UiA University of Agder

Master

Identifying protists along the Agder coastline using Illumina sequencing

MAIKEN TORKELSEN
M.SC COASTAL ECOLOGY

SUPERVISOR
Tove M. Gabrielsen

University of Agder, 2022
Faculty of Engineering and Science
Department of Natural Sciences

Everybody is a genius,
But if you judge a fish by its ability to climb a tree,
it will live its whole life believing that it is stupid
Albert Einstein

Universitetet i Agder
Fakultet for Teknologi og Realfag
Institutt for Naturvitenskapelige fag
Gimlemoen 25
4604 Kristiansand
http://www.uia.no

Abstract

Environmental quality, sustainability and functioning of marine ecosystems are closely linked to species richness and diversity. Microalgae play key roles in coastal ecosystems contributing significantly to carbon flux through the microbial loop and are the main suppliers of photosynthetic products that higher trophic levels of the marine food web depend upon. The Norwegian coastal water is a mixture of freshwater run-off from rivers, outflow of brackish water from the Baltic Sea through the Kattegat, and North Sea coastal water. This produces a low-saline coastal water which mixes in the north with Atlantic water and this forms the Norwegian coastal current (NCC).

In this study, we determined the different groups and species of protists present in spring, as well as comparing species composition in inner and outer coastal areas of Agder. We also evaluated the potential changes in protists community composition along two riverine influenced transects. Water samples were collected at four different depth layers, (sea surface, 5m, 15m and deep) in the beginning of April, DNA was extracted and Illumina sequences on the 18S rRNA gene were obtained. Our results also showed a large diversity of Dinoflagellata in every location, as well as Ochrophyta and Picozoa. In the more sheltered inner locations, there was signs for anoxic bottom water and low sequence number, while the more outer locations with colder water, more salinity and oxygen had a higher sequence count. The diversity measures showed a community with richness on each location with only a few exceptions. This methodology is useful tool for timeseries and could be a quick and cost-saving method for further research.

Norsk oppsummering

Mikroalger spiller en viktig rolle i kystøkosystemet og bidrar blant annet til i karbonfluks gjennom det mikrobielle næringsnettet. Gjennom sin fotosyntese er de hovedleverandørene av fotosyntetiske produkter som høyere trofiskenivåer i det marine næringsnettet er avhengig av. Det norske kystvannet er en blanding av ferskvannsavrenning fra elver, utløp av brakkvann fra Østersjøen gjennom Kattegat, og kystvann fra Nordsjøen. Dette gir et kystvann med lavt saltinnhold som blander seg i nord med atlantisk vann og dette danner den norske kyststrømmen (NCC). Grensen mellom Skagerrak og Nordsjøen går ved Lindesnes.

I denne studien har vi identifisert de ulike gruppene og artene av pelagiske protister som er til stede langs kysten av Agder våren 2021, og sammenlignet artssammensetningen i indre og ytre kystområder av Agder. Vi evaluerte også de mulige endringene i protistenes artssammensetning langs kysten av Agder. Vannprøver ble samlet i fire forskjellige dybdelag (havoverflate, 5m, 15m og dyp), disse ble sekvensert ved bruk av Illumina-sekvensering av V9 regionen av 18S rRNA-genet. Sammensetningen av protistsamfunnene viste til resultater av et stort sekvensnummer av Dinoflagellata på hvert sted, så vel som Ochrophyta og Picozoa. På de mer lune indre stedene var det tegn til anoksisk vann og også lavt sekvenstall, mens de mer ytre stedene med kaldere vann, mer saltholdighet og oksygen hadde høyere sekvenstall. Diversitet målene viste høyt artsmangfold på alle lokasjonene med få unntak. Denne metodikken er et nyttig verktøy for tidsserier, og kan være en rask og kostnadsbesparende metode for fremtidig forskning.

Table of contents

Preface

I would like to thank my supervisor professor Tove M. Gabrielsen for her guidance and supervision during my years at University in Agder. I am very grateful for your availability, even in super busy times in the office, you set some minutes and hours free to provide help and support, day or night. Supportive during hard personal times and a good preacher for motivation. Truly grateful for the world of algae you have introduced to me, it has been a journey, and I hope I can dive deeper into in it the future.

I would also like to say thank you to Tove´s team of PhD candidates, Post docs and fellow master students for all support. Especially Henriette G. Horn for all help with my plots and script in R, as well as collecting my samples with Tove when covid-19 hit us. Also want to say thank you to Cheshtaa Chitkara for her help with dada2 R script and all answers regards that.

To Stian Borg-Stoveland for pick me-ups, laughter and moral support, you have been a good friend during these years, and a friend I hope to have for many years to come. Thank you to Bjørg Karin Varnes for being a big support and friend during my years at UiA, for all assignments together, and coffee breaks. We did it!

Thank you to my amazing partner, Simen. I would not have made the finish line without you by my side to encourage and support me, letting me cry when it all got a bit too much, helping me out with the dogs, or just simply buying chocolate to me.
Lastly, I want to say thank you to all my fellow ADHD sisters out there. You have made me feel less alone. Made me realize that even though my brain might work different than others, does not mean I am not capable.

Kristiansand, Friday 20th of May
Maiken Torkelsen

1 Background

The oceans are the largest habitat on our planet and marine coastal areas are among the most productive ecosystems in the world (Bar-On et al., 2018). Environmental quality, sustainability and functioning of marine ecosystems are closely linked to the species richness and diversity (Bar-On et al., 2018). Ecosystems require a balance of energy to function. Energy in a food web flows from producers to consumers to decomposers. Consumers and decomposers are heterotrophs, eating other organisms to obtain energy (Kaiser et al., 2011). Decomposers consume organic material from dead plants and animals, break them down chemically into simpler molecules, and return the molecules to the environment (Kaiser et al., 2011). Plants and other producers such as algae use these molecules, which include carbon, nitrogen and minerals (Kaiser et al., 2011). Producers like algae form the basis of energy in a food web (Marquardt, 2016). Algae use light energy from the sun to convert inorganic carbon into sugars through the process of photosynthesis (Kaiser et al., 2011).

1.2 Marine microbes

During the past two decades there has been an increased understanding of how marine microbes influence the structure and function of the oceans (Heidelberg et al., 2010). Microbial eukaryotes are involved in several nutrient and energy acquisition mechanisms, as well as playing important roles in ocean food webs and biogeochemical cycling (Caron et al., 2009). Together with Cyanobacteria they are responsible for almost half of the global primary production in the ocean (Field et al., 1998).

Photosynthetic microbial eukaryote (phytoplankton) communities are characterized by a continuous change in their taxonomic composition and abundance (Egge et al., 2015). Protists include their heterotrophic relatives and represent organisms from the paraphyletic group Protista, comprising protozoa, algae, and lower-fungi, and including more than 200 000 described species (Pawlowski, 2014). Protists can be broken down into three groups; fungal-like protists that take in food through absorption; algae, which are plant-like protists that produce organic carbon through photosynthesis; and protozoa, which are animal-like protists that ingest their food. Protists play key roles in coastal ecosystems contributing significantly to carbon flux through the microbial loop and are the main suppliers of photosynthetic products that higher trophic levels of the

marine food web depend upon (Gran-Stadniczenko et al., 2019). Simon et al. (2019) described the protist community biomass on continental shelves to consist of predominantly diatoms, dinoflagellates, and haptophytes.

The community of bacteria, archaea, protist, fungi and viruses are responsible for half of the global biogeochemical flux of carbon, nitrogen, phosphor, sulphur and iron (Fuhrman et al., 2015). Microalgae play key roles in coastal ecosystems contributing significantly to carbon flux through the microbial loop and are the main suppliers of photosynthetic products that higher trophic levels of the marine food web depend upon. Protists are unicellular and multicellular algae and protozoans with a wide range of ecological functions (Gran-Stadniczenko et al., 2019). Protists are morphologically and genetically diverse and are present in all types of marine habitats. Protists can be heterotroph, phototroph and mixotroph, thus, giving a larger adaptability to adapt to environment. Heterotrophic protists feed on their surroundings e.g., bacteria, decaying organic matter and other protists. Mixotrophs are protists that can use both heterotrophy and phototrophy to gain energy and get nutrients, therefore, they are very prevalent across all areas (Leles et al., 2017; Nygaard & Tobiesen, 1993). Gran-Stadniczeñko et al., (2019) discussed the compositions of plankton dynamics in the Skagerrak and concluded that the dominated supergroups were Alveolates, Stramenopiles and Hacrobia (Fig 1.).

1.2.1 Alveolata

Alveolata is a supergroup that contains a wide range of heterotrophic, phototrophic and mixotrophic species. Dinoflagellata is one of three major phyla of Alveolata. Dinoflagellata contain mostly motile cells with two flagella and can be either thecate with cellulose plates or athecate (naked). In the marine system, Alveolata play important roles as grazers and parasites, as well as phototrophic primary producers (Levinsen & Nielsen, 2002). Dinophyceae is a class under the division of Dinoflagellata, and often found throughout the year (Marquardt et al., 2016). Ciliophora is another phylum of Alveolata and occur in the marine food web as both prey and grazers (Levinsen & Nielsen, 2002; Seuthe et al., 2011). Syndiniales, also called MALV (marine alveolate) group I and II, is an exclusive order composed of marine parasites (Guillou et

al., 2008). Syndiniales are known to be very opportunistic, and can infect many marine hosts from different trophic levels in the marine food webs (Guillou et al., 2008).

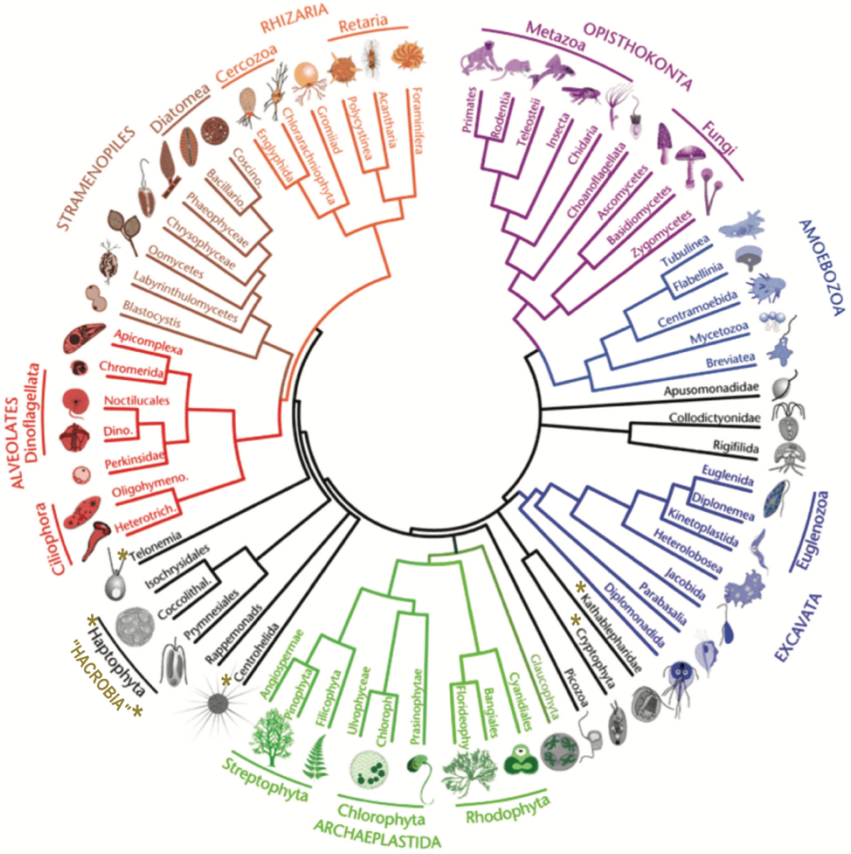


Fig 1. Phylogenomic tree of eukaryotes. Presentation of the seven supergroups where also protists are placed. Image from Gran Stadniczeñko (2019).

1.2.2 Stramenopiles

Stramenopiles are known for their two differently shaped flagellate. This supergroup includes brown algae (Phaeophyceae/Ochrophyta) and diatoms (Bacillariophyceae). Diatoms are abundant and ubiquitous in marine environments (Pawlowski et al., 2016) and dominate the spring bloom of artic waters (Hodal et al., 2012). Marine Stramenopiles (MASTs) is a very common and phylogenetically diverse group. MASTs are picoplankton grazers and MAST-1, -3, -4, -7, -12 are known MASTs to appear in Skagerrak (Logares et al., 2012). Region. MAST-7 and -12 have been associated with deep sea sediment and anoxic water (Logares et al., 2012)

1.2.3 Hacrobia

Haptophyta and Cryptophyta form the newest supergroup Hacrobia (Okamoto et al., 2009). Hacrobia also include Picozoa, which are hypothesized to be important in sedimentation processes (Seenivasan et al., 2013). Hacrobia is often identified with having two unequal flagella, and include mainly plastid-bearing phototrophs as well as heterotrophic and mixotrophic species (Marquardt, 2016). Haptophyta is known to be of ecological importance as they can form blooms that can be toxic and because they are important for the biogeochemical cycles in the ocean.

1.2.4 Archaeplastida

Archaeplastida are photosynthetic primary producers which originated from cyanobacteria (Burki et al., 2020). Archaeplastida is a supergroup that consist of red (Rhodophyta, mostly macroscopic) and green algae(Chlorophytes), land plants (Streptophyta) and also the unicellular group of Glaucophytes (Marquardt, 2016). Green algae contain marine micro- and macroalgae but dominate more in freshwater environments.

