	0.00	2	0.003	Only 1m	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S167	3	0.00	2	0.003	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S168	3	0.00	3	0.004	Only DCM	Only nano	Prymnesiaceae	*Haptolina*
OTU S169	3	0.00	3	0.004	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S170	3	0.00	2	0.003	Only 1m	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S171	3	0.00	3	0.004	Only DCM	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S172	3	0.00	3	0.004	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S173	3	0.00	3	0.004	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S174	3	0.00	3	0.004	Only DCM	Only pico	Clade HAP3-4	Clade HAP4
OTU S175	3	0.00	0	0.000	Only 1m	Only nano	Prymnesiophyceae	Prymnesiophyceae
OTU S176	3	0.00	3	0.004	Only 1m	Only pico	Prymnesiaceae	Prymnesiaceae
OTU S177	3	0.00	2	0.003	Only 1m	Both	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B5
OTU S178	3	0.00	3	0.004	Only DCM	Both	Haptophyta	Haptophyta
OTU S179	3	0.00	3	0.004	Only 1m	Only nano	Prymnesiaceae	*Prymnesium*
OTU S180	3	0.00	2	0.003	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S181	3	0.00	3	0.004	Only 1m	Only pico	Prymnesiophyceae	Braarudosphaeraceae
OTU S182	3	0.00	2	0.003	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S183	3	0.00	2	0.003	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*

OTU S184	3	0.00	2	0.003	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU S185	3	0.00	2	0.003	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S186	3	0.00	1	0.001	Both	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S187	3	0.00	2	0.003	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina rotalis*
OTU S188	3	0.00	2	0.003	Only DCM	Only nano	Prymnesiaceae	Prymnesiaceae
OTU S189	3	0.00	3	0.004	Only DCM	Only nano	Phaeocystaceae	*Phaeocystis*
OTU S190	3	0.00	2	0.003	Only DCM	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S191	3	0.00	2	0.003	Only DCM	Only pico	Haptophyta	Haptophyta
OTU S192	3	0.00	3	0.004	Only DCM	Only nano	Calcihaptophycidae	Calcihaptophycidae
OTU S193	3	0.00	3	0.004	Only DCM	Only nano	Prymnesiales Clade B3-B4-B5	Prymnesiales Clade B4
OTU S194	3	0.00	1	0.001	Only DCM	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU S195	3	0.00	3	0.004	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S196	3	0.00	3	0.004	Only DCM	Only pico	Prymnesiophyceae	Prymnesiophyceae
OTU S197	3	0.00	2	0.003	Only DCM	Only pico	Calcihaptophycidae	Calcihaptophycidae
OTU S198	3	0.00	3	0.004	Only 1m	Only pico	Prymnesiaceae	Prymnesiaceae
OTU S199	3	0.00	2	0.003	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S200	3	0.00	1	0.001	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU S201	3	0.00	3	0.004	Only DCM	Only pico	Haptophyta	Haptophyta
OTU S202	3	0.00	3	0.004	Only DCM	Only pico	Clade D-E-F	Clade F
OTU S203	3	0.00	3	0.004	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU S204	3	0.00	1	0.001	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S205	3	0.00	3	0.004	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina scutellum*
OTU S206	3	0.00	3	0.004	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S207	3	0.00	3	0.004	Only DCM	Only pico	Clade HAP3-4	Clade HAP3
OTU S208	3	0.00	3	0.004	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S209	3	0.00	3	0.004	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU S225	2	0.00	2	0.003	Only DCM	Only pico	Clade HAP3-4	Clade HAP4

OTU S228	2	0.00	0	0.000	Only 1m	Both	Clade HAP3-4	Clade HAP3
OTU S252	2	0.00	2	0.003	Only 1m	Both	Phaeocystaceae	*Phaeocystis*
OTU S253	2	0.00	0	0.000	Only DCM	Both	Prymnesiaceae	Prymnesiaceae
OTU S262	2	0.00	2	0.003	Only DCM	Only pico	Clade HAP3-4	Clade HAP4
OTU S291	2	0.00	2	0.003	Only DCM	Both	Clade HAP3-4	Clade HAP3

Table S2b. Haptophyte D1-D2 28S rRNA OTUs recorded in the Skagerrak in August 2013. Red OTUs were removed after subsampling. Taxonomic assignations are based on phylogenetic placement.

OTU ID	Total reads (N)	Total reads (%)	Total reads after subsampling (N)	Total reads after subsampling (%)	Depth	Size fraction	Group	Lowest taxonomic level possible to determine
OTU L001	1980	6.66	740.00	5.87	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L002	1830	6.15	685.00	5.43	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L003	1555	5.23	679.00	5.38	Both	Both	Phaeocystaceae	Phaeocystis
OTU L004	957	3.22	494.00	3.92	Both	Both	Chrysochromulinaceae	Chrysochromulina acantha
OTU L005	862	2.90	365.00	2.89	Both	Both	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L006	782	2.63	339.00	2.69	Both	Both	Prymnesiaceae	*Haptolina*
OTU L007	710	2.39	318.00	2.52	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L008	676	2.27	299.00	2.37	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii/ C. campanulifera*
OTU L009	659	2.22	319.00	2.53	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L010	648	2.18	269.00	2.13	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L011	645	2.17	224.00	1.78	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L012	597	2.01	232.00	1.84	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L013	576	1.94	257.00	2.04	Both	Both	Prymnesiaceae	*Prymnesium polylepis*
OTU L014	560	1.88	257.00	2.04	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L015	538	1.81	234.00	1.85	Both	Both	Prymnesiaceae	*Prymnesium kappa*
OTU L016	516	1.73	234.00	1.85	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L017	491	1.65	198.00	1.57	Both	Both	Prymnesiaceae	*Prymnesium*
OTU L018	468	1.57	160.00	1.27	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L019	467	1.57	188.00	1.49	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L020	426	1.43	180.00	1.43	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L021	412	1.38	186.00	1.47	Both	Both	Prymnesiaceae	*Haptolina*

OTU L022	406	1.36	174.00	1.38	Both	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU L023	367	1.23	137.00	1.09	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L024	355	1.19	178.00	1.41	Both	Both	Coccolithales	Coccolithaceae
OTU L025	352	1.18	155.00	1.23	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L026	351	1.18	130.00	1.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L027	339	1.14	140.00	1.11	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L028	327	1.10	157.00	1.24	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L029	305	1.03	117.00	0.93	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L030	291	0.98	133.00	1.05	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L031	289	0.97	140.00	1.11	Both	Both	Prymnesiaceae	*Haptolina*
OTU L032	286	0.96	124.00	0.98	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L033	274	0.92	127.00	1.01	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L034	268	0.90	125.00	0.99	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L035	265	0.89	93.00	0.74	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L036	260	0.87	118.00	0.94	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L037	254	0.85	116.00	0.92	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L038	238	0.80	102.00	0.81	Both	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L039	238	0.80	101.00	0.80	Both	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L040	235	0.79	84.00	0.67	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L041	234	0.79	87.00	0.69	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L042	234	0.79	95.00	0.75	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L043	224	0.75	89.00	0.71	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L044	223	0.75	110.00	0.87	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L045	219	0.74	94.00	0.75	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L046	213	0.72	101.00	0.80	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L047	207	0.70	77.00	0.61	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L048	205	0.69	91.00	0.72	Both	Both	Chrysochromulinaceae	*Chrysochromulina*

OTU L049	195	0.66	80.00	0.63	Both	Both	Prymnesiaceae	*Prymnesium neolepis*
OTU L050	192	0.65	100.00	0.79	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L051	189	0.64	74.00	0.59	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L052	163	0.55	61.00	0.48	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L053	141	0.47	63.00	0.50	Both	Both	Calyptrosphaeraceae	*Calyptrosphaera sphaeroidea*
OTU L054	139	0.47	64.00	0.51	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L055	133	0.45	55.00	0.44	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L056	118	0.40	57.00	0.45	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L057	118	0.40	55.00	0.44	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina rotalis*
OTU L058	105	0.35	42.00	0.33	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L059	105	0.35	41.00	0.32	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L060	100	0.34	48.00	0.38	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L061	99	0.33	51.00	0.40	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L062	95	0.32	32.00	0.25	Both	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L063	95	0.32	53.00	0.42	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L064	92	0.31	32.00	0.25	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L065	91	0.31	30.00	0.24	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L066	85	0.29	32.00	0.25	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L067	79	0.27	32.00	0.25	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L068	78	0.26	41.00	0.32	Both	Both	Coccolithales	Coccolithales
OTU L069	76	0.26	27.00	0.21	Both	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L070	76	0.26	22.00	0.17	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L071	75	0.25	36.00	0.29	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L072	72	0.24	38.00	0.30	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L073	71	0.24	26.00	0.21	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L074	68	0.23	25.00	0.20	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L075	67	0.23	31.00	0.25	Both	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*

OTU L076	62	0.21	33.00	0.26	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L077	61	0.21	26.00	0.21	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L078	58	0.19	26.00	0.21	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L079	57	0.19	27.00	0.21	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L080	57	0.19	28.00	0.22	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L081	51	0.17	25.00	0.20	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L082	49	0.16	19.00	0.15	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L083	48	0.16	21.00	0.17	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L084	48	0.16	27.00	0.21	Both	Both	Prymnesiaceae	*Prymnesium*
OTU L085	47	0.16	18.00	0.14	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L086	42	0.14	18.00	0.14	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L087	42	0.14	20.00	0.16	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L088	40	0.13	19.00	0.15	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L089	39	0.13	14.00	0.11	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L090	38	0.13	15.00	0.12	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L091	37	0.12	16.00	0.13	Both	Only pico	Prymnesiaceae	*Prymnesium*
OTU L092	37	0.12	14.00	0.11	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L093	36	0.12	18.00	0.14	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L094	33	0.11	11.00	0.09	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L095	33	0.11	19.00	0.15	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L096	32	0.11	11.00	0.09	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L097	32	0.11	13.00	0.10	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L098	31	0.10	11.00	0.09	Only DCM	Only nano	Syracosphaeraceae	*Coronosphaera mediterranea*
OTU L099	31	0.10	14.00	0.11	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L100	29	0.10	7.00	0.06	Only DCM	Both	Prymnesiaceae	*Prymnesium polylepis*
OTU L101	29	0.10	12.00	0.10	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L102	28	0.09	9.00	0.07	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina simplex*

OTU L103	27	0.09	10.00	0.08	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L104	26	0.09	8.00	0.06	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L105	26	0.09	13.00	0.10	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L106	25	0.08	10.00	0.08	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L108	20	0.07	9.00	0.07	Both	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU L109	20	0.07	6.00	0.05	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L110	19	0.06	10.00	0.08	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L111	19	0.06	8.00	0.06	Both	Both	Prymnesiaceae	*Prymnesium*
OTU L112	18	0.06	10.00	0.08	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L113	18	0.06	11.00	0.09	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L114	18	0.06	9.00	0.07	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L115	18	0.06	6.00	0.05	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L116	17	0.06	7.00	0.06	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L117	17	0.06	9.00	0.07	Both	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU L118	17	0.06	10.00	0.08	Both	Both	Coccolithales	Coccolithales
OTU L119	16	0.05	8.00	0.06	Both	Both	Chrysochromulinaceae	*Chrysochromulina leadbeateri*
OTU L120	15	0.05	5.00	0.04	Both	Both	Prymnesiaceae	*Haptolina brevifila*
OTU L121	15	0.05	3.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina leadbeateri*
OTU L122	14	0.05	14.00	0.11	Only 1m	Only nano	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L123	14	0.05	7.00	0.06	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L124	14	0.05	6.00	0.05	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L125	13	0.04	8.00	0.06	Both	Both	Prymnesiaceae	*Prymnesium*
OTU L126	13	0.04	7.00	0.06	Only 1m	Only nano	Prymnesiophyceae	Zygodiscales
OTU L127	13	0.04	5.00	0.04	Both	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L128	13	0.04	5.00	0.04	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L129	13	0.04	5.00	0.04	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L130	12	0.04	4.00	0.03	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)

OTU L131	12	0.04	12.00	0.10	Only 1m	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU L132	12	0.04	3.00	0.02	Only 1m	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU L133	12	0.04	3.00	0.02	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L134	12	0.04	7.00	0.06	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L135	11	0.04	7.00	0.06	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L136	11	0.04	6.00	0.05	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L137	11	0.04	5.00	0.04	Only DCM	Only nano	Coccolithales	*Coccolithus braarudii*
OTU L138	11	0.04	4.00	0.03	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L139	11	0.04	5.00	0.04	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L140	11	0.04	6.00	0.05	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L141	11	0.04	4.00	0.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L142	11	0.04	3.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L143	11	0.04	6.00	0.05	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L144	11	0.04	5.00	0.04	Both	Both	Prymnesiaceae	*Prymnesium*
OTU L145	11	0.04	5.00	0.04	Both	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU L146	10	0.03	7.00	0.06	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L147	10	0.03	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L148	10	0.03	4.00	0.03	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L149	10	0.03	3.00	0.02	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L150	10	0.03	6.00	0.05	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L151	10	0.03	3.00	0.02	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L152	10	0.03	5.00	0.04	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L153	10	0.03	4.00	0.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L154	9	0.03	8.00	0.06	Only 1m	Both	Calcihaptophycidae	Calcihaptophycidae incertae sedis
OTU L155	9	0.03	2.00	0.02	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L156	9	0.03	5.00	0.04	Only 1m	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L157	9	0.03	3.00	0.02	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)

OTU L158	9	0.03	3.00	0.02	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L159	9	0.03	2.00	0.02	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L160	9	0.03	4.00	0.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L161	9	0.03	3.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L162	9	0.03	3.00	0.02	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L163	9	0.03	4.00	0.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L164	9	0.03	0.00	0.00	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L165	9	0.03	6.00	0.05	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L166	9	0.03	3.00	0.02	Both	Only pico	Prymnesiaceae	*Prymnesium*
OTU L167	8	0.03	3.00	0.02	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L168	8	0.03	4.00	0.03	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L169	8	0.03	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L170	8	0.03	1.00	0.01	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L171	8	0.03	1.00	0.01	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L172	8	0.03	1.00	0.01	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L173	8	0.03	1.00	0.01	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L174	8	0.03	4.00	0.03	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii/ C. Campanulifera*
OTU L175	8	0.03	2.00	0.02	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L176	7	0.02	3.00	0.02	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L177	7	0.02	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L178	7	0.02	4.00	0.03	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L179	7	0.02	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L180	7	0.02	1.00	0.01	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L181	7	0.02	6.00	0.05	Both	Both	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L182	7	0.02	2.00	0.02	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L183	7	0.02	5.00	0.04	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*

OTU L184	7	0.02	3.00	0.02	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L185	7	0.02	4.00	0.03	Only 1m	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU L186	7	0.02	3.00	0.02	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L187	7	0.02	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L188	7	0.02	6.00	0.05	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L189	6	0.02	5.00	0.04	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L190	6	0.02	4.00	0.03	Only 1m	Only pico	Prymnesiaceae	*Prymnesium polylepis*
OTU L191	6	0.02	3.00	0.02	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L192	6	0.02	1.00	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L193	6	0.02	1.00	0.01	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L194	6	0.02	0.00	0.00	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L195	6	0.02	4.00	0.03	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L196	6	0.02	3.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L197	6	0.02	2.00	0.02	Both	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L198	6	0.02	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L199	6	0.02	3.00	0.02	Only DCM	Both	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L200	6	0.02	5.00	0.04	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L201	6	0.02	4.00	0.03	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L202	6	0.02	4.00	0.03	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L203	6	0.02	2.00	0.02	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L204	6	0.02	2.00	0.02	Only 1m	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L205	6	0.02	4.00	0.03	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L206	6	0.02	2.00	0.02	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L207	6	0.02	3.00	0.02	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L208	6	0.02	5.00	0.04	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L209	6	0.02	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L210	6	0.02	5.00	0.04	Only DCM	Both	Prymnesiaceae	*Prymnesium*

OTU L211	6	0.02	2.00	0.02	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L212	6	0.02	1.00	0.01	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU1
OTU L213	5	0.02	1.00	0.01	Only DCM	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L214	5	0.02	1.00	0.01	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L215	5	0.02	3.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L216	5	0.02	1.00	0.01	Both	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L217	5	0.02	2.00	0.02	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L218	5	0.02	2.00	0.02	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU L219	5	0.02	4.00	0.03	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L220	5	0.02	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L221	5	0.02	3.00	0.02	Both	Both	Prymnesiaceae	*Haptolina*
OTU L222	5	0.02	3.00	0.02	Only 1m	Only nano	Coccolithales	*Chrysotila stipitata*
OTU L223	5	0.02	2.00	0.02	Only 1m	Both	Prymnesiaceae	*Dicrateria rotunda*
OTU L224	5	0.02	5.00	0.04	Only 1m	Only nano	Coccolithales	*Umbilicosphaera*
OTU L225	5	0.02	4.00	0.03	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L226	5	0.02	3.00	0.02	Only 1m	Only nano	Prymnesiophyceae	Clade PRY-LSU1
OTU L227	5	0.02	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L228	5	0.02	2.00	0.02	Both	Only pico	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L229	5	0.02	2.00	0.02	Both	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L230	5	0.02	1.00	0.01	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L231	5	0.02	4.00	0.03	Both	Only nano	Prymnesiaceae	*Prymnesium*
OTU L232	5	0.02	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L233	5	0.02	3.00	0.02	Only 1m	Only nano	Calyptrosphaeraceae	*Calyptrosphaera sphaeroidea*
OTU L234	5	0.02	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L235	5	0.02	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L236	5	0.02	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L237	5	0.02	3.00	0.02	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)

OTU L238	5	0.02	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina leadbeateri*
OTU L239	5	0.02	1.00	0.01	Both	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L240	5	0.02	3.00	0.02	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L241	5	0.02	2.00	0.02	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L242	5	0.02	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina*
OTU L243	5	0.02	3.00	0.02	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L244	5	0.02	1.00	0.01	Both	Only nano	Phaeocystaceae	*Phaeocystis*
OTU L245	5	0.02	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L246	5	0.02	3.00	0.02	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L247	5	0.02	3.00	0.02	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L248	5	0.02	5.00	0.04	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L249	5	0.02	4.00	0.03	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L250	5	0.02	4.00	0.03	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L251	4	0.01	2.00	0.02	Only 1m	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L252	4	0.01	3.00	0.02	Only DCM	Only nano	Prymnesiaceae	*Haptolina ericina*
OTU L253	4	0.01	2.00	0.02	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina rotalis*
OTU L254	4	0.01	1.00	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L255	4	0.01	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L256	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L257	4	0.01	2.00	0.02	Both	Both	Prymnesiaceae	*Haptolina*
OTU L258	4	0.01	1.00	0.01	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L259	4	0.01	2.00	0.02	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L260	4	0.01	1.00	0.01	Both	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L261	4	0.01	1.00	0.01	Only 1m	Both	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L262	4	0.01	1.00	0.01	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L263	4	0.01	0.00	0.00	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L264	4	0.01	2.00	0.02	Only 1m	Both	Prymnesiaceae	*Haptolina*

OTU L265	4	0.01	1.00	0.01	Only 1m	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L266	4	0.01	0.00	0.00	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L267	4	0.01	2.00	0.02	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L268	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU L269	4	0.01	1.00	0.01	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L270	4	0.01	1.00	0.01	Only DCM	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L271	4	0.01	4.00	0.03	Only 1m	Only nano	Prymnesiaceae	*Haptolina ericina*
OTU L272	4	0.01	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L273	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU L274	4	0.01	1.00	0.01	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L275	4	0.01	2.00	0.02	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L276	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L277	4	0.01	1.00	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L278	4	0.01	0.00	0.00	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L279	4	0.01	2.00	0.02	Only 1m	Both	Prymnesiaceae	*Haptolina*
OTU L280	4	0.01	2.00	0.02	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L281	4	0.01	0.00	0.00	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L282	4	0.01	2.00	0.02	Only DCM	Only nano	Prymnesiaceae	*Prymnesium kappa*
OTU L283	4	0.01	3.00	0.02	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L284	4	0.01	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L285	4	0.01	0.00	0.00	Only 1m	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L286	4	0.01	3.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L287	4	0.01	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L288	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L289	4	0.01	1.00	0.01	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L290	4	0.01	3.00	0.02	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L291	4	0.01	3.00	0.02	Both	Both	Phaeocystaceae	*Phaeocystis*

OTU L292	4	0.01	1.00	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L293	4	0.01	0.00	0.00	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L294	4	0.01	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L295	4	0.01	1.00	0.01	Both	Only nano	Prymnesiaceae	*Dicrateria rotunda*
OTU L296	4	0.01	3.00	0.02	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L297	4	0.01	2.00	0.02	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L298	4	0.01	4.00	0.03	Both	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L299	4	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L300	4	0.01	2.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L302	4	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L303	4	0.01	4.00	0.03	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina throndsenii*
OTU L304	4	0.01	2.00	0.02	Both	Only pico	Prymnesiaceae	*Haptolina brevifila*
OTU L305	4	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L306	4	0.01	3.00	0.02	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU1
OTU L307	4	0.01	2.00	0.02	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU1
OTU L308	3	0.01	1.00	0.01	Only DCM	Both	Prymnesiaceae	*Haptolina*
OTU L309	3	0.01	1.00	0.01	Only 1m	Only nano	Noelaerhabdaceae	*Emiliania huxleyi*
OTU L310	3	0.01	1.00	0.01	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L311	3	0.01	0.00	0.00	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L312	3	0.01	3.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L313	3	0.01	0.00	0.00	Only DCM	Only pico	Prymnesiaceae	*Haptolina brevifila*
OTU L314	3	0.01	2.00	0.02	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L315	3	0.01	2.00	0.02	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L316	3	0.01	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L317	3	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L318	3	0.01	0.00	0.00	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L319	3	0.01	1.00	0.01	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)

OTU L320	3	0.01	2.00	0.02	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L321	3	0.01	3.00	0.02	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L322	3	0.01	2.00	0.02	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L323	3	0.01	2.00	0.02	Only DCM	Only pico	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L324	3	0.01	3.00	0.02	Only 1m	Only nano	Prymnesiophyceae	Clade PRY-LSU1
OTU L325	3	0.01	2.00	0.02	Both	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L326	3	0.01	1.00	0.01	Only 1m	Only nano	Prymnesiaceae	*Prymnesium*
OTU L327	3	0.01	0.00	0.00	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L328	3	0.01	3.00	0.02	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L329	3	0.01	0.00	0.00	Only DCM	Only pico	Prymnesiophyceae	Clade PRY-LSU1
OTU L330	3	0.01	1.00	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L331	3	0.01	1.00	0.01	Only DCM	Both	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L332	3	0.01	0.00	0.00	Only 1m	Only nano	Prymnesiaceae	*Haptolina*
OTU L333	3	0.01	1.00	0.01	Only DCM	Both	Coccolithales	Coccolithaceae
OTU L334	3	0.01	1.00	0.01	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L335	3	0.01	1.00	0.01	Only DCM	Only pico	Prymnesiaceae	*Dicrateria rotunda*
OTU L336	3	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L337	3	0.01	2.00	0.02	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L338	3	0.01	0.00	0.00	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L339	3	0.01	2.00	0.02	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU L340	3	0.01	1.00	0.01	Only DCM	Only pico	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L341	3	0.01	1.00	0.01	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L342	3	0.01	1.00	0.01	Only 1m	Only nano	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L343	3	0.01	1.00	0.01	Both	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L344	3	0.01	3.00	0.02	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L345	3	0.01	2.00	0.02	Only DCM	Only pico	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L346	3	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Dicrateria rotunda*

OTU L347	3	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Haptolina ericina*
OTU L348	3	0.01	0.00	0.00	Only DCM	Only pico	Prymnesiaceae	*Dicrateria rotunda*
OTU L349	3	0.01	2.00	0.02	Only DCM	Only pico	Prymnesiaceae	*Prymnesium*
OTU L350	3	0.01	1.00	0.01	Only DCM	Only nano	Calyptrosphaeraceae	*Calyptrosphaera sphaeroidea*
OTU L351	3	0.01	1.00	0.01	Both	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L352	3	0.01	0.00	0.00	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L353	3	0.01	2.00	0.02	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L354	3	0.01	1.00	0.01	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU L355	3	0.01	3.00	0.02	Only 1m	Only nano	Chrysochromulinaceae	*Chrysochromulina*
OTU L356	3	0.01	3.00	0.02	Only 1m	Only nano	Prymnesiaceae	*Prymnesium*
OTU L357	3	0.01	1.00	0.01	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L358	3	0.01	2.00	0.02	Only 1m	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L359	3	0.01	2.00	0.02	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L360	3	0.01	1.00	0.01	Only 1m	Only nano	Prymnesiaceae	*Prymnesium*
OTU L362	3	0.01	0.00	0.00	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L363	3	0.01	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L364	3	0.01	0.00	0.00	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L365	3	0.01	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L366	3	0.01	2.00	0.02	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L367	3	0.01	1.00	0.01	Only DCM	Only nano	Phaeocystaceae	*Phaeocystis*
OTU L368	3	0.01	2.00	0.02	Only DCM	Only nano	Phaeocystaceae	*Phaeocystis*
OTU L369	3	0.01	1.00	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L370	3	0.01	2.00	0.02	Only DCM	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L371	3	0.01	1.00	0.01	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L372	3	0.01	3.00	0.02	Only 1m	Both	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L373	3	0.01	3.00	0.02	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L374	3	0.01	1.00	0.01	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*

OTU L375	3	0.01	2.00	0.02	Only 1m	Only nano	Prymnesiaceae	*Haptolina brevifila*
OTU L376	3	0.01	0.00	0.00	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L377	3	0.01	2.00	0.02	Only DCM	Only pico	Helicosphaeraceae	Helicosphaera
OTU L378	3	0.01	1.00	0.01	Only DCM	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L379	3	0.01	0.00	0.00	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L380	3	0.01	1.00	0.01	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L381	3	0.01	1.00	0.01	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L382	3	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU L383	3	0.01	0.00	0.00	Only 1m	Only pico	Prymnesiaceae	*Prymnesium*
OTU L384	3	0.01	2.00	0.02	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L385	3	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L386	3	0.01	0.00	0.00	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L387	3	0.01	0.00	0.00	Both	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L388	3	0.01	0.00	0.00	Only 1m	Only pico	Prymnesiaceae	*Dicrateria rotunda*
OTU L389	3	0.01	0.00	0.00	Only DCM	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L390	3	0.01	1.00	0.01	Both	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L392	2	0.01	0.00	0.00	Only DCM	Both	Rhabdosphaeraceae	*Algirosphaera robusta*
OTU L397	2	0.01	0.00	0.00	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L401	2	0.01	2.00	0.02	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L411	2	0.01	1.00	0.01	Both	Only nano	Prymnesiaceae	*Prymnesium*
OTU L417	2	0.01	1.00	0.01	Only 1m	Only pico	Phaeocystaceae	*Phaeocystis*
OTU L418	2	0.01	1.00	0.01	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L422	2	0.01	2.00	0.02	Both	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L426	2	0.01	0.00	0.00	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L427	2	0.01	2.00	0.02	Only 1m	Only nano	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L431	2	0.01	2.00	0.02	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L434	2	0.01	0.00	0.00	Only DCM	Both	Rhabdosphaeraceae	*Algirosphaera robusta*

OTU L435	2	0.01	0.00	0.00	Both	Only pico	Prymnesiaceae	*Prymnesium kappa*
OTU L437	2	0.01	2.00	0.02	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina simplex*
OTU L440	2	0.01	0.00	0.00	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L443	2	0.01	0.00	0.00	Both	Only pico	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L445	2	0.01	1.00	0.01	Only 1m	Only pico	Prymnesiophyceae	Clade PRY-LSU3 (Clade E-F?)
OTU L451	2	0.01	1.00	0.01	Only 1m	Only pico	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L461	2	0.01	0.00	0.00	Only DCM	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L462	2	0.01	1.00	0.01	Only DCM	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L468	2	0.01	1.00	0.01	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L474	2	0.01	1.00	0.01	Only 1m	Both	Prymnesiaceae	*Prymnesium*
OTU L482	2	0.01	2.00	0.02	Only DCM	Only nano	Prymnesiaceae	*Prymnesium polylepis*
OTU L483	2	0.01	0.00	0.00	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L484	2	0.01	0.00	0.00	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU1
OTU L485	2	0.01	1.00	0.01	Both	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L487	2	0.01	1.00	0.01	Both	Both	Chrysochromulinaceae	*Chrysochromulina acantha*
OTU L488	2	0.01	1.00	0.01	Both	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L490	2	0.01	0.00	0.00	Only 1m	Both	Calcihaptophycidae	Calcihaptophycidae incertae sedis
OTU L492	2	0.01	2.00	0.02	Both	Only pico	Chrysochromulinaceae	*Chrysochromulina*
OTU L498	2	0.01	0.00	0.00	Both	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)
OTU L499	2	0.01	1.00	0.01	Only 1m	Both	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L500	2	0.01	0.00	0.00	Only DCM	Only nano	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L504	2	0.01	1.00	0.01	Both	Both	Phaeocystaceae	*Phaeocystis*
OTU L508	2	0.01	1.00	0.01	Both	Only pico	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L511	2	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Prymnesium*
OTU L512	2	0.01	1.00	0.01	Only 1m	Both	Prymnesiaceae	*Haptolina*
OTU L514	2	0.01	1.00	0.01	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina camella*
OTU L528	2	0.01	0.00	0.00	Only DCM	Both	Prymnesiophyceae	Clade PRY-LSU2 (Clade D?)

