Master's Thesis

Temporal and Spatial Variation in Microbial Communities of Aquatic Sediments in the Baltic Sea

Aleksi Kolehmainen

University of Jyväskylä

Department of Biological and Environmental Science

Cellular and Molecular Biology

24.5.2021

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science
Department of Biological and Environmental Science
Cellular and Molecular Biology

Aleksi Kolehmainen Temporal and Spatial Variation in Microbial Communities of Aquatic Sediments

MSc thesis: 49 p., 10 appendices (13 p.)
Supervisors: University lecturer Emily Knott and doctoral student Anna-Lotta Hiillos
Reviewers: Reetta Penttinen, PhD and Prof. Phillip Watts

May 2021

Keywords: Sequencing, Eukaryotes, Prokaryotes, Alpha diversity, 16S rDNA, 18S rDNA

Microbe populations are essential for every ecosystem and have various functions. However, protists are notably less studied compared to prokaryotic microbes. This master's thesis focuses on studying the diversity of microbe populations in the Baltic Sea using next-generation sequencing. The focus is especially on the protist diversity. The aim is to find out whether the salinity gradient and seasonal changes have an effect on the structure of the microbe population. It was hypothesized that the bacterial diversity would remain steady throughout the year regardless of surrounding conditions, while protists would experience more variation. The samples used in the thesis were taken from three different sites with varying salinity in the Baltic Sea. The sampling was repeated four times throughout the year. The samples were sequenced and the results were statistically analyzed with RStudio. Bacteria showed significant temporal variation in diversity, with *Cyanobacteria* and *Proteobacteria* being the most common phyla. Protist populations varied spatially. More variation could be seen on the class level. *Dinoflagellata* were the most abundant phylum in nearly all sites in August 2018, but declined throughout the year, while the abundance of *Ochrophyta* grew. This was most likely due to exceptionally high temperatures during August 2018.

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta
Bio- ja ympäristötieteiden laitos
Solu- ja molekyylibiologia

Aleksi Kolehmainen: Ajallinen ja paikallinen vaihtelu Itämeren pohjasedimenttien mikrobipopulaatioissa
Pro gradu -tutkielma: 49 s., 10 liitettä (13 s.)
Työn ohjaajat: Yliopistonlehtori Emily Knott ja tohtoriopiskelija Anna-Lotta Hiillos
Tarkastajat: Tohtori Reetta Penttinen ja Professori Phillip Watts

Toukokuu 2021

Hakusanat: Sekvensointi, Eukaryootit, Prokaryootit, Alfadiversiteetti, 16S rDNA, 18S rDNA

Mikrobipopulaatiot ovat olennainen osa jokaista ekosysteemiä. Tästä huolimatta alkueliöt ovat kuitenkin huomattavasti vähemmän tutkittuja prokaryootteihin verrattuna. Tämä maisteritutkielma tutkii mikrobipopulaatioiden diversiteettiä Itämeressä uuden sukupolven sekvensointimenetelmillä. Tutkimus keskittyy erityisesti alkueläinpopulaatioihin. Päämääränä on selvittää, kuinka Itämeren suolagradientti ja ajallinen vaihtelu vaikuttavat populaatioiden rakenteisiin. Hypoteesina on, että bakteeripopulaatioiden diversiteetti pysyisi tasaisena ajasta ja paikasta riippumatta, mutta alkueliöiden määrät vaihtelisivat enemmän. Työssä käytetyt näytteet kerättiin kolmelta eri näytteenottopaikalta neljästi vuodessa. Näytteet sekvensoitiin ja tulokset analysoitiin RStudio-ohjelmalla. Bakteereissa havaittiin merkittävää vaihtelua ajallisesti. Alkueliöiden diversiteetti vaihteli paikallisesti. *Dinoflagellata* oli suhteellisesti yleisin pääjakso elokuussa 2018 lähes kaikilla näytteenottopaikoilla, mutta määrä laski vuoden edetessä. *Ochrophyta* puolestaan kasvoi määrällisesti vuoden aikana. Syy tähän ilmiöön oli todennäköisesti vuoden 2018 elokuun korkea lämpötila. Tulokset eivät noudattaneet hypoteesia, jonka mukaan alkueliöt olisivat alttiimpia ympäristötekijöille, sillä bakteereiden diversiteetin vaihtelu ajallisesti oli vastoin odotuksia.

TABLE OF CONTENTS

TERMS AND ABBREVIATIONS

TERMS

16S	A highly conserved ribosomal gene of prokaryotes, often used for species identification. Used in this thesis to also refer to a primer pair amplifying the V1/V2 regions of 16S gene.
18S	A highly conserved ribosomal gene of eukaryotes, often used for species identification.
TarEuk	Primer pair, also used in this thesis to refer to samples amplified with this primer pair. Amplifies V4 region of 18S
UNonMet	Primer pair, also used in this thesis to refer to samples amplified with this primer pair. Amplifies V4 region of 18S.
V9	Primer pair, also used in this thesis to refer to samples amplified with this primer pair. Amplifies V9 region of 18S
alpha diversity	Measure of diversity; measures the number of OTUs
Shannon Index	Diversity index that reflects the abundance and distribution of OTUs

ABBREVIATIONS

NMDS	Non-metric multidimensional scaling
OTU	Operational taxonomic unit

1 INTRODUCTION

Microbes have a large influence on the surrounding ecosystem. Depending on the species, they can have a role in the carbon and oxygen cycles, be a part of the food chain and act as parasites (Caron et al. 2012). Microbes can be considered to include all unicellular, microscopical organisms. These can in turn be divided into prokaryotes and eukaryotes (Campbell 2014).

Prokaryotic microbes consist of Archaea and Bacteria. Prokaryotes are characterized by their lack of nucleus. In bacteria, the genetic material is typically stored in a circular plasmid. The size of the plasmid can vary between different bacterial species, with the average being 80 bp (Shintani et al. 2015). Bacteria also differ structurally from eukaryotic cells by having no membrane bound organelles, lacking mitochondria and endoplasmic reticulum. On the cell surface, bacteria have a cell wall consisting of peptidoglycan and flagella used for movement. In this master's thesis, only bacterial prokaryotes will be studied.

Eukaryotic microbes, on the other hand, make up a morphologically and phylogenetically diverse group including, for example, algae, ciliates and slime molds. The eukaryotic microbes are commonly grouped under the name Protista even though such grouping does not reflect their phylogeny. Typical characteristics of protists are a nucleus and single-celled structure, but due to the large species diversity, common features are nearly impossible to define (Hamilton 2006).

1.1 Documenting microbial diversity

There are various methods to determine microbial diversity, based on either a molecular or morphological approach. Morphological approaches identify the microbes based on their physical characteristics, while molecular approaches are mostly based on PCR applications and sequencing. The application of molecular high-throughput techniques on environmental sciences is commonly known as environmental omics (Martyniuk & Simmons 2016).

Morphological methods are largely based on physical properties. Characteristics used to identify microbes include cell shape and size, flagella and in the case of bacteria, Gram positivity (Sarethy et al. 2013). Many different microbial species can share the same morphological characteristics. As a consequence, the results of morphological analysis might not be as precise compared to molecular methods. The advantages, however, are that they are faster and cheaper.

A limiting factor in studying microbes morphologically is the inability to culture certain microbial species. The majority of microbes cannot be cultured on plates, because they require specific conditions to grow that cannot be replicated with the culturing medium. As an alternative, DNA sequencing techniques provide one way to document diversity of uncultivable species. However, these techniques are still limited by the inability to culture microbes and cannot be used to identify the microbes. The sequences are described as OTUs (operational taxonomic units) and exact identification requires culturing.

Metagenomics is the sequencing of a whole environmental sample and building a metagenome of all the microbes included in the sample (Thomas et al. 2012). This approach can provide valuable information on the ecological dynamics of the system and allows for new useful genetic information of enzymes and proteins to be discovered. The building of a metagenome starts by processing the sample. In the

sample processing, the DNA is extracted from the sample. The extraction method varies by the sample type (e.g. soil, tissue, blood). In some cases, the sample might need to be fractionated before extraction by filtration or centrifugation in order to exclude host DNA in the case of, for example, gut biota (Schmeisser et al. 2007). The utility of metagenomics can range from using just fragments of marker sequence to build an OTU library to assembling the whole genome of a species (Pérez-Cobas et al. 2020).

Following the extraction, the resulting DNA is sequenced. Various sequencing methods exist, with the most commonly used today being the next-generation sequencing methods (NGS), which has become cheaper and more accessible (Gu et al. 2019). The advantage of NGS is its speed, as it is capable of sequencing whole sequences in a small timeframe. NGS technologies vary, but all follow a similar concept (Kchouk et al. 2017). The strands of DNA are first fractioned and amplified multiple times to produce a stronger signal. After this, the method of sequencing varies depending on the technology. The results of the sequencing can be seen immediately after the sequencing procedure. For example, the Ion Torrent sequencing method used in this thesis utilizes chips covered in wells, which contain small beads. Each of these beads are covered with amplified fragments from the sample. When a nucleotide binds to these strands, a hydrogen ion is released. This causes a change in the pH of surrounding solution, which can be detected by the sequencer. Based on these changes, the instrument is capable of deducing the nucleotide sequence of each strand. Other NGS methods include, for example, Illumina and Roche/454 sequencing.

After sequencing, the short-read fragments are either assembled or binned according to the goals of the study. Assembly is utilized when building a whole genome. In amplicon sequencing, the DNA sequences are binned. The sequences are sorted into separated groups representing related organisms or genomes to form OTUs. Both of these processes can be achieved by suitable software. Following this, the data is

annotated and statistically analyzed depending on the aims of the study. (Thomas et al. 2012)

1.2 Functions of bacteria in aquatic systems

Bacteria have various roles in microbial communities. The most important roles in aquatic systems are photosynthesis and nitrification, decomposition of dead matter, being part of the food chain and acting as pathogens. In the nitrogen cycle, aquatic bacteria produce ammonium from nitric animal wastes by a process of ammonification. The ammonium is subsequently used by other bacteria in their metabolism, which produces nitrites and nitrates. The organic nitrogen in nitrates is freed by denitrifying bacteria. (Campbell 2014)

Photosynthesis has only been documented in the bacteria in the phylum Cyanobacteria. Cyanobacteria have a significant role in marine ecosystems, as they produce oxygen together with aquatic plants and algae. However, an influx of nitrogen and phosphorus in the ecosystem could cause an increase in the cyanobacteria population sizes. As several cyanobacteria species, such as the members of the genus *Microcystis* (Bláha et al. 2009), are highly poisonous to both humans and the environment, they pose a serious ecological threat. Baltic Sea, for example, has suffered greatly from cyanobacteria blooms due to nutrients released by agriculture (WWF 2020). Eutrophication is also partly caused by the benthic decomposer bacteria, as their populations increase along with the increase of dead organic matter in water. Increased activity of decomposers leads to depletion of oxygen. This in turn increases the populations of anaerobic bacteria that release more organic compounds, resulting in a self-feeding cycle that can lead to a death of large areas in the body of water (Sinkko et al. 2013).

However, the most widely known function of bacteria is their nature as pathogens. Pathogenic bacteria can have a significant effect on the population dynamics of other species in aquatic environments. In aquatic systems, bacteria can use a variety of eukaryotes as their hosts. For example, *Pasteuria ramosa* uses *Daphnia magna* water fleas as their host species (Frost et al. 2008). The effects of the infection include gigantism, sterilization and eventual death of the host. Another example is *Renibacterium salmoninarum,* which infects fish in the family Salmonidae (European commission 1999). The fish infected with the bacterium contract bacterial kidney disease, which often leads to death. The disease causes lesions to kidney tissue and, in later stages, other organs. Similar to *Renibacterium salmoninarum, Aeromonas salmonica* infects salmonids, among other fish species and has been documented in both seawater and fresh water (Austin 2005). It spreads from fish to fish by either sea lice or other fish. The pathogen is highly virulent and often leads to death.

1.3 Functions of protists in aquatic systems

Protists are also important for the function of the ecosystem, but they have been studied relatively little compared to bacteria (Keeling & del Campo 2017). As such, it is important to note that a large number of protist species are only characterized either morphologically or by their phylogenetic relationships to other more well-known species. Due to this, their actual functions are not well known. Protists are taxonomically divided into four groups: *Excavata, Achaeplastida, Unikonta* and the SAR group (Campbell 2014). The name of SAR group is derived from the clades that comprise it (Stramenopiles, Alveolata, Rizaria). Each of these groups except for *Unikonta* include species capable of photosynthesis. According to Thomas et al. (2012), protists produce no less than half of the world's oxygen. This makes them an important research subject from the perspective of nature preservation. Compared to bacteria, protists function in largely similar roles, such as by acting as decomposers. For

example, the members of the class *Labyrinthulomycetes* often act as decomposers in exclusively marine environments, where they have been observed to feed on dead plants (Raghukumar 2002)

Some protists function as parasites or symbionts in a variety of host species. A good example is *Ichthyophthirius multifiliis,* belonging to SAR group, which is a common parasite that uses fishes as their hosts and causes white spot disease (Xu and Klesius 2004). It can have serious economic effects on fisheries. Parasitism can impact the host's role in the food chain and also affect the population dynamics of the host species. For example, Kohler and Hoiland (2001), found that the population size of *Brachycentrus americanus* caddisfly was strongly driven by a protist-caused disease. Another example of a parasitic protist also from the SAR group, *Perkinsus marinus,* infects several species of oysters and causes death in 1 to 2 years (OIE 2019). It is especially prevalent in eastern oyster (*Crassostrea virginica)* and has three life stages, which are all dependent on oysters (Audemard 2004).

Of special interest from the perspective of this thesis is the phylum Apicomplexa. Apicomplexa are a part of the SAR group of protists and all members are obligate symbiotes, displaying either mutualistic or parasitic lifestyle. Several species, for example the members of genus *Toxoplasma*, also infect humans and cause severe symptoms (Rueckert et al. 2019). Apicomplexa have been documented in marine and terrestrial environments. Marine Apicomplexa species also include parasitic species. A good example is the order Archigregarinorida, in the class Conoidasida. Several of these species are marine parasites that infect invertebrates (Leander 2007). For example, *Selenidium vivax* lives in the intestinal tract of peanut worm *Phascolosoma agassizii.* However, it is still unknown whether all the species Archigregarinoridans are in fact parasitic. Symbiosis exists on a spectrum that ranges from free living organisms to parasites (Ruckert et al. 2019).

Protists also include various mutualistic species in addition to parasitic species. Mutualism offers a beneficial relationship to both parts of the relationship. Effects of mutualism can include, for example, supply of nutrients from both interacting species or protection of hosts against predators while simultaneously offering grazing protection to the mutualist. Several members of the aforementioned phylum Apicomplexa are mutualistic. In addition, ciliates, common members of eukaryotic microbial communities, are known to be often mutualistic. (Dziallas et al. 2012)

1.4 Microbes in the Baltic Sea

The Baltic Sea is a special ecosystem in many regards. It is a young body of water that was formed during the last ice age. It is also heavily affected by the surrounding freshwater systems with a constant runoff of fresh water that has created a salinity gradient. However, the freshwater runoff has also brought pollution and excess nutrients to the sea. As the Baltic Sea is one of the most polluted seas in the world, its natural population dynamics have shifted over time. Increased eutrophication has led to various 'dead zones' with lowered oxygen concentrations that are nearly uninhabitable, except for cyanobacteria and various other microbes. Generally, the species diversity of both animals and plants in the Baltic Sea is low (HELCOM 2018). These factors make the Baltic Sea an interesting setting for studying microbial diversity.

According to Klier et al. (2018), the bacterial species diversity is strongly affected by salinity. Baltic Sea is brackish, with the salinity being much lower compared to seawater. In addition, the influx of freshwater flowing into the sea from rivers creates a salinity gradient, where the salinity rises going southwest (Sjöqvist et al. 2015). The brackish nature of the Baltic Sea could indeed have a restrictive effect on species diversity, as the low salinity is not suitable for many marine species. However, several other factors also had an effect on the community composition, such as temperature and nutrients (Klier et al. 2018). In their study, the species composition was found to

be the highest in mesohaline conditions, where the species composition was a mix of both freshwater and marine species (Klier et al. 2018).

According to a study by Edlund (2007), the population structure of prokaryotes in the Baltic Sea is connected to the surrounding conditions. The largest factor affecting the population sizes of archaea was the oxygen concentration of water, while the bacteria were most affected by the depth. Other contributing factors were salinity, organic carbon and silicate. Ininbergs and colleagues (2015) found that the bacterial diversity of the Baltic Sea is large throughout, despite the salinity gradient, due to the fast adaptability of the bacteria. This, coupled with challenging conditions, leads to a large diversity. Edlund and colleagues (2005) earlier found that the diversity of archaea and bacteria are strongly affected by environmental factors and the community compositions to be different in sites of pollution. This implies that human activity has had a clear impact on the microbial composition of environment. This in turn causes changes in the cycle of nutrients.

There is also variation in microbial communities of aquatic environments according to seasonal changes. In studies done in Skagerrak (North Sea) (Gran-Stadniczeñko et al. 2019), there was seasonal variation in the diversity of protists. According to the study, the maximal variation in the protist diversity was during late summer. Different protist groups were also more abundant during different seasons, such as diatoms being most abundant during spring. Despite previous studies, more research is needed to better understand the dynamics of microbial populations.

1.5 Aims of the study

This master's thesis documents changes in species diversity in bacterial and protist communities in sediments from three coastal sites of the Baltic Sea. In particular, the aim was to find out how different microbial groups are associated in the microbial

communities and if their diversity changes depending on location or season. Based on previous studies, I hypothesized that diversity in bacterial communities would vary little, but diversity in protist communities would display more variation both temporally and spatially. Diversity of the different microbial groups would also be related to environmental factors (temperature and salinity). Focusing specifically on parasitic protists in the Phylum Apicomplexa, the microbial diversity was related to the species diversity of the benthic invertebrates sampled from the same areas. Since Apicomplexa species lay oocysts in the sediments and infect host invertebrates by being consumed, detection of Apicomplexa from sediment samples should be related to infection in the invertebrate populations.

2 MATERIALS AND METHODS

2.1 Sample collection

The sediment samples were collected from three different sites: Herslev (Denmark), Saltö (Sweden) and Öland (Sweden) (Figure 1). From these sites, two replicate sediment samples were collected at four different time points (8/2018, 11/2018, 4/2019, 8/2019) to include temporal variation. In total, there were 12 samples including the replicates. The samples were preserved in 95% EtOH at the site and transported to University of Jyväskylä.

Figure 1. The sites from which the samples were collected. 1: Saltö, Sweden; 2: Herslev, Denmark; 3: Öland, Sweden

2.2 DNA extraction and sample preparation

Approximately 250 µg of sediment was taken from each sediment sample for the extraction. Before extraction, the samples were prepared by adding 750 µl of PowerBead solution (from DNeasy Powerlyzer PowerSoil Kit) into the samples and lysing the cells with Bead Ruptor Elite (OMNI International) at 2500 rpm for 45 s. DNA was extracted from the samples using DNeasy Powerlyzer Powersoil kit (Qiagen). The final concentration of extracted DNA in the samples was quantified using a Qubit 4.0 fluorometer with 1X HS DNA reagents.

