Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Astrid Poleszynski Hoel

A year in the Arctic; Seasonal Dynamics of the Marine Prokaryote Community in Isfjorden, Svalbard

Master's thesis in Biotechnology

Supervisor: Ass. prof. Anna Vader and prof. Ingrid Bakke

September 2020

NTNU
Norwegian University of Science and Technology

UNIS
The University Centre in Svalbard

Astrid Poleszynski Hoel

A year in the Arctic; Seasonal Dynamics of the Marine Prokaryote Community in Isfjorden, Svalbard

Master's thesis in Biotechnology
Supervisor: Ass. prof. Anna Vader and prof. Ingrid Bakke
September 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of Science and Technology

UNIS
The University Centre in Svalbard

Abstract

This thesis represents a pilot study of the marine prokaryotic community at the IsA time series station (78°N, 15°E) located in Isfjorden, Svalbard. Water samples were taken on a monthly basis for a full year (2019) at different depths, together with measures of temperature and salinity profiles, inorganic nutrient concentrations, phytoplankton biomass (Chl a) and cell counts. This allowed for a better insight on the extreme seasonal dynamics found in the Arctic. Illumina 16S rDNA analysis revealed *Proteobacteria* and *Bacteroidetes* as the most abundant phyla, together with relatively high amounts of *Thaumarchaeota* during winter and autumn. *Bacteroidetes* showed a large increase during summer, with a single taxonomic unit (TU) belonging to *Polaribacter* making up to 44% of the reads in June. Winter was characterized by a diverse and stable community with high evenness and many low abundance TUs, while summer was dominated by a few highly abundant TUs and rapid changes in community structure. The main driver for this change was recognized as the annual phytoplankton bloom. Inorganic nutrient availability was coupled to heterotrophic cell counts and changes in phytoplacton composition. The effect of water masses was less clear. Stratification during summer could have enhanced differences in nutrient availability between the upper and lower depths.

Contents

1 Introduction

1.1 Prokaryotic community of the Arctic

All across the worlds oceans picoplancton play an essential role as the agents of nutrient cycling and provide an important source of energy for higher trophic levels (Doolittle, Li, and Wood 2008; Stockner 1988). This is especially prominent in the Arctic (Lee and Whitledge 2005; Tremblay et al. 2009; E et al. 2006; Christman et al. 2011). A survey from 2010 estimated there to be around 45,000 bacterial taxonomic units (TUs) in the Arctic ocean while only a tenth the number of Archea and Pico-Eucarya (Lovejoy, Pierre E. Galand, and David L. Kirchman 2011). Of these 60% have been found to be unique for the Arctic alone (Ghiglionea et al. 2012). Even so, studies concerning bacteria have been widely dispersed in either time or location, or does not span a full season.

The Arctic is an area greatly dominated by large seasonal variation, especially in light conditions, going form perpetual darkness during winter to midnight sun during the summer (Figure 1), but also temperature, ocean currents and nutrient availability change greatly (F. R. C. F. Nilsen et al. 2010; Wiedmann et al. 2016). This is reflected in large changes observed in the prokaryotic community of the epipelagic surface waters ($<$ 200 m) (Wilson et al. 2017; David L. Kirchman, Cottrell, and Lovejoy 2010; Christman et al. 2011; Connelly et al. 2014). Changes in light availability and the annual phytoplancton bloom have been indicated as the main drivers for this (Wilson et al. 2017; Garneau et al. 2008). Water masses have also been pointed to as an important factor for determining community composition (Pierre E Galand, Potvin, et al. 2010; Müller et al. 2018; Pierre E. Galand et al. 2009), but mainly as stable biogegraphic barriers. Seasonal changes in ocean currents and water masses could however, be hypothesised as important drivers for more dynamic shifts in community composition. Inorganic nutrient availability has been coupled to the seasonal variation of *Crenarchaeota* (Wuchter et al. 2006; Herfort et al. 2007; David L. Kirchman, Elifantz, et al. 2007), but is thought to play a grater indirect role by facilitating the annual phytoplancton bloom (Wilson et al. 2017; Connelly et al. 2014).

Spring is a period of spectacular change in the Arctic as the sun returns after many months of perpetual darkness. This powers a great phytoplancton bloom which follows the retreating sea ice. This initial boom accounts

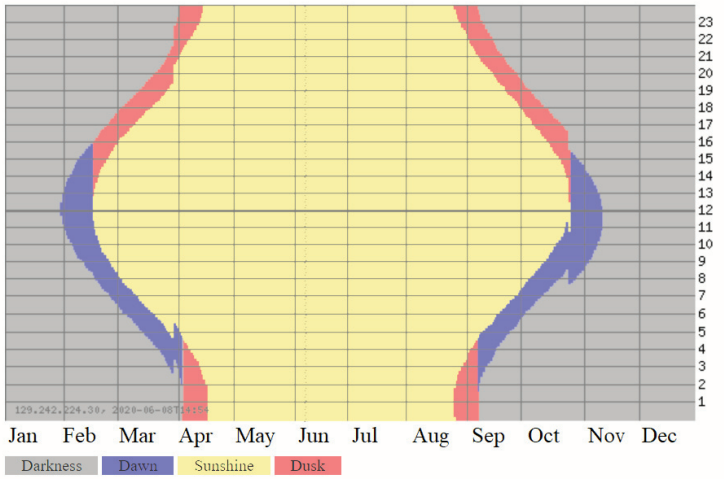


Figure 1: Yearly sun graph from the IsA station, the location for this study. The x-axis show the months of the year, while the y-axis shows the time of day (following a 24 h time cycle). The colours indicate which periods are dominated by sunshine, darkens, dusk and dawn. The graph was adapted from Gaisma.com.

for more than 50% of the annual primary production around Svalbard and in the Northern Barents Sea (E. Sakshaug 2004), and is followed by a successional postbloom stage that lasts throughout the summer. This provides the nutritional foundation for a parallel boom in the abundance of heterotrophic bacteria. *Bacteroidetes*, especially *Flavobacteria*, have often been found to show the greatest increase, together with *Gammaproteobacteria* (Wilson et al. 2017; Alonso-Sáez et al. 2008; Teeling et al. 2012; Lovejoy, Pierre E. Galand, and David L. Kirchman 2011). Despite the increase in light availability few *Cyanobacteria* have been detected in Arctic marine waters (Waleron et al. 2007). However, a recent study by Paulsen et al. 2016 found relatively high amounts of *Synechococcus* at 79°N.

Despite this flurr of activity during late spring and summer, studies have shown prokaryotes to be active throughout the seasons, including the mids of winter (Alonso-Sáez et al. 2008; Garneau et al. 2008; Alonso-Sáez, Wallerc, et al. 2012; Connelly et al. 2014). After the phytoplancton bloom dies away in late summer, chemolithoautotrophic processes come into play during the dark winter months (Christman et al. 2011; Alonso-Sáez et al. 2008; Grzymski et al. 2012) building up nitrate concentrations that will fuel the next spring bloom (Connelly et al. 2014). Chemoautotrophic ammonia oxidizing

Archaea (AOA) has been recognized as the major contributors to marine microbial ammonia oxidation (Wuchter et al. 2006; Valentine 2007), with Thaumarchaeote being one of the most notable members (Müller et al. 2018; Wilson et al. 2017; Pedneault et al. 2014; Alonso-Sáez, Wallerc, et al. 2012). During summer however, their abundance has been found to diminish in surface waters (Kalanetra, Bano, and Hollibaugh 2009; Pedneault et al. 2014; Alonso-Sáez, Wallerc, et al. 2012). The cause for this annual dissaperance has been hypothesized to be due to photoinhibition of ammonia oxidation (Guerrero and R. Jones 1996; Murray et al. 1998; Mincer et al. 2007; Merbt et al. 2012; Pedneault et al. 2014).

In general, *Bacteroidetes* and *Proteobacteria* have been found to be the most abundant phyla of the prokaryotic community in the Arctic (Lovejoy, Pierre E. Galand, and David L. Kirchman 2011; David L. Kirchman, Cottrell, and Lovejoy 2010; Comeau et al. 2011; Bowman et al. 2012; Ortega-Retuerta et al. 2013; Yergeau et al. 2017; Gorrasi et al. 2019; Wilson et al. 2017). Of the *Bacteroidetes*, *Flavobacteria* is often the most abundant (Gorrasi et al. 2019; Ghiglionea et al. 2012) with *Polaribacter* being one of the key members (Gorrasi et al. 2019; Lovejoy, Pierre E. Galand, and David L. Kirchman 2011) especially in coastal waters (Ghiglionea et al. 2012). Amongst the *Proteobacteria*, the *Gammaproteobacteria* and *Alphaproteobacteria* make out the greatest proportion. In most studies however, *Gammaproteobacteria* have been found to be the most abundant, which stands in contrast to more temperate regions where *Alphaproteobacteria* are usually more dominant (Wilson et al. 2017; Ghiglionea et al. 2012; Malmstrom et al. 2007). The ubiquitous SAR11 has also been detected in several studies as a commonly represented group of the *Alphaproteobacteria*, being particularly abundant in deep waters (Ghiglionea et al. 2012; Pierre E Galand, Potvin, et al. 2010; Lovejoy, Pierre E. Galand, and David L. Kirchman 2011; Malmstrom et al. 2007; Kraemer et al. 2020). Another bacterial phylum previously found to be abundant is *Verrucomicrobiales* (Lovejoy, Pierre E. Galand, and David L. Kirchman 2011), recently at lower depths in August on the west coast of Svalbard (Wilson et al. 2017). Amongst the Archea *Crenarchaeota* have previously been found to be the overall most abundant phyla, together with a decent proportion of *Euryarchaeota* in coastal surface waters (Pierre E Galand, Casamayor, et al. 2012). Recently however, the *Thaumarchaeota* has been established as a a new Archeal phylum separate from the *Crenarchaeota* (Brochier-Armanet et al. 2008). Today *Thaumarchaeota* is recognized as the most abundant Archeal phylum in the Arctic (Müller et al. 2018; Wilson et al. 2017).

1.2 Isfjorden

Arctic fjords are systems of special interest for several reasons. Most importantly they work as a border region between land and the open water, exchanging water masses between the continental shelf and inner coastal regions. This works as an important transport of temperature and nutrients between these vastly different habitats. The extreme seasonal variations of the high Arctic greatly affects this exchange (F. R. C. F. Nilsen et al. 2010).

Isfjorden is the largest fjord system in Svalbard far to the north of Norway. During summer and autumn months glacial melt water and river run-off provide a big influx of fresh water heavily saturated with sediments. This affects the density and turbidity of the inner water of the fjord and also lead to an unusually high increase in particulate organic carbon and other nutrients compared to that of temperate fjords (Wiedmann et al. 2016). Sea ice formation during winter lead to brine production and an increase in water density. Ice polynyas further increase this effect which makes it one of the most important contributions to fjord-shelf exchange of water masses (F. Nilsen et al. 2008). Adventfjorden, the site for this study, has commonly been an area of such sea ice formation. However, since 2007 this small inner fjord of Isfjorden has been mostly ice-free (Wiedmann et al. 2016), the reason being increased inflow of warm Atlantic water to the fjord (Skogseth et al. 2020).

Outside the mouth of Isfjorden the warm Atlantic water of the West Spitsbergen Current (WSC) meets the cold Artic water of the East Spitsbergen Current (ESC) (see Figure 2). The high difference in density between these two water masses prevent them from mixing, setting up a barrier with the cold ESC snaking its way along the west coast of Spitsbergen preventing access to the warm water of WSC. However, the unique topography of the seabed right outside Isfjorden causes these two currents to mix (see Figure 2). Normally the mixing of shelf water and inner fjord water is greatly limited by the so called geostrophic control due to density differences between the fjord water and outer water. However, this unique topography also facilitates an increased exchange with the fjord system (F. R. C. F. Nilsen et al. 2010). Isfjorden is also a very broad fjord, allowing the corrollis effect to increase the mixing of the fjord water masses laterally by rotational currents (F. Nilsen et al. 2008). The overall result is that the water of Isfjorden is generally well exchanged with that of the outer sea. However, the degree to which this happens and how much the WSC or ESC contribute varies across the seasons

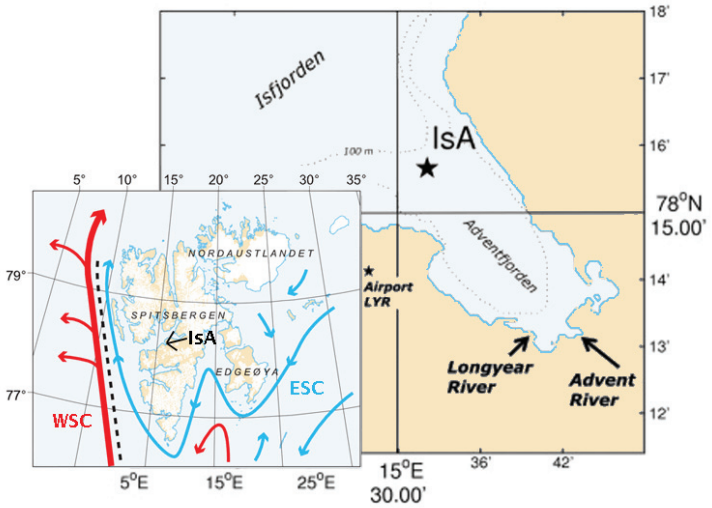


Figure 2: Map showing the location of IsA station (78°15.32 N, 15°32.04 E). The West Spitsbergen Current (WSC) and East Spitsbergen Current (ESC) are outlined. Adapted from Svendsen et al. 2002 and Wiedmann et al. 2016.

and from year to year. In summer, warm Atlantic water generally dominates, while in winter cold Arctic water dominates. The biggest driver for this is density changes, as discussed above. The last decades however, there has been an increased inflow of Atlantic water to the fjord, gradually increasing the mean sea surface temperature and mean air temperature by advection (Skogseth et al. 2020). This is thought to be one of the most important local drivers for climate change.

The sampling station used for this study is IsA (78°15.32 N, 15°32.04 E), located at the mouth of Adventfjorden in the inner part of Isfjorden (see Figure 2). This station is part of a standard time-series where water samples have been taken on a monthly basis together with various physiochemical measures. This has enabled the detailed study on both short term and long term changes in water masses and the impact on the microbial community . The biological focus so far however, has mainly concerned protists (Marquardt et al. 2016). In order to get a holistic understanding of the system and better predict the impact global warming, prokaryotes has now been included for the first time.

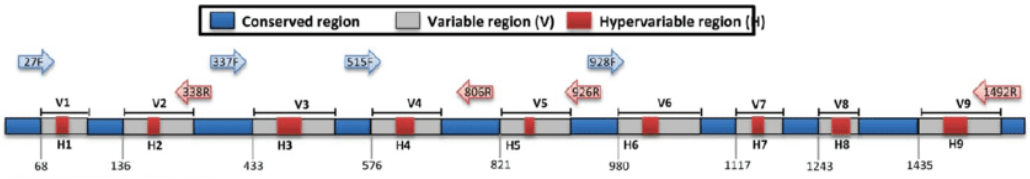


Figure 3: Illustration of the 16S rRNA gene with conserved, variable and hypervariable regions indicated. Common primers for barcode amplification are also shown. For this study the 515F(B) and 926R primers were used to amplify the v4 and v5 regions. The figure was adapted from Shahi, Freedman, and Mangalam 2017.

1.3 Metabarcoding

The emergence of high throughput sequencing (HTS) has allowed for the development of metabsrcoding as a swift and cost-efficient technique for identifying which organisms are present in an environmental sample. For bacteria and archea this is often the only available method as morphological analyses are virtually impossible with normal microscopy and culturing can be very difficult for many species (Kennedy et al. 2010; Kress et al. 2015; Pedrós-Alió et al. 2018).

