Marine protist diversity and harmful algae

revealed by metabarcoding

Hulda Kjellevold Bjørneklett

Master of Science thesis

Department of Biosciences

Section for Aquatic Biology and Toxicology

UNIVERSITY OF OSLO

2018

Marine protist diversity and harmful algae

revealed by metabarcoding

Hulda Kjellevold Bjørneklett
hbjorneklett@gmail.com
+47 45 66 23 54

Supervisors
Bente Edvardsen
bente.edvardsen@ibv.uio.no

Anders Kristian Krabberød
a.k.krabberod@ibv.uio.no

Section for Aquatic Biology and Toxicology
Department of Biosciences
University of Oslo
2018

Year: 2018

Title: Marine protist diversity and harmful algae revealed by metabarcoding

Author: Hulda Kjellevold Bjørneklett

http://www.duo.uio.no

Printed: Reprosentralen, University of Oslo

Abstract

Community structure and temporal variation of marine protists may affect higher trophic levels and thereby harvestable marine resources. Some microalgae can be harmful to animal-, human- and environmental- health, with potential economic disruptions. Monitoring of protist communities and harmful algae is therefore crucial. The overall aim of this study was to investigate what groups of pelagic protist could be detected through metabarcoding, focusing especially on harmful and toxic algae. Using environmental seawater samples from two depths, the main objectives were to reveal the seasonal variation of the detected protists and at what time of the year species of harmful algae appeared in the samples. Further, this thesis questions whether it is necessary to sample at two selected depths in monitoring of harmful algae and if the same species can be detected through cell counts by microscopy. The highly variable and conserved V4 region of the 18S of the small eukaryotic ribosomal RNA gene was amplified from DNA that was extracted from environmental seawater samples from inner Oslofjorden. The material was sequenced using the Illumina MiSeq paired-end technology. The present thesis demonstrates a large diversity of protists. Dinoflagellates were found to be the most abundant and diverse group, followed by the diatoms. Strong seasonal variation characterized the protist community and toxic species of *Alexandrium* and *Dinophysis* periodically appeared in high abundances. Moreover, representatives of 11 genera of harmful algae were discovered. The results indicate no significant difference between 0-2 m- and 5 m depth in the detection of harmful algae in Oslofjorden. Finally, comparison of microscopy cell counts and the present approach revealed major differences in detection of harmful species and only eight species were detected by both methods combined. This is the first long-time study to investigate the protist plankton community in inner Oslofjorden with the use of metabarcoding revealing high protist diversity and harmful algae.

Acknowledgement

The work presented in this thesis was accomplished at the section for Aquatic Biology and Toxicology at the Department of Biosciences at the University of Oslo. Main supervisor was professor Bente Edvardsen (UiO) and co-supervisor was head engineer Anders Kristian Krabberød (UiO). The project has been linked to the NFR (the research council of Norway) - supported Ph.D. research project "Plankton - Change in phytoplankton community structure in the inner Oslofjord during a century of sampling", project number 262960. On that regard, I would like to thank NFR for accessibility of CTD-data and the time spent at cruises. Elisabeth Lundsør has been employed at Norconsult as industrial Ph.D.-student in the project and she has collected some of the samples.

This thesis would not have been accomplished without my supervisor Bente, whom I would like to thank for her dedication to this project and her engagement for microalgae. You are a great inspiration. Further, I owe gratitude to my co-supervisor Anders for guiding me through the world of coding and statistics. Your willingness to always help has been indispensable.

Many thanks go to Bushra Abbas and Luka Supraha for assistance in lab-work. Your patience and encouragement have made the days in the lab an educational and fun experience. I would like to express an extra thanks to Luka for providing help in later analysis and for eagerly trying to solve whatever issue I encountered. Many thanks to the crew at R/V Trygve Braarud for great days at sea, to professor Josefin Titelman for guidance in writing and Silje Marie Kristiansen for sharing her artistic skills in helping me in illustrator.

Thank you, Karoline Saubrekka, for making my last year at UiO a great year.

Finally, I wish to thank my mother and father for their support and encouragement during this process.

Table of contents

1 Introduction

Protists are unicellular eukaryotes that represent a large variety of organisms exhibiting many different ecological roles. Apart from being unicellular and eukaryotic, protists do not necessarily share any genetic -or morphological traits and include all eukaryotes other than animals, fungi, and plants (Pawlowski, 2014). In other words, protists cannot be unified as *one* phylogenetic clade but include organisms from the five major eukaryotic supergroups: Archeplastida, Opisthokonta, Amoebozoa, Excavata, Rhizaria, Stramenopiles and Alveolata (Caron et al., 2012; Pawlowski, 2014). Archeplastida cover the three photosynthetic linages Rhodophyta (red algae), Chlorophyta (green algae) and glaucophytes. Opisthokonta contains the large groups Metazoa (multicellular animals) and fungi. Few protists are found within Opisthokonta however, the choanoflagellates are placed as sister group to Metazoa with strong genomic evidence (Suga et al., 2013). Moreover, Metazoa includes the fish parasites called ichthyosporea (within Mesomycetozoea) that are protists. Within Amoebozoa we find mainly amoebae, whereas Excavata includes many parasitic forms along with other free-living taxa (Pawlowski, 2014). The supergroup Rhizaria cover the diverse group Cercozoa and Radiolaria, including the important calcium carbonate covered Foraminifera. Finally, the Stramenopiles and Alveolata are both large supergroups with many protistan phyla. The latter contains the Dinoflagellates, Ciliates, and Apicomplexa, while the Stramenopiles, also called Heterokonta, is composed of several algal groups. Common for the algae within Stramenopiles is that they possess heterokont flagella, i.e. two differed flagella –one smooth and one with hairs.

As indicated above, the world of protists is complex and involves many different taxa and ecologic strategies. In temperate oceans, protist composition, amount and production go through constant changes throughout the year. Because of their short life cycles, many algal species can increase in density rapidly. When encountering favorable growth conditions, some species can even form massive blooms. Such changes in the community are essentially driven by abiotic factors like sunlight, temperature, salinity, currents and nutrient supply, in combination with biotic factors like grazing by zooplankton. Variation in composition and production has effects on higher trophic levels and thereby harvestable marine resources. Research on protist community composition and which groups are dominating in the different seasons is therefore important.

Most algae are photosynthetic organisms living in the aquatic environment (Graham, L. E. et al., 2016). The microalgae that perform photosynthesis are defined as phytoplankton. They are microscopic and spend their life floating in the free water masses as primary producers and fundamental members of the food web. Like plants, phytoplankton depends on sunlight and carbon dioxide to live and reproduce, while they generate oxygen and sugars for herbivores to exploit. In this way, phytoplankton (including planctonic cyanobacteria and phototrophic microalgae with chlorophyla a) makes up the foundation of the aquatic food web and acts as a major carbon dioxide sink (Ebenezer et al., 2012). Moreover, phytoplankton accounts for about half of all primary production and oxygen production in the world (Duarte & Cebrián, 1996; Field et al., 1998) and forms the basis of nearly all marine food webs (Not et al., 2012). Although most algae are photosynthetic, many are also heterotrophic and rely on eating other organisms to obtain energy. Further, some algae are mixotrophic, having the ability to obtain energy from both photosynthesis and by consuming other organisms.

1.1 The seasonal cycle of phytoplankton

Depending on the time of the year, temperate oceans are usually separated in two strata, caused by differences in temperature and salinity (Denny, 2008). However, during the winter the water masses hold approximately similar temperatures throughout the water column. This makes the water masses unstable, creating vertical mixing. In addition, strong winds in the winter months contribute to the mixing by pushing water out from land and consequently bringing bottom water up to the surface. As a result, nutrients are being brought up to the surface layer while the phytoplankton concentration is typically low because of constant transportation of cells down below the euphotic zone. Low microalgae biomass is therefore often observed in December to February in temperate seas (Figure 1).

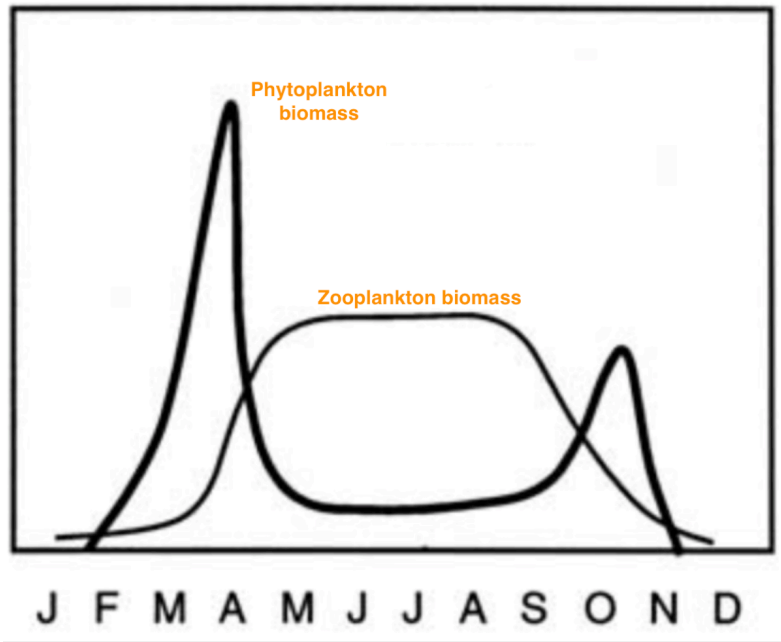


Figure 1: *Seasonal cycle of phytoplankton and herbivorous zooplankton in south Norwegian cosal waters. Figure modified from compendium by (Paasche, 2005).*

A salinity gradient starts forming in the early spring (February-March) when freshwater inputs from rivers establish on top of the more saline and heavy water. In combination with heating of the surface layer from the sun, the water column is stabilized. With the upper layer saturated with nutrients, it becomes a favourable place for algal growth (Paasche, 2005). As spring brings with it more sunlight and longer days an algal bloom usually occurs. In the course of spring, phytoplankton assimilate ammonia, nitrate, phosphate, and silicate, resulting in an upper layer with low levels of inorganic nutrients (Naustvoll & Dahl, 2002). A chlorophyll minimum is therefore often the situation in April. However, nutrient-rich floods and water from snow melting may give rise to a second small spring bloom in May. As the months of fall arrive, it can bring with it heavy rainfall and strong winds. The wind generates mixing of the water and freshwater input adds more nutrients to the top layer. The result from this can be a third bloom, termed fall bloom. In the winter, the two layers retain the same temperature and density, resulting in more or less complete mixing of the water column (Denny, 2008). Because of these distinct seasonal patterns, one can argue that the phytoplankton concentration in temperate oceans is to some extent predictable. However, what species that are present and dominating throughout the year can be highly irregular (Naustvoll & Dahl, 2002).

1.2 Harmful algal blooms (HABs)

Blooms of microalgae are a natural part of the seasonal cycle in marine food webs (Berdalet et al., 2016) and blooms are sometimes linked to human health problems and economic disruptions. Such issues are caused by HABs (Harmful Algal Blooms) (Berdalet et al., 2016; Zingone & Wyatt, 2004). Algal species causing HABs are either toxin-producing species or non-toxic species that become harmful at high densities. Some blooms of toxin-producing algae can accumulate in shellfish and fish making them toxic for humans to ingest. Others produce aerosolized toxins that impact human health in other ways. Further, some algae can be harmful only to fish and invertebrates causing death of these animals (Not et al., 2012; Zingone & Enevoldsen, 2000; Zingone & Wyatt, 2004). The variety of harmful effects is depicted in Table 1.

Table 1: *Harmful effects of microalgae (table modified from Zingone & Enevoldsen, 2000; Zingone & Wyatt, 2004).*

Effects on human health	**Causative algal phyla**
Paralytic shellfish poisoning (PSP) (Saxitoxin group)	*Dinoflagellates, Cyanobacteria*
Diarrheic shellfish poisoning (DSP) (Okadaic acid group and pecteno toxin group)	*Dinoflagellates*
Amnesic shellfish poisoning (ASP) (Domoic acid group)	*Diatoms*
Neurotoxic shellfish poisoning (NSP) (Breve toxin group)	*Dinoflagellates*
Ciguatera fish poisoning (CFP) (Neurotoxin)	*Dinoflagellates*
Azaspiracid poisoning (AZP) (Azaspir acid group)	*Dinoflagellates*
Effects on natural and cultured marine resources	**Causative algal phyla**
Haemolytic, hepatotoxic, osmoregu- latory effects and other unspecified toxicity	*Dinoflagellates, Raphidophytes, Prymnesiophytes, Pelagophytes, Cyanobacteria*
Mechanical damage	*Diatoms, Prymnesiophytes*
Gill clogging and necrosis	*Diatoms*
Effects on tourism and recreational activities	**Causative algal phyla**
Production of foams, mucilages, discolouration, repellent odours	*Dinoflagellates, Prymnesiophytes, Diatoms, Pelagophytes, Cyanobacteria*
Effects on the marine ecosystem	**Causative algal phyla**
Hypoxia, anoxia	*Dinoflagellates, Diatoms, Prymnesiophytes*
Negative effects on feeding behav- iour, reduction of water clarity	*Pelagophytes*
Toxicity to wild marine fauna	*Dinoflagellates, Diatoms*

As can be seen in the table above, harmful effects on humans are most often caused by toxin producing dinoflagellates and diatoms (Zingone & Enevoldsen, 2000), and mostly from species of the genera *Alexandrium, Dinophysis, Gymnodinium* and *Pseudo-nitzschia* (Farabegoli et al., 2018). The harmful affects on humans from dinoflagellates and diatoms include paralytic shellfish poisoning (PSP), diarrhoeic shellfish poisoning (DSP), neurotoxic shellfish poisoning (NSP) ciguatera fish poisoning (CFP), Amnesic shellfish poisoning (ASP) and Azaspiracid poisoning (AZP).

PSP originate from eating shellfish that have accumulated cells of toxin producing dinoflagellates, often species of *Alexandrium*. The symptoms are nausea, vomiting, tingling sensation in the mouth and paralysis that can, in some cases, be life threatening (Farabegoli et al., 2018). Symptoms like abdominal pain, nausea, and diarea are caused by DSP, while NSP and CFP can give gastrointestinal and neurological symptoms. Outbreaks of DSP have occurred in Norway more than once, with *Dinophysis acuminata*, *D. acuta*, and *D. Norvegica* as the causative species (Dittami, S. et al., 2013). Sympthoms of ASP is Nausea and diarea, in addition to short-therm memory loss and sometimes coma and brain damage. AZP can arise from ingestion of mussels and cause gastrointestinal symptoms (Farabegoli et al., 2018). In other words, the effects of HABs appear in large variety and can have considerable consequences in all areas of the world including Norway (Table 2). Some algae are ichtyotoxic meaning that they are toxic to fish. One example is the Haptophyte species *Prymnesium parvum* that have caused many problems like economic disruptions connected to fish farms in Norway and other places (Granéli et al., 2012; Roelke et al., 2016). This is one of the reasons why fich and mussel farmers are most often the ones that experience issues with HABs in Noway (Dittami, S. et al., 2013).

Table 2: *Table of algae that are known to cause harm in Norwegian waters and species that are found in Norway and found to be toxic other places, and therefore considered potentially toxic in Norway. Modified from (Edvardsen, 2017)*

Algal classes	Harmful in Norway	Potentially harmful in Norway
BLUE-GREEN ALGAE (Cyanophyceae)		*Nodularia spumigena*
DINOFLAGELLATES (Dinophyceae)	*Alexandrium tamarense*	*Prorocentrum minimum*
	Alexandrium ostenfeldii	*Pseudopfiesteria shumwayae*
	Alexandrium minutum	*Alexandrium pseudogonyaulax*
	Azadinium spinosum *Azadinium* spp.	*Tripos* (*Ceratium*) sp.
	Dinophysis acuminata	
	Dinophysis acuta	
	Dinophysis norvegica	
	Karenia mikimotoi	
	Karlodinium veneficum	
	Protoceratium reticulatum	
HAPTOPHYTES (Prymnesiophyceae)	*Chrysochromulina leadbeateri*	Prymnesiales spp. (=*Chrysochromulina* spp.)
	Prymnesium parvum	
	Prymnesium polylepis (=Chrysochromulina polylepis)	
DIATOMS (Bacillariophyceae)	*Pseudo-nitzschia* spp.	*Pseudo-nitzschia* spp.
RAPHIDOPHYTES (Raphidophyceae)	*Heterosigma akashiwo*	
DICTYOCHALES/ SILICOFLAGELLATES (Dictyochophyceae)	*Pseudochattonella farcimen*	*Dictyocha speculum*
	Pseudochattonella verruculosa	
	Vicicitus globosus	
	Fibrocapsa japonica	

1.3 Oslofjorden – topography and physical conditions

Seasonal variations in the protist community can hardly be explained without some knowledge about the topography and physical state of the environment. Oslofjorden is characteristic in the way that it possesses a shallow threshold marking the separation of the inner and outer fjord (Figure 2 and 3). This 20 m deep- and 1 km wide sill is the only passage for the exchange of water from the outer to the inner fjord (Thaulow & Faafeng, 2013). Each side of the sill is deep: about 200 m on the south side and 150 m on the north side (Figure 3). Further, the inner Oslofjorden has multiple sills and the depth varies between 160-50 m depth. These sills separate different basins, limiting water renewal in the inner deep parts (Thaulow & Faafeng, 2013).

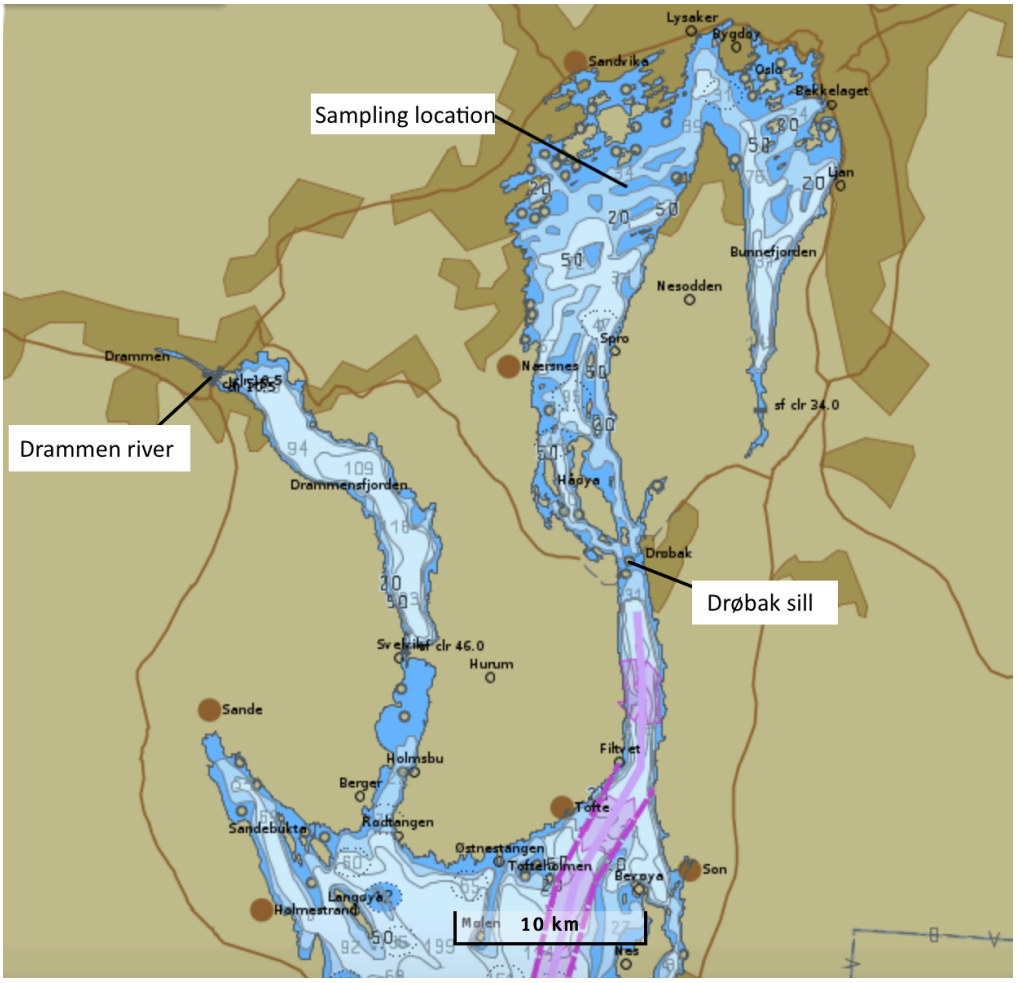


Figure 2: *Map of Oslofjorden pointing at the Drøbak sill, the river mouth of Drammenselva and the sampling location. Map created at kartverket.no*

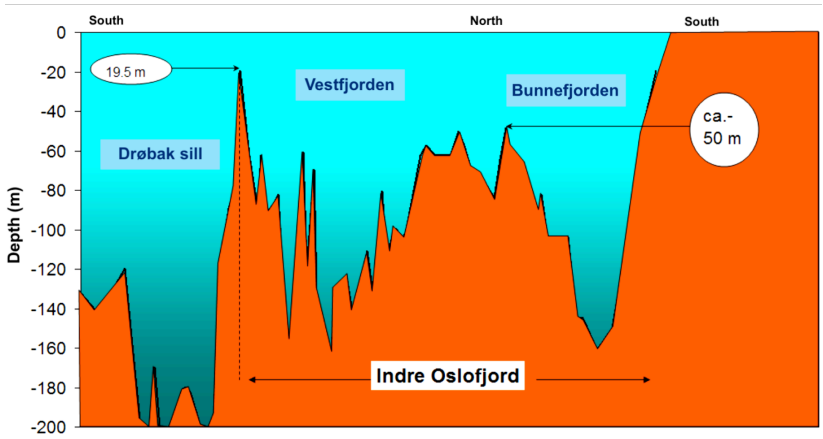


Figure 3: *Topography of inner Oslofjorden. Figure modified from (Thaulow & Faafeng, 2013).*

The water masses in inner Oslofjorden is influenced by water that enters from the outer Oslofjorden and Skagerak. The water masses in Skagerak interfere with water from the Atlantic Ocean brought by the Gulf Stream. On its way to Skagerak, the Gulf Stream passes the British islands and enters between Iceland and Norway. Some of these water masses breaks off and ends up in the North Sea, where it encounters the Skagerak. In addition, the Skagerak area is influenced by water from the Jylland Stream, which passes the English Canal and Germany. Last but not least, much of the water in Skagerak originate from the more or less brackish water from the Baltic sea (Baalsrud & Magnusson, 2002). This means that inner Oslofjorden is dynamic and interfere with many other bodies of water.

In inner Oslofjorden temperature and salinity varies throughout the year with normally low temperatures and high salinity in the winter and the opposite in the summer (Baalsrud & Magnusson, 2002). The heating of the sun primarily drives these seasonal changes. The main contributors to fresh water are the large rivers of Glomma and Drammenselva. Drammenselva runs out approximately in the middle of the fjord while Glomma flows out in the outermost part of the fjord. These inputs create a brackish water layer in the outer fjord, especially during times of snow melting in the spring. The brackish surface layer occasionally flows into the inner fjord by coastal winds. The input of freshwater affects the salinity and thereby the stratification that in turn has an impact on the physical conditions in the fjord (Baalsrud & Magnusson, 2002). In the spring and summer, the stratification is normally strong due to the freshwater inputs and warming of the surface water, whereas uniform temperature in the water column resulting from water mixing normally is the situation in the winter.

As a result of industrial growth and the introduction of water closets in the early 1900s, Oslofjorden experienced large inputs of pollutants, leading the fjord into a poor physical state in terms of too high levels of nutrients leading to high chlorophyll levels and low oxygen in the bottom waters (Baalsrud & Magnusson, 2002). On this basis a program for monitoring of the environmental conditions of Oslofjorden was started in the 1970s (Fagrådet for vann- og aløpsteknisk sammarbeid i indre Oslofjord, 2017). NIVA (Norwegian Institute for Water Research) has been performing the monitoring up until 2015 when Norconsult took over. Expansion of the sewage canals and construction of water treatment plants has led to an improvement of the water quality and the recovery has been going on for the last four decades. However, later years the improvement has been leveling off and a slight increase in nutrient inputs have been registered (Thaulow & Faafeng, 2013). Reasons for this is that the capacity of the sewage networks is starting to reach its bearing capacity due to population growth. In addition, heavier rainfall, caused by climate change, is predicted to challenge the capacity of the sewage canals and rivers further (Baalsrud & Magnusson, 2002). Today the condition of Oslofjorden is fairly good, but problems regarding oxygen limitations in the deep basins and occasional algal blooms are both problems that are predicted to increase (Thaulow & Faafeng, 2013).

1.4 Monitoring of microalgae

Because of the harmful effects of some microalgae and the effect of general protist composition on higher trophic levels, monitoring of the protist -and algae community is important. Monitoring of micoalgae has traditionally been performed by microscopy (Ebenezer et al., 2012). The most common method for microscopic cell counts is the Utermøhl-method (Utermöhl, 1931). This approach involves fixation of water samples (5-50 ml) and allowing cells to sink to the bottom of a counting chamber (ca 1 mL). Then the cells are identified under an inverted microscope and counted. This method is time-consuming and requires high taxonomic expertise because some species can be very hard to distinguish in a light microscope and some even impossible (cryptic species). The practice of counting cells in a microscope is limited to species larger than 15-20 μm because smaller species can be hard to see. In addition, some cells are easier to preserve (hard structures) than others, presenting a potential bias. Furthermore, species of low abundance can be hard to detect. On the other hand microscopy cell counts is a low-cost metod compared to many other teqniqes for detection of biodiversity. Another positive aspect of the microscopy approasch is that this

method has long traditions and generated years of data, which allows for comparisson to earlier findings. Today the monitoring of inner oslofjorden of microalgae includes cell counts in microscopy on samples taken from 0-2 m depth and 5 m depth.

Other methods, often used for quantification of toxic algae, are microarray and real-time PCR. Microarray are estimations that aim to discriminate genetic differences of different species with the use of species-specific DNA fragments (probes) (Dittami, S. M. et al., 2013). Quantitative polymerase chain reaction (qPCR) is another technique that can be used to detect specific harmful algae by amplification of target DNA molecules during PCR (real-time), meaning that the number of amplified DNA in the PCR reaction can be estimated in the samples (Dittami, S. M. et al., 2013) However, during the last decade, molecular sequencing methods for research on microbial diversity have had a rapid evolution (Kozich et al., 2013). Sanger sequencing has, for many years, been used to sequence genetic libraries and has been a common method used for detection of biodiversity of microscopic organisms. Later years new methods for high-throughput sequencing (HTS) has become a more efficient approach in that it can sequence millions of fragments at a time compared to hundreds using sanger sequencing (Deiner et al., 2017; Zinger et al., 2012).

In this study Metabarcoding was used to reveal the protist community in environmental samples and potentially harmful algal species. Metabarcoding is a teqnique to detect biodiversity in a pooled DNA sample and ere we used DNA from samples from inner Oslofjorden. The pooled DNA is used to aplify a specific region and the regions are sequenced using sequencing technologies, in this case, Illumina sequencing.

1.5 Illumina sequencing

The metabarcoding tecnique starts with extraction of DNA from environmental samples. Extracted DNA sequences are then amplified thorough polymerase chain reaction (PCR). The PCR-products, or amplicons, are rinsed and labelled with index adaptors. After this preparation of the so-called genetic library, it can be sequenced. The present study used Illumina MiSeq paired-end sequencing to amplify the clones by bridge amplification. The genetic marker used in this study was the V4 region of the small subunit (SSU) of the 18S rRNA gene. This region is on average 390-410 base pairs (bp) long, but the length varies widely. Moreover, the region is highly variable (Nickrent & Sargent, 1991; Zimmermann, J.

et al., 2011) which makes it a well-suited maker for studying the diverse community of protists.