2.3 Library preparation for amplicon sequencing

The primers used in the amplification targeted the V1/V2 regions of 16S rRNA gene in bacteria and the V4 and V9 areas of 18S rRNA gene in protists. In this way, it would be

possible to quantify the bacterial and protistan abundances in the samples separately. The different primers used in PCR I are described in Table 1.

Table 1. Primer pairs for PCR I amplification. Two different primers were used for the 18S V4 area, with UnonMet primers also requiring another pair for successful amplification (UNonMet(E572) + UNonMet(E1009R)).

		Primer	Sequences	Target size
18S	**V9**	1391F+ EukB	F: 5´-GTACACACCGCCCGTC-3´ R: 5´-TGATCCTTCTGCAGGTTCACCTAC-3´ (Stoeck et al. 2010)	200 bp
	V4	UNonMet(E572) + UNonMet(E1009R)	F: 5´-CYGCGGTAATTCCAGCTC-3´ R: 5´-CRAAGAYGATYAGATACCRT-3´ (Comeau et al. 2011)	400 bp
		UNonMetF+ UNonMetR	F: 5´-GTGCCAGCAGCCGCG-3´ R: 5´-TTTAAGTTTCAGCCTTGCG-3´ (Bower et al. 2004)	600 bp
		TarEuk454F+ TarEukR	F: 5´-CCAGCASCYGCGGTAATTCC-3´ R: 5´-ACTTTCGTTCTTGATYRA-3´ (Stoeck et al. 2010)	270 bp
16S		27F+338R	F: 5'-AGAGTTTGATCMTGGCTCAG-3' R: 5'-ATTACCGCGGCTGCTGG-3' (Nakatsu & Marsh 2007)	311 bp

Target genes from each sample were amplified using PCR. The PCR amplification step will be referred to as PCR I. For practical reasons, the primer pairs will henceforth be referred as V9, UNonMet, TarEuk and 16S. The amplification conditions for each primer pair were as described in Table 2.

Table 2. The protocols of 16S, TarEuk, UNonMet (both amplifications) and V9 primers for PCRI.

16S	TarEuk	UNonMet	UNonMet (E572+E1009R)	V9
1. 94°C, 2 min 2. 94°C, 30 s 3. 58 °C, 30 s } x35 4. 72°C, 30 s	1. 98°C, 3 min 2. 98°C, 15 s 3. 50 °C, 30 s } x36 4. 72°C, 30 s	1. 94°C, 2 min 2. 94°C, 10 s 3. 51,1 °C, 30 s } x35 4. 72°C, 1 min	1. 94°C, 2 min 2. 94°C, 30 s 3. 55 °C, 30 s } x31 4. 72°C, 1 min	1. 94°C, 2 min 2. 94°C, 30 s 3. 58 °C, 30 s } x35 4. 72°C, 30 s

5. 72°C, 5 min	5. 72°C, 10 min	5. 72°C, 5 min	5. 72°C, 5 min	5. 72°C, 5 min

Reactions were performed in 25 µl volumes containing 12.5 µl of either iQ™ SYBR® Green Supermix (2X) (for UNonMet and 1391F+ EukB) (Bio-Rad) or SsoAdvanced™ Universal SYBR® Green Supermix (2X) (for TarEuk) (Bio-Rad), 200 nM of each primer, 7.5-9.5 µl of nuclease-free water and 3 ng of template DNA. After the PCR, the amplicon size was checked on agarose gel. The UNonMet primer pair amplified a non-specific 450 bp long product. To remove it from the samples, the products from UNonMet amplification were used as templates in another PCR with E572 and E1009R primer pair following the reaction recipe described above. The results of amplification were verified to be successful using agarose gel electrophoresis. The gel concentration was 1% agarose in TA buffer with 2 µl of SYBR Safe (Thermo Fischer Scientific) as a staining reagent. The gels were visualized under an UV light.

Sequencing adapters were added to the amplified sequences in PCR II. The adapters consisted of a barcoded fusion primer IonA with an M13 linker, a target specific P1 fusion primer and a target-specific M13 linker primer. The function of these primers is illustrated in Figure 2. The barcodes were numbered from 1 to 96. For the sequences of fusion primers, refer to Appendix 1.

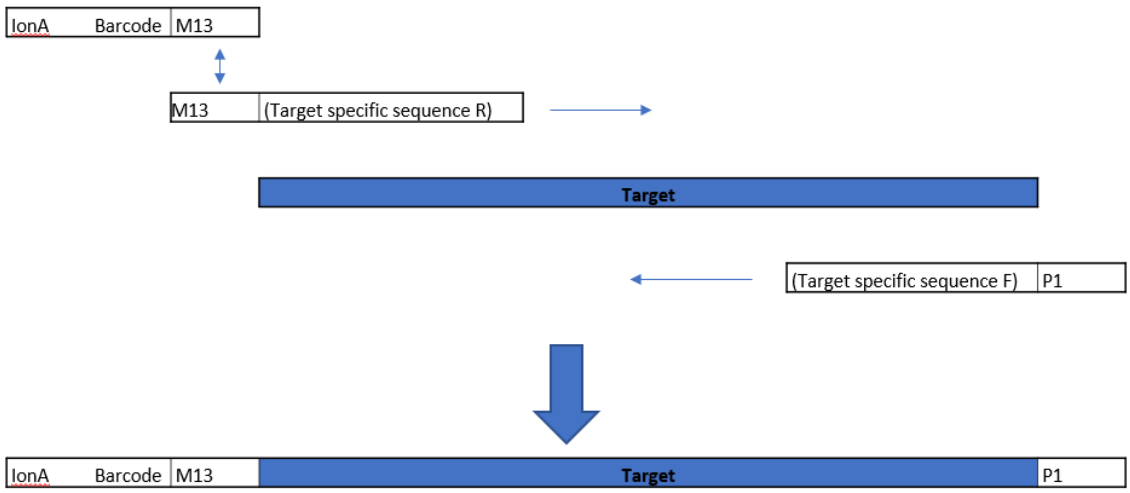


Figure 2. The function of primers in PCR II. P1 and M13 primers are paired with the target using target specific sequences. M13 then acts as a linker to link the IonA primers with barcode, which have a complementary M13 sequence. The resulting product is then ready for subsequent steps.

Once the primers were prepared, they were used in PCRII. The protocol is described in Table 3. All samples were amplified simultaneously on the same plate.

Table 3. PCR II protocol

Primer pairs	16S	TarEuk	UNonMet	V9
Program	1. 95°C, 5 min 2. 94°C, 45 s 3. 53°C, 1 min 4. 72°C, 1 min 5. Repeat 1-4 x13 6. 72°C, 5 min			
Reagents (per sample)	12,5 µl iQ SYBRgreen gPCR Super Mix (2X) 8,5 µl H_2O 1 µl IonA/barcode fusion primer (separate for each sample) 1 µl P1 reverse fusion primer (10 µM) 1 µl M13 linker primer (1 µM)			

Excess reagents and amplicons shorter than 100 bp were purified from the samples with Quanta SparQ PureMag Beads system (Quantabio). The purified samples were quantified again with Qubit in order to determine the volume needed for pooling. However, some amplifications were too low in concentration to be used. These amplifications were repeated, most of them successfully, and included in the analysis. In the end, three amplifications were left out of the results (see Appendix 2).

2.4 Sample pooling and Ion Torrent run

The successful amplification products were pooled together. 16S and TarEuk samples were erroneously pooled together with each other, while UNonMet and V9 were later pooled separately. 16S and TarEuk pools included 10 ng of each product, while UNonMet and V9 pools included only 5 ng of each product due to low concentrations. The pool was again purified using SparQ PureMag Beads system and the molarity was measured using the Agilent 2200 Tapestation. Based on the region molarity, the pools were combined in equimolar concentration of 22 pM to a volume of 25 µl. The pool was then sequenced.

Emulsion PCR to attach the library fragments to the sequencing beads (Ion Sphere Particles) was performed with Ion OneTouch 2 System using Ion PGM Hi-Q OT2 Kit (Life Technologies) After this, the beads were washed, quality controlled with Qubit 2.0 and enriched. The enriched library was sequenced by laboratory technician E. Virtanen with Ion PGM system using an Ion 318 Chip Kit version 2 (Ion Torrent, Life Technologies).

2.4 Bioinformatics

Single-end raw reads were demultiplexed by barcode prior to exporting from the Ion Torrent Suite™ Software. They were then trimmed to remove primers using Cutadapt

(version 1.18, Martin 2011). Length filtering was performed with Cutadapt as follows: 16S: min. 150bp, max. 400bp; V4 (TarEuk and UNonMet): min. 25bp max. 400bp; V9: min. 25bp max. 200bp. Quality filtering was performed using FASTX toolkit (Hannon 2010), with a minimum quality score of 20 for 80% of the bases in each read. Dereplication, singleton and chimera filtering were performed using VSEARCH (version 2.15.1, Rognes et al. 2016). All datasets were aligned to the SILVA database (version 138, Yilmaz et al. 2014) using Mothur (version 1.44.3, Schloss et al. 2009) before clustering using to 97% similarity threshold in VSEARCH. Amplicon sequence variants (ASVs) were mapped using VSEARCH, at a similarity of 95-97.5%. 16S datasets were classified to the SILVA SEED database (version 138, Yilmaz et al. 2014) using the Wang method (Wang et al. 2007) with kmer = 8 and 80% similarity in Mothur. V4 and V9 datasets were classified to the PR2 database (version 4.12.0, Guillou et al. 2013; del Campo et al. 2018) using the BLAST method with 80% similarity in Mothur. The 16S dataset generated 21,7% unclassified OTUs, UNonMet V4 dataset generated 24,8% unclassified OTUs, TarEuk V4 dataset generated 56,6% unclassified OTUs, and the V9 dataset generated 62,4% unclassified OTUs at phylum level. Archaeplastida, Metazoa and unknown reads were filtered from the results.

2.4 Statistical analysis

The results were analyzed with RStudio version 1.4.1106 (RStudio PBC) using the following packages: phyloseq **(**McMurdie & Holmes 2019) and vegan (Oksanen et al. 2020).

The data was divided into separate dataframes containing the metadata, the number of OTUs in each sample, taxonomy of the species identified to the phylum level and taxonomy of the species identified to the class level.

Analysis protocol was identical to every sample set (16s, V9, TarEuk and UNonMet). The data was first imported to R Studio and organized. The OTU table, metadata and

phylum-based taxonomy file were sorted and merged together into a phyloseq object. The merging of the data was proven to be successful by checking the contents of the merged phyloseq object.

The number of reads in each sample was checked with command '*sample_sums*'. To be able to compare relative abundance in each site, the samples were rarified to even sequencing depth with command '*rarify_even_depth*' (phyloseq package). The reads were removed using random subsampling. The preceding protocol was repeated with the class-level taxonomical data in eukaryote samples. The effect of rarefication is exemplified in Figure 3.

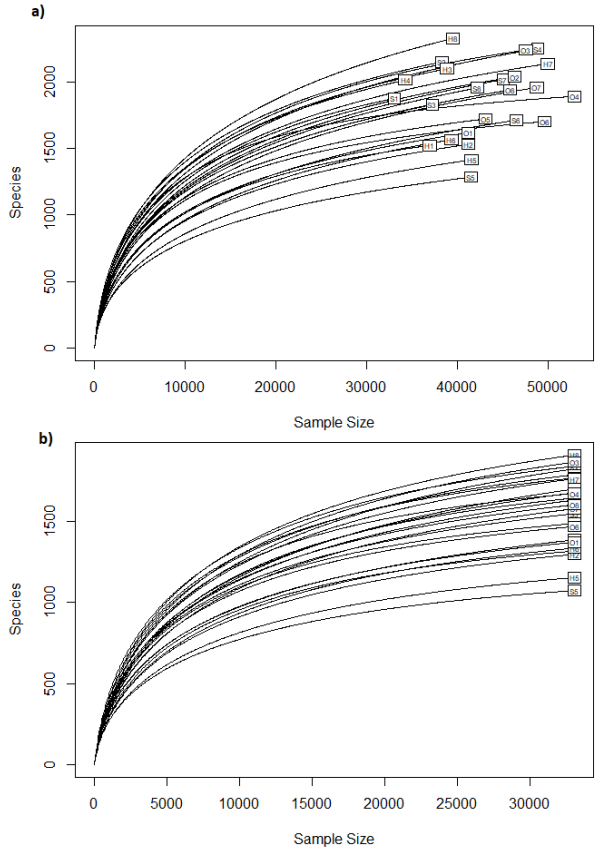


Figure 3. Example of the effect of rarefication using 16s OTUs identified to class level.

Barplots were used to visualize the relative abundance in the samples. First, the data were merged by groups. Due to the nature of the phyloseq package, the variables in the data were relabeled into numbers. New labels matching the original variables were assigned to the data to replace the numbers. Each column had to be done separately. The data were transformed and made into plots. This was repeated for both class and phylum data using both rarefied and unrarefied data. Only rarefied data is shown and discussed, as there was no qualitative difference in interpretation when using the unrarefied data.

The alpha diversity of samples was calculated using Shannon index. The observed alpha diversity (OTU richness) was also included for comparison. Compared to the observed alpha diversity, Shannon index also takes into account the richness and evenness of the OTUs. OTU richness estimates were done using the command '*estimate_richness*' (in phyloseq). The normal distribution of alpha diversity was checked first with Shapiro-Wilk test. If the data followed normality (Appendix 3) , it was further analyzed with ANOVA and Tukey's honest significance test in order to find out which samples had statistically significant differences. If the data were not normally distributed, a Kruskal-Wallis rank sum test was conducted. Beta diversity was quantified with Bray-Curtis dissimilarity index, and visualized with non-metric multidimensional scaling (NMDS). The significance of beta diversity was checked with PERMANOVA analysis using 'adonis' function (in vegan).

3 RESULTS

3.1. Sequencing depth

The depth of sequencing was documented by compiling the numbers of reads and OTUs to a table (Table 4). The number of both OTUs as well as reads was noticeably higher in bacteria compared to eukaryotes.

Table 4. Number of reads and OTUs by different targets.

	16S	TarEuk	UNonMet	V9
Number of reads (phylum)	33 562 – 57 164	3 595 – 18 831	2 970 – 93 064	2 083 – 17 594
Number of reads (class)	33 075 – 52 958	3 574 – 19 663	2 827 – 91 931	1 908 – 11 371
Number of reads (Apicomplexa)	-	22-788	1-2454	1-226
Number of OTUs (phylum)	4829	323	2032	1694
Number of OTUs (class)	4638	301	1757	1 362
Number of OTUs (Apicomplexa)	-	14	42	6

Percentage of unknown OTUs	21,7	56,6	24,8	62,4
Removed OTUs (phylum)	67	3	530	289
Removed OTUs (class)	52	4	459	181

The sequencing depth was also illustrated using species accumulation curves (Figures 4 and 5). The accumulation curves show the relationship between the number of OTUs and reads. The rarefied accumulation curves are included in the appendix (Appendix 4). For the legend on sample numbers, refer to Appendix 5.

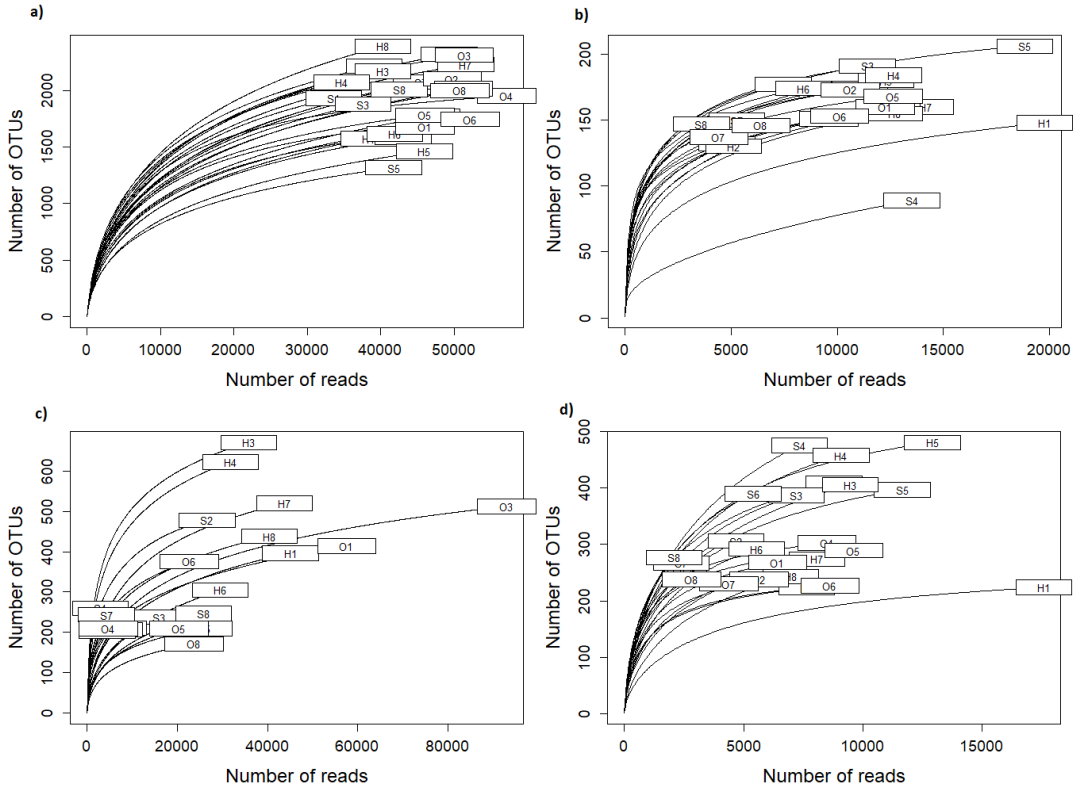


Figure 4. Unrarefied accumulation curves displaying the number of reads against the number of OTUs. OTUs identified to at least phylum level are included, with *Metazoa*, *Archaeplastida* and unclassified results filtered. a: 16S; b: TarEuk; c: UNonMet; d: V9

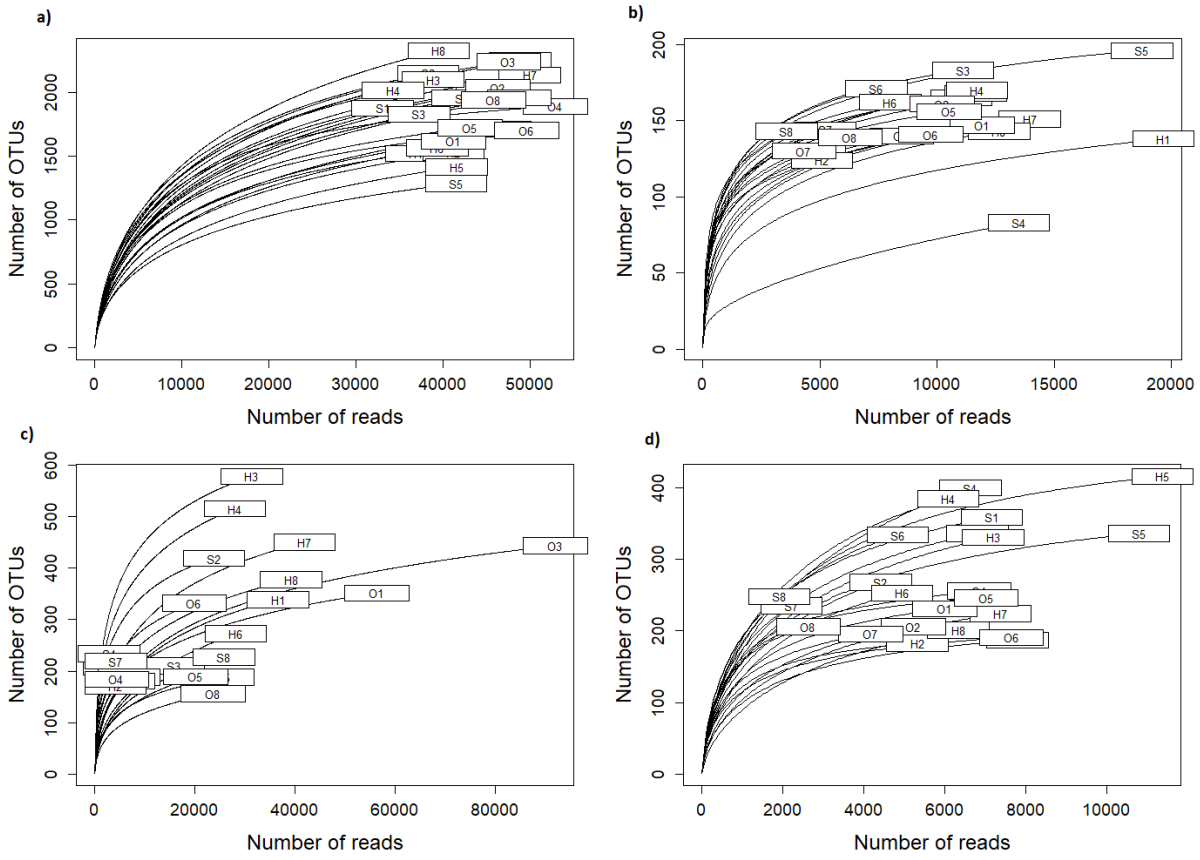


Figure 5. Unrarefied accumulation curves displaying the number of reads against the number of OTUs. OTUs identified to at least class level are included, with *Metazoa*, *Archaeplastida* and unclassified results filtered. a: 16S; b: TarEuk; c: UNonMet; d: V9

3.2. Bacterial communities and diversity

The most common bacterial phyla in all the samples were *Cyanobacteria* and *Proteobacteria.* The relative abundance in each sample is displayed in Figure 6. Unrarefied relative abundance plots of all samples are included in Appendix 6. As bacteria are not the focus of this study, only phylum-level variation is shown.