In metabarcoding all DNA is first extracted from an environmental sample, and then amplified by PCR targeting a region commonly referenced to as the "barcode". The amplified DNA is then sequenced by HTS and analysed (Pedrós-Alió et al. 2018; Madigan et al. 2014). A barcode is a sequence found in the genome of all the organisms of interest in the study with enough similarity that it can be amplified with a general primer set, but enough difference so that different organisms will have unique sequences. In order to identify which sequences belong to which organisms the amplified barcodes are compared to a reference database. For bacteria and archea the barcode is often chosen as subparts of the DNA sequence encoding the 16S rRNA of the 30S small subunit of their ribosomes. This gene is known to be highly preserved between all prokaryotes, but interspersed with some variable regions, making it ideal for metabarcoding (Figure 3).

When analysing the raw data from the HTS, the reads are first filtered; removing low quality sequences and potential amplification errors (Pedrós-Alió et al. 2018). The sequences are then traditionally clustered into so called Operational Taxonomic Units (OTUs). While the concept of species in the world of microbiology is hard to pinpoint and a much discussed subject, the argument is that very similar sequences originate from organisms that are

phylogenetically very close (Madigan et al. 2014). These are also likely to possess very similar properties and behaviour and can thus be considered the same "unit". These are historically thought to be much easier to operate with because it simplifies the data by reducing the number of groups or *units*. Traditionally the threshold for considering two sequences part of the same OTU is a sequence similarity of 97% or higher. Some have strongly argued for a similarity threshold of 99% as the biologically most meaningful (Edgar 2018). In recent years however, a new practice of zero radius OTUs (zOTUs), also referred to as Exact Sequence Variants (ESVs), has appeared (Callahan, Paul J McMurdie, and S. P. Holmes 2017). In this method no clustering is done except for exactly similar sequences. In other words a zOTU is roughly the same as an OTU with a similarity threshold at 100%.

There are two major arguments for using zOTUs (Callahan, Paul J McMurdie, and S. P. Holmes 2017; Porter and Hajibabaei 2018). The first is that this introduces the least bias to the biological interpretation. An zOTU is simply the occurrence of the an exact sequence. This is as fine a resolution as can be gained, of which the quality will depend directly on the barcode of choice and the primers used for amplification of this. In many ways the whole "species" discussion is thus avoided. Secondly, this makes the result comparable across different samples and studies. In order to compare OTUs across samples all samples have to be analysed at once under the same conditions, clustering OTUs across the different samples. A recent study by Glassman and Martiny 2018 gave the surprising result that clustering into OTUs instead of zOTUs not significantly simplified the data, thus removing one of the strongest arguments for using OTUs. In this thesis we will refer to OTUs and zOTUs in general as Taxonomic Units or "TUs" (meaning both or either).

A TU is simply the occurrence of a specific sequence in an environmental sample. Depending on the resolution of the chosen barcode and degree of clustering, a single TU can encompass several different types of organisms, usually with a common evolutionary past, or two different TUs might originate from the same group of organisms, representing slightly different strains, or even different copies of the barcode from the same genome (Pedrós-Alió et al. 2018).

Despite the clear advantages and accessibility of microbial metabarcoding several limitations still remain. First and foremost matabarcoding will only provide information about the relative abundance of certain TUs to that of all target TUs in the given sample. This means that if one TU get particularly

abundant, the relative abundance of another TU might decrease, despite keeping the same absolute abundance, or simply increasing less.

This problem can be faced by scaling the relative abundances with the absolute abundance of all target TUs in the community. For microbial communities scaling by cell count by flow cytometry has been pointed out as the best way to do so (Cao et al. 2016). However, several limitations still remain for such methods. In particular, the relative abundances of the TUs might not correctly reflect that of the target organisms. Depending on the barcode of choice, different TUs might have different copy numbers in their original genome. This is particularly common for the 16S rRNA gene off prokaryotes. Depending on the primers of choice the barcode might also be more easily amplified for certain organisms than for others, also leading to inflation or deflation of the TU relative abundances (Pedrós-Alió et al. 2018).

1.4 Aim of study

The aim of this study was to give a detailed characterization of the prokaryote community found at the IsA station in Isfjorden throughout a full year, thereby enabling a better understanding of the complete microbial community in this area and possible impacts of a changing ocean. To the authors knowledge, this also provides the highest temporal resolution of an Arctic prokaryotic community from the same location (as of this date), which enables a more in-depth understanding of the seasonal dynamics of such a system.

More specifically the objectives were to 1) identify which prokaryotic organisms dominate at different parts of the year, 2) how the prokaryotic community structure change with the seasons, and 3) how water mass, depth and nutrient availability drives this change. Most importantly, this study provide the foundation for future studies of temporal community dynamics, as a comprehensive reference and data source.

2 Materials and Methods

2.1 Sampling and filtration

Seawater samples were taken on a monthly basis at the IsA station during the full year of 2019. Water from the upper part of the water column (15 m depth) as well as close to the bottom (75 m depth) were taken using a Niskin bottle. In addition, surface water (0 m) was collected using a metal bucket. For all three depths about 200 mL of water were put directly in an acid washed bottle (10 % HCl, 24 h) for later nutrient analysis. Some water was also stored in containers rinsed with distilled water for *chlorophyll a* and flow cytometry measurements. Around 3 L of seawater from both 15 m and 75 m were kept in separate containers also rinsed with distilled water beforehand. These were later used for filtration, DNA extraction and metabarcoding targeting both Bacteria and Archaea.

In addition to collecting water samples, physicochemical water column profiles were taken using a conductivity-temperature-depth sensor (CTD). This was usually attached underneath the Niskin bottle going down to 75 m depth.

Depending on the vessel used for taking the samples filtrations were either done below deck or back at the lab at UNIS in Longyearbyen. It normally took around 1.5 h from leaving the sampling location until the first filtration was started at this lab. To keep the conditions as stable as possible for the microorganisms the samples were kept dark in an isothermal storage box. If weather conditions were good pre-filtration of the DNA samples (using a 10 µm filter) would be done on deck beforehand.

2.2 Chlorophyll a measurement

Seawater collected at all depths (0 m, 15 m and 75 m) was filtered using GF/F glass microfibre filters (0.7 µm) and vacuum pump, and the exact volume noted (normally 300 mL). In addition a separate filtration was done using 10 µm isopore membrane polycarbonate filters (Millipore, USA) and gravitational pull. This was done to collect particles of different size fractions so that the relative contribution of smaller microorganisms to the total *chlorophyll a* (Chl a) concentration could be estimated. The filters were wrapped in aluminium foil and immediately stored at -80 ℃ until further processing.

Chl a was later extracted from the filters by soaking each in 10 mL methanol for 24 h in a fridge (4 ℃). The solution was then filtered through a 0.22 µm filter and fluorescence determined using a 10-AU-005-CE Fluorometer (Turner, USA). After measuring total Chl a from each filter, non-degraded Chl a was degraded by the addition of hydrocloric acid (HCl, 10%, 2 drops), and fluorescence measurements were repeated. Acid corrected Chl a concentrations ($C_{corr.}$) were calculated using the following formula:

$$C_{corr.} = (C_0 - C_{acid})\frac{V_{extrat}}{V_{filtrate}}\sigma \quad (1)$$

Where C_0 and C_{acid} denotes the measured Chl a concentration before and after acid treatment respectively, $V_{extract}$ the volume of methanol used for extraction, and $V_{filtrate}$ the volume of seawater originally filtrated. σ is a correction factor calculated from a diluted standard of pure Chl a used to calibrate the fluorimeter. In this study it was determined to be 1.7.

2.3 Nutrients measurement

Seawater from all three depths (0 m, 15 m and 75 m) were kept in acid washed bottles (10 % HCl, 24 h) and stored at -20 ℃. The inorganic nutrients content of the seawater was later determined using a QuAAtro39 (SEAL Analytical) set up with four channels measuring nitrate, nitrite, phosphate and silicate. For each 4 standards were used to calibrate the instrument.

2.4 Flow cytometry

Seawater from 15 m and 75 m depth were each put in two cryotubes (2 x 1.8 mL) and fixated by adding glutaraldehyde (GLA, 50%, 36 µL, 1% final conc.). The tubes were then stored at -80 ℃ until later analysis.

The samples were sent to the Department of Biological Sciences (BIO) at Bergen University, and the heterotrophic bacterioplancton and viral abundance enumerated using a BD FACS Calibur according to the protocol in Robinson et al. 1999 updated with respect to Brussaard 2004 and Paulsen et al. 2016.

2.5 DNA extraction and barcoding

Seawater from 15 m and 75 m depth were filtered in two steps: First a pre-filtration step using a 10 µm mesh (KC Denmark) and gravitational pull to remove larger particles, then a filtration using a 0.22 µm filter (47mm Durapore polycarbonate membrane filters, Millipore) and a vacuum pump. The latter filter was cut in two and each filter part kept in a separate cryotube at -80 ℃ until further analysis.

The filter halves were later thawed and DNA extracted using QIAGEN's DNeasy® Plant Mini Kit (250). Each filter half were extracted in separate batches as a precaution in case of contamination. The extraction followed the kit protocol with the exception of the initial disruption step. The filter halves were first cut into 4 smaller pieces and put in an eppendorf tube. Buffer AP1 was added (400 µL) together with zirconium beads (200 µm, molecular grade, OPS diagnostics, half a tube) and the samples immediately beat with a bead-beater (1 min at 30 Hz for both orientations). The buffer was then transferred to a new tube. Buffer AP1 was again added (400 µL) and the beating repeated. The buffer was then pooled in the same tube as the previous one. This effectively doubled the volume of each sample, and in every subsequent lysis and washing steps the added volumes were also doubled with respect to the standard protocol. The final elution step was repeated twice using a volume of 75 µL, resulting in a total extract volume of approximately 150 µL.

The extraction was controlled using PCR targeting the v16 region of prokaryote 16S rRNA encoding gene followed by agarose gel electrophoresis. For the PCR the Thermo Scientiffic PCR kit was used according to the standard protocol with 25 µL reaction volume using one negative control. The forward and reverse primer used were Bact 341F (5'-CCT ACG GGN GGC WGC AG-3') and Bact 785R (5'-GAC TAC HVG GGT ATC TAA TCC-3') respectively (Klindworth et al. 2013). A melting temperature of 95 ℃, annealing temperature of 55 ℃ and elongation temperature of 72 ℃was used. In total 30 cycles were run (95 ℃ - 5'; 30x (95 ℃ - 40"; 55 ℃ - 2'; 72 ℃ - 1'); 72 ℃ 7'; 4 ℃ - inf). The PCR products were run on an agarose gel (1.5%, 1x TAE buffer, 1x GelRed) using the Thermo Scientific gel kit. Successful extraction was indicated by a strong band at about 300 bp for the filter samples and no visible band for the DNA extraction controls.

In the case of successful extraction of both filter halves, both extracts were pooled together and the DNA concentration determined using a Qubit™ 4

(ThermoFisher Sci.) fluorimeter according to the standard protocol. Sub samples of the extracts (10 µL) were put in separate PCR tubes and sent to the Integrated Microbiome Resource (IMR, Dalhousie University in Halifax, Nova Scotia, Canada) for amplicon sequencing using the Illumina MiSeq sequencing platform. The initial PCR amplification was done by the IMR using their standard forward and reverse universal primers, 515FB (5'-GTG YCA GCM GCC GCG GTA A-3') and 926R (5'-CCG YCA ATT YMT TTR AGT TT-3') respectively, targeting the v4 and v5 region of the 16S rRNA encoding gene of bacteria and archaea (Parada, Needham, and Fuhrman 2016).

2.6 Bioinformatics analysis

For the processing of the raw metabarcode sequences as obtained from the Integrated Microbiome Resource (IMR), *Usearch v11* was used (Edgar 2010). Two different analysis were conducted obtaining OTUs with 97% clustering and zOTUs respectively (Figure 4). Both started by stripping the primers and merging the pairs with a maximum difference in alignment of 15 bp and a minimum length of 350 bp in total (using the *fastq_mergepairs.pl* script). Quality filtering using a threshold of 1.0 were also done for all samples (using *fastq_filter.pl*). Next, the samples were pooled, de-replicated and the sequences sorted by the number of occurrences, in addition singletons were removed (*fastx_uniques-sortbysize.pl*).

For the OTU pipeline an implementation of the **UPARSE algorithm** (Edgar 2013) was then used to cluster the resulting sequences into OTUs *de novo* with a radius of 3, also removing chimera sequences (*cluster_otus.pl*). An OTU table was then produced, with the read number of reads of each OTU for the different samples, using an identity threshold of 97% (*usearch_global-otutab2biom.pl*).

For the zOTU pipeline the **UNOISE algorithm** (Edgar 2015) was instead used to predict correct biological sequences by removing sequencing errors and chimera sequences (using the *unoise3* command). The default of minimum 8 number of reads were used, discarding any sequences with lower number of reads, assuming them to be too noisy for interest in this study. An zOTU table was then produced, with the read number of each zOTU for the different samples, using the *otutab* command.

Taxonomic assignments were given to both OTUs and zOTUs by the SINTAX algorithm (Edgar 2016) using both the RDP 16S v16 (Cole et al.

2014) and SILVA 138 SSU, Ref NR99 (Quast et al. 2013) databases as a reference and a cutoff score of 0.80. The SILVA database showed far better coverage of the prokaryotic community compared to the RDP database (a maximum of 1% reads not assigned to any phyla in a sample versus 20%, respectively), and was therefore used as the primary reference database. The taxonomic assignments were translated into the more official format of the RDP database. For in depth study of single TUs a combination of RDP and SILVA were used to obtain as high level taxonomic assignments as possible. Where neither sufficed the NCBI Blast database was used. Which databases supported which assignemnts have been indicated. The RDP database is mainly based on well described and cultured prokaryotes and is therefore considered more reliable than the SILVA database, which includes environmental sequences. The NCBI is based on public user input of reference sequences and was considered the least reliable. It should be noted however, that all of the taxonomic assignments from using the SILVA database were consistent with that of the RDP database, for all of the most abundant TUs which were compared.

The resulting OTU and zOTU tables were further processed using the **vegan** (version 2.5-6) package for *R*. TUs assigned as Chloroplasts were removed, after quality assessment. Samples were then normalized by first finding the fraction of each TU for each individual sample and then multiplying so as to get as many "reads" per sample as the original sample with the lowest number of reads (so that all samples had the same number of reads). If we let M represent the TU table with the number of reads for m TUs over n samples, the new normalized matrix M^* can be expressed mathematically as:

$$M_n^* = \frac{M_n}{\sum_{i=1}^{m} M_{i,n}} \min_{j=1,\ldots,n} (\sum_{i=1}^{m} M_{i,j}) \tag{2}$$

Where M_n indicate the n^{th} column of M (results from sample nr. n) and $M_{i,j}$ entry number i on the j^{th} column of M (number of reads for TU nr. i in sample j).

According to Paul J. McMurdie and S. Holmes 2014 this is preferable to the traditional rarefaction normalization method. Both were tested, but no significant difference in the results of downstream analysis were noticed.