Sequencing starts with the genetic library being loaded into a flow cell. The flow cell is a glass plate with oligonucleotides (oligos) that are complementary to index adaptors attached to the PCR products. The process starts with the DNA fragments attaching to the complementary primers in the flow-cell by hybridisation, forming a bridge (Figure 4) A polymerase enzyme makes a complemetary DNA- strand and the double stranded DNA, is denaturazed. The original template is washed away and the oposite end of the sequence hybridizes with anoter oligo on the flow-cell surface. Next, polymerase synthesizes a complimentary strand and the outcome is a double stranded DNA structure. This double DNA strand is denaturized and each strand detaches from the surface on the one end, making it straighten up. Both strands form a new bridge and polymerase syntetizises both strands. This process is performed simulatiusly for all the DNA fragments in the flow-cell surface, which results in clusters of identical strands. After completed amplification, all the reverse-strands are washed away and the flow-cell contains only forward-strands.

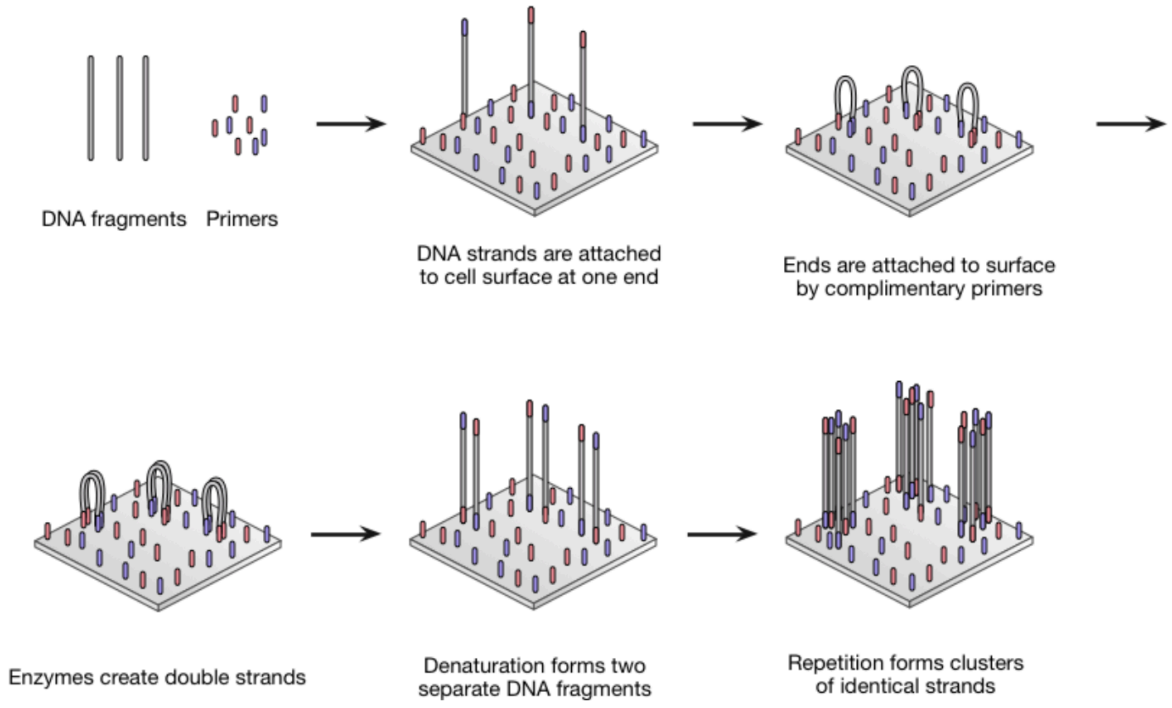


Figure 4: *The steps of bridge amplification. Copyright © ATDBio Ltd.2005-2018*

The forward strands are sequenced using deoxynucleotide triphosphates (dNTPs) that are fluorescently marked. Two lighters light upon the nucleotides; one with green light for the bases guanine (G) and thymine (T) and one red laser for the bases adenine (A) and cytosine (C). When lighted upon, the flotochrome in the fluorescently labelled dNTPs emits light. Clusters on the flow-cell that are registered as red or green are interpreted as C and T bases respectively. Clusters that reflect both green and red light are interpreted as A bases (seen in yellow). Unlabelled clusters are registered as G bases (Illumina, 2018). This is repeated for many cycles. Each nucleotide that is registered in the sequence is assigned a quality score that tells the probability that a given base has been called correctly. The sequenced strands are washed away and index one is registered. Then, the forward strands forms bridges and indiex 2 is registered. The reverse strands are synthesized, forward strands are washed away and the reverse strands are sequenced. This process is called paired-end sequencing.

1.6 Objectives

The present study aimed to determine what groups of single-celled eukaryotic organisms (protists) could be detected through metabarcoding of environmental samples, from inner Oslofjorden, over a two years period. Further, this study reveals the diversity and seasonal dynamics of protist groups. Temporal variation of potentially harmful algae was described and phylogenetically classified. Finally, it was tested whether sampling at the two depths of 0-2 m and 5 m is necessary in monitoring of harmful and toxic algae and if the same species detected by metabarcoding also were detected by cell counts by microscopy.

The principal questions addressed in this thesis were:

1. What groups of protists can be detected through metabarcoding of environmental seawater samples from inner Oslofjorden?

2. Does the composition and relative read abundance of the detected protist groups vary throughout the year?

3. What genera of potentially harmful algae can be detected in the samples and do they vary in relative read abundance throughout the year?

4. Is there a significant difference between the two depths (0-2 and 5 m) in the presence of harmful algae?

5. Can the same species of harmful algae as detected by microscopy cell counts be detected through metabarcoding?

2 Materials and methods

2.1 Sampling location

The fieldwork was conducted in the middle of Vestfjorden in inner Oslofjorden, Norway (Figure 5). The sampling station is called DK1 and the coordinates are 59° 48´54.0´´N, 10°34´09.8´´E. The station represents a deep basin in this area (100 m) and therefore considered to be representative of inner Oslofjorden.

Figure 5: *Sampling station (DK1) marked on map over Oslofjorden. Map created at kartverket.no*

2.2 Sampling procedures

Water samples and hydrographical data were collected from two depths (0-2 m and 5 m), during the period February 2016 to December 2017. Altogether 30 samples were collected (Table 3) in addition to backup-samples. Cruises were carried out on the University of Oslo research vessel Trygve Braarud.

Table 3: *The 30 samples collected at station DK1 in 2016 and 2017*

Date	0-2 m	5 m
24. February 2016	✔	✔
11. April 2016	✔	✔
22. June 2016	✔	✔
18. August 2016		✔
12. September 2016	✔	✔
25. November 2016	✔	✔
02. February 2017		✔
20. February 2017		✔
06. March 2017		✔
18. April 2017	✔	✔
15. May 2017	✔	✔
21. June 2017	✔	✔
14. August 2017	✔	✔
18. September 2017	✔	✔
18. October 2017	✔	✔
16. November 2017	✔	✔
15. December 2017	✔	✔

Water samples from 5 m depth were collected in 5 L Niskin bottles that were attached to a CTD rosette. The CTD is lowered into the water from the boat. The closing of the Niskin bottles is computer-controlled, allowing for bottles to be shut at desired depths. Surface water samples were collected using a 2 m long water hose (KC, Danmark). The water from both depths was transferred to rinsed plastic bottles.

On board the boat, the water samples were filtered through a Sterivex filter (Sterivex GV 0.22 µm, EMD Millipore, Italy) using an Ismatec peristaltic pump (Ismatec, Germany) and a silicon tube (Figure 6). The filter collects organisms larger than 0.22 µm. The pump was set at 500-700 rotations per minute (rpm), and time spent at filtering was never more than 2 hours. Two L of water from each of the two depths was filtered, in addition to one backup-filter for each sampling-depth. The filters were sealed with poster gum and kept cool in a cooling-bag with ice packs. Back at the lab, the Sterivex filters were stored in an -80°C freezer.

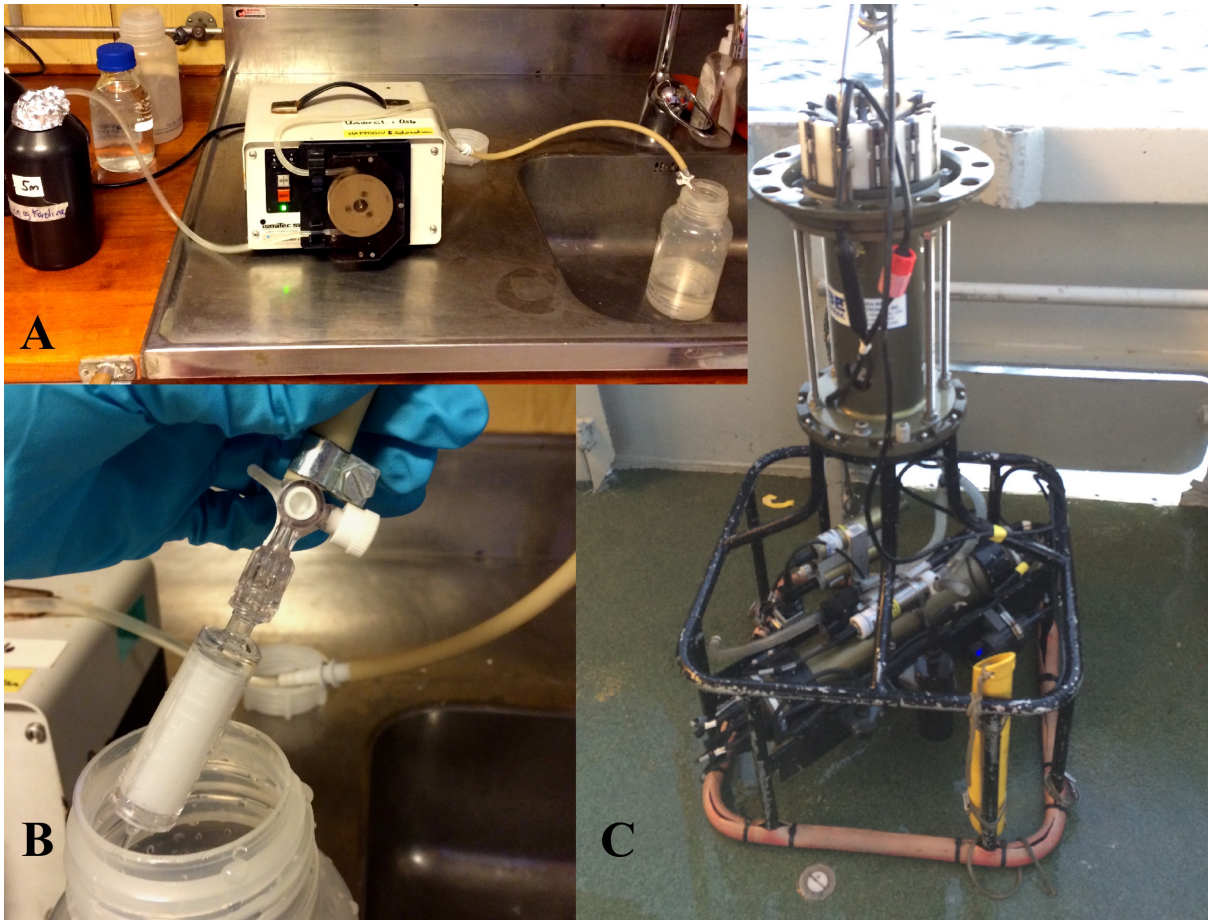


Figure 6: ***A)*** *Picture of pump set-up on board R/V Trygve Braarud.* ***B)*** *Picture of the Sterivex filter attached to silicon tube.* ***C)*** *Picture of the CTD (Niskin bottles detached).*

2.3 Hydrographical data

To be able to observe seasonal changes in the physical and chemical properties of the water, hydrographical data was collected on the same day as the samples were gathered. The data were obtained with the use of an SBE 9plus CTD (Sea-Bird Scientific, USA). The CTD is lowered into the water while it pumps seawater through a metal housing. This is equipped with sensors, which records conductivity (a measure for salinity), temperature and density for every meter as it descends. In addition, *in vivo* fluorescence, a measure for chlorophyll was recorded with another sensor on the CTD. Measurements were registered all the way from the bottom (100 m depth) to the surface and the data was simultaneously visualized on a CTD-profile.

2.4 Molecular methods

2.4.1 DNA isolation

DNA on the Sterivex filters was extracted in the lab using the Qiagen DNeasy Power Water Sterivex Kit (Qiagen, Germany). The extractions followed the protocol included in the kit. A modification of the protocol used for extractions are described in Appendix 1. Briefly: the Sterivex filters were thawed on ice and subsequently treated with cell release solution. This release solution made microbes detach from the filter membrane into solution. Then the filters were attached to a vortex-adaptor, specifically designed for attachment of sterivex filters, and then vortexed (Vortex Genie 2, Scientific industries, USA). When mixed, a pre-heated strong lysing reagent was added and the filters were incubated at 90°C and vortexed again. The lysing reagent helped to break the cell walls and remove non-DNA organic- and inorganic material. The lysate was carefully removed from the Sterivex filters with a syringe and transferred to a 5 ml bead beating tube that was vortexed and centrifuged. During mixing and centrifugation, the small beads shatter the cells in the sample, cracking them open, resulting in a release of intracellular material. The supernatant was removed from the beads, using a syringe, and transferred to a clean tube. To get as high DNA purity as possible, another reagent was added to help remove additional non-DNA organic and inorganic material, cell debris, and proteins. The tubes were vortexed, incubated and centrifuged. The sample was then transferred to a column with a silica filter membrane with a tube placed underneath (collection tube). At high salt concentrations DNA binds to silica, therefore a solution high on salt was added to the column. This made DNA bind to the silica filter membrane. Next, multiple washing steps (ethanol) combined with centrifugation ensured that

residual contaminants, salt, inhibitors, and proteins were removed. Finally, to make DNA detach from the silica filter membrane, a sterile EB buffer (Elution Buffer 250 ml, QIAGEN, Germany) was added to the membrane in the column. This released the DNA from the silica into the collection tube. The tube containing the pure DNA was put in the refrigerator, holding approximately 4°C. Samples were kept in the refrigerator for a few weeks up to approximately four months. Extractions were performed on four to six Sterivex filters per day in the lab, due to limited space on the vortex adaptor as well as the risk of DNA degradation.

2.4.2 DNA quantification

To make sure enough DNA was extracted for Illumina sequencing, the DNA was measured using a Qubit fluorometric quantification machine (Qubit 3.0 fluorometer, Thermo Fisher Scientific, USA). This method uses a Qubit reagent that is a fluorophore that attaches to DNA in a sample. This fluorophore emits light when exposed to it. As the Qubit machine lights the sample, it estimates the concentration of nucleic acid in the sample

To prepare the DNA samples for quantification, a high-sensitivity qubit kit (Qubit dsDNA HS Assay Kit, 500 assays, Invitrogen, Thermo Fisher Scientific, USA) was used, and the protocol from the manufacturer was followed. Briefly: the two standard tubes from the kit (standard DNA) were taken out of the refrigerator, allowing them to reach room temperature. Then DNA samples were thawed on ice, vortexed and spun down. Meanwhile, a mastermix for the qubit reaction was prepared according to Table 4.

Table 4: *Amount of qubit reagent and buffer used for preparation of DNA quantification*

Components (x = number of samples+2 standards+1 extra)	Volume
Buffer	x * 199 µl
Qubit reagent (fluorophore)	x * 1 µl

Next, 190 µl of this mix was added to each of the standard tubes, while 199 µl mix was partitioned in each of the DNA-sample tubes. Then, 10µl standard DNA was added to each of the standard tubes and 1 µl DNA sample was added to each of the DNA sample tubes. While protecting the tubes from light, they were incubated for 2 minutes before they were measured in the qubit machine. If the concentration of DNA showed to be too low, DNA from the backup filters was extracted and quantified.

2.5 Amplification and library preparation

The goal of library preparation is to obtain pure DNA in a pool containing all the samples, marked with indexes. Preparing a genetic library for sequencing requires a series of steps (Table 5). The first step was to make many copies of the targeted region V4 of 18S nuclear ribosomal DNA by polymerase chain reaction (PCR). Next, the PCR products were tested for purity on an agarose gel, by gel electrophoresis. Then the PCR products were cleaned, labeled with indexes and adaptors, and cleaned once more. Finally, the samples were quantified, pooled and sent for sequencing. The procedure for library preparation followed the protocol "16S Metagenomic Sequencing Library Preparation" from Illumina (version 15044223 Rev. B) and the kit Nextera XT (Illumina) was used.

Link to Illumina protocol:

https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Table 5: *Workflow of library preparation (following protocol "16S Metagentomic Sequencing Library Preparation (Illumina)*

Workflow of library preparation
PCR amplification
PCR purification 1
Index PCR
PCR purification 2
Quantification
Pooling
Quantification

First, the samples were diluted with PCR water (Water, Mol Bio grade DNase-, RNase-, and protease-free, Germany) to obtain equal DNA concentration in all the samples. To get the desired concentration of 5ng/µl, the equation below was used (Equation 1), where C_1 is measured DNA concentration in the sample (ng/µl), V_1 is the volume needed to obtain the desired concentration, C_2 is the desired concentration of 5ng/µl and V_2 is the final volume, 100 µl

Equation 1:

$$C_1 \times V_1 = C_2 \times V_2$$

Before PCR amplification different temperatures for annealing were tested in a temperature gradient. The temperatures tested were 62, 65.2, 67.9 °C. The lowest temperature showed the most distinct and strong band, hence, 62 °C was the temperature used for annealing in the PCR reactions. To obtain higher variation, and thereby less PCR bias in the samples, two more PCR products (amplicons) for each sample was made and tested for purity on a gel. The procedure followed the protocol 16S Metagenomic Sequencing Library Preparation from Illumina (p. 6-7).

2.5.1 PCR (Polymerase Chain Reaction) Amplification

The template was amplified out of the DNA sample using primers that are universal to eukaryotes (the V4 of the 18S). In addition, adaptor overhang nucleotide sequences were added to the sequences. The primers used were forward primer V4F_illumina (5' - TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCA GCA SCY GCG GTA ATT CC- 3') and reverse primer V4R_AZig_illumina: (5' -GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAC TTT CGT TCT TGA TYR ATG A-3') The primers were published by Piredda (Piredda et al., 2017).

The mastermix was made using 12.5µl 2x KAPA HiFi HotStart ReadyMix kit (Kapabiosystems, USA) and 5 µl times the number of samples of forward- and reverse primers, respectively (Table 6). The ReadyMix kit contains polymerase enzyme, PCR, buffer, and nucleotides. The mix was vortexed and put on ice. Next, the mix was distributed in separate tubes in a PCR strip each containing 22.5µl mix. 2.5µl of the extracted DNA template (5 ng μl^{-1}) was added and a total volume of 25µl in each tube was then put in a thermocycler (Mastercycle EP gradient, Eppendorf AG, Germany).

Table 6: *Contents in the master mix. n=number of samples*

	Contents	Amount in mastermix
KAPA HiFi HotStart Ready Mix	DNA-polymerase, dNTP, $MgCl_2$ and stabilisator	12.5 µl * n
Forward primer: V4F_illumina	Oligonucleotides	5 µl * n
Reverse primer: V4R_AZig_illumina	Oligonucleotides	5 µl * n

The PCR reaction was started with the following programme: initial denaturation at 95°C for 3 min, followed by 30 cycles of denaturation at 98°C for 20 sec, annealing at 62°C for 30 sec and extension at 72°C for 1:30 min. After completed cycles, the extension/elongation was conducted at 72°C for 10 min.

2.5.2 Gel electrophoresis

The PCR products were tested for purity and correct length on 0.8% agarose gel. The gels were made out of 0.8 g agarose powder (Lonza, USA) diluted in 100 ml 1xTAE buffer (preparation of 1xTAE buffer is described in Appendix 2). The mixture was heated in a microwave and thereafter cooled down to ca 50^{o}C. Four µl of Gelred (nucleic acid gel strain, Biotium, USA) was added and the gel was poured into an electrophoresis-plate with combs that created wells in the gel. Thereafter the gel was left to congeal for 40 minutes. When the gel was ready, the combs were carefully removed. The first well was loaded with the size marker Low Range DNA Ladder (Appendix 3). The next wells were loaded with a mix consisting of 1µl loading dye/buffer (Thermo Scientific, USA) and 5µl PCR product. The gel electrophoresis was run using an Electrophoresis machine (Electrophoresis Power Supply-EPS 301, Bio-science AB, Sweden) at 80 V for 40 min. After completed gel electrophoresis, the gel was studied under UV-transillumination (Gene Genus bioimaging system, Syngene, UK). Gel picture of all the samples are shown in Appendix 4.

2.5.3 PCR purification

After a PCR reaction, DNA samples may contain undesirable impurities such as free primers and primer dimer species. Therefore, the PCR products needed to be cleaned away from primers, nucleotides, enzymes, reagents and other contaminants. This step in the Illumina protocol (p. 8) is called “clean-up 1” and the goal is to obtain pure DNA (amplicon fragments) for index PCR down the line. To clean the DNA samples, AMPure XP beads

(Beckman Coulter inc., USA) was used. They are small magnetic particles that have the ability to bind to nucleic acid. When a tube or plate containing a DNA sample is left in contact with a magnet, the magnetic beads bound to DNA will get forced to the walls of the container. This makes it possible to pipette out the unwanted fluid and clean the DNA with ethanol (80%), leaving the DNA pure.

Starting the clean-up procedure, 25 µl from each of the 30 samples of PCR product was transferred to individual wells on a 96 well plate. Then 20 µl of AMPure XP beads were added to each well. Next, the plate was placed on a magnetic stand. This forced the beads bound to the DNA, to the walls of the wells. When the supernatant had cleared, it was carefully pipetted out. The beads were rinsed with 80% ethanol. After two such washing steps, elution buffer was added to make the beads detach from the purified DNA.

2.5.4 Index PCR

To be able to recognize what sequences belong to which sample, it is necessary to mark the samples. This was done through index PCR. Index PCR is a procedure that attaches index nucleotides to the amplicon fragments. In this step (p. 10-13 in the Illumina protocol) indexes and Illumina sequencing adapters were attached to the amplicon fragments in the library. The indexes are specific to each sample, which makes it possible to associate samples to reads after pooling and sequencing. The Illumina sequencing adapters are necessary for the sequences to attach to the flow-cell during cluster generation (ref. introduction).

After index PCR the library was cleaned once more with AMPure beads and ethanol. This step (p. 13-15 in illumine protocol) is called clean-up 2 and was a repetition of the procedure done in clean-up 1. Finally, the library was fluorometrically quantified using the qubit method described earlier (p.18).

2.5.5 Pooling and quantification

All the samples in the library were diluted with elution buffer (EB) to an equal concentration of 10nM, as described on p. 16 in the Illumina protocol. Then the samples were pooled together by pipetting out 5 µl from each sample and transferred to a new tube. The pooled library went through another quality check with a Nanodrop ND-1000 spectrometer (Thermo Scientific, USA). This was done to get a "purity ratio" between DNA, RNA, and proteins. In addition, the sample was tested on an agarose gel to check for purity (strong, single band) and

correct length of approximately 600pb (ca 400 bp PCR product (amplicon) and 200 bp index and adaptor). Gel picture of the pooled sample are shown in Appendix 5. Both results were sent away for sequencing (ref. introduction) with the sample to the Norwegian Sequencing Centre at Oslo university hospital.

2.6 Bioinformatics

2.6.1 Bioinformatic pipeline

From the sequencing center, two fastq files were received for each of the 30 samples, one containing forward sequences, the other containing reverse sequences. The adapter sequence and the primers had been removed and the zipped files were downloaded from the Norwegian Sequencing Centre (NSC) ftp site and then uncompressed. Before analysis, the files were renamed and separated into new folders marked with sample name. For all analyses, an in-house modified pipeline designed specifically for filtration and quality checking of Illumina MiSeq amplicon sequences (Ramiro Logares, 2017) was used (Table 7). The required software was installed on the computing cluster, Abel.

Table 7: *Workflow of pipeline (Ramiro Logares, 2017)*

Step	Software
Error correction (correcting raw reads)	BayesHammer (HAMMER)
Merge forward and reverse reads	PEAR
Removal of sequences under 100 bp	Usearch
Quality filtering	Usearch
Merge identical? sequences	Usearch
Put sequences in correct direction (5´-3´)	HMMER2
Add identifiers to sequences	
Dereplication	Usearch 64 bits
Sequence sorting	Uparse
Removal of singletons	Uparse
Removal of chimeras	Uparse
Clustering at 98% similarity	Uparse
Generating OTU table	BLAST

The first step in the pipeline was running BayesHammer (Nikolenko et al., 2013), following (Schirmer et al., 2015). BayesHammer is an error correction tool used to correct raw reads. BayesHammer operates using a clustering algorithm (HAMMER) and constructs clusters that are ranked by Illumina quality values of the reads (Nikolenko et al., 2013). The Illumina values are predictions of the probability of an error and are generated by a quality table that uses a set of quality predictor values (Illumina). Furthermore, BayesHammer makes use of Bayesian penalties to achieve extra sub-clustering parameters (Nikolenko et al., 2013). When clustered, a new algorithm corrects sequencing errors and discard sequences that are assigned bad quality scores.

During paired-end sequencing, each amplicon is sequenced from both ends producing a forward and a reverse sequence (ref. introduction). The software PEAR was used to merge these reads (Zhang et al., 2014). PEAR scans the overlapping areas of the pairs and merges these into *one* sequence. If individual bases in the forward and reverse sequence do not match, the quality score of the base is used to decide which base is correct. Then PEAR calculates a new quality score for the altered base. After the overlapping area of forward -and reverse reads were merged, all sequences shorter than 100 bases were discarded using Usearch. Because the V4 region of 18S rRNA is a variable region, the length of the sequences are irregular, but has an average length of 390-400 bp. The restriction was set to 100 bp since all sequences under 100 bp are, in all probability, false positives. False positives can be genuine, but too short to be used in classification, which makes them useless. In other words: for BLAST to be able to classify a sequence, it needs to be of a certain length.

Sequences that were merged with PEAR, went through a quality filter in Usearch (Edgar 2010). Usearch makes use of quality scores (Phred or Q scores) to calculate the probability of one sequence being false. If the probability of the base being incorrect is P, then:

$$P = 10^{-Q/10}$$

$$Q = -10 \log_{10}(P)$$

Sequences that passed this quality check were merged. Next, reads were put into the correct 5´-3´ direction using HMMER (miTags protocol). In this pipeline, a set of HMMER-profiles of 16S, 23S, 28S, and 18S rRNA gene sequences are already made. 16S and 23S are

components of the small subunit and the large subunit of prokaryotic ribosomes respectively. 18S is, as mentioned, a part of the SSU of eukaryotic ribosomes whereas 28S is a sequence of the large subunit of the eukaryotic ribosomes. Based on the HMMER sequence profiles 18S sequences were identified and kept.

Before collecting all the samples into *one* file, identifiers were added to the sequences. That is, sample names were added to the sequence header. Then a new file was created, containing all the samples. Next, sequences were checked for replicated reads. Replicates are sequences that are identical or almost identical. Replicates can be of same (full length) or different length, or the starting or the endings of the sequences can be missing, called prefix. Sometimes both starting *and* ending of the sequence are missing. This is called substring. The process of eliminating replicates is called dereplication (Figure 7) and was conducted using Usearch 64bits. The longest replicate is kept and the result is only unique sequences. These sequences are marked with the number of replicates it originally existed of that sequence.