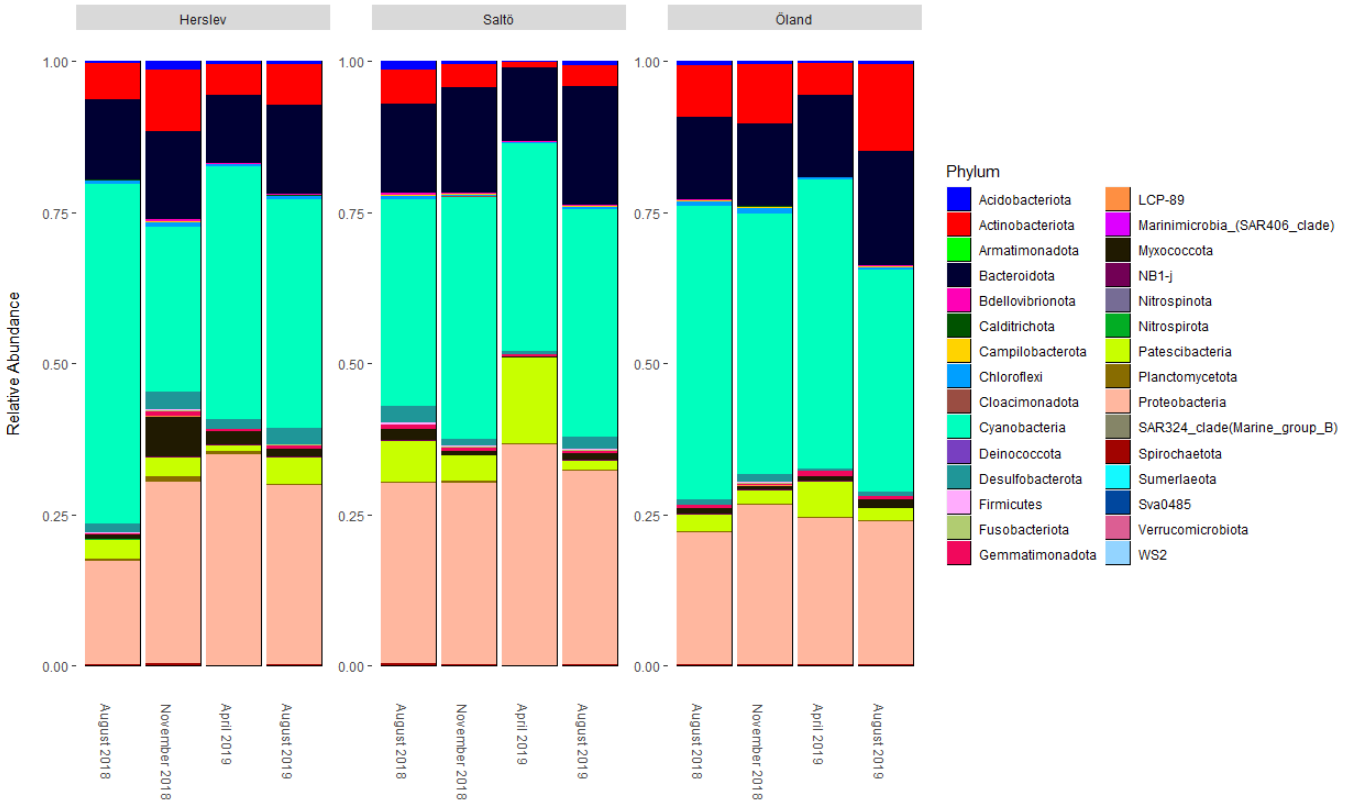


Figure 6. Relative abundance of phyla amplified with 16S primers. The samples have been grouped according to the sites and are shown chronologically.

In Herslev, Shannon's index was lowest in August 2018 (4,714) and highest in November 2018 (5,972). In Öland, Shannon's index was lowest in August 2018 (5,131) and highest in November 2018 (5,686). In Saltö, Shannon's index was lowest in April 2019 (5,018) and highest in August 2018 (6,2).

The lowest Shannon's index in August 2018 was in Herslev (4,713) and the highest was in Saltö (6,2). The lowest Shannon's index in November 2018 was in Öland (5,512) and the highest was in Herslev (5,972). The lowest Shannon's index in April 2019 was in Herslev (4,909) and the highest was in Öland (5,459). The lowest Shannon's index in August 2019 was in Öland (5,229) and the highest was in Herslev (5,876).

According to ANOVA, the alpha diversity differed significantly between time points (F= 9,501, df = 3, p = 0.002) (ANOVA estimates for Shannon index for each target can be found in Appendix 7). Based on Tukey's honest significance test, August -19 and April -19 (p=0,03), November -18 and April -19 (p=0,003) as well as November -18 and August -19 (p=0.009) had significant difference between them. August -18 had 0.384 higher alpha diversity in comparison to April -19, November had 0.545 higher alpha diversity than April -19 and November -18 was 0.468 higher than August -19.

The difference between sampling sites was also significant (F = 3.985, df = 2, p = 0,05), with Saltö having 0.268 higher alpha diversity than Herslev according to Tukey's test (p= 0,05). The interaction between time and site was also significant (F = 4.955, df = 6, p = 0,009), though only two pairs of samples (Herslev: November 2018-Herslev: April 2019, p= 0,019 and Herslev: August 2019-Herslev: August 2018, p= 0.024) differed significantly. The alpha diversity is illustrated in Figure 7.

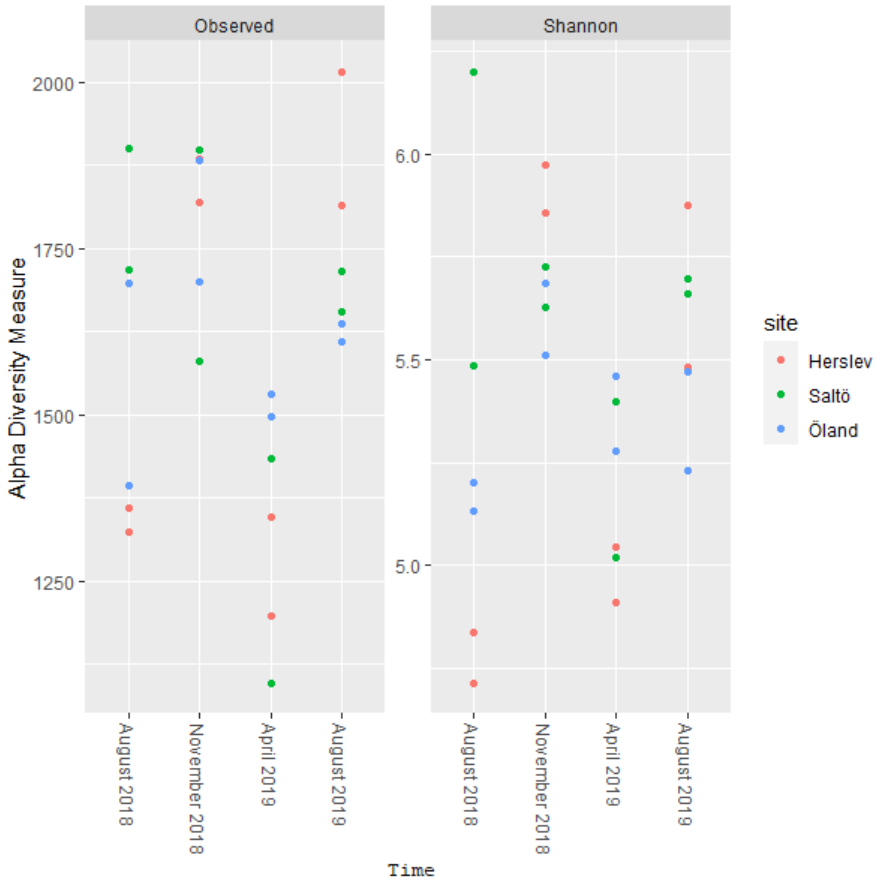


Figure 7. The alpha diversity of 16S samples visualized as a plot of alpha diversity measure against sampling time. The plot on the left describes the observed diversity, while the plot on the left uses Shannon index.

Bacterial beta diversity difference was statistically significant between sites, time points and the interaction between these factors (p=0,001). The difference between samples can also be seen in the 16S beta diversity plot (Figure 8). For permanova estimates for Bray-Curtis dissimilarity index with site, time point and their interaction for each target, see Appendix 8.

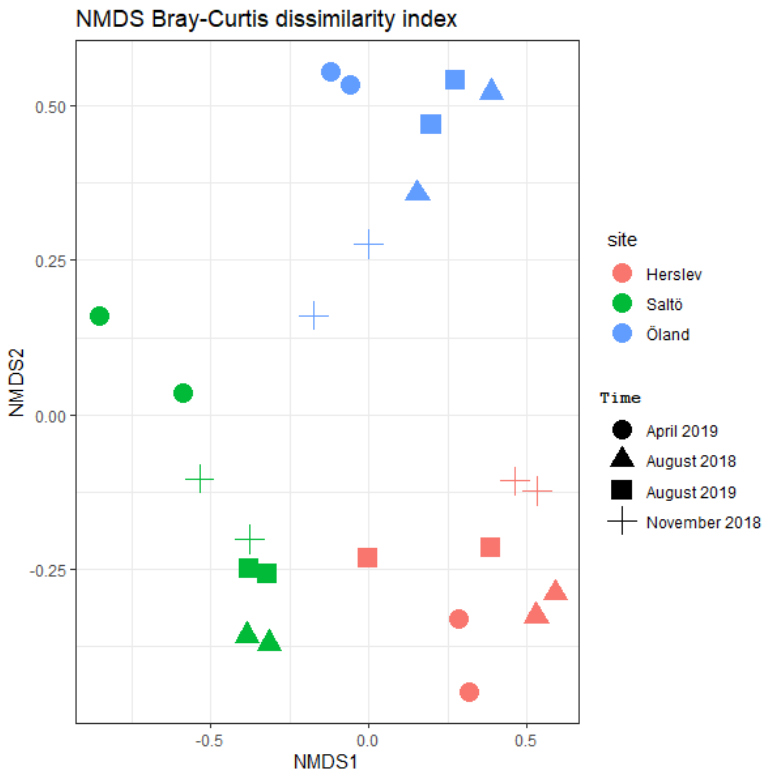


Figure 8. Non-metric multidimensional scaling plot with Bray-Curtis dissimilarity index for 16S samples. The different sampling sites are distinguished by color, while the shape indicates sampling time.

3.3 Eukaryotic microbial diversity

The most abundant phyla in TarEuk samples were Ochrophyta (31,2 %, 101 OTUs), Ciliophora (34,6%, 112 OTUs) and Dinoflagellata (11,4%, 37 OTUs) (Figure 9). These three phyla are present in all timepoints and sites. *Dinoflagellata* are notably common in Herslev in August 2018. The abundance of *Ochrophyta* steadily increases between the four timepoints in Herslev, while the population of *Dinoflagellata* can be seen simultaneously declining. In Saltö and Öland, *Ciliophora* are the most abundant during August 2018.

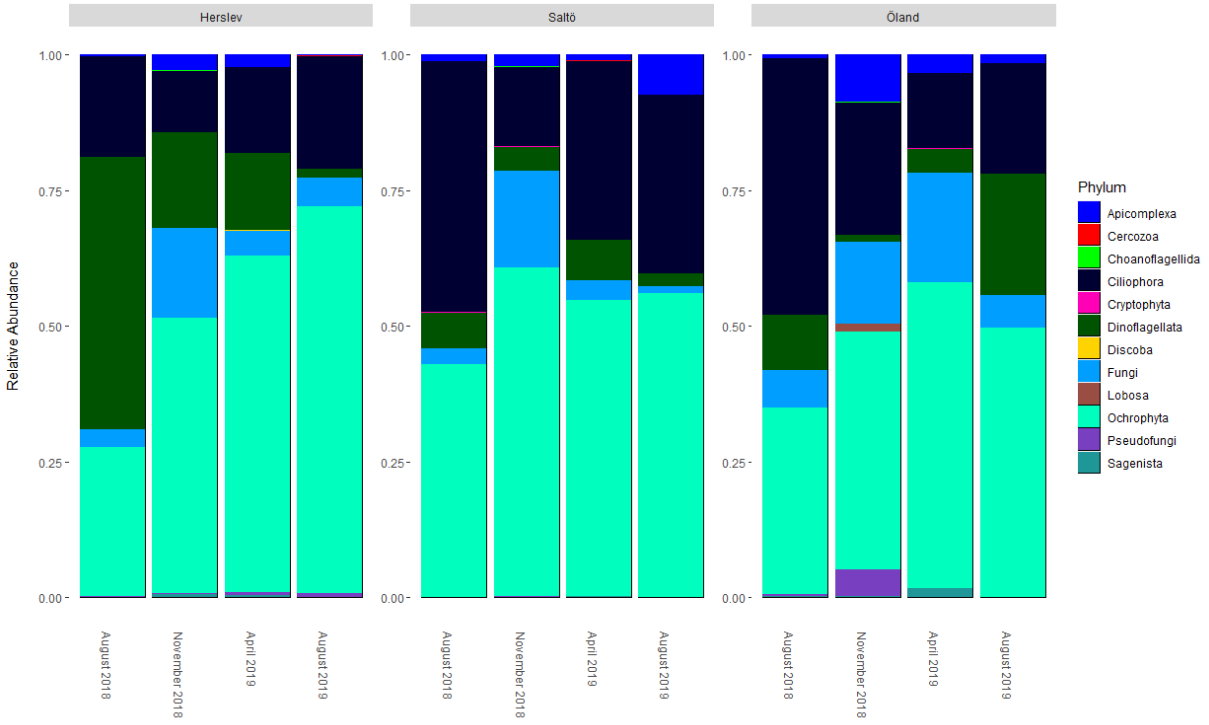


Figure 9. Relative abundance of phyla amplified with TarEuk primers

Bacillariophyta (28,5 %, 86 OTUs) was the most abundant class in nearly all samples. However, in August 2018 in Herslev, *Dinophyceae* was notably abundant. The abundance of *Dinophyceae* did not reach the same numbers in the following year and is seen declining steadily. The relative abundance is shown in Figure 10. Unrarefied relative abundance plots for classes are included in Appendix 9.

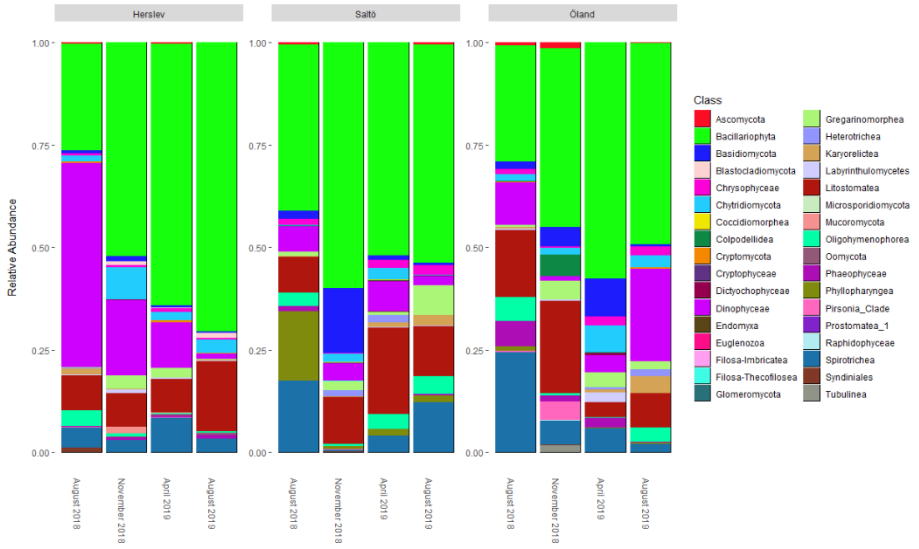


Figure 10. Relative abundance of classes amplified with TarEuk primers

In Herslev, Shannon's index was lowest in August 2018 (2.181) and highest in April 2019 (3,681). In Öland, Shannon's index was lowest in August 2018 (3,276) and highest in August 2019 (3,921). In Saltö, Shannon's index was lowest in November 2018 (2,270) and highest in August 2019 (4,066).

was in Herslev (2,180) and the highest was in Saltö (3,756). The lowest Shannon's index in November 2018 was in Saltö (2,270) and the highest was in Saltö (3,7). The lowest Shannon's index in April 2019 was in Herslev (3,681) and the highest was in Saltö (3,925). The lowest Shannon's index in August 2019 was in Herslev (2,446) and the highest was in Saltö (4,066).

According to the Shapiro-Wilk test, alpha diversity calculated from TarEuk data did not follow normality. Because of this, a Kruskal-Wallis test was utilized. Time did not have a statistically significant effect on the diversity (p=0,09), but the diversity between sites was found to be significantly different (p=0,040). The alpha diversity is shown as a plot in Figure 11, which includes the observed alpha diversity and the alpha diversity with Shannon index.