As a measure for alpha diversity of the samples Shannon entropy (i.e. exponentiated Shannon index) was chosen, as this have been indicated as

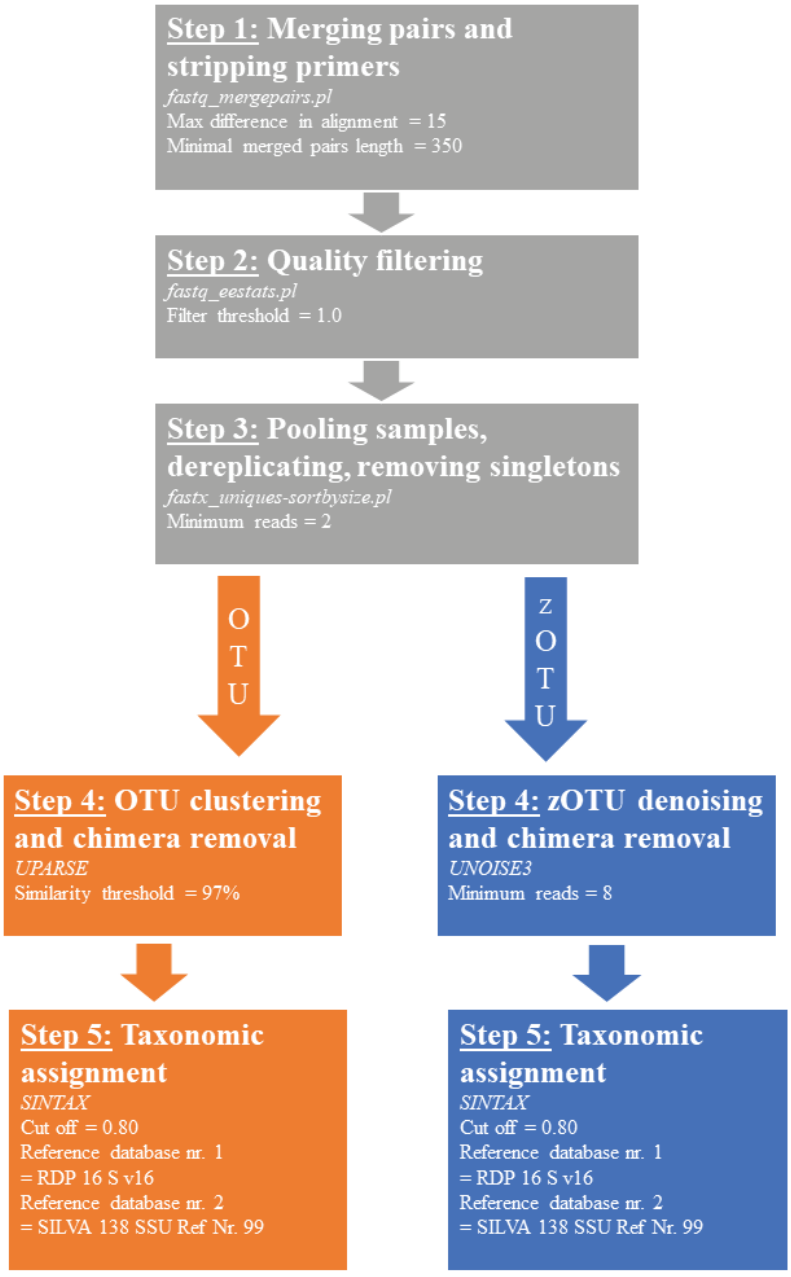


Figure 4: Flow diagram depicting the OTU and zOTU pipeline as done in *Usearch v11* (orange and blue respectively). Common steps are showed in grey.

one of the best existing indices for this purpose (Lucas et al. 2016), while also being closely related to the popular Shannon index. Put simple Shannon entropy quantifies the uncertainty in predicting the TU identity of a read that is taken at random from the sample. It thus reflects both the TU diversity and evenness of the reads in the sample; the higher the number of TUs and more even in abundance they are, the higher the Shannon entropy. In this measure both rare and abundant TUs are weighted by their particular abundance and are thus treated equally.

As a measure for beta diversity Bray-Curtis distance was chosen as this is both widely used and strongly supported as one of the most biologically meaningful indices (ibid.). For visualizing the Bray-Curtis distance between the different samples a cluster dendrogram was creating using the *pvclust* function from the **pvclust** (version 2.2-0) package for *R*. The default "average" method was used to create the tree and 10,000 bootstraps conducted to get the bootstrap probability for each branch. In addition an non-metric multidimensional scaling (NMDS) ordination plot was made with fitted environmental factors on top using the *monoMDS* and *envfit* function of the **vegan** package. NMDS was chosen in front of other ordination methods as this is one of the few common unconstrained techniques that allow for Bray-Curtis distance to be used directly (Kindt and Coe 2005). That NMDS is unconstrained means that the resulting plot tries to maximize the variance in the distance shown among the samples without biasing for environmental factors. These were fitted independently on top afterwards.

Correlation analyses of the Prokaryotic community was also conducted. This was done by simply finding the Pearson correlation between the different TUs across all the samples. This was then used to create correlation networks in *Cytoscape v3.8.0*. A cut off value of > 0.85 correlation was found to reveal the clearest structure when inspecting all TUs together. Smaller networks of the most abundant TUs were also extracted, this time with cut off values at > 0.50 and $<$ -0.50.

3 Results

3.1 Sequence results and quality of data

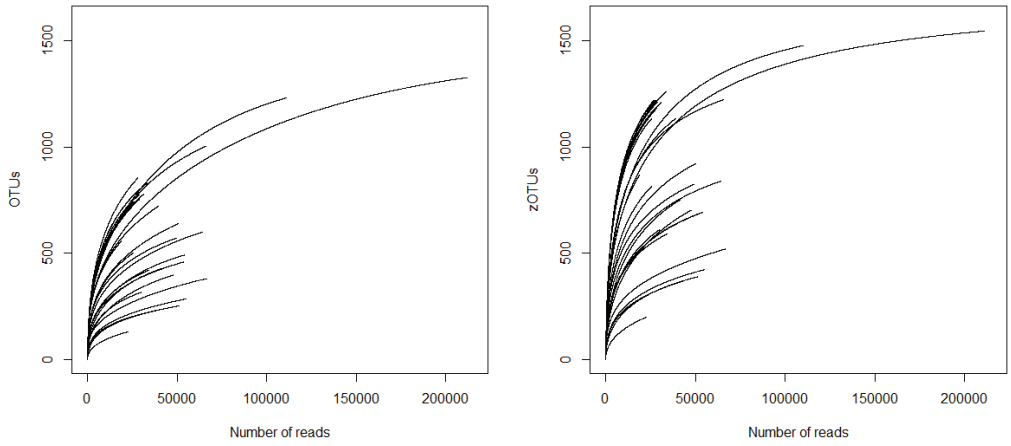
The total number of reads after merging pairs and stripping primers was 1.54 million. Which was reduced to 1.44 million after quality filtering. The proportion of reads kept after filtration was stable for all samples ranging from 93% to 94%. The number of unique reads were 225 000 of which 168 000 singletons were removed. The final OTU and zOTU table contained a total of 1,354,641 and 1,349,733 reads respectively, comprising 2370 OTUs and 1928 zOTUs. After normalization this was reduced to 386,038 and 384,859 "reads" (13,821 and 13,773 "reads" per sample).

Rarefaction curves for the zOTU pipeline showed a steeper initial growth and plateaued higher than for the OTU pipeline (Figure 5a and 5b). This showed the zOTU pipeline to be more sensitive, detecting more TUs per sample. Both pipelines had non-zero end slopes for all samples, with a maximum of 0.017 and 0.023 respectively. This indicate incomplete sequence coverage, but were deemed sufficiently good for this study.

By pooling together all the 28 samples, they can be considered one combined sample of the IsA station. The end slopes of the rarefaction curves for this combined sample were zero for both pipelines (Figure 6a and 6b). However, the zOTU pipeline reached this saturation point much earlier than that of the OTU pipeline.

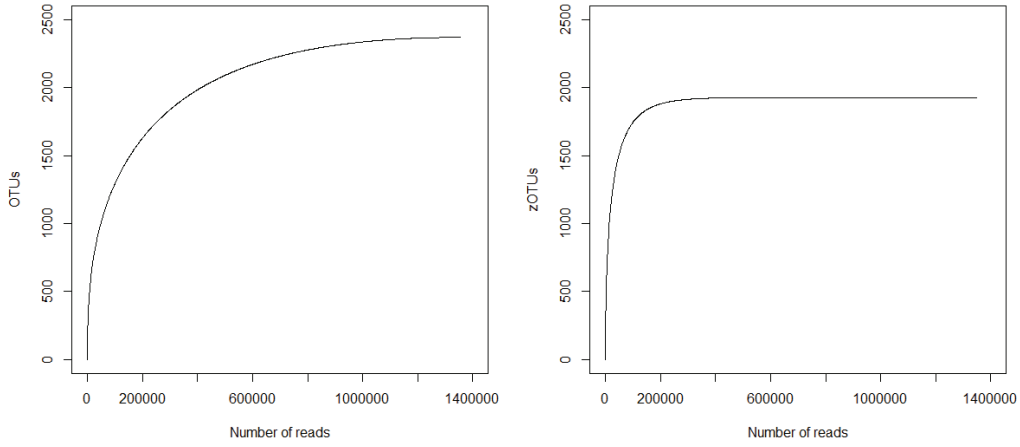
Accumulation curves were also made showing the number of TUs accounted for when adding one by one sample to the "combined IsA sample". By doing this 1000 times, adding the samples in a random order each time, the mean number of TUs and standard deviation were found (Figure 7a and 7b). The OTU pipeline showed a relatively high standard deviation even when most of the samples were accounted for. The zOTU pipeline on the other hand reached near-zero deviation when only half of the samples were accounted for, indicating around 14 samples to suffice for covering the whole IsA community.

The rank abundance curves of the combined IsA sample revealed a difference in the relative species abundance of the OTU and zOTU pipeline (Figure 8a and 8b). As expected both curves followed an approximately lognormal distribution, with a few highly abundant TUs and the majority having an intermediate or low abundance. However, the mean number of reads for the TUs of the OTU pipeline was lower than that of the zOTU pipeline. The



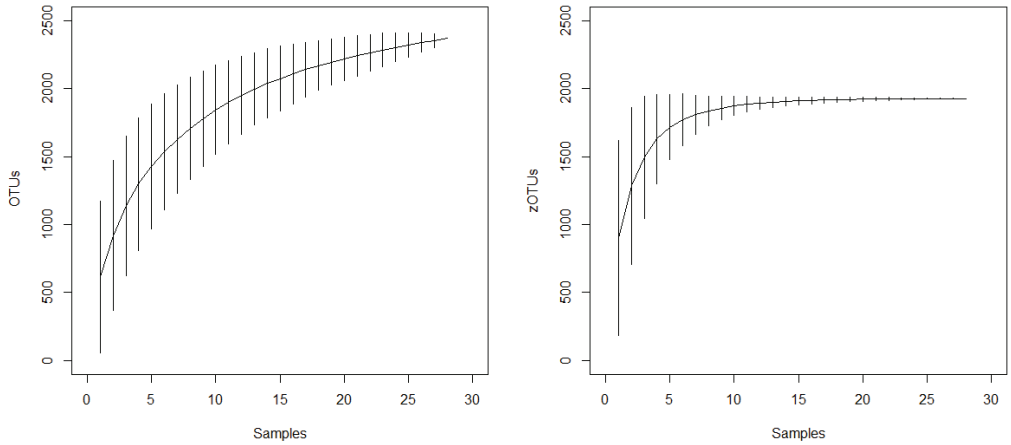
(a) OTU (97%) **(b)** zOTU

Figure 5: Rarefaction curves for the individual samples before normalization. Produced using the *rarecurve* function of **vegan** with a step size of 20 and default subsample size.



(a) OTU (97%) **(b)** zOTU

Figure 6: Rarefaction curves for all samples combined. Produced using the *rarecurve* function of **vegan** with a step size of 1,000 and subsample size of 10,000.



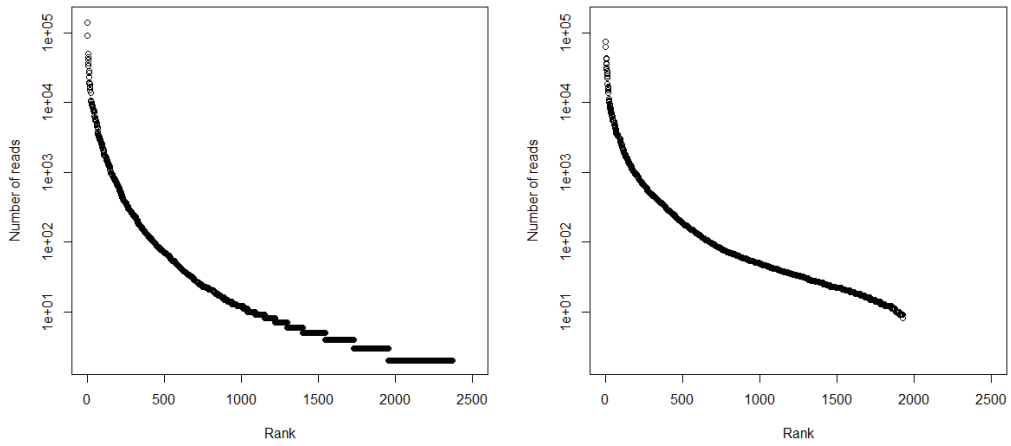
(a) OTU (97%) **(b)** zOTU

Figure 7: Ugland's accumulation curves for all samples combined. Produced using the *specaccum* function of **vegan** with the method set to "random" and number of permutations 1000. The vertical lines show the unconditional standard deviation.

end of the zOTU curve also shows the beginning of a drop, but is cut short by the minimum abundance limit of 8.

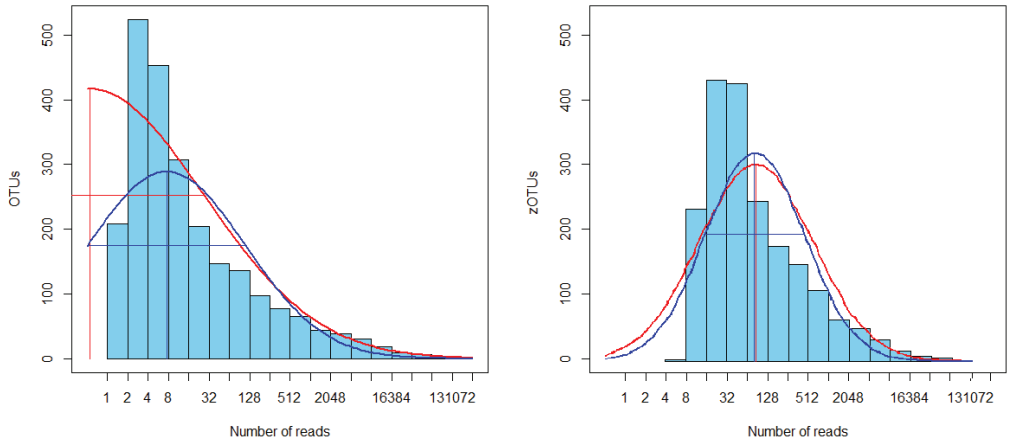
The evenness of the community can also be inspected by making Preston plots. This displays the same information as the rank abundance curves, but structured as histograms of the TU abundance with exponentially increasing intervals (Preston 1948). This is also expected to follow a lognormal distribution, but will end up looking bell-shaped or normal due to the scaling of the x-axis. If all the TUs of the community are accounted for, the entire bell-shape should in theory be visible. This can be used to estimate the number of species *not* accounted for in a real dataset, so-called "veiled species". In this study this was found to be 440 using the OTU pipeline and only 2 using the zOTU pipeline (using Preston's truncated lognormal model).