OTU L530	2	0.01	1.00	0.01	Only DCM	Both	Prymnesiaceae	*Prymnesium*
OTU L532	2	0.01	2.00	0.02	Only DCM	Only nano	Prymnesiophyceae	Clade PRY-LSU4 (Clade B3-4-5?)
OTU L538	2	0.01	1.00	0.01	Only DCM	Only nano	Syracosphaeraceae	*Syracosphaera pulchra*
OTU L539	2	0.01	1.00	0.01	Only DCM	Only nano	Prymnesiaceae	*Haptolina ericina/hirta /fragaria*
OTU L554	2	0.01	1.00	0.01	Only 1m	Both	Chrysochromulinaceae	*Chrysochromulina*
OTU L555	2	0.01	0.00	0.00	Only DCM	Both	Phaeocystaceae	*Phaeocystis*
OTU L557	2	0.01	0.00	0.00	Only DCM	Both	Phaeocystaceae	*Phaeocystis*

Table S3. Total and proportional read abundances and OTUs within each major clade for 18S and 28S rRNA genes.

	18S rRNA				28S rRNA			
Family/Clade	**Total reads (N)**	**Total reads (%)**	**Total OTUs (N)**	**Total OTUs (%)**	**Total reads (N)**	**Total reads (%)**	**Total OTUs (N)**	**Total OTUs (%)**
Chrysochromulinaceae	34586	33.1	52	24.2	10039	33.7	134	31.0
Noelaerhabdaceae	18869	18.1	4	1.9	121	0.4	6	1.4
Prymnesiaceae	15262	14.6	44	20.5	5047	17.0	91	21.1
Syracosphaeraceae	12648	12.1	2	0.9	5172	17.4	38	8.8
Phaeocystaceae	9630	9.2	27	12.6	3415	11.5	49	11.3
Prymnesiales Clade B3-B4-B5	5667	5.4	14	6.5	0	0.0	0	0.0
Prymnesiophyceae_sp.	3187	3.1	18	8.4	4526	15.2	94	21.8
Calyptrosphaeraceae	1378	1.3	1	0.5	149	0.5	3	0.7
Haptophyta_sp.	805	0.8	17	7.9	0	0.0	0	0.0
Rhabdosphaeraceae	803	0.8	2	0.9	796	2.7	8	1.9
Calcihaptophycidae	750	0.7	8	3.7	9	0.0	1	0.2
Braarudosphaeraceae	310	0.3	2	0.9	0	0.0	0	0.0
Clade D-E-F	192	0.2	5	2.3	0	0.0	0	0.0
Clade HAP3-4	163	0.2	16	7.4	0	0.0	0	0.0
Isochrysidaceae	71	0.1	1	0.5	0	0.0	0	0.0
Coccolithales	24	0.0	2	0.9	474	1.6	7	1.6
Helicosphaeraceae	0	0.0	0	0.0	3	0.0	1	0.2
TOTAL	**104345**		**215**		**29751**		**432**	

Table S4. **a)** Overview over matching of 18S OTUs to other databases. Total: Total number of OTUs in each group. ≥ 99% any sequence: Number of OTUs that have ≥ 99% BLAST match with either any sequence in the Haptophyta-PiP database, or an OTU from Oslofjorden from Egge et al. 2015a. ≥ 99% Hapto-PiP_ENV: Number of OTUs that have ≥ 99% BLAST match with an "environmental sequence" in the Haptophyta-PiP database. ≥ 99% Hapto-PiP_CULT: Number of OTUs that have ≥ 99% BLAST match with a sequence from a cultured species in the Haptophyta-PiP database. ≥ 99% OF-OTUs: Number of OTUs that have ≥ 99% BLAST match with an OTU previously obtained by HTS of samples from Oslofjorden (these may represent either cultured species, environmental sequences obtained by Sanger sequencing, or novel sequences from the Egge et al. 2015a study). ≥ 99% OF-OTU & < 99% with any Hapto-PiP sequence: Number of OTUs that have ≥ 99% match to an OTU from Egge et al. (2015 a,b), but at the same time is < 99% similar to any sequence present in Hapto-PiP. The numbers from ≥ 99% Hapto-PiP_ENV and ≥ 99% Hapto-PiP_CULT may not add up, because environmental sequences in the Haptophyta-PiP database may also come from species that exist in culture.
b) Overview over matching of 28S OTUs to the 28S haptophyta reference database, consisting of sequences from cultured strains.

a)

Group	Total	≥ 99% any sequence	≥ 99% Hapto-PiP_ENV	≥ 99% Hapto-PiP_CULT	≥ 99% OF-OTUs	≥ 99% OF-OTU & <99% with any Hapto-PiP sequence
Clade HAP3-4	16	7	5	NA	5	2
Clade D-E-F	5	2	2	NA	2	0
Phaeocystaceae	27	8	7	2	6	1
Noelaerhabdaceae	4	1	1	1	1	0
Isochrysidaceae + Clade EV	4	2	1	1	2	1
Rhabdosphaeraceae	2	1	1	1	1	0
Coccolithales	2	1	1	1	1	0
Calyptrosphaeraceae	1	1	0	1	1	0
Syracosphaeraceae	2	1	1	1	1	0
Helicosphaeraceae	1	0	0	0	0	0
Calcihaptophycidae *incertae sedis*	5	2	0	1	1	1
Braarudosphaeraceae	2	1	1	0	1	0
Prymnesiales Clade B3-B4-B5	14	7	3	NA	7	4
Prymnesiaceae	44	17	8	7	16	9
Chrysochromulinaceae	52	18	13	4	18	5

Prymnesiophyceae sp.	17	3	1	0	3	2
Haptophyta sp.	17	3	2	0	2	1
Sum	215	75	47	20	68	26

b)

Group	**Total**	**≥ 97% sequence from cultured strain**
Phaeocystaceae	49	9
Noelaerhabdaceae	6	4
Rhabdosphaeraceae	8	1
Coccolithales	7	1
Calyptrosphaeraceae	3	1
Syracosphaeraceae	38	1
Helicosphaeraceae	1	0
Calcihaptophycidae *incertae sedis*	2	1
Prymnesiaceae	91	18
Chrysochromulinaceae	134	10
Prymnesiophyceae_sp.	93	1
Sum	432	47

File S1. Description of how the haptophyte 28S reference database was created.

The 28S D1-D2 reference database was created based on the 28S database from Bittner et al. 2013. All the sequences from cultured species were selected. In addition to these sequences, we searched GenBank for haptophyte 28S sequences using the following custom script in Biopython (Cock et al. 2009):

```
import sys
from Bio import Entrez
Entrez.email = "your@email.here"

#Search nucleotide database using Entrez for entries with these terms
handle = Entrez.esearch(db="nucleotide", term="(Haptophyceae OR Haptophyta)
AND (large subunit ribosomal RNA OR 28s OR 23s) NOT WGS[keyword]", usehis-
tory="y", retmax=5000)
result = Entrez.read(handle)

#Get genbank-entries for ids in "IdList"
net_handle = Entrez.efetch(db="nucleotide", rettype="gb", id=re-
sult["IdList"], retmode="text")
data = net_handle.read()
net_handle.close()
out_handle = open("haptosLSUor28Sor23s_genbank.txt", "w")
out_handle.write(data)
out_handle.close()
```

The genbak-file was parsed into a table with the following script:

```
#!/usr/bin/env python

import sys
from Bio import SeqIO
from Bio.Blast import NCBIXML
#Usage: $python gbparse_euk.py outfile.txt infile.gb
OUT = open(sys.argv[1], 'w')
OUT.write("Accno\tLength\tOrganism\tTaxonomy\tStrain\tIsolate\tCulture_col-
lection\tVoucher\tIsolation_source\tClone\tOrganelle\tGene\tProduct\tMole-
cule_type\tNote\tDB_xref\tPrimers\tCountry\tYear_submitted\tAuthor\tKey-
word\tSequence\n")
result_handle = open(sys.argv[2])
gbfiles = SeqIO.parse(result_handle, 'gb')
for rec in gbfiles:
    acc = rec.id
    sequence = str(rec.seq)
    length = str(len(rec.seq))
    recfeat1 = rec.features[1]
    source = rec.features[0]
    if 'organism' in rec.annotations:
        organism = rec.annotations['organism']
    if 'taxonomy' in rec.annotations:
        taxonomy = "_".join(rec.annotations['taxonomy'])
    if 'clone' in source.qualifiers:
        clone = source.qualifiers['clone'][0]
    else:
        clone = ""
    if 'source' in rec.annotations:
        gensource = rec.annotations['source']
    else:
        gensource = ""
    if 'isolation_source' in source.qualifiers:
        isolation_source = source.qualifiers['isolation_source'][0]
    else:
        isolation_source = ""
```

```
    if 'country' in source.qualifiers:
        country = source.qualifiers['country'][0]
    else:
        country = ""
    if 'strain' in source.qualifiers:
        strain = source.qualifiers['strain'][0]
    else:
        strain = ""
    if 'isolate' in source.qualifiers:
        isolate = source.qualifiers['isolate'][0]
    else:
        isolate = ""
    if 'culture_collection' in source.qualifiers:
        cultcol = source.qualifiers['culture_collection'][0]
    else:
        cultcol = ""
    if 'specimen_voucher' in source.qualifiers:
        voucher = source.qualifiers['specimen_voucher'][0]
    else:
        voucher = ""
    if 'PCR_primers' in source.qualifiers:
        primers = ".".join(source.qualifiers['PCR_primers'])
    else:
        primers = ""
    if 'gene' in recfeat1.qualifiers:
        whichgene = recfeat1.qualifiers['gene'][0]
    else:
        whichgene = ""
    if 'product' in recfeat1.qualifiers:
        product = recfeat1.qualifiers['product'][0]
    else:
        product=""
    if 'mol_type' in source.qualifiers:
        molecule_type = source.qualifiers['mol_type'][0]
    else:
        molecule_type = ""
    if 'organelle' in source.qualifiers:
        organelle = source.qualifiers['organelle'][0]
    else:
        organelle = ""
    if 'note' in source.qualifiers:
        note = source.qualifiers['note'][0]
    else:
        note = ""
    if 'db_xref' in source.qualifiers:
        db_xref=source.qualifiers['db_xref'][0]
    else:
        db_xref=""
    if 'references' in rec.annotations:
        pubref = rec.annotations['references'][0]
        authors = pubref.authors
        firstaut = authors.split(",")[0]
    else:
        firstaut = ""
    if 'date' in rec.annotations:
        date = rec.annotations['date']
        submyear = rec.annotations['date'][7:11]
    if 'keywords' in rec.annotations:
        keyword = rec.annotations['keywords'][0]
    fields = [acc, length, organism, taxonomy, strain, isolate,
cultcol,voucher, isolation_source, clone, organelle, whichgene, product,
molecule_type, note, db_xref, primers, country, submyear, firstaut, keyword,
sequence]
    OUT.write("\t".join(fields)+ "\n")
OUT.close()
```

From this table we selected all sequences from cultured strains. These were aligned in Geneious 8.1.8, and sequences that did not cover the region spanned by the primer pair LSU1 were removed. The alignment was trimmed at the start of the forward end end of the reverse primer. Identical sequences were removed. This left 184 haptophyte ”LSU1” reference sequences from cultured strains.

The taxonomy of the sequences (from class to genus level) was determined according to (Edvardsen et al. 2016)

To create a reference alignment for the phylogenetic analyses, the reference sequences were aligned in MAFFT v.6 with the Q-INS-I method (Katoh & Toh 2008).

References:

Edvardsen B., Egge E. S. & Vaulot D. 2016. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review. *Perspectives in Phycology*.

Cock PA, Antao T, Chang JT, Bradman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B and de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422-1423

Katoh, K. & Toh, H. 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT- based framework. BMC Bioinformatics, 9:212. doi:10.1186/ 1471-2105-9-21

PAPER III

Harmful Algae 75 (2018) 105–117

Contents lists available at ScienceDirect

Harmful Algae

journal homepage: www.elsevier.com/locate/hal

Development of a qPCR assay to detect and quantify ichthyotoxic flagellates along the Norwegian coast, and the first Norwegian record of *Fibrocapsa japonica* (Raphidophyceae)

Anette Engesmo[a,*], David Strand[b], Sandra Gran-Stadniczeñko[c], Bente Edvardsen[c], Linda K. Medlin[d], Wenche Eikrem[a,c]

[a] Norwegian Institute for Water Research, Gaustadallèen 21, 0349, Oslo, Norway
[b] Norwegian Veterinary Institute, P.O. box 750 Sentrum, 0106, Oslo, Norway
[c] University of Oslo, Department of Biosciences, P. O. box 1066 Blindern, 0316, Oslo, Norway
[d] Marine Biological Association of the UK, The Citadel, Plymouth, Pl1 2PB, United Kingdom

ARTICLE INFO

Article history:
Received 3 January 2018
Received in revised form 18 April 2018
Accepted 18 April 2018

Keywords:
Karenia mikimotoi
Karlodinium veneficum
Heterosigma akashiwo
Molecular monitoring
rDNA
454 high throughput environmental sequencing
SEM

ABSTRACT

Blooms of ichthyotoxic microalgae pose a great challenge to the aquaculture industry world-wide, and there is a need for fast and specific methods for their detection and quantification in monitoring programs. In this study, quantitative real-time PCR (qPCR) assays for the detection and enumeration of three ichthyotoxic flagellates: the dinoflagellate *Karenia mikimotoi* (Miyake & Kominami ex Oda) Hansen & Moestrup and the two raphidophytes *Heterosigma akashiwo* (Hada) Hada ex Hara & Chihara and *Fibrocapsa japonica* Toriumi & Takano were developed. Further, a previously published qPCR assay for the dinoflagellate *Karlodinium veneficum* (Ballantine) Larsen was used. Monthly samples collected for three years (Aug 2009–Jun 2012) in outer Oslofjorden, Norway were analysed, and the results compared with light microscopy cell counts. The results indicate a higher sensitivity and a lower detection limit (down to 1 cell L^{-1}) for both qPCR assays. Qualitative and semi-quantitative results were further compared with those obtained by environmental 454 high throughput sequencing (HTS, metabarcoding) and scanning electron microscopy (SEM) examination from the same samplings. All four species were detected by qPCR and HTS and/or SEM in outer Oslofjorden (Aug 2009–Jun 2012); *Karlodinium veneficum* was present year-round, whereas *Karenia mikimotoi*, *Heterosigma akashiwo* and *Fibrocapsa japonica* appeared mainly during the autumn in all three years. This is the first observation of *Fibrocapsa japonica* in Norwegian coastal waters. This species has previously been recorded off the Swedish west coast and German Bight, which may suggest a northward dispersal.
© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

During the last few decades, significant attention has been paid to harmful algal bloom (HAB) events. Most coastal regions in the world are affected, and the number of described causative species and the toxins they produce are increasing (Tillmann et al., 2009). Over the past few decades HABs have increased in frequency (Anderson et al., 2012), which imposes financial constraints on the aquaculture industry. To reduce financial losses and improve seafood safety, most countries that trade in seafood have an algal monitoring system in place (Medlin, 2013).

The current standard method in monitoring of microalgae is based on the counting technique described by Utermöhl (1958), where a volume of water sample (usually 5–50 mL) is preserved with a fixative, such as Lugol's solution, and left to settle in a sedimentation chamber before enumeration in an inverted microscope. The accuracy of this method is dependent on several factors, such as the sampling procedure, the fixative chosen, and the taxonomic expertise of the researcher conducting the survey (Bott et al., 2010). Another factor is the morphology of the species of interest e.g. small size, lack of hard cell components, and fixative induced changes to the morphology can make many flagellates difficult to detect and enumerate correctly under a light microscope (LM). Recent investigations indicate that the species

* Corresponding authors at: Norwegian Institute for Water Research, Gaustadallèen 21, 0349, Oslo, Norway.
E-mail addresses: anette.engesmo@niva.no (A. Engesmo), david.strand@vetinst.no (D. Strand), s.g.stadniczenko@ibv.uio.no (S. Gran-Stadniczeñko), bente.edvardsen@ibv.uio.no (B. Edvardsen), lindli@MBA.ac.uk (L.K. Medlin), wenche.eikrem@niva.no (W. Eikrem).

https://doi.org/10.1016/j.hal.2018.04.007
1568-9883/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

diversity is larger than observed by microscopy for dinoflagellates (Nézan et al., 2014), haptophytes (Egge et al., 2015) and protists in general (de Vargas et al., 2015). Time- and financial restraints make it desirable to develop molecular methods to compliment LM cell counts in monitoring programs (Medlin, 2013). Implementing molecular methods in monitoring for certain ichthyotoxic species may lower detection limits, increase sensitivity and accuracy, and reduce both costs and processing time per sample. Several molecular techniques have been developed for the detection and quantification of microalgae, such as fluorescent in situ hybridization (FISH-probes) or microarrays with molecular probes and high throughput sequencing, but quantitative real-time PCR (qPCR) is currently considered the most advantageous for detection and quantification of a restricted number of target species (Ebenezer et al., 2012). The initial tasks of designing, testing, and validating qPCR primers and hydrolysis probes is a major effort, but once an assay is established, it is highly sensitive, specific, and cost-effective. It can also be applied to preserved environmental samples (Bott et al., 2010; Eckford-Soper and Daugbjerg, 2015a). Recently qPCR has been utilized in monitoring of ichthyotoxic *Prymnesium parvum* Carter in USA (Zamor et al., 2012) and several toxic species in New Zealand, e.g. *Alexandrium catenella* (Whedon & Kofoid) Balech (Rhodes et al., 2013; Smith et al., 2014), and the results are promising. One of the drawbacks of qPCR is that it only detects targets actively being searched for. Consequently, untargeted and invasive species will go unnoticed, and it will not give information about the phytoplankton community as a whole. There are other molecular techniques, viz., microarrays and environmental high throughput sequencing (HTS) of marker genes, also called metabarcoding (Dittami et al., 2013; Kegel et al., 2016, de Vargas et al., 2015), which used in combination with qPCR could facilitate and improve monitoring in the future.

The present study focused on four ichthyotoxic flagellates, known- or suspected to occur in Norwegian coastal waters: *Karenia mikimotoi, Karlodinium veneficum, Heterosigma akashiwo* and *Fibrocapsa japonica*. The two dinoflagellates, *K. mikimotoi* and *K. veneficum,* have formed recurrent blooms in the Oslofjorden (Throndsen et al., 2003). One of the first major algal blooms to cause public attention in Norway was caused by the raphidophyte, *H. akashiwo,* in 1964 (Braarud and Nygaard, 1967), and since then, it has been reported regularly (Naustvoll et al., 2002). The other targeted raphidophyte, *F. japonica,* has not previously been reported in Norwegian coastal waters. It has, however, been reported from several other European locations in the North Sea (Elbrächter, 1999) and from the Swedish west coast (www.smhi.se/klimatdata/oceanografi/havsmiljodata).

The aim of this study was to develop a rapid detection and enumeration method for these ichthyotoxic species, which can be utilized in algal monitoring as a compliment to LM. We further wanted to improve our knowledge about the seasonal distribution of ichthyotoxic species present in the Skagerrak.

2. Material and methods

2.1. Field sampling

Field sampling was carried out monthly using the University of Oslo's research vessel R/V Trygve Braarud over the course of three years (Aug 2009–Jun 2012) at station OF2 (59.18 N, 10.69 E) in outer Oslofjorden (Fig. 1). Water samples were collected from 1 m depth using Niskin bottles attached to a rosette and used for all samples described below. For LM cell counts, 100 mL samples were collected directly from the Niskin bottles and preserved with 1 mL neutral Lugol's solution (Throndsen, 1978). Samples for DNA-isolation and subsequent qPCR were collected in two ways: during

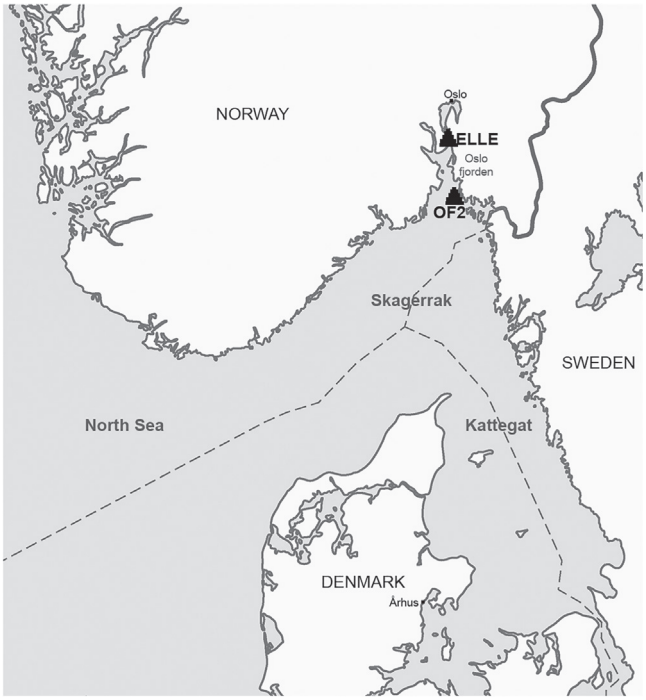


Fig. 1. Map of sampling area, with stations OF2 and ELLE in outer Oslofjorden marked.

the first two years (Aug 2009–Jun 2011) 20 L seawater were pre-filtered through a 180 μm mesh to remove large zooplankton, before being filtered by peristaltic pumping (Masterflex 07523-80, ColeParmer, IL, USA), on to 0.45 μm pore size, 142 mm diameter Durapore filters (Polyvinylidene fluoride (PVDF), Millipore, Billerica, MA, USA), placed in a Millipore stainless steel tripod. Filters were cut into four approximately equal pieces on board and frozen separately in liquid nitrogen. They were kept at −80 °C until further processing. In the last year (Aug 2011–Jun 2012), 1 L sea water samples were pre-filtered through a 180 μm mesh and then filtrated down on 25 mm nitrate cellulose filters (Sartorius-stedim, Göttingen, Germany) with 1.2 μm pore size and frozen directly in liquid nitrogen. The filters were kept at −80 °C until further processing.