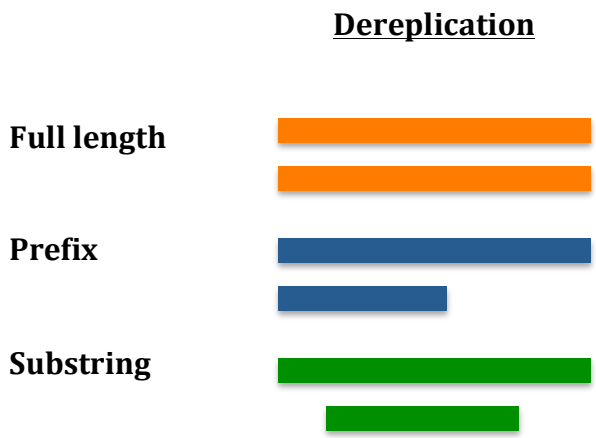


Figure 7: *Dereplication*

Using Uparse, the unique sequences were sorted by the number of replicates the sequence had before dereplication. Then singletons were discarded. A singleton is a read that occurs once in *one* sample. A sequence showing up *once* in all of the samples is most likely a sequencing error, PCR error or a contaminant and these were therefore removed.

Chimeras were removed from the dataset using Uparse. Chimeras are DNA sequences originating from two different organisms. They can form early in the PCR reaction, during annealing, and they can, therefore, be amplified during the rest of the PCR reaction. Formation of chimeras usually happens in the conserved regions of the gene where the PCR

product can anneal to a DNA fragment from another organism. In that way, chimeras are hybrid products from two or multiple parent DNA sequences that should be removed.

Finally, the data was ready to be grouped into clusters of similar sequence variants, Operational Taxonomic Units (OTUs). An OTU is a proxy for a species since the organism has not been observed. The sequence that has been replicated the most times (the one with the most replicates) is called a representative sequence and forms the basis for the first cluster. If a read matches the existing OTU within a similarity threshold, the OTU abundance is updated (Edgar, 2013). The similarity threshold used in this study was 98%. The decision on defining the OTUs at 98% similarity level was made based on Caron et al., 2009, which found that 98% similarity is suitable for discrimination on species-level. Another study, by Gérikas Ribeiro et al., 2018, also used the clustering level of 98%. The latter study has used the same sequencing primers as used in the present study. This is considered to constitute sufficient reason to use this level of clustering.

A taxonomic assignment of eukaryotic OTUs was generated by BLASTing (Altschul et al., 1990) OTUs or representative sequences against the Protist Ribosomal Reference database (PR2) (Guillou et al., 2013). The abundance (number of Illumina reads) of each unique sequence was mapped back to OTUs in each sample. Based on this mapping an OTU table with the abundance for each OTU in all samples was generated.

2.6.2 Data filtering

From the complete OTU table, OTUs with <10 reads and OTUs with a frequency of one were manually eliminated from the dataset. This is because these sequences most probably are formed from PCR errors happening during amplification. Further, the OTUs belonging to Metazoa, Fungi, and Streptophyta (higher plants) was excluded. This was done to obtain a dataset free of the large non-protist groups.

Sub-sampling of the dataset was done to normalize the dataset to equal sample sizes by rarifying using the "rrarefy" package: vegan (Oksanen et al., 2018) function in R software v.0.98.1087 (R Development Core Team 2017). The workflow of sub-sampling starts with identifying the lowest number of reads found in a single sample. Then, starting at zero sequences, random sequences are added until reaching the desired amount. The probability of the representation of an OTU in the selection depends on how many sequences this OTU

initially had. The result is a sub-sampled dataset with an equal amount of reads per sample. R-script made for sub-sampling are shown in Appendix 6.

2.6.3 Statistical Analysis

Line plots depicting hydrographical data and bar graphs displaying reads of potentially harmful algae was made in Microsoft Excel for Mac 2011, v.14.7.3. All other statistical analyses were carried out using R software v.0.98.1087 (R Development Core Team 2017).
For constructions of the treemaps, the package "treemap" (Tennekes & Ellis, 2017) was used. For making histograms, bar plots and heat maps the packages "ggplot2" (Wickham et al., 2018), “rlang” (Henry, L. et al., 2018) and “reshape” (Wickham, 2018) were used.

OTUs assigned to classes of potentially harmful algae were sorted out on the basis of table 2 (in the introduction). These OTUs were extracted from the sub-sampled dataset. A check for OTUs belonging to some of the harmful taxa, in the original dataset, was performed. Two OTUs of the classes *Karenia* and *Karlodinium* were discovered in the original dataset and not in the sub-sampled dataset. However, these OTUs had negligibly 3 reads each with a frequency of 1 and are therefore not further dealt with. Percent of total reads for the OTUs of harmful taxa, in each sample, was presented in a heatmap. To make numbers fit into the figure and become more readable, the proportional abundances were transformed to the square root of the percent of total reads.

2.6.4 Construction of phylogenetic trees

To confirm classification, a phylogeny of 13 of the most abundant OTUs of harmful algal taxa was made. Reference sequences were found by manually BLASTing (Altschul et al., 1990) the OTU sequences up against the reference database PR2 (Guillou et al., 2013). One to four of the best matches were added as reference sequences. Further reference sequences were found by BLASTing the relevant algal genus in PR2 and picking out common species. All reference sequences were aligned using MAFFT v 7.3.88 (Katoh & Standley, 2013). OTU sequences were aligned to the existing alignment of reference sequences. All phylogenetic trees were constructed as PhyML trees (settings: GTR, estimate) with 500 bootstraps in the programme Geneious Prime v.2019.0.3.

3 Results

After initial filtration of raw reads, removal of OTUs with less than 10 reads and removal of OTUs with a frequency of one (OTUs that appeared only in one sample), the dataset consisted of 5.048.846 reads assigned to 1.584 OTUs (Appendix 7).

3.1 Taxonomic composition and relative abundance

Figure 8 displays all the groups, at phylum level, that were detected in all the 30 samples before removal of large non-protist groups and sub-sampling. In addition, relative abundances (relative proportions of reads) are shown. All the OTUs were grouped into 26 taxonomic groups divided on six supergroups.

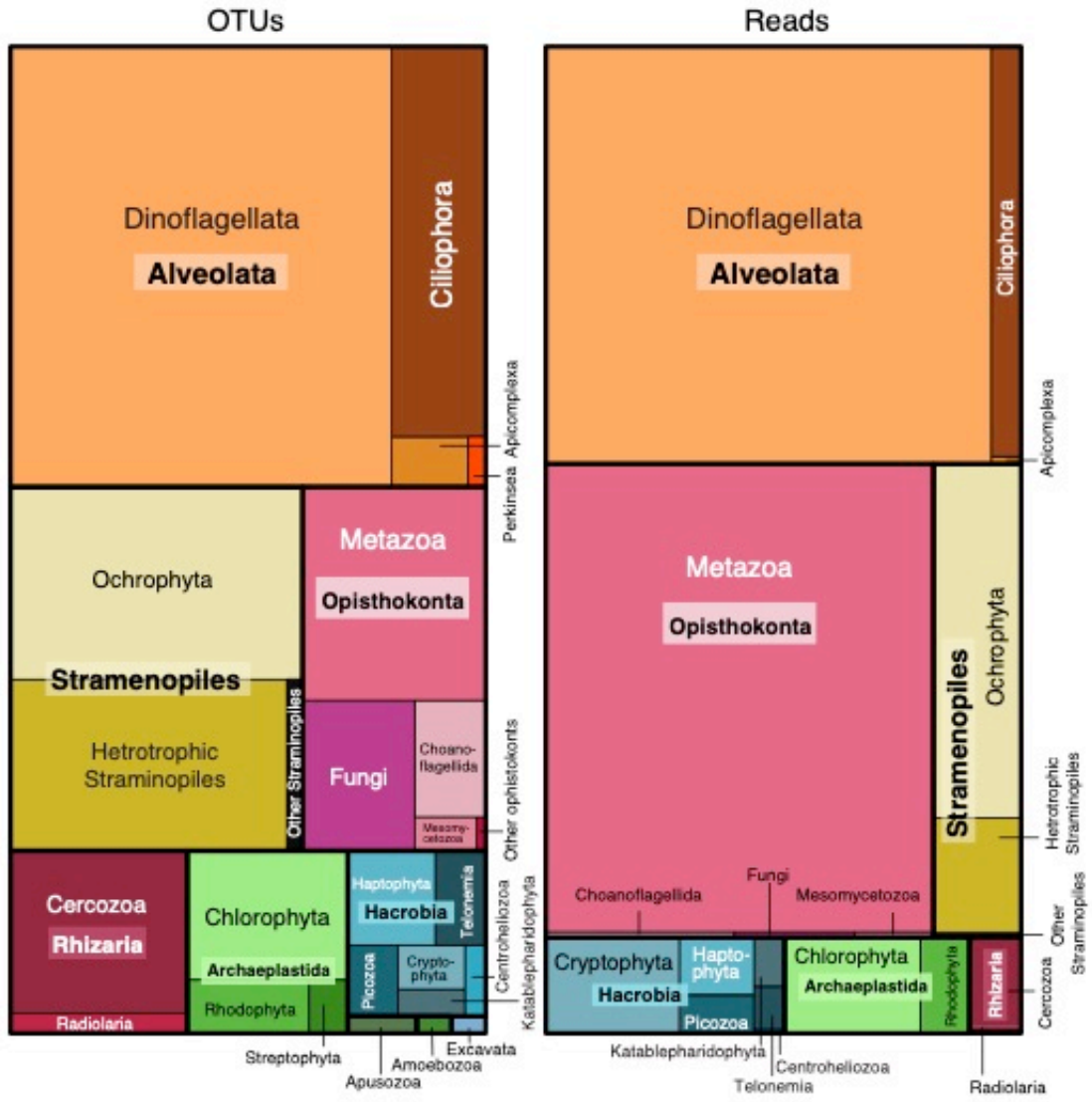


Figure 8: *Tree-maps displaying relative OTU richness (left) and the proportion of read abundance (right) for all the phyla/taxonomic groups in the original dataset. This includes samples from both sampling depths.*

The number of OTUs indicates the species richness of the groups. The proportion of reads represents a measure of their relative abundance in all samples together. In the original dataset, the supergroup Alveolata had the highest OTU richness and abundance. In total 707 (≈45%) of all the OTUs and 2135784 (≈42%) of the reads belonged to this supergroup. The largest group within Alveolata was Dinoflagellata that dominated both in richness, with 567 (≈36%) OTUs, and abundance, with 2000859 (≈40%) of reads. After Alveolata, the supergroup Stramenopiles made up the next largest fraction in terms of OTU richness. Within this group, we find the important group of microalgae, the diatoms, within phylum Ochrophyta. The diatoms were represented by 107 OTUs in this original dataset.

In total 225 OTUs were assigned to the supergroup Ophistokonta. These OTUs were distributed between phylum Metazoa (multicellular animals), Fungi, choanoflagellates, Mesomycetozoea (including the parasitic Ichthyosporea), and one OTU belonging to an unclassified opisthokont. All together kingdom Opisthokonta had an OTU richness of ≈14% and the relative abundance was ≈39% of reads. Noteworthy, the phylum Metazoa accounted for 99% of the reads within Ophistoconta. The most important OTUs making up the group Metazoa was assigned to the crustacean class Maxillophora accounting for 34 OTUs and 1.805.478 reads. In fact, the two most abundant OTUs in the whole original dataset belonged to this class that covers copepods amongst others. A manual BAST search in the database NCBI, on the two most abundant OTUs, resulted in copopods as best hits for both of these sequneces.

3.1.1 Filtered dataset (protist dataset)

After removing the large non-protist phyla of Metazoa, Fungi, and Streptophyta (Table 8), the dataset had 1,388 OTUs and 3,088,389 reads (Appendix 7). This resulted in a different composition, displayed in Figure 9. From originally making up a ≈14% of OTUs and ≈39% of reads in the original dataset, both in richness and abundance, Opisthokonta now only had 36 OTUs and ≈0.5% of reads in what is here called the protist dataset.

Table 8: *Number of OTUs and reads removed in the process of data filtering*

Data filtering		Number of OTUs	Number of reads
Opisthokonta	OTUs of kingdom **Metazoa**	132	1,952,516
	OTUs of kingdom **Fungi**	57	7479
	OTUs of kingdom **Streptophyta** (higher plants)	7	462
	Sum	≈12.4% of total	≈38.8% of total

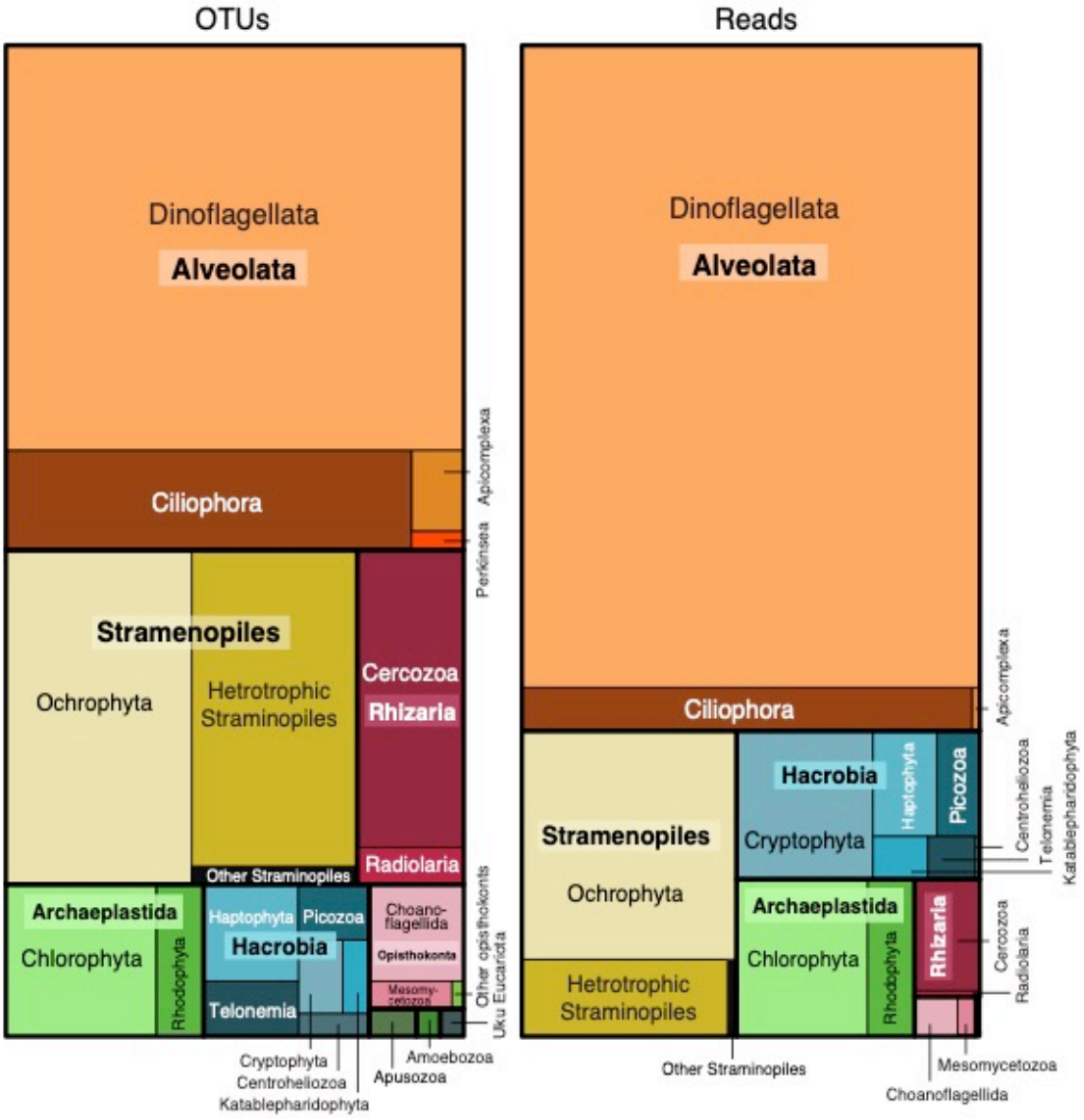


Figure 9: Tree maps displaying OTU richness (left) and relative abundance of reads (right) of the major protist phyla in the dataset. This includes both depths.

In the dataset where the large non-protist phyla have been removed (the protist dataset), the Dinoflagellates had ≈65% of the reads. Further, Ochrophyta accounted for ≈11% of the reads followed by Chlorophyta, Cryptophyta, Ciliophora and Heterotrophic Stramenopiles with ≈4.5%, 4.5%, 4.3%, and 3.5% reads respectively.

3.1.2 Sub-sampled dataset

After normalization to equal samples size (the smallest sample size) by rarefication the dataset was reduced to 1.315 OTUs and 315.900 reads (Appendix 7). In this sub-sampled dataset, each sample had an abundance of equal amounts of reads, 10.530, making the 30 samples comparable. A table displaying the 30 most abundant OTUs (Table 9) gives an indication of the fact that very few OTUs had many reads while many OTUs had few reads. The top five OTUs alone accounted for ≈41% of all reads.

Table 9: *Table showing the sum of reads and % of total reads in the 30 most abundant OTUs in the sub-sampled dataset. The taxa with the best mach to the sequence and the % suport for this match is included.*

OTUs	Sum of reads	% of total reads	Phylum	Taxa with best match	% similarity to best match
OTU 3	39425	12,48	Dinoflagellata	*Tripos* sp.	99
OTU 4	33705	10,67	Dinoflagellata	*Heterocapsa* sp.	100
OTU 29	32421	10,26	Dinoflagellata	*Dinophyceae* sp.	99
OTU 10	14257	4,51	Dinoflagellata	*Suessiales* sp.	100
OTU 8	11074	3,51	Dinoflagellata	*Alexandrium hiranoi*	100
OTU 15	9420	2,98	Ochrophyta	*Thalassiosira* sp.	100
OTU 11	8111	2,57	Cryptophyta	*Teleaulax amphioxeia*	100
OTU 1033	5930	1,88	Dinoflagellata	*Dinophyceae* sp.	99
OTU 14	5911	1,87	Dinoflagellata	*Dinophyceae* sp.	100
OTU 103	5134	1,63	Dinoflagellata	*Tripos fusus*	99
OTU 12	4608	1,46	Ochrophyta	*Cyclotella* sp.	100
OTU 22	3822	1,21	Dinoflagellata	*Dinophysis acuminata*	99
OTU 13	3792	1,20	Ochrophyta	*Chaetoceros* sp.	100
OTU 51	3174	1,00	Dinoflagellata	*Dinophyceae* sp.	100
OTU 16	3119	0,99	Chlorophyta	*Bathycoccus prasinos*	100
OTU 38	2887	0,91	Dinoflagellata	*Dinophyceae* sp.	95
OTU 56	2785	0,88	Het. Stramenopiles	MAST-1C X sp.	100
OTU 43	2654	0,84	Dinoflagellata	Dino-Group-I sp.	100
OTU 37	2557	0,81	Dinoflagellata	*Alexandrium margalefii*	99
OTU 18	2380	0,75	Rhodophyta	*Stylonema alsidii*	89
OTU 30	2379	0,75	Dinoflagellata	*Gyrodinium* sp..	100
OTU 19	2354	0,75	Chlorophyta	*Chlorodendrales* sp.	100
OTU 25	2255	0,71	Cryptophyta	*Cryptomonadales* sp.	100
OTU 17	2219	0,70	Ciliophora	*Cyclotrichium* sp.	100
OTU 20	2191	0,69	Dinoflagellata	*Dinophyceae* sp.	100
OTU 70	2159	0,68	Dinoflagellata	*Dino-Group-I* sp.	100
OTU 21	2138	0,68	Ochrophyta	*Skeletonema marinoi*	100
OTU 27	2122	0,67	Haptophyta	*Emiliania* sp.	100
OTU 638	2078	0,66	Dinoflagellata	*Dinophyceae* sp.	99
OTU 31	2077	0,66	Ciliophora	*Strombidiidae* sp.	100

A histogram displaying the number of reads per OTU visualize how few OTUs in the sub-sampled dataset had many reads. Similarly, there are many OTUs that have very few reads (Figure 10). The span ranged from OTU 3, with the most reads, 39.425 to OTU 2522 with only one read.

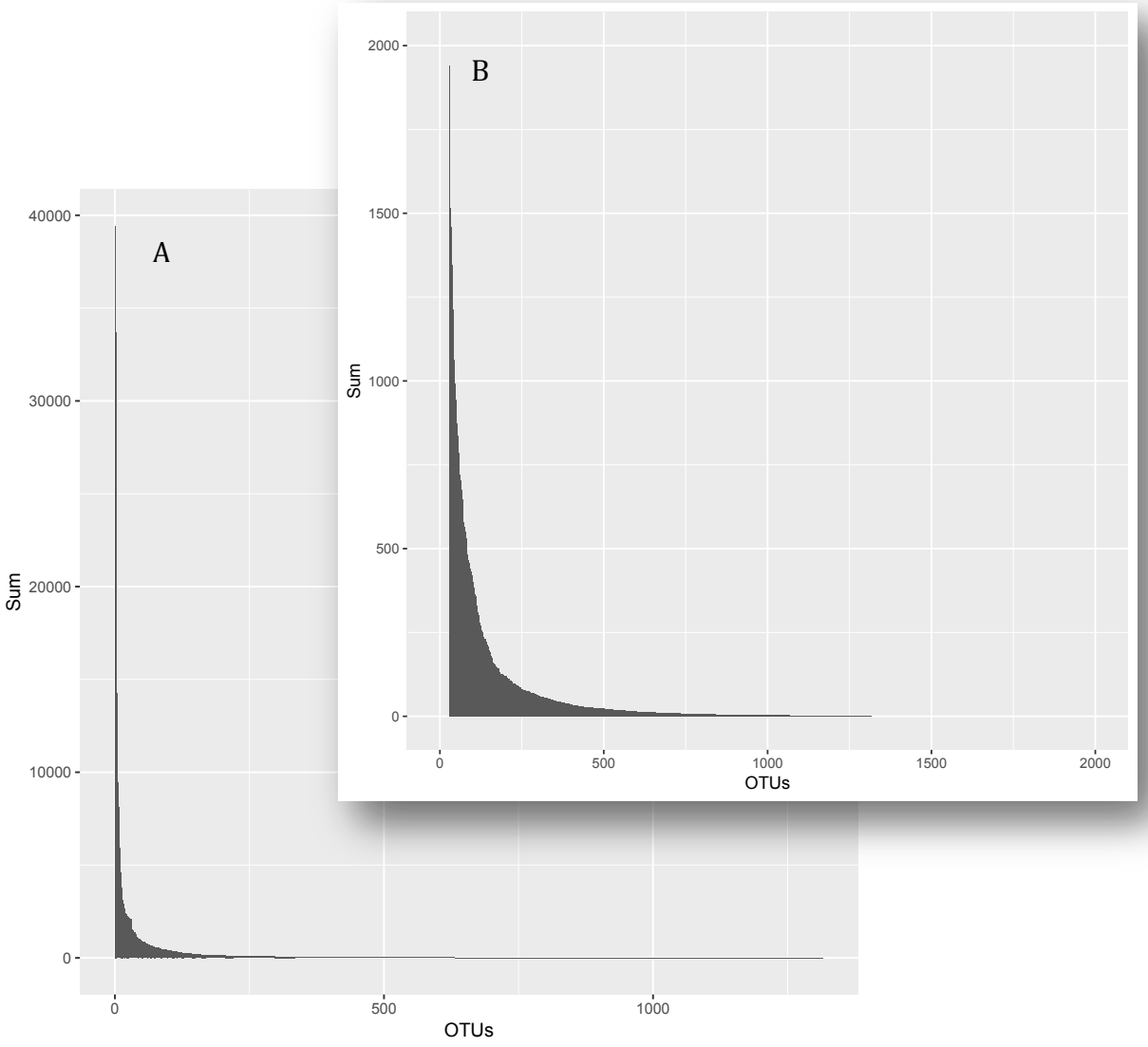


Figure 10: *Number of reads per OTU in the sub-sampled dataset. A) The complete plot with the lowest to highest values on x- and y-axis. B) Same plot with x-axis cropped to max 2000 OTUs and y-axis cropped to max sum of 2000*

All further analyses were performed on the sub-sampled protist dataset since this dataset was comparable between samples and was free of non-protist groups. The composition of supergroups and phyla in this dataset are depicted in Figure 11.

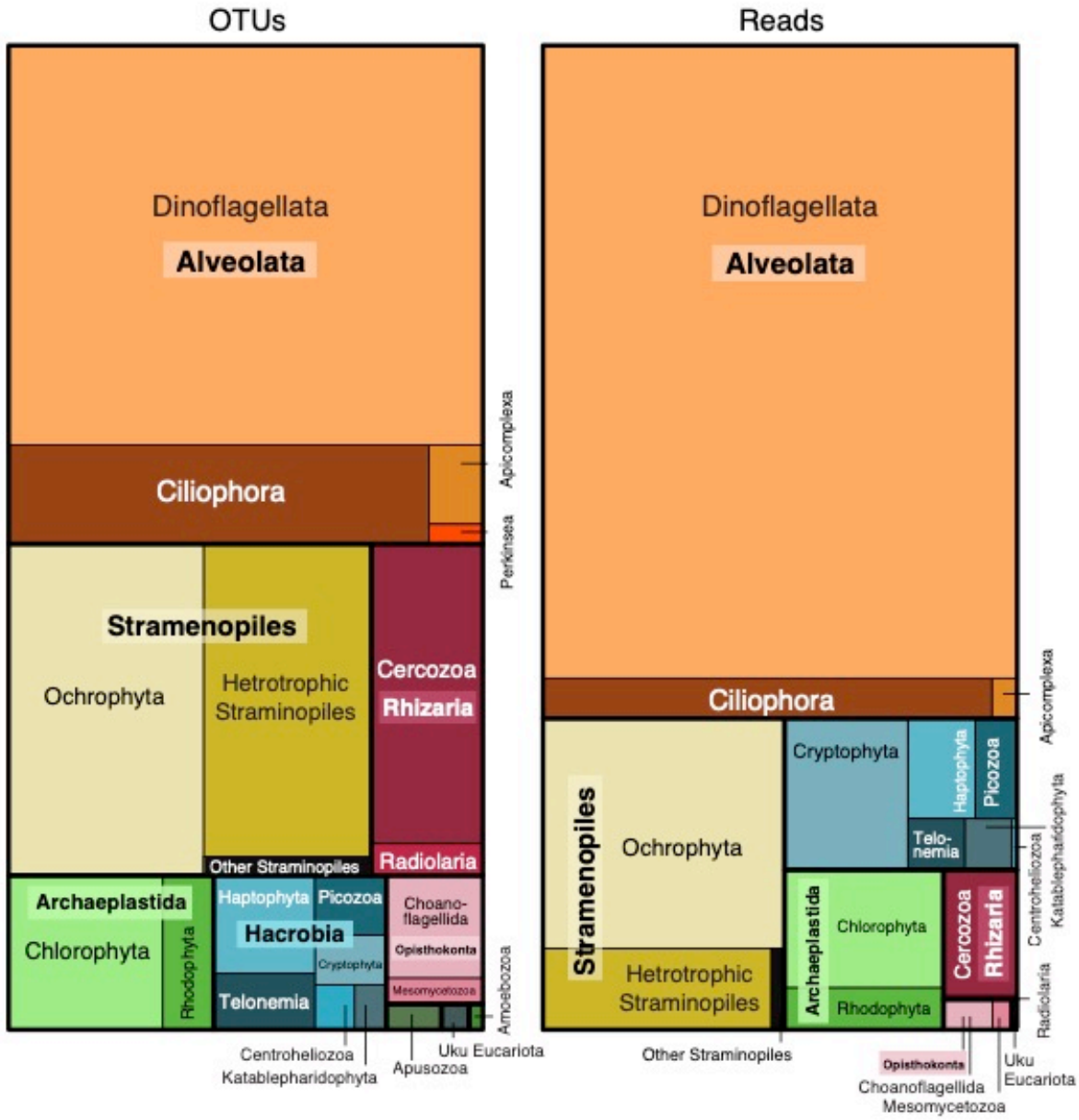


Figure 11: Tree maps displaying OTU richness (left) and relative abundance of reads (right) *for the major groups in sub-sample of protist dataset. Samples from both depths are included.*

The Alveolates still dominated both in terms of richness (≈51% of OTUs) and relative read abundance (≈68% of reads). Of the alveolate OTUs, ≈80% belonged to the dinoflagellates and in total the dinoflagellates had ≈41 % of OTUs and ≈64 % of reads. Further, 18% of OTUs within Alveolata belonged to the heterotrophic group of protists called Ciliophora,

named after their covering of hair-like cilia. The ciliates had an abundance of 4% of reads in this dataset and were represented by 117 OTUs. The remaining alveolate OTUs belonged to the parasitic groups Apicomplexa and Perkinsea, accounting for 12 and 3 OTUs respectively.