From the results above the zOTU pipeline was concluded to produce the best dataset (see Discussion), and was therefore used for the remaining part of the study. It should be mentioned however, that the data from the OTU pipeline was analysed in parallel and produced very similar or identical results (see Appendix A). Hence the choice of pipeline did not affect the final conclusion.



(a) OTU (97%) **(b)** zOTU

Figure 8: The ranked number of reads of the different TUs for the OTU and zOTU pipeline.



(a) OTU (97%) **(b)** zOTU

Figure 9: Preston plots with fitted curves for the combined samples. Red curve shows the fit using Preston's lognormal model, while the blue curve shows the fit using Preston's truncated lognormal model using the *prestonfit* and *prestondistr* function of **vegan**.

3.2 Community seasonality and influence of environmental factors

3.2.1 Water masses

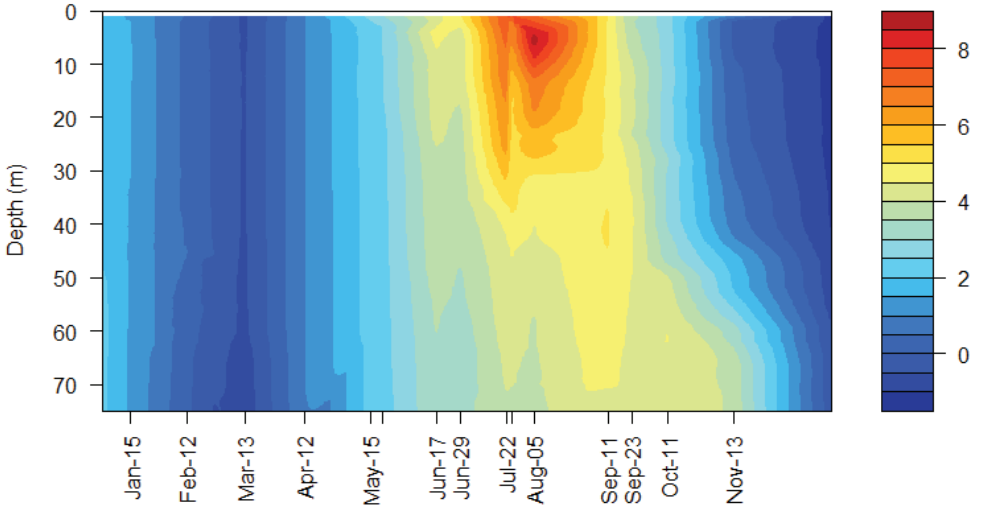
The beginning of 2019 was characterized by cold saline water. During late spring and summer there was a gradual increase in temperature spreading from the surface down to the deeper layers (Figure 10a). In July a rapid decrease in salinity in the upper layers was also observed, gradually spreading down the water column (Figure 10b). The HYCOM global oceanography model revealed a change in the ocean currents with warm Atlantic water from the WSC reaching into Isfjorden and the IsA station, replacing the cold Artic water of the ESC (Figure 11a to 11k). During late autumn the water temperature dropped back close to original levels (Figure 10a) and the ESC again dominated along the mouth of Isfjorden (Figure 11k to 11n). The salinity also rose during this period, but displayed a second drop in January 2020 (Figure 10b). Unfortunately no CTD data are available from the December 12th sampling.

3.2.2 Nutrient concentrations

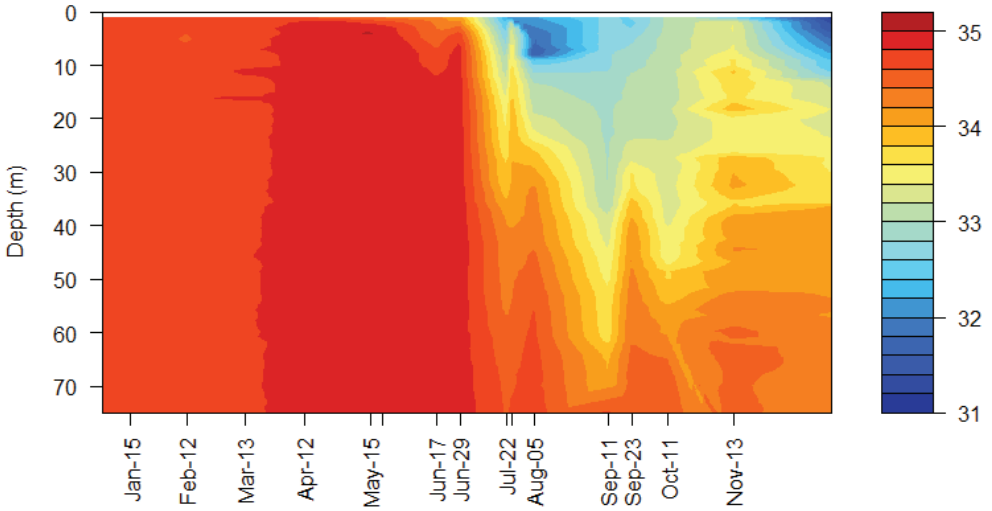
The nutrient concentrations of nitrate, nitrite, phosphate and silicate were relatively high during winter and early spring, but experienced a substantial drop in May-June (Figure 12). This was particularly prominent at 15 m depth where the highest and lowest nutrient values were measured on May 21st and June 17th, respectively. After this initial drop the concentrations rose slightly, but stayed low until the last measurement on November 13th. The nitrate concentrations had a second drop on September 11th at 15 m. At 75 m depth the drop in all nutrient concentrations during spring and second drop in nitrate concentrations were observed around one month earlier and were not as prominent. The rise in nutrient concentration during autumn (Oct-Nov) however, was more pronounced.

3.2.3 Heterotrophic cell count an viruses

The heterotrophic cell count showed a more than a five-fold increase in just 1-2 months during late spring (Figure 12). This started around a month earlier for 15 m depth with a more gradual initial increase. At 75 m depth the cell count was highest during May and June with a subsequent gradual

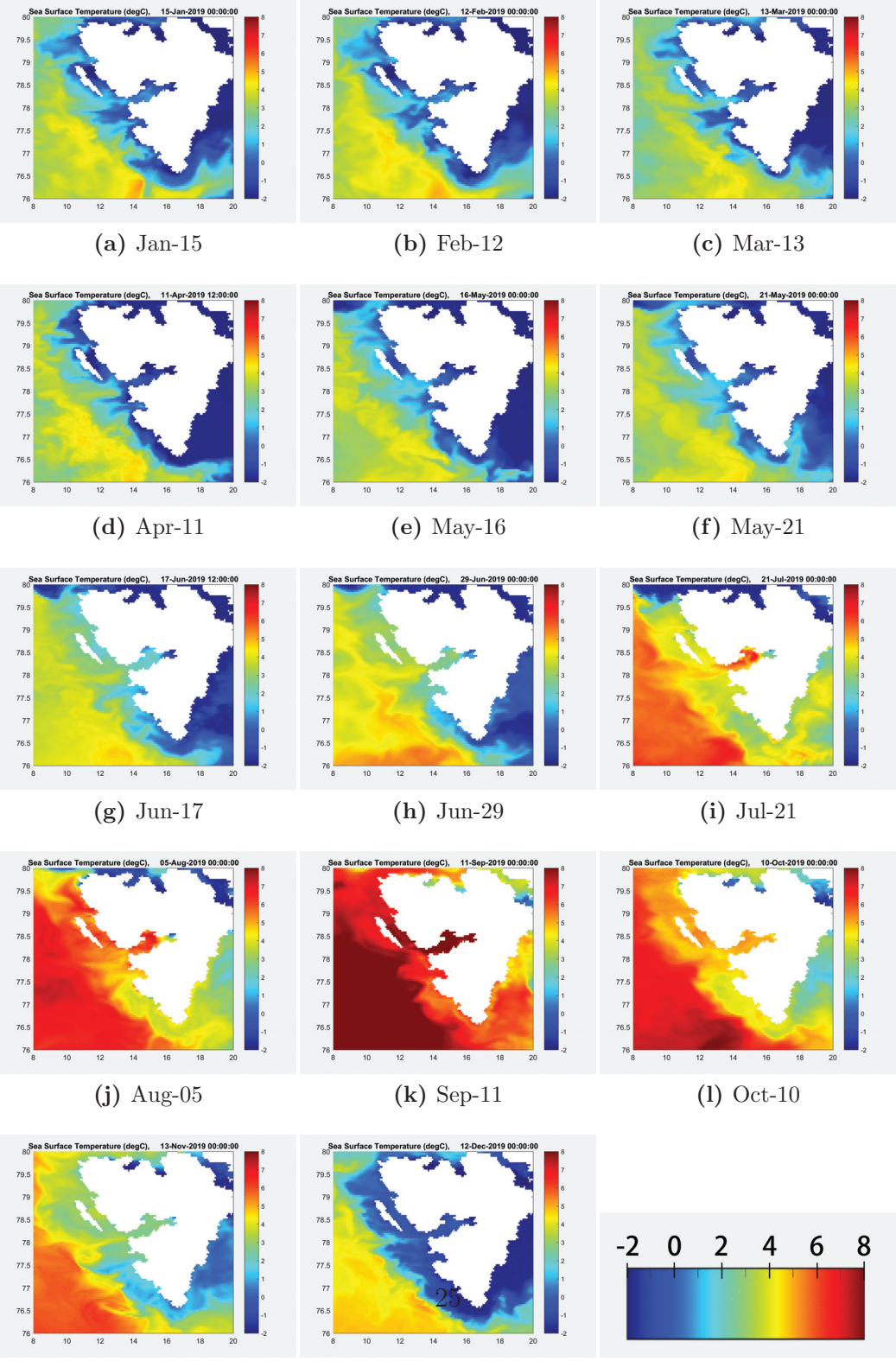


(a) Temperature (℃)



(b) Salinity (PSU)

Figure 10: Temperature (a) and salinity profile (b) throughout 2019. Sampling dates with CTD probes are indicated on the x-axis. The rest were filled in based on linear interpolation using the *interp* function of the **akima** package for *R*.



(a) Jan-15 **(b)** Feb-12 **(c)** Mar-13

(d) Apr-11 **(e)** May-16 **(f)** May-21

(g) Jun-17 **(h)** Jun-29 **(i)** Jul-21

(j) Aug-05 **(k)** Sep-11 **(l)** Oct-10

(m) Nov-13 **(n)** Dec-12 **(o)** Scale (°C)

Figure 11: Surface temperatures around Spitsbergen in 2019 based on satellite data as provided by the HYCOM consortium. The figures were made using the *OPeNDAP* package for **MatLab**.

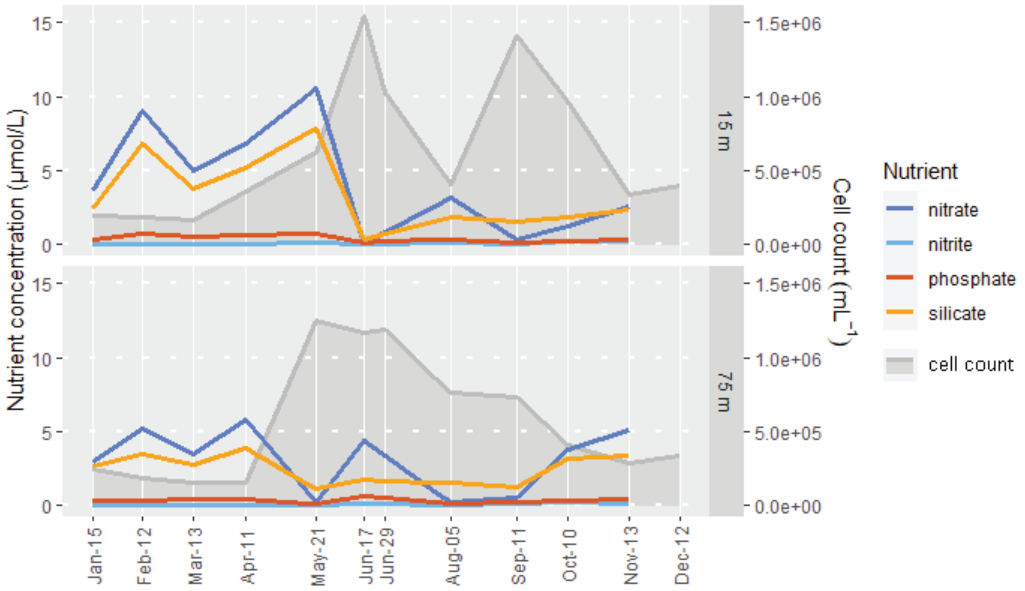


Figure 12: Nutrients concentrations and heterotrophic cell counts measured throughout 2019. Sampling dates are indicated on the x-axis. Nutrients were not measured on June 29th and December 12th.

decrease back to near original levels in November. At 15 m depth the peak cell count on June 17th was followed by a rapid decline, reaching a minima in August. A second peak in abundance of similar magnitude was observed on September 11th, before a quick decline to near original winter levels. The peaks two in cell count at 15 m coincided with the minimas in nitrate concentrations.

The number of counted viruses increased steadily throughout the year for both depths (see Appendix A, Figure A.1). This was most prominent for small virus, going from an abundance of around 5 x10^6 mL^{-1} in January to 2 x10^7 mL^{-1} in December.

3.2.4 Chlorophyll a concentrations

During winter there were almost no detectable levels of phytoplanton biomass. In late spring and early summer a rapid increase was observed for both depths, reaching high levels earlier at 15 m (Figure 13). The smallest microorganisms (< 10 µm) dominated at the very beginning of the bloom. However, at the peak it was the larger organisms (>10 µm) that contribute most

of the photosynthetic biomass. During the autumn the smaller microorganisms again dominated with a large contribution to the Chl a concentration, until the bloom eventually subsided with the light into winter. July, August and September the phytoplancton biomass was very unevenly distributed between the depths, with low amounts at 75 m.

3.2.5 Alpha diversity

The alpha diversity of the prokaryotic community showed a clear change through the seasons. The same trend was evident at both depths (Figure 13). The Shannon entropy was relatively high during winter, with the highest values observed on February 12th for both 15 m and 75 m depth. During spring there was a significant drop, with the lowest value observed on June 17th, also for both depths. After that the alpha diversity gradually increased during the autumn back to near original levels. The drop during spring correlates well with the strong increase in observed *chlorophyll a* (Chl a) concentration during April and May (Figure 13).

3.2.6 Beta diversity and correlation to environmental factors

The beta diversity revealed a clearly structured change in community composition between the different seasons. A cluster dendrogram based on the Bray-Curtis distance between the samples showed that samples in close temporal proximity of each other tended to cluster together (Figure 14). There was a clear division between the winter and summer months, and especially May stood out with highly supportive bootstrap probability. Although the winter samples from early and late 2019 clustered separately, they were still more similar than to the spring and summer samples. While samples taken at the same date generally clustered together independent of depth, the October 10th samples were an exception. Here the 15 m sample clustered with the summer samples, while the 75 m sample clustered with the winter samples. This marks October as a transition month.