2.2. Algal culturing

All cultures used in this study are listed in Table 1 and were obtained from the following culture collections: The Norwegian Culture Collection of Algae (NORCCA), Roscoff Culture Collection (RCC), CMS Algal Research Collection (ARC), National Institute for Environmental Studies (NIES), National Center for Marine Algae and Microbiota (CCMP), and Microalgae Culture Collection of the Department of Plant Biology and Ecology of the University of the Basque Country (EHU). Culture conditions are detailed in Table S1, and for media recipes, readers are referred to Andersen (2005).

2.3. DNA-isolation

All samples were defrosted on ice and sodium phosphate buffer (provided by the MPBio Fast DNA Spin Kit) was added before cell lysis. Filter disruption was performed with a Precellys 24 homogenizer for 2 × 15 s at 6000 rpm (Bertin, Montigny le Bretonneux, France). Two negative controls were employed to ensure no contamination took place during DNA-isolation, negative environmental control (NEC), and negative sample control (NSC). The NEC consisted of a tube with 200 μL molecular

Table 1
Algal strains used in this study, with results of specificity testing.

Strain	Species	Class	Isolator	Location	Specificity:			
					K. mik	*K. ven*	*H. aka*	*F. jap*
UIO019*	*Karenia mikimotoi*	Dinophyceae	K. Tangen	Oslofjorden, Norway	+	–	n/a	n/a
SCCAP K-1274	*Karenia brevis*	Dinophyceae	J. Rogers	Gulf of Mexico, FL, USA	–	–	n/a	n/a
UIO254*	*Karlodinium veneficum*	Dinophyceae	K. Tangen	Oslofjorden, Norway	–	+	n/a	n/a
UIO297	*Karlodinium* cf. *veneficum*	Dinophyceae	S. Ota	Oslofjorden, Norway	–	+	n/a	n/a
SCCAP K-1471	*Alexandrium tamarense*	Dinophyceae	A. Godhe	Lysekil, Sweden	–	–	n/a	n/a
SCCAP K-0675	*Levanderina fissa*	Dinophyceae	E. Silva	San Andre lagoon, Portugal	–	–	n/a	n/a
SCCAP K-1332	*Gymnodinium catenatum*	Dinophyceae	S. Ribeiro	Sines, Portugal	–	–	n/a	n/a
CCMP2088	*Polarella glacialis*	Dinophyceae	C. Lovejoy	Baffin Bay, Canada	–	–	–	–
UIO296	*Azadinium* cf. *Spinosum*	Dinophyceae	S. Ota	Oslofjorden, Norway	–	–	n/a	n/a
SCCAP K-1137	*Prorocentrum micans*	Dinophyceae	T. Berge	Flekkefjorden, Norway	–	–	n/a	n/a
UIO284	*Scrippsiella trochoidae*	Dinophyceae	S. Ota	Black Sea, Bulgaria	–	–	n/a	n/a
UIO081	*Amphidinium carterae*	Dinophyceae	K. Tangen	Unknown	–	–	n/a	n/a
RCC1502	*Heterosigma akashiwo*	Raphidophyceae	J. Fresnel	La Rochelle, France	n/a	n/a	+	–
ARC HA0309-2	*Heterosigma akashiwo*	Raphidophyceae	C. Tomas	Offats Bayou, TX, USA	n/a	n/a	+	–
SCCAP K-1549*	*Heterosigma akashiwo*	Raphidophyceae	G. Hansen	Århus harbour, Denmark	n/a	n/a	+	–
RP02EHU	*Heterosigma akashiwo*	Raphidophyceae	S. Seone	Bay of Biscay, Spain	–	–	+	–
RCC1501	*Fibrocapsa japonica*	Raphidophyceae	I. Probert	English Channel	n/a	n/a	–	+
SCCAP-K0542*	*Fibrocapsa japonica*	Raphidophyceae	J.Ø. Nielsen	North Sea, Germany	n/a	n/a	–	+
ARC CS0707-1	*Chattonella subsalsa*	Raphidophyceae	C. Tomas	Inland Bay, DE, USA	n/a	n/a	–	–
ARC CA0800	*Chattonella marina* var. *antiqua*	Raphidophyceae	S. Yoshimatsu	Kagawa Prefecture, Japan	n/a	n/a	–	–
ARC HD0110	*Haramonas dimorpha*	Raphidophyceae	S. Yoshimatsu	Kagawa Prefecture, Japan	n/a	n/a	–	–
NIES-15	*Olisthodiscus luteus*	Incerta sedis	I. Inouye	Seto Inland Sea, Japan	–	–	–	–
UIO109	*Pseudochattonella farcimen*	Dictyochophyceae	B. Edvardsen	Langesund, Norway	–	–	–	–

*Strains used for primer/probe design, specificity testing, qPCR standards and SEM examination.

grade water (Promega, Madison, WI, USA), which was left open on the bench-top during DNA-isolation. The NSC consisted of a sample where DNA was isolated from molecular grade water.

DNA-isolation from Durapore membranes and algal culture pellets were carried out using MPBio Fast DNA Spin Kit (MP Biomedicals, Santa Ana, USA), following the manufacturer's protocol with the following modifications: after step four, the supernatant was transferred to a fresh tube, and 200 μL of protein precipitation solution (PPS) were added, the sample was then gently mixed by hand before centrifugation (14,000 rpm for 5 min) in an Eppendorf centrifuge 5424 (Hamburg, Germany) to pellet the debris. In step five, the supernatant and binding matrix were mixed in a 15 mL tube to facilitate optimal binding of DNA. After the washing steps, filters were air dried for 5 min at room temperature before elution of DNA.

DNA-isolation from nitrocellulose filters was carried out with MPBio Fast DNA Spin Kit for Soil (MP Biomedicals), following the manufacturer's protocol, with one modification: an additional elution step was included where the samples were incubated for 5 min at 55 °C to increase yield.

2.4. Assay design

The DNA sequence of selected strains (Table 1) was obtained from cultured algal cells in the exponential growth phase. DNA was isolated as described in Section 2.3 before PCR, and sequencing were performed following the protocol in Engesmo et al. (2016). The generated sequences are available at GenBank with accession-numbers: *K. mikimotoi* strain UIO019: KU314866, *K. veneficum* strain UIO254: KU314867, *H. akashiwo* strain SCCAP K-1549: KP702879, KP702897 and *F. japonica* strain SCCAP K-0542: KU314865. DNA from the remaining strains listed in Table 1 was used for specificity testing of the qPCR assays.

Species-specific primers and hydrolysis probes were designed manually using Geneious version 7.1.7 (Biomatters, Auckland, New Zealand). Known raphydophyte and dinoflagellate sequences were imported from NCBI (https://www.ncbi.nlm.nih.gov/genbank/) and aligned with sequences generated in this study using MAFFT (Katoh et al., 2002) plugin for Geneious. The alignments were examined by eye and sequences compared to reveal intra-species genetic variation and select potential probe and primer sites. The specificity was then examined in silico by Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and the suitability of the sequence determined using OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) and OligoAnalyzer 3.1 available from Integrated DNA Technologies (https://eu.idtdna.com/calc/analyzer). All primers and probes designed for this study were located in the 28S rDNA (Table 2). The assay used for *K. veneficum* was published previously by Park et al. (2009) and targeted the ITS1 region. The probe was designed first and primers were designed on both sides to generate an amplicon of 50–150 base pairs (bp). When possible, one primer was placed as close as possible to the probe sequence, without overlap. The probe and primers were designed to be species-specific. Primers were synthesised by Eurofins Genomics (Ebersberg, Germany). All primers and probes are listed in Table 2.

Optimization of qPCR working conditions was established for each assay by running a temperature gradient of 10 °C, starting 3 °C below the lowest primers melting temperature (Tm), and testing four different primer concentrations (125, 250, 500 and 1000 nM) and three probe concentrations (75, 125 and 250 nM) on the dilution series used for standard curves. Assays were then tested for specificity against a matrix of relevant species (Table 1) to determine if they amplified only the desired target.

2.5. qPCR

All qPCR reactions were performed on a BioRad CFX96 or CFXTOUCH 96 (Bio-Rad, Hercules, CA, USA) using 96-well plates (blue plates with clear wells) sealed with transparent adhesive. All qPCR reactions consisted of 7.5 μL 2× TaqMan® environmental mastermix (Life Technologies, Carlsbad, CA, USA), 250–500 nM primers (Table 2), 125–200 nM probe (Table 2), 1 μL DNA template (Tables 3, S3), and molecular grade water to a final volume of 15 μL. Distribution of master mix and addition of template DNA was carried out using a Biomek 3000 pipetting robot (Beckman Coulter, Brea, CA, USA). The qPCR assays were run with the cycling conditions: 10 min initial denaturation (hot-start) at 95 °C

Table 2
Properties of qPCR primers and probes used in this study: Melting temperature (Tm), proportion of probe or primer sequence rDNA GC content (GC), amplicon length, qPCR annealing temperature (Ta), qPCR primer and probe concentration.

Name	Sequence (5-3)	Target species	rDNA	Tm	GC	Amplicon	Ta	Primer	Probe
Kmik2-F	CTCGCCTTGCATGTCAACGTCAGTT	*Karenia mikimotoi*	28S	62 °C	52%	178 bp	62 °C	500 nM	250 nM
Kmik4-R	TCT GCT CTG CAT GAA GGT TGT TGG T	*Karenia mikimotoi*	28S	61 °C	48%	178 bp			
Kmik1-P	FAM- CAC TGC TTC ATG TGC T –MGB	*Karenia mikimotoi*	28S	50 °C	50%				
KVITSF3*	CTGTGAACTTCTTTGTGAGCTCTT	*Karlodinium veneficum*	ITS1	55 °C	42%	128 bp	60 °C	500 nM	250 nM
KVITSR3*	TAGCGATAGCTTCGCAGACA	*Karlodinium veneficum*	ITS1	56 °C	50%	128 bp			
KVITSP3*	FAM-AGGTGAATCCCAATGCTGCTCCACTA-TAMRA	*Karlodinium veneficum*	ITS1	62 °C	50%				
Haka9-R	TGC AAT CCC AAG CAA CAC	*Heterosigma akashiwo*	28S	59 °C	52%	161 bp	62 °C	250 nM	125 nM
Haka8-F	AGC TTG CTG GCG AAT TGT AGT C	*Heterosigma akashiwo*	28S	60 °C	60%	161 bp			
Haka1-P	FAM- AAG GTG CGT GCT CAG TCG TGG TCC –TAMRA	*Heterosigma akashiwo*	28S	65 °C	63%				
Fjap7-F	GAAAGGGAAGCGAAGGAAGTCA	*Fibrocapsa japonica*	28S	58 °C	52%	171 bp	62 °C	250 nM	125 nM
Fjap6-R	CACGACATGCCACAGGGTT	*Fibrocapsa japonica*	28S	58 °C	58%	171 bp			
Fjap2-P	FAM- CAT ATT TCG TGC CTT –MGB	*Fibrocapsa japonica*	28S	50 °C	44%				

*Previously published Park et al., 2009

followed by 50 cycles: 15 s at 95 °C and 30 s annealing time at the primer-specific temperatures (Table 2). All samples were run in technical triplicates, two of the three replicates had to amplify for a sample to be considered positive, and all positive results with quantification cycle (Cq) higher than 40 was considered negative. All qPCR results are given as the average of the three technical replicates, with error bars indicating standard deviation. DNA templates were diluted ×10 in molecular grade water (Wilson, 1997) to reduce the influence of natural PCR inhibitors present in seawater and to avoid false negative results or underestimation of cell numbers.

The negative DNA-isolation controls (NEC and NSC) were tested with all primer-probe sets. In addition, a negative template control (NTC) was included with all qPCR runs to ensure that no contamination occurred during preparation of the sample plate.

2.5.1. Construction of standard curves

Standard curves were constructed from DNA isolated from 10 mL of culture harvested by filtration during the exponential growth phase (Table 3). The concentration of each species was determined using a hemacytometer (*H. akashiwo* and *K. veneficum*: Fuchs-Rosenthal, *F. japonica* and *K. mikimotoi*: Sedgewick-Rafter) under a Nikon Eclipse TE300 inverted microscope (Nikon, Tokyo, Japan). DNA was isolated as described in Section 2.3. The standard curve was constructed as a 5-step, 4-fold dilution series for all species, except OF2 samples with the *K. veneficum* assay, which were run as a 5-step 10-fold dilution series (Fig. 2). To avoid degradation of DNA during multiple thawing cycles the standards was diluted ×10 in molecular grade water (Promega), aliquoted, and stored at −20 °C. Diluted DNA standards were discarded after the first use.

2.5.2. Calibration

The accuracy of the qPCR assays was tested by adding cultures of known cell concentrations of the four target species to a sea water sample collected July 2015 at station Elle (59°37 N, 10°37 E, Fig. 1) in Outer Oslofjorden, following the procedure described in Section 2.1. The spiked sea water samples were filtered onto 25 mm polycarbonate filters with 1 μm pore size (Millipore) and DNA was isolated as described in 2.3. Two samples (500 mL) were processed without the addition of cultured cells and two samples were each spiked with 5 mL cultured cells of each target species to a total volume of 500 mL. At the same time aliquots of each culture were fixed in neutral Lugol's solution, final concentration 1%, and enumerated using a hemacytometer. QPCR was performed on the isolated DNA of the non-spiked samples and a 5-steps, 10-fold dilution series of the spiked samples. Cell concentrations where calculated for each target species in the same manner as the sea water samples (Section 2.5.3).

2.5.3. Data analysis

Amplification data were handled in Bio Rad CFX manager v 3.0 (Bio-Rad), with Cq determination mode set to single threshold, and the baseline decided by baseline subtracted curve fit. Unknown cell concentrations were derived directly from the standard calibration curve by Bio Rad CFX manager v 3.0. Raw data were extracted to Microsoft Excel Professional Plus 2010 where they were inspected manually.

2.6. Verification

2.6.1. Molecular verification

The PCR products from all qPCR reactions were run on a 3% agarose gel in TBE buffer at 5 kV cm^{-1} for 25 min to check that the PCR products were of the expected length. Two PCR products (high and low Cq) were selected for sequencing. The products were diluted ×5 with molecular grade water, purified with ExoSAP-IT (Affymetrix) and Sanger sequenced using the same primers as for qPCR (Table 1) by the GATC sequencing service. The resulting DNA-sequences were inspected manually in Geneious and taxonomically assigned with BLASTn (http://blast.ncbi.nlm.nih.gov).

2.6.2. Cell counts in light microscopy

For all sample dates, two replicate 5 mL Lugol's preserved sea water samples were settled for approximately 24 h and subsequently enumerated following the Utermöhl's sedimentation technique (Utermöhl, 1958). The naked dinoflagellates were counted as unidentified naked dinoflagellates (UND). Five selected

Table 3
Properties of the standards used to quantify qPCR results.

Species	Strain	Cells/10 mL	DNA cons (ng/μL)	Purity A260/A280
Karenia mikimotoi	UIO019	7450	53.89	1.56
Karlodinium veneficum	UIO254	227692	38.58	1.62
Heterosigma akashiwo	SCCAP K1549	521000	22.31	1.93
Fibrocapsa japonica	SCCAP K0542	8440	12.54	1.96

sample dates were re-examined to perform a more thorough LM examination where UND was identified to species level when possible and raphidophytes were searched for especially (Table 4). Aliquots (10 mL) of Lugol's preserved sea water samples were left to settle for approximately 12 h and enumerated following the Utermöhl's technique (a 50 mL sample was used for Oct 2010).

2.6.3. Scanning electron microscopy

Samples for scanning electron microscopy (SEM) were pre-filtered through a 45 μm sieve, concentrated with tangential flow filtration (Vivaflow 200; VivaScience, Hannover, Germany) and preserved in 1% osmium tetroxide (OsO_4) (Sigma-Aldrich, St. Louise, USA) and diluted in sterile filtered (0.2 μm mesh) natural sea water. The samples were mounted on glass cover slips covered in poly-L-lysin (Sigma-Aldrich) and left to settle overnight in a moist chamber before dehydration, followed by critical point drying and sputter coating in accordance with Engesmo et al. (2016). The samples were examined in a S-4800 Hitachi Field Emission Scanning Electron Microscope (Hitachi, Tokyo Japan). Cultures of *K. mikimotoi*, *K. veneficum* and *H. akashiwo* were prepared and examined as above for comparison. It was necessary to prepare *Fibrocapsa japonica* for SEM separately, because cell morphology became distorted when fixed in OsO_4. To prevent the trichocysts from discharging, cells were added to the Lugol's solution (1% final concentration), quickly followed by the addition of glutaraldehyde (GLA, 1% final concentration) and then gently mixed by inverting the tube. Cells were left to sink 1 h before they were collected from the bottom of the tube, mounted on a glass slide and prepared as above.

2.6.4. 454 High throughput sequencing

Field sampling (from 1 m depth, cell size fraction 3–45 μm), RNA extraction, reverse transcription to cDNA, PCR, and sequencing followed in large the protocol by Egge et al. (2015), with one modification: PCR amplification targeted the hypervariable V4 region (~380 bp) of the 18S rRNA gene was performed using the universal eukaryote primers TAR454-F3: 5′-CCAGCASCYGCGGTAATTCC-3′ and TAReukREV3: 5′-ACTTTCGTTCTTGATYRA-3′ described by Stoeck et al. (2010). Analyses of 454 reads were carried out as described by Logares et al. (2014) with some modifications: AmpliconNoise v.1.6.0 (Quince et al., 2011) was used to denoise the amplicons (~400 bp), and Perseus was used to remove the putative chimeras (Quince et al., 2011) as implemented in QIIME pipeline v.1.4 (Caporaso et al., 2010). UCLUST v.1.2.22 (Edgar, 2010) was used to cluster the reads at a 98% similarity threshold. All generated OTUs that contained singletons (only one read) or doubletons (only two reads that were both present in the same sample) were removed. Operational taxonomic units (OTUs) assigned to Raphidophyceae were aligned to the Engesmo et al. (2016) Raphidophyceae alignment, using MAFFT-add in v.7.1.9 with the Q-INS-I strategy (http://mafft.cbrc.jp/alignment/server/). The alignment was checked and manually edited in Geneious v.7.1.9. Phylogeny was inferred with MrBayes v. 3.2.2 in Geneious, using the substitution model GTR and invariable gamma rate variation, MCMC settings 2 000 cycles, four heated chains, and subsampling frequency of 500 (Huelsenbeck and Ronquist, 2001). OTUs assigned to Dinophyceae were further taxonomically assigned as far as possible by local Blast in Geneious against a local database consisting of 1593 dinoflagellate reference sequences from PR2 (Guillou et al., 2013), followed by phylogenetic analyses (Gran-Stadniczeñko et al., unpublished). The nucleotide sequences of the OTUs were submitted to ENA and have the accession numbers PRJEB20755 (study) and ERZ407999 (analysis).

3. Results

The specificity of the qPCR assays was tested by running the two dinoflagellates assays (targeting *Karenia mikimotoi* and *Karlodinium veneficum*) on cultures of other dinoflagellate strains (Table 1). The two raphidophyte assays (targeting *Heterosigma akashiwo* and *Fibrocapsa japonica*) were tested on other raphidophytes and dictyochophytes (Table 1). No unspecific amplification occurred for any species. All qPCR products were checked by agarose gel electrophoresis and only one clear band of the expected length was visible on the gel, indicating no unspecific binding masked by the probe. Sequences of the qPCR-products were blasted and matched their intended targets. The qPCR properties (efficiency, slope and r^2) as calculated from the standard curves are shown in Fig. 2 for all sample runs. The PCR efficiency was generally high (>90%). No bands of DNA were visible in the negative controls.

Table 4
Comparison of all results for selected sample dates (2009–2012).

Karenia mikimotoi	Sep 09	Oct 10	Jun 11	Aug 11	Jun 12
qPCR (cells L^{-1})	54	1665	0	662511	444
SEM (recorded)	X			X	
LM (cells L^{-1})	200	1000	0	6800	600
LM UND* 25–40 μm (cells L^{-1})	1888	0	10500	8400	1200
Karlodinium veneficum:					
qPCR (cells L^{-1})	189	7604	438210	82	34665
SEM (recorded)	X		X		X
LM (cells L^{-1})	600	4600	9000	100	6000
LM UND* 8–24 μm (cells L^{-1})	8800	16500	110400	8400	30800
Heterosigma akashiwo:					
qPCR (cells L^{-1})	581	8402	0	219	0
SEM (recorded)		X			
LM (cells L^{-1})	0	3800	0	0	0
454 HTS (recorded)	X	X		–	–
Fibrocapsa japonica					
qPCR (cells L^{-1})	257	27	0	225	0
SEM (recorded)					
LM (cells L^{-1})	0	100	0	0	0
454 HTS (recorded)	X			–	–

UND = Unidentified naked dinoflagellate, – = No data for this date, X = species recorded, no quantifiable data available, blank space = data analysed but species not recorded.

3.1. Calibration experiment

Samples of seawater spiked with known concentrations of the four species were accurately detected using the qPCR assays (Fig. 3). Detection limits were determined based on this experiment and on the results from the field samples. There was only slight variation between the two biological replicates for all species. The detection limit for *Karenia mikimotoi* (spiked with 110,000 cell L^{-1}, Fig. 3a), *Karlodinium veneficum* (spiked with 270,000 cells L^{-1}, Fig. 3b) and *Fibrocapsa japonica* (spiked with 16,000 cells L^{-1}, Fig. 3d) were found to be approximately 1 cell L^{-1}. The assay for *Heterosigma akashiwo* (spiked with 110,000 cells L^{-1}, Fig. 3c) was not able to detect the lowest concentrations, giving a detection limit of approximately 100 cells L^{-1}.

3.2. Field samples

The qPCR assays confirmed the presence of all four species in Oslofjorden (Fig. 4). The dinoflagellates appeared frequently in the samples, *Karenia mikimotoi* was recorded during most summer and autumn months in low quantities (<1000 cells L^{-1}). There was one peak in August 2011 (665,000 cells L^{-1}); the population was then present in low cell numbers throughout 2012 (Fig. 4a). In June 2011 there was a peak of *Karlodinium veneficum* with cell numbers reaching 438,000 cells L^{-1}, and it was recorded almost every

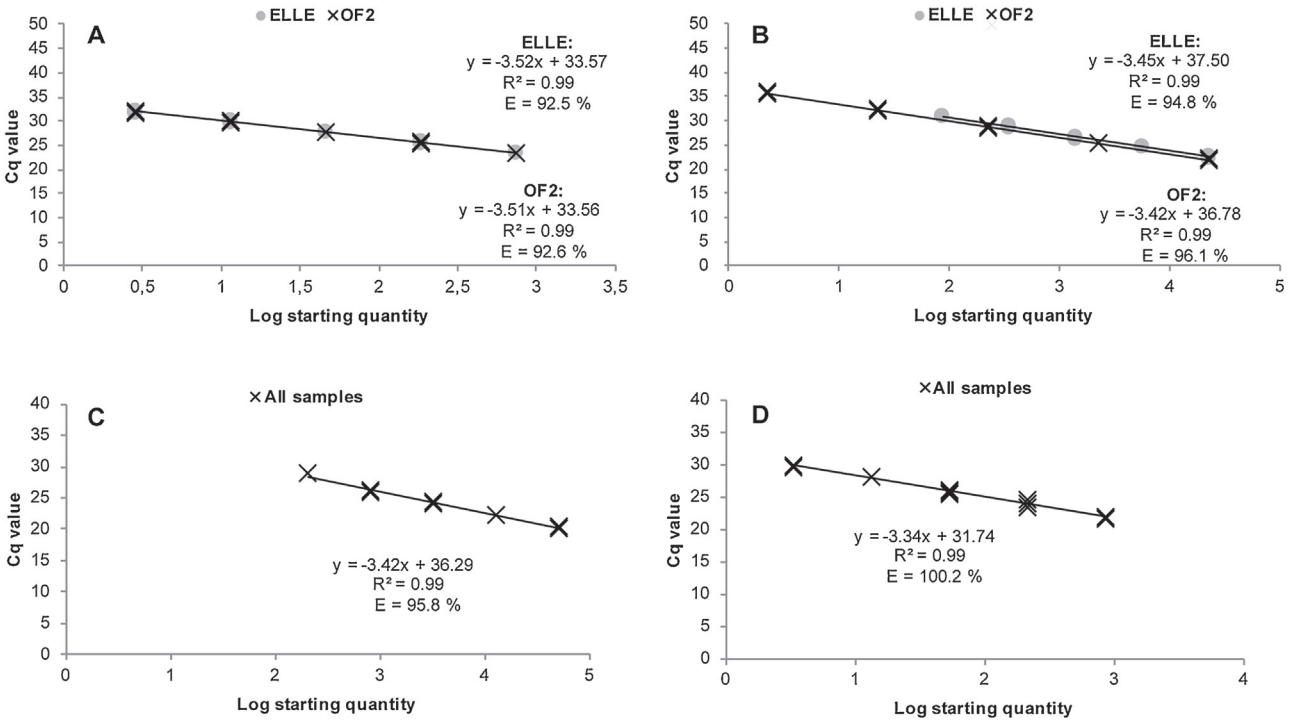


Fig. 2. qPCR properties showing efficiency, slope and r^2 as calculated from the standard curves for a) *Karenia mikimotoi* b) *Karlodinium veneficum* c) *Heterosigma akashiwo* and d) *Fibrocapsa japonica.*

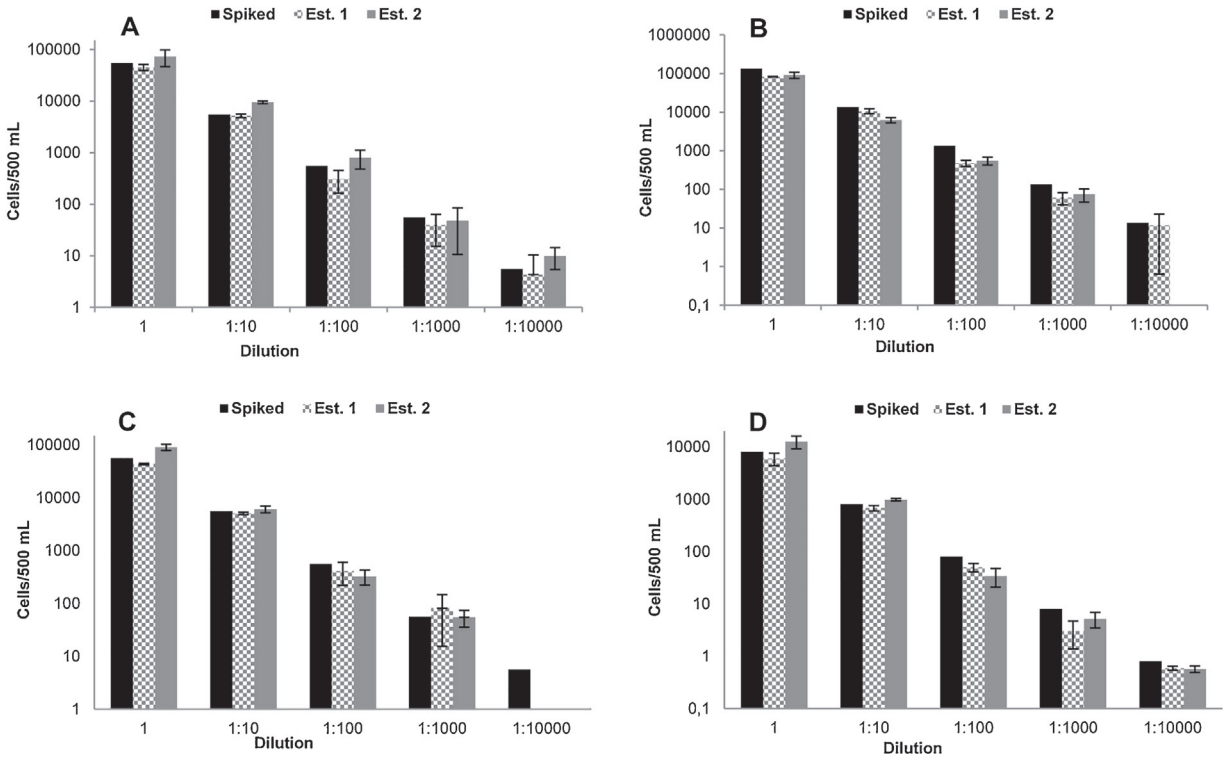


Fig. 3. Results from the experiments with spiked samples from ELLE: a) *Karenia mikimotoi* b) *Karlodinium veneficum* c) *Heterosigma akashiwo* and d) *Fibrocapsa japonica*. The left column (spiked) indicates the number of cells added as counted in a hemocytometer. The estimations are two biological replicates of qPCR estimates of the number of cells. All cell estimates are given as the average of three technical replicates, with error bars indicating standard deviation.

month from August 2009 until August 2011 (except Oct 09) in relatively low abundances, 100-10,000 cells L^{-1}. During the last year it occurred less frequently except in June 2012 when there was a small peak of 35,000 cells L^{-1} (Fig. 4a).