1.3 Agder Coast

The Norwegian coastal water is a mixture of freshwater run-off from rivers, outflow of brackish water from the Baltic Sea through the Kattegat, and North Sea coastal water (Albretsen et al., 2011). This produces a low-saline coastal water which mixes in the north with Atlantic water and this forms the Norwegian coastal current (NCC) (Albretsen et al., 2011) (Figure 2). Skagerrak is a part of the transition area between the Baltic Sea and the North Sea (Gustafsson & Stigebrandt, 1996). The Skagerrak is around 200 by 100 km and with an average depth of 210 km. The NCC on the Norwegian west coast is mainly influenced by Atlantic water, while the Skagerrak water is mixed water between Atlantic water, central North Sea water and coastal North Sea water from the Danish side and a blend with brackish water on the Norwegian side. The Norwegian shelf is important for several commercially fish species for spawning and hatching ground (Sætre et al., 2003). Skagerrak, off the coast of Norway undergoes strong seasonal environmental variations due to changes in meteorological and hydrological conditions and irradiance. The balance of hydrological forces from Baltic currents, saline north Atlantic currents and land runoff lead to considerable salinity and temperature flucations and seasonal water column stratification (Egge et al., 2015).

Skagerrak is location on the North Sea shelf, and there controlled by the strong flow created by the divergence between inflowing high-saline water from the North Atlantic and the North sea, and the outflow low-saline water from the Baltic sea (Skov & Durinck, 1998). The border between Skagerrak and the North Sea is at Lindesnes. Studies in Skagerrak (Oslofjorden) have revealed the dynamics and distribution of organisms belonging to different trophic and taxonomic groups such as dinoflagellates, diatoms, haptophytes, cryptophytes, prasinophytes, dictyochophytes and euglenoids (Egge et al., 2015; Gran-Stadniczenko et al., 2019).

Along the Norwegian waters, over 700 diverse species of phytoplatonic groups have been found; 178 species of dinoflagellates, 177 species of diatoms, and 291 species from other algal/protist groups (euglenoids, ciliates and haptophytes) (Kuylenstierna, 2006).

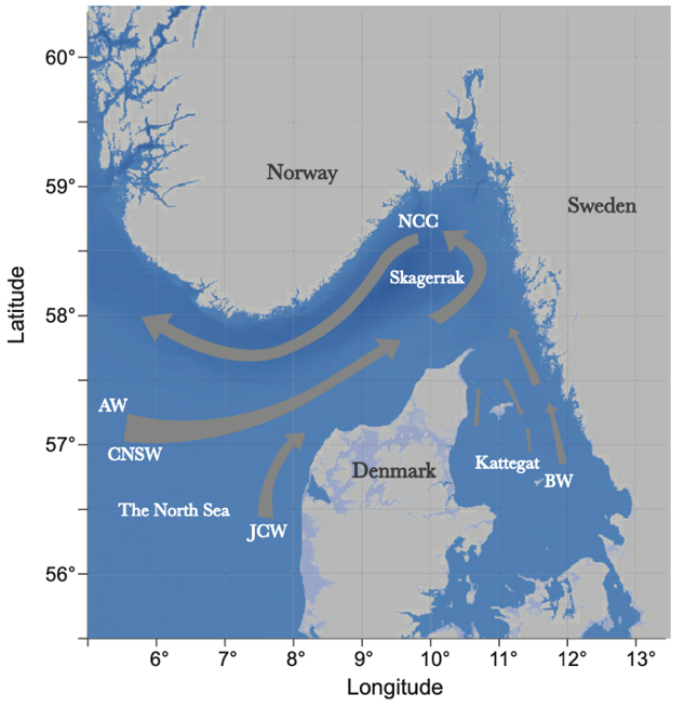


Fig 2: Current circulation pattern in Skagerrak and Kattegat on the coast of Norway with the Norwegian coastal current (NCC), Atlantic water (AW), Central North Sea water (CNSW), and Baltic water (BW). The Agder coast consist of Skagerrak on the east and the North Sea on the west, with the border being at Lindesnes. Modified map from Gran-Stadniczeñko (2019)

1.6 18S rDNA gene metabarcoding

High-resolution biodiversity assessments of environmental samples have been made possible due to next generation sequencing technology (Hadziavdic et al., 2014). Through high throughput sequencing (HTS) it is possible to have massive parallel sequencing of clonally amplified single DNA molecule or DNA templates (Margulies et al., 2005). Thus, giving of over hundreds of reads in a single run, which are assembled into genotypes by using computer tools (Margulies et al., 2005). The standard approach for exploring microbial diversity in the environment is by using high-throughput sequencing of phylogenetic markers of environmental samples, also called metabarcoding (Decelle et al., 2014). Using metabarcoding enables the identification of multiple species from a DNA samples by comparing the DNA sequences to databases of known species (Liu et al., 2020). While Sanger sequencing was long the gold standard due to its high accuracy, HTS allow for large set of parallel DNA sequencing, therefore, providing extremely high throughput from several samples simultaneously and for a reduced cost (Young & Gillung, 2019). Metabarcoding is a cost-effective and time-effective approach for large-scale studies, therefore also one of the most used techniques in environmental identification studies where conventional morphology-based species identification is logistically or financially impractical (Liu et al., 2020). DNA metabarcoding uses standardized DNA regions as a tag for quick and accurate species identification (Valentini et al., 2009). Several studies over the years have shown that DNA barcoding is effective both cost and time wise (Hebert et al., 2003; Saunders, 2005; Ward et al., 2005).

In Illumina sequencing molecules of DNA are hybridized to oligonucleotides already attached to polymer-coated glass surface of a flow cell. In the flow cell, amplification is performed by flowing enzymes and reagents through the flow cell and attaching to the DNA. Once amplification is done, molecules form clusters of amplicons that are derived from a single template molecule (Quail et al., 2009). The choice of primers can impact the results from the biodiversity assessments, and using primers that targets all prokaryotes and eukaryotes it will limit the depth of the assessment (Hadziavdic et al., 2014). Thus, by limiting the universality of the primer one may also exclude important groups in the analysis and thus potentially introduce biases (Hadziavdic et al., 2014). To be able to obtain an accurate taxonomic profiling of protist communities, the choice of

primers is crucial (Vaulot et al., 2022). The goal is to amplify the target community with minimal biases with the use of both forward and reverse primers. Therefore, the amplified region must be long enough to differentiate between closely related taxa by including enough variable positions. For eukaryotes the small subunit ribosomal RNA gene (SSU) is commonly used. The 18S rRNA gene contains variable regions which are used to assign taxonomy, in eukaryotes the variable regions which are the most targeted is the V4 and the V9 region (Decelle et al., 2014; Vaulot et al., 2022). The V9 region was previously due to the limitation in sequence size (restricted to 2 x 75bp). Thus, with the development of illumina Miseq the V4 region (2 x 300 bp) is now favored due to being longer, more variable, and better covered in reference databases (Pawlowski et al., 2012).

1.7 Location

The location for the project is outside of Agder, in the Skagerrak and North Sea. The Skagerrak, off the coast of Norway have a strong seasonal environmental variation due to changes in meteorological and hydrological conditions and irradiance. The balance of saline north Atlantic currents, hydrological forces from Baltic currents, and land runoff lead to considerable salinity and temperature flucations and seasonal water column stratification. Earlier studies in Skagerrak have revealed the dynamics and distribution of organisms belonging to different trophic and taxonomic groups such as dinoflagellates, diatoms, haptophytes, cryptophytes, prasinophytes, dictyochophytes and euglenoids. Due to the location of Skagerrak on a continental shelf makes the biomass of the protist community mainly comprises of diatoms, dinoflagellates and haptophytes (Stadniczeñko, 2019). The reason for interest in this location is that our area of project is in an important place where there are two different waters, on one side you have Skagerrak which runs along the east coast of Norway, from Lindesnes and west coast of Norway, the North Sea current runs, there we have potential for different species to prefer the different waters and temperature. There is also very limited research on these locations.

1.8 Aims of the project

- To determine the different groups and species of protists present in spring along the Agder coast.
- Compare species composition in inner and outer coastal areas of Agder.
- Study the potential changes in protist community composition along two fjord transects in the Agder coast.

2 Methods

2.1 Sample collection

Seawater samples were collected in Niskin bottles on 8-20 April 2021 from 12 different locations along the Agder coast from the research vessel G.M Dannevig, ranging from Farsund in the west to Kristiansand in the east (Fig 3). Samples were collected at four standard depths; sea surface (0m), shallow (5m), intermediate (15m), deep (5-10m from bottom) (Table 1) and filtered through 0.22µl Sterivex ™ filters (Merck KGaA, Darmstadt, Germany) using a peristaltic pump (Master Flex L/S®, Cole-Parmer, Illinois, USA). Filters were stored at -80 C° before DNA isolation. Sampling depths were chosen based on the aim of potentially sampling different habitats for microbial protists. Temperature, salinity, oxygen, and PAR (photosynthetic active radiation) profiles from each station was also collected using SBE 911 CTD (Sea-Bird Electronics INC, Bellevue, Washington, USA). The CTD equipment was attached to the Niskin carousel and was lowered to 5-10m from the bottom.

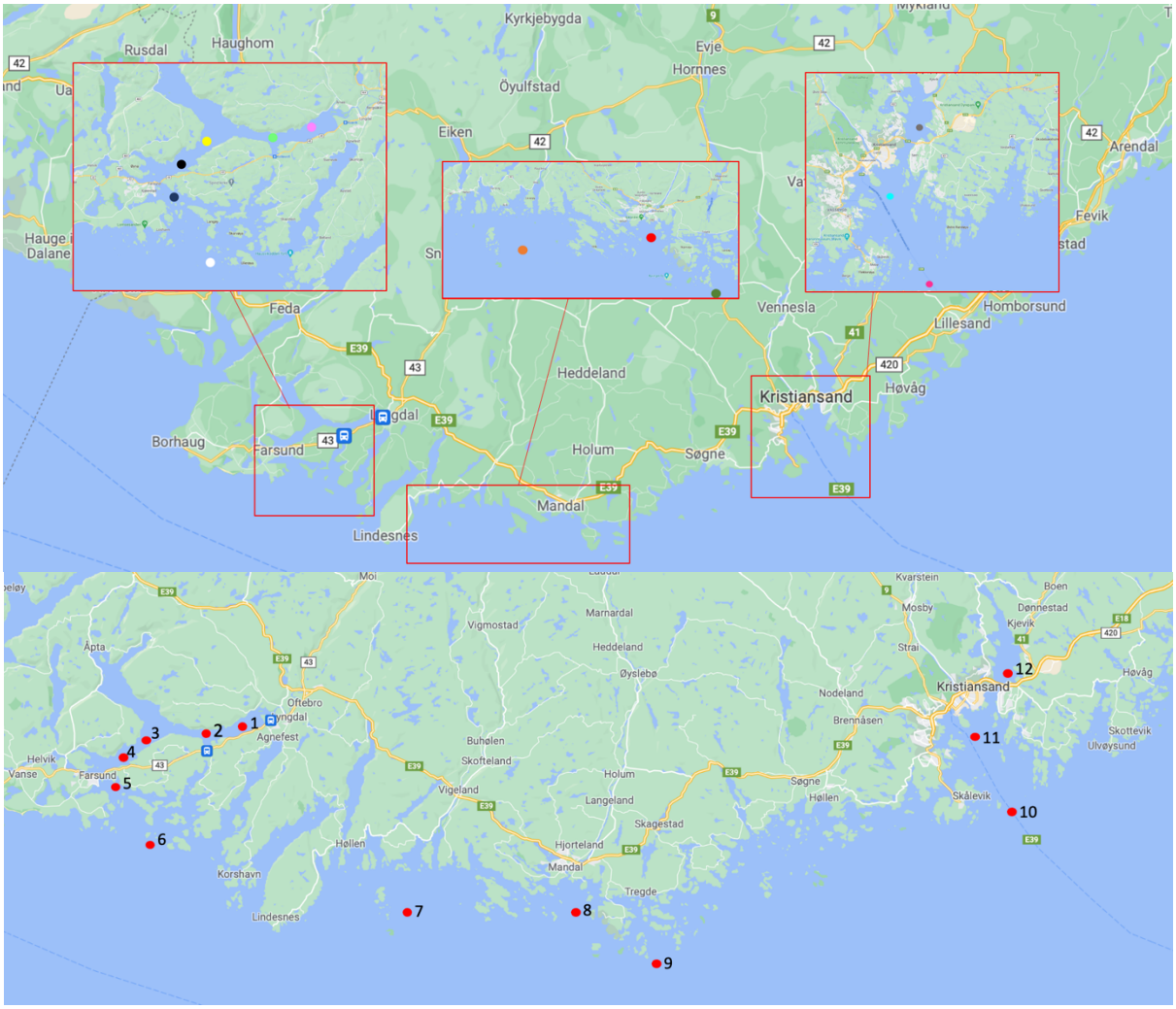


Figure 3: Sample stations along the Agder coast in South of Norway. 1/ light pink – Kvavik, 2/light green – Kollevoll, 3/yellow – Lok1, 4/black – Lok2, 5/marine blue – Midtfjordssjær, 6/white – Lindholmen, 7/orange – Springen, 8/red – Mannefjorden, 9/green – Oddhausen, 10/dark pink – Oksøy, 11/turquoise – Byfjorden, 12/purple – Topdalsfjorden. Modified map extracted from Google.com (2021).