The stramenopiles, characterized by their heterokont flagella (two different flagella), had an OTU richness of 26% and an abundance of 16%. Ochrophyta and heterotrophic stramenopiles dominated this phylum, accompanied with a small fraction of a cluster of OTUs here assigned to "other stramenopiles" meaning stramenopiles that cannot be placed with certainty within heterotropic stramenopiles or Ocrophyta. This is because they might be chimeras or too short sequences. Almost half of the OTUs within the heterotrophic Stramenopiles were assigned to the groups MAST (MArine STramenopiles). This group are stramenopiles that is lacking a cultured representative (Gran-Stadniczeñko et al., 2018; Massana et al., 2004)

Archeplastida is a group consisting of algae and higher plants comprising red alga (Rhodophyta), green algae (Chlorophyta), higher plants (removed) and a group of unicellular freshwater algae called glaucophytes, not detected here. The red- and green-algae had a richness of 2%- and 5% of OTUs respectively. The relative abundance of red algae was 1% and the green algae abundance was of 4% of reads.

Hacrobia had an OTU richness of 6% of OTUs and relative abundance of 8% of reads. Hacrobia comprises haptophytes, cryptophytes, Katablepharidophyta, Picozoa, Telonemia and Centroheliozoa. Lastly, the super-group Rhizaria (cercozoans and radiolarians) had a richness of 8% of OTUs, and abundance of only 2% of the reads in the dataset. The remaining 7 OTUs were assigned to two OTUs of unclassified eukaryotes, one OTU of Amoebozoa and four OTUs of Apusozoa. Altogether the last 7 OTUs was represented by 0.03% reads in the dataset.

3.2 Seasonal variation of hydrographical data

The CTD logged measurements of conductivity (salinity), temperature, and density, as well as fluorescence by depth down to 100 m. Values were extracted from 0-2 m depth and 5 m depth. Figure 12 display curves based on the mean measurements logged from the two sampling depths for the mentioned parameters. Curves for each sampling depth separately are shown in Appendix 8.

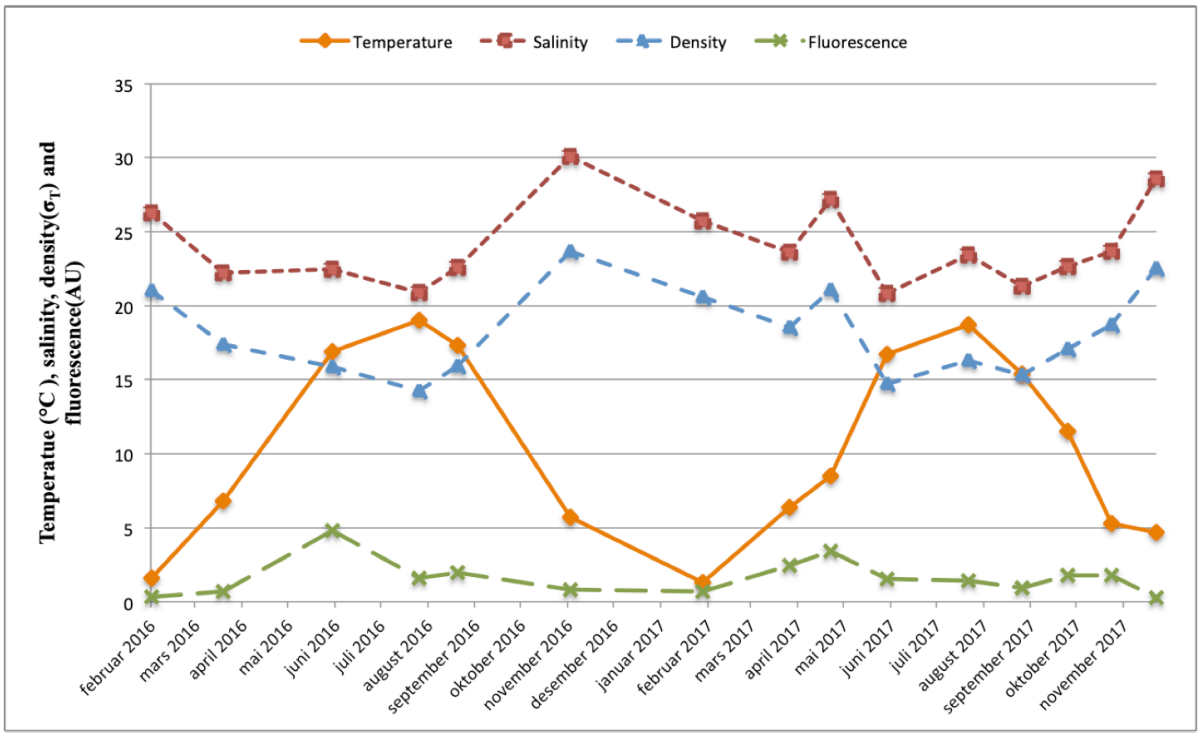


Figure 12: *The variation in temperature (°C) (orange line), salinity (brown line), density (σ_T) (blue line) and fluorescence (Arbitrary Units) (green line) over the two-year sampling period.*

The water temperature increased from February to August and decreased from August to February both years. The highest temperatures were observed in August with measurements of about 19°*C* both years. Lowest temperature measured was in February 2017 with 0.7°*C*. The temperature data showed little to no difference between the two depths, although the upper water layer was slightly warmer than at 5 m depth in April 2016 and May and June 2017 (Appendix 8).

Salinity and density varied through the year, peaking in November both years and being lowest during summer (June-September). The latter parameters showed a strong correlation throughout the year. Salinity varied between 19.9 and 30.3 whereas density varied between

14.2 and 23.8 *(σ_T)*. There was little difference between the two depths however; the salinity (and density) was generally a little lower in the upper water layer (Appendix 8). Fluorescence had a peak in the measurements from June 2016 and May 2017. Neither of the mentioned parameters showed significant differences in measurements between the two depths (Table 10).

Table 10: *Summary of correlation test (Pearson`s product-moment correlation) between measurements from the two sampling depths. No measurements differed significantly, $p<0.05$.*

	Temperature	Salinity	Density	Fluorescence
t	23.679	13.04	16.004	4.3263
df	13	13	13	13
p-value	4.467e-12	7.662e-09	6.191e-10	0.0008222

Fluorescence is a measure of algal biomass and shows large variations throughout the year. A closer look at the measurements shows that it was high in June 2016 and May 2017 (Figure 13). Also, the measurements show small peaks in September 2016 and November 2017. Noteworthy, for most of the date's, fluorescence had higher values in the deeper water layer. In June 2016 and May 2017 the difference was up to 3 AU between the surface (0-2 m depth) and 5 m depth. Only in November 2016 and September 2017 fluorescence measurements were somewhat higher at the surface than at 5 m depth.

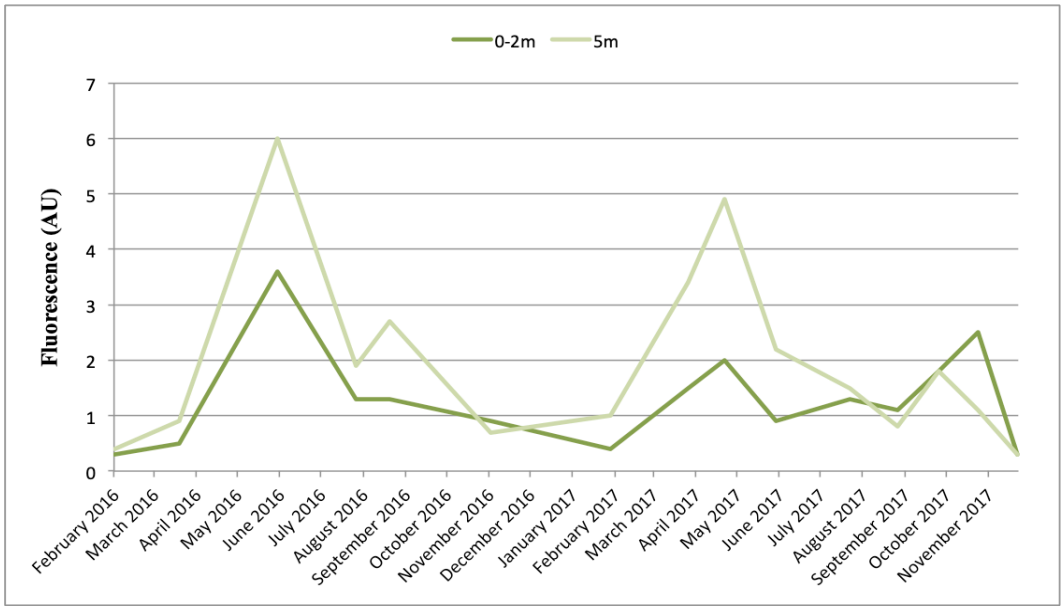


Figure 13: *The variation of fluorescence (in Arbitrary Units) measured at the two sampling depths over the two-year sampling period.*

The variation in oxygen saturation through the sampling period is shown in Figure 14. The saturation at 5 m depth exceeded 100% in the spring months of April, May and June both years. This correlates with the trends seen in the fluorescence measurements, which was high in the same period. The saturation was lower at the surface than at 5 m in the spring (April-June) in 2016. A similar trend was observed in September – November 2017.

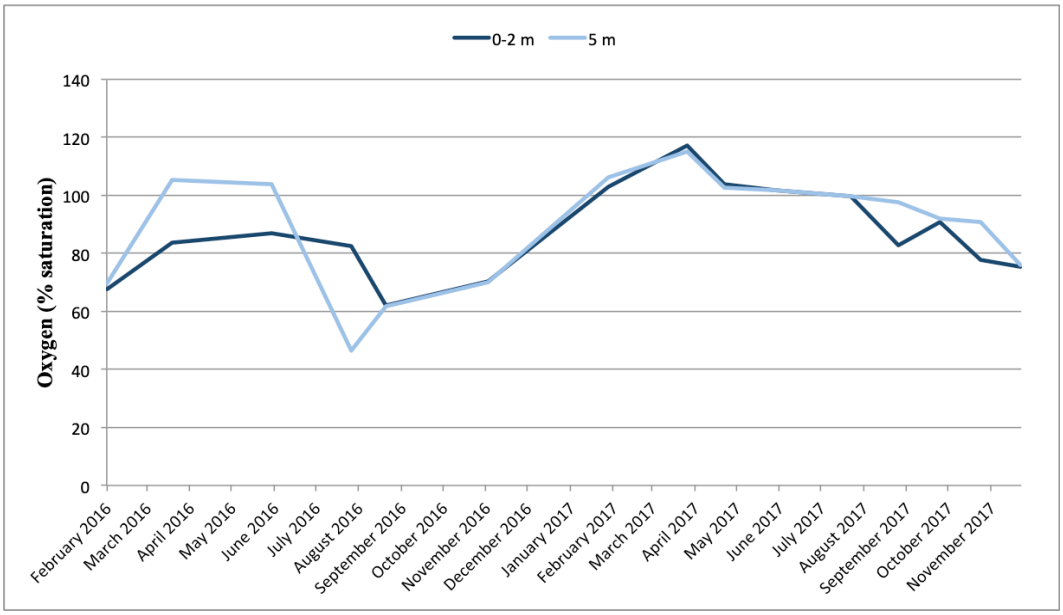


Figure 14: *The variation of oxygen saturation measured at the two sampling depths over the two-year sampling period.*

3.3 Seasonal dynamics of major groups

3.3.1 Richness of operational taxonomic units (OTUs)

The numbers of OTUs for each taxonomic group retrieved from the different sampling dates are illustrated in bar plots, Figure 15. The samples are separated by depth, visualizing the differences in richness between the surface samples (0-2 m) and 5 m depth samples. All Phyla was detected at both depths and the bars show number of OTUs per sample of the major groups (and the sum of OTUs in all samples).

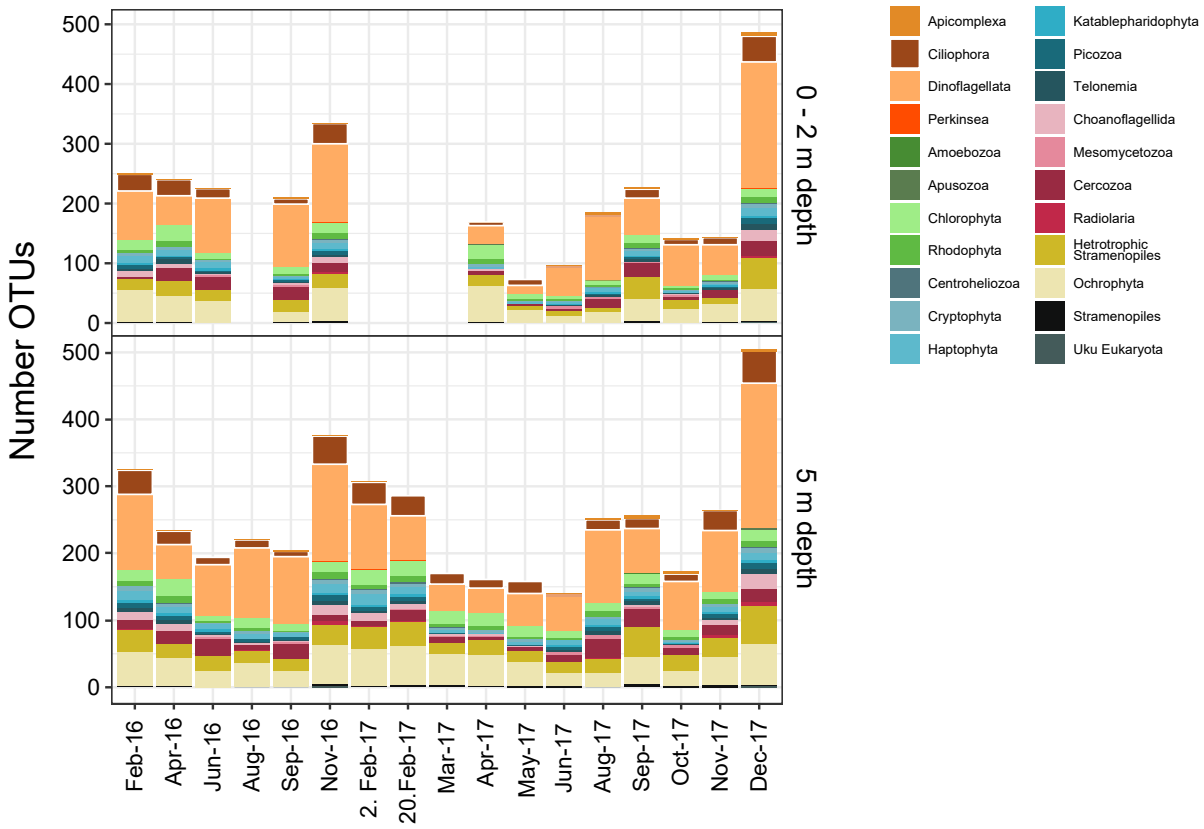


Figure 15: *Dynamics of 22 major taxonomic groups in the dataset. Top: number of OTUs per sample from 0-2 m depth. Bottom: Number of OTUs per date from 5 m depth. The descriptions to the right of the bar plot are in the same order as in the figure.*

The total number of OTUs at both depths peaked in the samples from November 2016 and December 2017. At 5 m depth, the samples from February 2016, November 2016, February 2017 and December 2017 had the highest richness of OTUs. The December-samples collected in 2017 were the ones with the highest total richness, counting almost 500 OTUs in both the surface-sample and the 5 m sample. Samples from February both years were also relatively high (above 250). The only samples with less than 100 OTUs were the ones

collected from 0-2 m depth in May and June 2017. Analogous samples from 5 m depth were also low, but still above 100 OTUs.

The phyla within the supergroup Alveolata (Apicomplexa, Ciliophora and Dinoflagellata) had the highest diversity (number of OTUs) in November 2016 and December 2017 at both depths. Lowest diversity was observed in the spring months of March to June at both depths in 2017. Dominance in richness of the phylum Dinoflagellata was applicable for most of the samples and the overall trends for this group reflect the tree-map (Figure 11). The general tendency of the heterotrophic Stramenopiles and the Ochrophyta was high diversity in November to February both years. In the months March to June 2017 OTU richness was low for these groups before they experienced a small jump in September. Relatively low richness of Stramenopile supergroup was also seen in the samples from April to September in 2016.

3.3.2 Abundance of operational taxonomic units (OTUs)

Relative read abundance of the major taxonomic groups in the dataset is shown in Figure 16. Percentage of reads of OTUs in a sample is a measure of the relative abundance of the OTUs of the different phyla/taxonomic groups.

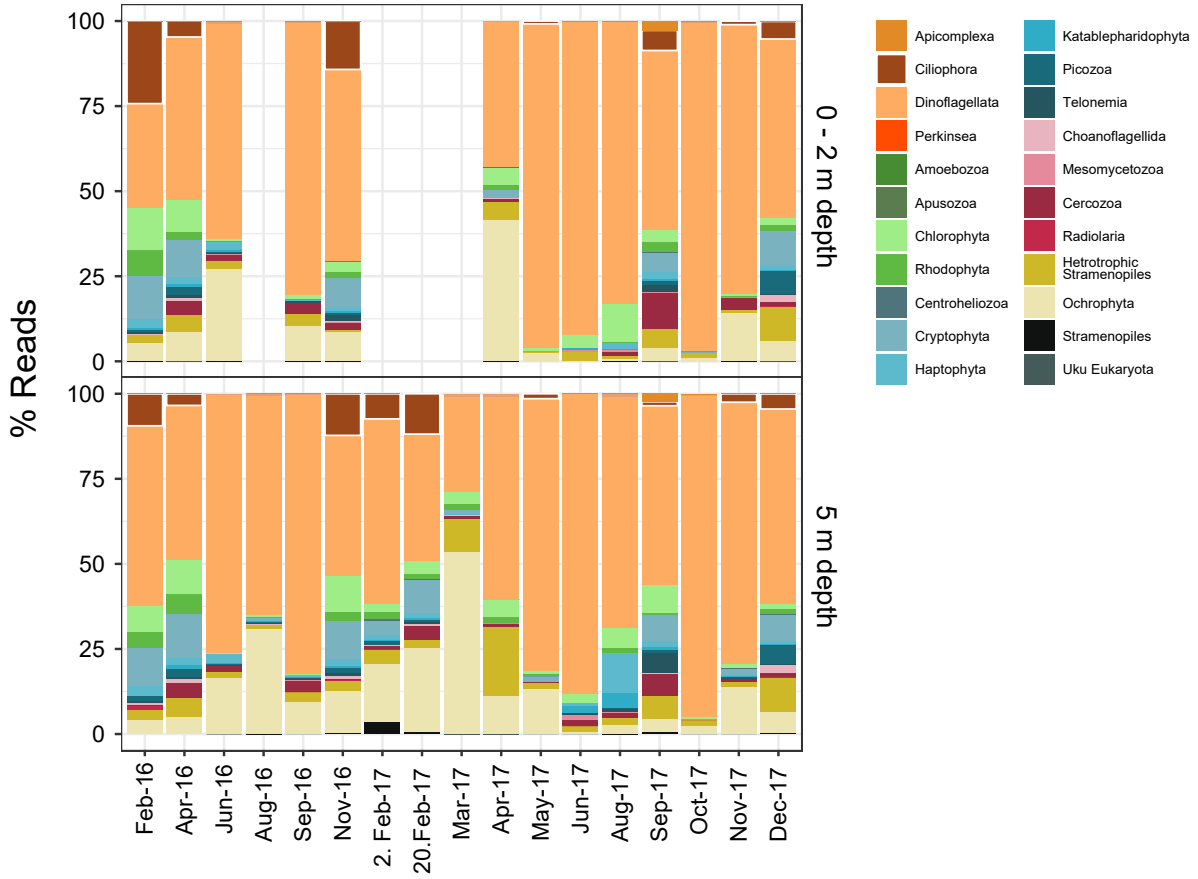


Figure 16: *Proportion of reads across the 22 major taxonomic groups in the dataset. Top: the proportion of reads per sample from 0-2 m depth. Bottom: the proportion of reads per sample from 5 m depth. The descriptions to the right of the bar plot are in the same order as in the figure.*

This figure depicts strong seasonal trends in the abundance of protist OTUs. The Alveolates dominated with the highest percentage of reads (highest relative abundance) in most of the samples, with dinoflagellates as the most important group. Highest abundance of dinoflagellate OTUs (over 30% of reads) was detected in the samples from May, June, and October and November 2017 as well as in June and September 2016. The lowest percentage of reads of dinoflagellate OTUs was detected in the March-samples from 2017. When analyzing this figure it is important to have in mind that a high percentage of reads can result from a few OTUs having very many reads.

In March and April 2017 the heterotrophic Stramenopiles and Ocrophyta generally had a high percentage of reads at both depths. April 2017 was the samples were the heterotrophic Stramenopiles had by far the highest percentage of reads compared to the rest of the samples. The phylum Ocrophyta demonstrates a trend of increasing percentage of reads in the period February to March 2017 followed by a drop in April 2017.

Comparing the two representations of OTU richness and relative abundance of OTUs show that for some of the dates they are corresponding. However, in some samples, like May, June, October and November 2017, the number of OTUs of Dinoflagellates is very low, while the relative abundance is over 75% for Dinoflagellates in these samples from 5 m depth. This indicates that one or just a few OTUs of Dinoflagellates were very dominant in these particular periods.

3.4 Temporal variation of harmful algae

In total, 26 OTUs of 11 genera of algae that are known to be harmful or potentially harmful in Norwegian waters was detected in the data. Proportional abundances of these OTUs are shown in Figure 17 for the surface samples (0-2 m) and in Figure 18 for the 5 m depth samples. Number of reads for each OTU can be seen in Appendix 9 and 10 and classification in the OTU table, given by the database PR2, is shown in Appendix 11.

Figure 17: *Heat-map representing the temporal variation of the 26 OTUs representing harmful algae sampled at 0-2 m depth, over the two-year sampling period. Proportional abundances are transformed to the square root of the % of all reads.*

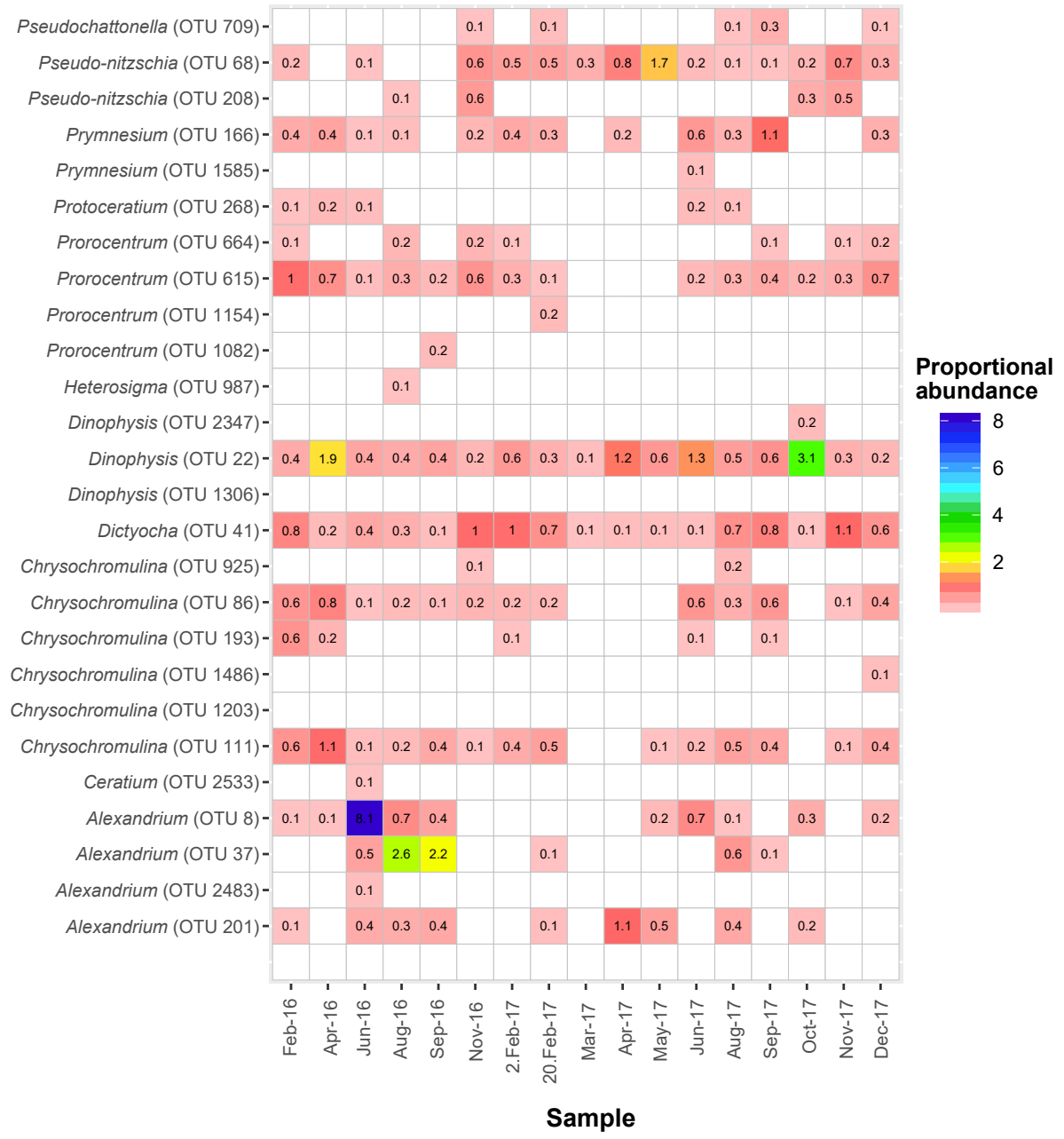


Figure 18: *Heat-map representing the temporal variation of the 26 OTUs of harmful algal classes from 5 m depth, over the two-year sampling period. Proportional abundance is transformed to the square root of the % of total reads.*

Out of all the OTUs within genera of harmful algae OTU 8 (*Alexandrium*) had the highest abundance, peaking in June 2016 and 2017, with the highest abundance in the upper water layer. Four OTUs of the genus *Alexandium* was detected and three of these (OTU 8, 37 and 201) showed very high relative abundance at times (Figure 17 and 19).