The raphidophytes appeared less frequently than the dinoflagellates (Fig. 4b), and both *Heterosigma akashiwo* and *Fibrocapsa japonica* was recorded during the autumn of 2009, 2010 and 2011 in low abundances. The highest recorded cell estimate of *H. akashiwo* was in October 2010 with 8400 cells L^{-1}, while cell estimates of *F. japonica* never went above 250 cells L^{-1}.

3.3. Light microscopy cell counts

None of the species in this study were initially registered in the microscopic cell counts (data not shown); however, *Karenia mikimotoi* and *Karlodinium veneficum* were included as UND in their respective size groups. Five samples (Sep 09, Oct 10, Jun 11, Aug 11 and Jun 12) were re-examined in LM (cell counts) with special emphasis on the four species included in this study (Table 4). The two dinoflagellates, *K. mikimotoi* and *K. veneficum*, were identified and counted in LM for all dates that had a positive qPCR signal. There was good correlation between the qPCR and LM cell estimates, with the exceptions of June and August 2011, where the qPCR estimates were two orders of magnitude larger than the LM estimates (Table 4). The only LM registration of *Heterosigma akashiwo* and *Fibrocapsa japonica* were from October 2010 with 3800 cells L^{-1} and 100 cells L^{-1}, respectively.

3.4. Morphology

Material from the five sampling dates that were chosen and re-counted in LM were also examined in SEM. In September 2009 and August 2011 *Karenia mikimotoi* was identified and compared with cultured cells of strain UIO019 (Fig. 5). Cells were dorso-ventrally

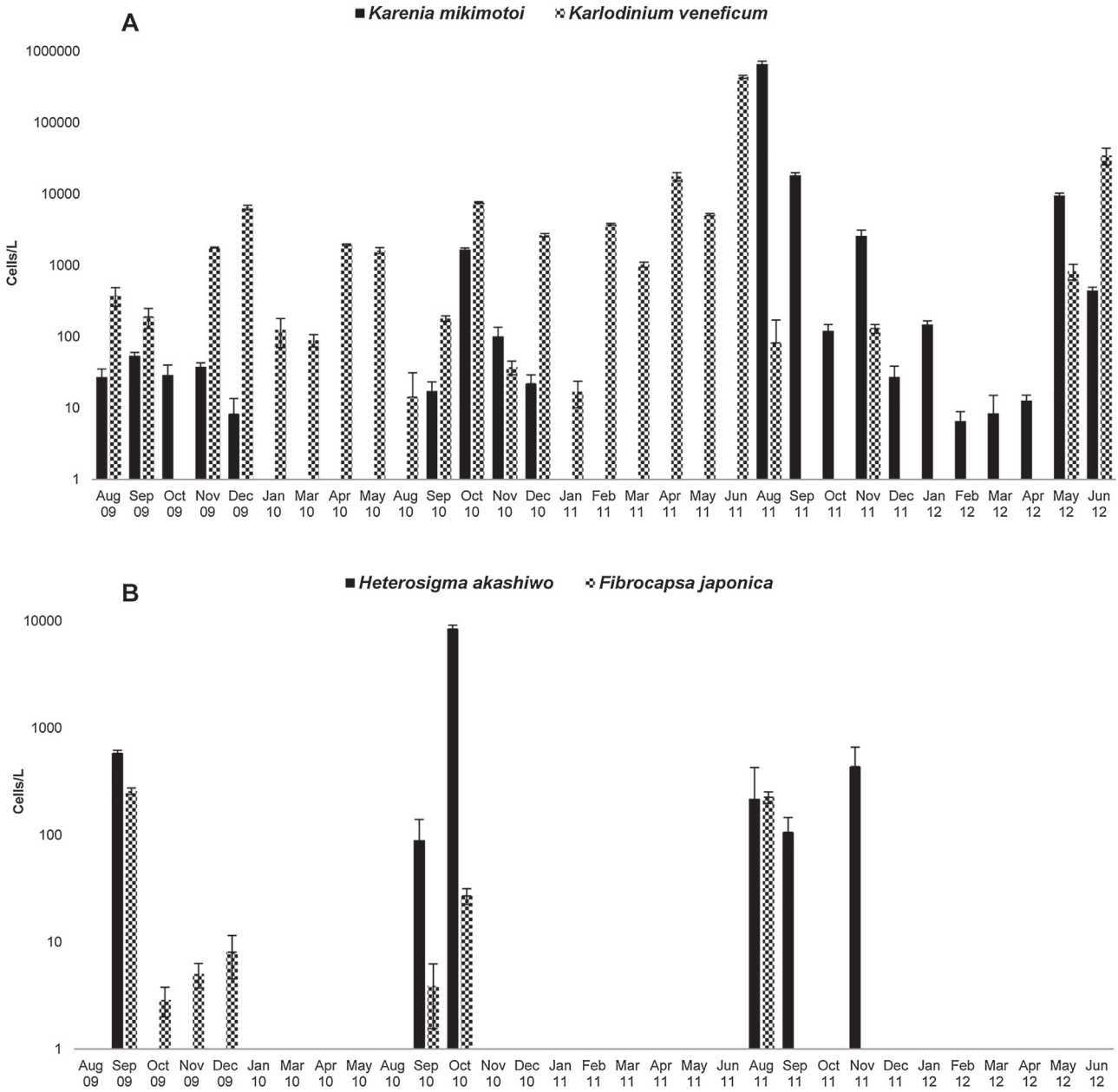


Fig. 4. qPCR results from station OF2 in the outer Oslofjorden for a) *Karenia mikimotoi* and *Karlodinium veneficum* and b) *Heterosigma akashiwo* and *Fibrocapsa japonica*. All estimates are given as cells L^{-1} presented as the average of three technical replicates, with error bars indicating standard deviation.

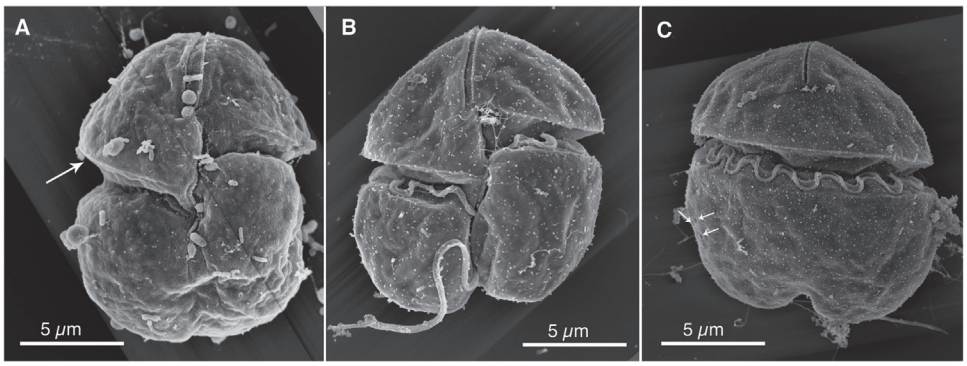


Fig. 5. SEM images of *Karenia mikimotoi*, a) ventral view of cell from field material (OF2 Aug 11), note the distinct edge on the epitheca (arrow) which is relatively visible despite the cell not having preserved well. b) Ventral view of strain SCCAP K-0260 with the distinct epithecal edge. c) Dorsal view of strain SCCAP K-0260 with three antapical pores (arrows).

compressed, somewhat taller than wide, and displayed large cingulum displacement. The width:height ratio varied, but cells were always taller than wide. They had a distinct, straight apical groove, extending from the ventral side of the epicone (Fig. 5a–b), over apex and down into the dorsal side of the epicone (Fig. 5c). The epicone had a distinct edge, which is easily recognisable in both LM and SEM. Three antapical pores were seen in left dorsal view (Fig. 5c). The cells from the field samples closely resembled those in culture (Fig. 5b–c).

Cells of *Karlodinium veneficum* were identified from samples collected in September 2009, June 2011 and June 2012 (Table 3) and compared with cultured cells of strain UIO254. Cells had large cingulum displacement, with sulcal intrusion into the epitheca and longitudinal rows or depressions beneath the amphiesma vesicles (Fig. 6). The apical groove was straight, but less pronounced than in *Karenia mikimotoi* (Fig. 5); in dorsal view, the termination of the apical groove was discernible close to the apex (Fig. 6c). Two apical pores were visible (Fig. 6a–b). Cells from the field material closely resembled cells from culture as shown in Fig. 6a, which is from a field sample collected in June 2012.

The morphology of the two Raphidophycean species, *Heterosigma akashiwo* and *Fibrocapsa japonica,* were difficult to detect in field samples with SEM (Figs. 7 and 8); however, *H. akashiwo* was identified from October 2010. The cell was not well preserved, but the heterokont flagella were intact, and it was apparent that it was naked because of the tear in the cell. The cell had a rounded outline with surface structures, discharged mucocysts, and rod-like structures (Fig. 8b). No identification of *F. japonica* was done in SEM (Table 4), but the morphology is depicted by cultured cells. The cells were rounded to oval with a variable outline and two apically inserted flagella. Trichocysts were concentrated in the posterior end of the cells.

3.5. 454 high throughput sequencing

The 454 HTS demonstrated the presence of minimum four different genotypes from the dinoflagellate Family Kareniaceae. It also successfully documented the presence of raphidophytes in Oslofjorden: In September 2009 and October 2010 *Heterosigma akashiwo* occurred, and *Fibrocapsa japonica* was detected in September 2009 (Fig. 9 and Table 4). Aligning all Raphidophyceaen 454 HTS generated sequences to a raphidophyte reference alignment also revealed previously undocumented genetic variety within Raphidophyceae (Fig. 9). An unknown genotype, probably representing a novel genus, was recorded during early spring of 2010 and 2011 (Mar 10 and Feb–Apr 11).

4. Discussion

In this study, we were able to detect and enumerate the ichthyotoxic flagellates *Karenia mikimotoi*, *Karlodinium veneficum*, *Heterosigma akashiwo,* and *Fibrocapsa japonica* over the course of three years (Aug 2009–Jun 2012) from environmental water samples collected in Outer Oslofjorden using qPCR. The results document the first occurrence of *F. japonica* in Norwegian waters and demonstrate the potential of qPCR as a monitoring method.

The two dinoflagellates, *Karenia mikimotoi* and *Karlodinium veneficum,* are known to be relatively common components of the coastal phytoplankton community along the Norwegian coast (Throndsen et al., 2003). During the autumn of 2009 and 2010, *K.*

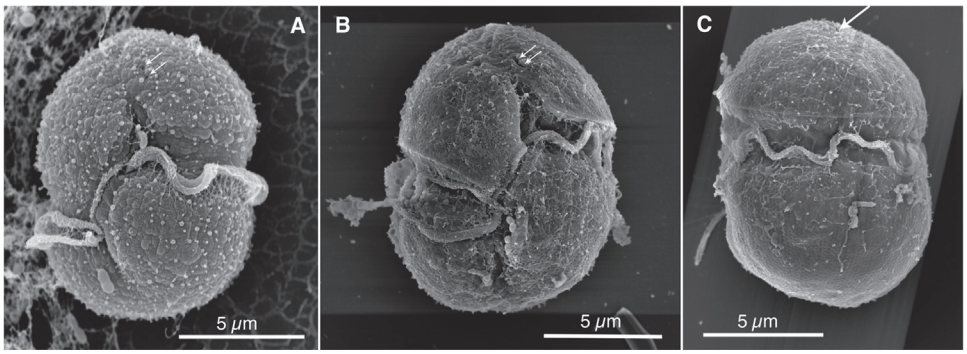


Fig. 6. SEM images of *Karlodinium veneficum,* a) ventral view of cell from field material (OF2 June 2012), note the sulcal intrusion into the epitheca and the two apical pores (arrows). b) Ventral view of strain SCCAP K-1640 with sulcal intrusion into the epitheca, two apical pores (arrows) and visible rows of depressions beneath the aphiesmal vesicles. c) Dorsal view of strain SCCAP K-1640 with visible rows of depressions beneath the aphiesmal vesicles, also notice the short apical groove (arrow).

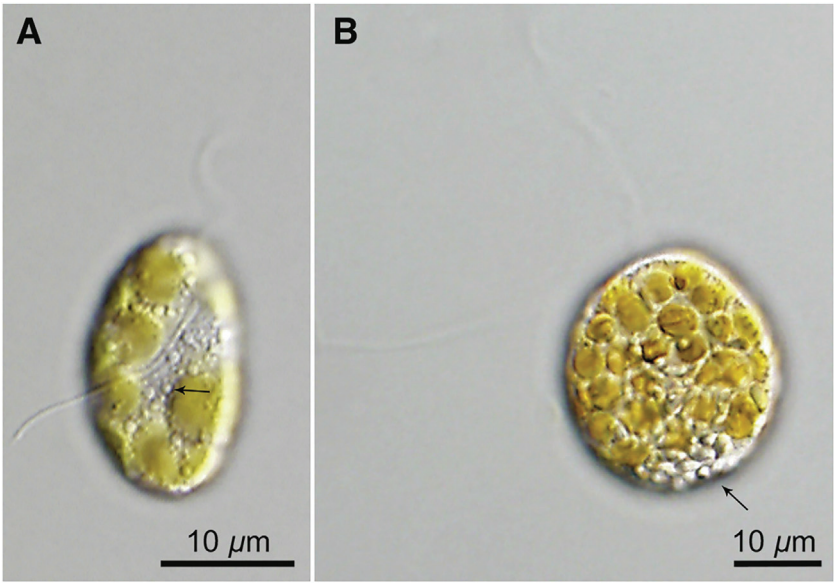


Fig. 7. LM of a) *Heterosigma akashiwo* strain SCCAP K-1549 showing both anterior (beating) and posterior (trailing) flagella, peripheral chloroplasts and mucocysts (arrow). The central nucleus is visible as a grey area surrounded by chloroplasts. b) *Fibrocapsa japonica* strain SCCAP K-0542 showing both the anterior (beating) and posterior (trailing) flagella, multiple chloroplasts and the mucocysts located in the posterior end (arrow). Both images kindly provided by Gert Hansen.

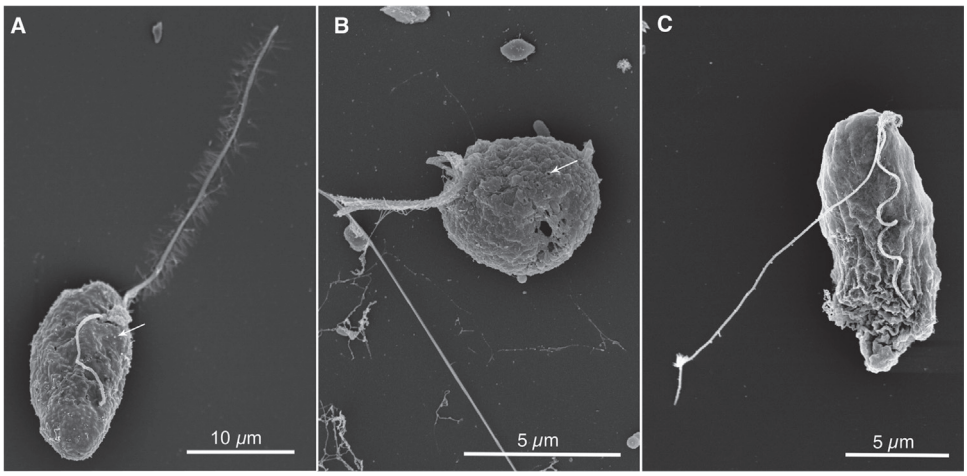


Fig. 8. SEM images of a) *Heterosigma akashiwo* strain SCCAP K-1549 with hairy (anterior) flagellum, smooth (posterior) flagellum and unreleased mucocysts (arrow). b) *Heterosigma akashiwo* cell from field material (OF2 Oct 10) with both flagella and discharged mucocysts (arrow). Note the rupture in the cell revealing the fragile plasma membrane. c) *Fibrocapsa japonica* strain SCCAP K-0542 with hairy (anterior) flagellum, smooth (posterior) flagellum and smooth cell surface. Note the "wrinkled" end, which contains the trichocysts.

mikimotoi appears in low concentrations, but it was not recorded during winter, spring, or summer. The concentration peaked in August 2011 and appeared to linger until June 2012, which marked the end of the sampling period. The other dinoflagellate, *K. veneficum,* was present year around in the period August 2009 to June 2011 and its concentration peaked in June 2011; however, it was not common from August 2011 to June 2012, when *K. mikimotoi* was frequent. Neither *K. mikimotoi* nor *K. veneficum* were identified from initial LM cell counts because they were included as unidentified naked dinoflagellates (UND) in their respective size groups. Field samples from five sample dates were re-examined in LM, and new cell counts were performed with special emphasis on finding the species targeted in the present qPCR assays. Upon re-examination in LM, cells that complied with the morphology of both *K. mikimotoi* and *K. veneficum* were found in all samples with positive qPCR signals. Results showed a positive correlation between the cell estimates given by qPCR and the LM cell counts. In the two months with the highest qPCR signals (Aug 2011 for *K. mikimotoi* and Jun 2011 for *K. veneficum*), the discrepancy between qPCR and LM estimates were notable, with LM estimates being two orders of magnitude below the qPCR estimates. Both LM examinations of the sample from August 2011 recorded less than 10,000 cells L^{-1} (UND 25–40 μm and *K. mikimotoi*), whereas qPCR estimated 660 000 *K. mikimotoi* cells L^{-1}. In June 2012, the discrepancy between the two LM examinations was also large, with the first examination recording approximately 110,000 cells L^{-1} (UND) and the second only 9000 cells L^{-1} of *K. veneficum*. The reason for the observed discrepancy between qPCR and LM cell estimates may be that both species are relatively difficult to detect in LM, and could be overlooked. They are also sensitive to fixation and may have their morphology distorted, rendering them unidentifiable. Another possibility is that there are closely related, novel species present in Oslofjorden, which are picked up by the qPCR assays, but not recorded in LM. Because we

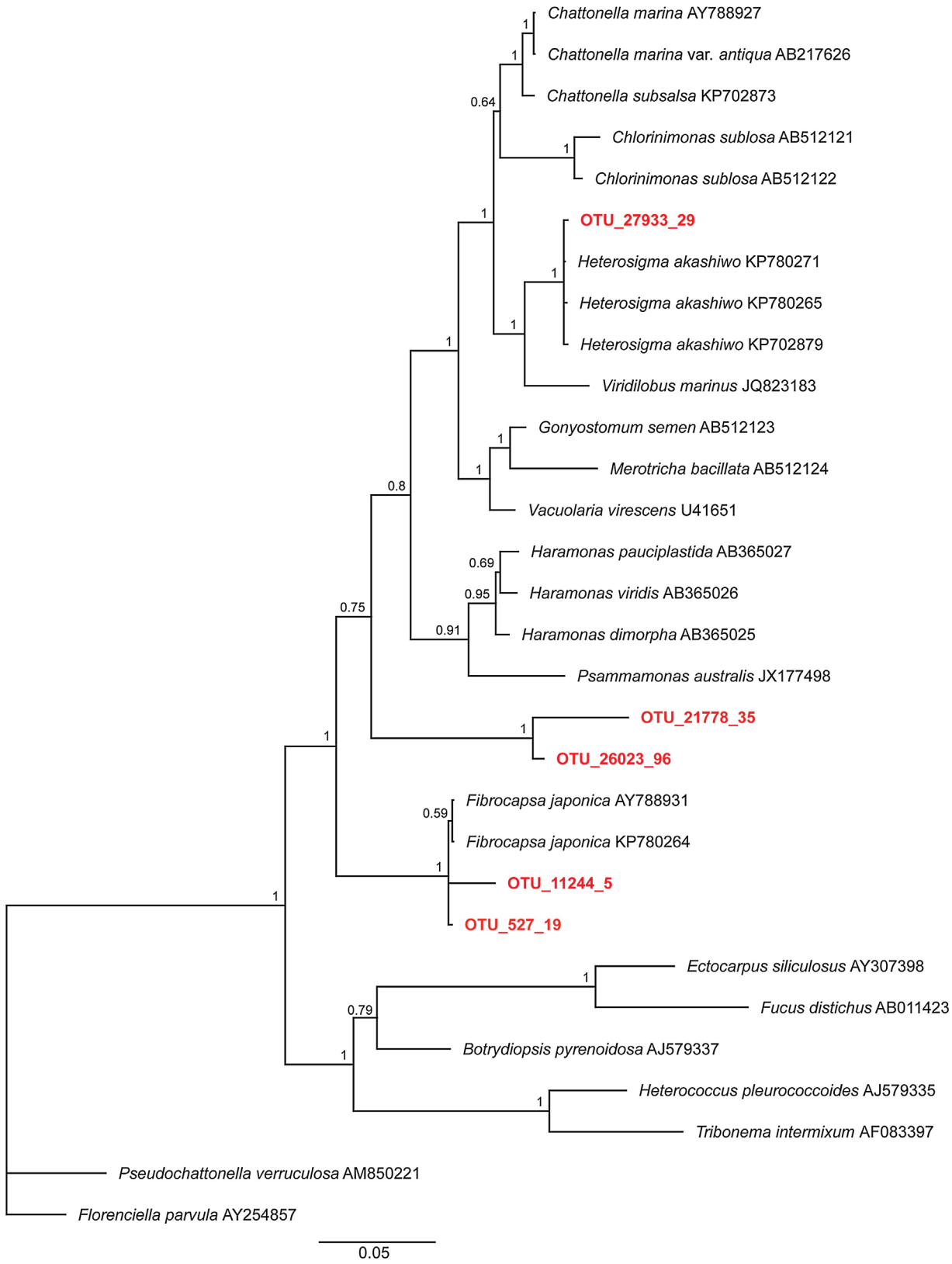


Fig. 9. Phylogeny of Raphidophyceae based on 18S rDNA (1988 characters), aligned with 454 environmental sequencing OTUs (371–421 bp in the V4 region) using Bayesian Interference (MrBayes). Supporting values are given as Bayesian posterior probability. OTUs are marked in bold. The nucleotide sequences of the OTUs are available from ENA and have the accession numbers PRJEB20755 (study) and ERZ407999 (analysis).

chose to work with DNA, instead of RNA, there is also the possibility that we are detecting eDNA from cells that are no longer viable, which may contribute to overestimation of qPCR cell estimates and lead to false positive samples (Goldberg et al., 2015).