Table 1: Sample collection location name, date, depth and GPS coordination. Depth color: yellow – 0m, pink – 5m, blue – 15m, green – deep (5m from bottom)

Number	Location	Date	Depth	GPS COORD
1	Kvavik	09.04.2021	0M	N58°7.879 E7°0.204
1	Kvavik	09.04.2021	5M	N58°7.879 E7°0.204
1	Kvavik	09.04.2021	15M	N58°7.879 E7°0.204
1	Kvavik	09.04.2021	81M	N58°7.879 E7°0.204
2	Kollevoll	09.04.2021	0M	N58°07.532 E6°57.117
2	Kollevoll	09.04.2021	5M	N58°07.532 E6°57.117
2	Kollevoll	09.04.2021	15M	N58°07.532 E6°57.117
2	Kollevoll	09.04.2021	100M	N58°07.532 E6°57.117
3	Lok 1	09.04.2021	0M	N58°07.216 E6°52.279
3	Lok 1	09.04.2021	5M	N58°07.216 E6°52.279
3	Lok 1	09.04.2021	15M	N58°07.216 E6°52.279
3	Lok 1	09.04.2021	240M	N58°07.216 E6°52.279
4	Lok 2	09.04.2021	0M	N58°06.299 E6°49.969
4	Lok 2	09.04.2021	5M	N58°06.299 E6°49.969
4	Lok 2	09.04.2021	15M	N58°06.299 E6°49.969
4	Lok 2	09.04.2021	140M	N58°06.299 E6°49.969
5	MFS	09.04.2021	0M	N58°05.881 E6°49.881
5	MFS	09.04.2021	5M	N58°05.881 E6°49.881
5	MFS	09.04.2021	15M	N58°05.881 E6°49.881
5	MFS	09.04.2021	90M	N58°05.881 E6°49.881
6	Lindholmen	10.04.2021	0M	N58°02.368 E6°52.734
6	Lindholmen	10.04.2021	5M	N58°02.368 E6°52.734
6	Lindholmen	10.04.2021	15M	N58°02.368 E6°52.734
6	Lindholmen	10.04.2021	90M	N58°02.368 E6°52.734
7	Springen	10.04.2021	0M	N 57°59.453 E 7°16.020
7	Springen	10.04.2021	5M	N 57°59.453 E 7°16.020
7	Springen	10.04.2021	15M	N 57°59.453 E 7°16.020
7	Springen	10.04.2021	150M	N 57°59.453 E 7°16.020
8	Mannefjord	10.04.2021	0M	N 58°00.176 E 7°28.198
8	Mannefjord	10.04.2021	5M	N 58°00.176 E 7°28.198
8	Mannefjord	10.04.2021	15M	N 58°00.176 E 7°28.198
8	Mannefjord	10.04.2021	100M	N 58°00.176 E 7°28.198
9	Oddhausen	10.04.2021	0M	N 57°57.156 E7°34.169
9	Oddhausen	10.04.2021	5M	N 57°57.156 E7°34.169
9	Oddhausen	10.04.2021	15M	N 57°57.156 E7°34.169
9	Oddhausen	10.04.2021	310M	N 57°57.156 E7°34.169
10	Oksøy	11.04.2021	0M	N58°03.391 E8°05.188
10	Oksøy	11.04.2021	5M	N58°03.391 E8°05.188
10	Oksøy	11.04.2021	15M	N58°03.391 E8°05.188
10	Oksøy	11.04.2021	230M	N58°03.391 E8°05.188
11	Byfjorden	11.04.2021	0M	N58°07.187 E8°01.860

11	Byfjorden	11.04.2021	5M	N58°07.187 E8°01.860
11	Byfjorden	11.04.2021	15M	N58°07.187 E8°01.860
11	Byfjorden	11.04.2021	180M	N58°07.187 E8°01.860
12	Topdalsfjorden	11.04.2021	0M	58 10.28 N 008 04.00 E
12	Topdalsfjorden	11.04.2021	5M	58 10.28 N 008 04.00 E
12	Topdalsfjorden	11.04.2021	15M	58 10.28 N 008 04.00 E
12	Topdalsfjorden	11.04.2021	65M	58 10.28 N 008 04.00 E

2.2 DNA Isolation and Illumina sequencing

DNA was extracted from 48 samples using the DNeasy® Plant minikit (Qiagen GmbH, Hilden, Germany) according to manufacturer's protocol. All steps were followed except for a few alterations. The volume of AP1 was changed to 500µl, and approximately 0.3 of Biospec Silica beads. The samples were treated to 1 round of MagNA Lyser Rotor (Roche Molecular Systems Inc, Switzerland) (1 time at speed 2800 for 48 seconds). The lysate was carefully pipetted out (around 500 µl). At the last step of the protocol 80 µl of buffer AE were added to elute, this step was repeated using the flow-through to potentially gain a higher DNA concentration. Products of DNA isolation were quantified spectrophotometrically (NanoDrop One microvolume UV-Vis spectrophotometer, Thermo fisher Scientific Inc.) prior to polymerase chain reaction (PCR). Test PCRs were to identify amplifiable DNA after extraction was carried out in a total volume of 25µl containing 12,5µl DreamTaq Green PCR Master Mix, 0.5µl Forward primer V4F (GTGCCAGCMGCCGCGGTAA), 0.5µl Reverse primer V4R (GGACTACHVGGGTWTCTAAT), 9.5µl Water, 2 µl DNA sample. Reactions were run on Thermo Scientific™ Owl™ EasyCast™ B1 Mini Gel Electrophoresis Systems (Thermo Fisher Scientific Inc., USA). The cycle conditions were as followed; 95°C for 3 min; 35 cycles of 95°C for 1 min, 52°C for 30 s, 72°C for 1 min; 72°C for 5 min. PCRs were carried out on all negative extractions controls to confirm their negative concentration and also random samples prior to being shipped to Integrated Microbiome Resource, Dalhouise University, Halifax, for library preparation and Illumina Miseq sequencing. The sequenced region was the variable V4 region of 18S DNA using the primers TAReuk454FWD1A (CCAGCASCYGCGGTAATTCC) and V4 18S Next.Rev (ACTTTCGTTCTTGATYRATGA) (Piredda et al., 2016).

2.3 Data analysis

Analysis of the Illumina data was performed using RStudio (version 4.1.3;R Core team (2021)), and the CTD data were visualized using Ocean Data View (ODV) (Schlitzer, 2018). The illumina data were analyzed using dada2 (Callahan et al., 2016) following the tutorial by Daniel Vaulot (https://vaulot.github.io/tutorials/R_dada2_tutorial.html) with a few exceptions described below. The following packages were installed in R Studio and used the analyses: readr, readxl, dplyr, tibble, tidyr, stringr, ggplot, dada2, phyloseq and Biostrings. Due to computer capacity the new script decipher was used to assign taxonomy based on the PR2 database (pr2_version_4.14.0_dada2_deciher.fasta.gz) rather than more commonly used assignTaxonomy. In addition, primers were removed not by using cut-adapt but by removing 20 nucleotides from the beginning and 21 reads from the end of the sequences. Phyloseq is a package used to store larger amounts of data and combine several files into one file, especially phylogenetic sequencing data which are in an OTU table. Non-metric Multi-dimensional Scaling (NMDS) analyses based on the dissimilarity matrix was used to explore community patterns. Alpha diversity measures were calculated to get a visual of the observed taxonomy richness or evenness. Chao1 is a non-parametric method of calculating species in a community (Bo-Ra et al., 2017). Chao1 gives more weight to low abundance species in a community. Therefore, in dataset were the samples are skewed or very different in sequence amount, all taxonomy will count and it may show a more accurate picture of the community deposition (Bo-Ra et al., 2017). Shannon-weaver (Shannon) look at population index and population variety (Hennink & Zeven, 1990).

3. Results

Here vi have identified and compared the different depths in the inner and outer fjord along the Agder coast which represents both Skagerrak and the North Sea. The results show a difference of species abundance on the difference levels of depths. There is a higher abundance on 5m and 15m, than on sea surface and deep samples. In the next section we will go further into the different taxa identified and the environmental parameters collected.

3.1 Species abundances

The dataset included 2 868 992 sequences after filtering and removal of the ones which occurred in less than 10 copies. The 16 most abundant OTUs represented 38% of the total sequences (Table 2.).

The 5 most abundant divisions identified in the dataset were Chlorophyta, Ciliophora, Dinoflagellata, Opalozoa, and Picozoa. Dinoflagellata were more profound of a larger abundant than other divisions in all depth layers (Fig 6 and 7.). When we go into class level on the OTU table the division of Dinoflagellata includes the classes Syndiniales, Dinophyceae and dinophyta. Ciliophora division includes class level of Spirotrichea, Oligohymenophorea, and Litostomatea. Opalozoa contains mainly of MAST-3 and MAST-12. The division of Chlorophyta includes, Chlorophyceae, Mamiellophyceae, Ulvophyceae, and Pyramimonadophyceae. Dinoflagellata was more abundant on 15m depth and deep, while Picozoa, Ciliophora and Chlorphyta was more abundant on 15m and seasurface. Opalozoa was most abundant on 15m. On the NMDS plots (fig 8.) there is a clustering of deep water samples, as well as Intermediate 15m samples.

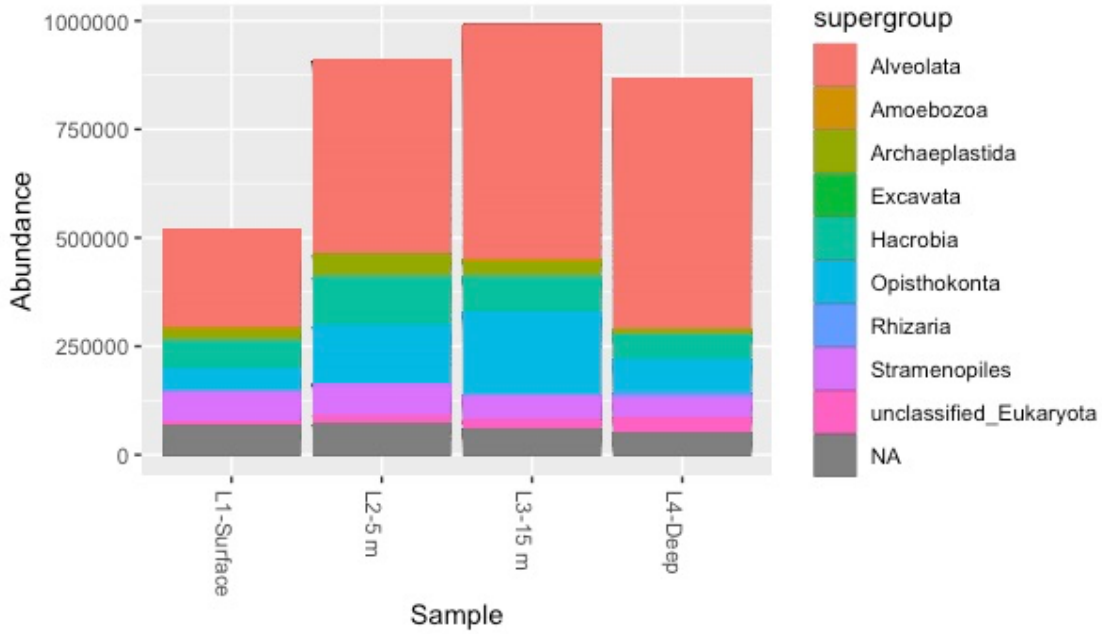


Fig 4. Supergroups identified in the total data set divided into depth layers. L1-surface = sea surface, L2-5m = shallow (5m), L3-15m = intermediate (15m), L4-deep = deep (5-10m from bottom). This plot is based on normalized data.

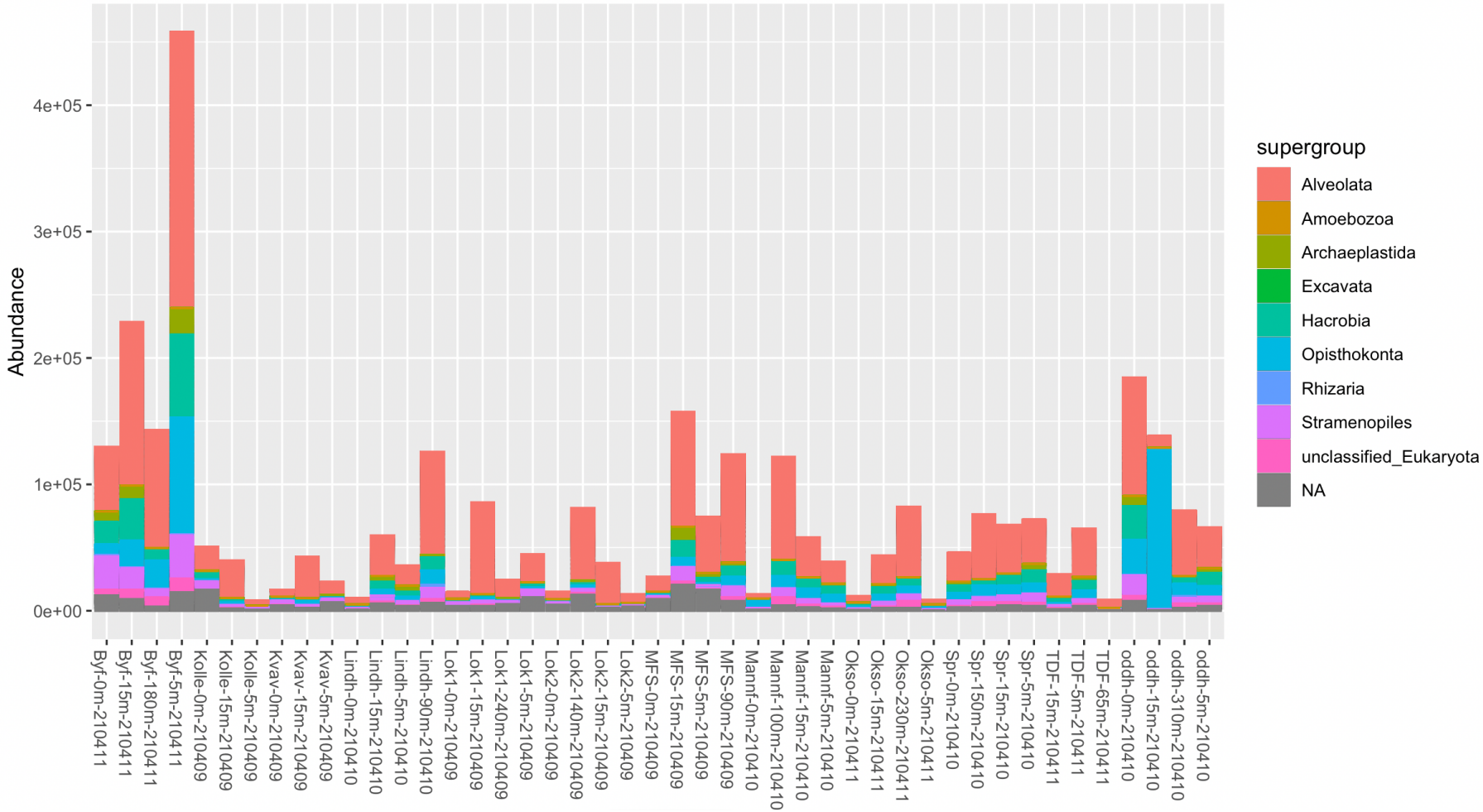


Fig 5. Supergroups identified into stations sorted in alphabetical order. Coloring represents the different supergroups as stated in Fig 4. This plot is based on non-normalized data.