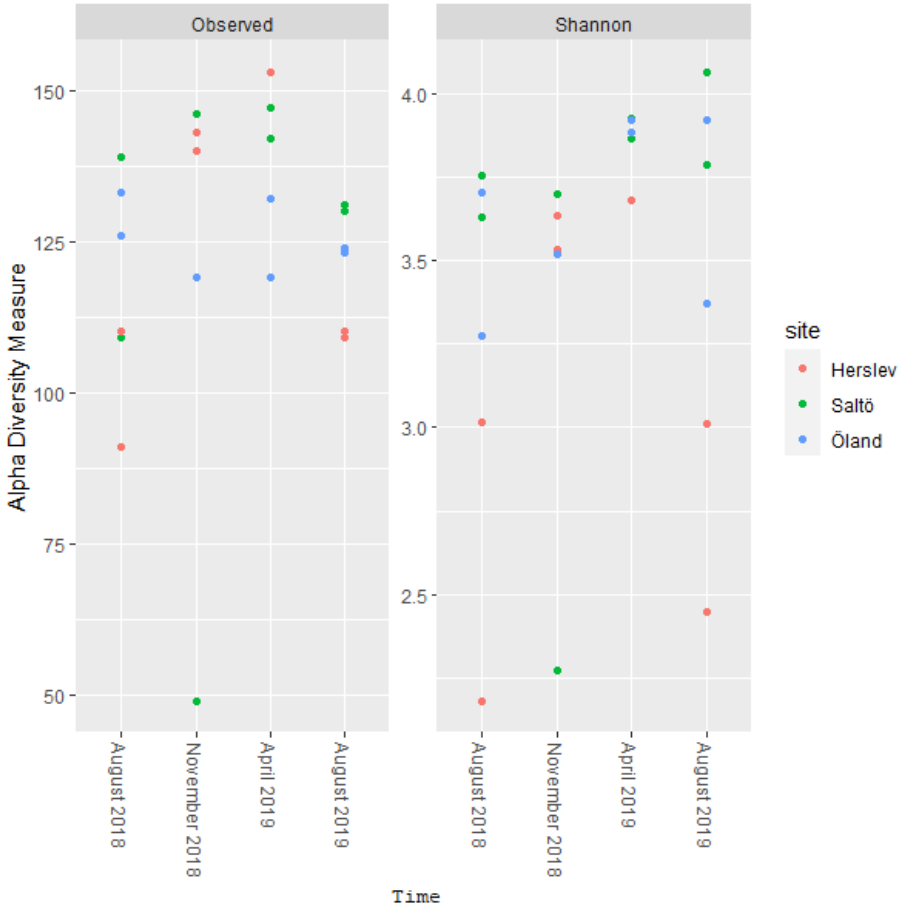


Figure 11. The alpha diversity of TarEuk samples visualized as a plot of alpha diversity measure against sampling time. The plot on the left describes the observed diversity, while the plot on the left uses Shannon index.

The beta diversity was calculated to be significantly different between sites, times and the interaction of them (p=0,001). The NMDS plot showing the diversity in the samples is shown in Figure 12.

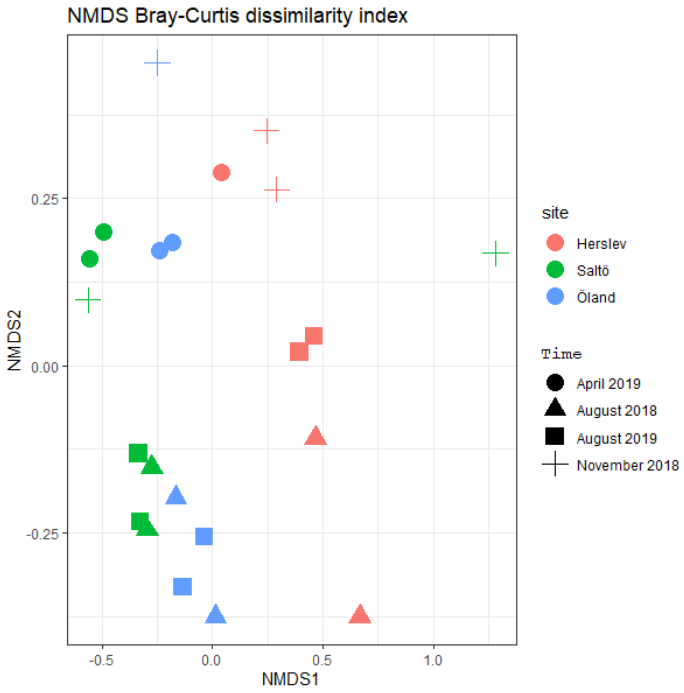


Figure 12. Non-metric multidimensional scaling plot with Bray-Curtis dissimilarity index for TarEuk samples. The different sampling sites are distinguished by color, while the shape indicates sampling time.

Ochrophyta (29,9 %, 610 OTUs) is the most common phylum in UNonMet samples and most common in April and August of 2019 in all sites (Figure 13). Herslev displays a high abundance of *Dinoflagellata* in August 2018, but the abundance is significantly smaller in the following timepoints. Simultaneously, the abundance of *Ochrophyta* can be seen increasing (this pattern is not, however, present in Öland). Also worth noting is the abundance of *Fungi* in Öland in November 2018. As with *Dinoflagellata* in Herslev, such high abundance is not detected in any other samples. Öland also displays a high abundance of *Ciliophora* in August 2018. This abundance declines through the year.

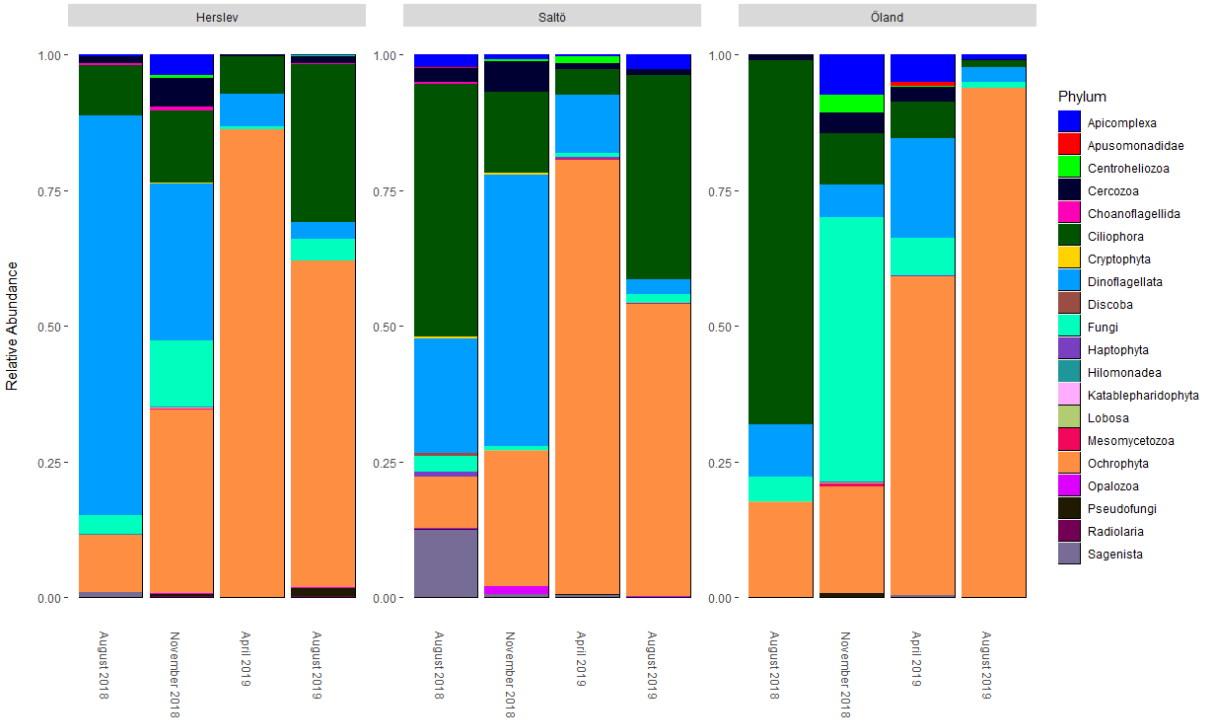


Figure 13. Relative abundance of phyla amplified with UNonMet primers.

The relative abundance (Figure 14) showed *Chrysophyceae* (11,7 %, 206 OTUs) to be the most abundant class in UNonMet during 2019 in all sites. Conversely, *Dinophyceae* was more abundant in 2018 in both Saltö and Herslev. Öland, on the other hand, differed by having a larger abundance of *Spirotricheae* (August 2018) and *Chytridiomycota* (November 2018).

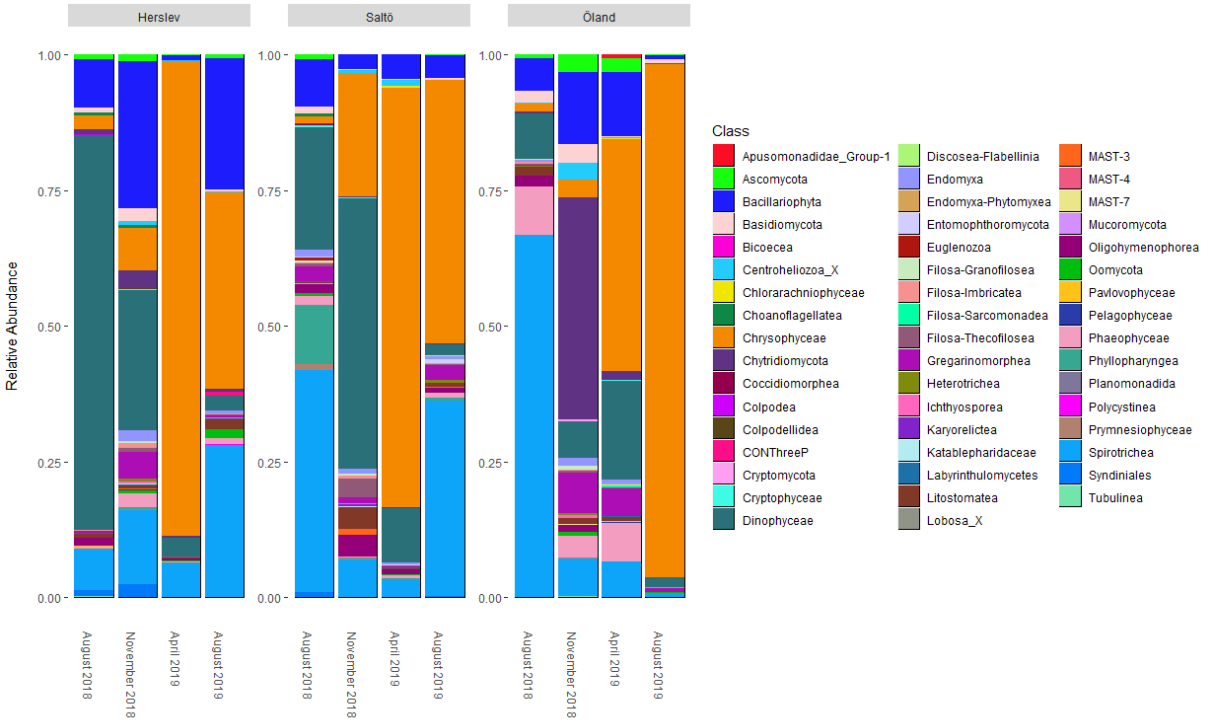


Figure 14. Relative abundance of classes amplified with UNonMet primers.

In Herslev, Shannon's index was lowest in April 2019 (1,491) and highest in November 2018 (4,67). In Öland, Shannon's index was lowest in August 2019 (1,342) and highest in November 2018 (4,147). In Saltö, Shannon's index was lowest in November 18 (1,44) and highest in August 2018 (4,351).

The lowest Shannon's index in August 2018 was in Herslev (2.888) and the highest was in Saltö (4,352). The lowest Shannon's index in November 2018 was in Saltö (1,44) and the highest was in Herslev (4,669). The lowest Shannon's index in April 2019 was in Herslev (1,491) and the highest was in Öland (4,121). The lowest Shannon's index in August 2019 was in Öland (1,342) and the highest was in Saltö (3,807).

The alpha diversity of UNonMet samples followed normality according to Shapiro-Wilk test (p=0,217). ANOVA revealed the alpha diversity to not differ significantly between sites (F = 0.196, df = 2, p = 0,825), time points (F = 1,419, df = 3, p=0,294) or the interaction of time and site (F = 1,328, df = 6, p=0,329). This was further proven by the

Tukey's honest significance test, where no comparison yielded a significant difference. Alpha diversity plot is shown in Figure 15.

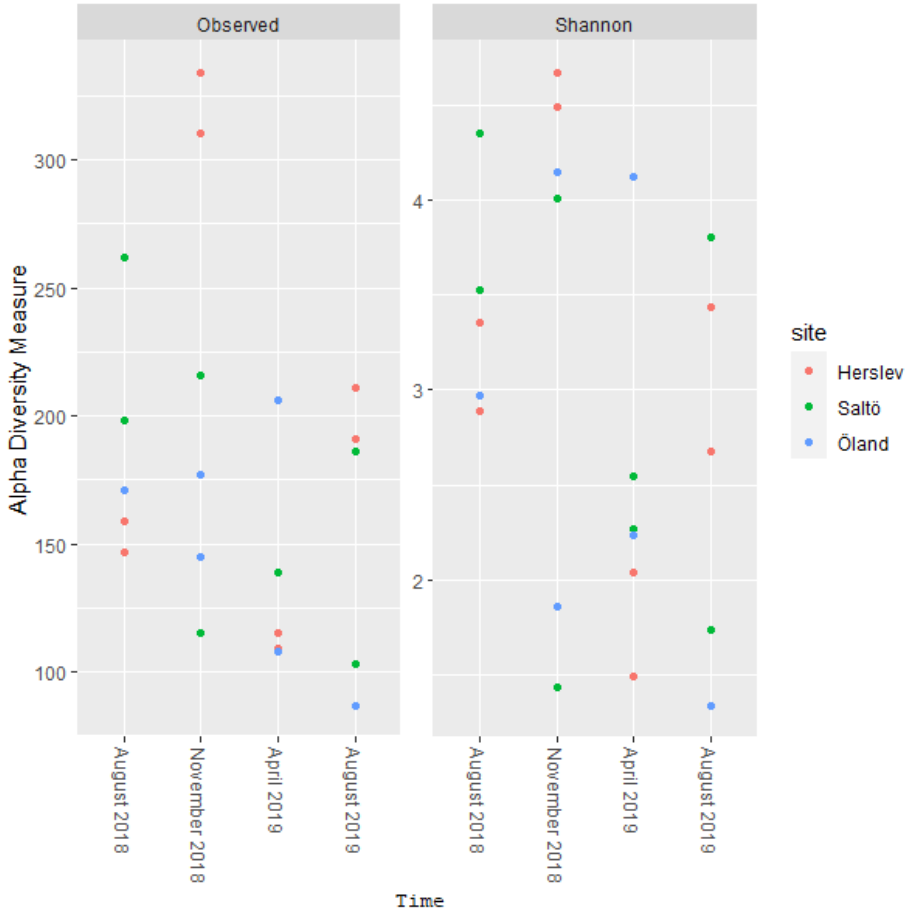


Figure 15. The alpha diversity of UNonMet samples visualized as a plot of alpha diversity measure against time (campaign). The plot on the left describes the observed diversity, while the plot on the left uses Shannon index.

The beta diversity between UNonMet samples was found to differ significantly with time (p=0,001), site (p=0,020) and their interaction (p=0,025). The beta diversity plot is shown in figure 16.

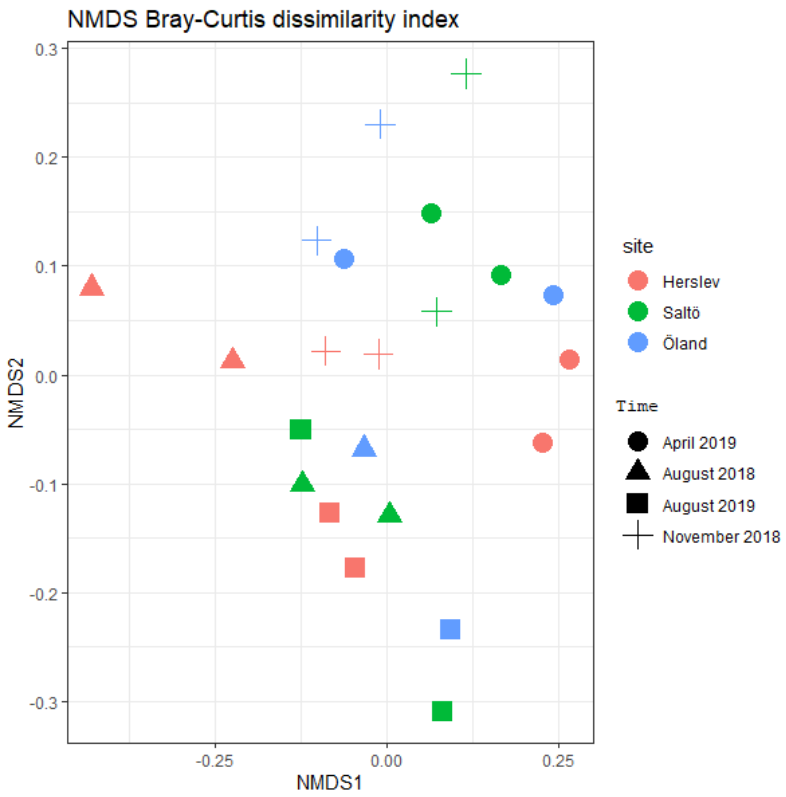


Figure 16. Non-metric multidimensional scaling plot with Bray-Curtis dissimilarity index for UNonMet samples. The different samples are distinguished by color, while the shape indicates time.

The most common phyla in nearly all V9 samples were *Ochrophyta* (35,4 %, 394 OTUs), *Dinoflagellata* (22,5 %, 251 OTUs) and *Ciliophora* (27,7 %, 309 OTUs) (Figure 17). *Dinoflagellata* were found to be prominent during August 2018 in Herslev. The number of *Ochrophyta* increases temporally in Herslev, while the number of *Dinoflagellata* declines. In other sites the abundance of *Ochrophyta* remains largely unchanged, while the abundance of *Dinoflagellata* increases in Öland as time goes on. Compared to other targets, the relative abundance remains more stable between sites and times.