Three OTUs of *Dinophysis* was present in the OTU table. Several species within this genus are known to cause gastrointestinal illness in humans from diarrhetic shellfish poisoning (DSP) (Reguera et al., 2014). Two out of these was only present in one sample each, however, one OTU, 22, was very abundant and present all year at 5 m depth. Genus *Dictyocha* was found by one OTU (OTU 41) also present in all samples from 5 m depth and many samples taken from the surface. This OTU was classified by BLAST searches in PR2 to *Dictyocha* sp. Furthermore, six OTUs, assigned to genus *Chrysochromulina,* was detected. Of these, OTU 86 and 111 was the ones with the highest proportional abundance.

Pseudochattonella was represented by one OTU in the dataset. In the database PR2, this OTU was assigned to the ichthyotoxic species *Pseudochattonella farcimen*. Ichthyotoxic means toxic to fish and blooms of this species has earlier been associated with fish kills (Andersen, N. G. et al., 2015). However, this OTU had very low proportional abundance and these results are interpreted to show no bloom of this species. Further, two OTUs of the diatom genus *Pseudo-nitzschia* was detected in the data. Of these, OTU 68 was the most abundant and was present in all of the samples, except in September 2016. PR2 identified this OTU as *Pseudo-nitzschia seriata.* However, a manual BLAST search revealed that the sequence had best match to many species of *Pseudo-nitzschia*, including *P. subcurvata* and *P. arenysensis*. OTU 208 was assigned no further than *Pseudo-nitzschia* sp. This algal genus can produce the toxin domoic acid causing amnesic shellfish poisoning (ASP) (Trainer et al., 2012), and blooms of species within this genus can, therefore, cause health problems. OTU 68 showed peaks at both depths in March 2017 and was also relatively abundant at 5 m depth in April 2017. The haptophyte genus *Prymnesium* was represented by two OTUs in the dataset. In the OTU table, both of these were assigned to *Prymnesium* sp. OTU 166 was the most abundant one, peaking in September 2017.

One OTU (268) of *Protoceratium* was detected in the data; this OTU had quite low abundances, but was present in February to November 2016 and appeared once more in June and August 2017. Genus *Prorocentrum* was represented by four OTUs in the dataset, were one of these (OTU 615) was abundant, appearing in most samples from both depths, except March to May 2017. The raphidophyte genus *Heterosigma* (phylum Ochrophyta) was represented by one OTU in the sub-samples dataset. This was assigned by PR2 to the toxic phytoplankton, *Heterosigma akashiwo* (Appendix 11). However, this OTU was in very low relative abundances and appeared only in two samples: 5 m depth in August 2016 and 0-2 m depth in November 2016.

In summary, the vast majority of the OTUs of harmful algae were detected at both depths. Exceptions were OTU 1306 (*Dinophysis*) and 1203 (*Chrysocromulina*) that was detected only at 0-2 m depth, and OTU 1154 (*Prorocentrum*), 2347 (*Dinophysis*), 1486 (*Chrysocromulina*), 2533 (*Ceratium*) –and 2483 (*Alexandrium*) that was detected only at 5 m depth.

The temporal variations of the eight most abundant OTUs of harmful algal genera were sorted out from the OTU table. These are displayed in bar charts depicting the percent of total reads that they constituted in the different samples (Figure 19).

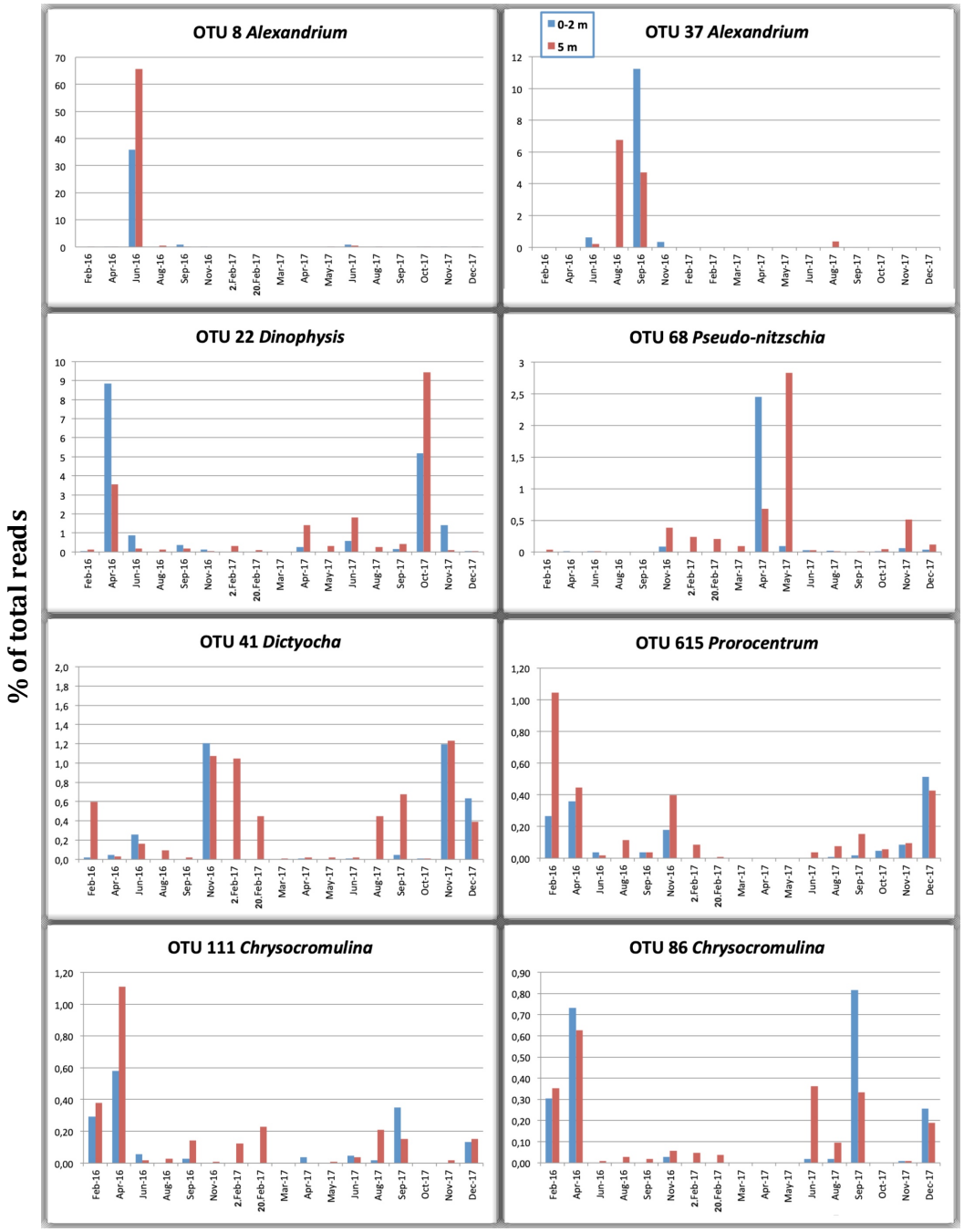


Figure 19: *Temporal variation in % of total reads of the 8 most abundant OTUs assigned by PR2 to genera of potentially harmful algae. Blue bars represent samples collected at 0-2 m depth. Red bars represent samples collected at 5 m depth.*

OTU 8, *Alexandrium*, accounted for 66% of total reads in the 5 m sample from June 2016. In the surface sample the same date it accounted for 36% of all reads. This OTU was not substantially present in almost none of the remaining samples. In the following samples (August and September 2016) **OTU 37** appeared in quite high abundances: 6.7%- and 11% in August 2016 and September 2016 respectively.

OTU 22, *Dinophysis*, was detected in all gathered samples except in the surface sample from August 2017. In two of the samples this OTU, of genus *Dinophysis*, had a very high abundance. It accounted for around 8.9% of total reads in April 2016 and 9.4% October 2017. Noteworthy abundance was highest in the surface sample in April 2016 while in 2017 the abundance was highest in the 5 m depth sample.

Pseudo-nitzschia, represented by **OTU 68**, peaked in April and May 2017. The percent of reads was 2.5% in the surface sample in April and 2.8% in the 5 m depth sample in May. This OTU showed generally low abundances in the rest of the samples.

OTU 41 of *Dictyocha* peaked in the November-samples of both years. When peaking, this OTU was about equally high on abundance at both depths. Highest % of total reads was measured to 1.2 in 0-2 m depth in November 2016 and similarly, 1.2 in 5m depth in November 2017. After the peak in 2016, it seems that this OTU decreased gradually during the coming month of February. Before the peak in 2017, abundance increased during August and September. This OTU was present in most of the samples, and was, for the most part, detected at 5 m depth.

OTU 615 (*Prorocentrum*) was present in most of the samples, except March, April and May 2017. At these dates, this OTU was absent at both depths. It showed a peak in February 2016 at 5 m depth by 1% of total reads.

Finally, the two most abundant OTUs of *Chrysocromulina* (**OTU 86 and 111**) both showed peaks in April 2016 and September 2017. OTU 111 was the one with the highest abundance of the two, reaching 1.2% of total reads at 5 m depth in April 2016.

3.4.1 Difference between sampling depths

A statistical analysis of similarities (ANOSIM) was run to check for differences in harmful algae between the two depths (Figure 20). The test is based on a dissimilarity matrix between two samples (Bray-Curtis dissimilarity). The two samples in this case are total reads of harmful algae from each depth. This test is a non-parametric statistical test, showing that there is no statistical significant difference in terms of reads in harmful algae between the two depths (R=-0.037, P-value: P>0.05). The R-value indicates how strongly they are different from each other. The R-value is close to zero, which represents the null hypothesis of no difference between the depths.

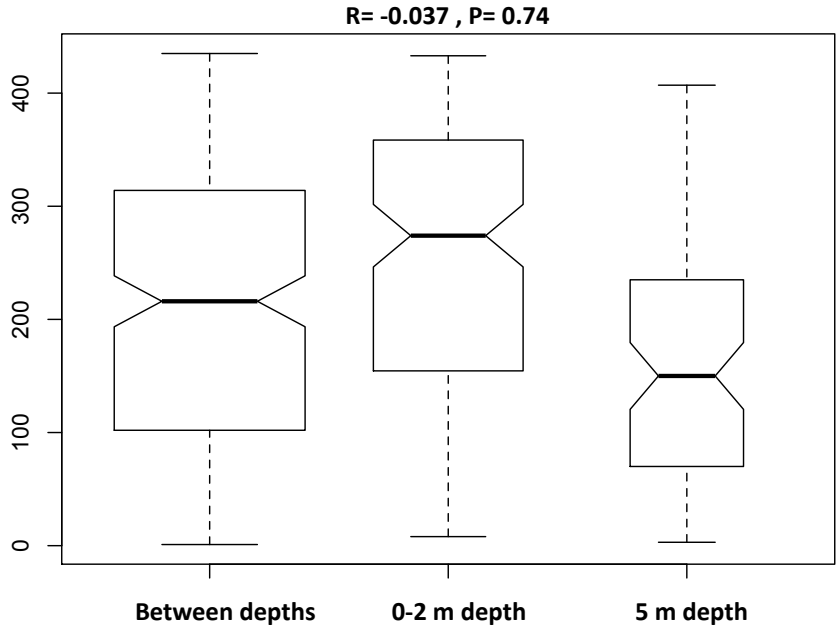


Figure 20: *Boxplot of ANOSIM test of the difference between the two depths in the total number of reads of the 26 OTUs of harmful algae. The P-value is the significance level. The R-value indicates the strength of the factors.*

3.5 Taxonomic affiliation

To confirm species classification, 13 of the most abundant OTUs were identified by alignment and construction of phylogenetic trees. Alignments were created as MAFFT-alignments of the reference sequences before adding OTU sequences to the existing alignment. For the 13 OTUs, a total of 7 phylogenetic trees of the genera *Dinophysis (one OTU), Dictyocha (one OTU), Pseudo-nitzschia (two OTUs), Prymnesium (one OTU), Prorocentrum (tow OTUs), Chrysochromulina (three OTUs)* and *Alexandrium (three OTUs)* were constructed (Figures 21-27).

OTU 22, *Dinophysis*, was placed within a monophyletic group together with two reference sequences of the (cultured) dinoflagellate *Dinophysis norvegica* supported by low values (Figure 21). The alignment revealed that apart from one base difference in the very beginning of the sequence, OTU 22 was identical to all the reference sequences in the uppermost monophyletic group with the support value 93. On the basis of the alignment and the phylogenetic tree, OTU 22 cannot be unambiguously identified as any of the following species: *D. fortii, D. infundibulum, D. acuta, D. acuminata, D. norvegica D. caudata, D. tripos, and D. miles*.

OTU 41 was placed in a monophyletic group with *Dictyocha speculum*, *D. octonaria*, and *D. fibula*. Support for this group was high, 92 (Figure 22). Further, OTU 41 was placed in a monophyletic group with *D. octonaria and D. speculum.* The MAFFT-alignment showed that the reference sequence of *D. speculum* and the sequence of OTU 41 were identical in the V4 region and that two base differences separated *D. Octonaria* from *Dictyocha speculum*. There is a high probability that OTU 41 is a sequence belonging to the species *Dictyocha speculum.*

OTU 68 was placed in a monophyletic group shared with two reference sequences of a diatom, *Pseudo-nitzschia delicatissima* (Figure 23). Support value for this position was 86. The branch below this monophyletic group was an environmental sequence of a Stramenopile that differed from OTU 68 with 7 base pairs in the alignment. The latter sequence was the best BLAST hit for the OTU sequence in PR2. The alignment showed that the two sequences of *P. delicatissima* were identical to OTU 68 in the V4 region. On this basis, OTU 68 is most likely a *P. delicatissima.* **OTU 208** fell into a monophyletic group with *P. calliantha* with support value 96. No base pair differences were found between the OTU sequence and the reference sequence of *P. calliantha.*

OTU 166 was assigned by BLAST in PR2 to a species of the genus *Prymnesium*. In the phylogenetic tree (Figure 24) this sequence was placed beside environmental sequences of Haptophyta. The closest cultured species sequence was of *P. kappa.* Two base pairs distinguish this species from the sequence of OTU 166.

OTU 664 and **615** were both classified closest to an environmental sequences of dinoflagellate genus *Prorocentrum* sp. (Figure 25). Even though both of these sequences were close to the common species *Prorocentrum minimum* in the tree, manual BLAST searches in the database NCBI resulted in best hits of uncultured eukaryote clones for both OTUs. The affiliation of OTU 664 and 615 therefore remains unknown.

Three OTUs of the taxon *Chrysochromulina* were placed in a taxonomic tree (Figure 26). This phylogeny classifies **OTU 86** to *Chrysochromulina simplex* with support value of 100. The alignment shows two differences in base pairs between the OTU sequence and the *C. simplex* in the V4 region of the 18S. Further, in the tree, **OTU 193** was positioned close to *C. leadbeateri and C. simplex*. When comparing the sequence of OTU 193 and the sequence of *C. leadbeateri,* 18 base differences were detected. OTU 193 and *C. simplex* differed by 23 base pair differences. **OTU 111** fell into a sister branch to the large monophyletic group possessing the mentioned OTUs among other common *Chrysochromulina* species. Closest to OTU 111 we find sequences of cultured species of *C. cymbium, C. campanulifera, and C. strobilus.* In the alignment, the three latter sequences were identical in the V4 region and OTU 111 differed from these with two base pairs. The conclusion is that the sequence of OTU 111 most probably belongs to one of these three species.

Three most frequent OTUs of dinoflagellate genus *Alexandrium* were placed in a phylogenetic tree seen in Figure 27. **OTU 37** was located beside *A. margalefii* supported by a value of 100 however, the sequences were not entirely identical in the V4 region, differing with two bases. **OTU 201** was placed in a well-supported group with *A. ostenfeldii* and was believed to be this species. The very abundant **OTU 8** was placed next to sequences of *A. hiranoi* and *A. Pseudogonyaulax.* All sequences of the latter species were similar in the alignment and OTU 8 was no different from these in any positions. This means that OTU 8 is either *A. hiranoi* or *A. pseudogonyaulax*.

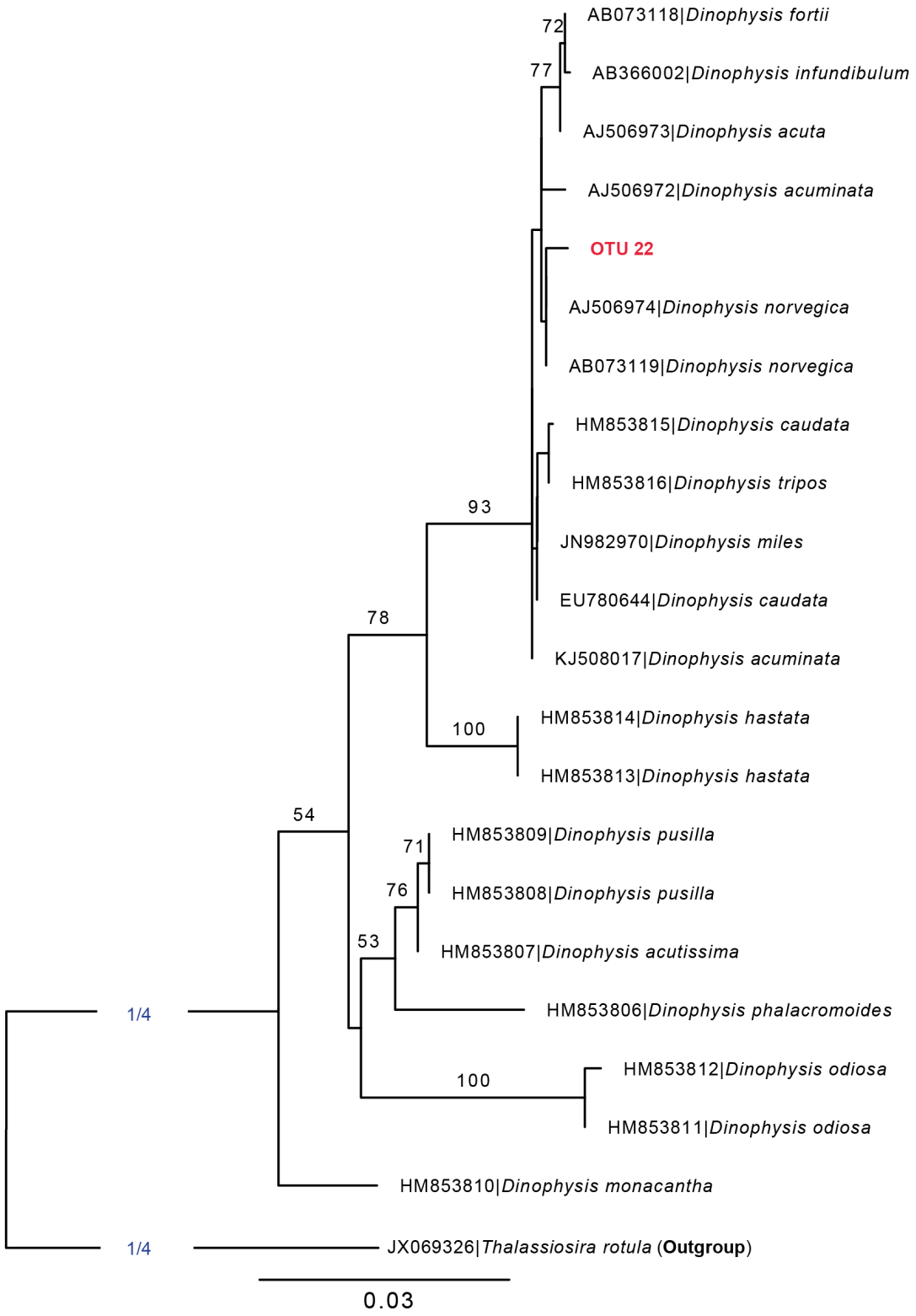


Figure 21: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from sequenced species of genus Dinophysis. The best BLAST hits on the OTU sequence in PR2 are included. The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequences retrieved in this study (OTU sequences) are highlighted in red. Scale-bar indicates a genetic distance of 0.03 base pairs in the sequence.*

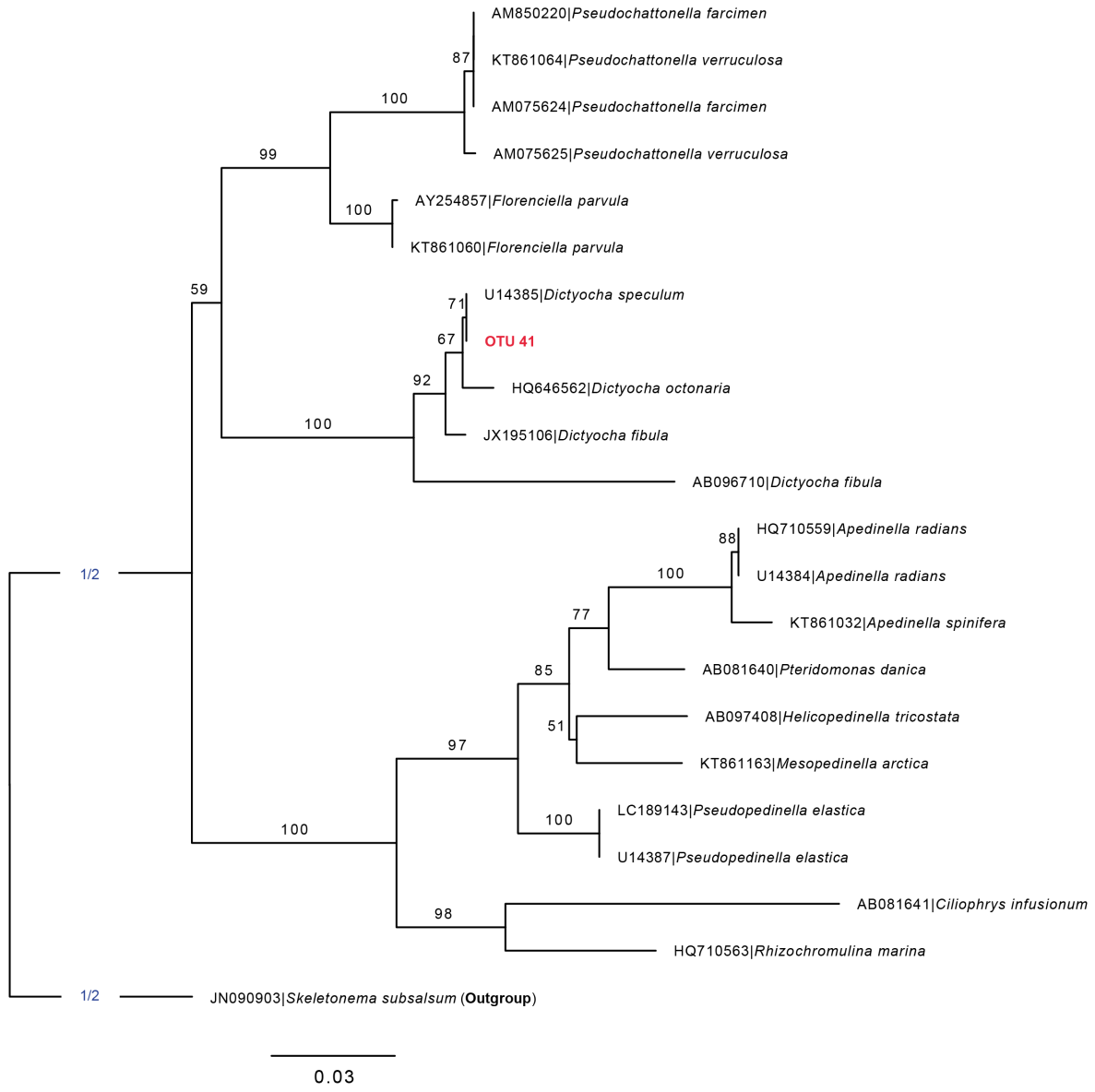


Figure 22: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from cultured species of Stramenopiles. The best BLAST hits on the OTU sequence in PR2 are included. The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ½. The sequence retrieved in this study (OTU sequences) is highlighted in red. Scale-bar indicates a genetic distance of 0.03 base pairs in the sequence.*

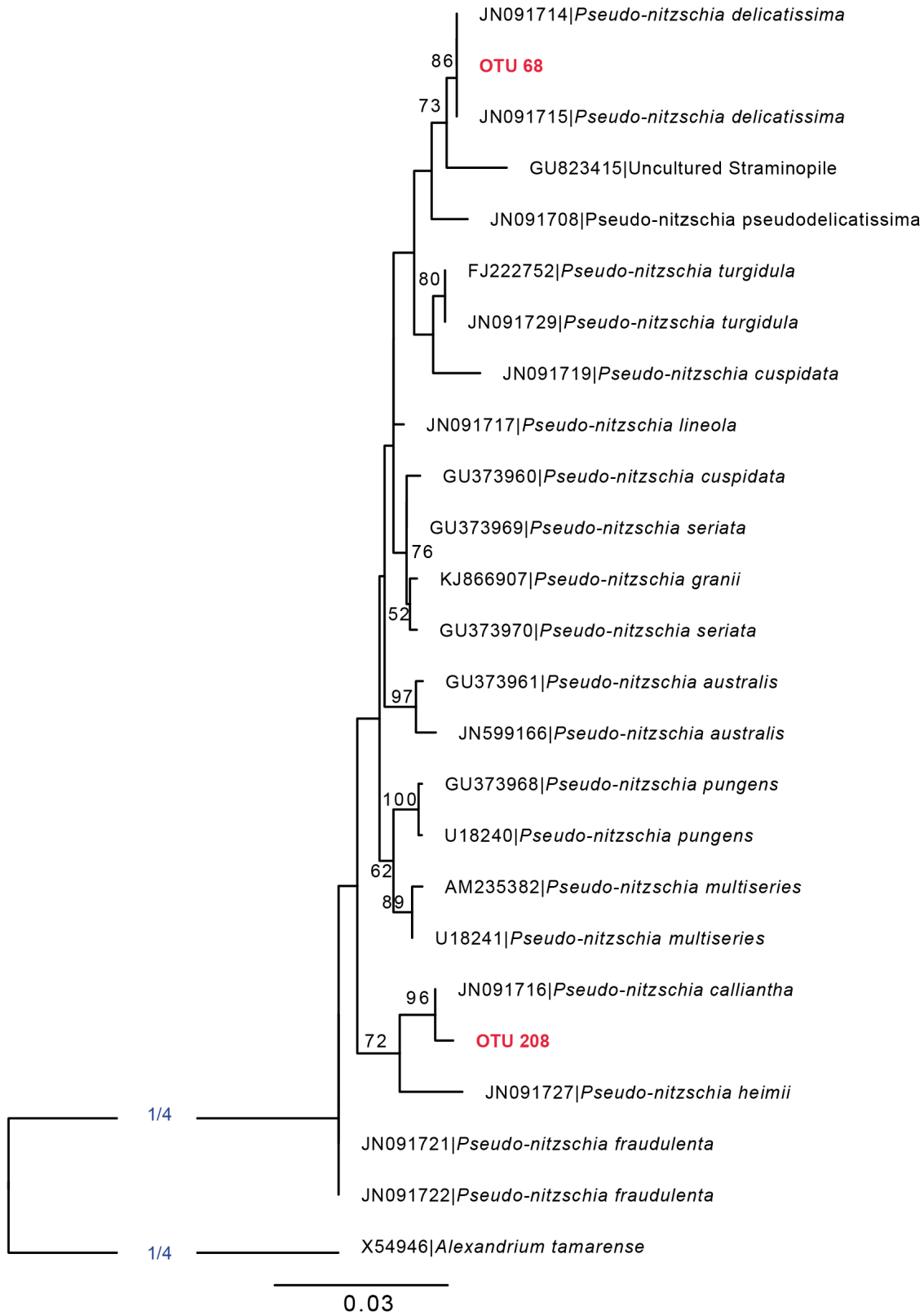


Figure 23: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from cultured species of diatoms. The best BLAST hits on the OTU sequence in PR2 are included. The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequences retrieved in this study (OTU sequences) are highlighted red. Scale-bar indicates a genetic distance of 0.03 base pairs in the sequence.*