Table 2: Identification of top 16 abundant sequences. Sequences were assigned using the decipher algorithm in R and using blastn searches on NCBI. OTU number is from the OTU table resulting from the dada2 pipeline. Accession ID and % are the NCBI sequence ID and the percentage of sequence similarity in blastn searches.

OTU	Decipher Species	Decipher Division	BLAST Species	Accession ID	%
00002	Dino-group-II-clade-10-and 11_X_sp	Dinoflagellata	Uncultured marine alveolate	KC488506.1	100%
00003	NA	NA	Amoebophrya sp	KY980047.1	100%
00004	NA	Dinoflagellata	Gyrodinium helveticum	FJ024299.1	99.7%
00005	NA	Dinoflagellata	Heterocapsa rotundata	KY980285.1	100%
00006	Unclassified_Heterocapsa	Dinoflagellata	Heterocapsa rotundata	KY980397.1	100%
00007	unclassified_Gymnodiniales	Dinoflagellata	Gyrodinium fusiforme	AB120002.1	100%
00009	Dino-Group-II-Clade-7_X_sp.	Dinoflagellata	Uncultered marine syndinales	FJ431620.1	100%
00010	Unclassified_micromonas	Chlorophyta	Micromonas commoda	MT117943.1	100%
00011	NA	NA	Prorocentrum sp.	MN824022.1	99.7%
00012	NA	Cryptophyceae	Teleaulax amphioxeia	MK956825.1	100%
00013	Unclassified_Heterocapsa	Dinoflagellata	Heterocapsa rotundata	KY980397.1	99.7%
00014	Picozoa_XXXX_sp	Picozoa	Picomonas sp.	MZ687537.1	99.2%
00016	Picozoa_XXXX_sp	Picozoa	Picobiliphyta sp.	JN934893.1	100%
00017	NA	NA	Picobiliphyta sp.	JN934892.1	89.7%
00018	Unclassified_gyrodinium	Dinoflagellata	Gyrodinium sp.	MZ687483.1	100%
00019	Dino-Group-I-Clade-1_X_sp.	Dinoflagellata	Karlodinium veneficum	KY979983.1	100%

3.1.1 Layer; Sea surface – 0m

Out of all sequences obtained, the sea surface samples represented 16,5% (475 296) of the sequences, which was the lowest number of all depths. Still, a considerable number of sequences were found. The Supergroup level mostly consisted of Alveolata and Stramenopiles (Fig 4.). On division level (Fig 7.) in sea surface layer Dinoflagellata dominated in abundance of sequences, while Chlorophyta, Stramenopiles_x, Ochrophyta and Picozoa was also abundant in less abundance. The NMDS (fig 8.) shows that sea surface samples were quite spread out, in comparison to the other depth layers are more clustered. These results suggest a dissimilarity between the sea surface locations.

3.1.2 Layer; Shallow – 5m

The shallow samples collected at 5m depth represented 27,3% (748 001) of the total sequences. In this layer (5m) most of the sequences represented Alveolata, having near half the sequence abundance in all locations on 5m depths (Fig 4.). There was also a higher amount of Ophisthokonta, Archaeplastida, Stramenopiles and Hacrobia. When looking into the division layer of the sequences, Dinoflagellata was still the most abundant taxa, closely followed by Ochrophyta, Chlorophyta, Picozoa and Ciliophora (Fig 7.). NMDS of the shallow area showed a clustering of 4 locations, while the rest were more scattered around the ordination plot (Fig 8.)

3.1.3 Layer; Intermediate – 15m

The intermediate layer had the highest number sequences with of 28% (805 777). Out of supergroups Alveolata represented almost half the sequences (Fig 4.). The supergroups following Alveolata was Ophisthokonta, Hacrobia, Stramenopiles, and Archaeplastida. The division level showed a high abundance of sequenced Dinoflagellata, followed by Ochrophyta, Chlorophyta, Ciliophora, Picozoa and Cryptophyta (Fig 7.). The NMDS (Fig 8.) plot showed that the intermediate layer (15m) had a clustering of 6 locations while the remaining 6 were very scattered.

3.1.4 Layer; Deep – 5-10m from bottom.

The deep water samples in the dataset represented around 28% (803 918) of the total sequences obtained. The samples taken at the deep (5-10m from bottom) showed a high number of sequences and showed the second highest sequence abundance. It showed a varied result of species from several divisions. The supergroups represented were Alveolata with the highest number of sequences. After Alveolata, the next supergroups

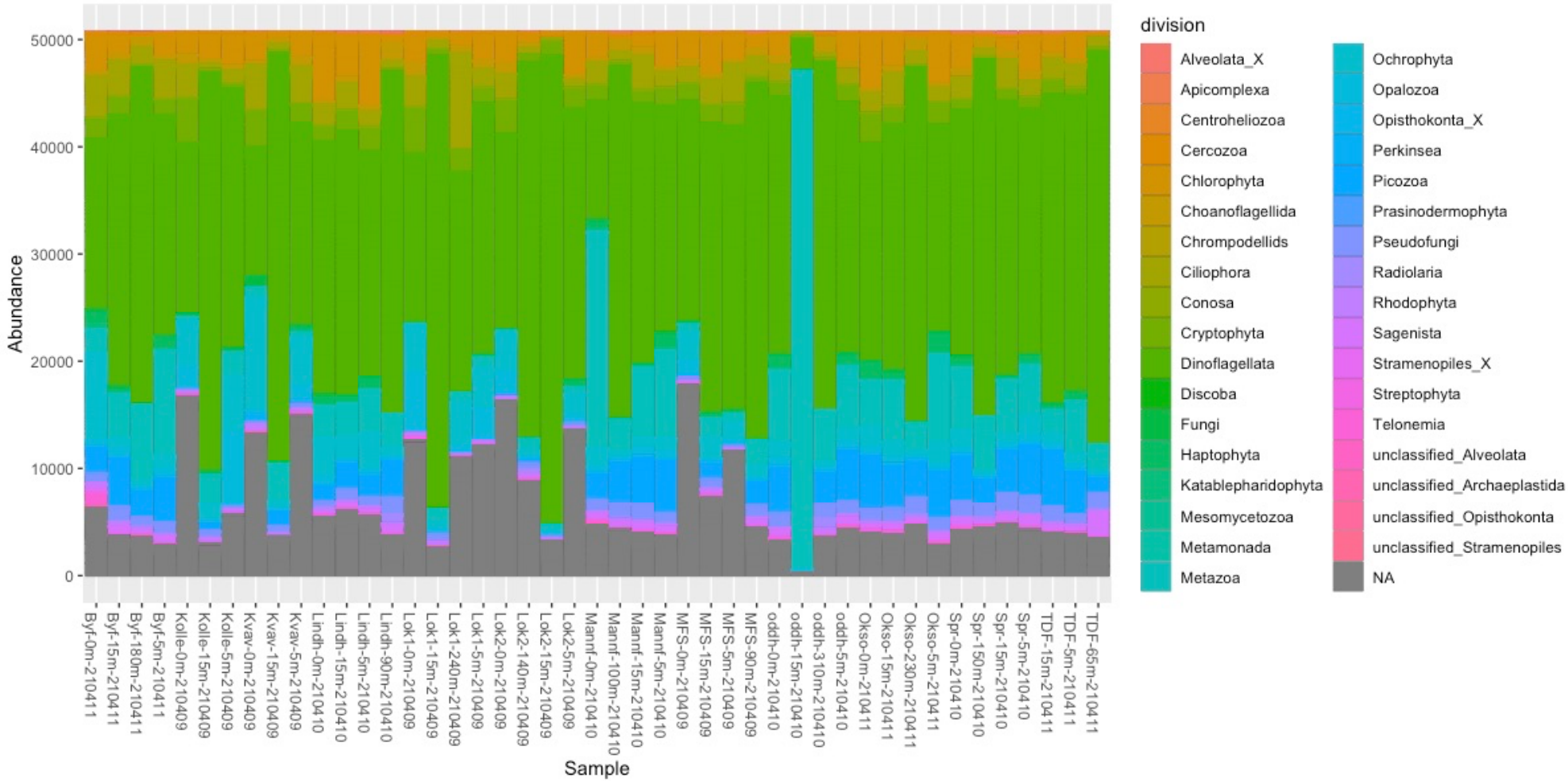


Fig 6. Bar plot of the identified divisions categorized by location sorted in alphabetical order. The plot is based on normalized data.

were Ophisthokonta, Stramenopiles, and Hacrobia (Fig 4.). At division level, Dinoflagellata had the highest sequence number, over half of the sequences were Dinoflagellata (Fig 7.). Ochrophyta was the second highest sequence abundance, followed by Picozoa, Chlorophyta, Ciliophora, Opalozoa and Ophisthokonta_X. Deep samples had the largest clustering on the ordination plot NMDS (Fig 8.), with 9 clustering locations and only 3 locations with a dissimilarity.

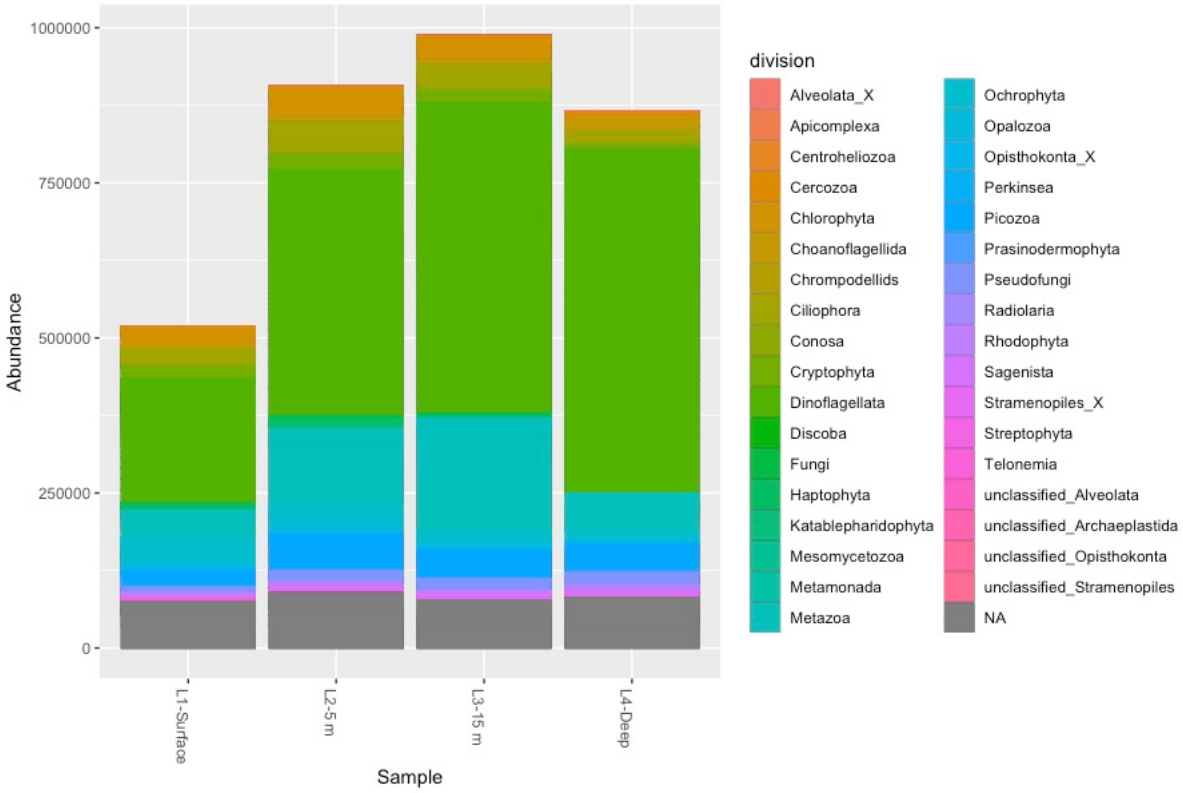


Fig 7. Plot divided into depth layer to represent which division is present. Four sample layers; L1 (sea surface), L2 (5m), L3 (15m) and L4 (deep). Coloring represents the division. Dinoflagellata (green) is by far the most abundant in the intermediate 15m layer. Chlorophyta (red) is second most abundant, following by Picozoa (purple) and Opalozoa (blue). The plot is based on normalized data.