An NMDS ordination suggested a cyclic annual pattern consistent with the different seasons (Figure 15). The largest gap being between May and June samples. When correlating the measured environmental factors to the ordination, all factors were found to be significant ($p < 0.05$) except for phosphate and nitrite, which had a p-value of 0.06 and 0.14 respectively (Figure 15). Nitrate, phosphate and salinity all pointed towards spring,

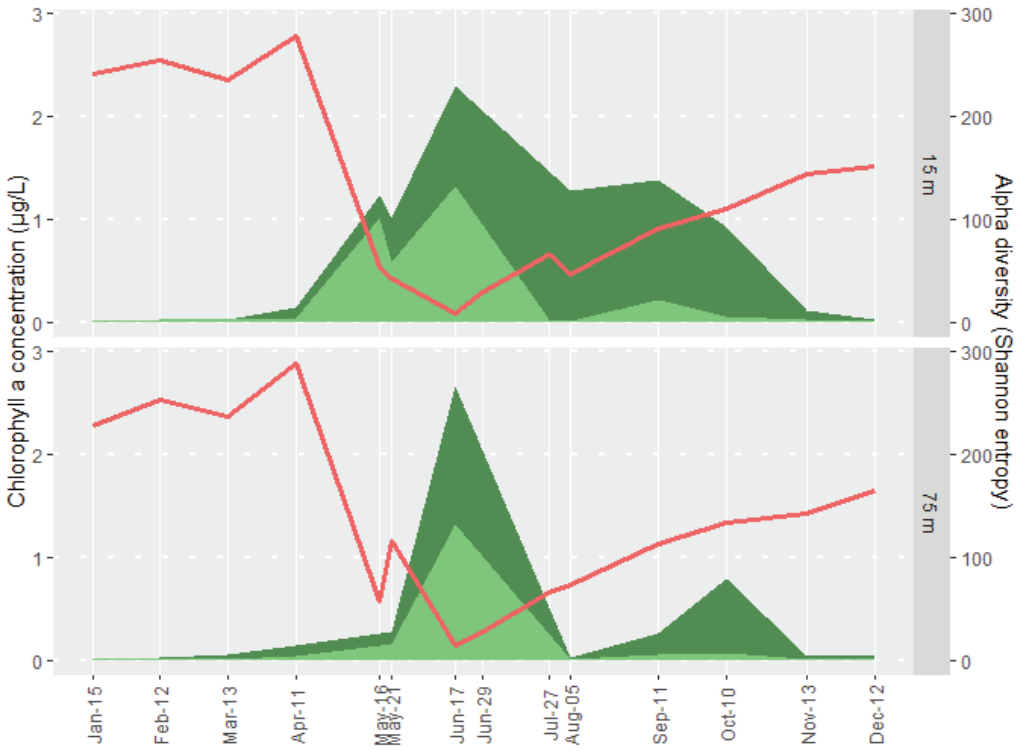


Figure 13: Alpha diversity and measured *chlorophyll a* (Chl a) concentrations throughout 2019. The red line shows the alpha diversity as measured by the Shannon entropy of each sample. The green areas represents the Chl a values measured through the seasons, with light and dark green representing the contribution from organisms greater and smaller than 10 µm, respectively. The values from both 15 m and 75 m depth are shown (top and bottom panel respectively).

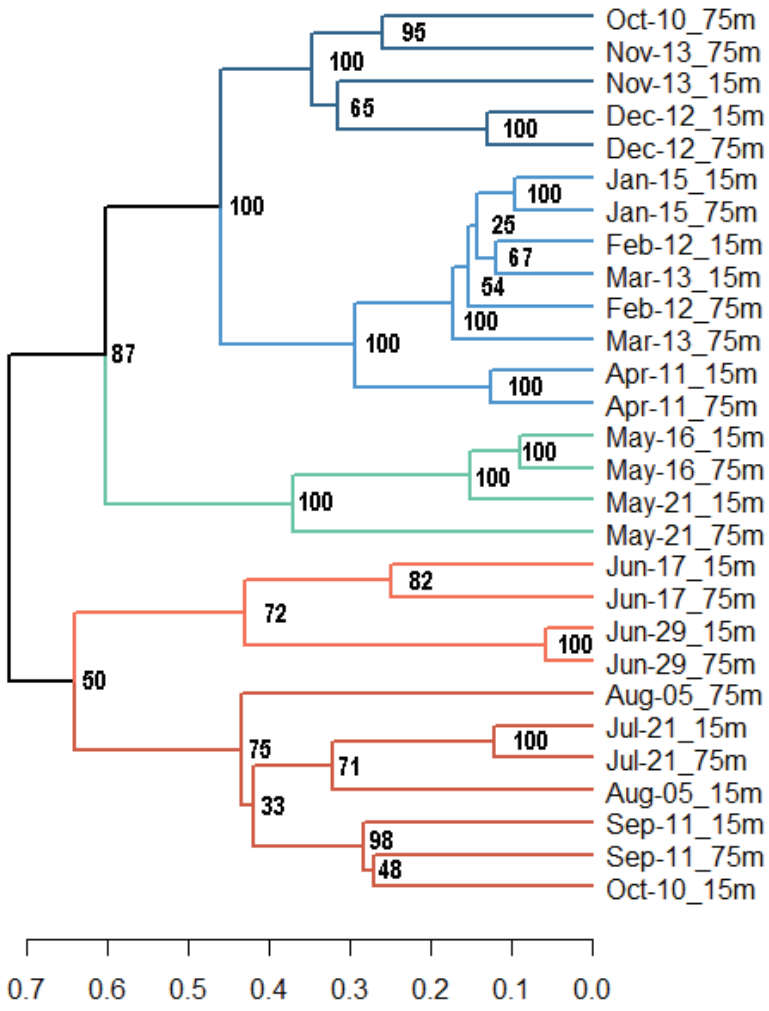


Figure 14: Dissimilarity of samples based on community 16S rDNA metabarcoding. The cluster dendrogram is based on Bray-Curtis distance between the different samples after zOTU processing. The distance is indicated by the lower axis. The values shown at the nodes are the bootstrap probability value in percent as calculated using the *pvclust* function in the **pvclust** package for R with number of bootstraps equal to 10,000.

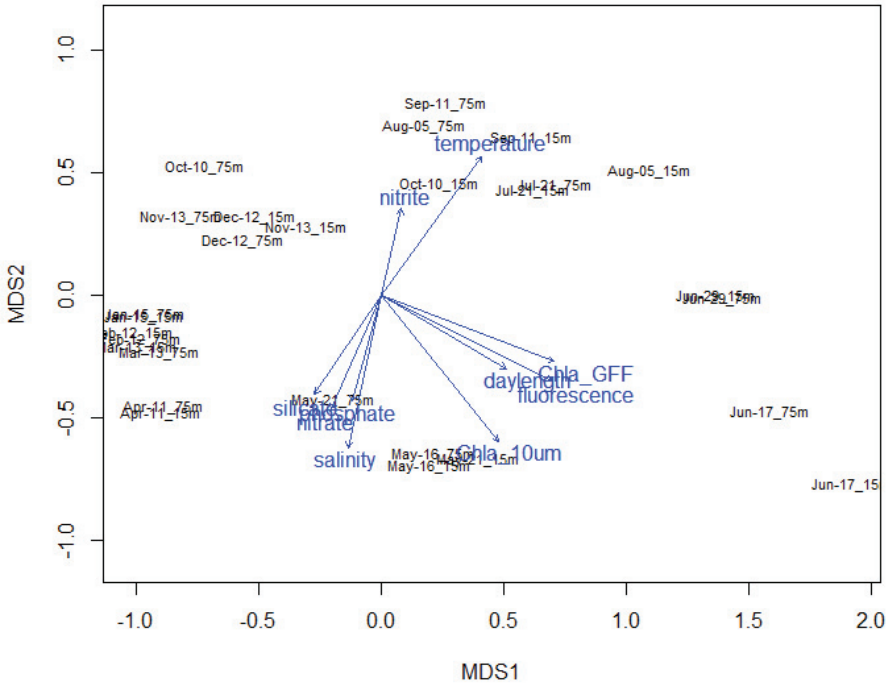


Figure 15: NMDS ordination plot of the different samples with environmental factors fitted on top. The ordination was based on the Bray-Curtis distance between the different samples after zOTU processing. The length and orientation of the arrows indicate the strength and direction of each factors role relative to the NMDS plot. Made using the *monoMDS* and *envfit* (999 permutations) functions of **vegan**.

indicating those to be the most positively correlated with that season. The day length, Chl a and fluorescence (proxy for Chl a) all pointed towards the early summer as the peak, while temperature pointed towards autumn.

3.3 Community composition and abundance through the seasons

3.3.1 Taxonomic composition and structure

The 16S rDNA results suggested that *Proteobacteria* and *Bacteroidetes* phyla greatly dominated the prokaryotic community (Figure 16a). *Bacteroidetes*

made up only 15-25% of the reads during winter, but increased during spring to peak at 75% and 80% on June 17th, 15 m and 75 m depth respectively. The proportion of *Proteobacteria* on the other hand sank during spring from 50% during winter to around 30% during summer.

The most abundant class of *Bacteroidetes* was *Flavobacteriia* making up 10% of the total reads in winter and more than 70% of the total reads on June 17th. Other recognized classes were *Sphingobacteriia*, *Bacteroidia* and *Cytophagia*. The two major classes of *Proteobacteria* were the *Alphaproteobacteria* and *Gammaproteobacteria*. *Alphaproteobacteria* was found to be most abundant during winter, while *Gammaproteobacteria* was more abundant during late spring and summer. June made an exception with the lowest recorded abundances of *Gammaproteobacteria* at only 1-6%. *Betaproteobacteria* and *Deltaproteobacteria* was also found throughout the year, but in far less numbers.

Less abundant bacterial phyla were *Actinobacteria*, *Chloroflexi*, *Dadabacteria*, *Gemmatimonadetes*, *Marinimicrobia*,*Nitrospirae*, *Planctomycetes* and *Verrucomicrobia*. They all kept a stable, but low abundance during the winter, sinking close to or below detection level during summer. *Verrucomicrobia* had an exceptional spike on August 5th at 75 m depth where it made up 15 % of the total reads.

Only 4 zOTUs were reliably assigned as *Cyanobacteria*, the three most abundant being *Synechococcus sp.* of *Family II* and genus *GpIIa* (the least abundant having no recognized affiliation). Their abundances were close to, or below detection level for all samples except on September 11th at 75 m where they made out 0.15% of the reads.

Thaumarchaeota was recognized as the most abundant archeal phylum. It had stable abundances of 7-10% during winter, but declined considerably during spring to below detection level on July 17th. In the autumn there was an exceptional high abundance of *Thaumarchaeota* at 75 m depth making up 15% of the reads on September 11th. The second most abundant archeal phylum, the *Euryarchaeota*, followed a similar pattern, but made up only 2-5% of the total reads during winter. By far the most abundant class of *Euryarchaeota* was *Thermoplasmata*. Some of the winter samples contained very small fractions of the *Nanoarchaeota* phylum.

An estimate of the absolute abundance of the different taxa was obtained by scaling the relative abundances of the different phyla with the total bacterial abundance, obtained from flow cytometry counts (Figure 16b). It should be noted that such an approach has several major weaknesses and should

only be considered a very rough estimate (see Introduction). By doing so however, the increase in *Bacteroidetes* during summer became even more evident. This also suggested *Proteobacteria* to become *more* abundant during summer, instead of *less*.

3.3.2 Abundant taxonomic units

At the zOTU level the same pattern persisted; a few units dominated during summer, especially during the peak abundance, while winter was characterised by many rare zOTUs (Figure 17a). Also here the composition at the end of the year seems to roughly reflect that of the beginning.

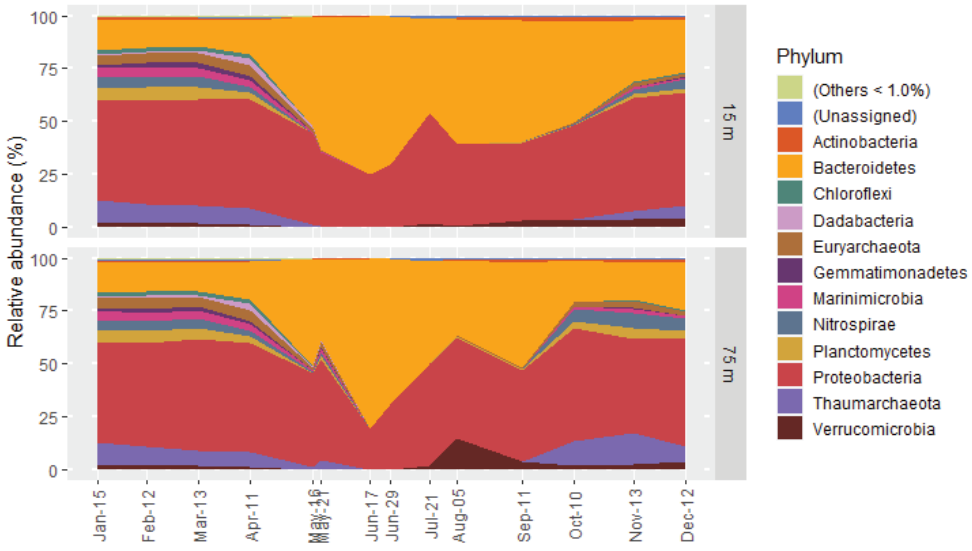
Flavobacterieae (*Bacteroidetes*) greatly dominated during the summer with seven of the most abundant zOTUs (Figure 17a). Amongst these a zOTU assigned to *Polaribacter* (named "Type 1" in this study) had the most pronounced peak, making up 44% and 35% of total read number on June 17th for 15 m and 75 m depth respectively. Fifteen other zOTUs were also recognized as *Polaribacter*, two of which were also amongst the most abundant.

Another family with a marked increase during summer was *Rhodobacteraceae* (*Proteobacteria*) with three of the most abundant zOTUs. Of these the zOTU recognized as *Sulfitobacter* genus was the most abundant, making up 14% and 20% of the reads on June 29th, at 15 m and 75 m depth respectively.

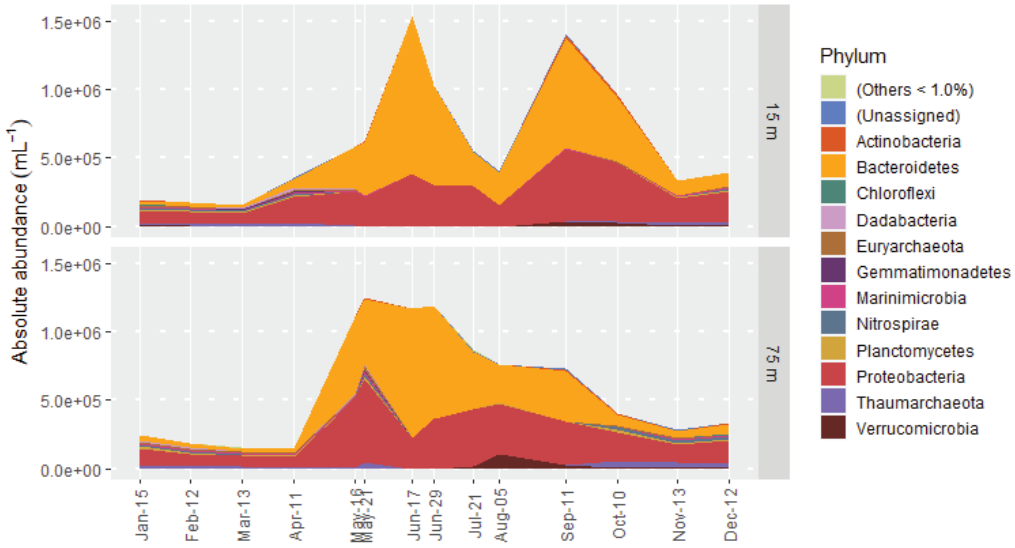
One zOTU from the *SAR11* family (*Proteobacteria*) classified as *Candidatus Pelagibacter*, kept a remarkably stable abundance at 5% of the reads throughout the seasons, with the exception of a decline in June. Its relative abundance seemed to slightly increase during the autumn.