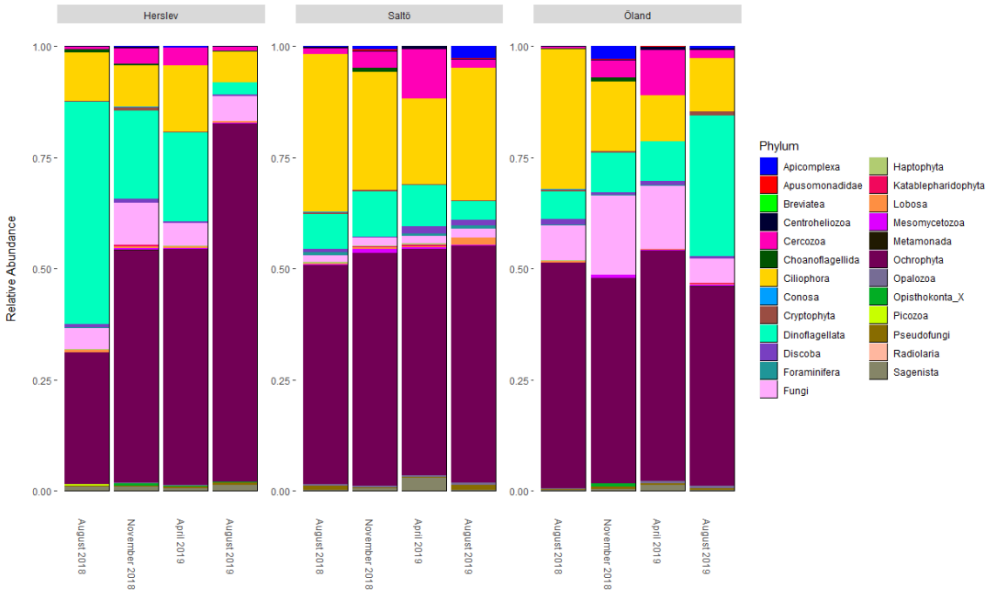


Figure 17. Relative abundance of classes amplified with V9 primers.

Bacillariophyta (43 %, 586 OTUs) was the most overwhelmingly abundant class in all samples (Figure 18). Other notable classes were *Spirotrichea* (7,7 %, 105 OTUs), *Dinophyceae* (8,2 %, 113 OTUs) and *Litostomatea* (4,3%, 59 OTUs).

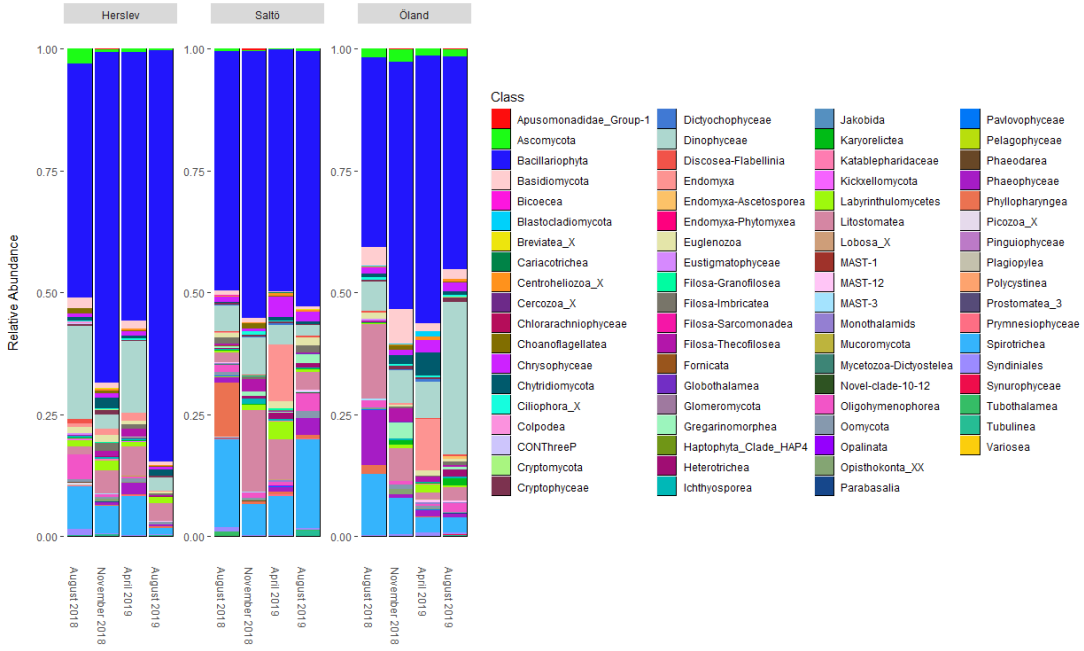


Figure 18. Relative abundance of classes amplified with V9 primers.

In Herslev, Shannon's index was lowest in August 2018 (2,071) and highest in November 2018 (4,147). In Öland, Shannon's index was lowest in August 2019 replicate 1 (3,093) and highest in August 2019 replicate 2 (4,33). In Saltö, Shannon's index was lowest in November 2018 (3,831) and highest in August 2018 (4,486).

The lowest Shannon's index in August 2018 was in Herslev (2,071) and the highest was in Saltö (4,486). The lowest Shannon's index in November 2018 was in Saltö (3,831) and the highest was in Öland (4,247). The lowest Shannon's index in April 2019 was in Herslev (3,722) and the highest was in Saltö (4,301). The lowest Shannon's index in August 2019 was in Herslev (2,546) and the highest was in Saltö (4.491).

The alpha diversity data of V9 did not follow normality (p=0,001) according to the Shapiro-Wilk test, so Kruskal-Wallis test was used. Time points showed no significant difference in alpha diversity (p=0,834). However, sites had a significant difference (p=0,024). Alpha diversity table is shown in Figure 19.

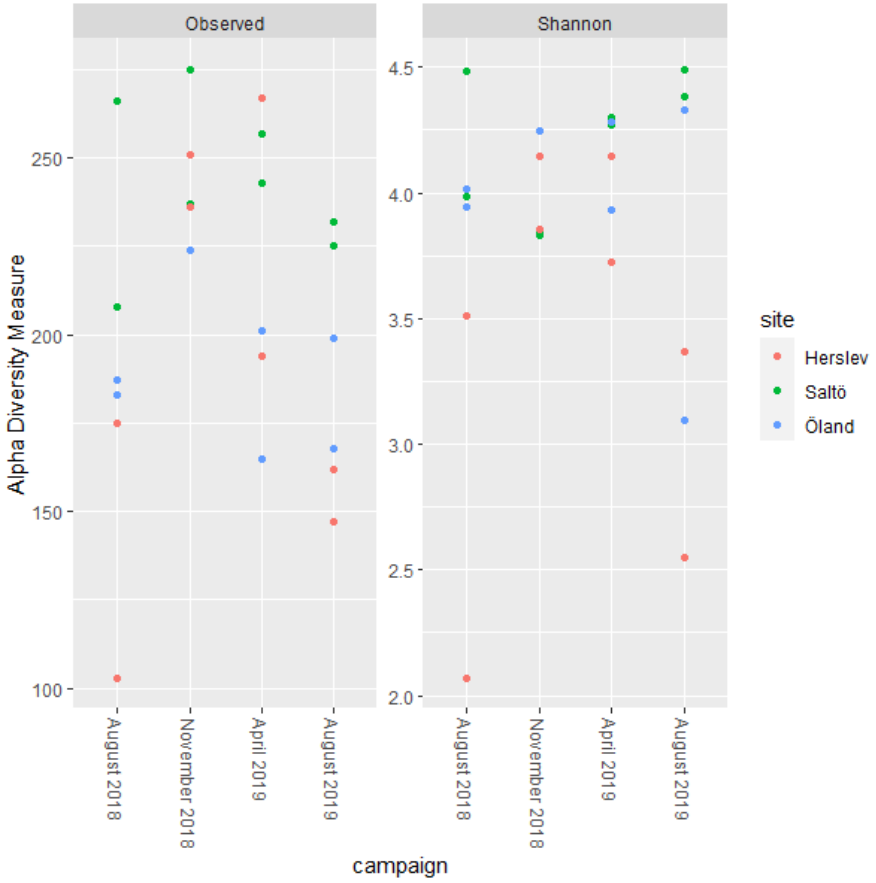


Figure 19. The alpha diversity of V9 samples visualized as a plot of alpha diversity measure against time (campaign). The plot on the left describes the observed diversity, while the plot on the left uses Shannon index.

The difference in beta diversity was significant in all variables (site, time, site*sample, p=0,001). The beta diversity is shown as a plot in Figure 20.

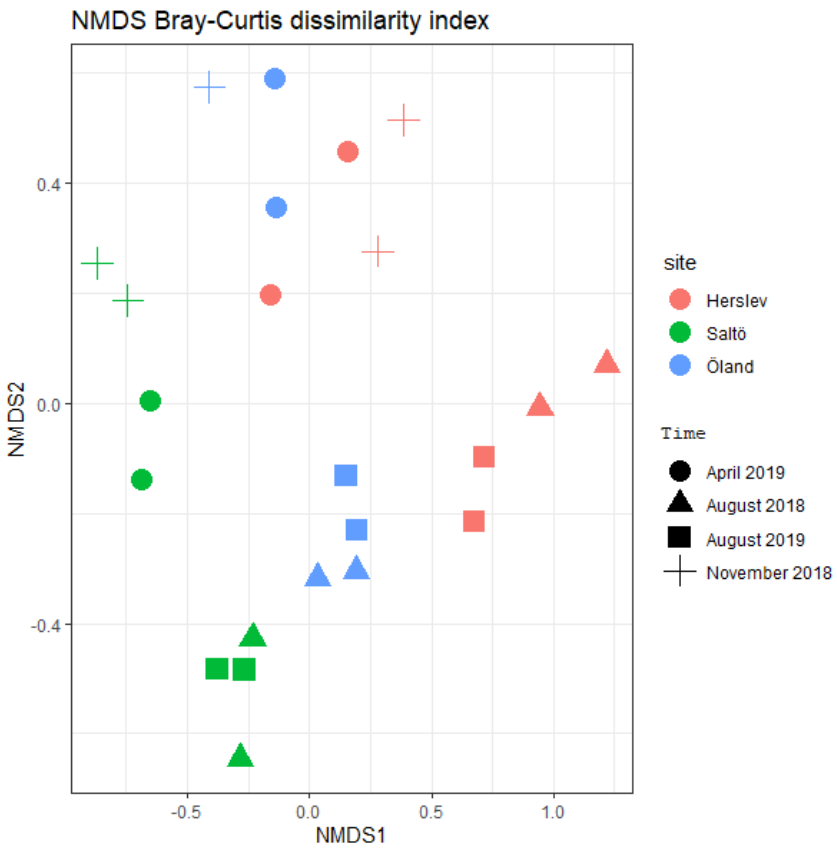


Figure 20. Non-metric multidimensional scaling plot with Bray-Curtis dissimilarity index for V9 samples. The different sampling sites are distinguished by color, while the shape indicates sampling time.

3.4 Apicomplexan Diversity

Apicomplexans were relatively rare among the samples. They were the most abundant when using TarEuk primers (Figure 21). The number of reads is approximately halved when using UNonMet primers, while V9 primers gave the least results. The number of observed apicomplexans when using V9 primers was noticeably low, with several samples containing no reads.

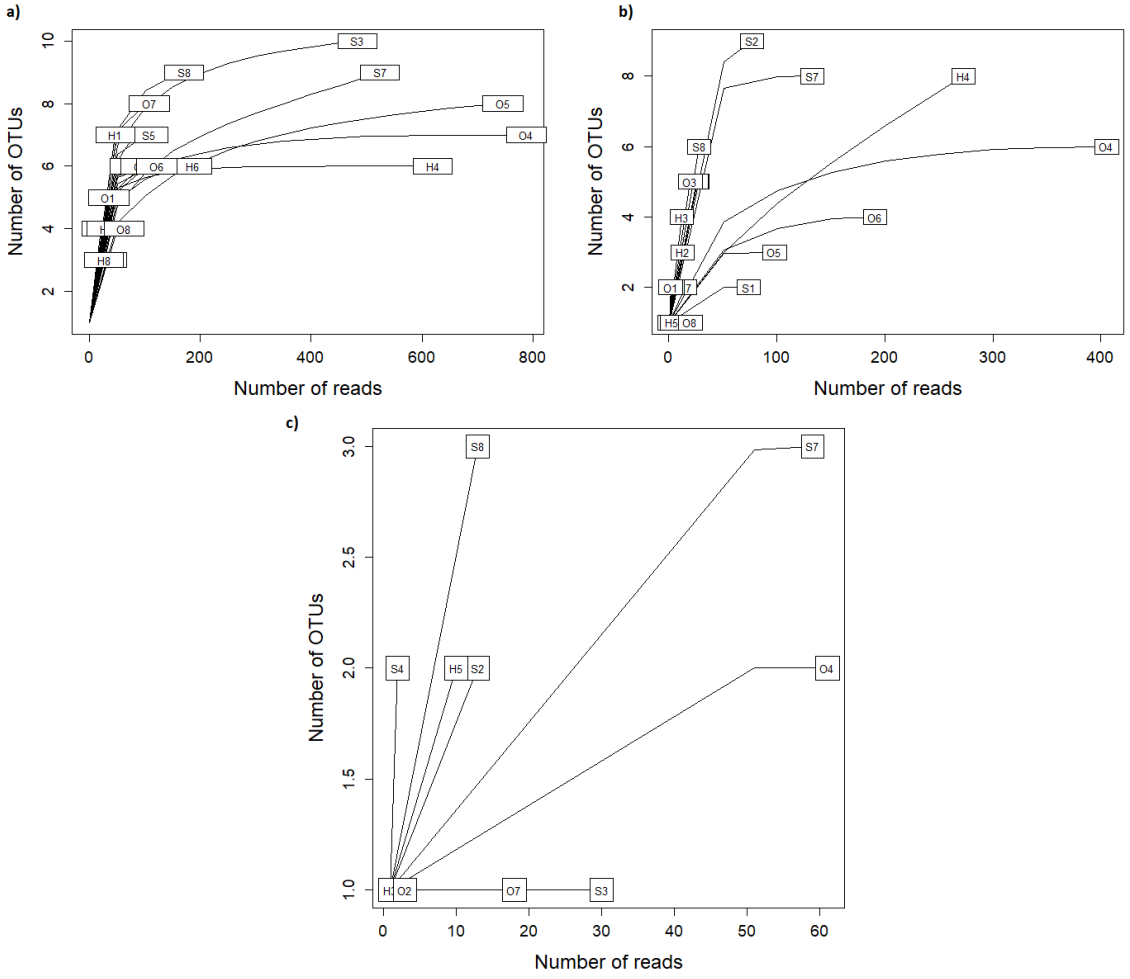


Figure 21. Accumulation curves of phylum *Apicomplexa* (unrarified), shown with the number of OTUs found against the number of reads. a: TarEuk; b: UNonMet; c: V9

The most common class regardless of time or site was Gregarinomorphea. Additionally, Coccidiomorphea and Colpodellidea were also detected in lesser numbers. Gregarinomorphea was the only class detected with V9 primers. The relative abundance charts for TarEuk, UNonMet and V9 are shown in Figure 22. The unrarefied relative abundance plots are included in Appendix 10.

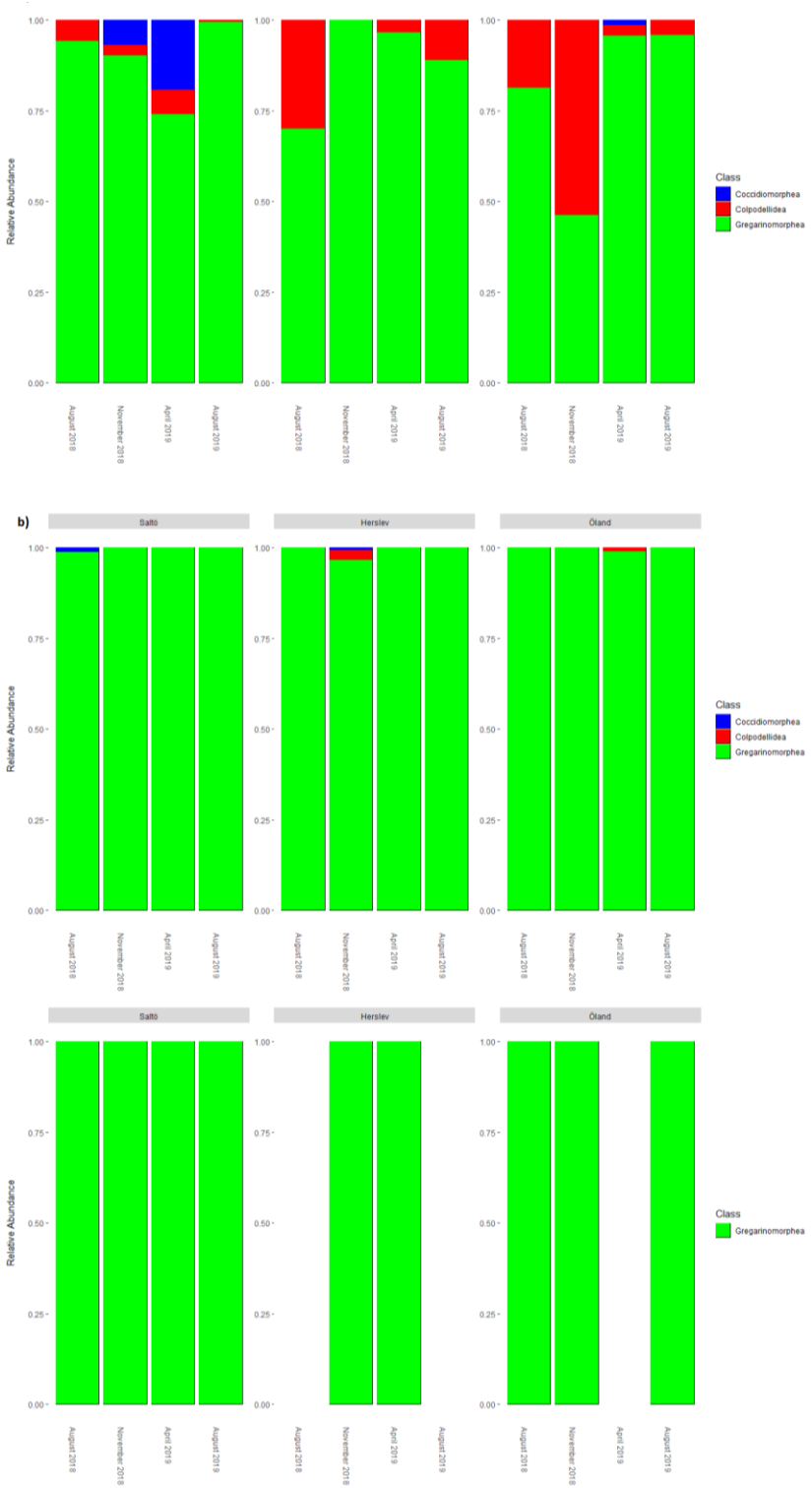


Figure 22. Relative abundances of phylum Apicomplexa found with different primer pairs. a: TarEuk; b: UNonMet; c: V9

4 DISCUSSION AND CONCLUSIONS

The aim of this thesis was to see how the microbial composition differs in coastal sediments of the Baltic Sea both spatially and temporally. It was hypothesized that bacteria would display less variation compared to protists. However, the results ended up not following the hypothesis.

The level of identification was better in bacteria. This was to be expected, as bacteria are better documented compared to protists. The number of reads in bacteria was multiple times larger than the number of reads in protists. One exception was UNonMet, although this was because it included one sample (November -18, Öland, replicate 1) that was an outlier. The reason for the exceptional read might have been an error in pooling. The difference between bacteria and protists is also evident in the number of OTUs identified, as the sequencing was able to identify a noticeably larger number of OTUs in bacteria. The reason for using primers for both V4 and V9 was to acquire a better idea of the diversity. The different primers could give different results due to sequencing bias and databases. The sequencing bias could thus have led to metazoan reads interfering with the results, as 18S is highly conserved even in multicellular eukaryotes.