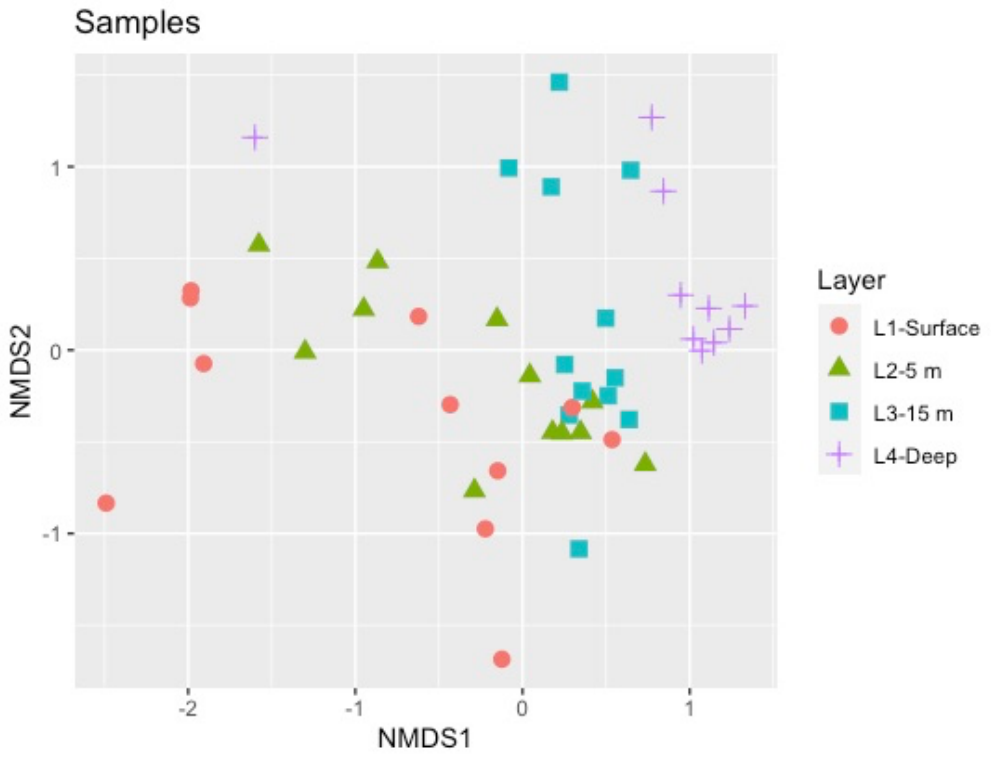


Fig 8. NMDS of full data set with all locations and depths. Deep (round, red) is clustering, as well as intermediate (triangle, green), with some sea surface mixing with the intermediate samples.

3.2 Diversity measures

Shannon-Weaner (aka Shannon) alpha diversity measure suggest that most of the locations had a similarity of diversity except for location Oddhausen (15m) which was a clear outlier sample (Fig 9). Especially the deep and intermediate samples were closely related and even. Chao 1 showed a more diverse and scattered sample set on all locations, the several depths were still on the same scale. Therefore, it is an evenness of the same depth samples and they show a more alike results. When comparing temperature and salinity in Chao1 calculation, there is a sign that colder temperatures give a higher result of salinity (Fig 10.). There are some offset samples like in shallow (5m) and intermediate (15m), there are two locations on both layers that have a warmer temperature (light blue) in the middle of several dark blue locations. This also occur in the salinity figure.

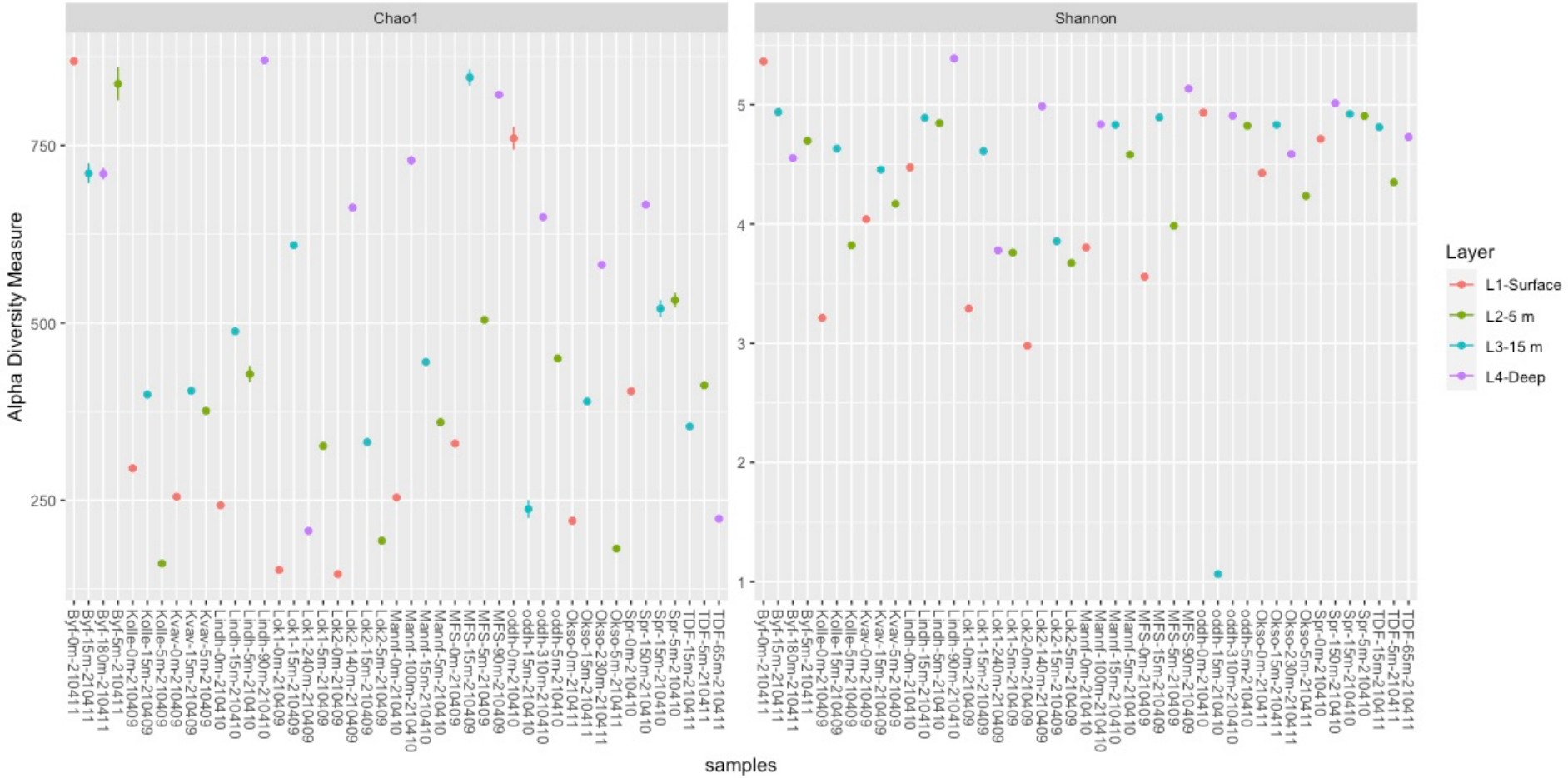


Fig. 9. Alpha diversity measure Chao1 and Shannon-Weaver . Each location and depth are represented on the X-axis, and coloring represent the depth layer. Red = sea surface, Green = shallow (5m), Blue = intermediate (15m) and Purple = deep (5-10m from bottom).

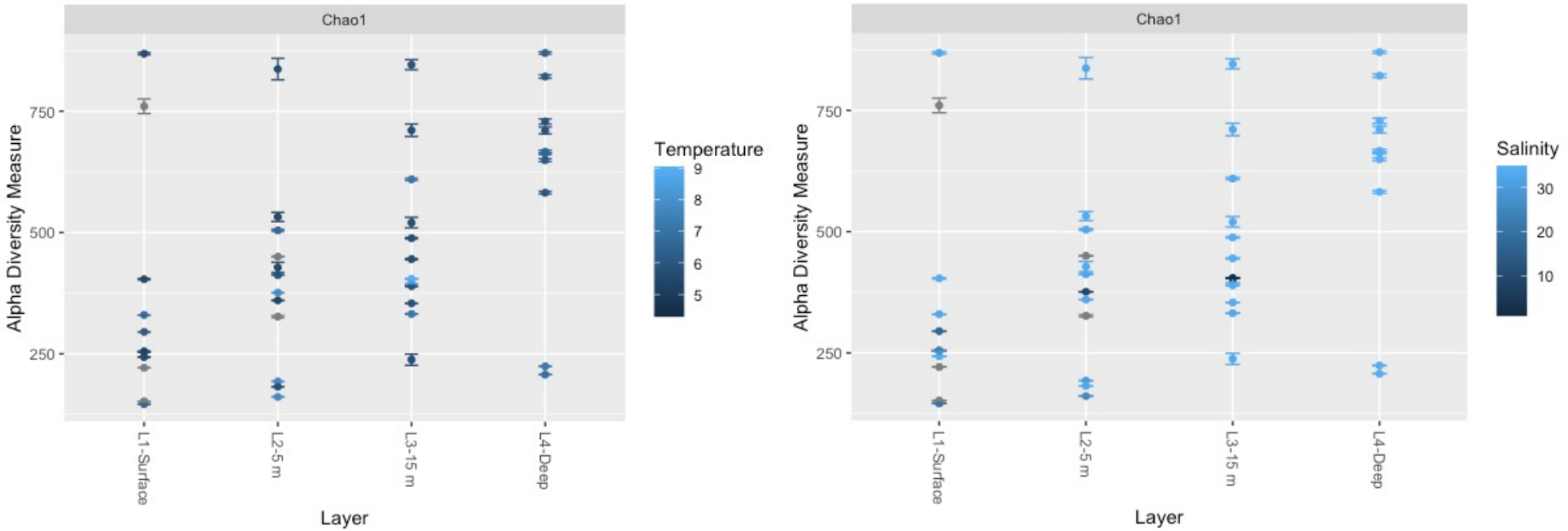


Fig 10. Chao1 diversity measures of temperature and salinity on the four depth layers. Temperature ranging from 9° to 5° degrees Celsius, going from light (9°C) to dark (5°C). Salinity measures goes from 30 (light blue) to 10 (dark blue) as well as some NA (grey).

3.3 Environmental factors

The CTD data showed warmer temperatures in more sheltered locations (Fig 11.) like Kollevoll and Kvavik in Farsund, as well as Topdalsfjorden on the other side of the transect. Temperatures was 5°C in the upper part of the water column, warmer in the 5-10m depth 11°C to 10°C, and towards the bottom of the ocean it gets colder, down to about 6 C°. Figures of full range/depths of temperature and salinity can be found in Appendix 1. In the more outer locations, which are more open and not sheltered there was a stabile temperature around 6 °C on all depths (Fig 11.). The CTD data showed anoxic water (Fig 11.) below 20m depth with as little as 0 to 0.5 oxygen in the inner locations towards the river Lygna. Overall, the locations all showed decreased amount of oxygen in the top sea surface layer (0-2m depth). When it comes to salinity (Fig 11.) we can see in the upper 0-5m water surface there is a decreased amount of salinity in the Farsund stations that are more inner fjord and sheltered. This is highly due to lakes running down causing brackish water. We can also see this pattern in the more sheltered inner locations in Topdalsfjorden, Kristiansand. There was also some irradiance in all locations in the upper 5-6m water column (Fig. 11). This shows algae biomass and might interpret that it may be a spring bloom in Lindholmen, Springen and Mannefjorden. At the same locations there was a higher amount of irradiance at 15m. To look especially for a spring bloom of protists we had two fluorescence plots made in ODV. Fluorescence CDOM (Fig. 12) show a bloom in the first 3 locations in Farsund; Kvavik, Kollevoll and Lok1 as well as some higher fluorescence in Topdalsfjorden in the upper water column. Fluorescence Chelsea UV looks for phototrophic activity and it is noted in the Farsund locations, Kvavik, Kollevoll, Lok1 and Lok2 (Fig 13.)