Two zOTUs assigned as *Nitrosopumilaceae Nitrosopumilus* (*Thaumarchaeota*) and *Verrucomicrobieae Luteibacter* (*Verrucomicrobia*) respectively, both showed an exceptional peak at 75 m only. The *Nitrosopumilus* zOTU made up 9% of the reads on October 10th and 12% on November 13th, around 4 times more than on the same dates at 15 m. The *Luteibacter* zOTU had an abundance close to or below detection level except for on August 5th at 75 m, where it made up 12% of the reads.

Scaling for absolute abundance made the dominance of a few zOTUs during late spring / early summer even more evident (Figure 17b). Especially that of *Polaribacter* and *Sulfitobacter*. The community was evidently more diverse for the second peak in absolute abundance at 15 m in September, with many zOTUs of lesser abundance.

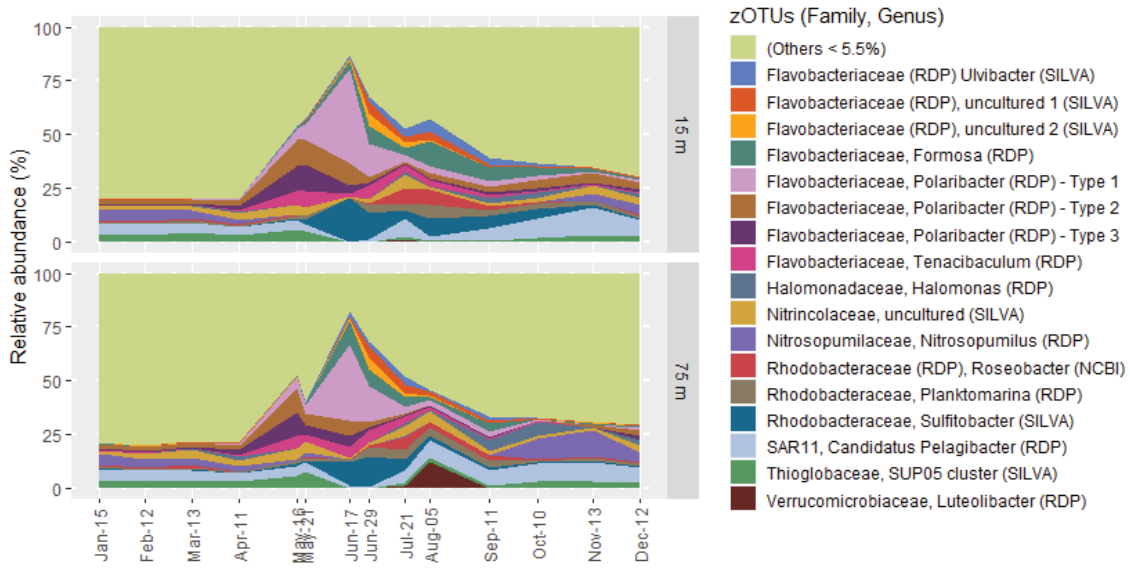


(a) Phylum relative abundance

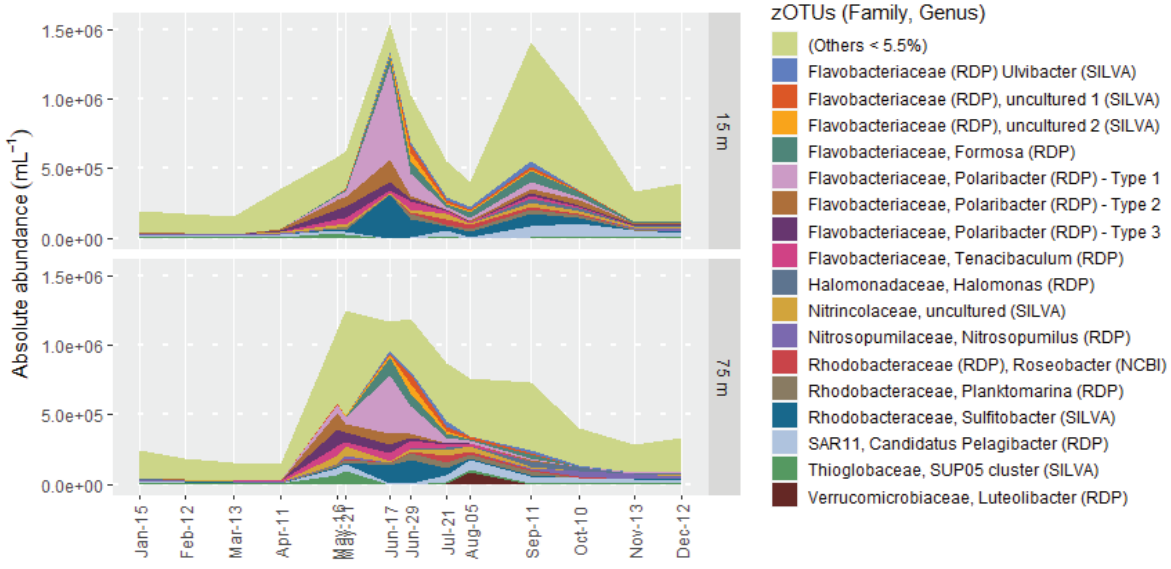


(b) Phylum absolute abundance

Figure 16: Phyla composition of samples taken throughout 2019. Relative **(a)** and absolute **(b)** abundances are shown. The taxonomic assignments were made based on the zOTU pipeline and SILVA reference database. The date of each sample is indicated along the x-axis, the areas in between are filled based on linear interpolation for illustrational purposes. For certain dates and depths the cell counts were unavailable and are then also based on linear interpolation (Figure 12). Phyla making up less than 1% of all the samples were pooled together as "Others".



(a) zOTU relative abundance



(b) zOTU absolute abundance

Figure 17: Abundances of most numerous zOTUs throughout 2019. Both the relative abundance **(a)** and absolute abundance based on scaling by flow cytometry **(b)** are shown. The dates of each sample are indicated along the x-axis, the areas in between are filled based on linear interpolation for illustrational purposes. For certain dates and depths the cell counts were unavailable and are then also based on linear interpolation (Figure 12). zOTUs with a relative abundance less than 5.5% for all samples were pooled into "Others".

3.4 Community correlation analysis

3.4.1 Correlation of the complete community

The Pearson correlation for the relative abundance between the different zOTUs across the samples can be represented in a network graph (Figure 18). Since 15 m and 75 m had very similar community composition throughout the year they were both treated as samples from the same system. The network had a power law degree distribution implying it to be scale free. There was a clear structure with some densely connected groups branched by less connected zOTUs. It is in these less dense areas of the network the most abundant zOTUs were located.

The great majority of the zOTUs clustered together in one densely connected group (group 1). These were mainly low abundance zOTUs (<1% mean relative abundance) most dominant during winter and spring (Jan-May and Sep-Dec). Three other minor, but also highly connected groups were recognized (groups 2-4). One contained low abundance zOTUs that were dominant in winter, but not in spring (group 2). It was roughly made up of half *Bacteroidetes* half *Proteobacteria*. The most densely connected group mainly consisted of *Proteobacteria* with an unique relative abundance distribution peaking in March, September and December (group 3). Group 4 mainly consisted of zOTUs originating from Chloroplasts. These showed a high relative abundance in July, but low during the rest of the year. A network constructed using absolute zOTU abundances showed a similar topology (see Appendix A).

3.4.2 Correlation of the most abundant taxonomic units

The correlation between the most abundant zOTUs were further studied, this time lowering the cut off value to 0.5. Depending on whether relative or absolute abundances were considered the resulting correlation networks had quite different topography (Figure 19a and 19b). For the relative abundance network the zOTUs that dominated during summer formed one large positively correlated group, opposed by the winter dominant. Three zOTUs (belonging to *Verrucomicrobiaceae*, *Halomonadaceae* and *Nitrincolaceae*) showed no connection to any of the other major zOTUs. Scaling for absolute abundance however, made most of the zOTUs positively correlated, with the exception of two disconnected zOTUs (belonging to *Verrucomicrobiaceae* and *Nitrosopumilaceae*).

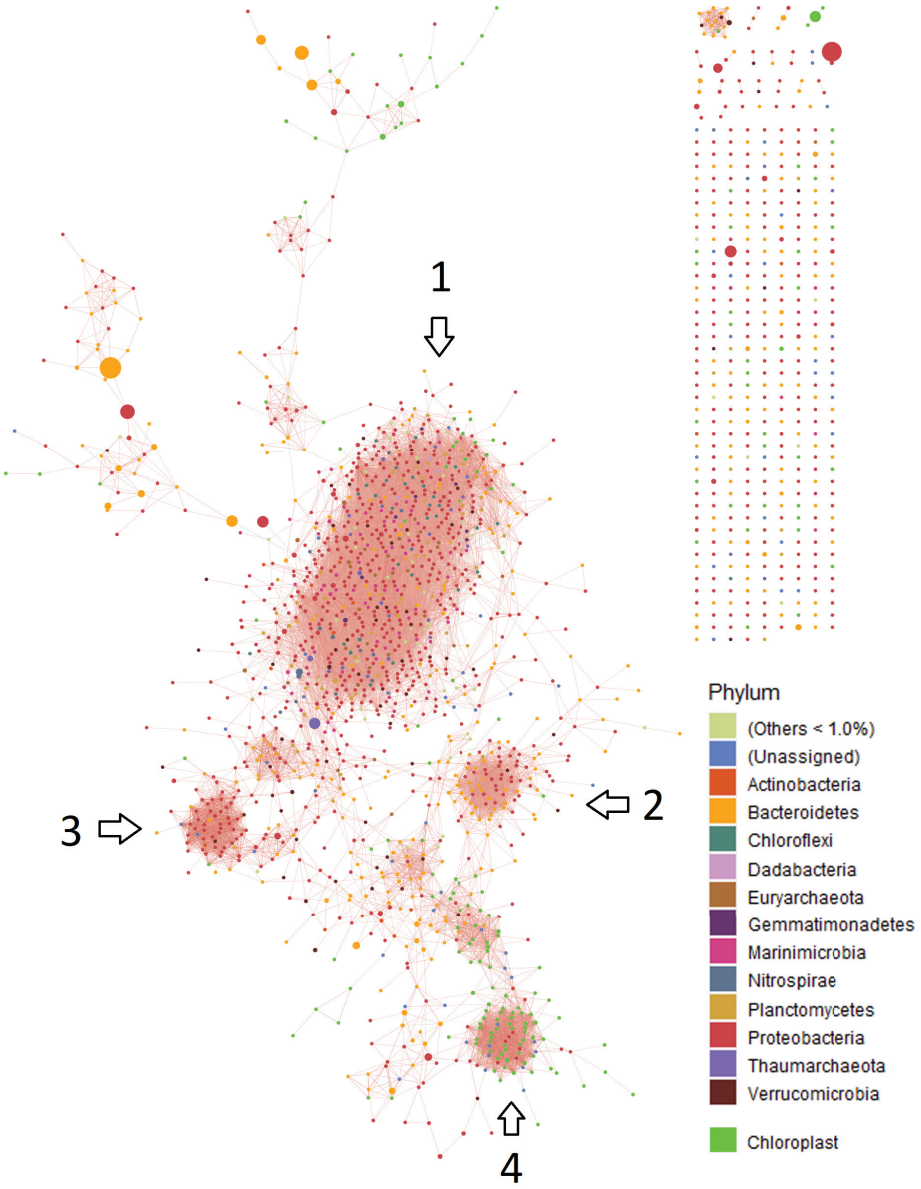
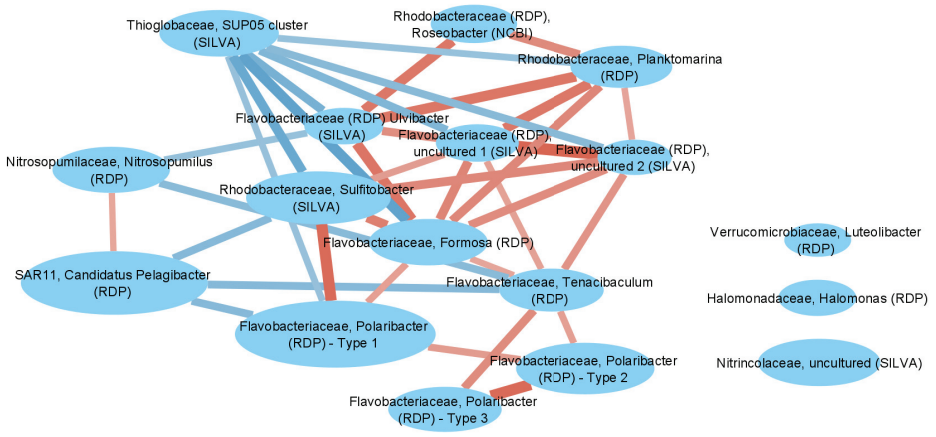
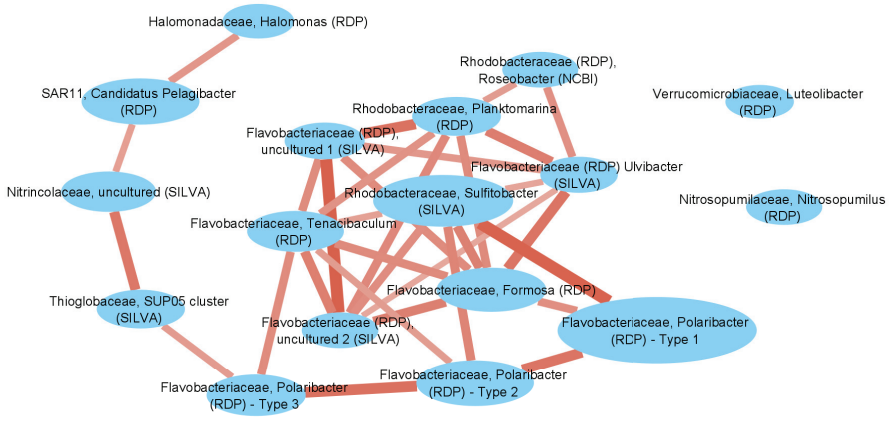


Figure 18: Correlation analysis of the IsA community. The nodes represent individual zOTUs with sizes scaled to the mean relative abundance across all samples. Nodes are coloured according to their phylum (similarly to Figure 16a and 16b). zOTUs originating from Chloroplasts were also included. The edges represents the Pearson correlation between the zOTUs across all samples using their relative abundance. A cut off value of 0.85 was used (no correlations were below -0.85 and thus no negative correlations are shown). Highly connected groups are indicated by arrows and numbers (groups 1-4). *Cytoscape v3.8.0* was used for visualization.



(a) Relative abundance



(b) Absolute abundance

Figure 19: Correlation networks for abundant zOTUs (>5.5% relative abundance for at least one sample). The nodes represent zOTUs with taxonomic assignments and the edges the Pearson correlation between them across all samples. Both network resulting from relative abundance **(a)** and absolute (scaled by cell count) **(b)** are shown. The node sizes are scaled by mean relative and absolute abundances respectively. Blue edges represents negative correlation and red edges represents positive. More intense colour and broader edges represents higher correlation. A correlations between -0.5 and 0.5 were excluded. *Cytoscape v3.8.0* was used for visualization.

4 Discussion

4.1 Processing and quality of data

4.1.1 Choice of pipeline

The zOTU pipeline was chosen in front of the traditional OTU pipeline with 97% clustering since it proved both more sensitive and simplified the data and downstream interpretation. Both pipelines performed similarly, keeping a similar number of reads. The zOTU pipeline however, detected more TUs per sample, while less in total across all samples. The resulting evenness distributions was also more inn line with that expected from a well covered community (Preston 1948; Haynes 2009), with higher mean TU abundances and very few predicted veiled TUs (in contrast to the OTU pipeline). Some of these results might seem counter-intuitive considering the lack of clustering for zOTUs, but can be explained by the UNOISE algorithm recognizing more TUs as amplification errors, removing them completely or combining them with their expected parent (Edgar 2015).