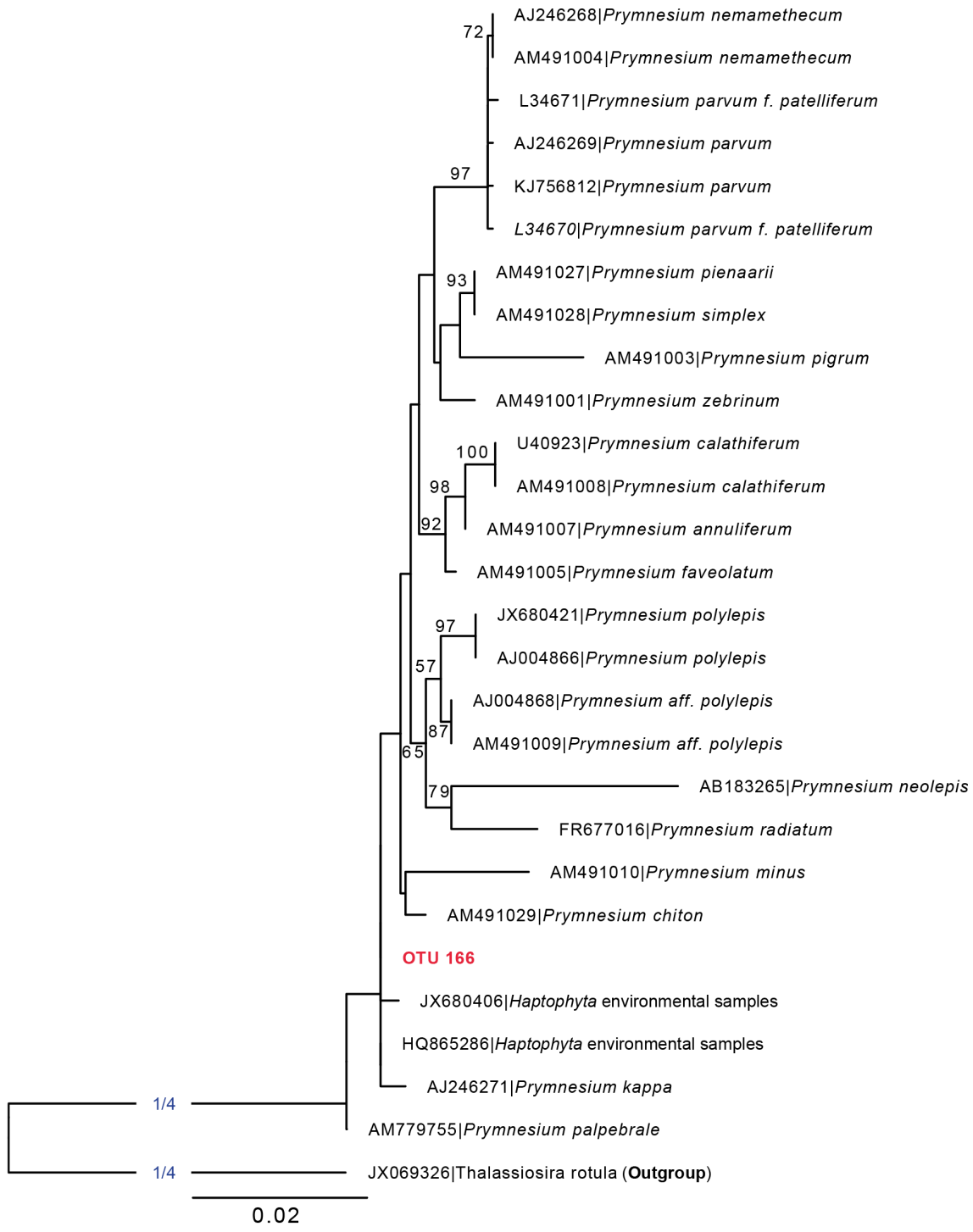


Figure 24: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from cultured species of Prymnesium. The best BLAST hits on the OTU sequence in PR2 are included, in addition to environmental sequences (Haptophyta environmental samples). The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequence retrieved in this study (OTU sequence) is highlighted in red. Scale-bar indicates a genetic distance of 0.02 base pairs in the sequence.*

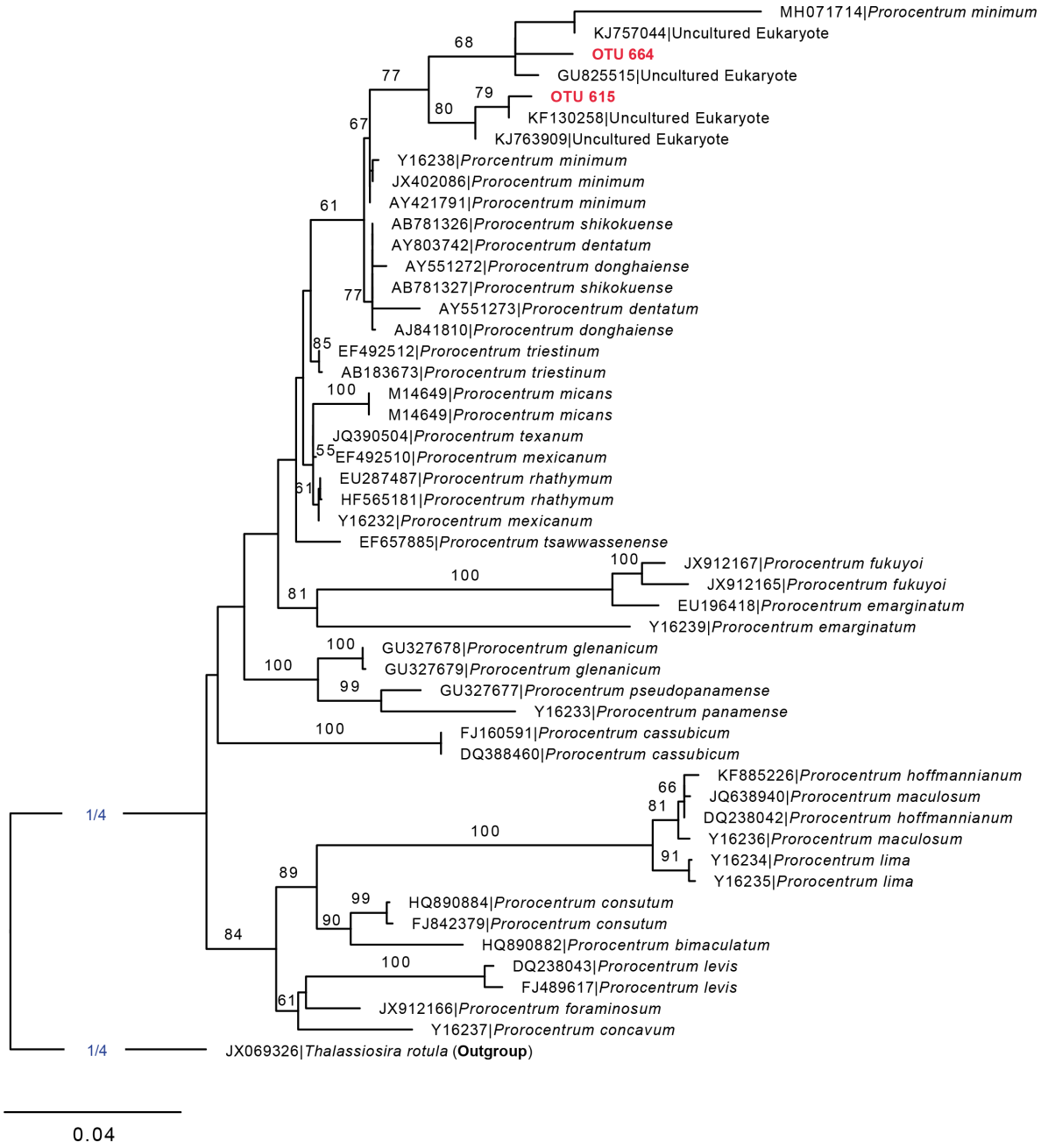


Figure 25: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from cultured species of Prorocentrum. The best BLAST hits on the OTU sequence in PR2 are included in addition to environmental sequences (uncultured eukaryotes). The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequences retrieved in this study (OTU sequences) are highlighted in red. Scale-bar indicates a genetic distance of 0.04 base pairs in the sequence.*

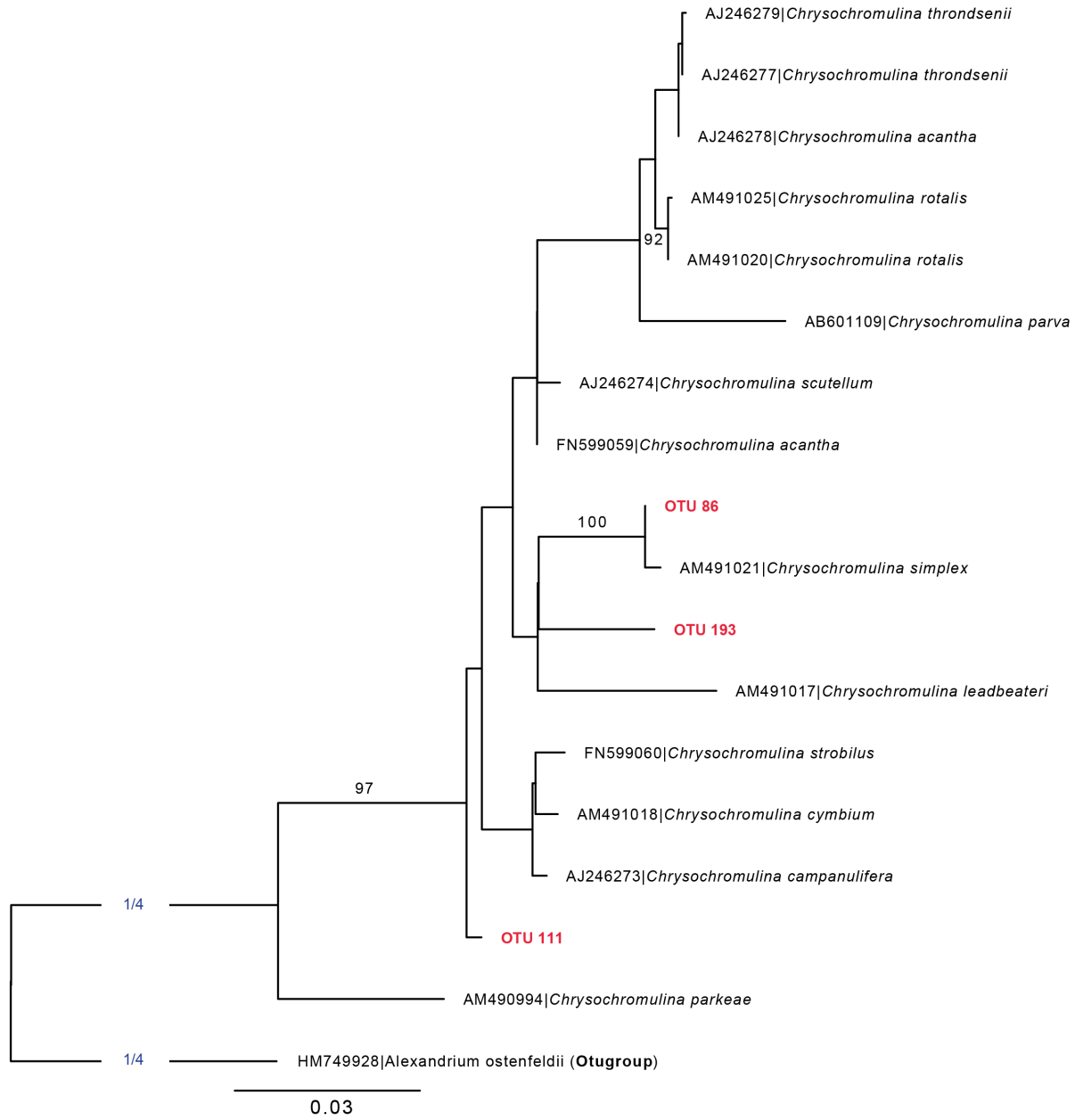


Figure 26: *Maximum likelihood (PhyMl) phylogeny of 18S rDNA sequences from cultured species of the haptophyte genus Chrysochromulina. The best BLAST hits on the OTU sequences in PR2 are included. The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequences retrieved in this study (OTU sequences) are highlighted in red. Scale-bar indicates a genetic distance of 0.03 base pairs in the sequence.*

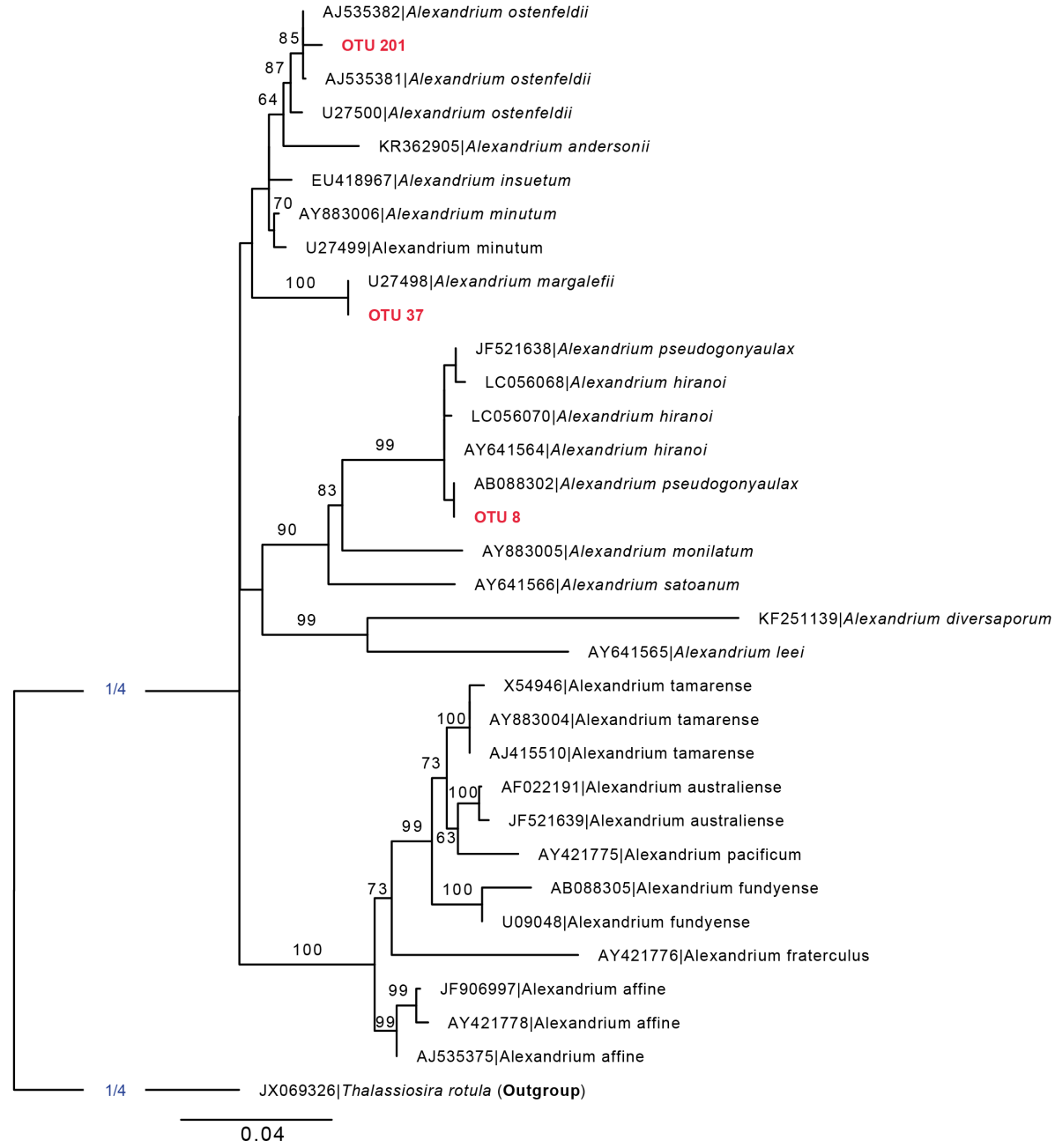


Figure 27: *Maximum likelihood (PhyML) phylogeny of 18S rDNA sequences from cultured species of Alexandrium. The best BLAST hits on the OTU sequences in PR2 are included, in addition to environmental sequences (uncultured eukaryotes). The tree was constructed with 500 Bootstraps. Only bootstrap support values above 50 are shown on the branches and the out-group-branch is cropped by ¼. The sequences retrieved in this study (OTU sequences) are highlighted in red. Scale-bar indicates a genetic distance of 0.04 base pairs in the sequence.*

3.6 Comparisons to light microscopy observations

Comparisons of the present data to light microscopy cell counts conducted by NIVA in 2017 (NIVA, 2018) are shown in Table 11. The compared data obtained by both methods are sampled from 5 m depth in 2017 at station DK1 in inner Oslofjorden. 37 species of harmful algae were detected by both methods combined. Eght of these were observed by both metabarcoding using high throughput sequencing (HTS) and counting in light microscopy (LM).

Table 11: *Table of harmful algal classes detected in the OTU table generated by the database PR2 from sequenced samples (HTS) and species detected by light microscopy (LM) cell counts. Species found by both methods are highlighted. All samples are collected at 5 m depth at station DK1, inner Oslofjorden during 2017.*

Class	Genus	Species	HTS	LM
Dinophyceae	*Prorocentrum*	*Prorocentrum* sp.	✔	
		Prorocentrum triestinum	✔	✔
		Prorocentrum cf. *balticum*		✔
		cf. *Prorocentrum balticum*		✔
		Prorocentrum micans		✔
		Prorocentrum cf. *minimum*		✔
	Ceratium	*Ceratium* sp.	✔	
	Alexandrium	*Alexandrium fundyense*	✔	
		Alexandrium hiranoi	✔	
		Alexandrium margalefii	✔	
		Alexandrium ostenfeldii	✔	
		Alexandrium pseudogonyaulax	✔	✔
	Protoceratium	*Protoceratium reticulatum*	✔	
	Dinophysis	***Dinophysis acuminata***	✔	✔
		Dinophysis acuta		✔
		Dinophysis norvegica	✔	✔
		Dinophysis rotundata		✔
		Dinophysis spp.		✔
		Dinophysis tripos		✔
		Diplopsalis (group)		✔
		Diplosalis spp.		✔
		cf. *Diplosalis bomba*		✔
	Karlodinium	cf. *Karlodinium veneficum*		✔
	Karenia	cf. *Karenia mikimotoi*		✔
Raphidophyceae	*Heterosigma*	*Heterosigma akashiwo*	✔	
Prymnesiophyceae	*Chrysochromulina*	***Chrysochromulina* sp.**	✔	✔

		Prymnesiales <5 µm		✔
		Prymnesiales, <5 µm		✔
		Prymnesiales 5-10 µm		✔
		Prymnesiales 10-15 µm		✔
	Prymnesium	*Prymnesium* sp.	✔	
Bacillariophyta	*Pseudo-nitzschia*	***Pseudo-nitzschia seriata***	✔	✔
		Pseudo-nitzschia sp.		
		Pseudo-nitzschia delicatissima	✔	✔
Dictyochophyceae	*Pseudochattonella*	*Pseudochattonella farcimen*	✔	
		Pseudochattonella spp.		✔
	Dictyocha	***Dictyocha speculum***	✔	✔

4 Discussion

This study aimed to reveal the protist community of inner Oslofjorden through metabarcoding by high throughput sequencing. The result was a dataset consisting of six supergroups of 21 protist phyla assigned to 315.900 reads of sequences. Dinoflagellates dominated the dataset, and the protist groups demonstrated seasonal variations in their composition. Several taxa of algae that are known to cause harm were detected and of these OTUs of *Alexandrium* and *Dinophysis* had the highest relative abundances. OTUs within harmful algal taxa showed significant shifts in relative abundance from one month to the other but did not differ significantly between 0-2 m- and 5 m depth. Comparisons to the traditional method of cell counts by microscopy revealed wide disparities from the present findings.

4.1 Taxonomic composition and relative abundance

The data obtained in this study revealed a broad diversity of protist taxa. Overall OTUs of 21 protist phyla were detected, represented by six supergroups. Alveolata and Stramenopila dominated the 30 samples (Figure 11). The dinoflagellates showed the highest relative abundance and was the phylum demonstrating the highest richness of OTUs. Together with the diatoms (within Ochrophyta), these two groups are significant primary producers in marine waters and constitute the most diverse taxa among all eukaryotic phytoplankton (Not et al., 2012). In access to nutrient supply from upwelling and/or runoff from land, dinoflagellates can reach high densities. This is one of the reasons why they are typically detected in coastal waters and estuaries (Not et al., 2012).

Further, dinoflagellates exhibit a wide range of ecological strategies and adaptions to different environments, which gives them the ability to form blooms when nutrients are abundant (Graham, L. E. et al., 2016). The diatoms are generally very common in Norwegian waters, typically dominating the spring bloom in Oslofjorden (Throndsen et al., 2007). The composition of dominating species observed in this study coincides with Not et al., 2012 stating that dinoflagellates and diatoms usually dominate eutrophic coastal areas and estuaries. Cell counts performed by NIVA in 2017 also found high numbers of diatoms and dinoflagellates in Vestfjorden in 2017 (NIVA, 2018), which further supports these findings.
The next largest supergroup in the dataset was the stramenopiles with an OTU richness of

26% and a total of 16% of reads in the sub-sampled dataset. These findings coincide with findings from a similar study (Gran-Stadniczeñko et al., 2018) where they found the stramenopiles to be the next largest supergroup of protists in samples from outer oslofjorden, both in OTU richness and in relative abundance. The stramenopiles are a major phylogenetic group within the eukaryotic lineage and include many ecologically important organisms (Logares et al., 2014; Massana et al., 2004). The large share of OTUs that were assigned to MAST (marine stramenopiles without a cultured representative) in the present dataset also coincides with (Gran-Stadniczeñko et al., 2018). The study found this group to be the most abundant within the heterotrophic stramenopiles in samples from outer oslofjorden. The richness of unknown marine stramenopiles further coincides with findings by (Massana et al., 2004) stating that marine stramenopiles are an abundant and poorly known group of protists.

While the minority of OTUs in the dataset was very abundant, the dataset revealed a large share of rare taxa (Table 8 and figure 12). Similar studies like (Vargas et al., 2015) have observed the same pattern of very few OTUs being highly abundant and many OTUs having a low relative abundance of reads. This agreement suggests that some few eukaryote taxa are especially crucial to the ecosystem acting as key species (Vargas et al., 2015).

In the dataset, non-protist groups originally made up 39% of all reads, divided on 196 OTUs. Metazoa, the multicellular animals, represented a significant fraction in abundance of OTUs and had 99% of reads within the Opisthokonta. The most abundant Metazoan OTU belonged to the crustacean group Maxillopoda, and manual BLAST search in the NCBI database revealed the best match to a copepod. Copepods are zooplankton, generally measuring 1-2 mm in body size. Since the collection of samples in this study was done by direct filtration on filters with a mesh size of 0.22 μm, organisms larger than this will attach to the filter. To reduce the large share of Metazoan representatives in the samples, pre-filtering of water samples on a filter of larger mesh size could have been included as a step in the sampling method. However, many diatom species such as *Chaetoceros* and *Thalassiosira* can form long chains or colonies. Likewise, the dinoflagellate species *Alexandrium catenella* and the haptophyte *Phaeocystis* possess the ability to form larger colonies (Not et al., 2012). Pre-filtering of water samples could potentially introduce the risk of missing these organisms. In the study by Gran-Stadniczeñko et al., 2018, pre-filtering on samples was performed using a 45μm nylon mesh with the purpose of removing larger size plankton from the samples.

Furthermore, another study by Massana has included pre-filtering as a step in their water sampling method (Massana et al., 2015). However, by sequencing with a high number of reads such as in this study, we still have a high number of protists reads in all samples and thus do not risk to miss out colony-forming phytoplankton species.

4.2 Seasonal dynamics of major groups

The composition and relative read abundance of the detected protist groups demonstrated marked seasonal variation at both sampling depths. The results showed monthly fluctuations in richness and abundance of the groups (Figure 15 and 16). Total OTU richness was generally highest in the winter months of November, December, and February. Dinoflagellates, Ciliates, Ocrophytta, and Heterotrophic stramenopiles dominated most of the samples regarding OTU richness and showed the highest diversity in November 2016 and December 2017. The high richness in the winter months can be explained by the mixing of the water column that takes place in Oslofjorden during this period (Baalsrud & Magnusson, 2002). Mixing of the water column due to for example absence of a strong pycnocline and unstable water masses may bring about species usually thriving at deeper water. This may have lead to the high richness of OTUs observed in the winter months. The generally low richness in March, April, May, and June in 2017 can be explained by this being the period we expect spring blooms. This is supported by detection by microscopy cell counts of a spring bloom occurring in Vestfjorden these exact months (NIVA, 2018). The spring bloom is often dominated by a few bloom-forming species of dinoflagellates and diatoms, which can explain the low richness of OTUs in this period.

The seasonal variation in percent of reads of OTUs showed almost complete dominance of dinoflagellates in the spring/summer months of May and June in 2017 at both depths. A high abundance of dinoflagellates was also observed in the same period in the monitoring performed by NIVA the same year (NIVA, 2018). The abundance of dinoflagellates was also very high in the fall months of October 2017 and September 2016, which also coincides with data the report from NIVA for 2017 showing that the density of dinoflagellates was high these months in 2017.

The results of the percent of reads were correlated to the fluorescence measurements (Figure 13), which showed high levels in May to June both summers and small peaks in October-November 2017. Fluorescence is a proxy of the phytoplankton biomass and indicates when the density of phytoplankton is high or low. The results are interpreted to show a spring bloom in May-June both years and a smaller fall-bloom in October to November. This coincides with observations published from inner Oslofjord (Fagrådet for vann- og aløpsteknisk sammarbeid i indre Oslofjord, 2016; Fagrådet for vann- og aløpsteknisk sammarbeid i indre Oslofjord, 2017). However, the early spring-bloom, typically appearing in February-March (Figure 1, introduction) does not show in our results from CTD-measurements of fluorescence. The phytoplankton spring bloom occurs typically in a short period, and the chances are high that we simply missed this exact period. In the mentioned monitoring reports from 2016 and 2017 measurements of fluorescence from surface waters in Vestfjorden show spring blooms in the middle of March 2016 and a big bloom in March-April 2017. It seems that the present data did not capture measurements from the early spring bloom in March. No measurements were gathered in March 2016, but in the beginning and end of February. However, the early spring blooms was missed that year. One cruise was performed the 6th of March 2017, but no bloom was detected in the Fluorescence data that was gathered this month.