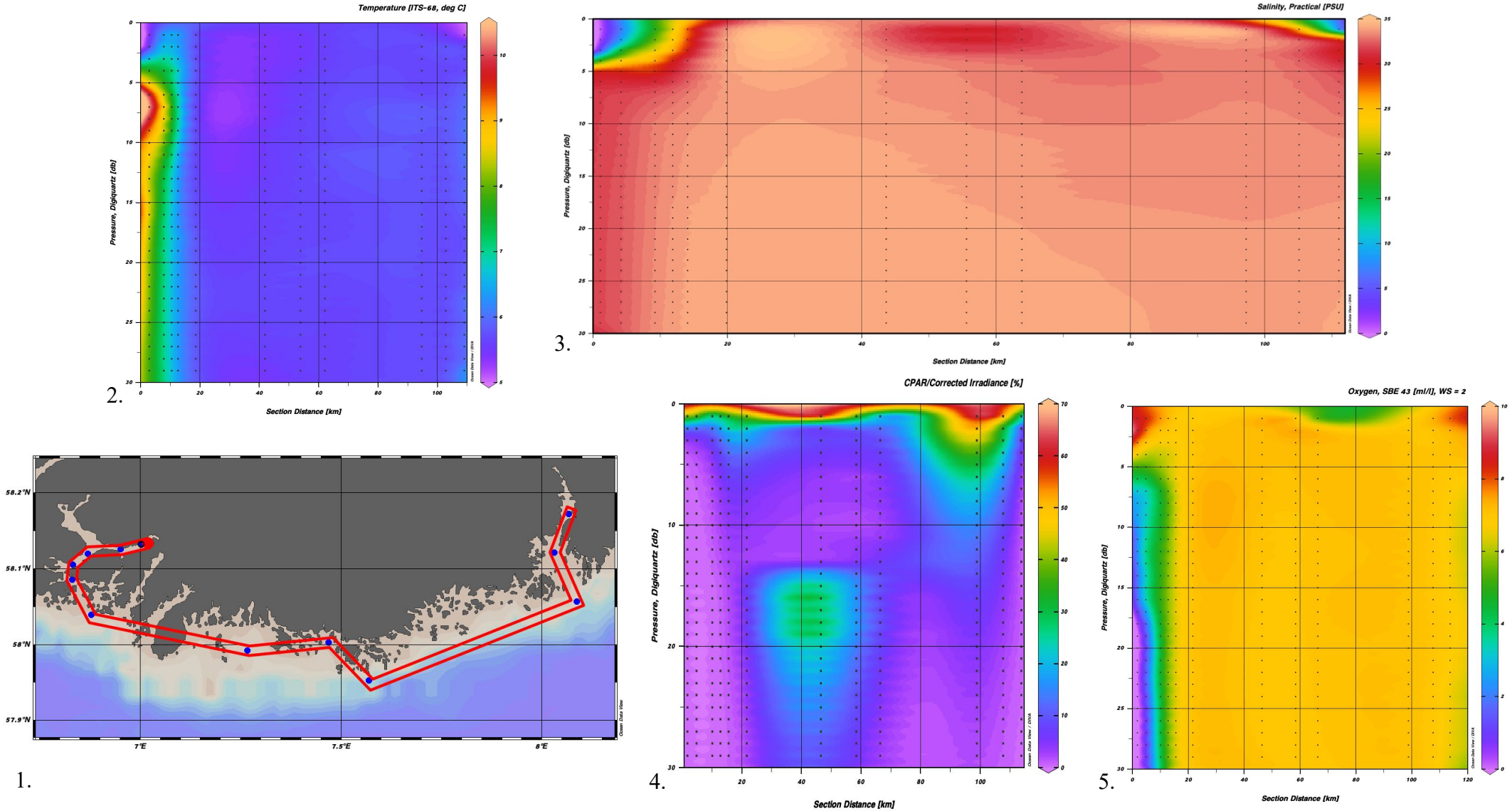


Figure 11: Environmental data obtained through CTD, made in Ocean data view (ODV). 1. Map obtain through ODV over transect, Kvavik in the west and Topdalsfjorden in the east. 2. Temperature measured against depth (max 30m). 3. Salinity measured against depth (max 30m). 4. Irradiance measured against depth (max 30m). 5. Oxygen measured against depth (max 30m).

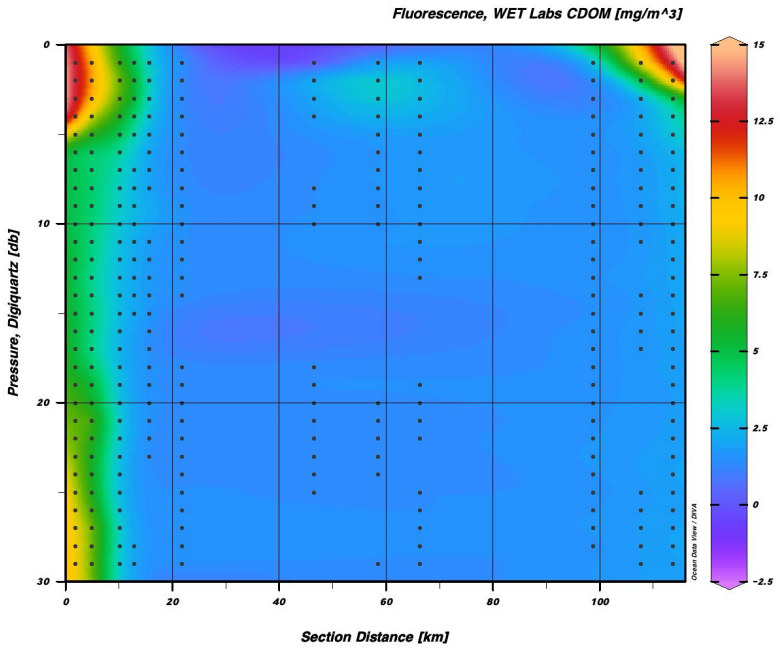


Fig 12. Fluorescence of all locations measured against depth and section distance (distance between locations and transect). CDOM visible (green, yellow, and red coloring) in the more sheltered inner locations than the more outer locations (blue). The plot is capped at 30m.

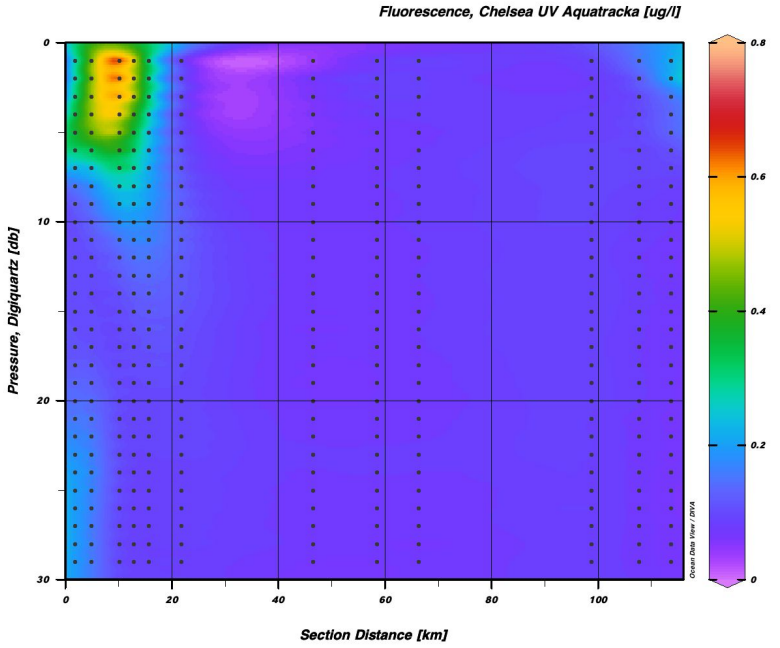


Fig 13. Cl a fluorescence (Chelsea UV) along the studied locations. Green and yellow with a tint of red suggested a higher biomass of phototrophic species in those inner locations, compared to the more outer locations (blue). The plot is capped at 30m.

4. Discussion

4.1 Community composition

Protists from four different depths representing 12 locations and two fjord transects along the Agder coastline were determined in this study using Illumina metabarcoding. The protist communities at different depths and locations were compared and connected to obtained information about the environmental conditions. The five supergroups of abundant protists, Chlorophyta, Ciliophora, Dinoflagellata, Opalozoa and Picozoa were found on all locations and were of a greater abundance than other groups. Especially Dinoflagellata were largely dominating the sequence OTU abundances as often found in studies based on rDNA metabarcoding of protists. A similar study were conducted in Northern part of Skagerrak, were Dinoflagellates were the most abundant phylum (Gran-Stadniczenko et al., 2019), as well as there is hypnotized that Dinoflagellata, Ciliophora and Chlorophyta are present in spring blooms (Niemi et al., 2011; Seuthe et al., 2011). Marine alveolates group named MALV are placed in the dinoflagellates, in this group *Amebophyra*, has been identified as a dominant species in marine metabarcoding assignments (Gran-Stadniczenko et al., 2019; López-García et al., 2001). This is similar to our results, were *Amebophyra* was a dominated species sequence and second highest OTUs. One of the most sequenced phylum in this dataset was *Heterocapsa Rotundata*, this is a phylum that is known to take up a higher proportion of proliferating dinoflagellates cells (Pertola et al., 2006). The results of the community composition in these locations showed an evenness in diversity across the two transects. This is supported by several studies in similar conditions, which also have a diverse but even results (Egge et al., 2015; Gran-Stadniczenko et al., 2019; Pertola et al., 2006).

4.2 Community composition in relation to environmental factors

In the inner fjord location of Kvavik, Lyngdal, there was more anoxic conditions below 6m depth. This location also had a lower level of salinity. When it comes to abundance in OTU sequences, the innermost locations Kollevoll and Kvavik in Lyngdalsfjorden had less sequence abundance of all depths. Kollevoll and Kvavik had a percentage of 6% of identified sequences in total on all depths, to compare Byfjorden had 28% of identified sequences. The reason for this large difference is most likely due to being a decreasing amount of oxygen from 7m depth. Also, in the upper water column there is a higher

amount of freshwater. Therefore, in combination with less salinity in the water and the raised temperature, there is not the best living conditions for any species of seawater origin. Kollevoll and Kvavik also did not have a sequenced deep layer (5-10m from bottom) sample due to being failed. For a sample to be failed it does not have any viable DNA present due to either anoxic sample or broken sample, e.g., failure during DNA extraction or during Illumina sequencing. It may appear that Kollevoll and Kvavik had a case of hypoxia or more commonly known as dead zone, where oxygen is cut off from the bottom water column, normally due to increase in temperature or/and salinity which cause a stratification between the water columns. This cause the oxygen levels to decline below levels which life can sustain (Diaz, 2016; Winder & Sommer, 2012). These changes can affect indirectly on the phytoplankton community through nutrient and light availability and stratification (Winder & Sommer, 2012). In the case of Kvavik and Kollevoll the dead zone is caused by front from Lok1 separating for water to go through, therefore, the water in these locations have had not much circulation or stream in many years, this has also been seen in by Røed and Albretsen (2007). It causes a dividing of less saline water from the more saline water. There was a high Cl a fluorescence in the inner locations especially in the upper water column. A larger biomass of phytoplankton is visible in the upper column in all Farsund locations but only on the upper water column (Fig 11, 12 and 13). This build up under the hypothesis that there is a stratification in these locations. It is suggested that with warmer temperatures the production of light-saturated phytoplankton may increase (Padilla-Gamiño & Carpenter, 2007; Sommer & Lewandowskal, 2011). In our two transects there was an increase of temperature in the inner locations on both transects; Lyngdal/Farsund; Kollevoll, Kvavik, Lok1 and Lok2, and Kristiansand; Topdalsfjorden. In the first four locations there is visible river runoff (Fig 12.) which is a result that is expected, due to being close to rivers on each side (Polovodova Asteman et al., 2018; Røed & Albretsen, 2007). The CDOM (Chromophoric dissolved organic matter) identifies organic matter by absorbing light over a broad range of UV and visible wavelengths, and is caused by *in situ* biological production, terrestrial sources, photochemical degradation and microbial consumption (Para et al., 2010). Along our two transects there is visible CDOM in the inner locations in each transect.

4.3 Sea surface layer lowest in OTUs sequence richness, but diversity is still quite even with the other layers

Sea surface layer had the lowest number of sequences. Shannon-weaver showed an evenness of sea surface locations with the rest of the dataset, as well as an evenness in species richness in this layer (Fig 9.). Therefore, sea surface was not the layer where we expected to find the most unique and diverse species composition. There was on some locations findings of Syndiniales on sea surface. Thus, the number was quite low and in random locations. On sea surface and 5m, Chlorophyta represented around 1.7% of the total sequences and *Micromonas commoda* was among the most abundant taxa found in this study. Chlorophyta is a division that is known to be species living in freshwater as well as seawater and being robust to be able to withstand currents and rough weather. *Micromonas commoda* was found in Skagerrak for the first time in 2018 (Gran-Stadniczenko et al., 2019), prior to this there was no evidence of *Micromonas commoda* in Skagerrak/Kattegat waters, the species that was suggested and identified to be living there was *Micromonas pusilla* (Sahlsten, 1998). In the locations Midtfjordssjær, Lindholmen and Springen, the upper water column showed signs of an algae bloom, these locations also had higher abundance of Chlorophyta present at sea surface, 5m and 15m depths. Our CTD data also confirmed the potential bloom at this location down to about 15m. we observed a higher Chl a fluorescens at sea surface and it decreases down to about 20m, but still a higher amount than other locations.

4.3 Intermediate layer 15m had the highest OTUs richness

The 15m layer had the most results over the dataset. This is proven to be the depth most of the species identified prefer and inhabit the most. The total OTU sequences percentage on this depth was 28%. Dinoflagellata was more abundant on 15m depth, which shows that they thrive better on that depth than the other 3 depths. Thus, the abundance of Dinoflagellata are still quite high on the other depths. Oddhausen 15m was a unique sample. In all analyses this samples were unique and stood out. In identification of sequences this sample had a very high abundance of Ochrophyta. Phylum Ochrophyta have been observed to be present during spring blooms (Sahu et al., 2022). Syndiniales was in high sequencing abundance on both 15m and deep on almost all locations with few exceptions. This indicate that they thrive on deeper locations. Oddhausen which was one of our deepest locations had a higher abundance of Cryptophyta which contain representatives which can form toxic blooms. Oddhausen

was an odd result on all plots, therefore, looking into Shannon-weaver, this sample have very little richness in it and little diversity. This sample undermine the limitation that samples with less richness and mass of species is plotted against locations with high diversity and species richness. This also show how important alpha-diversity measures is to determine the composition of the community (Kleine Bardenhorst et al., 2022).