4.1.2 Sequencing depth

The sequencing depth for the combined samples seemed very good when considering the zOTU pipeline. Very few TUs were predicted to have been left out, and the accumulation curve analysis revealed that about half the samples would have been sufficient to detect al the TUs of the entire community (Figure 7b).

When considering the individual samples however, most showed an incomplete coverage with end rarefaction slopes larger than zero (Figure 5b). These were still relatively low, and thus only very rare TUs are expected to have been missed out or had their abundances distorted (Haynes 2009). Since the focus of this study where on the most abundant TUs and overall community structure, this would not had any noticeable effect on the results.

4.1.3 Taxonomic coverage and resolution

The 515FB/926R primer pair used for amplification and sequencing in this study have been reported to have an excellent coverage of both the Bacteria and Archea domain with little amplification bias in marine environments compared to previously standard primer sets (Parada, Needham, and

Fuhrman 2016).

The taxonomic resolution acquired in this study were down to genus level for most of the abundant zOTUs, using the SILVA (138 SSU, Ref NR99) reference database. This is mainly due to the limitation of the database rather than the target barcode. Resolution down to species level requires an extensive database with many represented sequences within the same genus (Edgar 2016), of which no curated version currently exists.

Of the most abundant zOTUs in this study the two assigned as *Polaribacter* "Type 2" and "Type 3" exhibited very similar distribution patterns (Figure 17a), and are therefore likely to originate from the same organisms or two closely related strands. "Type 1" on the other hand, though assigned the same phylogeny, had a very different distribution and was clearly representing another group of organisms. This suggests a potential resolution beyond "species level".

4.1.4 Frequency of samples

The changes in relative abundances seemed gradual for most of the TUs, especially during winter. This indicated monthly sampling to be sufficient in capturing the dynamic changes of the community. During late spring and summer however, the changes in abundances were quite rapid with large differences between samples. This was particularly prominent on May 16th and 21st at 75 m depth where some of the TUs showed relatively large differences in their abundance, despite these samples only differing by 5 days (Figure 17a).

Another striking example at 75 m was found on August 5th where a zOTU belonging to the *Luteibacter* genus (*Verrucomicrobia*) had a single peak in abundance making up 12% of all the reads. For all other samples this was close to 0% for this zOTU. The duration of this peak in relative abundance can have been anywhere from more than a month to a few days. A study by Alonso-Sáez, Zeder, et al. 2014 outlined such a peak for a single bacterial TU in the Arctic ocean lasting only 2 weeks. Similar peaks might very well have been missed out between two samples in this study. The temporal resolution of this study, while sufficient for a general outline of the seasonal dynamics and trends, is still too low for a complete coverage.

4.2 Drivers of the prokaryotic community

4.2.1 Spring bloom and race for energy

The winter prokaryotic community of the IsA station was found to be quite stable and diverse, with a low cell count (Figure 12 and 13). The return of the Sun set in motion a series of large changes, most notably a strong increase in phototrophic biomass, consistent with previous descriptions of the annual spring bloom (Marquardt et al. 2016). This would have greatly increased the amount of available organic nutrients and is likely the direct cause of the tandem increase in the number of heterotrophic prokaryotes, as suggested by previous studies (Teeling et al. 2012; Wilson et al. 2017). The subsequent decrease in alpha diversity, and dominance of a few zOTUs, indicate that some TUs were able to exploit this rapid boom much better than others. This could have been due to direct competition or other independent factors such as the hypothesized light inhibition of the *Thaumarchaeota* (Guerrero and R. Jones 1996; Murray et al. 1998; Mincer et al. 2007; Merbt et al. 2012; Pedneault et al. 2014).

A study by Teeling et al. 2012 from more temperate regions claims the successional dominance of a few TUs to be due to niche adaptation to substrate availability, which change as the bloom change in composition. Indeed, the peaks of several abundant zOTUs were observed at different dates from the initiation of the bloom and throughout summer (Figure 17a). Changes were also observed for the composition of the phototrophic biomass with smaller phytoplancton taking over the dominance of larger as the bloom developed (Figure 13).

4.2.2 Nutrient limitations

The spring bloom seems to have rapidly exhausted the amount of available nutrients, especially near the surface. This is likely what led to the large drop in heterotrophic cell count at 15 m (Figure 12). Especially the nitrate concentrations were directly negatively correlated to the cell count, with opposing minimas and maximas. This could be due to a direct effect where the zOTUs abundant in the initial boom do not get enough inorganic nutrients to sustain their own growth, which in turn lead to a complete shift in community structure, promoting a more diverse community with different zOTUs being abundant (Figure 17b). This new community would presumably cope better with the nutrient limitations, with the result of a second

peak in heterotrophic cell count at 15 m depth. For instance, the dominance of different TUs at different stages of the bloom have been coupled to different acquisition strategies for phosphate, which will be in direct competition with the algae (Teeling et al. 2012).

While this does seem as a plausible explanation, heterotrophic organisms do acquire much of their nutrients through the digestion of organic compounds, and prefer ammonia in front of nitrate as a source of inorganic nitrogen (D. L. Kirchman 1994). Other studies in the Arctic have suggested inorganic nutrients to play a more indirect role on the prokaryotic community by affecting the development of the phytoplancton bloom (Connelly et al. 2014; Wilson et al. 2017).

The gradual build up of nutrient concentrations during autumn was likely due to a combination of declining phototrophic biomass and heterotrophic cell count releasing organic nutrients that got broken down to inorganic, and an increase of chemolithoautotrophic processes, as suggested by previous studies (Connelly et al. 2014; Grzymski et al. 2012; Alonso-Sáez et al. 2008). This is strongly supported by the tandem increase in *Thaumarchaeota* and nitrate concentrations at 75 m depth in October (Figure 12 and 16a). The zOTU making up most of the *Thaumarchaeota* at this time and depth belonged to the *Nitrosopumilus* genus, which is known for its strict autotrophic ammonia and urea oxidating members (Bayer et al. 2016; Walker et al. 2010; Könneke et al. 2005). The increased amounts of viruses late in the year can have facilitated this process by releasing more organic nutrients into the water by lysis. Other subtle effects of virus counts on prokaryotic communities have previously been implied (Jardillier et al. 2005; Winter et al. 2007), but were not evident in this study.

4.2.3 The effect of water masses

The role of water masses on the prokaryote community remained somewhat unclear due to the covariance with other environmental factors. The CTD data of this study suggests heating of the surface water directly powered by the sun spreading to lower depths (Figure 10a). The following decrease in salinity was likely due to the inflow of melt-water from surrounding glaciers and rivers, leading to a stratification of the water column (Figure 10b and 2). This drastic change in conditions, somewhat delayed to the initial spring bloom, could also account for the changes in community structure previously noted, either directly or indirectly.

Some studies have pointed out salinity in particular as playing an important and direct role on prokaryotic community composition, by affecting the growth of the TUs differently (Dupont et al. 2014; Ortega-Retuerta et al. 2013; Barghini et al. 2018). The range in salinity of these studies however, was far greater than for this. The high similarity in community composition between 15 m and 75 m depth, also during the stratification in summer (Figure 14), suggests neither salinity nor temperature played a significant role on this range. The stratification however, was likely sufficient to limit vertical mixing of the water column. This has been suggested to play an important role by limiting the nutrient cycling between depths (Smith and Sakshaug 1990), and could have enhanced the differences in nutrients and phototrophic biomass between the different depths as discussed above.

Satellite data from the HYCOM consortium's oceanic model also sowed a change in the ocean currents surrounding Spitsbergen (Figure 11a to 11o). Considering the unique topography of Isfjorden it is therefore possible that the observed changes in the water profile at the IsA station were also caused by an exchange of water masses with Atlantic water from the WSC getting access to the fjord (F. Nilsen et al. 2008; Skogseth et al. 2020). In this case the import and export of prokaryotes carried by the currents could be a possible explanation for the strong change in community structure. To which degree this played a role compared to *in situ* growth remains unclear.

4.3 Community composition and structure

4.3.1 Taxonomic composition

The taxonomic composition of the prokaryotic community was largely in accordance with previous studies of Arctic waters (see Introduction). Even at zOTU level the high amounts of *Polaribacter* and *Sulfitobacter* during summer and prevailance of *Pelagibacter* (*SAR11*) throughout the year was consistent with previous findings in Arctic coastal waters (Ghiglionea et al. 2012). The high abundance of *Polaribacter* and *Sulfitobacter* during summer however, stood in contrast to findings for the Arctic open ocean (Ghiglionea et al. 2012; Wilson et al. 2017).

Surprisingly *Alphaproteobacteria* was found to be more abundant than *Gammaproteobacteria*, which is more reminiscent of temperate marine communities (Malmstrom et al. 2007; Ghiglionea et al. 2012). This could imply a higher influence of Atlantic water from the WSC, yet the abundance of

Alphaproteobacteria was found to be largest during winter when the Arctic water of the ESC dominated the mouth of Isfjorden (Figure 11a to 11o). Some standard primers previously used for metabarcoding, such as by Wilson et al. 2017, have been known to bias for *Gammaproteobacteria* (Parada, Needham, and Fuhrman 2016), which could have exaggerated these differences. Interestingly the same study had a singular peak in the relative abundance of *Verrucomicrobia* in August for deeper mesophelagic waters similar to this study. The zOTU behind this peak belonged to *Luteibacter*, a genus associated with the decomposing of dead wood (Tláskal et al. 2017).

4.3.2 Community structure

The majority of the TUs were most dominant during winter with highly correlating abundance distributions (Figure 18). This could suggest a winter community of highly specialized TUs dependant on each others metabolic processes for coping with the extreme energy depleted conditions. The most abundant TUs however, dominated during summer with little correlation to other TUs, including others that were abundant. This suggests a very different survival strategy consistent with the earlier notion that the overall most abundant TUs are those which are able to exploit short windows of the phytoplankton bloom. These are TUs which proliferation in the Arctic might rely upon a single growth burst when the composition of the phytoplankton bloom is favourable (Teeling et al. 2012), in line with the Seed Bank theory (Lennon and S. E. Jones 2011).

To further the understanding of community structure and interactions among the TUs more complex models should be made, leading to a more mechanistic description of the system. For this purpose estimates of absolute TU abundances are essential for the outcome of the analysis (Cao et al. 2016; Xiao et al. 2017), as illustrated by simple correlation networks of the most abundant TUs (Figure 19a and 19b). The data provided by this study should therefore be sufficient for simplified models such as suggested by Ovaskainen et al. 2017.

Conclusion

This thesis represents a pilot study of the prokaryotic community in Isfjorden throughout a full year, providing higher temporal and taxonomic resolution than previous studies, with a good community coverage. Zero radius OTUs (zOTUs) were chosen in front of traditional OTU clustering as it produced better data quality with easier biological interpretation and comparability across studies.

Proteobacteria and *Bacteroidetes* were found to be the most abundant phyla, with the latter being the particularly abundant during summer. Surprisingly, *Alphaproteobacteria* was generally more abundant than *Gammaproteobacteria*, with the highest relative abundance being in winter. *Thaumarcheota* was recognized as the most abundant archeal phylum. The most abundant zOTU belonged to *Polaribacter* making up to 44 % of the reads in June. Other notable zOTUs belonged to *sulfitobacter*, *Pelagibacter* (*SAR11*) and *Nitrosopumilus*.

The prokaryotic community exhibited a cyclic change, with similar community composition at the beginning and end of the year. The largest differences in community structure was found between winter and summer. Winter was characterized by a diverse and stable community with high evenness and many low abundance taxonomic units (TUs), while summer was dominated by a few highly abundant TUs and rapid changes in community structure.

The main driver for this difference was recognized as the annual phytoplancton bloom heralding the return of the Sun. Inorganic nutrient availability was also coupled to changes in cell counts and could possibly explain the transition to a more diverse and even community in late summer and autumn. The roles of temperature and salinity were more elusive, but stratification of the water column during summer could explain some of the differences in nutrient composition and cell count between upper and lower depths. The community composition however, remained highly similar between the two depths throughout the year, with some few exceptions.

Satellite data revealed a change in ocean currents which could have affected community structure by replacement of water masses and suspended prokaryotes. To elucidate the importance of this compared to *in situ* growth a comparative study could be made with parallel samples taken from the WSC and ESC off the coast of Svalbard. Modelling of the system can provide a more mechanistic understanding of the community dynamics and interactions amongst the TUs.

Acknowledgements

A big thanks to ass. prof. Anna Vader[1] and prof. Ingrid Bakke[2] for supervision and support for my work. Without you there would have been no thesis. Thanks also to PhD cand. Cheshtaa Chitkara[1]. Working with you on the IsA station made life so much more enjoyable. Thanks to lab technician Stuart Thomson[1] for making sure we did not have to worry about lab supplies and safety, and for taking care of the analysis of the nutrient samples. Lastly, a big thanks to all the people working on the IsA time series station and which took the samples of the first half year of this study.

[1]The University Centre in Svalbard (UNIS)
[2]The Norwegian University of Science and Technology (NTNU)

References

Alonso-Sáez, Laura, Alison S. Wallerc, et al. (2012). "Role for urea in nitrification by polar marine Archaea". In: *PNAS.* 109 (44), pp. 17989–17994. DOI: 10.1073/pnas.1201914109.

Alonso-Sáez, Laura, Michael Zeder, et al. (2014). "Winter bloom of rare betaproteobacterium in the Arctic Ocean". In: *Frontiers in Microbiology* 5. DOI: 10.3389/fmicb.2014.00425.

Alonso-Sáez, Laura et al. (2008). "Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes". In: *Environmental Microbiology* 10 (9). DOI: 10.1111/j.1462-2920.2008.01674.x.

Barghini, Paolo et al. (2018). "Bacteria from the "Saline di Tarquinia" marine salterns reveal very atypical growth profiles with regards to salinity and temperature". In: *Mediterranean Marine Science* 19 (3). DOI: 10.12681/mms.15514.

Bayer, Barbara et al. (2016). "Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation". In: *The ISME Journal* 10, pp. 1051–1063. DOI: 10.1038/ismej.2015.200.

Bowman, Jeff S et al. (2012). "Microbial community structure of Arctic multiyearsea ice and surface seawater by 454 sequencingof the 16S RNA gene". In: *The ISME Journal* 6, pp. 11–20. DOI: 10.1038/ismej.2011.76.

Brochier-Armanet, Céline et al. (2008). "Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota". In: *Nature Reviews Microbiology* 6, pp. 245–252. DOI: 10.1038/nrmicro1852.

Brussaard, Corina P. D. (2004). "Optimization of Procedures for Counting Viruses by Flow Cytometry". In: *Applied and Environmental Microbiology* 70 (3). DOI: 10.1128/AEM.70.3.15061513.2004.

Callahan, Benjamin J, Paul J McMurdie, and Susan P Holmes (2017). "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis". In: *The ISME Journal* 11, pp. 2639–2643. DOI: 10.1038/ismej.2017.119.

Cao, Hong-Tai et al. (2016). "inferring human microbial dynamics from temporal metagenomics data: Pitfalls and lessons". In: *Bioessays* 39.2, p. 1600188. DOI: 10.1002/bies.201600188.

Christman, Glenn D. et al. (2011). "Abundance, Diversity, and Activity of Ammonia-Oxidizing Prokaryotes in the Coastal Arctic Ocean in Sum-

mer and Winter". In: *Applied and Environmental Microbiology* 77 (6), pp. 2026–2034. DOI: 10.1128/AEM.01907-10.