The most common bacterial phyla were *Cyanobacteria* and *Proteobacteria.* Based on previous knowledge, these two phyla are two of the most common among bacteria (Madigan et al. 2019). The relative abundance figures suggest the relative abundance of phyla in bacterial communities remain steady both temporally and spatially. Permutational analysis of variance suggested that the beta diversity between the sites and time points is significantly different. Moreover, the alpha diversity was found to be temporally significant, with no significant variation between sites. This points to the bacteria having a higher degree of variation on a lower taxonomical level, while the relation of phyla remains constant. These results also imply bacterial community to be

dependent on the season, contrary to the hypothesis. Based on this, it could be deduced that bacterial diversity is affected differently by the environmental conditions compared to protists. Bacterial diversity could be less susceptible to differing salinity levels. This is in strong contrast to previous studies, such as one by Klier et al. (2018). According to this study, bacteria should be strongly impacted by changing salinity. Herlemann et al. (2011) also came to a conclusion that salinity is the most deciding factor in bacterial species composition. Because of this, the results can be considered inconclusive.

The relative abundance of protists varied more greatly between samples. A common pattern seen especially in Herslev was the high abundance of *Dinoflagellata* in August 2018, which subsequently declines in the following timepoints. Possible reason for this could be abnormal weather conditions during the summer. Indeed, August 2018 was measured to have the warmest temperature in Baltic Sea in recent years (Sea Temperature.info, 2021). Other factors could also have had an impact, albeit smaller. *Dinoflagellata* are known to thrive in higher temperatures, with many species being also photosynthetic and in correct conditions, *Dinoflagellata* are known to form harmful blooms (Madigan et al. 2019). This could also correlate to the rising abundance of *Ochrophyta.* As *Dinoflagellata* thrive, the space and resources for other phyla are limited. The decline of *Dinoflagellata* populations could allow for the other phyla to increase in numbers. In V9 samples, the number of *Dinoflagellata* is notably low and *Ochrophyta* is seemingly dominant throughout the year in all sites, with the exception of August 2018 in Herslev. The seasonal fluctuation of protists is in line with previous studies. A similar study conducted by Gran-Stadniczeñko et al. in Skagerrak (2019) also found temporal variation in protists. The results are also collaborated by a study by Mironoma et al. (2011), which concentrated especially on ciliates of Neva Estuary. The study also found the ciliate community to significantly fluctuate according to season in both abundance and composition.

Time was a lesser factor for alpha diversity variation compared to location with protists. The alpha diversity displayed no significant temporal variation when using TarEuk, UNonMet or V9 primers. As such, the protist species composition could be linked to the salinity. Beta diversity varied significantly both spatially and temporally. TBased on relative abundance tables, the OTU composition could be assumed to stay relatively similar throughout the year while their relative numbers fluctuate. This is supported by the relative abundance figures of classes, which display a greater amount of variation compared to the relative abundance figures of phyla.

Apicomplexans were relatively sparse in numbers compared to other phyla. The V9 primer pair, in particular, was unsuccessful at finding apicomplexans, while UNonMet was the most effective. A significant factor in the low number of Apicomplexans was the lack of data in the database and the actual number could have been higher. *Gregarinomorphea* was the most common class found in all sites. *Coccidiomorphea* and *Colpodellidea* were also found in low numbers with TarEuk and UNonMet. *Gregarinorphea* consists of parasites that use invertebrates as hosts (Rueckert et al. 2019). A part of their life cycle involves the parasite producing oocysts, which are most often found in the bottom sediments, waiting to be ingested by a new host. This would explain the relative abundance of this particular class. The parasitic nature of Apicomplexans could also in some capacity explain the low number of reads for them, as they are not usually found free roaming in the sediment.

The acquired data provide estimates of microbial diversity from coastal sites of the Baltic Sea and give more insight into how changing environmental factors affect these communities. In addition, previous studies on protist communities of the Baltic Sea have been minimal, which is why this thesis provides more insight into their diversity and structure.

4.1 Possible sources of errors

There exist various possible error sources that should be taken into account. 16S and TarEuk samples were pooled together before sequencing, instead of being pooled separately by the target. This could have affected the calculated regional molarity, which in turn could have had an effect on the results. In UNonMet and V9 samples the pooling was done as described in the methods. However, the pooled amount was only 5 µg instead of 10 µg, due to low concentration.

There were many environmental factors that weren't taken into account with the analyses that could have an impact on the outcome. The analyses didn't have data on temperatures, soil types, larger eukaryotes or oxygen levels. These factors could all have an effect on how the microbe communities are built. For example, oxygen levels are proven to be a significant factor in microbial composition of coastal sediments (Broman et al. 2017)

The reliability of the results could have been increased by increasing the number of samples. The amount of sample used (250 µg) was also small and might not have been representative of a larger area. However, the results give a glimpse into the larger microbial communities of the Baltic Sea.

ACKNOWLEDGEMENTS

I would like to thank my supervisors doctor Emily Knott and doctoral student Anna-Lotta Hiillos, who have been supporting me throughout this thesis, as well as Cecilie Petersen for additional support.

REFERENCES

Audemard C., Reece K. S. & Burreson E. M. 2004. Real-Time PCR for Detection and Quantification of the Protistan Parasite *Perkinsus marinus* in Environmental Waters. *Applied and Environmental Microbiology* 70(11): 6611-6618

Austin B. 2005. Bacterial Pathogens of Marine Fish. In: Belkin S. & Colwell R. R. *Oceans and Health: Pathogens in the Marine Environment.* Springer, Boston, MA

Bláha L., Babica P. & Maršálek B. 2009. Toxins produced in cyanobacterial water blooms- toxicity and risks. *Interdisciplinary toxicology* 2(2): 36-41

Bower S. M., Carnegie R. B., Goh B., Jones S. R. M., Lowe G. J. & Mak M. W. S. 2004. Preferential PCR Amplification of Parasitic Protistan Small Subunit rDNA from Metazoan Tissues. *The Journal of Eukaryotic Microbiology* 51(3): 325-332

Broman E., Sachpazidou V., Pinhassi J. & Dopson M. 2017. Oxygenation of Hypoxic Coastal Baltic Sea Sediments Impacts on Chemistry, Microbial Community Composition, and Metabolism. *Frontiers of Microbiology* 8:2453, doi: 10.3389/fmicb.2017.02453

Campbell N. A. ym. 2014. *Biology: A Global Approach.* Pearson, Harlow

Caron D. A., Countway P. D., Jones A. C., Kim D. Y., Schnetzer A. 2012. *Annual Review of Marine Science* 4: 467-93

Comeau A. M., Li W. K. W., Tremblay J., Carmack E. C. & Lovejoy C. 2011. Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLOS ONE 6(11), doi: https://doi.org/10.1371/journal.pone.0027492

del Campo J, Kolisko M, Boscaro V, Santoferrara LF, Nenarokov S, Massana R, Guillou L, Simpson A, Berney C, de Vargas C, et al. 2018. EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLoS Biol. 16(9):1–14. doi:10.1371/journal.pbio.2005849.

Dorigo U., Volatier L. & Humbert J. 2005. Molecular approaches to the assessment of biodiversity inaquatic microbial communities. *Water Research* 39: 2207-2218

Dziallas C., Allgaier M., Monaghan M. T. & Grossart H. 2012. Act together—implications of symbioses in aquatic ciliates. *Frontiers in microbiology* 3(288), doi: https://doi.org/10.3389/fmicb.2012.00288

Edlund A., Soule T., Sjöling S. & Jansson J. K. 2005. Microbial community structure in polluted Baltic Sea sediments. *Environmental microbiology* 8(2): 223-232

Edlund, A. (2007). *Microbial diversity in Baltic Sea sediments* (väitöskirja, Sveriges lantbruksuniversitet). Available at https://pub.epsilon.slu.se/1396/

European Commission. 1999. *Bacterial Kidney Disease.* Report of the Scientific Committee on Animal Health and Animal Welfare, available at https://ec.europa.eu/food/sites/food/files/safety/docs/sci-com_scah_out36_en.pdf

Frost P. C., Ebert D. & Smith V. H. 2008. Responses of a bacterial pathogen to phosphorus limitation of its aquatic invertebrate host. *Ecology* 89: 313-318

Gran-Stadniczeñko 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. *The Journal of Eukaryotic Microbiology* 66(3): 494-513

Gu W., Miller S. & Chiu C. Y. 2019. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annual Review of Pathology: Mechanisms of Disease* 14: 319-338

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, De Vargas C, Decelle J, et al. 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41(D1):597–604. doi:10.1093/nar/gks1160.

Hamilton G. 2006. *Kingdoms of Life: Protista.* Milliken Publishing Company, Dayton, Ohio

Hannon GJ. 2010. No Title FASTX-Toolkit: FASTQ/a short-reads pre-processing tools.

HELCOM. 2018. HELCOM Thematic assessment of biodiversity 2011-2018. *Baltic Sea Environment Proceedings* 158, available at https://helcom.fi/baltic-sea-trends/holistic-assessments/state-of-the-baltic-sea-2018/reports-and-materials/

Herlemann D. P. R., Labrenz M., Jürgens K., Bertilsson S., Waniek J. J. & Andersson A. F. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal* 5:1571-1579, doi: https://doi.org/10.1038/ismej.2011.41

Ininbergs K., Bergman B., Larsson J. & Ekman M. 2015. Microbial metagenomics in the Baltic Sea: Recent advancements and prospects for environmental monitoring. *Ambio* 44(Suppl. 3): 439-450

Kchouk M., Gibrat J. & Elloumi M. 2017. Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine* 9(3), doi: 10.4172/0974-8369.1000395

Keeling P. J. & del Campo J. 2017. Marine Protists Are Not Just Big Bacteria. *Current Biology* 11: R541-R549

Klier J., Dellwig O., Leipe T., Jürgens K. & Herlemann D. P. R. 2018. Benthic Bacterial Community Composition in the Oligohaline-Marine Transition of Surface Sediments in the Baltic Sea Based on rRNA Analysis. *Frontiers in Microbiology* 9: 236, doi: 10.3389/fmicb.2018.00236

Kohler S. L. & Hoiland W. K. 2001. Population regulation in an aquatic insect: the role of disease. *Ecology* 82: 2294-2305

Leander B. S. 2006. Ultrastructure of the archigregarine Selenidium vivax (Apicomplexa)-A dynamic parasite of sipunculid worms (host: Phascolosoma agassizii). *Marine Biology Research* 2: 178-190

Madigan M. T., Bender K. S., Buckley D. H., Sattley W. M. & Stahl D. A. 2019. Brock biology of microorganisms. Pearson, London, Great Britain

Martin M. 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet.journal. 17(1):10–12.

Martyniuk C. J. & Simmons D. B. 2016. Spotlight on environmental omics and toxicology: a long way in a short time. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 19: 97-101

McMurdie P. J. & Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* 8(4): e61217, doi: 10.1371/journal.pone.0061217

Medlin L. K. & Kooistra W. H. C. F. 2010. Methods to Estimate the Diversity in the Marine Photosynthetic Protist Community with Illustrations from Case Studies: A Review. *Diversity* 2(7): 973-1014

Mironova K., Telesh I. & Skarlato S. O. 2011. Diversity and seasonality in structure of ciliate communities in the Neva Estuary (Baltic Sea). *Journal of Plankton Research* 34(3):208-220, doi: 10.1093/plankt/fbr095

Mäki A., Rissanen A. J. & Tiirola M. 2016. A practical method for barcoding and size-trimming PCR templates for Amplicon sequencing. *BioTechniques* 60(2), doi: 10.2144/000114380

Nakatsu C. H. & Marsh T. L. 2007. Analysis of Microbial Communities with Denaturing Gradient Gel Electrophoresis and Terminal Restriction Fragment Length Polymorphism. *Methods for General and Molecular Microbiology, 3rd Edition*, https://doi.org/10.1128/9781555817497.ch41

OIE. 2019. *Manual of Diagnostic Tests for Aquatic animals.* Office International des Epizooties, Paris, France

Oksanen J., Blanchet F. G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P. R., R. B. O'Hara, Simpson G. L., Solymos P., Stevens M. H. H., Szoecs E. & Wagner H. (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Pérez-Cobas A. E., Gomez-Valero L. & Buchrieser C. 2020. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics* 6(8), doi: https://doi.org/10.1099/mgen.0.000409

Raghukumar S. 2002. Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology* 38: 127-145

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016(10):1–22. doi:10.7717/peerj.2584.

Rueckert S., Betts E. L. & Tsaousis A. D. 2019. The Symbiotic Spectrum: Where Do the Gregarines Fit? *Trends in Parasitology* 35(9): 687-694

Sarethy I. P., Sharadwata P. & Danquah M. K. 2013. Modern Taxonomy for Microbial Diversity. In: Grillo O. *Biodiversity-The Dynamic Balance of the Planet*. IntechOpen

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 75(23):7537–7541. doi:10.1128/AEM.01541-09.

Schmeisser C. Steele H. & Streit W. R. 2007. Metagenomics, biotechnology with non-culturable microbes. *Applied Microbiology and Biotechnology* 75: 955-962

Sea Temperature.Info. 2021. Baltic Sea Water temperature in August. Retrieved 5.4.2021 from https://seatemperature.info/august/baltic-sea-water-temperature.html

Shintani M., Sanchez Z. K. & Kimbara K. 2015. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology* 6: 242, doi: 10.3389/fmicb.2015.00242

Sinkko H., Lukkari K., Sihvonen L. M., Sivonen K., Leivuori M., Rantanen M., Paulin L. & Lyra C. 2013. Bacteria Contribute to Sediment Nutrient Release and Reflect Progressed Eutrophication-Driven Hypoxia in an Organic-Rich Continental Sea. *PLoS One* 8(6): e67061, doi: 10.1371/journal.pone.0067061

Sjöqvist C., Godhe A., Jonsson P. R., Sundqvist L. & Kremp A. 2015. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea–Baltic Sea salinity gradient. *Molecular Ecology* 24(11): 2871-2885

Stoeck T., Bass D., Nebel M., Christen R., Jones M. D. M., Breiner H. & Richards T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* 19(1): 21-31

Thomas T., Gilbert J. & Meyer F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2:3

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 73(16):5261–5267. doi:10.1128/AEM.00062-07.

WWF. 2020. Itämeren rehevöityminen. Retrieved 9.4.2020 from https://wwf.fi/alueet/itameri/rehevoityminen/

Xu D. & Klesius P. H. 2004. Two year study on the infectivity of *Ichthyophthirius multifiliis* in channel catfish *Ictalurus punctatus*. *Diseases of Aquatic Organisms* 59: 131-134

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and "all-species Living Tree Project

(LTP)" taxonomic frameworks. Nucleic Acids Res. 42(D1):643–648. doi:10.1093/nar/gkt1209.

APPENDIX 1: Table of fusion primer sequences

Primer	Sequence
M13_27F	**TGTAAAACGACGGCCAGT**AGAGTTTGATCMTGGCTCAG
M13_E572F	**TGTAAAACGACGGCCAGT**CYGCGGTAATTCCAGCTC
M13_ TAReuk454FWD1	**TGTAAAACGACGGCCAGT**CCAGCASCYGCGGTAATTCC
M13_1391F	**TGTAAAACGACGGCCAGT**GTACACACCGCCCGTC
IonP1_E1009R	**CCTCTCTATGGGCAGTCGGTGAT**CRAAGAYGATYAGATACCRT
IonP1_ TAReukREV3	**CCTCTCTATGGGCAGTCGGTGAT**ACTTTCGTTCTTGATYRA
IonP1_EukB	**CCTCTCTATGGGCAGTCGGTGAT**TGATCCTTCTGCAGGTTCACCTAC
IonP1_338R	**CCTCTCTATGGGCAGTCGGTGAT**TGCTGCCTCCCGTAGGAGT
IonA _key_**bc**_ M13	**CCATCTCATCCCTGCGTGTCTCCGAC**TCAG**NNNNNNNNNN**TGTAAAACGACGGCCAGT

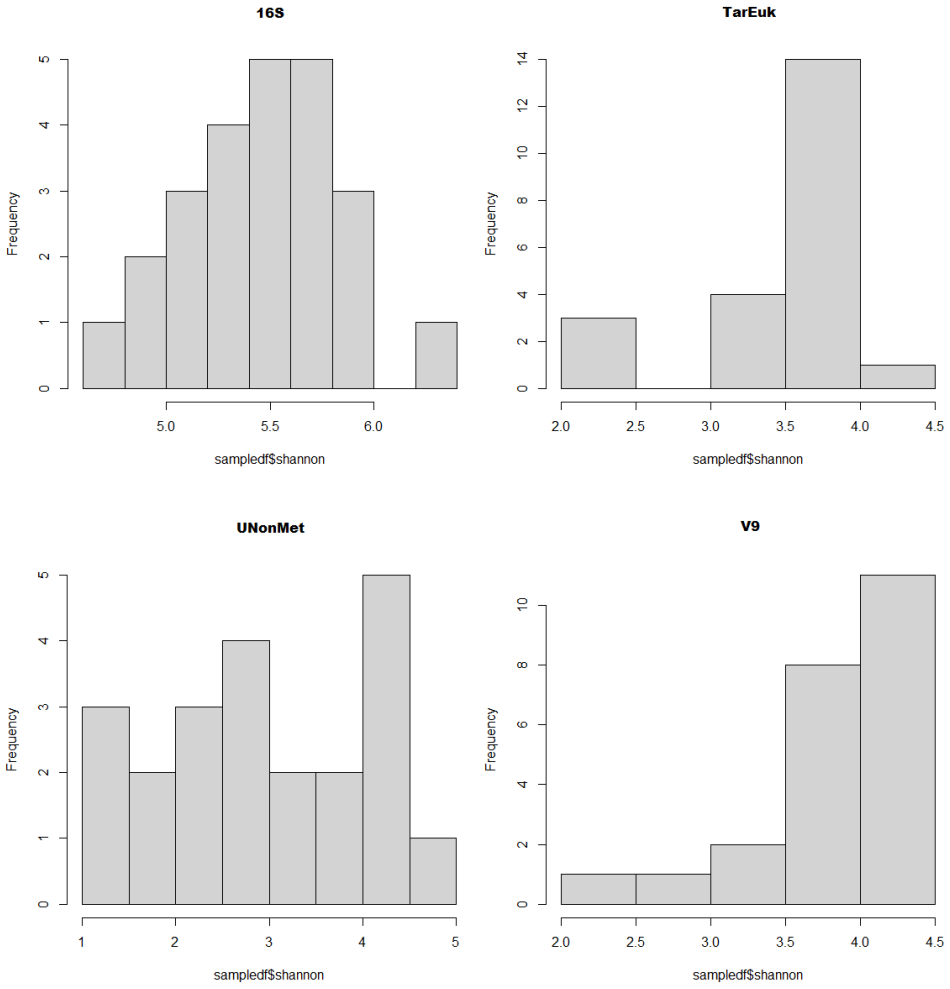
APPENDIX 2: Table of redone samples

The samples marked with blue were not included in the final pooling.