4.4 Changes in the transect from west to east.

Looking into the whole transect, there is a change from the inner sheltered locations into the more outer locations. The inner Lyngdal and Farsund locations had a warmer temperature, lower salinity, and decreased oxygen. These locations also had low Chao1 measures compared to other locations in the transects, this measure is to treat each species as one and also for it to be expected that there cannot be found any additional species when all species in the sample are represented by at least two individuals (Gotelli & Colwell, 2011). There is a change in runoff, suggested by CDOM levels, that there is more CDOM and discolored water in the inner more sheltered locations, these locations also have river connected to them. The locations in the middle of the transect have colder temperature, higher salinity and oxygen levels. Closer to Kristiansand and Topdalsfjorden there is sign of the same as inner Lyngdal, warmer temperatures, less oxygen and salinity. Thus, in the more outer area there was evidence of CDOM, to explain this it might be more farming in area around Lindholmen and Springen. There was a higher Chl a fluorescence at these two locations and it can also be seen in the results for identifications.

5. Conclusion

5.1 Concluding remarks

This study aimed to investigate the community diversity and potential richness in 12 locations with 4 depths in two fjord transects along the Agder coast. To do so water samples were collected and sequenced at the V9 region on 18S rRNA using Illumina metabarcoding. Our results showed a bloom of Dinoflagellates, Picozoa, Chlorophyta and Ochrophyta. Diversity measures supported these results by showing a high abundance of these divisions on almost all locations. We documented a higher level of runoff and discoloring (CDOM) in the inner locations on each side of the transect, Kollevoll and Kvavik, and Topdalsfjorden, in the upper water column. As well as anoxic

and high saline water in the bottom in these locations, and low salinity in the upper column. The diversity on these locations was lower than the rest of the locations in the sample set, which undermines the suggestion that the environmental conditions in the locations are not preferred for growth and living for protists.

5.2 Limitations and further research

There were a few limitations to this research project. Firstly, error sources were not removed during taxonomy assignment, therefore, samples with few sequences count as much as high abundant sequences, except for in diversity plots that that this in account. As well as there are several steps were technical error can occur and that will affect the results greatly. The CTD measurement tool is a very sensitive one, we experienced 1-2°C degrees in the water compared to what we expected due to weather on the day of sampling.

Further research should be focused on a timeseries, to look for a change in community diversity and mass over several years. Seasonal dynamics would also be interesting to have more research in these locations and identify the potential times for blooms in this area throughout a year E.g., similar to Gran-Stadniczeñko research in the Northern Skagerrak (2019).

Bibliography

Albretsen, J., Aure, J., Sætre, R., & Danielssen, D. S. (2011). Climatic variability in the Skagerrak and coastal waters of Norway. *ICES Journal of Marine Science*, *69*(5), 758-763. https://doi.org/10.1093/icesjms/fsr187

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, *115*(25), 6506-6511. https://doi.org/doi:10.1073/pnas.1711842115

Bo-Ra, K., Jiwon, S., Robin, B. G., Jun Hyung, L., Doo Wan, K., Kuk-Hwan, S., Ju-Hoon, L., & Hyeun Bum, K. (2017). Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *Journal of Microbiology and Biotechnology*, *27*(12), 2089-2093. https://doi.org/10.4014/jmb.1709.09027

Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. (2020). The new tree of eukaryotes. *Trends in ecology & evolution*, *35*(1), 43-55.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581-583. https://doi.org/10.1038/nmeth.3869

Caron, D. A., Worden, A. Z., Countway, P. D., Demir, E., & Heidelberg, K. B. (2009). Protists are microbes too: a perspective. *ISME J*, *3*(1), 4-12. https://doi.org/10.1038/ismej.2008.101

Decelle, J., Romac, S., Sasaki, E., Not, F., & Mahé, F. (2014). Intracellular Diversity of the V4 and V9 Regions of the 18S rRNA in Marine Protists (Radiolarians) Assessed by High-Throughput Sequencing. *PLOS ONE*, *9*(8), e104297. https://doi.org/10.1371/journal.pone.0104297

Diaz, R. J. (2016). Anoxia, Hypoxia, And Dead Zones. In M. J. Kennish (Ed.), *Encyclopedia of Estuaries* (pp. 19-29). Springer Netherlands. https://doi.org/10.1007/978-94-017-8801-4_82

Egge, E. S., Johannessen, T. V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R. A., & Edvardsen, B. (2015). Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol Ecol*, *24*(12), 3026-3042. https://doi.org/10.1111/mec.13160

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), 237-240.

Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, *13*(3), 133-146.

Gotelli, N., & Colwell, R. (2011). Estimating species richness. In (Vol. 12, pp. 39-54).

Gran-Stadniczenko, S., Egge, E., Hostyeva, V., Logares, R., Eikrem, W., & Edvardsen, B. (2019). Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as Revealed by Metabarcoding and Microscopy. *J Eukaryot Microbiol*, *66*(3), 494-513. https://doi.org/10.1111/jeu.12700

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R., Kirkham, A., Massana, R., Scanlan, D. J., & Worden, A. Z. (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental microbiology*, *10*(12), 3349-3365.

Gustafsson, B., & Stigebrandt, A. (1996). Dynamics of the freshwater-influenced surface layers in the Skagerrak. *Journal of Sea Research*, *35*(1-3), 39-53.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLOS ONE*, *9*(2), e87624. https://doi.org/10.1371/journal.pone.0087624

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313-321. https://doi.org/doi:10.1098/rspb.2002.2218

Heidelberg, K. B., Gilbert, J. A., & Joint, I. (2010). Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microbial biotechnology*, *3*(5), 531-543.

Hennink, S., & Zeven, A. (1990). The interpretation of Nei and Shannon-Weaver within population variation indices. *Euphytica*, *51*(3), 235-240.

Hodal, H., Falk-Petersen, S., Hop, H., Kristiansen, S., & Reigstad, M. (2012). Spring bloom dynamics in Kongsfjorden, Svalbard: nutrients, phytoplankton, protozoans and primary production. *Polar Biology*, *35*(2), 191-203.

Kaiser, M. J., Attrill, M. J., Williams, P. J. B., Jennings, S., Thomas, D. N., & Barnes, D. K. A. (2011). *Marine Ecology: Processes, Systems, and Impacts*. OUP Oxford. https://books.google.no/books?id=WYKcAQAAQBAJ

Kleine Bardenhorst, S., Vital, M., Karch, A., & Rübsamen, N. (2022). Richness estimation in microbiome data obtained from denoising pipelines. *Computational and Structural Biotechnology Journal*, *20*, 508-520. https://doi.org/https://doi.org/10.1016/j.csbj.2021.12.036

Kuylenstierna, M. K., Bengt. (2006). *Checklist of phytoplankton in the Skagerrak-Kattegat*. Retrieved March from https://www.smhi.se/oceanografi/oce_info_data/plankton_checklist/ssshome.htm

Leles, S. G., Mitra, A., Flynn, K. J., Stoecker, D. K., Hansen, P. J., Calbet, A., McManus, G. B., Sanders, R. W., Caron, D. A., Not, F., Hallegraeff, G. M., Pitta, P., Raven, J. A., Johnson, M. D., Glibert, P. M., & Våge, S. (2017). Oceanic protists with different forms of acquired phototrophy display contrasting biogeographies and abundance. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170664. https://doi.org/doi:10.1098/rspb.2017.0664

Levinsen, H., & Nielsen, T. G. (2002). The trophic role of marine pelagic ciliates and heterotrophic dinoflagellates in arctic and temperate coastal ecosystems: a cross-latitude comparison. *Limnology and Oceanography*, *47*(2), 427-439.

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, *45*(3), 373-385.

Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C., & Massana, R. (2012). Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *The ISME Journal*, *6*(10), 1823-1833. https://doi.org/10.1038/ismej.2012.36

López-García, P., Rodriguez-Valera, F., Pedrós-Alió, C., & Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, *409*(6820), 603-607.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., . . . Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376-380. https://doi.org/10.1038/nature03959

Marquardt, M. (2016). *Marine microbial eukaryotes in Svalbard waters* The artic university of Norway]. https://munin.uit.no/handle/10037/9804.

Marquardt, M., Vader, A., Stubner, E. I., Reigstad, M., & Gabrielsen, T. M. (2016). Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). *Appl Environ Microbiol*, *82*(6), 1868-1880. https://doi.org/10.1128/AEM.03208-15

Niemi, A., Michel, C., Hille, K., & Poulin, M. (2011). Protist assemblages in winter sea ice: setting the stage for the spring ice algal bloom. *Polar Biology*, *34*(12), 1803-1817.

Nygaard, K., & Tobiesen, A. (1993). Bacterivory in algae: a survival strategy during nutrient limitation. *Limnology and Oceanography*, *38*(2), 273-279.

Okamoto, N., Chantangsi, C., Horák, A., Leander, B. S., & Keeling, P. J. (2009). Molecular phylogeny and description of the novel katablepharid Roombia truncata gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLOS ONE*, *4*(9), e7080.

Padilla-Gamiño, J. L., & Carpenter, R. C. (2007). Seasonal acclimatization of Asparagopsis taxiformis (Rhodophyta) from different biogeographic regions. *Limnology and Oceanography*, *52*(2), 833-842. https://doi.org/https://doi.org/10.4319/lo.2007.52.2.0833

Para, J., Coble, P., Charrière, B., Tedetti, M., Fontana, C., & Sempéré, R. (2010). Fluorescence and absorption properties of chromophoric dissolved organic matter (CDOM) in coastal surface waters of the Northwestern Mediterranean Sea (Bay of Marseilles, France). *Biogeosciences*, *7*, 4083-4103. https://doi.org/10.5194/bg-7-4083-2010

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., . . . de Vargas, C. (2012). CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol*, *10*(11), e1001419. https://doi.org/10.1371/journal.pbio.1001419

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12-25. https://doi.org/https://doi.org/10.1016/j.ejop.2016.02.003

Pertola, S., Faust, M. A., & Kuosa, H. (2006). Survey on germination and species composition of dinoflagellates from ballast tanks and recent sediments in ports on the South Coast of Finland, North-Eastern Baltic Sea. *Marine Pollution Bulletin*, *52*(8), 900-911. https://doi.org/https://doi.org/10.1016/j.marpolbul.2005.11.028

Piredda, R., Tomasino, M. P., D'Erchia, A. M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W. H. C. F., Sarno, D., & Zingone, A. (2016). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiology Ecology*, *93*(1). https://doi.org/10.1093/femsec/fiw200

Polovodova Asteman, I., Risebrobakken, B., Moros, M., Binczewska, A., Dobosz, S., Jansen, E., Sławińska, J., & Bąk, M. (2018). Late Holocene palaeoproductivity changes: a multi-proxy study in the Norwegian Trench and the Skagerrak, North Sea. *Boreas*, *47*(1), 238-255.

Quail, M. A., Swerdlow, H., & Turner, D. J. (2009). Improved protocols for the illumina genome analyzer sequencing system. *Current protocols in human genetics*, *Chapter 18*, 10.1002/0471142905.hg0471141802s0471142962-0471142918.0471142902. https://doi.org/10.1002/0471142905.hg1802s62

Røed, L. P., & Albretsen, J. (2007). The impact of freshwater discharges on the ocean circulation in the Skagerrak/northern North Sea area Part I: model validation. *Ocean Dynamics*, *57*(4), 269-285. https://doi.org/10.1007/s10236-007-0122-5

Sætre, R., Aure, J., & Danielssen, D. (2003). Long-term hydrographic variability patterns off the Norwegian coast and in the Skagerrak. ICES Marine Science Symposia,

Sahlsten, E. (1998). Seasonal abundance in Skagerrak-Kattegat coastal waters and host specificity of viruses infecting the marine photosynthetic flagellate Micromonas pusilla. *Aquatic microbial ecology*, *16*(2), 103-108.

Sahu, G., Mohanty, A. K., Sarangi, R. K., & Satpathy, K. K. (2022). Asterionellopsis glacialis (Family: Fragilariaceae, Class: Bacillariophyceae, Phylum: Ochrophyta) bloom and its impact on plankton dynamics at Kalpakkam (Bay of Bengal, Southeast coast of India). *Oceanologia*, *64*(1), 145-159. https://doi.org/https://doi.org/10.1016/j.oceano.2021.04.005

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1879-1888.

Schlitzer, R. (2018). *Ocean Data View*. In (Version 5.6.0) https://odv.awi.de

Seenivasan, R., Sausen, N., Medlin, L. K., & Melkonian, M. (2013). Picomonas judraskeda gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'picobiliphytes'. *PLOS ONE*, *8*(3), e59565.

Seuthe, L., Iversen, K. R., & Narcy, F. (2011). Microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): II. Ciliates and dinoflagellates. *Polar Biology*, *34*(5), 751-766. https://doi.org/10.1007/s00300-010-0930-9

Skov, H., & Durinck, J. (1998). Constancy of frontal aggregations of seabirds at the shelf break in the Skagerrak. *Journal of Sea Research*, *39*(3), 305-311. https://doi.org/https://doi.org/10.1016/S1385-1101(98)00003-3

Sommer, U., & Lewandowskal, A. (2011). Climate change and the phytoplankton spring bloom: warming and overwintering zooplankton have similar effects on phytoplankton. *Global Change Biology*, *17*(1), 154-162. https://doi.org/https://doi.org/10.1111/j.1365-2486.2010.02182.x

Stadniczeñko, S. G. (2019). Diversity and community dynamics of protists and their viruses in the Skagerrak.

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in ecology & evolution*, *24*(2), 110-117. https://doi.org/https://doi.org/10.1016/j.tree.2008.09.011

Vaulot, D., Geisen, S., Mahé, F., & Bass, D. (2022). pr2-primers: An 18S rRNA primer database for protists. *Molecular Ecology Resources*, *22*(1), 168-179. https://doi.org/https://doi.org/10.1111/1755-0998.13465

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1847-1857.