Considering data of other hydrographical parameters, temperature demonstrated seasonal changes similar to what can be expected in Oslofjorden with low temperatures in the winter months raising gradually, reaching a top in August (Baalsrud & Magnusson, 2002). This temperature pattern was proven for both depths and reflects previously reported trends from Oslofjorden (Baalsrud & Magnusson, 2002). A sudden change in salinity in May-June 2017 indicates an input of freshwater, most likely due to snow melting. An input of water from land usually brings with it nutrients. Nutrient inputs can explain the high abundance of dinoflagellates in May-June 2017 since autotrophic dinoflagellates thrives in high nutrient water. The oxygen saturation at 5 m exceeded 100% in the spring months of April, May, and June both years. Oxygen produced by photosynthetic phytoplankton is a contributor to oxygen saturation in the ocean since they produce oxygen by photosynthesis. When oxygen saturation is high, and light is available, it indicates high biomass of photosynthetic organisms. The oxygen saturation trends (Figure 14) correlate with the measurements of fluorescence (Figure 13), also depicting high values in April, May, and June.

4.3 Temporal variation of harmful algae

Eleven taxa of harmful algal species were detected in the dataset. The most abundant of these were OTUs of *Alexandrium, Dinophysis, Pseudo-nitzschia, Dictyocha, Chrysocromulina*, and *Prorocentrum*. The Dinoflagellate genus *Alexandrium* was found by 4 OTUs in the dataset. Two of these OTUs were very abundant, accounting for more than over 60% of total reads one period in June 2016 (Figure 19). Generally, species of *Alexandrium* are frequently detected in coastal waters of Norway (Throndsen et al., 2007), which supports these findings. One OTU of *Dinohysis* was also very abundant reaching almost 10% of total reads in April 2016 and October 2017 (Figure 19). *Dinophysis* spp. frequently occurs in Norwegian waters and is found throughout the whole year (Edvardsen, B. et al., 2003).

OTU 8 of *Alexandrium*, was the one OTU of harmful algae with the highest proportional abundance with 60% of all reads in June 2016 in the sample from 5 m depth and 36% in the sample taken from 0-2 m depth. Many species of *Alexandrium* are known to produce toxins or have other harmful effects (Anderson et al., 2012). The three different toxins; saxitoxins, spirolides, and goniodomins have been linked to species of this genera (Anderson et al., 2012). Effects from ingestion of saxitoxin, by eating affected shellfish, can be PSP (paralytic shellfish poisoning) causing diarrhea, nausea and abdominal pain in humans. Effects of the toxins spirolides and goniodomins are described as severe neurological symptoms in humans (Munday et al., 2012) and fish mortality (Hsia et al., 2006) respectively.

The database PR2 classified OTU 8 to *Alexandrium hiranoi* (Appendix 11), and MAFFT-alignments revealed no differences in base pairs between the OTU 8-sequence and reference sequences of *A. hiranoi*. However, the sequence was also similar to reference sequences of *A. pseudogonyaulax* and the present phylogenetic analysis (Figure 27) placed OTU 8 in a monophyletic group with *A. hiranoi* and *A. pseudogonyaulax* with support value 99. *A. pseudogonyaulax* is known to be common along the Norwegian coast (Throndsen et al., 2007) and was observed in relatively high amounts in outer Oslofjorden in July 2016 (NIVA, 2016). On this basis, OTU 8 most likely is a sequence origination from *A. pseudogonyaulax.* This species is widespread, found at many locations around the world, which suggests it to be a cosmopolitan species (Zmerli Triki et al., 2016). It has been shown that *A. pseudogonyaulax* contains the toxin goniodomin (Zmerli Triki et al., 2016) and the high density of this species detected in June 2016 is, therefore, an important finding.

Two other abundant OTUs of *Alexandrium* were detected in the dataset. One of these, OTU 37, reached high abundances is *September* 2016 (above 10% of total reads) and was here classified to *A. margalefii*. The other, OTU 201, reaching its highest relative abundance in May 2017 (Figure 17), was classified to *A. ostenfeldii, which* is a species known to cause PSP (Salgado et al., 2015). Humans ingesting shellfish that have accumulated cells of *A. ostenfeldii* can experience neural system failure, which in some cases can cause death (Anderson et al., 2012).

OTU 22 *(Dinophysis)* was the OTU representing the next most abundant genera of potentially harmful algae. It showed high abundances in two periods during the sampling period: April 2016 and October 2017. Species of *Dinophysis* is widely distributed in the open oceans, but also in coastal areas (Raho et al., 2013) as well as in Norway (Throndsen et al., 2007). In the OTU table generated by the database PR2, this OTU was defined as *D. acuminata.* The most common *Dinophysis* species in Scandinavian waters are *D. acuta*, *D. acuminata*, *D. and norvegica* (Edvardsen, B. et al., 2003). By studying the alignment and the phylogenetic tree (Figure 21) it was not possible to classify OTU 22 further than *Dinophysis* sp. Since OTU 22 was almost entirely identical (one bp difference) to reference sequences of *D. acuta*, *D. acuminate* and *D. Norvegica,* it most probably belongs to one of these species. This is based on them being common in Scandinavian waters. Species of *Dinophysis* possess an extremely low interspecific variability within their nuclear ribosomal genes (Edvardsen, B. et al., 2003; Raho et al., 2013). This low variation makes it difficult to identify sequences originating from this genus to species, with the use of the marker from the 18S SSU. Therefore, other makers must be used to identify species of *Dinophysis* sp. The mitochondrial gene marker mtcox1 has been proposed as a better marker for taxonomic affiliation of the *Dinophysis* genera (Raho et al., 2013).

All three species of *D. acuta, D. acuminata,* and *D. Norwegica* are known to produce diarrheic shellfish toxins causing diarrhetic shellfish poisoning (DSP) (Lee et al., 1989). Poisoning can result in gastrointestinal illness in humans ingesting shellfish that have accumulated toxins from the mentioned species, even at low cell densities (Reguera et al., 2014). Due to its potential influence on human health, speeis of genus *Dinophysis* is important in the protist communities (Qiu et al., 2011).

OTUs of the class Dinophyceae were further represented by four OTUs of *Prorocentrum* sp. (Figure 17 and 18). Of these, OTU 615 was the most abundant, showing its presence in the surface and 5 m depth in all months except in March-May 2017. The results showed that frequent OTUs of *Prorocentrum* probably belong to *Prorocentrum minimum*, but could also be more related to an unknown species (Figure 25). HABs of *Prorocentrum minimum* have been recorded in Norway and monitoring of this species is essential due to its potentially harmful effects on humans through shellfish poisoning (Heil et al., 2005). However, this phylogenetic analysis failed to classify the most abundant OTU of this genera to species.

Genus *Dictyocha* was represented in all samples taken from 5 m depth (Figure 17 and 18). PR2 classified this OTU (41) no further than *Dictyocha* sp. but was found to originate from *Dictyocha speculum* in the present taxonomic affiliation (Figure 22). This classification is supported by the finding that species of *Dictyocha* can be successfully classified to species with the use of markers from the small subunit ribosomal DNA (Chang et al., 2017). Further, *Dictyocha speculum* is often found in cold and temperate areas and can form blooms in Oslofjorden (Throndsen et al., 2007). In the present data, it demonstrated peaks in November both years (2016 and 2017), which is supported by the fact that *Dictyocha speculum* is capable of forming blooms in the autumn in Oslofjorden (Throndsen et al., 2007). This species can cause ichthyotoxic HABs, i.e. cause death of fish, which was recorded from Denmark in 1983 through fish kills caused by *D. Speculum* (Henriksen et al., 1993).

Some marine pennate diatoms of the genus *Pseudo-nitzschia* can produce domoic acid (DA) causing amnesic shellfish poisoning (Miller, P. E. & Scholin, 1996). Two OTUs of this genus were present in the dataset and the phylogenetic analysis placed these within *P. delicatissima* and *P. calliantha* (Figure 23). However, there is some uncertainty concerning this phylogenetic assignment in that the best hit by BLAST in PR2 for OTU 68 was an environmental sequence of a Stramenopile. Anyhow, both species of *P. delicatissima* and *P. calliantha* have been recorded by light microscopy in 2009 and 2010 in outer oslofjorden (Hostyeva et al., 2012). In the study by Hostyeva et al., 2012 *P. delicatissima* and *P. calliantha* were the ones detected with the highest frequency throughout the sampling period, which may indicate that the present phylogenetic affiliation is correct. The neurotoxin domoic acid (DA) has caused amnesic shellfish poisoning (ASP) in human on several

occations (Bates, S. et al., 2018), but no humans have been poisoned in Norway even though DA-producing pseudo-nitzschia species have been recorded (Hostyeva et al., 2012).

The most abundant of the OTUs within genus *Prymnesium* (OTU 166) peaked in September 2017 at both depths. Some species of *Prymnesium* can form HABs that historically have caused tremendous consequences in Norway and other places (Granéli et al., 2012; Roelke et al., 2016). The most known in our waters is probably the massive bloom in Skagerak in 1988 which caused fish kills resulting in significant economic losses (Edvardsen, E. & Paasche, 1998).

4.4 The abundance of harmful algae between depths

Except from 6 quite rare OTUs, all species of genera of harmful algae were detected in samples from both 0-2 m and 5 m. Furthermore, a statistical comparison of the reads of the harmful algae OTUs showed no significant difference between the two sampling depths. Pelagic microalgae are more or less dependent on currents; therefore vertical mixing of the water column down to 5 m is probably the most plausible explanation for the homogeneous distribution of algae in the water collumn down to 5 m depth. The OTUs only appearing at one depth included *Dinophysis* (OTU 1306) and *Chrysocromulina* (OTU 1203) for the surface (0-2 m), and *Prorocentrum* (OTU 1154), *Dinophysis* (OTU 2347), *Chrysocromulina* (OTU 1486), *Ceratium* (OTU 2533) and *Alexandrium* (OTU 2483) that were only detected at 5 m. As mentioned, these OTUs were rare, meaning that they all appeared in low abundances and only in one sample each. For this reason, it is most likely that it is, to some degree, random at which depth they occurred at, rather than the species thriving at one specific depth. As earlier mentioned neither of the parameters temperature, salinity nor density varied significantly between 0-2 m depth and 5 m depth which indicates that vertical stirring of the water masses down to at least 5 m depth takes place most of the year. The homogeneity of the hydrographical measurements supports the fact that few differences in abundance of harmful algae were found between the depths. Sampling from depths bellow 5m, could possibly have revealed larger differences. In the similar study by Gran-Stadniczeñko et al., 2018 sampling was performed on surface water as well as the depth of chlorophyll maximum. Chlorophyll maximum is the depth at which the chlorophyl concentration is at its highest and the study revealed differences between the depths.

4.5 Comparison to light microscopy observations

Substantial dissimilarities between detection of harmful species in the present dataset and the species counted in light microscopy by NIVA were found. Both the two methods detected only eight species combined. With that said, many OTUs could not be sufficiently classified to species level with the use of the V4 of the 18S SSU marker. One example of this is OTUs of the genera *Dinophysis* that could not be distinguished to species by the present method. This presents a bias when comparing to microscopy cell counts. In the study by Gran-Stadniczenko et al., 2018, results from the same two approaches were compared which similarly showed marked differences.

The practice of counting cells in a microscope is often limited to larger species. Moreover, species with distinct morphological structures are easier to detect than others. Besides, the traditional way of counting using the Utermøhl-method (Utermöhl, 1931), covers only a small volume of water, which also could present a bias with this method. The present finding revealed a large share of rare species that may not be detectable by microscopy surveys. However, shortcomings in therms of species classification are still a major constraint in the present method compared to the traditional way of manually counting cells in microscopy. It is also worth mentioning that naming of algae at a species level can differ. This is because some species possess multiple taxonomic names due to alternations in classification during the molecular genetic evolution that has taken place the last decades.

Notably, the samples counted by light microscopy and those analyzed with metabarcoding in this study did not originate from the same samples, and this would, to some extent, have provided a more solid basis for comparison. However, the samples that were compared were gathered at the same sampling station, at the same depth, throughout the same year and the only difference was the sampling dates. This is believed to provide a somewhat good basis for comparison.

4.5.1 Metabarcoding for monitoring of harmful algae

In summary, the present method revealed a high diversity of OTUs, even though many were not sufficiently identified to species. Light microscopy surveys can often more easily distinguish many organisms to species level than molecular methods using a universal marker such as here. On this basis, a combination of the two methods is preferable when studying the

protist diversity and harmful algae and other studies have come to the same conclusion (Xiao et al., 2014).

Alexandrium and *Dictyocha* clades could be distinguished to species, while for example *Dinophysis, Prorocentrum, Prymnesium* could hardly be distinguished further than to genus. This shows that the marker used in this study is less suitable for the identification of species within the latter genera and more applicable for species of for example *Alexandrium* and *Dictyocha*. The present method is further limited in that it is not a quantitative method meaning that the result is in the relative abundance of sequences rather than the number of cells or the number of sequences. The results from counting cells in microscopy can be transformed to cells per. volume, which in some cases can be valuable. Another method often used for quantification of toxic algal species is qPCR. This is a quantitative method that estimates the amount of DNA that is amplified during a PCR reaction. The result from this is a quantitative estimation of how many sequences of toxic algae were detected in the samples rather than relative abundance in % of reads as in the present study.

4.6 Summary and concluding remarks

The data obtained in this study revealed a broad diversity of protistan taxa and dinoflagellates and diatoms dominated the general composition of protist groups. The results of the overall composition of protist taxa coincide with earlier findings by both metabarcoding (Gran-Stadniczeñko et al., 2018) and microscopic surveys (NIVA, 2018). A large share of rare OTUs was detected, indicating that some few species of protists are especially important to the ecosystem. The detected protist groups varied in composition and relative abundance throughout the year and a spring bloom dominated by dinoflagellates was detected, followed by a smaller bloom in October-November. The results matched the observations of hydrographical data that indicated variations similar to what can be expected in inner oslofjorden.

Multiple taxa of harmful algae were detected in this study. The genera that was dominating the dataset was *Alexandrium, Dinophysis, Pseudo-nitzschia, Dictyocha, Chrysocromulina*, and *Prorocentrum.* The most abundant species was the toxin-producing *Alexandrium pseudogonyaulax* appearing in alarming abundances in June 2016. High abundances of an OTU of *Dinophysis* was further detected and was found to belong to eighter of the toxic

species *D. acuta, D. acuminata,* and *D. Norwegica.* No significant difference in abundance of the OTUs of harmful algae was detected in the dataset and mixing of the upper water layer is supposedly the reason for this.

The present approach, metabarcoding, can reveal an overall detailed picture of the broad diversity of marine protists. However, identification to species is narrow due to limitations in the databases. Universal markers as used in this study, are well suited to identify a broad diversity of protist, but, when studying composition at the species level, choice of marker must be more carefully picked out for suitability for the various genera. This is proven here by for example species of *Alexandrium* being fairly well defined to species whereas, for example, *Dinophysis* is hard to classify with the use of the V4 region of the 18S. Metabarcoding with the use of HTS of universal primers can provide a good snapshot of the community of protists in a given sample. However, sturdy classification to species level is probably still best done with identification by microscopy. A combination of the two methods is preferable in analyzation of marine communities and especially when studying the community of harmful algae or other more narrow groups of species, where it can be crucial telling the organisms to species. For further studies, it is suggested to analyze the same samples by using both approaches: microscopy and metabarcoding.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215 (3) October, pp. 403–410.

Andersen, N. G., Hansen, P. J., Engell-Sørensen, K., Nørremark, L. H., Andersen, P., Lorenzen, E. & Lorenzen, N. (2015) Ichthyotoxicity of the Microalga Pseudochattonella Farcimen under Laboratory and Field Conditions in Danish Waters. *Diseases of Aquatic Organisms*, 116 (3) October, pp. 165–172.

Anderson, D. M., Alpermann, T. J., Cembella, A. D., Collos, Y., Masseret, E. & Montresor, M. (2012) The Globally Distributed Genus Alexandrium: Multifaceted Roles in Marine Ecosystems and Impacts on Human Health. *Harmful Algae*, 14 February, pp. 10–35.

Baalsrud, K. & Magnusson, J. (2002) *Indre Oslofjord, Natur Og Miljø*. Fagrådet for vann- og avløpsteknisk samarbeid i indre Oslofjord.

Bates, S., A. Hubbard, K., Lundholm, N., Montresor, M. & Pin Leaw, C. (2018) Pseudo-Nitzschia, Nitzschia, and Domoic Acid: New Research since 2011. *Harmful Algae*, October.

Berdalet, E., Fleming, L. E., Gowen, R., Davidson, K., Hess, P., Backer, L. C., Moore, S. K., Hoagland, P. & Enevoldsen, H. (2016) Marine Harmful Algal Blooms, Human Health and Wellbeing: Challenges and Opportunities in the 21st Century. *Journal of the Marine Biological Association of the United Kingdom*, 96 (1) February, pp. 61–91.

Caron, D. A., Countway, P. D., Jones, A. C., Kim, D. Y. & Schnetzer, A. (2012) Marine Protistan Diversity. *Annual Review of Marine Science*, 4 (1), pp. 467–493.

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 (18) September, pp. 5797–5808.

Chang, F. H., Sutherland, J. & Bradford-Grieve, J. (2017) Taxonomic Revision of Dictyochales (Dictyochophyceae) Based on Morphological, Ultrastructural, Biochemical and Molecular Data. *Phycological Research*, 65 (3) July, pp. 235–247.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N. de, Pfrender, M. E. & Bernatchez, L. (2017) Environmental DNA Metabarcoding: Transforming How We Survey Animal and Plant Communities. *Molecular Ecology*, 26 (21) November, pp. 5872–5895.

Denny, M. (2008) *How the Ocean Works: An Introduction to Oceanography*. Princeton: Princeton University Press.

Dittami, S., Hostyeva, V., Egge, E., Kegel, J., 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 (10), pp. 6719–6732.

Dittami, S. M., Pazos, Y., Laspra, M. & Medlin, L. K. (2013) Microarray Testing for the Presence of Toxic Algae Monitoring Programme in Galicia (NW Spain). *Environmental Science and Pollution Research International*, 20 (10) October, pp. 6778–6793.

Duarte, C. M. & Cebrián, J. (1996) The Fate of Marine Autotrophic Production. *Limnology and Oceanography*, 41 (8) December, pp. 1758–1766.

Ebenezer, V., Medlin, L. K. & Ki, J.-S. (2012) Molecular Detection, Quantification, and Diversity Evaluation of Microalgae. *Marine Biotechnology (New York, N.Y.)*, 14 (2) April, pp. 129–142.

Edgar, R. C. (2013) UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads. *Nature Methods*, 10 (10) October, pp. 996–998.

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

Edvardsen, B. (2017) Lecture Presentation in the Course BIO4381. Presented at: *2017, University of Oslo*.

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 RDNA1. *Journal of Phycology*, 39 (2) April, pp. 395–408.

Edvardsen, E. & Paasche, E. (1998) Bloom Dynamics and Physiology of Prymnesium and Chrysochromulina. In: *Physiological Ecology of Harmful Algal Blooms*. Springer, Berlin, pp. 193–208.

Fagrådet for vann- og aløpsteknisk sammarbeid i indre Oslofjord (2016) *Årsberettning 2016*.

Fagrådet for vann- og aløpsteknisk sammarbeid i indre Oslofjord (2017) *Årsberettning 2017*.

Farabegoli, F., Blanco, L., Rodríguez, L. P., Vieites, J. M. & Cabado, A. G. (2018) Phycotoxins in Marine Shellfish: Origin, Occurrence and Effects on Humans. *Marine Drugs*, 16 (6) June, p. 188.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281 (5374) July, pp. 237–240.

Gérikas Ribeiro, C., Lopes dos Santos, A., Marie, D., Pereira Brandini, F. & Vaulot, D.

(2018) Small Eukaryotic Phytoplankton Communities in Tropical Waters off Brazil Are Dominated by Symbioses between Haptophyta and Nitrogen-Fixing Cyanobacteria. *The ISME Journal*, 12 (5) May, pp. 1360–1374.

Graham, L. E., Graham, J. M., Wilcox, L. W. & Cook, M. E. (2016) *Algae*. 3rd ed. LJLM Press, LLC.

Granéli, E., Edvardsen, B., Roelke, D. L. & Hagström, J. A. (2012) The Ecophysiology and Bloom Dynamics of Prymnesium Spp. *Harmful Algae*, 14 February, pp. 260–270.

Gran-Stadniczeñko, S., 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* [Online], 0 (0). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jeu.12700> [Accessed 7 December 2018].

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., Vargas, C. de, Decelle, J., Campo, J. del, 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 Research*, 41 (Database issue) January, pp. D597–D604.

Heil, C. A., Glibert, P. M. & Fan, C. (2005) Prorocentrum Minimum (Pavillard) Schiller: A Review of a Harmful Algal Bloom Species of Growing Worldwide Importance. *Harmful Algae*, 4 (3) March, pp. 449–470.

Henriksen, P., Knipschildt, F., Moestrup, Ø. & Thomsen, H. A. (1993) Autecology, Life History and Toxicology of the Silicoflagellate Dictyocha Speculum (Silicoflagellata, Dictyochophyceae). *Phycologia*, 32 (1) January, pp. 29–39.

Henry, L., Wickham, H. & RStudio (2018) *Rlang: Functions for Base Types and Core R and 'Tidyverse' Features* [Online]. Available from: <https://CRAN.R-project.org/package=rlang> [Accessed 20 November 2018].

Hostyeva, V., Eikrem, W., Edvardsen, B. & Hasle, G. R. (2012) Annual Cycle of Pseudo-Nitzschia Species in Outer Oslofjorden, Norway. pp. 81–83.

Hsia, M. H., Morton, S. L., Smith, L. L., Beauchesne, K. R., Huncik, K. M. & Moeller, P. D. R. (2006) Production of Goniodomin A by the Planktonic, Chain-Forming Dinoflagellate Alexandrium Monilatum (Howell) Balech Isolated from the Gulf Coast of the United States. *Harmful Algae*, 5 (3) April, pp. 290–299.

Illumina, 2018. Illumina | Sequencing and array-based solutions for genetic research [WWW Document]. URL https://www.illumina.com/ (accessed 12.21.18).

Katoh, K. & Standley, D. M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and*

Evolution, 30 (4) April, pp. 772–780.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. (2013) Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, 79 (17) September, pp. 5112–5120.

Lee, J.-S., Igarashi, T., Fraga, S., Dahl, E., Hovgaard, P. & Yasumoto, T. (1989) Determination of Diarrhetic Shellfish Toxins in Various Dinoflagellate Species. *Journal of Applied Phycology*, 1 (2) August, pp. 147–152.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W. H. C. F., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M. C., Romac, S., Shalchian-Tabrizi, K., Simon, N., Stoeck, T., Santini, S., Siano, R., Wincker, P., Zingone, A., Richards, T. A., Vargas, C. de & Massana, R. (2014) Patterns of Rare and Abundant Marine Microbial Eukaryotes. *Current biology: CB*, 24 (8) April, pp. 813–821.

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 (6) June, pp. 3528–3534.

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. C. F., Logares, R., Mahé, 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. & Vargas, C. de (2015) Marine Protist Diversity in European Coastal Waters and Sediments as Revealed by High-Throughput Sequencing: Protist Diversity in European Coastal Areas. *Environmental Microbiology*, 17 (10) October, pp. 4035–4049.

Miller, P. E. & Scholin, C. A. (1996) IDENTIFICATION OF CULTURED PSEUDO-NITZSCHIA (BACILLARIOPHYCEAE) USING SPECIES-SPECIFIC LSU RRNA-TARGETED FLUORESCENT PROBES1. *Journal of Phycology*, 32 (4) August, pp. 646–655.

Munday, R., Quilliam, M. A., LeBlanc, P., Lewis, N., Gallant, P., Sperker, S. A., Ewart, H. S. & MacKinnon, S. L. (2012) Investigations into the Toxicology of Spirolides, a Group of Marine Phycotoxins. *Toxins*, 4 (1), pp. 1–14.

Naustvoll, L.-J. & Dahl, E. (2002) *Kompendium i Planktonalger*. Forskningsstasjonen Flødevigen: Havforskningsinstituttet.

Nickrent, D. L. & Sargent, M. L. (1991) An Overview of the Secondary Structure of the V4 Region of Eukaryotic Small-Subunit Ribosomal RNA. *Nucleic Acids Research*, 19 (2) January, pp. 227–235.

Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. (2013) BayesHammer: Bayesian Clustering for Error Correction in Single-Cell Sequencing. *BMC Genomics*, 14 (1) January, p. S7.

NIVA (2016) *Overvåking Av Ytre Oslofjord 2014-2018*. Årsrapport 7169–2017.

NIVA (2018) *Årsovervåking Med FerryBox - Indre Oslofjord 2017*. 7266–2018.

Not, F., Siano, R., Kooistra, W. H. C. F., Simon, N., Vaulot, D. & Probert, I. (2012) Diversity and Ecology of Eukaryotic Marine Phytoplankton [Online]. In: *Advances in Botanical Research*. vol. 64. Elsevier, pp. 1–53. Available from: <http://linkinghub.elsevier.com/retrieve/pii/B9780123914996000013> [Accessed 15 November 2018].

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. & Wagner, H. (2018) *Vegan: Community Ecology Package. R Package Version 2.5-3* [Online]. Available from: <https://cran.r-project.org/web/packages/vegan/vegan.pdf> [Accessed 29 October 2018].

Paasche, E. (2005) Undervisningskompendium. FORELESNINGER I MARIN BIOLOGI BOTANISK DEL. Universitetet i Oslo.

Pawlowski, J. (2014) Protist Evolution and Phylogeny [Online]. In: *eLS*. American Cancer Society. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0001935.pub2> [Accessed 5 December 2018].

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 Microbiology Ecology*, 93 (1) January, p. fiw200.

Qiu, D., Huang, L., Liu, S. & Lin, S. (2011) Nuclear, Mitochondrial and Plastid Gene Phylogenies of Dinophysis Miles (Dinophyceae): Evidence of Variable Types of Chloroplasts. *PLOS ONE*, 6 (12) des, p. e29398.

Raho, N., Rodriguez, F., Reguera, B. & Marin, I. (2013) Are the Mitochondrial Cox1 and Cob Genes Suitable Markers for Species of Dinophysis Ehrenberg? *Harmful Algae*, 28 August, pp. 64–70.

Ramiro Logares (2017) *Ramalok/Amplicon_processing: Workflow for Analysing MiSeq Amplicons Based on Uparse* [Online]. Zenodo. Available from: <https://zenodo.org/record/259579#.W1beFdUzbIU> [Accessed 24 July 2018].

Reguera, B., Riobó, P., Rodríguez, F., Díaz, P. A., Pizarro, G., Paz, B., Franco, J. M. & Blanco, J. (2014) Dinophysis Toxins: Causative Organisms, Distribution and Fate in Shellfish. *Marine Drugs*, 12 (1) January, pp. 394–461.

Roelke, D. L., Barkoh, A., Brooks, B. W., Grover, J. P., Hambright, K. D., LaClaire, J. W., Moeller, P. D. R. & Patino, R. (2016) A Chronicle of a Killer Alga in the West: Ecology, Assessment, and Management of Prymnesium Parvum Blooms. *Hydrobiologia*, 764 (1) January, pp. 29–50.