The raphidophyte, *Heterosigma akashiwo,* was recorded with qPCR from all three sampling years, albeit at very low concentrations, and the highest concentration (8400 cells L^{-1}) was recorded in October 2010 (Table 4). It went undetected in the initial LM cell counts, but when the samples were re-examined, 3800 cell L^{-1}, possibly corresponding to *H. akashiwo's* morphology were counted from October 2010. One cell of *H. akashiwo* was identified in SEM from October 2010 (Fig. 8b). It was not recorded in LM from any other sampling date. The presence of the dictyochophyte, *Pseudochattonella farcimen* (Eikrem, Edvardsen & Throndsen) Eikrem, was recorded in September 2009 and October 2010, and *H. akashiwo* has previously been reported to appear alongside this species (Edvardsen et al., 2007; Naustvoll et al., 2002). In the present study, *Fibrocapsa japonica* was detected with the qPCR assay in the autumn of 2009, 2010 and 2011, albeit in very low concentrations (the highest concentration was 257 cells L^{-1} in Sep 09). One cell was recorded in LM from October 2010. This is the first time *F. japonica* was recorded in Norwegian waters, but there is a LM record from the Swedish coast at the island of Åstol, which is only 160 km away from the Norwegian border (www.smhi.se/klimatdata/oceanografi/havsmiljodata). Alongside other toxic raphidophytes, *Fibrocapsa japonica* has been observed in French and Dutch coastal waters since 1991 (Billard, 1992; Vrieling et al., 1995). It is now well established in the North Sea and has occurred in bloom concentrations in the German Bight (Rademaker et al., 1998) and in the northern Adriatic Sea (Cucchiari et al., 2008), causing kills of farmed fish, and it has on one occasion been linked, although not unequivocally to the death of seals (Leftley and Hannah, 2009). So far, *Fibrocapsa japonica,* has not been detected by the Norwegian Surveillance Programme (www.algeinfo.imr.no) that conducts light microscopy examinations of water samples collected along the Norwegian coast. A possible introduction and establishment of *F. japonica* in Norwegian waters may pose a future challenge to fish-farmers and wild life. Verifying the presence of *F. japonica* with microscopy proved difficult, therefore we used available 454 HTS data (Gran-Stadniczeñko et al., unpublished) for the first two sampling years (Aug 09–Jun 11), in which *F. japonica* was recorded from September 2009. The 454 HTS data did not detect *F. japonica* from any other positive sample, suggesting that the qPCR detection limit is lower than 454 HTS. The 454 HTS also confirmed the presence of *H. akashiwo,* and it suggested the presence of a novel Raphidophyceae genus in Oslofjorden. Interestingly, this unknown, novel taxon, expected to represent a novel genus appeared during early spring (Mar 10, Feb–Apr 11), unlike the two other raphidophytes, which only occurred in the autumn (data not shown).

The concentrated samples of small phytoplankton (<45 μm) were examined in SEM, which allowed the morphology of cells from field samples to be compared with cells from cultures (Figs. 5–8). The morphology of the *K. mikimotoi* cells found in field samples was compared to the morphology of strain UIO019, which was isolated from Oslofjorden in 1977. Both cultured cells and cells from field material clearly conformed to previous descriptions of *K. mikimotoi* (Daugbjerg et al., 2000; Haywood et al., 2004). Certain cells, both in culture and field material exhibited the morphology of *K. mikimotoi,* but were smaller than the previously published size range (18–40 μm in length and 13–35 μm in width). This indicates that the currently published size range for *K. mikimotoi* should probably be amended to include smaller cells. The field material also contained cells as small as 10 μm fitting *Karenia* morphology from sample dates that were negative for *K. mikimotoi*, indicating that there is a novel *Karenia*-species commonly present in Oslofjorden. A recent investigation from French coastal waters indicate that the diversity of the dinoflagellate family Kareniaceae is much larger than previously recorded (Nézan et al., 2014). The *Karlodinium veneficum* strain UIO254 was isolated from Oslofjorden in 1977 and was used here to compare morphology with the cells found in the field samples. The cultured cells conform to previous descriptions of the species (Daugbjerg et al., 2000; Bergholtz et al., 2006). Although it was outside the scope of this project to identify and describe novel species, it was clear from LM, SEM and 454 HTS that the diversity of naked dinoflagellates in Oslofjorden was greater than currently recognized.

No identification of *Fibrocapsa japonica* were obtained from field samples in SEM, which could be due to its low cell concentrations or because *F. japonica* does not preserve well in OsO_4-fixation. Strain SSCAP K-0542 isolated from the North Sea (Helgoland, Germany) was examined in SEM, and provided some further insights into why SEM detection of this species is particularly difficult. Satisfactory results were not obtained with OsO_4 fixation (Fig. S2), therefore a separate procedure had to be followed for *F. japonica* (see Section 2.6.3). The "wrinkled" end of *F. japonica* is where the trichocysts are located. When disturbed, they readily discharge, causing disruption to the cell, covering it in mucus and discharge as seen in Fig. S2.

Absolute quantification of microalgal targets in qPCR assays are usually achieved using standard curves. The curve is generated using a serially diluted DNA standard of a known quantity, which creates a linear relationship between the threshold cycle (Cq) and the logarithm of the starting quantity of DNA in the standard (Heid et al., 1996). Two types of standards are typically used: DNA from cultured cells of the targeted species (Park et al., 2007; Handy et al., 2008; Park et al., 2009; Eckford-Soper and Daugbjerg, 2015b) or a cloned plasmid of the targeted gene (Galluzzi et al., 2008; Galluzzi et al., 2010; Yuan et al., 2012; Zamor et al., 2012). The plasmid approach will generate the copy number of the targeted gene; it is therefore essential to know how many copies of the targeted gene are present per cell, and if this number is constant. In a eukaryote, nuclear genome rDNA will typically consist of hundreds of tandemly repeated copies, but it can consist of as few as one copy or up to several thousand (Hillis and Dixon, 1991). It is also previously documented that the amount of rRNA can vary between both strains and life stages (Galluzzi et al., 2010), growth conditions (Dittami and Edvardsen, 2013) and for dinoflagellates especially, biovolume (Godhe et al., 2008). In this study, standards with known quantity of cultured cells were used, therefore bypassing the problem of determining the gene copy number per cell. Using cultured cells as a standard also has its shortcomings. DNA in whole cells is much less stable than in cloned plasmids, meaning new standards must be prepared from live and fresh cells. Keeping live cultures is time consuming, and so is enumeration and DNA isolation. In the present study, DNA degradation of the cultured standard appeared as a major challenge. If a dilution series was thawed twice, the lower concentrations would start to degrade, resulting in lower qPCR efficiencies (E), which results in over-estimation of cell numbers. Subsequently, TE buffer was used for diluting samples for the standard curves and for storing standards.

A guideline for minimum information of publication of qPCR results (MIQE) was published in 2009 (Bustin et al., 2009). Several different qPCR assays have been published since the early 2000's, including the micro algal species of the present study (i.e. Coyne et al., 2005; Bowers et al., 2006; Yuan et al., 2012). In order to comply with the MIQE guidelines, new primers and probes were designed for all species except *K. veneficum*, where an assay developed by Park and co-workers (Park et al., 2009) was used. The region of a recently published assay for *K. mikimotoi* (Smith et al.,

2014) is located slightly downstream of the assay presented herein, but targeting largely the same region of 28S rDNA. The detection limit reported by Smith et al. is lower (0,007 cells L^{-1}) than for the present assay (1 cell L^{-1}), however there is no reason to believe the assay presented herein could not detect lower concentrations, if applied.

The results demonstrate that qPCR is a sensitive tool for the quantification of the four species targeted in the present study and would be suitable and valuable as a compliment to LM-based monitoring as the qPCR assays had a higher sensitivity and lower detection limit than LM cell counts.

Acknowledgements

The authors are very grateful to colleagues at the University of Oslo, mainly Sissel Brubak for assistance with molecular biological lab work, Vladyslava Hostyeva for the initial LM cell counts and Elianne Egge for allowing us to use her DNA and HTS reads for the first two sampling years. We also thank the curator of SCCAP culture collection, Gert Hansen, for allowing us the use of his LM picture of strain SCCAP K-1549 and SCCAP K-0542 (Fig. 7). The Norwegian Research Council is thanked for funding project 196702/E40 Toxic Algae. [SS]

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.hal.2018.04.007.

References

Andersen, R., 2005. Algal Culturing Techniques. Elsevier Academic Press, Amsterdam, pp. 596.
Anderson, D.M., Cembella, A.D., Hallegraeff, G.M., 2012. Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. Annu. Rev. Mar. Sci. 4, 143–176.
Bergholtz, T., Daugbjerg, N., Moestrup, O., Fernandez-Tejedor, M., 2006. On the identity of *Karlodinium veneficum* and description of *Karlodinium armiger* sp. nov. (Dinophyceae), based on light and electron microscopy, nuclear-encoded LSU rDNA, and pigment composition. J. Phycol. 42 (1), 170–193.
Billard, C., 1992. *Fibrocapsa japonica* (Raphidophyceae) algue planctonique nouvelle pour les cotes de France. Cyptogamie Algologie 13, 225–231.
Bott, N.J., Ophel-Kellener, K.M., Sierp, M.T., Rowling, K.P., Mckay, A.C., Loo, M.G.K., Tanner, J.E., Deveney, M.R., 2010. Toward routine, DNA-based detection methods for marine pests. Biotechnol. Adv. 28, 706–714.
Bowers, H.A., Tomas, C.R., Tengs, T., Kempton, J.W., Lewitus, A.J., Oldach, D.W., 2006. Raphidophyceae (Chadefaud ex Silva) systematics and rapid identification: sequnce analyses and real-time PCR assays. J. Phycol. 42 (6), 1333–1348.
Braarud, T., Nygaard, I., 1967. Oslofjorden Og Dens Forurensningsproblemer. Report Norwegian Institute for Water Research, Oslo, pp. 1–171.
Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55 (4), 611–622.
Caporaso, G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7 (5), 335–336.
Coyne, K.J., Handy, S.M., Demir, E., Whereat, E.B., Hutchins, D.A., Portune, K.J., Doblin, M.A., Cary, S.C., 2005. Improved quantitative real-time PCR assays for enumeration of harmful algal species in field samples using an exogenous DNA reference standard. Limnol. Oceanogr. Methods 3, 381–391.
Cucchiari, E., Guerrini, F., Penna, A., Totti, C., Pistocchi, R., 2008. Effect of salinity, temperature, organic and inorganic nutrients on growth of cultured *Fibrocapsa japonica* (Raphidophyceae) from the northern Adriatic Sea. Harmful Algae 7, 405–414.
Daugbjerg, N., Hansen, G., Larsen, J., Moestrup, Ø., 2000. Phylogeny of some of the major genera of dinoflagellates based on ultrastructure and partial LSU rDNA sequence data, including the erection of three new genera of unarmoured dinoflagellates. Phycologia 39 (4), 302–317.
de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Lukes, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., Tara Oceans, C., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348 (6237), 1261605.
Dittami, S.M., Edvardsen, B., 2013. Culture conditions influence cellular RNA content in ichthyotoxic flagellates of the genus *Pseudochattonella* (Dictyochophyceae). J. Phycol. 48 (4), 1050–1055.
Dittami, S.M., Hostyeva, V., Egge, E.S., Kegel, J.U., Eikrem, W., Edvardsen, B., 2013. Seasonal dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. Environ. Sci. Pollut. Res. 20 (10), 6719–6732.
Ebenezer, V., Medlin, L.K., Ki, J.-S., 2012. Molecular detection, quantification, and diversity evaluation of microalgae. Mar. Biotechnol. 14, 129–142.
Eckford-Soper, L.K., Daugbjerg, N., 2015a. Examination of six commonly used laboratory fixatives in HAB monitoring programs for their use in quantitative PCR based on Taqman probe technology. Harmful Algae 42, 52–59.
Eckford-Soper, L.K., Daugbjerg, N., 2015b. Development of a multiplex real-time qPCR assay for simultaneous enumeration of up to four marine toxic bloom-forming microalgal species. Harmful Algae 48, 37–43.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26 (19), 2460–2461.
Edvardsen, B., Eikrem, W., Shalchian-Tabrizi, K., Riisberg, I., Johnsen, G., Naustvoll, L., Throndsen, J., 2007. *Verrucophora farcimen* gen. et sp nov (Dictyochophyceae, Heterokonta) – a bloom-forming ichthyotoxic flagellate from the Skagerrak, Norway. J. Phycol. 43 (5), 1054–1070.
Egge, E.S., Eikrem, W., Edvardsen, B., 2015. Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. J. Eukaryot. Microbiol. 62 (1), 121–140.
Elbrächter, M., 1999. Exotic flagellates of coastal North Sea waters. Helgoländer Meeresunters 52, 235–242.
Engesmo, A., Eikrem, W., Seoane, S., Smith, K., Edvardsen, B., Hofgaard, A., Tomas, C. R., 2016. New insights into the morphology and phylogeny of *Heterosigma akashiwo* (Raphidophyceae), with the description of *Heterosigma minor* sp. nov. Phycologia 55 (3), 279–294.
Galluzzi, L., Bertozzini, E., Penna, A., Perini, F., Pigalarga, A., Graneli, E., Magnani, M., 2008. Detection and quantification of *Prymnesium parvum* (Haptophyceae) by real-time PCR. Lett. Appl. Microbiol. 46 (2), 261–266.
Galluzzi, L., Bertozzini, E., Penna, A., Perini, F., Garcés, E., Magnani, M., 2010. Analysis of rRNA gene content in the Mediterranean dinoflagellate *Alexandrium catenella* and *Alexandrium taylori*: implications for the quantitative real-time PCR-based monitoring methods. J. Appl. Phycol. 22, 1–9.
Godhe, A., Asplund, M.E., Härnström, K., Saravanan, V., Tyagi, A., Karunasagar, I., 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microb. 74, 7174–7182.
Goldberg, C.S., Strickler, K.M., Pilliod, D.S., 2015. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biol. Conserv. 183, 1–3.
Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Lebescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small SubUnit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41(DI, D597–D604.
Handy, S.M., Demir, E., Hutchins, D.A., Portune, K.J., Whereat, E.B., Hare, C.E., Rose, J.M., Warner, M., Farestad, M., Cary, S.C., Coyne, K.J., 2008. Using quantitative real-time PCR to study competition and community dynamics among Delaware Inland Bays harmful algae in field and laboratory studies. Harmful Algae 7 (5), 599–613.
Haywood, A.J., Steidinger, K.A., Truby, E.W., Bergquist, P.R., Bergquist, P.L., Adamson, J., Mackenzie, L., 2004. Comparative morphology and molecular phylogenetic analysis of three new species of the genus *Karenia* (Dinophyceae) from New Zealand. J. Phycol. 40 (1), 165–179.
Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M., 1996. Real time quantitative PCR. Genome Res. 6 (10), 986–994.
Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA – molecular evolution and phylogenetic iInference. Q. Rev. Biol. 66 (4), 411–453.
Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17 (8), 754–755.
Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30 (14), 3059–3066.
Kegel, J.U., Guillebault, D., Medlin, L.K., 2016. Application of microarrays (phylochips) for analysis of community diversity by species identification. Perspectives Phycol. 3 (2), 93–106.
Leftley, J.W., Hannah, F., 2009. A Literature Review of the Potential Health Effects of Marine Microalgae and Macroalgae. Environment Agency, United Kingdom (ISBN: 978-1-84911-134-8. 70pp).
Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.M., Decelle, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W.H., Mahe, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M.C., Romac, S., Shalchian-Tabrizi, K., Simon, N., Stoeck, T., Santini, S., Siano, R., Wincker, P., Zingone, A., Richards, T.A., de Vargas, C., Massana, R., 2014. Patterns of rare and abundant marine microbial eukaryotes. Curr. Biol. 24 (8), 813–821.
Medlin, L.K., 2013. Molecular tools for monitoring harmful algal blooms. Environ. Sci. Pollut. Res. 20, 6683–6685.

Nézan, E., Siano, R., Boulben, S., Six, C., Bilien, G., Cheze, K., Duval, A., Le Panse, S., Quere, J., Chomerat, N., 2014. Genetic diversity of the harmful family Kareniaceae (Gymnodiniales, Dinophyceae) in France, with the description of *Karlodinium gentienii* sp. nov.: a new potentially toxic dinoflagellate. Harmful Algae 40, 75–91.

Naustvoll, L., Dahl, E., Danielssen, D., Aure, J., Skogen, M., Budgell, P., 2002. Chattonella i Skagerrak –en ny trussel for oppdrettsnæringen? Havets Miljø Fisken og Havet særnummer 2, 126–129.

Park, T.-G., De Salas, M.F., Bolch, C.J.S., Hallegraeff, G., 2007. Development of a real-time PCR probe for quantification of the heterotrophic dinoflagellate *Cryptoperidiniopsis brodyi* (Dinophyceae) in environmental samples. Appl. Environ. Microbiol. 73 (8), 2552–2560.

Park, T.-G., Park, Y.-T., Lee, Y., 2009. Development of a SYTO9 based real-time PCR probe for detection and quantification of toxic dinoflagellate *Karlodinium veneficum* (Dinophyceae) in environmental samples. Phycologia 48 (1), 32–43.

Quince, C., Lanzen, A., Davenport, R.J., Turnbaugh, P.J., 2011. Removing noise from pyrosequenced amplicons. BMC Bioinf. 12, 38.

Rademaker, M., Reckermann, M., Tillmann, U., Tillmann-Mayer, A., Colijn, F., Zevenboom, W., Houpt, P., 1998. *Fibrocapsa japonica* and *Heterosigma akashiwo*: new observations. Harmful Algae News 8–10.

Rhodes, L., Smith, K., Moisan, C., 2013. Shifts and stasis in marine HAB monitoring in New Zealand. Environ. Sci. Pollut. Res. 20, 6872–6877.

Smith, K.F., de Salas, M., Adamson, J., Rhodes, L.L., 2014. Rapid and accurate identification by real-time PCR of biotoxin-producing dinoflagellates from the family Gymnnodinisceae. Mar. Drugs 12, 1361–1376.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W., Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. 19 (Suppl. 1), 21–31.

Throndsen, J., Hasle, G.R., Tangen, K., 2003. Norsk Kystplanktonflora. Almater Forlag, Oslo, pp. 341.

Throndsen, J., 1978. Preservation and storage. In: Sournia, A. (Ed.), Phytoplankton Manual. UNESCO, Paris, pp. 59–69.

Tillmann, U., Elbrächter, M., Krock, B., John, U., Cembella, A., 2009. *Azadinium spinosum* gen. et sp. nov. (Dinophyceae) identified as a primary producer of azaspiracid toxins. Eur. J. Phycol. 44 (1), 63–79.

Utermöhl, H., 1958. Zur Vervollkommung der quantitativen Phytoplankton-Methodik. Mitt. int. Verein. theor. angew. Limnol. 9, 1–38.

Vrieling, E.G., Koeman, R.P.T., Nagasaki, K., Ishida, Y., Peperzak, L., Gieskes, W.W.C., Veenhuis, M., 1995. *Chattonella* and *Fibrocapsa* (Raphidophyceae): novel, potentially harmful red tide organisms in Dutch coastal waters. Neth. J. Sea Res. 33, 183–191.

Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. 63 (10), 3741–3751.

Yuan, J., Mi, T., Zhen, Y., Yu, Z., 2012. Development of a rapid detection and quantification method of *Karenia mikimotoi* by real-time quantitative PCR. Harmful Algae 17, 83–91.

Zamor, R.M., Glenn, K.L., Hambright, K.D., 2012. Incorporating molecular tools into routine HAB monitoring programs: using qPCR to track invasive *Prymnesium*. Harmful Algae 15, 1–7.

Development of a qPCR assay to detect and quantify ichthyotoxic flagellates along the Norwegian coast, and the first Norwegian record of *Fibrocapsa japonica* (Raphidophyceae)

SUPLEMENTARY INFORMATION

Anette Engesmo[a], David Strand[b], Sandra Gran-Stadniczeñko[c], Bente Edvardsen[c], Linda K. Medlin[d], Wenche Eikrem[a,c]

[a] Norwegian Institute for Water Research, Gaustadallèen 21, 0349, Oslo, Norway
[b] Norwegian Veterinary Institute, P.O. box 750 Sentrum, 0106, Oslo, Norway
[c] University of Oslo, Department of Biosciences, P. O. box 1066 Blindern, 0316, Oslo, Norway
[d] Marine Biological Association of the UK, The Citadel, Plymouth, Pl1 2PB, United Kingdom

Table S1. Culturing conditions of algal strains used in this study.

Table S2: PCR and Sanger Sequencing primers used in this study.

Table S3: DNA concentration and purity of Oslofjorden field samples.

Table S1. Culturing conditions of algal strains used in this study.

Strain:	Synonym:	Species:	Class:	Media:	Salinity:	Temp:	Light cycle:	Light (photon flux):	Culture collection:
UIO019	SCCAP K-0260	*Karenia mikimotoi*	Dinophyceae	ES	25 PSU	17 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
SCCAP K-1274		*Karenia brevis*	Dinophyceae	K½	25 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
UIO254	SCCAP-1640	*Karlodinium veneficum*	Dinophyceae	IMR½	30 PSU	17 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
UIO297		*Karlodinium* cf. *veneficum*	Dinophyceae	ES	25 PSU	12 °C	12:12	100 mmol m^{-2} s^{-1}	UIO
SCCAP K-1471		*Alexandrium tamarense*	Dinophyceae	K½	25 PSU	16 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
SCCAP K-0675		*Levanderina fissa*	Dinophyceae	IMR½	30 PSU	16 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
SCCAP K-1332		*Gymnodinium catenatum*	Dinophyceae	K½	25 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
CCMP2088		*Polarella glacialis*	Dinophyceae	ES	30 PSU	4 °C	12:12	100 mmol m^{-2} s^{-1}	CCMP
UIO296	RCC2538	*Azadinium cf. Spinosum*	Dinophyceae	ES	34 PSU	16 °C	12:12	100 mmol m^{-2} s^{-1}	RCC
SCCAP K-1137		*Prorocentrum micans*	Dinophyceae	K½	25 PSU	16 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
UIO284		*Scrippsiella trochoidae*	Dinophyceae	ES	25 PSU	12 °C	12:12	100 mmol m^{-2} s^{-1}	UIO
UIO081		*Amphidinium carterae*	Dinophyceae	ES	25 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	UIO
RCC1502		*Heterosigma akashiwo*	Raphidophyceae	IMR½	30 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	RCC
ARC HA0309-2		*Heterosigma akashiwo*	Raphidophyceae	IMR½	30 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	ARC
SCCAP K-1549		*Heterosigma akashiwo*	Raphidophyceae	IMR½	30 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
RP02EHU		*Heterosigma akashiwo*	Raphidophyceae	ES	34 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	EHU
RCC1501		*Fibrocapsa japonica*	Raphidophyceae	IMR½	30 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	RCC
SCCAP-K0542		*Fibrocapsa japonica*	Raphidophyceae	IMR½	30 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	SCCAP
ARC CS0707-1		*Chattonella subsalsa*	Raphidophyceae	ES	34 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	ARC
ARC CA0800		*Chattonella antiqua*	Raphidophyceae	ES	34 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	ARC
ARC HD0110		*Haramonas dimorpha*	Raphidophyceae	ES	34 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	ARC
NIES-15		*Olisthodiscus luteus*	Incerta sedis	ES	25 PSU	19 °C	12:12	100 mmol m^{-2} s^{-1}	NIES
UIO109		*Pseudochattonella farcimen*	Dictyochophyceae	IMR½	22 PSU	16 °C	12:12	100 mmol m^{-2} s^{-1}	UIO

RCC : Roscoff Culture Collection http://roscoff-culture-collection.org
SCCAP: Scandinavian Culture Collection of Algae and Protozoa http://www.sccap.dk
UIO: University of Oslo Culture Collection
NIES: National Institute for Environmental Studies http://mcc.nies.go.jp/localeAction.do?lang=en
EHU: Culture Collection of the Department of Plant Biology and Ecology of the University of the Basque Country
ARC: CMS Algal Research Collection (Carmelo Tomas) http://www.algalresourcescollection.com
CCMP: National Center for Marine Algae and Microbiota https://ncma.bigelow.org/

Table S2: PCR and Sanger Sequencing primers used in this study

Primer:	rDNA:	Sequence:	Used:	Reference:
DIR-F	28S	ACC CGC TGA ATT TAA GCA TA	PCR and sequencing	Scholin *et al*. 1994[1]
D2C-R	28S	CCT TGG TCC GTG TTT CAA GA	PCR and Sequencing	Scholin *et al*. 1994[1]
1F	18S	AAC CTG GTT GAT CCT GCC AGT	PCR and sequencing	Medlin *et al*. 1988[2]
1528-R	18S	TGA TCC TTC TGC AGG TTC ACC TAC	PCR and sequencing	Medlin *et al*. 1988[2]
300R	18S	TCAGGCTCCCTCTCCGG	Sequencing	Elwood *et al*. 1985[3]
1055F	18S	GGTGGTGCATGGCCG	Sequencing	Elwood *et al*. 1985[3]
1055R	18S	CGGCCATGCACCACC	Sequencing	Elwood *et al*. 1985[3]
528F	18S	CGGTAATTCCAGCTCC	Sequencing	Elwood *et al*. 1985[3]
1400-F	ITS	TGYACACACCGCCCGTC	PCR and sequencing	Elwood *et al*. 1985[3]
ITS3-F	ITS	GCATCGATGAAGAACGCAGC	Sequencing	White *et al*. 1990[4]
4618-F	ITS	GTAGGTGAACCTGCAGAAGGATCA	PCR and sequencing	Bowers *et al*. 2006[5]
LSU1-R	ITS	ATATGCTTAAATTCAGCGGGT	PCR and sequencing	Bowers *et al*. 2006[5]

[1] Scholin, C.A., Herzog, M., Sogin, M., Anderson, D.M., 1994. Identification of group- and strain specific genetic markers for globally distributed *Alexandrium* (Dinophyceae) II: Sequence analysis of a fragment of the LSU rRNA gene. J. Phycol. 30(6): 999-1011.