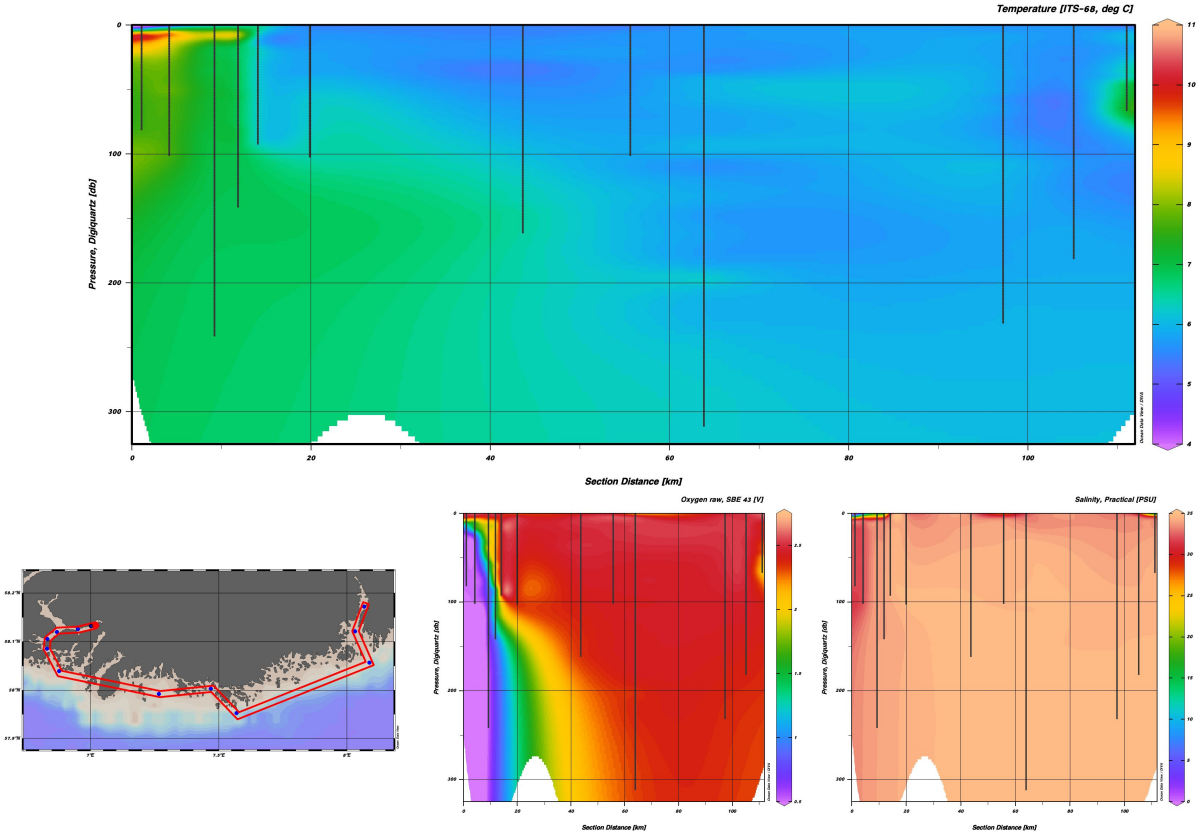
Winder, M., & Sommer, U. (2012). Phytoplankton response to a changing climate. *Hydrobiologia*, *698*. https://doi.org/10.1007/s10750-012-1149-2

Young, A., & Gillung, J. (2019). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, *45*. https://doi.org/10.1111/syen.12406

Appendix

1.ODV full range figures

Appendix figure 1: CTD shown in full depth in ODV.



2. R- SOP

> install.packages("devtools")

> if (!requireNamespace("BiocManager", quietly=TRUE))
  install.packages("BiocManager")

> BiocManager::install("dada2")

> BiocManager::install("Rcpp")

> library("dada2")

>library("phyloseq")

>library("Biostrings")

>library("ggplot2")

>library("dplyr")

>library("tidyr")

>library("tibble")

```
>library("readxl")
>library("readr")
>library("stringr")
>library("kableExtra")
>library("tidyverse")
>library("plyr")
>library("R.utils")

>.zipf <- list.files()
>ldply(.data = zipf, .fun=gunzip)
>fns <- sort(list.files(fastq_dir, full.names = T))

>fns <- fns[str_detect(basename(fns), ".fastq")]
>fns_R1 <- fns[str_detect(basename(fns), "R1")]
>fns_R2 <- fns[str_detect(basename(fns), "R2")]

>sample.names <- str_split(basename(fns_R1), pattern = "_", simplify = TRUE)
>sample.names <- sample.names[, 1]

>fastq_dir <- "fastq_unzipped"
>database_dir <- "databases/"
https://github.com/vaulot/metabarcodes_tutorials/tree/master/databases
>filtered_dir <- "fastq_filtered/"
>qual_dir <- "qual_pdf/"  # quality scores plots
>dada2_dir <- "dada2_results/"
>blast_dir <- "blast/"

>dir.create(filtered_dir)
>dir.create(qual_dir)
>dir.create((dada2_dir))
>dir.create(blast_dir)

>filt_R1 <- str_c(filtered_dir, sample.names, "_R1_filt.fastq")
```

```
>filt_R2 <- str_c(filtered_dir, sample.names, "_R2_filt.fastq")

>out <- filterAndTrim(fns_R1, filt_R1, fns_R2, filt_R2, truncLen = c(240, 210), trimLeft =
c(20, 21), maxN = 0, maxEE = c(2, 2), truncQ = 2, rm.phix = TRUE, compress = FALSE,
multithread = TRUE)

>err_R1 <- learnErrors(filt_R1, multithread = TRUE)
>plotErrors(err_R1, nominalQ = TRUE)

>err_R2 <- learnErrors(filt_R2, multithread = T)
plotErrors(err_R2, nominalQ = TRUE)

>derep_R1 <- derepFastq(filt_R1, verbose = FALSE)
>derep_R2 <- derepFastq(filt_R2, verbose = FALSE)

>names(derep_R1) <- sample.names
>names(derep_R2) <- sample.names

>dada_R1 <- dada(derep_R1, err = err_R1, multithread = TRUE, pool = FALSE)
>dada_R2 <- dada(derep_R2, err = err_R2, multithread = TRUE, pool = FALSE)

>dada_R1[[1]]
>dada_R2[[1]]

>mergers <- mergePairs(dada_R1, derep_R1, dada_R2, derep_R2, verbose = TRUE)
head(mergers[[1]])

>seqtab <- makeSequenceTable(mergers)

>seqtabmerged <- mergeSequenceTables(repeats="sum", seqtab1, seqtab2)

> dim(seqtabmerged)
> dim(seqtab1)
```

```
> dim(seqtab2)

>t_seqtabmerged <- t(seqtabmerged) # the function t() is a simple transposing of the matrix
table(nchar(getSequences(seqtabmerged)))

>plot(table(nchar(getSequences(seqtabmerged)))) #simple plot of length distribution
>hist(nchar(getSequences(seqtabmerged)), main="Distribution of sequence lengths")

>st1 <- readRDS
>st2 <- readRDS
>st.all <- mergeSequenceTables(st1, st2)

>seqtab.nochim <- removeBimeraDenovo(seqtabmerged, method = "consensus", multithread = FALSE, verbose = TRUE)
>saveRDS(seqtab.nochim)

>paste0("% of non chimeras : ", sum(seqtab.nochim)/sum(seqtab) * 100)
>paste0("total number of sequences : ", sum(seqtab.nochim))

>paste0("% of non chimeras : ", sum(seqtab.nochim)/sum(seqtab) * 100)
> paste0("total number of sequences : ", sum(seqtab.nochim))

>getN <- function(x) sum(getUniques(x)) # example of a function in R

>track <- cbind(out, sapply(dada_R1, getN), sapply(mergers, getN), rowSums(seqtabmerged), rowSums(seqtab.nochim))

>colnames(track) <- c("input", "filtered", "denoised", "merged", "tabled", "nonchim")
>rownames(track) <- sample.names

>write_tsv(data.frame(track), str_c(dada2_dir, "read_numbers_dada2.tsv"))
```

```
>seqtab.nochim_trans <- as.data.frame(t(seqtab.nochim)) %>%
rownames_to_column(var = "sequence") %>% rowid_to_column(var = "OTUNumber")
%>% mutate(OTUNumber = sprintf("OTU_%05d", OTUNumber)) %>%
mutate(sequence = str_replace_all(sequence, "(-|\\.)", ""))

>df <- seqtab.nochim_trans
>seq_out <- Biostrings::DNAStringSet(df$sequence)
>names(seq_out) <- df$OTUNumber
>seq_out

>Biostrings::writeXStringSet(seq_out, str_c(dada2_dir, "ASV_no_taxonomy.fasta"),
compress = FALSE, width = 20000)
>PR2_tax_levels <- c("Kingdom", "Supergroup", "Division", "Class",
                    "Order", "Family",
                    "Genus", "Species")

>pr2_file <- paste0(database_dir, "pr2_version_4.14.0_SSU_dada2.fasta.gz")

>save.image("name.file.RData")
>load("name.file.RData")
>taxa <- assignTaxonomy(seqtab.nochim, refFasta = pr2_file, taxLevels = PR2_tax_levels,
minBoot = 0, outputBootstraps = TRUE, verbose = TRUE, multithread = TRUE)

>saveRDS(taxa, str_c(dada2_dir, "dada2.taxa.reads"))
>taxa <- readRDS(str_c(dada2_dir, "taxa.rds"))

>write_tsv(as_tibble(taxa$tax), file = str_c(dada2_dir, "taxa.txt"))

>taxa_tax <- as.data.frame(taxa$tax)
>taxa_boot <- as.data.frame(taxa$boot) %>% rename_all(funs(str_c(., "_boot")))
>seqtab.nochim_trans <- taxa_tax %>% bind_cols(taxa_boot) %>%
>bind_cols(seqtab.nochim_trans)
```

```
>unique(seqtab.nochim_trans$Kingdom)
>unique(seqtab.nochim_trans$Supergroup)

>bootstrap_min <- 80

>seqtab.nochim_18S <- seqtab.nochim_trans %>% dplyr::filter(Supergroup_boot >=
bootstrap_min)
>seqtab.nochim_18S <-
seqtab.nochim_trans[which(seqtab.nochim_trans$Supergroup_boot>80),]

>unique(seqtab.nochim_18S$Division)

>seqtab.nochim_18S_noMetazoa <-
seqtab.nochim_18S[which(seqtab.nochim_18S$Division!="Metazoa"),]
>seqtab.nochim_18S_lowsupport<- seqtab.nochim_trans %>%
>dplyr::filter(Supergroup_boot <= bootstrap_min)



>write_tsv(seqtab.nochim_18S, str_c(dada2_dir, "OTU_table.tsv"))
>write_tsv(seqtab.nochim_18S_noMetazoa, str_c(dada2_dir,
"OTU_table_noMetazoa.tsv"))
>write_tsv(seqtab.nochim_18S_lowsupport, str_c(dada2_dir,
"OTU_table_lowsupport.tsv"))

>df <- seqtab.nochim_trans
>seq_out <- Biostrings::DNAStringSet(df$sequence)
>names(seq_out) <- str_c(df$OTUNumber, df$Supergroup, df$Division, df$Class,
df$Order, df$Family, df$Genus, df$Species, df$Species_boot1, sep = "|")
>Biostrings::writeXStringSet(seq_out, str_c(blast_dir, "OTU.fasta"), compress = FALSE,
width = 20000)

> samples_df <- samples_df %>%
+ tibble::column_to_rownames("sample-id")
```

```
> samples = sample_data(samples_df)
> ps6 <- phyloseq(samples)

> ps9 <- merge_phyloseq(ps, ps6)

> ps9 <- subset_taxa(ps9, ! supergroup %in% c("Bacteria_X")) #Fjerner bakterier fra
datasettet

> ps9_Layer <- merge_samples(ps9, "Layer")
Warning messages:
1: In asMethod(object) : NAs introduced by coercion
2: In asMethod(object) : NAs introduced by coercion
3: In asMethod(object) : NAs introduced by coercion
4: In asMethod(object) : NAs introduced by coercion

> plot_bar(ps9_Layer, fill = "supergroup") +
+ geom_bar(aes(color=supergroup, fill=supergroup), stat="identity", position="stack")



> ps9
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 2980 taxa and 45 samples ]
sample_data() Sample Data:       [ 45 samples by 11 sample variables ]
tax_table()   Taxonomy Table:    [ 2980 taxa by 9 taxonomic ranks ]

> rank_names()
[1] "root"       "kingdom"    "supergroup" "division"   "class"
[6] "order"      "family"     "genus"      "species"
> plot_bar(ps9, fill="supergroup") + geom_bar(aes(color=supergroup, fill=supergroup),
stat="identity", position="stack")

> total=median(sample_sums(ps9))
```

```
> standf=function(x, t=total) round(t * (x/sum(x)))
> ps9=transform_sample_counts(ps9, standf)

> plot_bar(ps9_Layer, fill="division") + geom_bar(aes(color=division, fill=division),
stat="identity", position="stack")

> plot_bar(ps9, fill="division") + geom_bar(aes(color=division, fill=division),
stat="identity", position="stack")

> plot_richness(measures = c("Chao1", "Shannon"))

> plot_richness(measures=c("Chao1", "Shannon"), x="Layer")

> plot_richness(measures=c("Chao1", "Shannon"), color="Layer")

> plot_richness(measures=c("Chao1"), x="Layer", color="Salinity")

> plot_richness(measures=c("Chao1"), color="Salinity")
```