Salgado, P., Riobó, P., Rodríguez, F., Franco, J. M. & Bravo, I. (2015) Differences in the

Toxin Profiles of Alexandrium Ostenfeldii (Dinophyceae) Strains Isolated from Different Geographic Origins: Evidence of Paralytic Toxin, Spirolide, and Gymnodimine. *Toxicon: Official Journal of the International Society on Toxinology*, 103 September, pp. 85–98.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T. & Quince, C. (2015) Insight into Biases and Sequencing Errors for Amplicon Sequencing with the Illumina MiSeq Platform. *Nucleic Acids Research*, 43 (6) March, pp. e37–e37.

Suga, H., Chen, Z., Mendoza, A. de, Sebé-Pedrós, A., Brown, M. W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., Torruella, G., Derelle, R., Manning, G., Lang, B. F., Russ, C., Haas, B. J., Roger, A. J., Nusbaum, C. & Ruiz-Trillo, I. (2013) The Capsaspora Genome Reveals a Complex Unicellular Prehistory of Animals. *Nature Communications*, 4 August, p. 2325.

Tennekes, M. & Ellis, P. (2017) *Treemap: Treemap Visualization* [Online]. Available from: <https://CRAN.R-project.org/package=treemap> [Accessed 20 November 2018].

Thaulow, H. & Faafeng, B. (2013) *Indre Oslofjord 2013 – status, trusler og tiltak* [Online]. Norsk institutt for vannforskning. Available from: <https://brage.bibsys.no/xmlui/handle/11250/194149> [Accessed 21 November 2018].

Throndsen, J., Hasle, G. R. & Tangen, K. (2007) *Phytoplankton of Norwegian Costal Waters*. Almater forlag AS.

Trainer, V. L., Bates, S. S., Lundholm, N., Thessen, A. E., Cochlan, W. P., Adams, N. G. & Trick, C. G. (2012) Pseudo-Nitzschia Physiological Ecology, Phylogeny, Toxicity, Monitoring and Impacts on Ecosystem Health. *Harmful Algae*, 14 February, pp. 271–300.

Utermöhl, H. (1931) Neue Wege in Der Quantitativen Erfassung Des Plankton.(Mit Besonderer Berücksichtigung Des Ultraplanktons.). *SIL Proceedings, 1922-2010*, 5 (2) January, pp. 567–596.

Vargas, C. de, Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N. L., 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., Coordinators, T. O., 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 (6237) May, p. 1261605.

Wickham, H. (2018) *Reshape: Flexibly Reshape Data* [Online]. Available from: <https://CRAN.R-project.org/package=reshape> [Accessed 20 November 2018].

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K. & RStudio (2018) *Ggplot2: Create Elegant Data Visualisations Using the Grammar of*

Graphics [Online]. Available from: <https://CRAN.R-project.org/package=ggplot2> [Accessed 20 November 2018].

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 (8) August, p. e106510.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zimmermann, J., Jahn, R. & Gemeinholzer, B. (2011) Barcoding Diatoms: Evaluation of the V4 Subregion on the 18S RRNA Gene, Including New Primers and Protocols. *Organisms Diversity & Evolution*, 11 (3) July, p. 173.

Zinger, L., Gobet, A. & Pommier, T. (2012) Two Decades of Describing the Unseen Majority of Aquatic Microbial Diversity. *Molecular Ecology*, 21 (8) April, pp. 1878–1896.

Zingone, A. & Enevoldsen, H. (2000) Zingone A, Enevoldsen HO.. The Diversity of Harmful Algal Blooms: A Challenge for Science and Management. Ocean Coast Manage 43: 725-748. *Ocean & Coastal Management*, 43 January, pp. 725–748.

Zingone, A. & Wyatt, T. (2004) Harmful Algal Blooms: Keys to the Understanding of Phytoplankton Ecology. *The Sea*, 13 January, pp. 867–926.

Zmerli Triki, H., Laabir, M., Moeller, P., Chomérat, N. & Kéfi Daly-Yahia, O. (2016) First Report of Goniodomin A Production by the Dinoflagellate Alexandrium Pseudogonyaulax Developing in Southern Mediterranean (Bizerte Lagoon, Tunisia). *Toxicon*, 111 March, pp. 91–99.

Appendix

Appendix 1: Procedure for DNA extraction from Sterivex filters using Qiagen Sterivex Kit

This is a modified protocol for DNA extraction from Sterivex filters using the Qiagen DNeasy Power Water Sterivex Kit. The protocol does not require Tube Extenders, VacConnector, VacValve and the Manifold (Steps 18-26) of the Quick-Start Protocol provided in the Kit.

Before starting:

1. Read the Qick-Start Protocol (steps 1-18, 27-29) and safety protocols at the UiO
2. Keep all the solutions at room temperature. After mixing the ST1B solution, keep it at 2-8°C
3. Clean the working surfaces with 20% deconex solution
4. Set the two incubators in the post-pcr lab at 65°C and 90°C
5. Defrost the filters on ice. Make sure that there is no seawater remaining in the filters. If there is seawater remaining, connect a large syringe with fully pulled-back plunger to the inlet and pour the remaining seawater in a separate tube through the outlet. Save the tube in the freezer.
6. The protocol is conducted at a working desk under room temperature except when stated otherwise

Protocol:

1. Warm the MBL and MR solutions in the incubator at 65°C for 5-10 min
2. Pour the ST1A liquid into the ST1B powder and mix
3. Close the both ends of the filter by placing the inlet and outlet cups. Use a gum for the outlet if the cup is not fitting
4. Add 900 µl of ST1B to the filter through the inlet opening. Place the pipette past the neck of the inlet and above the white membrane. Tilt the filter to avoid the accumulation of water in the inlet
5. Close the inlet opening and place the filter on the vortex with the inlet facing out. Vortex for 5 min, at minimum speed (500). Rotate the filter while attached so that the inlet is facing in and vortex again for 5 min. at minimum speed

6. Open the inlet and add 900 µl of warm (65°C) MBL solution. Place the pipette past the neck of the inlet and above the white membrane. Tilt the filter to avoid the accumulation of water in the inlet
7. Close the inlet, wrap the filter in the aluminium foil and incubate at 90°C for 5 min. (no warmer and no longer)
8. Cool the filter for 2 min. at room temperature and then vortex (inlet facing out) for 5 min, at maximum speed (the speed will be ~2100). Make sure that both caps are firmly on as the gum may fall off during vortex
9. Set the plunger of the 3 ml syringe at 1ml mark and connect the syringe to the inlet (remove the cap). Hold the filter and the syringe with outlet facing up and pull the plunger to extract the lysate from the syringe. Transfer the lysate to the Power Bead Tube (blue cap) and repeat until most of the lysate is collected in the tube
10. Vortex the Power Bead Tube (cap facing out) for 5 min. at maximum speed (final speed~2300)
11. Centrifuge the Power Bead Tube 1 min/ 4000 g (use the large centrifuge in the chemistry lab)
12. Collect the supernatant from the Power Bead Tube in a 2.2 ml collection tube using a micropipette (1000 micrliter). It is important to push the pipette to the bottom of the tube and release slowly to fill the pipette with the liquid. Note that the small amount of beads does not affect the extraction but larger amount of beads can clog the pipette
13. Add 300 µl of IRS solution, vortex briefly and then incubate at 4°C for 5 min
14. Centrifuge for 1 min. / 13,000 rpm. Transfer the supernatant in 5 ml collection tube. Avoid the beads
15. Add 3 ml of MR (warmed to 65°C) and vortex the tube for few seconds
16. Add 750 µl of lysate to a MB Spin Column attached to the with appropriate collection tube. Use two Spin Columns per sample. Centrifuge for 1 min. / 8,000 rpm, discard the filtered liquid and repeat the process using the same Spin Column. It takes three(or four) rounds of centrifugation with two MB Spin Columns to filter all of the lysate from the 5 ml collection tube. The DNA is now bonded to the column and the liquid in steps 17-19 can be discarded
17. Add 750 µl of ethanol to the Spin Column and centrifuge for 1 min. / 8,000 rpm. Discard the filtered liquid

18. Add 750 µl of PW1 to the Spin Column and centrifuge for 1 min. / 8,000 rpm. Discard the filtered liquid
19. Add 750 µl of ethanol to the Spin Column and centrifuge for 3 min. / 14,000 rpm. Discard the filtered liquid and change the collection tube. Leave to dry at room temperature and make sure there is no leftover ethanol in the collection tube
20. Add 50 µl of EB solution to the centre of the Spin Column. Centrifuge for 1 min. / 10,000 rpm. Do not discard the liquid after this step since it contains the DNA washed from the column. Collect the liquid and add it to the same Spin Column one more time. Centrifuge for 1 min. / 10,000 rpm
21. Pool the contents of the two Spin Columns that were used for the same sample
22. Label the tubes and store them in the freezer.

Appendix 2:

TAE x 50 Stock solution:

1.Mix:
242 g Tris-acetat
57,1 mL glacial acetic acid
100 mL 0,5 M EDTA (pH 8,0)

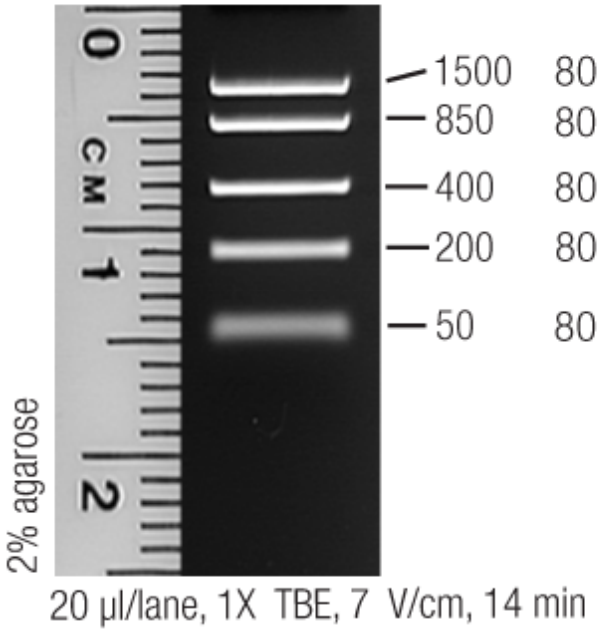
2.Add dH_2O for total volume of 1L

3.Autoclave

TAE x1 buffer:

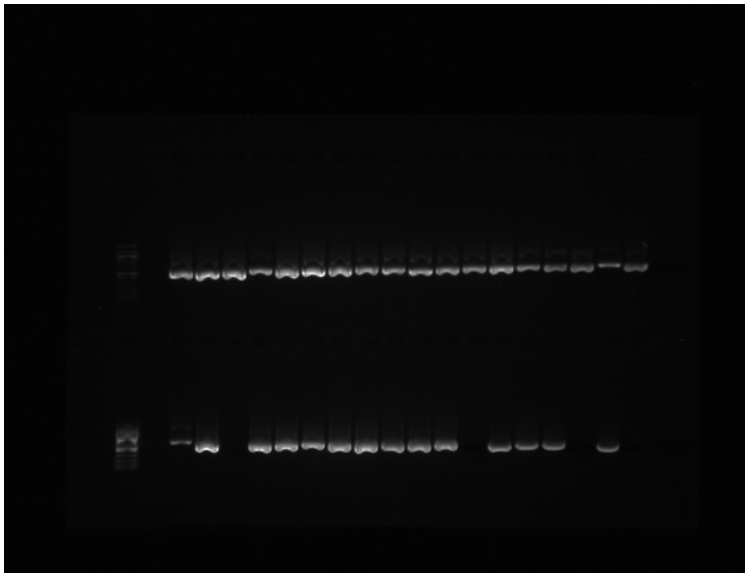
Mix:
20 mL TAE x50 Stock solution
980 mL dH_2O

Appendix 3: Amount of DNA (ng) in each band of FastRuler™ Low Range DNA Ladder ready-to-use (Thermo Fisher Scientific, MA, USA)

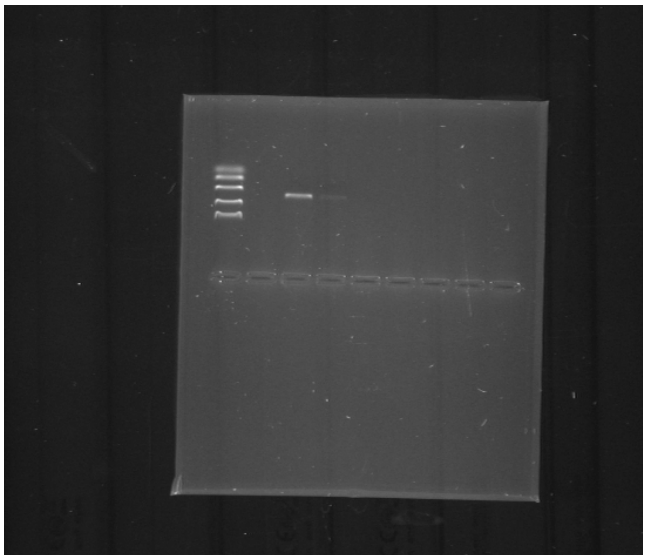


bp	ng/20µl	ng/15µl	ng/10µl	ng/5µl	ng/3µl
1500	80	60	40	20	12
850	80	60	40	20	12
400	80	60	40	20	12
200	80	60	40	20	12
50	80	60	40	20	12

Appendix 4: Picture of electrophoreses gel with bands of all the samples



Appendix 5: Picture of electrophoresis gel with band of the pooled sample



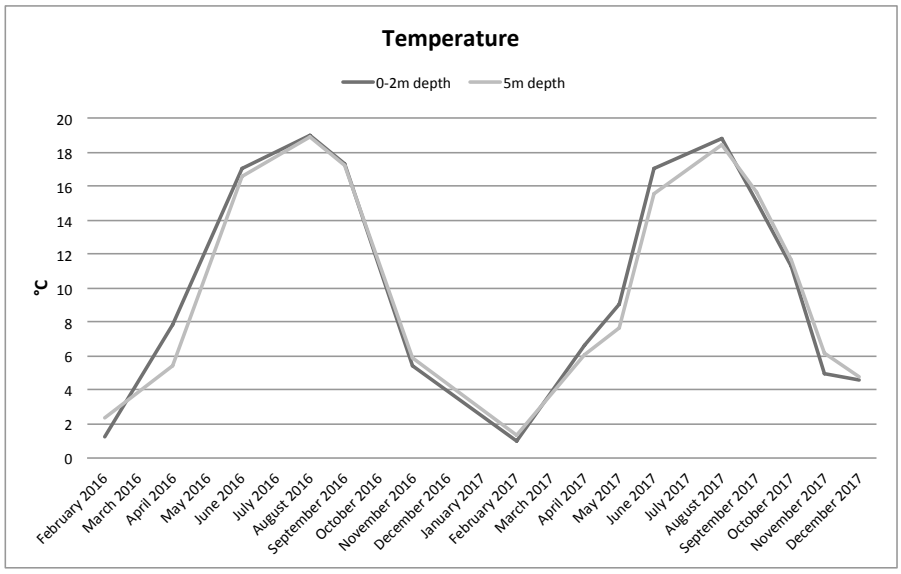
Appendix 6: R-script for sub-sampling of data

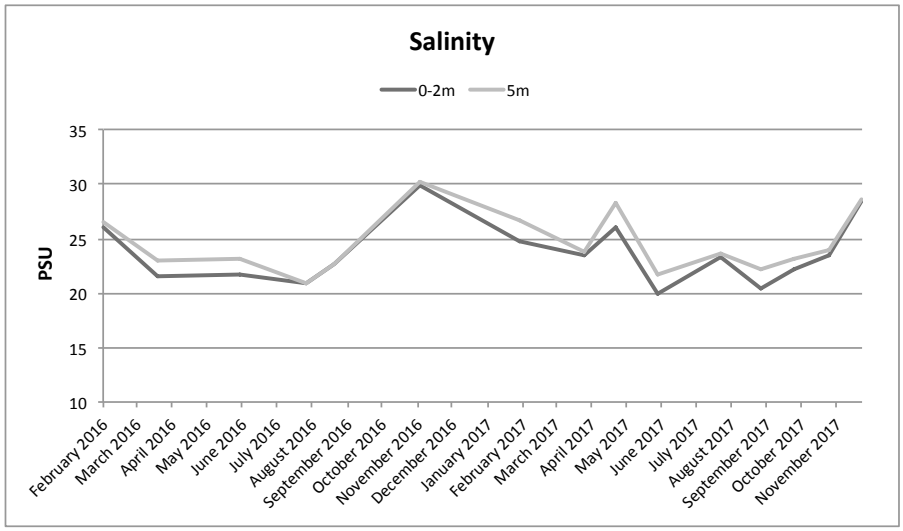
```
###SUB-SAMPLING###

library(vegan)
subsample<-read.table("Subsample_input.txt", header=T, row.names = 1)
subsample

colSums(subsample)
min(colSums(subsample)) #The sample with fewest reads

sort(colSums(subsample))
subsample$X61_S61
sort (subsample$X61_S61)

otu_rrarefy_subsample<-rrarefy(t(subsample), sample=41960)
otu_rrarefy_subsample2<-t(otu_rrarefy_subsample)

otu_rrarefy_subsample2

#write out the subsampled dataset to txt file
write.csv(otu_rrarefy_subsample2,"Subsample18.juni.csv")
```

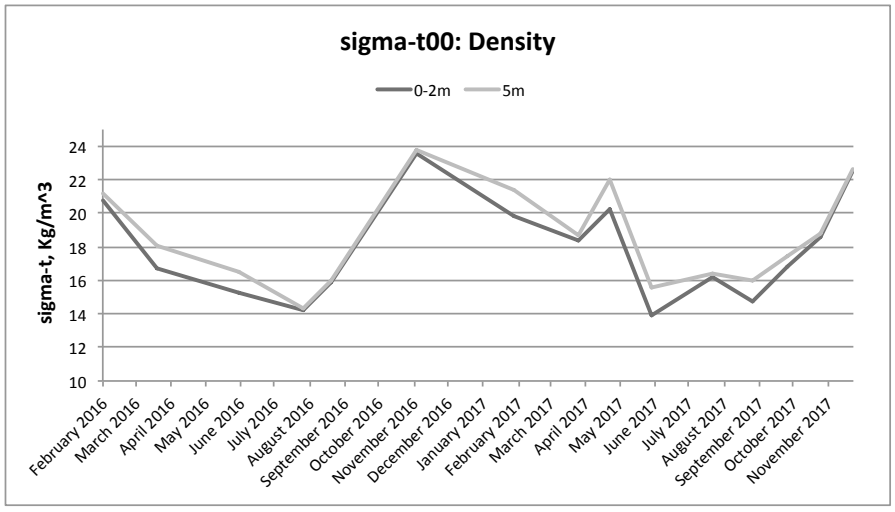
Appendix 7: Total number of OTUs and reads in all three versions of dataset. The percent of OTUs and percent of reads for each taxa in the subsampled dataset are included.

		Original dataset		**Protist dataset**		**Sub-sampled dataset**			
Supergroup	Phylum/Group	OTUs	Reads	OTUs	Reads	OTUs	% of OTUs	Reads	% of reads
Alveolata	Apicomplexa	13	2325	13	2325	12	0,913	677	0,214
	Ciliophora	124	132536	124	132536	117	8,897	12171	3,853
	Dinoflagellata	567	2000859	567	2000859	534	40,608	203100	64,292
	Perkinsea	3	64	3	64	3	0,228	9	0,003
Amoebozoa	Amoebozoa	2	30	2	30	1	0,076	3	0,001
Apusozoa	Apusozoa	4	64	4	64	4	0,304	9	0,003
Archeplastida	Chlorophyta	70	140025	70	140025	67	5,095	12698	4,02
	Rhodophyta	22	50124	22	50124	22	1,673	4496	1,423
	Streptophyta	7	462	-	-	-	0,304	-	0,063
Eukaryota	Unc Eukaryota	2	1064	2	1064	2	0,76	96	4,008
Hacrobia	Centroheliozoa	5	1407	5	1407	4	2,129	198	1,448
	Cryptophyta	10	137956	10	137956	10	0,38	12660	0,515
	Haptophyta	28	46074	28	46074	28	0,913	4573	0,903
	Katablepharidophyta	6	16034	6	16034	5	1,217	1628	0,623
	Picozoa	12	30812	12	30812	12	2,129	2852	0,318
	Telonemia	16	13771	16	13771	16	0,532	1967	0,126
Opisthokonta	Choanoflagellida	28	11652	28	11652	28	7,148	1004	1,988
	Fungi	57	7479	-	-	-	0,76	-	0,056
	Mesomycetozoa	7	4757	7	4757	7	11,255	398	3,981
	Metazoa	132	1952516	-	-	-	13,84	-	11,902
	Other Opisthokonta	1	115	1	115	-	0,684	-	0,231
Rhizaria	Cercozoa	97	49363	97	49363	94	0,152	6281	0,03
	Radiolaria	12	2621	12	2621	10	0,913	176	0,214
Stramenopiles	Het Straminopiles	160	107880	160	107880	148	8,897	12576	3,853
	Other Straminopiles	9	4900	9	4900	9	40,608	729	64,292
	Ochrophyta	190	333956	190	333956	182	0,228	37599	0,003
Sum		**1584**	**5048846**	**1388**	**3088389**	**1315**		**315900**	

Appendix 8: CTD measurements from 0-2 m -and 5 depth







Appendix 9: Number of reads for each OTU of harmul algal taxa, in each sample from 0-2 m

		2016						2017										
	Date	24/02	11/04	22/06	18/08	12/09	25/11	02/02	20/02	06/03	18/04	15/05	21/05	14/08	18/09	18/10	16/11	15/12
Taxa	**OTUs**																	
Prorocentrum	**615**	28	38	4	NA	4	19	NA	NA	NA	0	0	0	1	2	5	9	54
	664	0	0	0	NA	0	1	NA	NA	NA	0	0	0	0	0	0	0	5
	1154	0	0	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
	1082	0	0	0	NA	1	0	NA	NA	NA	0	0	0	0	0	0	0	0
Ceratium	**2533**	0	0	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
Alexandrium	**8**	0	1	3790	NA	86	21	NA	NA	NA	0	0	82	0	0	14	2	0
	2483	0	0	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
	37	0	0	66	NA	1184	35	NA	NA	NA	0	0	0	0	0	0	0	0
	201	7	0	9	NA	30	0	NA	NA	NA	31	68	0	4	0	3	57	2
Protoceratium	**268**	3	2	11	NA	2	1	NA	NA	NA	0	0	12	1	0	0	0	0
Heterosigma	**987**	0	0	0	NA	0	1	NA	NA	NA	0	0	0	0	0	0	0	0
Dinophysis	**22**	4	932	92	NA	38	13	NA	NA	NA	26	1	61	0	16	547	149	3
	2347	0	0	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
	1306	0	2	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
Chrysochromulina	**925**	0	0	0	NA	0	0	NA	NA	NA	0	0	1	1	0	0	0	0
	111	31	61	6	NA	3	0	NA	NA	NA	4	0	5	2	37	0	0	14
	86	32	77	0	NA	0	3	NA	NA	NA	0	0	2	2	86	0	1	27
	193	59	9	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	6
	1203	0	1	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
	1486	0	0	0	NA	0	0	NA	NA	NA	0	0	0	0	0	0	0	0
Prymnesium	**166**	8	31	1	NA	0	2	NA	NA	NA	5	0	7	0	61	0	0	15
Prymnesium	**1585**	0	0	0	NA	0	0	NA	NA	NA	0	0	0	2	0	0	0	0
Pseudo-nitzschia	**68**	0	1	1	NA	0	9	NA	NA	NA	258	10	3	2	0	1	6	4
	208	0	0	0	NA	0	24	NA	NA	NA	0	0	0	0	0	11	44	3
Pseudochattonella	**709**	0	2	1	NA	0	1	NA	NA	NA	0	0	0	0	7	0	0	1
Dictyocha	**41**	2	5	27	NA	0	127	NA	NA	NA	1	0	1	0	5	1	126	67

Appendix 10: Number of reads for each OTU of harmul algal taxa, in each sample from 5 m depth

		2016						2017										
	Date	24/02	11/04	22/06	18/08	12/09	25/11	02/02	20/02	06/03	18/04	15/05	21/06	14/08	18/09	18/10	16/11	15/12
Taxa	**OTUs**																	
Prorocentrum	**615**	110	47	2	12	4	42	9	1	0	0	0	4	8	16	6	10	45
	664	1	0	0	5	0	6	1	0	0	0	0	0	0	1	0	1	4
	1154	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
	1082	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Ceratium	**2533**	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alexandrium	**8**	1	1	6922	55	21	0	0	0	0	0	4	57	1	0	12	0	4
	2483	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	37	0	0	23	712	497	0	0	1	0	0	0	0	38	1	0	0	0
	201	1	0	13	12	16	0	0	2	0	127	26	0	16	0	3	0	0
Protoceratium	**268**	1	4	2	0	0	0	0	0	0	0	0	3	2	0	0	0	0
Heterosigma	**987**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Dinophysis	**22**	13	375	19	13	20	5	33	9	2	148	34	191	27	44	993	9	5
	2347	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
	1306	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chrysochromulina	**925**	0	0	0	0	0	1	0	0	0	0	0	0	4	0	0	0	0
	111	40	117	2	3	15	1	13	24	0	0	1	4	22	16	0	2	16
	86	37	66	1	3	2	6	5	4	0	0	0	38	10	35	0	1	20
	193	42	6	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
	1203	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1486	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Prymnesium	**166**	13	20	1	2	0	5	16	10	0	4	0	34	8	133	0	0	10
Prymnesium	**1585**	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Pseudo-nitzschia	**68**	4	0	1	0	0	40	25	22	10	72	298	3	1	1	5	54	12
	208	0	0	0	1	0	32	0	0	0	0	0	0	0	0	12	23	0
Pseudochattonella	**709**	0	0	0	0	0	1	0	2	0	0	0	0	1	9	0	0	1
Dictyocha	**41**	63	3	17	10	2	113	110	47	1	2	2	2	47	71	1	130	41

Appendix 11: Taxonomic affiliation of OTUs of harmful algal taxa generated by the database PR2

OTU	Class	*Speceis identification by database PR2*
615	Dinophyceae	*Prorocentrum sp.*
664	Dinophyceae	*Prorocentrum sp.*
1154	Dinophyceae	*Prorocentrum sp.*
1082	Dinophyceae	*Prorocentrum triestinum*
2533	Dinophyceae	*Ceratium sp.*
8	Dinophyceae	*Alexandrium hiranoi*
2483	Dinophyceae	*Alexandrium hiranoi*
37	Dinophyceae	*Alexandrium margalefii*
201	Dinophyceae	*Alexandrium ostenfeldii*
268	Dinophyceae	*Protoceratium reticulatum*
987	Raphidophyceae	*Heterosigma akashiwo*
22	Dinophyceae	*Dinophysis acuminata*
2347	Dinophyceae	*Dinophysis acuminata*
1306	Dinophyceae	*Dinophysis acuminata*
925	Prymnesiophyceae	*Chrysochromulina leadbeateri*
111	Prymnesiophyceae	*Chrysochromulina sp.*
86	Prymnesiophyceae	*Chrysochromulina sp.*
193	Prymnesiophyceae	*Chrysochromulina sp.*
1203	Prymnesiophyceae	*Chrysochromulina sp.*
1486	Prymnesiophyceae	*Chrysochromulina sp.*
166	Prymnesiophyceae	*Prymnesium sp.*
1585	Prymnesiophyceae	*Prymnesium sp.*
68	Bacillariophyta	*Pseudo-nitzschia seriata*
208	Bacillariophyta	*Pseudo-nitzschia sp.*
709	Dictyochophyceae	*Pseudochattonella farcimen*
41	Dictyochophyceae	*Dictyocha sp.*