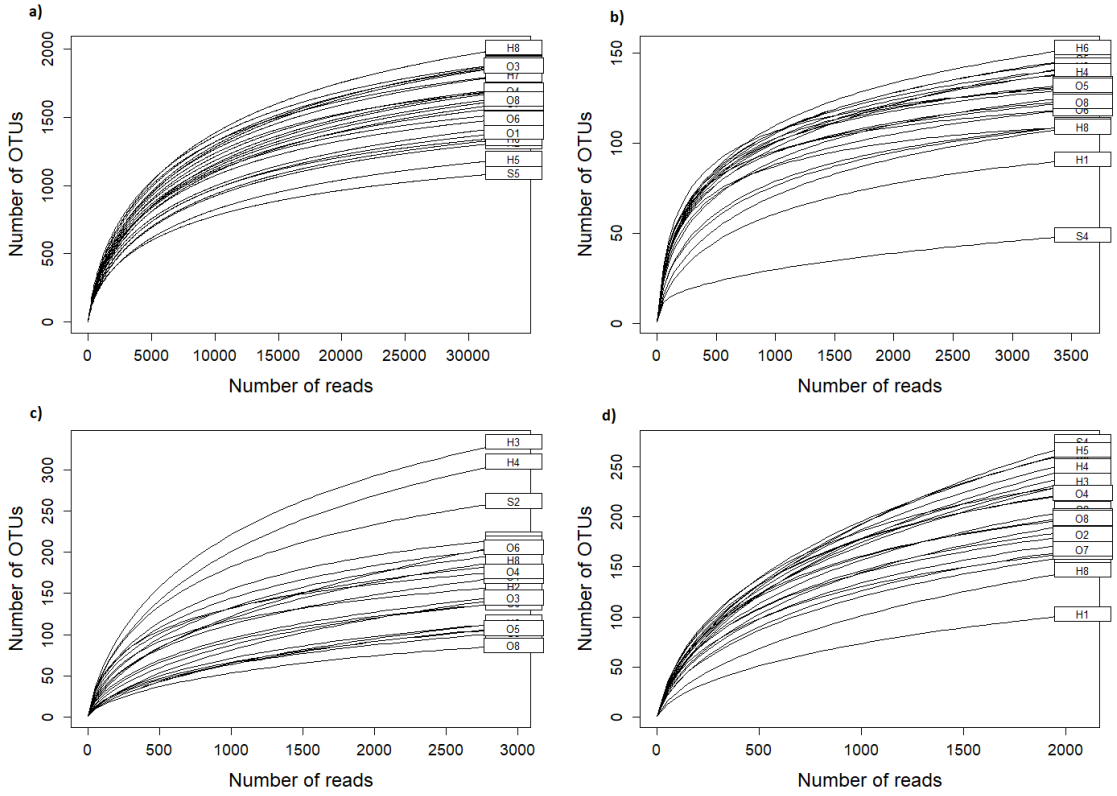
Primer Pair	Time	Replicate	Site
TarEuk	Nov 18	1	Öland
TarEuk	Apr 19	1	Herslev
UNonMet	Aug 18	1	Öland
UNonMet	Nov 18	1	Herslev
UNonMet	Nov 18	1	Öland
UNonMet	Aug 19	1	Herslev
UNonMet	Aug 19	1	Öland
UNonMet	Nov 18	2	Herslev
UNonMet	Apr 19	2	Öland
UNonMet	Aug 19	2	Herslev
V9	Apr 19	1	Öland

V9	Aug 18	2	Herslev
V9	Aug 18	2	Saltö
V9	Aug 18	2	Öland
V9	Nov 18	2	Herslev
V9	Aug 19	2	Öland

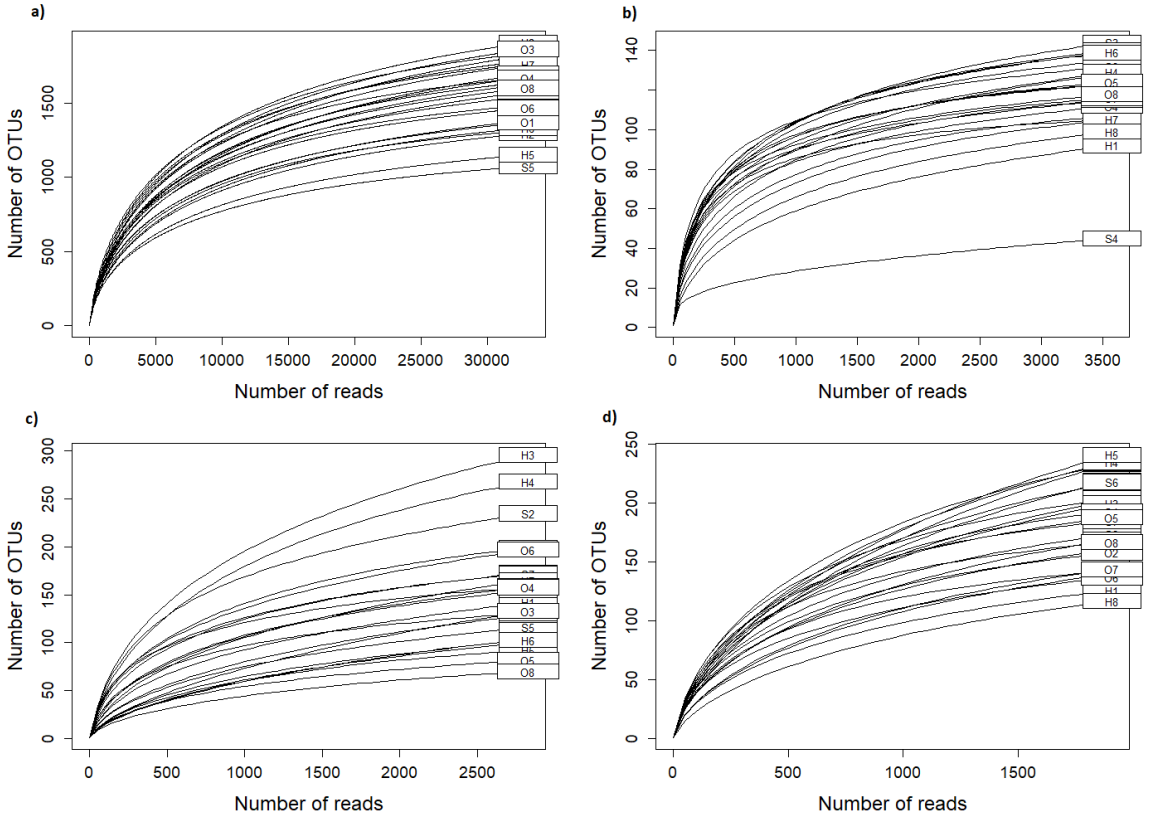
APPENDIX 3: Shannon index normality



APPENDIX 4: Rarefied accumulation curves



Rarefied accumulation curves for OTUs identified to phylum level. a: 16S; b: TarEuk; c: UNonMet; d: V9



Rarefied accumulation curves for OTUs identified to class level. a: 16S; b: TarEuk; c: UNonMet; d: V9

APPENDIX 5. Legend for sample names in accumulation curves

Sample	Time	Site	Replicate
H1	August -18	Herslev	1
H2	August -18	Herslev	2
H3	November -18	Herslev	1
H4	November -18	Herslev	2
H5	April -19	Herslev	1
H6	April -19	Herslev	2
H7	August -19	Herslev	1
H8	August -19	Herslev	2
S1	August -18	Saltö	1
S2	August -18	Saltö	2
S3	November -18	Saltö	1
S4	November -18	Saltö	2
S5	April -19	Saltö	1
S6	April -19	Saltö	2
S7	August -19	Saltö	1
S8	August -19	Saltö	2
O1	August -18	Öland	1
O2	August -18	Öland	2
O3	November -18	Öland	1
O4	November -18	Öland	2
O5	April -19	Öland	1
O6	April -19	Öland	2
O7	August -19	Öland	1
O8	August -19	Öland	2

APPENDIX 6: Unrarefied barplots (phylum)

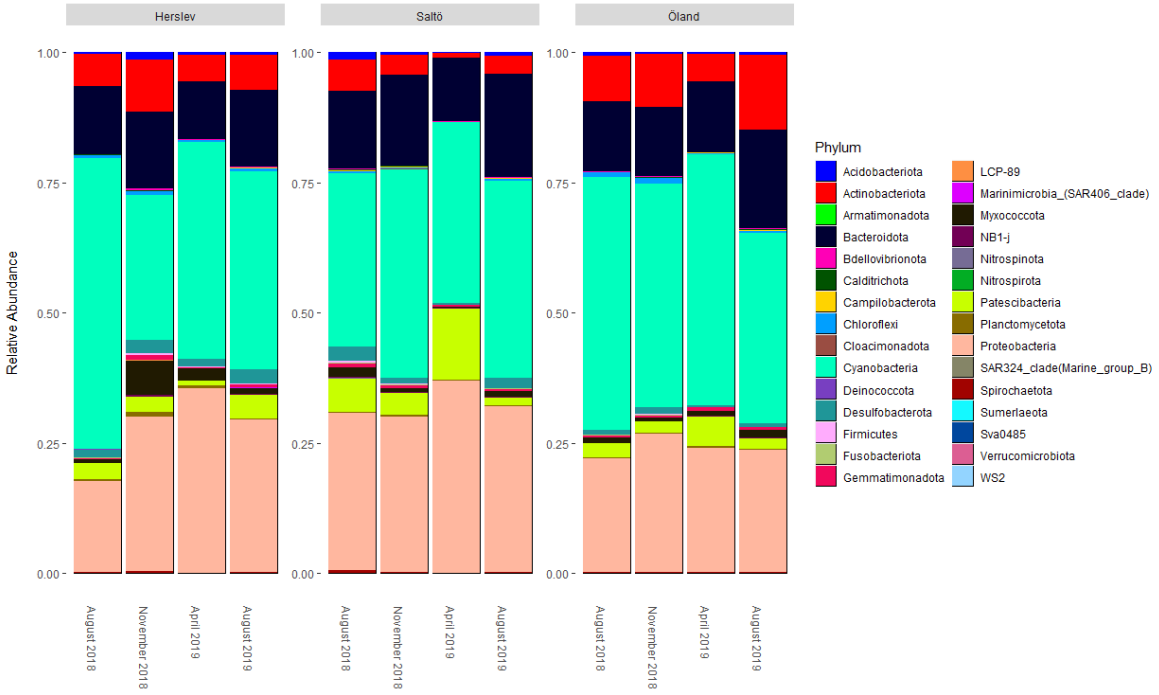
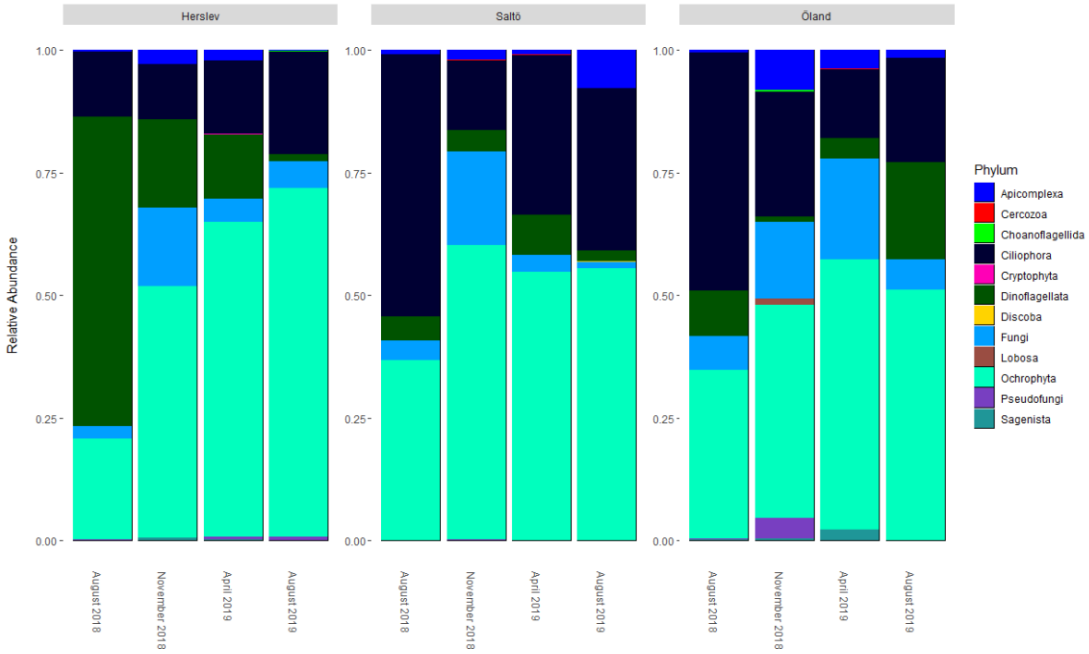
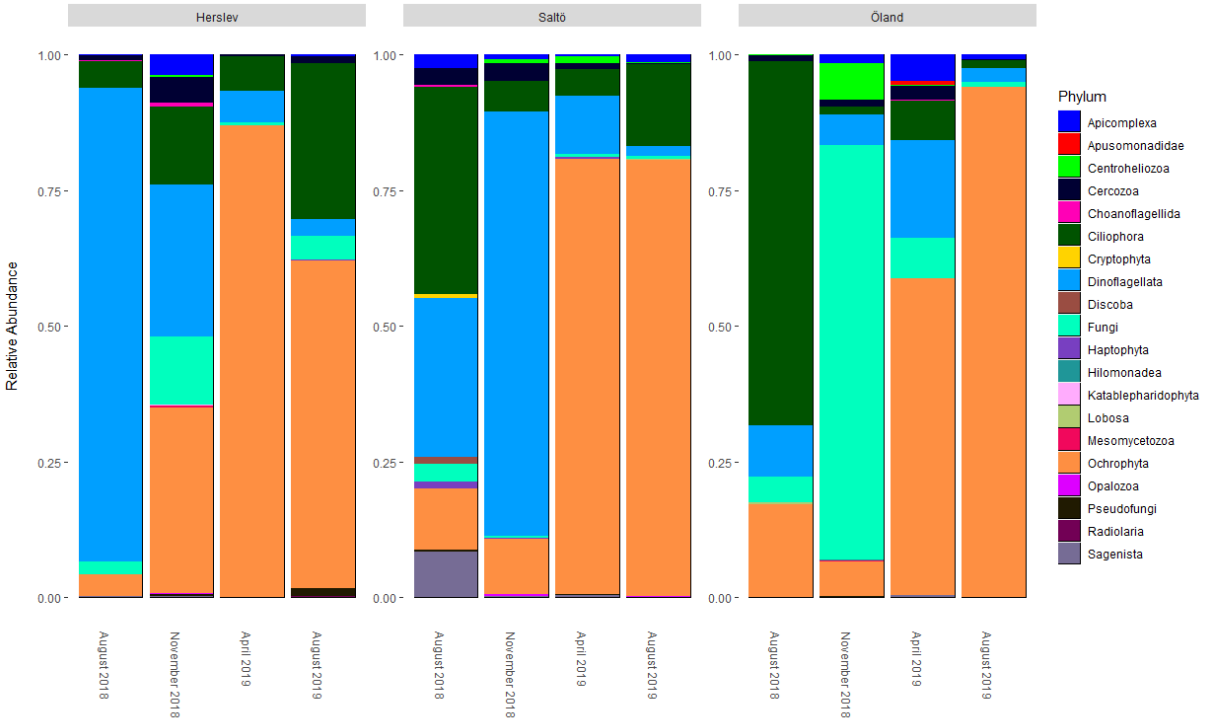


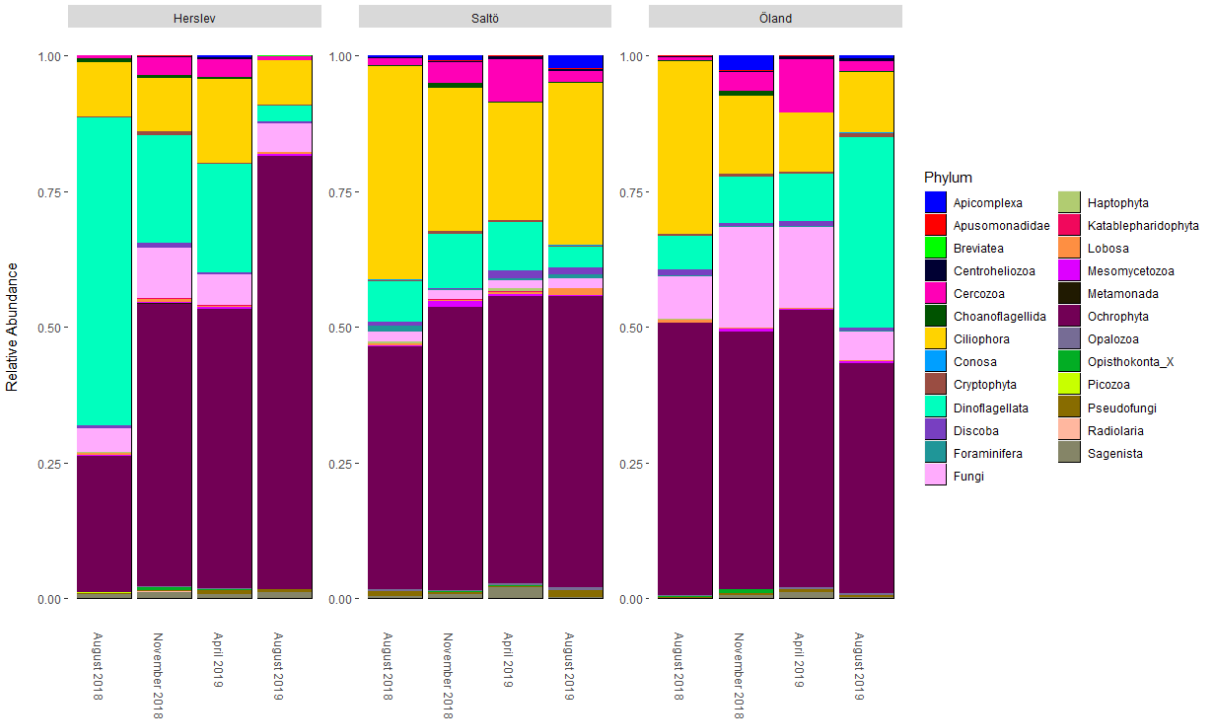
Figure 4 16S samples



TarEuk samples



UNonMet samples



V9 samples

APPENDIX 7: Alpha diversity estimates for Shannon index

ANOVA/Kruskal-Wallis test estimates for Bray-Curtis dissimilarity index with site, time point and their interaction (when applicable). TarEuk and V9 were analyzed with Kruskal-Wallis test and the significance between interactions couldn't therefore be estimated