[2]Medlin, L.K., Elwood, H.J., Stickel S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71: 491-499.

[3]Elwood, H.J., Olsen G.J., Sogin, M., 1985. The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Mol. Biol. Evol. 2: 399-410.

[4]White T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: A Guide to Methods and Applications (Ed. by M.A. Innis, D.H. Gelfand, J.J. Sninsky & T.J. White), pp 315-322. Academic Press, San Diego.

[5]Bowers, H. A., Tomas, C.R., Tengs, T., Kempton J.W., Lewitus, A.J., Oldach, D.W., 2006. Raphidophyceae (Chadefaud ex Silva) systematics and rapid identification: Sequnce analyses and real-time PCR assays. J. Phycol. 42: 1333-1348.

Table S3: DNA concentration and purity of Oslofjorden field samples.

Sample date:	Station:	DNA conc. (ng/µl):	Purity $A_{260}/_{A280}$:
Aug-09	OF2	107.82	1.54
Sep-09	OF2	95.21	1.51
Oct-09	OF2	105.16	1.28
Nov-09	OF2	115.32	1.46
Dec-09	OF2	90.58	1.46
Jan-10	OF2	91.40	1.40
Mar-10	OF2	52.87	1.51
Apr-10	OF2	65.82	1.50
May-10	OF2	137.34	1.50
Aug-10	OF2	59.11	1.52
Sep-10	OF2	53.60	1.65
Oct-10	OF2	67.91	1.62
Nov-10	OF2	81.28	1.34
Dec-10	OF2	44.70	1.75
Jan-11	OF2	29.51	1.62
Feb-11	OF2	65.50	1.69
Mar-11	OF2	68.38	1.73
Apr-11	OF2	74.88	1.69
May-11	OF2	38.83	1.67
Jun-11	OF2	133.25	1.49
Aug-11	OF2	51.48	1.98
Sep-11	OF2	13.22	1.87
Oct-11	OF2	11.18	3.10
Nov-11	OF2	24.44	2.01
Dec-11	OF2	13.53	1.90
Jan-12	OF2	15.81	1.88
Feb-12	OF2	17.60	1.87
Mar-12	OF2	26.75	1.70
Apr-12	OF2	24.73	1.83
May-12	OF2	33.69	1.82
Jun-12	OF2	22.35	1.79
Jul-15	ELLE	18.19	1.85
Jul-15	ELLE	25.09	1.95
Jul-15	ELLE	26.41	2.03
Jul-15	ELLE	38.37	1.93

PAPER IV

Seasonal dynamics of algal infecting viruses in the Skagerrak, North Atlantic and inferred interactions with protists.

Sandra Gran-Stadniczeñko[1], Anders K. Krabberød[1], Ruth-Anne Sandaa[2], Sheree Yau[3], Elianne Dunthorn Egge[1,4], Bente Edvardsen[1].

[1] Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway.

[2] Department of Biological Sciences, University of Bergen, 5020 Bergen, Norway.

[3] Integrative Biology of Marine Organisms Laboratory (BIOM), CNRS UMR7232, 66650 Banyuls-sur-Mer, France.

[4] University of Duisburg-Essen, Biology, 45141 Essen, Germany.

[*] Correspondence: bente.edvardsen@ibv.uio.no

ABSTRACT

Viruses are a highly abundant, dynamic and diverse component of the planktonic communities playing key roles in marine ecosystems. We aimed to reveal the diversity and dynamics of marine algal viruses in the Northern Skagerrak, South Norway through the year by metabarcoding and how this is correlated to the protist diversity and dynamics. The Major Capsid Protein (MCP) gene was extracted from the total DNA. A high diversity of algal viruses was revealed compared to previous metabarcoding surveys in Norwegian coastal waters. We obtained 313 putative algal infecting virus OTUs, all classified by phylogenetic analyses to either the *Phycodnaviridae* or *Mimiviridae* families, most of them in clades without any cultured or environmental reference sequences. The viral community showed a clear temporal variation, with some OTUs persisting for several months. There was a coexistence between virus and host during long periods. This study gives new insights into the virus-algal host dynamics and provides a baseline for future studies of algal virus diversity and temporal dynamics.

Keywords: *Phycodnaviridae*; *Mimiviridae*; metagenomics; Oslofjorden; viral-host co-occurrences.

INTRODUCTION

Virus-like particles (VLPs) are widely spread throughout the world's oceans, and considered to be highly abundant (5-15x10^6 viruses ml^{-1}), genetically diverse and dynamic components of the planktonic community (Bergh et al. 1989). As viruses follow the host abundance, total viral abundance is highest in productive coastal waters (~10^8 viruses ml^{-1}) decreasing off-shore and deeper in the water column (Suttle 2005, 2007; Boxshall 2006).

Viruses play a major role in marine ecosystems. They are substantial agents of mortality in marine microbial communities, influencing species distribution and abundance and maintain the coexistence of competing species (Thingstad and Lignell 1997; Thingstad 2000). Consequently, they affect the nutrient and carbon cycling by converting microbial biomass into dissolved and particulate organic matter (Bratbak et al. 1994) which heterotrophic prokaryotes and other degraders will make use of (Fuhrman 1999; Suttle 2005). Viruses are also linked to preventing and terminating algal blooms (Bratbak et al. 1995; Jacobsen et al. 1996; Castberg et al. 2001; Baudoux and Brussaard 2005). Indirectly, viruses will therefore affect the climate on the Earth by increasing dimethyl sulfide (DMS) emissions from phytoplankton by viral lysis (Danovaro et al. 2011). Yet, the role of marine viral communities in marine ecosystems is largely unknown and far from being understood.

New molecular methods have enabled investigations of the diversity of viruses and their hosts. Studies of algal viruses suggest that they may be species-specific or strain-specific to their host, but many different viruses can infect the same species (Brussaard 2004; Baudoux 2007; Johannessen et al. 2015). Algal viruses infecting different species have, however, also been described (Johannessen et al. 2015). Most algal viruses so far isolated are large double-stranded DNA (dsDNA), often referred to as giant viruses due to their large (>200 kb) genome size, with icosahedron viron structure. They are assigned to two families; the *Phycodnaviridae* (Dunigan et al. 2006; Brussaard and Martinez-Martinez 2008) and the *Mimiviridae* (Claverie and Abergel 2018). These two families, together with the *Asfarviridae*, *Iridoviridae*, and *Poxiviridae* make up a monophyletic group (Boyer et al. 2010) referred to as Nucleo-Cytoplasmic Large DNA Viruses (NCLDV, Iyer et al. 2001). All viruses within the *Phycodnaviridae* consists of viruses infecting algae, while members of the *Mimiviridae* infect both photosynthetic and non-photosynthetic protists. Viruses infecting photosynthetic protists all fall within one subcluster within the *Mimiviridae* family, e.g. viruses infecting the haptophytes *Phaeocystis pouchetii, Prymnesium kappa and Haptolina ericina* (Gallot-Lavallée et al. 2015; Johannessen et al. 2015), while those members infecting heterotrophic

protists form another subclade (Fischer et al. 2010; Claverie and Abergel 2018). Even though the taxon richness of protist-infecting viruses may exceed that of prokaryotes and archaea in the ocean (Mihara et al. 2018), very few genetically characterised taxa with a known host of these two families are available to date.

Virus-host interactions between algal viruses and their phytoplankton hosts have been described both as acute boom-bust infections, where a specific virus can lyse a dense bloom of host cells within hours, and as more persistent infections where viruses and hosts can co-exist (Johannessen et al. 2017; Sandaa and Bratbak 2018). The latter can be explained by viral resistance, immunity and/or strain specificity, or the virus becoming less virulent (Suttle and Chan 1994; Tarutani et al. 2001; Thyrhaug et al. 2003; Dimmock et al. 2016). One approach for measuring these interactions is by correlating viral diversity and abundances with host diversity and abundances to better understand the underlying mechanisms driving the dynamics. Algal virus diversity and dynamics within their natural habitats are increasingly studied by metabarcoding using e.g. the major capsid protein (MCP) as the marker gene (e.g. Johannessen et al. 2017).

Our aim was to reveal the diversity and dynamics of marine algal viruses in the Northern Skagerrak, South Norway through the year by metabarcoding and how this is correlated to the protist diversity and dynamics. We addressed the following questions: 1) Which dsDNA algal viruses are found in the Outer Oslofjorden by metabarcoding of the MCP gene? 2) How is the algal-virus community composition and relative abundances changing over the year? 3) How do the algal viruses co-occur with various co-existing protists? 4) Can co-occurrence analyses give new information about potential virus-algal host relationships?

MATERIAL AND METHODS

Sampling

Sea-water samples from the Outer Oslofjorden were collected at the OF2 monitoring station (59.17 N, 10.69 E) during 15 monthly sampling campaigns between March 2010 and June 2011. Niskin bottles (4x5L) attached to a CTD rosette were used to collect water samples from 1m depth (subsurface) for virus collection (20 L). The water samples were first filtrated through a 200 µm mesh sieve to remove zooplankton and then through a 0.45 µm pore-size, low-protein-binding, Durapore membrane filters of 142 mm in diameter (Millipore) mounted in a steel tripod (Millipore, Billerica, USA) by peristaltic pumping (Masterflex, Cole-Parmer,

IL, USA). Further, samples were concentrated (10 psig, high speed ~) to a final volume of 50 ml using the QuixStand benchtop system with hollow 100 000 pore size fibre cartridges (NMWC) and stored in sterile falcon tubes. Then, 10 x 250 µL of the virus-concentrate were transferred into sterile cryo-tubes, fast-frozen in liquid nitrogen and stored at -80^{o}C until further processing.

High-throughput sequencing

DNA extraction, purification and PCRs amplifications of a part of the Major Capsid Protein gene (MCP) were done as described in detail in Johannessen et al. (2017), with the exception that for PCR re-amplification 2.5 µl template was used. Four replicates from each sample were then pooled and purified with AMPure XP beads purification kit (Beckman Coulter, Brea, USA). A pooled sample of 75 µl with a concentration of 22.5 ng µl^{-1} DNA, quantified by Nanodrop, was sent for sequencing. Library preparation and Lib-A unidirectional amplicon sequencing was performed on 20 µl of the DNA sample (11 ng^{-1} µl) in a Roche 454 GS-FLX Titanium (Microsynth AG, Balgach, Switzerland). A total of 203229 viral reads were obtained.

Bioinformatic pipeline

Viral 454 reads were processed through the QIIME v.1.6 pipeline (Caporaso et al. 2010). AmpliconNoise v.1.6.1 and Perseus (Quince et al. 2011) were used to correct errors of the 454 reads and remove putative chimeras. Clustering of reads into OTUs was performed using UCLUST v1.2.22 (Edgar 2010) with 97% sequence similarity. A total of 582 OTUs were obtained based on 127348 reads. Putative spurious data (single singletons, non-algal virus OTUs and OTUs that could not be aligned, possibly non-MCP genes), were removed leaving a total of 313 OTUs. Subsampling was then performed to the minimum read number 1696 with the "rarefy" option in Vegan package in R (Oksanen et al. 2011).

Taxonomic classification and phylogenetic analyses

A first taxonomic assignation of the OTU nucleotide sequences was done with BLASTn against the Viral GenBank database in ViroBLAST with default parameters (Deng et al. 2007). In addition, a phylogenetic placement of OTUs was performed, based on the amino

acid (aa) sequence of the OTUs. Representative nucleotide sequences from the virus OTUs were translated into aa with GeneMark (Besemer and Borodovsky 1999, 2005; Zhu et al. 2010). When the aa sequence predicted by GeneMark was shorter than the expected length based on the nucleotide sequence (i.e. the aa sequence was less than one third the length of the nucleotide sequence) they were manually inspected. The manual inspection consisted in aligning the nucleotide sequence to the closest hit in GenBank and checking for mutations, insertions and deletions leading to premature stop codons or frameshifts in the predicted aa. These differences in the nucleotide sequence are most likely due to sequencing errors. Sequences that had wrongly inserted stop codons, or frameshifts were manually corrected and translated into aa sequence. The resulting aa sequences were aligned against a previously published reference alignment (Johannessen et al. 2017) of 15 *Mimiviridae* and 16 *Phycodnaviridae* viruses, the closest hits in ViroBLAST, and the top hits from non-redundant protein database in the NCBI database (O'Leary et al. 2016). The top hits from the non-redundant protein database was determined using blastp in Geneious version 10.2.3 (Kearse et al. 2012).

Alignment constructing started with the sequences longer than 200 aa using MAFFT L-ins-i (Katoh and Standley 2013, 2016). The shorter fragments were then added to the reference alignment using the *addfragment* algorithm that takes shorter sequences and adds it to an alignment, keeping the gaps and the relative position of the characters in the original alignment (Katoh and Frith 2012). Ambiguously aligned positions were removed with trimAl using the *gappyout* setting (Capella-Gutierrez et al. 2009). This alignment consisted of 403 virus sequences. A second alignment consisting of only the 22 most abundant virus OTUs and the sequences of 15 *Mimiviridae* and 16 *Phycodnaviridae* from Johannesen et al. (2017) was constructed with the same approach as the larger dataset. Phylogenetic trees were constructed for both alignments with FastTree2 (Price et al. 2010), implemented in Geneious 10.2.3 (Kearse et al. 2012) and visualized in FigTree (Rambaut 2016).

Network construction

Association networks were constructed to analyse the co-occurrence of algae and viruses. The dataset was prepared for network construction after first filtering OTUs not present in at least two of the samples. The dataset consisted of 342 protist OTUs (Gran-Stadniczeñko et al. 2018), and 42 virus-OTUs. Co-occurrence networks were constructed using *SparCC* as implemented in the R package *SpiecEasi* (Kurtz et al. 2015). *SparCC* was run with default settings and with 500 bootstraps. Two-sided pseudo p-values were calculated for both datasets

independently, and edges with p-value > 0.05 and correlation score < |0.5| were deleted. Protist-protist associations were also deleted, since we were primarily interested in the associations between viruses and protists. Visualization of networks were done with *Cytoscape* v3.6.1(Su et al. 2014).

RESULTS

We assessed the *Phycodnaviridae* and *Mimiviridae* taxonomic composition, OTU richness and relative abundance through the year by high-throughput amplicon sequencing (metabarcoding) in 15 monthly samples from the outer Oslofjorden, Skagerrak, Southern Norway. The conserved MCP gene was amplified from all samples and resulted in 126775 reads of 200-400 bp length, representing 313 putative algal infecting virus OTUs after filtration, 97% similarity clustering, removal of singletons and translation into amino acids (aa). After subsampling to the lowest number of reads per sample (1696 reads), to be able to compare samples, 243 OTUs remained. None of the rarefaction curves for the 15 samples reached a plateau, indicating that the sequencing depth was not sufficient to capture the full diversity at outer Oslofjorden (Fig. S1). The most abundant OTU (OTU 1) contained 53% of all reads. Twenty-two OTUs had more than 50 reads (0.2% of total reads per OTU) conforming 95% of the total reads and were considered the most abundant OTUs, whereas 208 OTUs had each less than 0.1% of the reads (Table S1).

Virus diversity

Of the 313 OTUs, 95 had ≥ 90% similarity to reference sequences obtained by BLASTn against Viral Genbank on the ViroBLAST platform (Table S1). Of these, 36 had best match with a cultured viral sequence. The putative hosts were two haptophytes (*Haptolina ericina* and *H. hirta*) and four prasinophytes (*Micromonas pusilla*, *Micromonas* sp. (strain RCC1109), *Ostreococcus lucimarinus* and *O. mediterraneus*). Fifty-nine OTUs had best match to uncultured environmental MCP viral sequences. The rest had lower similarity to cultured or uncultured reference sequences, meaning they had not been found elsewhere by metabarcoding.

Phycodnaviridae and *Mimiviridae* phylogenetic trees based on MCP amino acid OTU sequences were constructed including all 313 OTUs (Fig. S2, File S1), and the 22 most abundant OTUs (>50 reads per OTU, Fig. 1, File S2). The table with accession numbers of

reference sequences is presented in Table S2. Fifty-four OTUs were taxonomically placed within the *Mimiviridae* and 259 within the *Phycodnaviridae* families. Several clades in both families did not cluster with any reference sequence, and some OTUs clustered with uncultured environmental MCP sequences only (Fig. S2, File S1). We found that several uncultured environmental reference sequences placed in the *Mimiviridae* clade were originally classified as *Phycodnaviridae*. Fifteen of the 22 most abundant OTUs were placed in the *Phycodnaviridae* family and seven in *Mimiviridae* (Fig. 1). The two most abundant *Phycodnaviridae*, OTU 1 and 3 clustered together with OTU 7, OTU 10 and OTU 20, close to an environmental OTU from Puddefjorden (Western Norway, OTU/P0604, Larsen et al. 2008). They were also placed in the same major clade as OTU 9 and the sequence of a virus infecting the picoplankton chlorophyte *Micromonas pusilla* (MpV1) with a bootstrap value of 84%. Three OTUs (OTU 14, 21 and 22) clustered with the environmental OTU/M0501 from Raunefjorden (Western Norway , Larsen et al. 2008) and a virus infecting *Haptolina hirta* (HhV-Of01) with high bootstrap values (>80%). Further, OTU 12 and 15 clustered with the sequences from a virus infecting *Ostreococcus tauri* (OtV1 165). The most abundant *Mimiviridae* OTUs were OTU 2 and 4. The OTU 4 clustered with *Chrysochromulina ericina* virus (CeV, infecting *Haptolina ericina*) with high bootstrap support (95%), whereas OTU 2 did not cluster with any known viral reference sequence. Further, a new branch consisting of two OTUs (OTU 5 and 16) was obtained as sister to the *Mimiviridae* family (Fig. 1).

Temporal variation

Virus richness was highest from September to November 2010 and lowest in August 2010 (~55 and 9 OTUs respectively, Fig. 2a and Fig. S3) whereas Shannon diversity index was highest in April 2010 and lowest in August 2010 (2.06 and 0.07 respectively, Fig. S3). The *Phycodnaviridae* family was more diverse than the *Mimiviridae* in all samples (Fig. 2a), but *Mimiviridae* dominated in abundances in March, August and September 2010, and was almost neglectable from April 2010 to June 2010 and from December 2010 to April 2011 (Fig. S4). Six OTUs were present in ten or more samples (OTU 1, 3, 7, 12, 13 and 15), and four of the 22 most abundant ones were present in only one sample (OTU 8, 11, 22, 30, Table S1, Fig. 2b). The *Phycodnaviridae* OTU 1, which clustered close to MpV1, dominated for eight months from October 2010 to May 2011 (Fig. 2b). The *Phycodnaviridae* OTU 3, also clustering close to MpV1 dominated in May and June 2010 and in June 2011 samples. The

Mimiviridae OTUs 2 and 4 dominated in the March, August and September 2010 samples and were rare or absent in the other samples.

Hierarchical clustering of samples based on Bray-Curtis dissimilarities showed three well-distinguished clusters. The August and September 2010 samples significantly differed from the other 13 ones (Fig. 3). All samples between December 2010 and May 2011 formed a cluster with highly similar samples, which may be explained by the strong dominance of OTU 1 in these samples (Fig. 2b).

Temporal variation of virus and their potential host

Co-occurrence network analyses were done to detect positive and negative protists-virus correlations between our protist (Gran-Stadniczeñko et al. 2018) and virus datasets (Fig. 4). Networks with correlation score > |0.5| and $p<0.05$) showed no clear positive or negative patterns for *Phycodnaviridae* and *Mimiviridae* with protist OTUs. The *Phycodnaviridae* OTU 1, which clustered with a virus infecting the chlorophyte *Micromonas pusilla* (MpV1), had negative correlations with two chlorophyte OTUs that had best match to *Pycnoccocus provasolii*. (OTU 58) and *Pyramimonas* sp. (OTU 209), together with two dinoflagellate OTUs (Fig. 4, Table 1), and positive correlation with the most abundant protist, *Karenia papillonaceae* (OTU 1). The *Mimiviridae* OTU 2, that did not cluster with a known reference sequence had positive correlations with a picobilizoan (heterotrophic picoplankton) and a negative correlation with an uncultured dinoflagellate. The *Phycodnaviridae* OTU 3 only presented negative correlations with two ciliates, one diatom and one chrysophyte. OTU 4, which clustered close to *Chrysocromulina ericina* virus CeV, did not show any significant correlation with a haptophyte OTU, but showed positive co-occurrences with two diatoms and a choanoflagellate (heterotrophic opisthokont protist). Also, the *Phycodnaviridae* OTU 7, showed negative correlations with a MAST (uncultured heterotrophic stramenopile) and the haptophyte *Emiliania huxleyi,* and negative correlations to two uncultured alveolates and a choanoflagellate. The remaining *Phycodnaviridae* and *Mimiviridae* OTUs did not show significant correlation with a potential viral protist host.

The seasonal dynamics of protists at the Outer Oslofjorden OF2 sampling campaigns was described by Gran-Stadniczeñko et al. (2018). In Fig. 5 temporal dynamics of the four most abundant virus OTUs with that of their co-occurring protists are shown. The patterns observed explain the results in Fig. 4. Protist OTUs that had positive correlations with the viruses showed similar temporal distribution, whereas protist OTUs with negative correlations

showed an opposite temporal pattern to those of the virus. Also, for *Phycodnaviridae* OTU 1 and OTU 3, the protists that showed negative co-occurrences with the virus, had a similar temporal variation.

We also combined the distribution pattern of the most abundant virus OTUs that were assigned to a cultured virus with a known host, with that of their putative protist host (Fig. 6). Relative abundances over time of virus OTU 1, 3 and 9 were compared to that of the possible host *Micromonas commoda* OTU 5 and *Micromonas* sp. clade-B-subarctic OTU 22 (the two most abundant *Micromonas* OTUs in Gran-Stadniczeñko et al. (2018), in addition to *Micromonas pusilla* OTU 107 and *Micromonas bravo* OTU 328. Also, virus OTUs 2 and 4 assigned to the *Mimiviridae* CeV (with the host *Haptolina ericina*) was compared to the dynamics of the haptophyte OTU 44 assigned to *Haptolina* sp. A temporal variation for all virus and protist OTUs was observed. For *M. commoda*, *M.* sp. clade-B-subarctic, *M. pusilla* and *Haptolina* sp., the virus OTUs occurrence occasionally overlapped with that of the hosts' and peaked after a drop of the hosts' abundances, whereas no overlap between abundances of *Micromonas bravo* and MpV1 was observed.

DISCUSSION

Viral diversity

This is the first metabarcoding study on algal viruses from the Skagerrak and the Oslofjorden. Here we have studied viruses in seawater passing a 0.45 μm pore-size filter, which include both viruses in the water mass and within host cells (< 200 μm) that might have been disrupted during filtration. A relatively high diversity of algal viruses was revealed compared to previous metabarcoding studies in Norwegian coastal waters (Johannessen et al. 2017). However, in the latter study the clustering level 95% was used. Based on the MCP gene we obtained 313 OTUs, all classified by phylogenetic analyses to either the *Phycodnaviridae* or *Mimiviridae* families with high bootstrap support (>60%). Twelve percent (36 OTUs) had best match with Blast to a cultured algal virus, whereas 70% (218 OTUs) had 89% or lower similarity to any available virus sequences. Also, by phylogenetic placement, the majority of the OTUs formed clades without any reference sequence. This is in accordance with previous studies by e.g. Clerissi et al. (2015), pinpointing the lack of characterised algal-infecting viruses and the need for more cultured and characterised reference strains. Also, several of the uncultured *Phycodnaviridae* reference sequences were placed in the *Mimiviridae* clade:

ABU23699, ABU23704 (Larsen et al. 2008), AHN92263, AHN92262, AHN92288, AHN92249 (Rozon and Short 2013) and AGI16567 (Zhong and Jacquet 2014). These misclassifications were also pointed out by Claverie and Abergel (2018). These authors further recommend the separation of different genera composing the *Phycodnaviridae* into different families. All algal viruses were initially placed in the family *Phycodnaviridae* (e.g. Larsen et al 2008), which explains some of the misclassifications. The taxonomy of algal viruses is at present under revision in ICTV. A revision of the taxonomy of algal virus reference sequences in gene databases will therefore be needed once a new approved classification is available.

Of the 22 most abundant viral OTUs seven were assigned (by phylogeny and BLAST, ≥90% similarity) to cultured marine viruses infecting different protist taxa such as *Micromonas pusilla* virus (MpV1), *Haptolina hirta* virus (HhV-Of01), *Ostreococcus* spp. viruses (OtV, OlV, and OmV viruses) and *Chrysochromulina ericina* viruses (CeV, *Haptolina ericina* was previously named *Chrysochromulina ericina)*. Members of the genera *Micromonas* and *Haptolina* are common phytoplankton species in the Oslofjorden (Throndsen et al. 2007; Johannessen et al. 2017; Gran-Stadniczeñko et al. 2018). *Micromonas pusilla* was until recently the only described species of this genus, but was divided into several species by Simon et al. (2017). *Micromonas commoda* was the most abundant *Micromonas* species in Outer Oslofjorden OF2 station, but *M. pusilla*, *M. bravo* and *M*. sp clade-B-subarctic were also detected (Gran-Stadniczeñko et al. 2018). We found several OTUs assigned to MpV1, which may turn out to have different *Micromonas* species as their host. The picoplanktonic prasinophyte genus *Ostreococcus* has a wide distribution in temperate to tropical marine waters, but was not recorded by metabarcoding north of 60^{o}N by Tragin and Vaulot (2018). The virus OTUs 12 and 15 were in our phylogenetic analysis clustering with the cultured and genome characterised strain OtV-165, isolated from the English Channel and infecting *Ostreococcus tauri,* strain OTH95 (Weynberg et al. 2009). With BLAST, however, the OTUs 12 and 15 had best match to OmV1 virus infecting *Ostreococcus mediterraneus*. Due to this incongruence, we classified OTU 12 and 15 as *Ostreococcus* sp. virus (Table S1). Both *O. tauri* and *O. mediterraneus* were recorded for the first time from Outer Oslofjorden station OF2 by Gran-Stadniczeñko et al. (2018), and are thus potential hosts of these viruses, despite that they previously have been found mostly in warmer waters (Tragin and Vaulot 2018).

Viruses infecting *Emiliania huxleyi* (EhV) are a divergent group within the *Phycodnaviridae* family. *Emiliania huxleyi* was the most abundant haptophyte in the Oslofjorden OF2 metabarcoding survey (Egge et al. 2015; Gran-Stadniczeñko et al. 2018). None of our

OTUs had, however, best match to a virus infecting this alga as the primer-pairs used here do not target the MCP gene of EhV (Larsen et al. 2008).

Finally, several dsDNA viruses infecting diverse protists taxa have been characterised such as *Heterocapsa circularisquama* virus (Takano et al. 2018), *Heterocapsa pygmaeaea* (Kim et al. 2012), *Heterosigma akashiwo* virus (Nagasaki and Yamaguchi 1997) and *Chaetoceros salsuginemum* virus (Nagasaki et al. 2005). However, no reference sequences for the MCP region are available for these viruses, possibly preventing a better characterisation of protist-infecting viruses in our study.

Succession of viral OTUs and potential hosts

The viral community at the Oslofjorden showed a clear temporal variation, but not a recurring seasonal pattern as has been demonstrated in other studies (e.g. Pagarete et al. 2013). The lack of seasonality could be due to methodological limitations such as primer pairs targeting only a small fraction of the algal virus population or PCR-biases towards amplification of certain genotypes. Too low sampling frequency is also to be considered, as viral communities change fast (e.g. due to high virus decay rate), and thus boom-bust episodes may be overlooked. Also, the fact that some protist OTUs are not observed in two consecutive samples could be due to different abiotic or biotic factors (e.g. grazing or nutrient availability), and not to viral lysis. Therefore, we may need closer sampling intervals (e.g. weekly or even daily) to catch a direct link between host-virus. Another explanation may be the interannual variation of the host community structure. Algal viruses may require a certain density of the host for infection which may vary from year to year (see review by Short 2012).

The newly described species *Micromonas commoda,* recently separated from *Micromonas pusilla* (Simon et al. 2017), showed highest relative abundance during May 2010 and then decreased abruptly in the following months (Gran-Stadniczeñko et al. 2018). Our temporal comparisons with potential hosts showed that the virus OTU 1 (clustering with MpV1) had an opposite pattern compared to *M. pusilla*, which decreased in abundance from September to October 2010, at the same time as the virus OTU 1 was gradually increasing in abundance. The MpV1-like OTU 1 was the most abundant virus OTU and dominated for eight months. Also, OTU 3 had an affinity to MpV1 and was among the most abundant OTUs from March to June, and increased in abundance when the potential host *M. commoda* decreased. In a previous study of *Micromonoas pusilla* virus (MpV) in the Skagerrak -Kattegat area, Sahlsten (1998) similarly found the highest abundance during spring and lasting for

several months. Zingone et al. (1999) found likewise the *Micromonas* infecting virus MpV to persist in the water for a long period. Also, Zingone et al. (2006) found that some strains of this algae were resistant to viral infection. A coexistence between virus and host during long periods is the most observed pattern in our study, as most of our virus OTUs are present at several timepoints (see. fig. 2b, 5 and 6). This can be explained by long virulence periods were the virus coexists with its host, infecting just a part of the population and thus ensuring virus proliferation. This persistent relationship is in agreement with the evolution theory on virus-host interactions (see Opinion by Sandaa and Bratbak 2018). According to this theory, any ecologically successful parasite will ensure host survival rather than mortality and viruses may hence be expected to evolve from causing acute infections with high mortality towards less virulent and more persistent infections. Assuming that this theory is correct, most virus-host interactions should be persistent since they have evolved over a long period (Goic and Saleh 2012; Longdon et al. 2014). Another explanation would be that these viruses might infect several similar or even different hosts, allowing them to proliferate on different host species (Johannessen et al. 2015).

In contrast, the most abundant *Mimiviridae* OTU 2 and 4 appeared to dominate during shorter periods or specific time-points. This pattern is generally described for lytic viruses and suggests that these *Mimiviridae* viruses have a boom-bust relationship with their host, causing acute infections and killing nearly all host cells. This viral-host interaction can be observed once or twice during the study period. Johannessen et al. (2017) found a similar temporal distribution pattern with some algal virus OTUs highly abundant only at specific time-points, although most viral OTUs were persistent. Most studies of algal viruses and their hosts have focused on the boom-bust system, probably as these are easier to detect.

Finally, the temporal distribution in abundance of co-occurring virus-protists (Fig. 5) explains our network results. Virus OTU 1 (MpV1), for instance, had a positive relationship with *Karenia papillonaceae* in addition to four negative correlations to photosynthetic taxa (two chlorophytes and two dinoflagellates) that varied similarly through the year. These positive/negative correlations to very diverse protist groups do not necessarily imply that these represent viruses able to infect such diverse hosts, but that they may influence one host group, consequently changing the growth conditions for another group, or the protist groups show similar or opposite responses to other environmental factors than viruses.

Concluding remarks

All our viruses could be classified to either *Phycodnaviridae* or *Mimiviridae* based on their major capsid protein amino acid sequence. *Phycodnaviridae* showed both highest richness and relative abundance. We obtained 313 viral OTUs and a majority (70% at ≥90% similarity) did not match with any cultured or environmental sequences available in gene databases. Twelve percent of the virus OTUs had affiliation to a known cultured virus with either a putative prasinophyte or haptophyte algal host. These hosts have all been found in the outer Oslofjorden. The viral dynamics show clear temporal variation, with some OTUs persisting for several months, suggesting a co-existence with their host and others present at specific time points, suggesting a possible boom and bust relationship with their host. The co-occurrence analysis could not with any certainty reveal new host-relationships, but suggests some relationships to be examined in future studies. The comparison between the relative abundance of viruses and of their potential host can give new insight into the virus-algal host dynamics and the ecological role of algal viruses. This study also provides a baseline for future studies of algal virus diversity and temporal dynamics. With more MCP reference sequences available, more of the viruses reported here may be connected to a host and contribute to a better understanding of the ecological importance of algal viruses and their distribution in space and time.

Acknowledgements

Financial support was given by the University of Oslo, Faculty of Mathematics and Natural Sciences to SGS and BE, by the Research Council of Norway through grant 190307 HAPTODIV to BE, EDE, and WE, by grant 225956/E10 MIKROPOLAR to EDE and BE, and by the EU project BioMarKs (2008-6530, ERA-net Biodiversa, EU) to BE and RL. We thank Rita Amundsen, captain Sindre Holm and the crew onboard R/V Trygve Braarud for assistance during sampling and Colomban de Vargas for leading the BioMArks project. High-throughput sequencing was performed Microsynth AG, Balgach, Switzerland.

REFERENCES

Baudoux A. C. 2007. The role of viruses in marine phytoplankton mortality. Royal Netherlands Institute for Sea Research (NIOZ). 148 p.

Baudoux A. C. & Brussaard C. P. D. 2005. Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. *Virology*, **341**:80–90.

Bergh Ø., Børsheim K. Y., Bratbak G. & Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature*, **340**:467–468.

Besemer J. & Borodovsky M. 1999. Heuristic approach to deriving models for gene finding. *Nucleic Acids Research*, **27**:3911–3920.

Besemer J. & Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*, **33**:W451–W454.

Boxshall G. A. 2006. La vida en suspensión: El plancton. *In*: Duarte C. M. (ed.), La exploración de la biodiversidad marina: Desafíos científicos y tecnológicos. Bilbao, Fundación BBVA. p. 93–115.

Boyer M., Madoui M. A., Gimenez G., La Scola B. & Raoult D. 2010. Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses.Martin D. P. (ed.). *PLoS ONE*, **5**:e15530.

Bratbak G., Levasseur M., Michaud S., Cantin G., Fernández E., Heimdal B. & Heldal M. 1995. Viral activity in relation to *Emiliania huxleyi* blooms:a mechanism of DMSP release? *Marine Ecology Progress Series*, **128**:133–142.

Bratbak G., Thingstad F. & Heldal M. 1994. Viruses and the microbial loop. *Microbial Ecology*, **28**:209–221.

Brussaard C. P. D. 2004. Viral Control of Phytoplankton Populations-a Review. *The Journal of Eukaryotic Microbiology*, **51**:125–138.

Brussaard C. P. D. & Martinez-Martinez J. 2008. Algal bloom viruses. *Plant Viruses*, **1**:1–3.

Capella-Gutierrez S., Silla-Martinez J. M. & Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**:1972–1973.

Caporaso J. G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F. D., Costello E. K., Fierer N., Gonzalez Peña A., Goodrich J. K., Gordon J. I., Huttley G. A., Kelley S. T., Knights D., Koenig J. E., Ley R. E., Lozupone C. A., McDonald D., Muegge B. D., Pirrung M., Sevinsky J. R., Turnbaugh P. J., Walters W. A., Widmann J., Yatsunenko T., Zaneveld J. & Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**:335–336.

Castberg T., Larsen A., Sandaa R., Brussaard C., Egge J., Heldal M., Thyrhaug R., van Hannen E. & Bratbak G. 2001. Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Marine Ecology Progress Series*, **221**:39–46.

Claverie J.-M. & Abergel C. 2018. *Mimiviridae*: An Expanding Family of Highly Diverse Large dsDNA Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. *Viruses*, **10**:506.

Clerissi C., Desdevises Y., Romac S., Audic S., de Vargas C., Acinas S. G., Casotti R., Poulain J., Wincker P., Hingamp P., Ogata H. & Grimsley N. 2015. Deep sequencing of amplified *Prasinovirus* and host green algal genes from an Indian Ocean transect reveals interacting trophic dependencies and new genotypes: Marine algae and their viruses in the Indian Ocean. *Environmental Microbiology Reports*, **7**:979–989.

Danovaro R., Corinaldesi C., Dell'Anno A., Fuhrman J. A., Middelburg J. J., Noble R. T. & Suttle C. A. 2011. Marine viruses and global climate change. *FEMS Microbiology Reviews*, **35**:993–1034.

Deng W., Nickle D. C., Learn G. H., Maust B. & Mullins J. I. 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*, **23**:2334–2336.

Dimmock N. J., Easton A. J. & Leppard K. N. 2016. Introduction to Modern Virology. 7th ed. Wiley-Blackwell.

Dunigan D. D., Fitzgerald L. A. & Van Etten J. L. 2006. Phycodnaviruses: A peek at genetic diversity. *Virus Research*, **117**:119–132.

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

Egge E. S., Johannessen T. V., Andersen T., Eikrem W., Bittner L., Larsen A., Sandaa R. A. & Edvardsen B. 2015. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol Ecol*, **24**:3026–3042.

Fischer M. G., Allen M. J., Wilson W. H. & Suttle C. A. 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proceedings of the National Academy of Sciences*, **107**:19508–19513.

Fuhrman J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature*, **399**:541–548.

Gallot-Lavallée L., Pagarete A., Legendre M., Santini S., Sandaa R.-A., Himmelbauer H., Ogata H., Bratbak G. & Claverie J.-M. 2015. The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting *Haptolina* (*Chrysochromulina*) *ericina* (Prymnesiophyceae). *Genome Announcements*, **3**:e01413-15.

Goic B. & Saleh M.-C. 2012. Living with the enemy: viral persistent infections from a friendly viewpoint. *Current Opinion in Microbiology*, **15**:531–537.

Gran-Stadniczeñko S., Dunthorn Egge E., Hostyeva V., Logares R., Eikrem W. & Edvardsen B. 2018. Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as Revealed by Metabarcoding and Microscopy. *Journal of Eukaryotic Microbiology*. doi: 10.1111/jeu.12700.

Iyer L. M., Aravind L. & Koonin E. V. 2001. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *Journal of Virology*, **75**:11720–11734.

Jacobsen A., Bratbak G. & Heldal M. 1996. Isolation and characterization of a virus infecting Phaeocystis pouchetii (Prymnesiophyceae). *Journal of Phycology*, **32**:923–927.

Johannessen T., Larsen A., Bratbak G., Pagarete A., Edvardsen B., Egge E. & Sandaa R.-A. 2017. Seasonal Dynamics of Haptophytes and dsDNA Algal Viruses Suggest Complex Virus-Host Relationship. *Viruses*, **9**:84.

Johannessen T. V., Bratbak G., Larsen A., Ogata H., Egge E. S., Edvardsen B., Eikrem W. & Sandaa R. A. 2015. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology*, **476**:180–188.

Katoh K. & Frith M. C. 2012. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*, **28**:3144–3146.

Katoh K. & Standley D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**:772–780.

Katoh K. & Standley D. M. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*, **32**:1933–1942.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P. & Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**:1647–1649.

Kim J., Kim C.-H., Takano Y., Jang I.-K., Kim S. W. & Choi T.-J. 2012. Isolation and Physiological Characterization of a New Algicidal Virus Infecting the Harmful Dinoflagellate *Heterocapsa pygmaea*. *The Plant Pathology Journal*, **28**:433–438.

Kurtz Z. D., Müller C. L., Miraldi E. R., Littman D. R., Blaser M. J. & Bonneau R. A. 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks.von Mering C. (ed.). *PLOS Computational Biology*, **11**:e1004226.

Larsen J. B., Larsen A., Bratbak G. & Sandaa R. A. 2008. Phylogenetic analysis of members of the *Phycodnaviridae* virus family, using amplified fragments of the major capsid protein gene. *Appl Environ Microbiol*, **74**:3048–3057.

Longdon B., Brockhurst M. A., Russell C. A., Welch J. J. & Jiggins F. M. 2014. The Evolution and Genetics of Virus Host Shifts.Hobman T. C. (ed.). *PLoS Pathogens*, **10**:e1004395.

Mihara T., Koyano H., Hingamp P., Grimsley N., Goto S. & Ogata H. 2018. Taxon Richness of "Megaviridae" Exceeds those of Bacteria and Archaea in the Ocean. *Microbes and Environments*, **33**:162–171.

Nagasaki K., Tomaru Y., Takao Y., Nishida K., Shirai Y., Suzuki H. & Nagumo T. 2005. Previously unknown virus infects marine diatom. *Appl Environ Microbiol*, **71**:3528–3535.

Nagasaki K. & Yamaguchi M. 1997. Isolation of a virus infectious to the harmful bloom causing microalga Heterosigma akashiwo (Raphidophyceae). *Aquatic Microbial Ecology*, **13**:135–140.

Oksanen J., Blanchet F. G., Kindt R., Legendre P., O'Hara R. B., Simpson G. L., Solymos P., Stevens M. H. H. & Wagner H. 2011. vegan: Community Ecology Package. R package version 1.17-2. Vienna, R Foundation for Statistical Computing.

O'Leary N. A., Wright M. W., Brister J. R., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse B., Smith-White B., Ako-Adjei D., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell C. M., Goldfarb T., Gupta T., Haft D., Hatcher E., Hlavina W., Joardar V. S., Kodali V. K., Li W., Maglott D., Masterson P., McGarvey K. M., Murphy M. R., O'Neill K., Pujar S., Rangwala S. H., Rausch D., Riddick L. D., Schoch C., Shkeda A., Storz S. S., Sun H., Thibaud-Nissen F., Tolstoy I., Tully R. E., Vatsan A. R., Wallin C., Webb D., Wu W., Landrum M. J., Kimchi A., Tatusova T., DiCuccio M., Kitts P., Murphy T. D. & Pruitt K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, **44**:D733–D745.

Pagarete A., Chow C.-E. T., Johannessen T., Fuhrman J. A., Thingstad T. F. & Sandaa R. A. 2013. Strong Seasonality and Interannual Recurrence in Marine Myovirus Communities. *Applied and Environmental Microbiology*, **20**:6253–6259.

Price M. N., Dehal P. S. & Arkin A. P. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.Poon A. F. Y. (ed.). *PLoS ONE*, **5**:e9490.

Quince C., Lanzen A., Davenport R. J. & Turnbaugh P. J. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**:38.

Rambaut A. 2016. FigTree v1.4.3.

Rozon R. M. & Short S. M. 2013. Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. *Freshwater Biology*, **58**:2648–2663.

Sahlsten E. 1998. Seasonal abundance in Skagerrak-Kattegat coastal waters and host specificity of viruses infecting the marine photosynthetic flagellate *Micromonas pusilla*. *Aquatic Microbial Ecology*, **16**:103–108.

Sandaa R.-A. & Bratbak G. 2018. Is the Virus Important? And Some Other Questions. *Viruses*, **10**:442.

Short S. M. 2012. The ecology of viruses that infect eukaryotic algae: Algal virus ecology. *Environmental Microbiology*, **14**:2253–2271.

Simon N., Foulon E., Grulois D., Six C., Desdevises Y., Latimier M., Le Gall F., Tragin M., Houdan A., Derelle E., Jouenne F., Marie D., Le Panse S., Vaulot D. & Marin B. 2017. Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M.*

pusilla (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and. *Protist*, **168**:612–635.

Su G., Morris J. H., Demchak B. & Bader G. D. 2014. Biological Network Exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, **47**:8.13.1-8.13.24.

Suttle C. A. 2005. Viruses in the sea. *Nature*, **437**:356–361.

Suttle C. A. 2007. Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology*, **5**:801–812.

Suttle C. A. & Chan A. M. 1994. Dynamics and Distribution of Cyanophages and Their Effect on Marine *Synechococcus* spp. *Applied and Environmental Microbiology*, **60**:3167–3174.

Takano Y., Tomaru Y. & Nagasaki K. 2018. Visualization of a Dinoflagellate-Infecting Virus HcDNAV and Its Infection Process. *Viruses*, **10**:554.

Tarutani K., Nagasaki K., Itakura S. & Yamaguchi M. 2001. Isolation of a virus infecting the novel shellfish-killing dinoflagellate *Heterocapsa circularisquama*. *Aquatic Microbial Ecology*, **23**:103–111.

Thingstad T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, **45**:1320–1328.

Thingstad T. & Lignell R. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, **13**:19–27.

Throndsen J., Hasle G. R. & Tangen K. 2007. Phytoplankton of Norwegian Coastal Waters. Oslo, Almater Forlag AS.

Thyrhaug R., Larsen A., Thingstad T. F. & Bratbak G. 2003. Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series*, **254**:27–35.

Tragin M. & Vaulot D. 2018. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. doi: 10.1101/449298.

Weynberg K. D., Allen M. J., Ashelford K., Scanlan D. J. & Wilson W. H. 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environmental Microbiology*, **11**:2821–2839.

Zhong X. & Jacquet S. 2014. Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes. *Environmental Microbiology*, **16**:759–773.

Zhu W., Lomsadze A. & Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, **38**:e132–e132.

Zingone A., Natale F., Biffali E., Borra M., Forlani G. & Sarno D. 2006. Diversity in morphology, infectivity, molecular characteristics and induced host resistance between two viruses infecting *Micromonas pusilla*. *Aquatic Microbial Ecology*, **45**:1–14.

Zingone A., Sarno D. & Forlani G. 1999. Seasonal dynamics in the abundance of *Micromonas pusilla* (Prasinophyceae) and its viruses in the Gulf of Naples (Mediterranean Sea). *Journal of Plankton Research*, **21**:2143–2159.

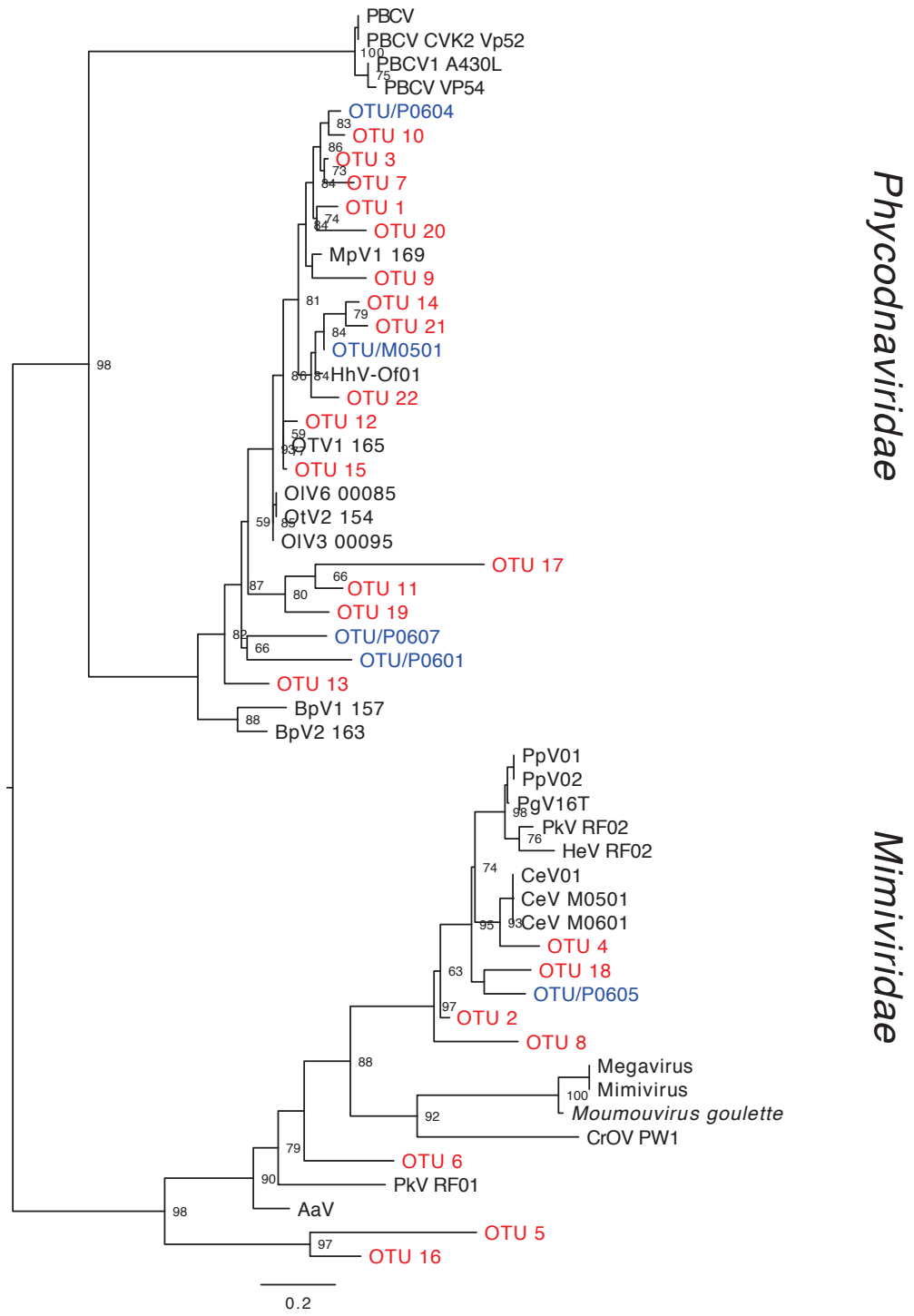


Figure 1. Phylogenetic tree based on available *Phycodnaviridae* and *Mimiviridae* MCP reference sequences (black), virus OTUs found in a previous study in Raunefjorden (blue, Larsen et al. 2008) and the 22 most abundant OTU algal viruses (>50 reads) from Outer Oslofjorden (red). Bootstrap values >50 are shown at the nodes.