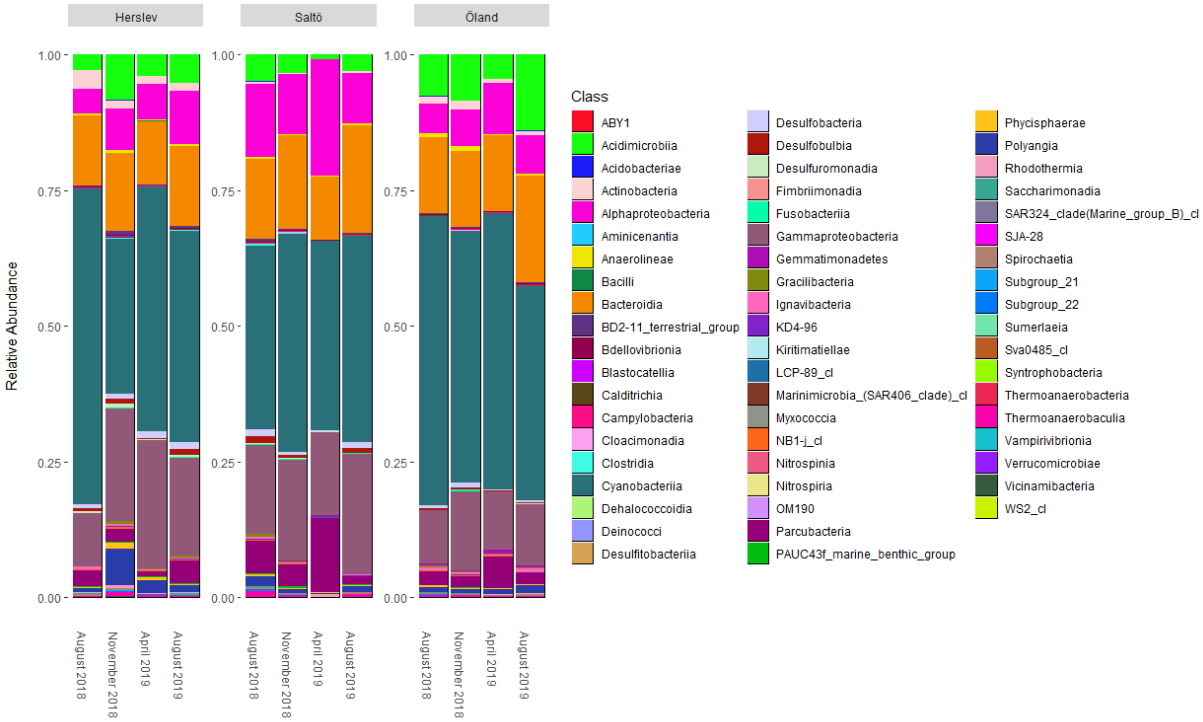
16S						
	Df	Sums of Squares	Mean Squares	F -model	$\chi 2$	p-value
Site	2	0.3321	0.1661	3.985	-	0.04707
Time	3	1.1877	0.3959	9.501	-	0.00171
Site * Time	6	1.2388	0.2065	4.955	-	0.00901
Residuals	12	0.5000	0.0417	-	-	-
Total	23	3.2586	-	-	-	-
18S TarEuk						
Site	2	-	-	-	6.4252	0.04025
Time	3	-	-	-	6.4411	0.09201
Residuals	-	-	-	-	-	-
Total	5	-	-	-	-	-
18S UNonMet						
Site	2	0.422	0.2111	0.196	-	0.825
Time	3	4.580	1.5268	1.419	-	0.294
Site * Time	6	8.571	1.4285	1.328	-	0.329
Residuals	10	10.757	1.0757	-	-	-
Total	21	24.33	-	-	-	-
18S V9						
Site	2	-	-	-	7.4592	0.024
Time	3	-	-	-	0.86594	0.8336
Residuals	-	-	-	-	-	-
Total	5	-	-	-	-	-

APPENDIX 8: Beta diversity estimated for Bray-Curtis dissimilarity index

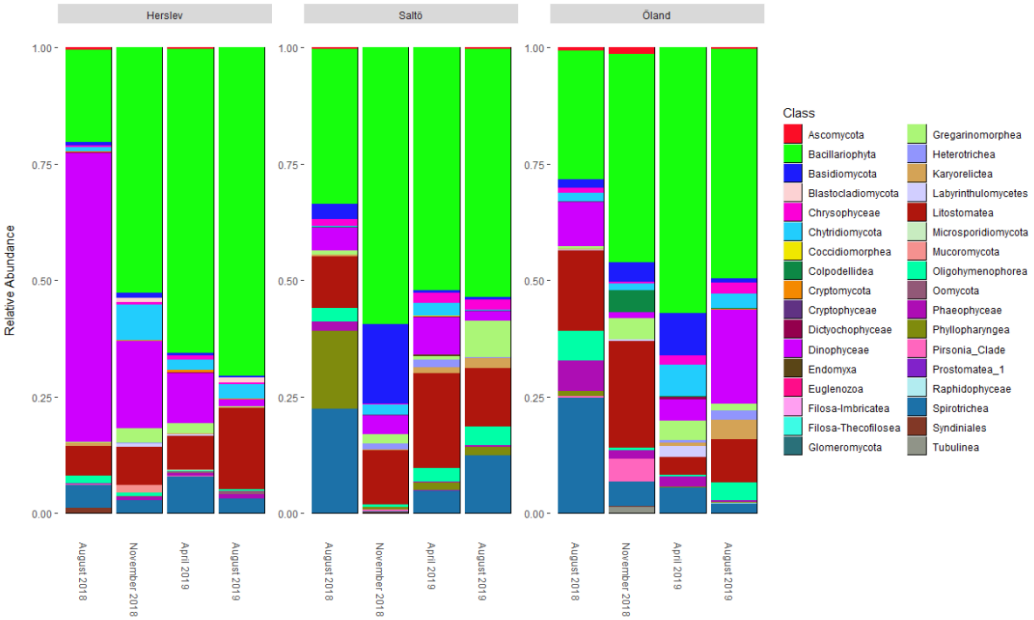
Permanova estimates for Bray-Curtis dissimilarity index with site, time point and their interaction for each target. Models were calculated with 999 permutations.

16S						
	Df	Sums of Squares	Mean Squares	F -model	R^2	p-value
Site	2	1.6116	0.80581	14.0413	0.41353	0.001
Time	3	0.7621	0.25403	4.4266	0.19555	0.001
Site * Time	6	0.8349	0.13915	2.4246	0.21422	0.001
Residuals	12	0.6887	0.05739	-	0.17670	-
Total	23	3.8973	-	-	1.00000	-
18S TarEuk						
Site	2	1.5770	0.78848	5.2318	0.25854	0.001
Time	3	1.1504	0.38346	2.5444	0.18860	0.001
Site * Time	6	1.8651	0.31085	2.0626	0.30578	0.001
Residuals	10	1.5071	0.15071	-	0.24708	-
Total	21	6.0996	-	-	1.00000	-
18S UNonMet						
Site	2	0.9888	0.49441	1.6119	0.11081	0.020
Time	3	2.3954	0.79846	2.6032	0.26842	0.001
Site * Time	6	2.4724	0.41207	1.3434	0.27706	0.025
Residuals	10	3.0673	0.30673	-	0.34371	-
Total	21	8.9239	-	-	1.00000	-
18S V9						
Site	2	1.5624	0.78118	5.8382	0.23574	0.001
Time	3	1.4614	0.48713	3.6406	0.22050	0.001
Site * Time	6	2.1320	0.35533	2.6556	0.32168	0.001
Residuals	11	1.4719	0.13380	-	0.22208	-
Total	22	6.6276	-	-	1.00000	-

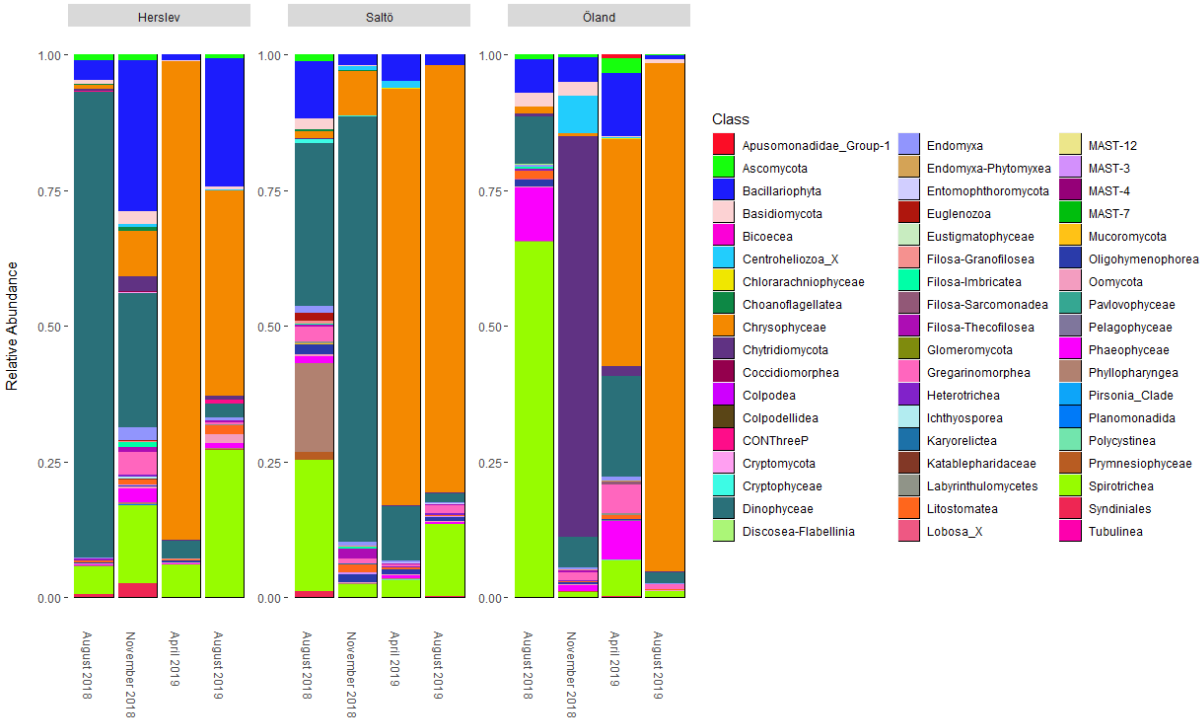
APPENDIX 9: Unrarefied barplots (class)



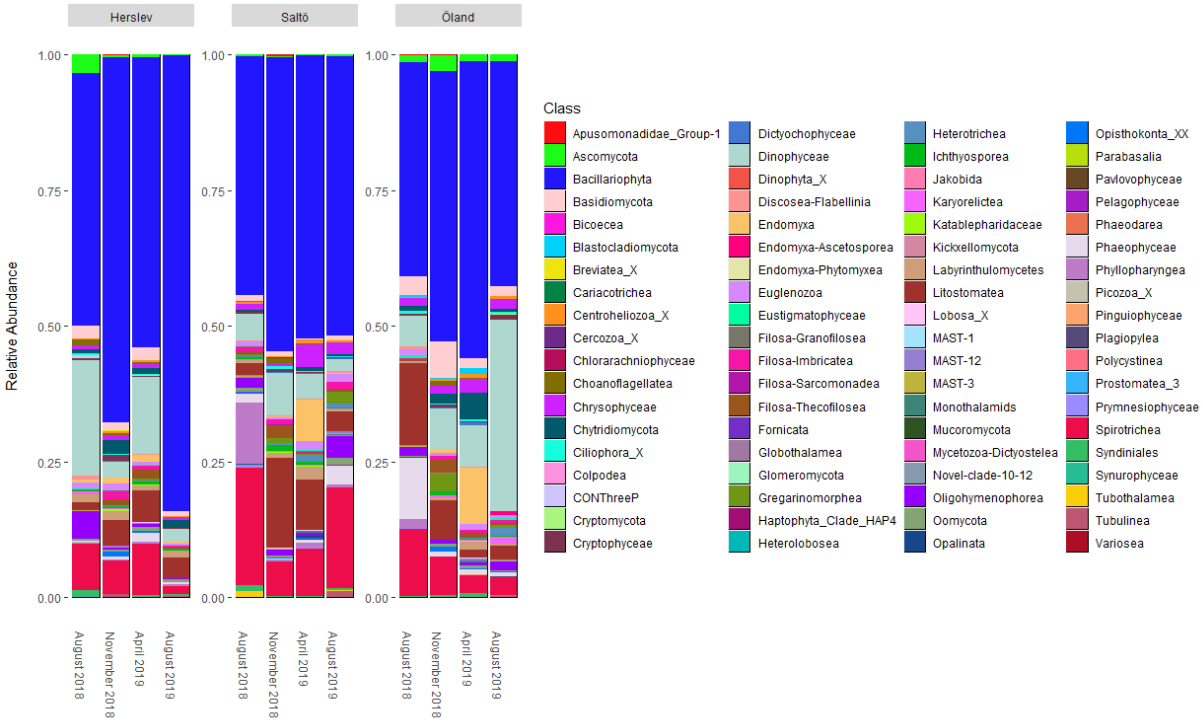
16S samples



TarEuk samples

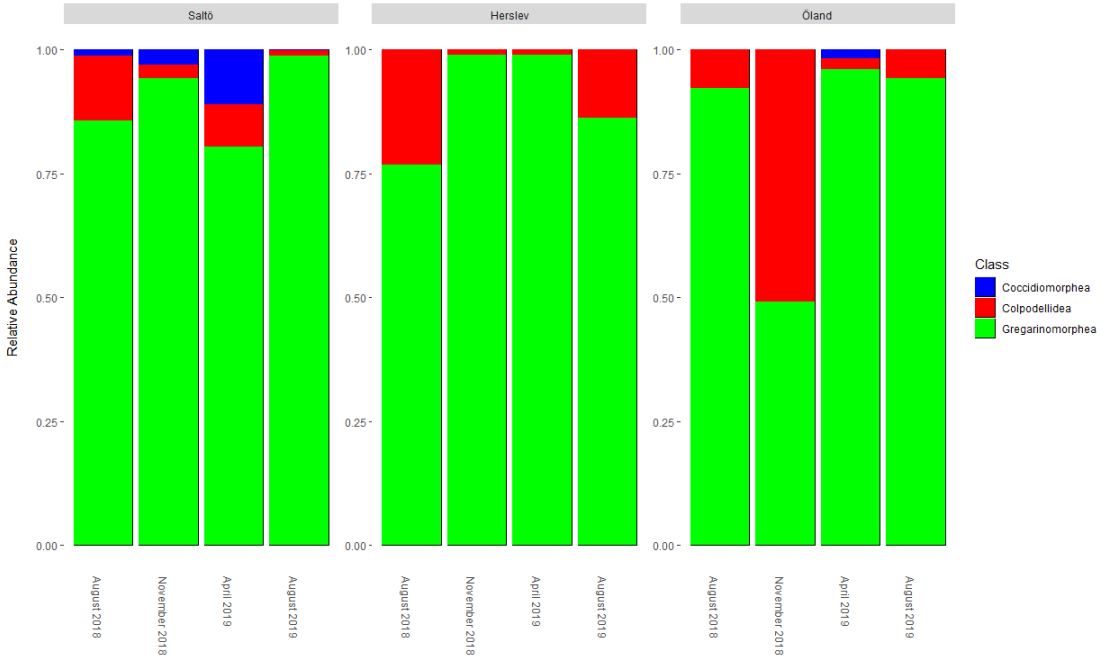


UNonMet samples

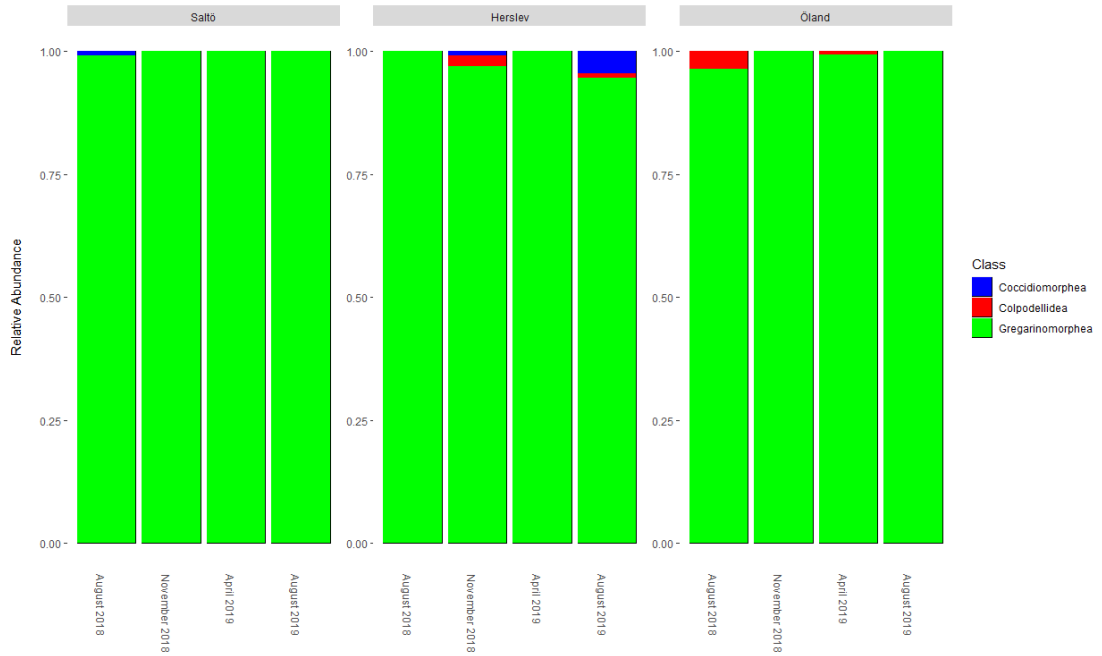


V9 samples

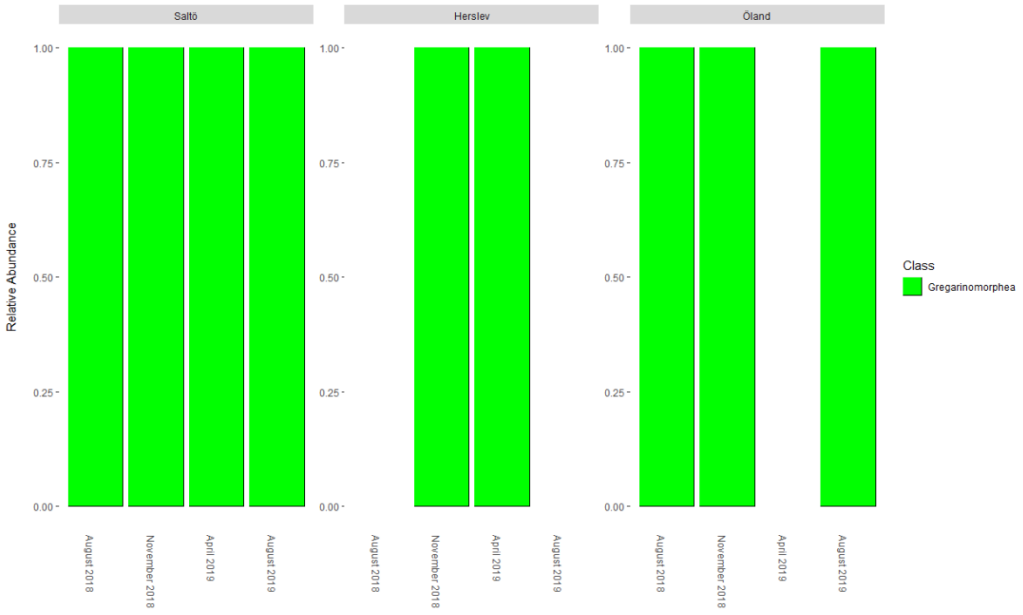
APPENDIX 10: Unrarefied barplots (Apicomplexa)



TarEuk



UNonMet


Saltö
Herslev
Öland
Relative Abundance
1.00
0.75
0.50
0.25
0.00
August 2018
November 2018
April 2019
August 2019
Class
Gregarinomorphea

V9