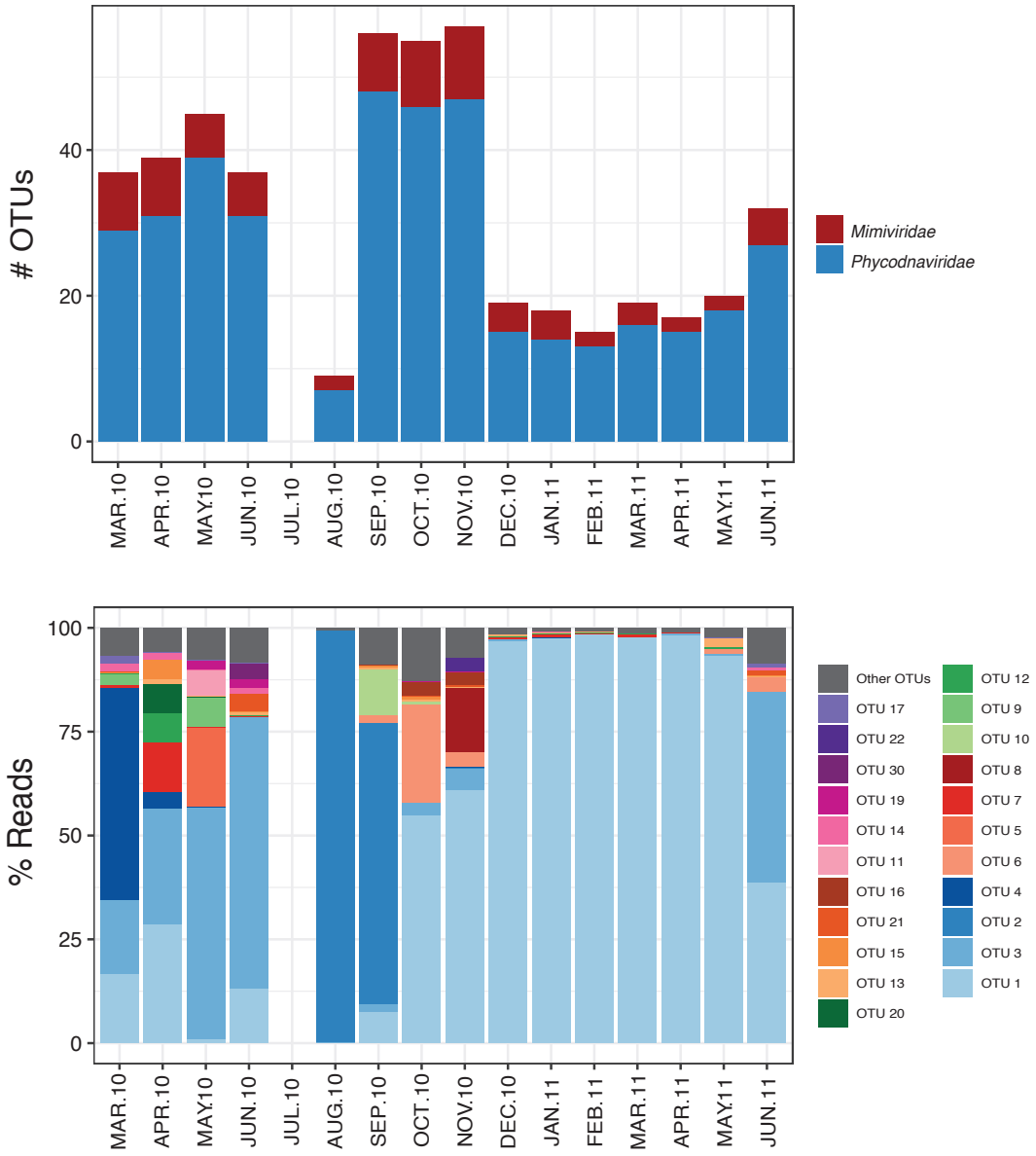


Figure 2. Barplots representing a) number of OTUs per family and b) Proportional abundance (percentage reads) of the most abundant OTUs (>50 reads per OTU) over time at Outer Oslofjorden station.

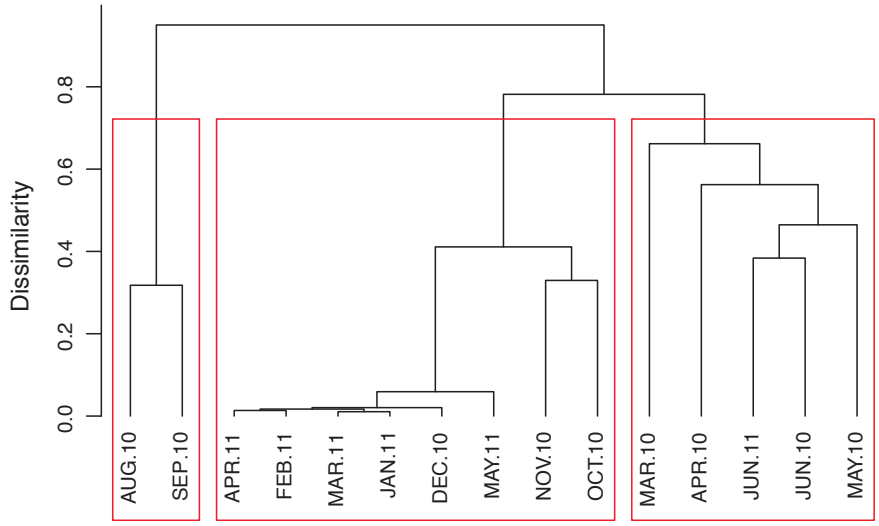


Figure 3. Bray Clustering analyses representing Bray-Curtis similarities in virus community composition based on OTU abundances.

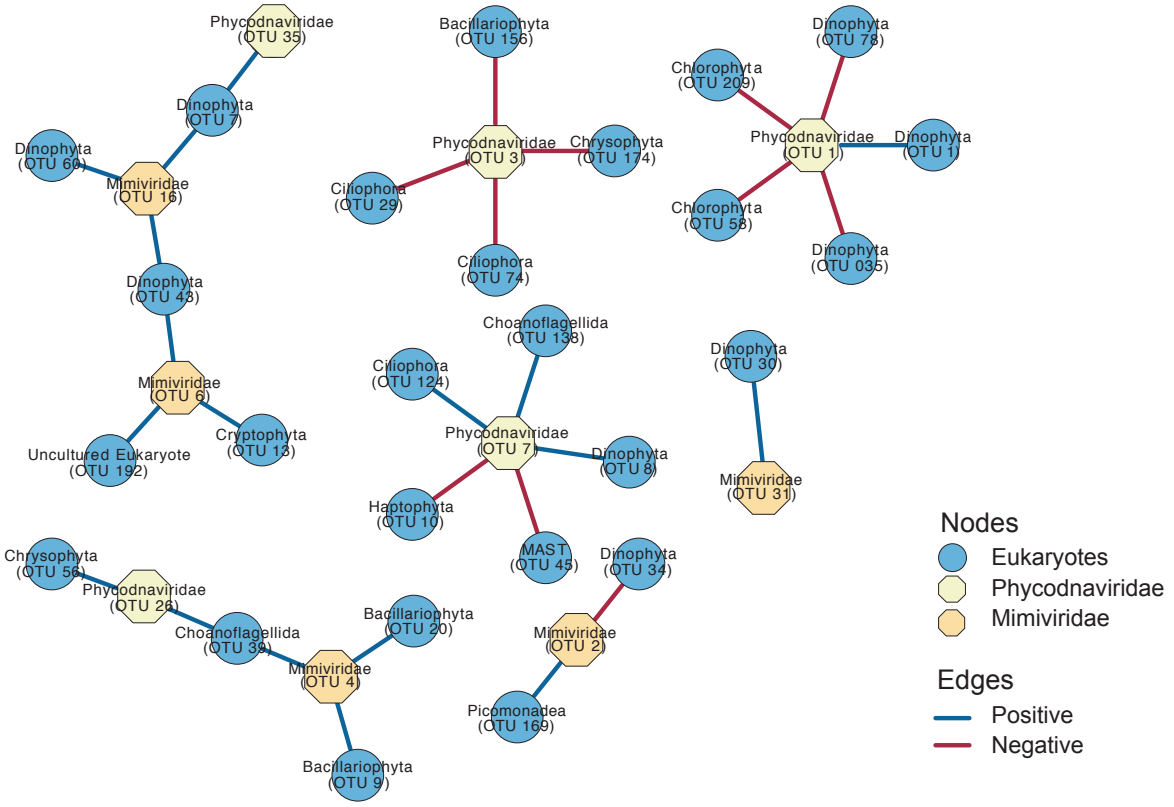


Figure 4. Network analysis showing the co-occurrence between virus and protist taxa (represented as blue and beige nodes respectively). Lines between nodes indicate positive (green) and negative (red) SparCC correlation > |0.5| (two-sided pseudo p-value < 0.05) between the abundances of linked taxa. Network was visualized by Cystoscape V3.3.0.

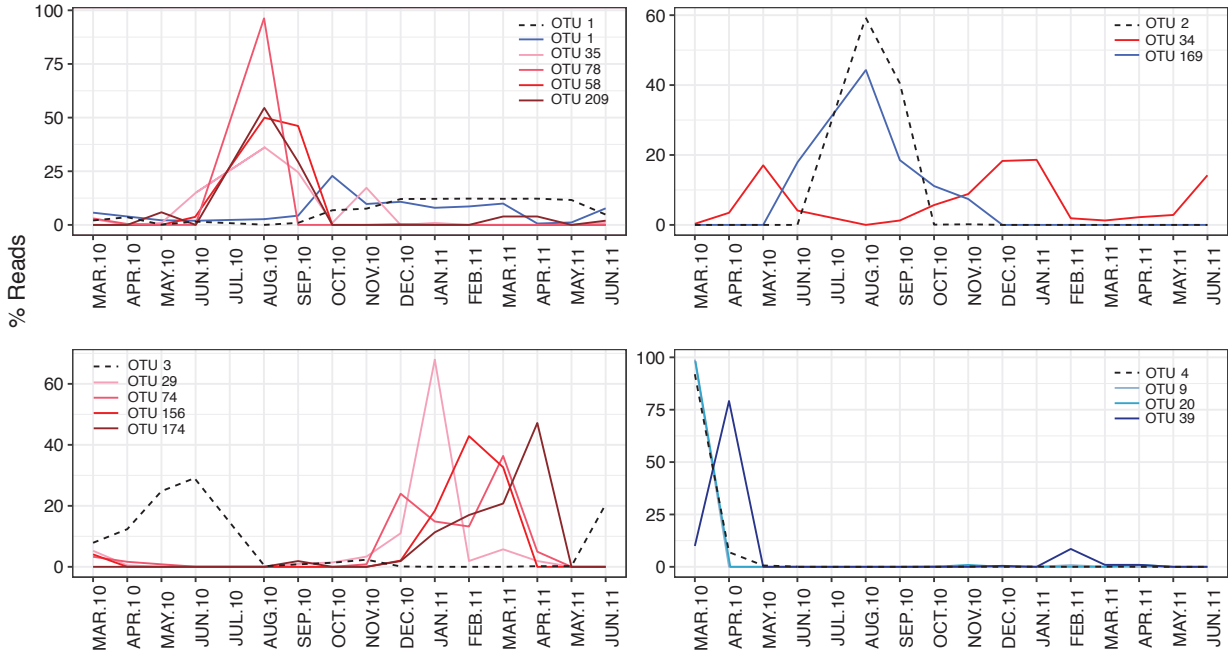


Figure 5. Temporal distribution of the four most abundant viruses and their co-occurring protists. Virus OTU relative abundances are represented with dotted lines, those of protists OTUs with continuous lines. Red lines indicate negative correlations between virus and protists, blue lines indicate positive correlations.

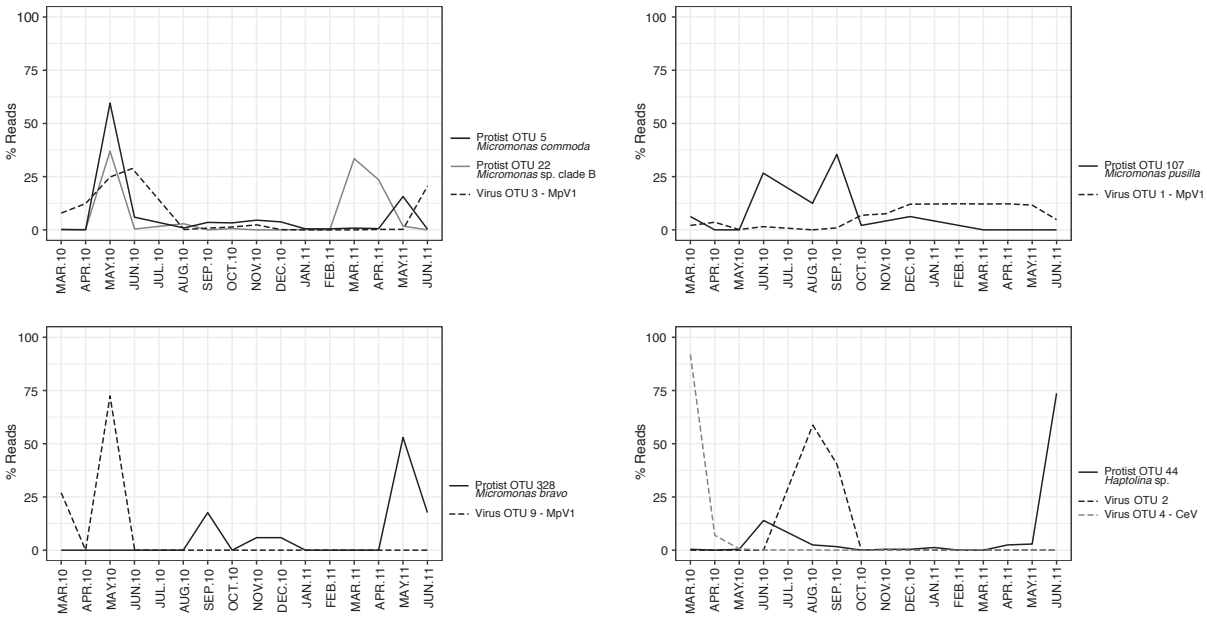


Figure 6. Relative abundances over time of OTUs matching known viruses (MpV1 and CeV) and their possible host *Micromonas commoda* and *M.* sp. clade-B-subarctic, *M. pusilla*, *M. bravo* and *Haptolina* sp. obtained by Gran-Stadniczeñko et al. (2018).

Table 1. Details on positive and negative co-occurrences between virus MCP gene OTUs obtained in this study and protist 18S rRNA gene OTUs obtained by Gran-Stadniczeñko et al. (2018).

VIRUS				PROTIST			
OTU ID	**Family**	**Lowest taxonomic level**	**Co-occurrence***	**OTU ID**	**Supergroup**	**Group**	**Lowest taxonomic level**
OTU 1	Phycodnaviridae	*Micromonas pusilla* virus (MpV1)	+	**OTU 1**	Alveolata	Dinophyta	*Karenia papillonaceae*
			-	**OTU 35**	Alveolata	Dinophyta	MALV-III
			-	**OTU 78**	Alveolata	Dinophyta	Uncultured dinoflagellate
			-	**OTU 58**	Archaeplastida	Chlorophyta	*Pycnococcus provasolii*
			-	**OTU 209**	Archaeplastida	Chlorophyta	Pyramimonadales
OTU 2	Mimiviridae	Uncultured	-	**OTU 34**	Alveolata	Dinophyta	Uncultured dinoflagellate
			+	**OTU 169**	Picozoa	Picomonadea	Picobiliphyta
OTU 3	Phycodnaviridae	*Micromonas pusilla* virus (MpV1)	-	**OTU 74**	Alveolata	Ciliophora	Cyclotrichia
			-	**OTU 29**	Alveolata	Ciliophora	Strombidiidae
			-	**OTU 156**	Stramenopila	Bacillariophyta	Coscinodiscophyceae
			-	**OTU 174**	Stramenopila	Chrysophyta	Chrysophyceae-Clade-C
OTU 4	Mimiviridae	*Chrysochromulina ericina* virus (CeV)	+	**OTU 39**	Opistokonta	Choanoflagellida	Stephanoecidae-Group-D
			+	**OTU 20**	Stramenopila	Bacillariophyta	*Pseudo-nitzschia multiseries*
			+	**OTU 9**	Stramenopila	Bacillariophyta	*Thalassiosira nordenskioel-dii*
OTU 6	Mimiviridae	Uncultured	+	**OTU 13**	Hacrobia	Cryptophyta	*Teleaulax gracilis*
			+	**OTU 192**	Uncultured eukaryote	Uncultured Eukaryote	Uncultured eukaryote
OTU 7	Phycodnaviridae	*Micromonas pusilla* virus (MpV1)	+	**OTU 124**	Alveolata	Ciliophora	Haptoria
			+	**OTU 8**	Alveolata	Dinophyta	*Gyrodinium helveticum*
			-	**OTU 10**	Hacrobia	Haptophyta	*Emiliania huxleyi*
			+	**OTU 138**	Opistokonta	Choanoflagellida	Stephanoecidae-Group-D
			-	**OTU 45**	Stramenopila	MAST	MAST-1C
OTU 16	Mimiviridae	Uncultured	+	**OTU 7**	Alveolata	Dinophyta	*Lepidodinium chlorophorum/L. viride*

			+	**OTU 60**	Alveolata	Dinophyta	MALV-III
			+	**OTU 43**	Alveolata	Dinophyta	Uncultured dinoflagellate
OTU 26	Phycodnaviridae	Uncultured	+	**OTU 39**	Opistokonta	Choanoflagellida	Stephanoecidae-Group-D
			+	**OTU 56**	Stramenopila	Chrysophyta	Chrysophyceae-Clade-C
OTU 31	Mimiviridae	*Chrysochromulina ericina* virus (CeV*)*	+	**OTU 30**	Alveolata	Dinophyta	Uncultured dinoflagellate
OTU 35	Phycodnaviridae	*Micromonas pusilla virus* (MpV1)	+	**OTU 7**	Alveolata	Dinophyta	*Lepidodinium chlorophorum/L. viride*

* "+" positive correlations, "-" negative correlations.

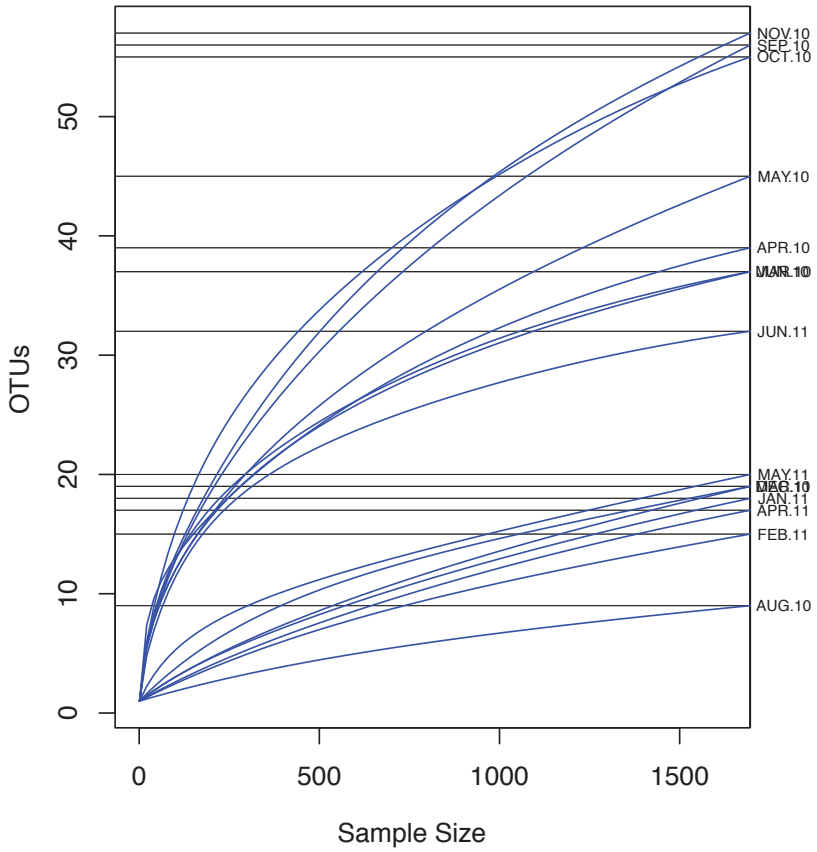


Figure S1. Rarefaction curves for the MCP reads at the different samples obtained at Outer Oslofjorden.

Figure S2. Phylogenetic tree based on available Phycodnaviridae and Mimiviridae MCP reference sequences (black), virus OTUs found in a previous study in Raunefjorden (blue, (Larsen et al. 2008)) and all algal viruses OTUs obtained from Outer Oslofjorden (red). Available at: https://figshare.com/s/7ff5a7e14308dea9394f

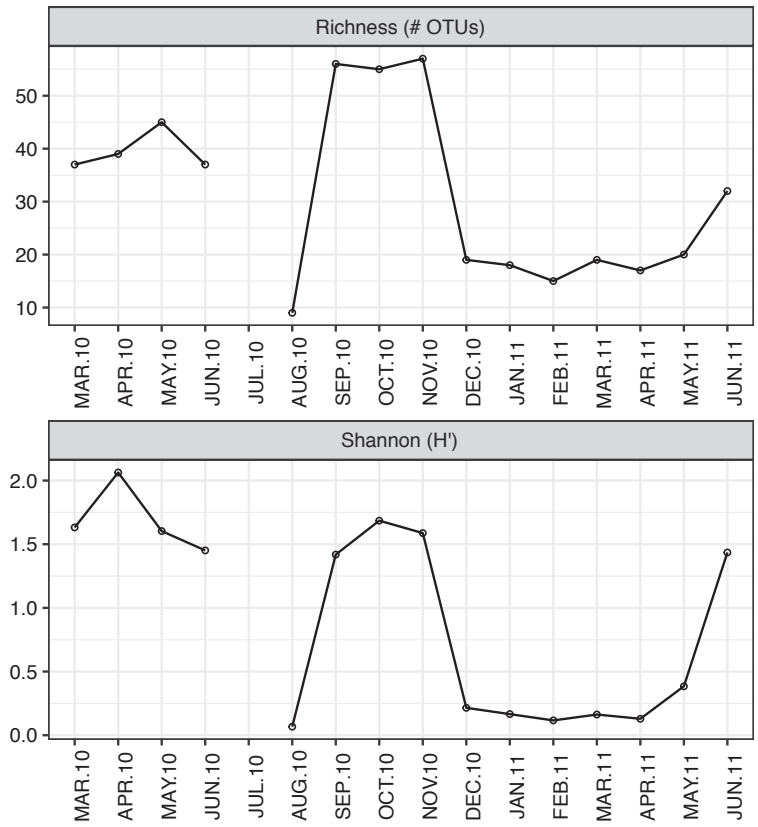


Figure S3. Virus OTU Richness and Diversity over time at the OF2 station.

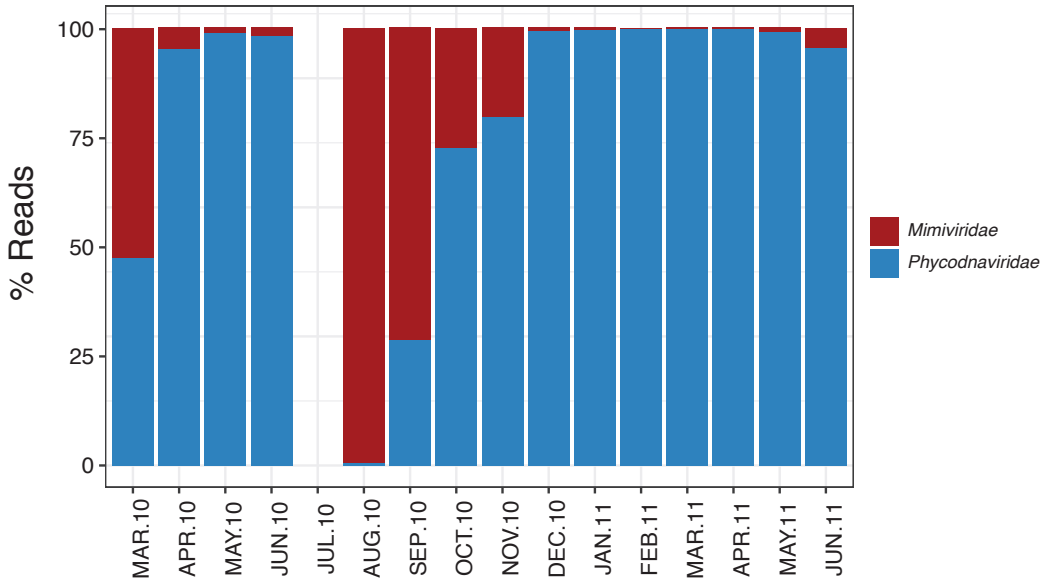


Figure S4. Proportional abundances of *Phycodnaviridae* and *Mimiviridae* overtime at Outer Oslofjorden.

Table S1. Virus MCP OTUs recorded at station OF2 during the sampling period March 2010 to June 2011 with number of reads in each sample and taxonomic placement. OTUs marked in red were removed after subsampling.
Available at: https://figshare.com/s/7ff5a7e14308dea9394f

Table S2. Isolated members of the *Phycodnaviridae* and *Mimiviridae families* used in Figure 1 study with GenBank accession numbers for the MCP.

Virus isolate	GenBank accession no.	Host species
BpV1_157	YP_004061587	*Bathycoccus sp.*
BpV2_163	ADQ91330	*Bathycoccus sp.*
CeV-M0501	EU006612	*Haptolina ericina*
CeV-M0601	EU006613	*Haptolina ericina*
CeV01	EU006628	*Haptolina ericina*
HeV_RF02	AHZ86982	*Haptolina ericina*
HhV-Of01	AHZ86983	*Haptolina hirta*
MpV1_169	YP_004062052	*Micromonas pusilla*
OlV3_00095	AFK66095	*Ostreococcus lucimarinus*
OlVG_00085	AFK65839	*Ostreococcus lucimarinus*
OtV1_165	YP_003212988	*Ostreococcus tauri*
OtV2_154	YP_004063587	*Ostreococcus tauri*
PBCV	AB018579	*Chlorella* sp.
PBCV_CVK2	BAA35143	*Chlorella* sp.
PBCV_VP54	AB006978	*Chlorella* sp.
PBCV-1	AAA88828	*Chlorella* sp.
PgV16T	EU006624	*Phaeocystis globosa*
PkV_RF01	KJ558372	*Prymnesium kappa*
PkV_RF02	AHZ86984	*Prymnesium kappa*
PpV-01	EU006631	*Phaeocystis pouchetii*
PpV-02	EU006623	*Phaeocystis pouchetii*
CrOV-PW1	YP_003969975BV	*Cafeteria roenbergensis*
Aureococcus anophagefferens virus	YP_009052173	*Aureococcus anophagefferens*
Megavirus	AGD92382	Unknown
Mimivirus	AGW18172	Unknown
Moumouvirus goulette	AGF85360	Unknown
OTU/M0501	EU086758	Unknown
OTU/P0601	EU006616	Unknown
OTU/P0604	EU006619	Unknown
OTU/P0605	EU006620	Unknown
OTU/P0607	EU006622	Unknown

File S1. Fasta file with MCP AA alignment used for the complete phylogenetic tree presented in Figure S2.
Available at: https://figshare.com/s/7ff5a7e14308dea9394f

File S2. Fasta file with MCP AA alignment used for the phylogenetic tree with the most abundant OTUs presented in Figure 1.
Available at: https://figshare.com/s/7ff5a7e14308dea9394f