Cole, J. R. et al. (2014). "Ribosomal Database Project: data and tools for high throughput rRNA analysis". In: *Nucleic Acids Research* 42.D633-D642. DOI: 10.1093/nar/gkt1244.

Comeau, André M. et al. (2011). "Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum". In: *PLoS One* 6 (11). DOI: 10.1371/journal.pone.0027492.

Connelly, Tara L. et al. (2014). "Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea during the Arctic Summer and Winter Seasons". In: *Applied and Environmental Microbiology* 80 (19), pp. 6013–6022. DOI: 10.1128/AEM.01431-14.

Doolittle, DF, WKW Li, and AM Wood (2008). "Wintertime abundance of picoplankton in the Atlantic sector of the Southern Ocean". In: *Nova Hedwig* 133, pp. 147–160.

Dupont, Chris L. et al. (2014). "Functional Tradeoffs Underpin Salinity-Driven Divergence in Microbial Community Composition". In: *PLoS ONE* 9.e89549 (2). DOI: 10.1371/journal.pone.0089549.

E, Teira et al. (2006). "Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic". In: *Limnology and Oceanography* 51, pp. 2131–2144.

Edgar, Robert C. (2010). "Search and clustering orders of magnitude faster than BLAST". In: *Bioinformatics* 26 (19), pp. 2460–2461. DOI: 10.1093/bioinformatics/btq461.

— (2013). "UPARSE: highly accurate OTU sequences from microbial amplicon reads". In: *Nature Methods* 10, pp. 996–998. DOI: 10.1038/nmeth.2604.

— (2015). "Error filtering, pair assembly and error correction for next-generation sequencing reads". In: *Bioinformatics* 31 (21), pp. 3476–3482. DOI: 10.1093/bioinformatics/btv401.

— (2016). "SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences". In: *bioRxiv*. DOI: 10.1101/074161.

— (2018). "Updating the 97% identity threshold for 16S ribosomal RNA OTUs". In: *Bioinformatics* 34 (14), pp. 2371–2375. DOI: 10.1093/bioinformatics/bty113.

Galand, Pierre E, Emilio O Casamayor, et al. (2012). "Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequenc-

ing". In: *The ISME Journal* 3, pp. 860–869. DOI: 10.1038/ismej.2009.23.

Galand, Pierre E, Marianne Potvin, et al. (2010). "Hydrography shapes bacterial biogeography of the deep Arctic Ocean". In: *The ISME Journal* 4, pp. 564–576. DOI: 10.1038/ismej.2009.134.

Galand, Pierre E. et al. (2009). "Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation". In: *Environmental Microbiology* 11 (4), pp. 971–980.

Garneau, Marie-Ève et al. (2008). "Seasonal dynamics of bacterial biomass and production in a coastal arctic ecosystem: Franklin Bay, western Canadian Arctic". In: *Journal of Geophysical Research* 113. DOI: 10.1029/2007JC004281.

Ghiglionea, Jean-François et al. (2012). "Pole-to-pole biogeography of surface and deep marine bacterial communities". In: *PNAS*. DOI: 10.1073/pnas.1208160109.

Glassman, Sydney I. and Jennifer B. H. Martiny (2018). "Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units". In: *mSphere* 3.e00148-18. DOI: 10.1128/mSphere.00148-18.

Gorrasi, Susanna et al. (2019). "Structure and diversity of the bacterial community of an Arctic estuarine system (Kandalaksha Bay) subject to intense tidal currents". In: *Journal of Marine Systems* 95 (4), pp. 77–85. DOI: 10.1016/j.jmarsys.2019.04.004.

Grzymski, Joseph J et al. (2012). "A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters". In: *The ISME Journal* 6, pp. 1901–1915. DOI: 10.1038/ismej.2012.31.

Guerrero, MA and RD Jones (1996). "Photoinhibition of marine nitrifying bacteria. I. Wavelength-dependent response". In: *MEPS* 141, pp. 183–192. DOI: 10.3354/meps141183.

Haynes, Matthew (2009). *Encyclopedia of Ecology (Second Edition)*. Elsevier.

Herfort, Lydie et al. (2007). "Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables". In: *FEMS Microbiology Ecology* 62 (3), pp. 242–257. DOI: 10.1111/j.1574-6941.2007.00397.x.

Jardillier, Ludwig et al. (2005). "Effects of Viruses and Predators on Prokaryotic Community Composition". In: *Microbial Ecology* 50 (4), pp. 557–569. DOI: 10.2307/25153280.

Kalanetra, Karen M., Nasreen Bano, and James T. Hollibaugh (2009). "Ammonia-oxidizing Archaea in the Arctic Ocean and Antarctic coastal waters". In: *Environmental Microbiology* 11 (9). DOI: 10.1111/j.1462-2920.2009.01974.x.

Kennedy, J et al. (2010). "Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism". In: *Marine Drugs* 8 (3), pp. 608–628. DOI: 10.3390/md8030608.

Kindt, Roeland and Richard Coe (2005). "CHAPTER 10: Analysis of ecological distance by ordination". In: *Tree Diversity Analysis: A Manual and Software for Common Statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), pp. 153–196. URL: http://www.worldagroforestry.org/output/tree-diversity-analysis.

Kirchman, D. L. (1994). "The uptake of inorganic nutrients by heterotrophic bacteria". In: *Microbial Ecology* 28, pp. 255–271. DOI: 10.1007/BF00166816.

Kirchman, David L., Matthew T. Cottrell, and Connie Lovejoy (2010). "The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes". In: *Environmental Microbiology* 12 (5), pp. 1132–1143. DOI: 10.1111/j.1462-2920.2010.02154.x.

Kirchman, David L., Hila Elifantz, et al. (2007). "Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean". In: *Limnology and Oceanography* 52 (2). DOI: 10.4319/lo.2007.52.2.0495.

Klindworth, Anna et al. (2013). "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies". In: *Nucleic Acids Research* 41 (1). DOI: 10.1093/nar/gks808.

Könneke, Martin et al. (2005). "Isolation of an autotrophic ammonia-oxidizing marine archaeon". In: *Nature* 437, pp. 543–546. DOI: 10.1038/nature03911.

Kraemer, Susanne et al. (2020). "Diversity and biogeography of SAR11 bacteria from the Arctic Ocean". In: *The ISME Journal* 14, pp. 79–90. DOI: 10.1038/s41396-019-0499-4.

Kress, W. John et al. (2015). "DNA barcodes for ecology, evolution, and conservation". In: *Cell press* 30 (1), pp. 25–35. DOI: 10.1016/j.tree.2014.10.008.

Lee, Sang H. and Terry E. Whitledge (2005). "Primary and new production in the deep Canada Basin during summer 2002". In: *Polar Biology* 28, pp. 190–197. DOI: 10.1007/s00300-004-0676-3.

Lennon, Jay T and Suart E Jones (2011). "Microbial seed banks: The ecological and evolutionary implications of dormancy". In: *Nature Reviews Microbiology* 9 (2), pp. 119–30. DOI: 10.1038/nrmicro2504.

Lovejoy, Connie, Pierre E. Galand, and David L. Kirchman (2011). "Picoplankton diversity in the Arctic Ocean and surrounding seas". In: *Marine Biodiversity* 41, pp. 5–12. DOI: 10.1007/s12526-010-0062-z.

Lucas, Rico et al. (2016). "A critical evaluation of ecological indices for the comparative analysis of microbial communities based on molecular datasets". In: *FEMS Microbiology Ecology* 93. DOI: 10.1093/femsec/fiw209.

Madigan, Michael T. et al. (2014). *Brock Biology of Microorganisms (14th Edition)*. Pearson. ISBN: 9780321897398.

Malmstrom, Rex R. et al. (2007). "Diversity, abundance, and biomass production ofbacterial groups in the western Arctic Ocean". In: *Aquatic Microbial Ecology* 47 (1), pp. 45–55. DOI: 10.3354/ame047045.

Marquardt, Miriam et al. (2016). "Strong Seasonality of Marine Microbial Eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway)". In: *Applied Environmental Microbiology* 82 (6). DOI: 10.1128/AEM.03208-15.

McMurdie, Paul J. and Susan Holmes (2014). "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible". In: *PLOS Computational Biology* 10.e1003531 (4). DOI: 10.1371/journal.pcbi.1003531.

Merbt, Stephanie N. et al. (2012). "Differential photoinhibition of bacterial and archaeal ammonia oxidation". In: *FEMS Microbiology Letters* 327 (1), pp. 41–46. DOI: 10.1111/j.1574-6968.2011.02457.x.

Mincer, Tracy J. et al. (2007). "Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre". In: *Environmental Microbiology* 9 (5). DOI: 10.1111/j.1462-2920.2007.01239.x.

Müller, Oliver et al. (2018). "Spatiotemporal Dynamics of Ammonia-Oxidizing Thaumarchaeota in Distinct Arctic Water Masses". In: *Frontiers in Microbiology* 9. DOI: 10.3389/fmicb.2018.00024.

Murray, A. E. et al. (1998). "Seasonal and Spatial Variability of Bacterial and Archaeal Assemblages in the Coastal Waters near Anvers Island, Antarctica". In: *MEPS* 64 (7), pp. 2585–2595. DOI: 10.1128/AEM.64.7.2585-2595.1998.

Nilsen, F. et al. (2008). "Fjord-shelf exchanges conrolled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Sval-

bard". In: *Continental Shelf Reseach* 28, pp. 1838–1853. DOI: 10.1016/j.csr.2008.04.015.

Nilsen, F. R. Cottierm F. et al. (2010). "Arctic fjords: a review of the oceanographic environment and dominant physical processes". In: *Geological Society* 344, pp. 35–50. DOI: 10.1144/SP344.4.

Ortega-Retuerta, E. et al. (2013). "Spatial variability of particle-attached and free-living bacterial diversity in surface waters from the Mackenzie River to the Beaufort Sea (Canadian Arctic)". In: *Biogeosciences* 10 (4), pp. 2747–2759. DOI: 10.5194/bg-10-2747-2013.

Ovaskainen, Otso et al. (2017). "How are species interactions structures in species-rich communities? A new methood for analysing time-series data". In: *The Royal Society* 284. DOI: 10.1098/rspb.2017.0768.

Parada, Alma E., David M. Needham, and Jed A. Fuhrman (2016). "Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples". In: *Environbmental Microbiology* 18 (5), pp. 1403–1414. DOI: 10.1111/1462-2920.13023.

Paulsen, Maria L. et al. (2016). "Synechococcus in the Atlantic Gateway to the Arctic Ocean". In: *Frontiers in Marine Science* 3. DOI: 10.3389/fmars.2016.00191.

Pedneault, Estelle et al. (2014). "Archaeal amoA and ureC genes and their transcriptional activity in the Arctic Ocean". In: *Scientiffic Reports* 4. DOI: 10.1038/srep04661.

Pedrós-Alió, Carlos et al. (2018). "Marine Microbial Diversity as seen by High-Throughput Sequencing". In: *Microbial Ecology of the Oceans (Third Edition)*. .2. John Wiley & Sons, Inc.

Porter, Teresita M. and Mehrdad Hajibabaei (2018). "Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis". In: *Molecular Ecology* 27 (2). DOI: 10.1111/mec.14478.

Preston, F. W. (1948). "The Commonness, And Rarity, of Species". In: *Ecology* 29 (3). DOI: 10.2307/1930989.

Quast, C et al. (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools". In: *Nucleic Acids Research* 41.D590-D596. DOI: 10.1093/nar/gks1219.

Robinson, J. Paul et al. (1999). "Current protocols in cytometry". In: Supplement 10.

Sakshaug, E. (2004). "Primary and Secondary Production in the Arctic Seas". In: *The Organic Carbon Cycle in the Arctic Ocean*. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-18912-8_3.

Shahi, Shailesh K., Samantha N. Freedman, and Ashutosh K. Mangalam (2017). "Gut microbiome in multiple sclerosis: The players involved and the roles they play". In: *Gut Microbes*. DOI: 10.1080/19490976.2017.1349041.

Skogseth, R. et al. (2020). "Variability and decadal trends in the Isfjorden (Svalbard) ocean climate and circulation – An indicator for climate change in the European Arctic". In: *Progress in Oceanography* 187.102394. DOI: 10.1016/j.pocean.2020.102394.

Smith, WO and E Sakshaug (1990). "Polar phytoplankton". In: *Polar oceanography, part B. Academic, San Diego*, pp. 475–525.

Stockner, John G. (1988). "Phototrophic picoplankton: An overview from marine and freshwater ecosystems". In: *Limnology and Oceanography* 33 (4). DOI: 10.4319/lo.1988.33.4part2.0765.

Svendsen, Harald et al. (2002). "The physical environment of Kongsfjorden-Krossfjorden, an Arctic fjord system in Svalbard". In: *Polar Research* 21.1, pp. 133–166.

Teeling, Hanno et al. (2012). "Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom". In: *Science* 336 (6081), pp. 608–611. DOI: 10.1126/science.1218344.

Tláskal, Vojtěch et al. (2017). "Bacteria associated with decomposing dead wood in a natural temperate forest". In: *FEMS Microbiology Ecology* 93 (12). DOI: 10.1093/femsec/fix157.

Tremblay, Geneviève et al. (2009). "Late summer phytoplankton distribution along a 3500 km transect in Canadian Arctic waters: Strong numerical dominance by picoeukaryotes". In: *Aquatic Microbial Ecology* 54 (1), pp. 55–70. DOI: 10.3354/ame01257.

Valentine, David L. (2007). "Adaptations to energy stress dictate the ecology and evolution of the Archaea". In: *Nature Reviews Microbiology* 5, pp. 316–323. DOI: 10.1038/nrmicro1619.

Waleron, Małgorzata et al. (2007). "Allochthonous inputs of riverine picocyanobacteria to coastal waters in the Arctic Ocean". In: *FEMS Microbiology Ecology* 59 (2), pp. 356–365. DOI: 10.1111/j.1574-6941.2006.00236.x.

Walker, C. B. et al. (2010). "Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed ma-

rine crenarchaea". In: *PNAS* 107 (19), pp. 8818–8823. DOI: 10.1073/pnas.0913533107.

Wiedmann, Ingrid et al. (2016). "Seasonality of vertical flux and sinking particle characteristics in an ice-free high arctic fjord - Different from subarctic fjords?" In: *Marine Systems* 154, pp. 192–205. DOI: 10.1016/j.jmarsys.2015.10.003.

Wilson, Bryan et al. (2017). "Changes in Marine Prokaryote Composition with Season and Depth Over an Arctic Polar Year". In: *Frontiers in Marine Science* 4. DOI: 10.3389/fmars.2017.00095.

Winter, Christian et al. (2007). "Relationship of geographic distance, depth, temperature, and viruses with prokaryotic communities in the eastern tropical Atlantic Ocean". In: *Microbial Ecology* 56 (2), pp. 383–389. DOI: 10.1007/s00248-007-9343-x.

Wuchter, Cornelia et al. (2006). "Archaeal nitrification in the ocean". In: *PNAS* 103 (33), pp. 12317–12322. DOI: 10.1073/pnas.0600756103.

Xiao, Yandong et al. (2017). "Mapping the ecological networks of microbial communities". In: *Nature Communications* 8.2042. DOI: 10.1038/s41467-017-02090-2.

Yergeau, Etienne et al. (2017). "Metagenomic survey of the taxonomic and functional microbial communities of seawater and sea ice from the Canadian Arctic". In: *Scientific Reports* 7 (42242). DOI: 10.1038/srep42242.

A Supplementary figures

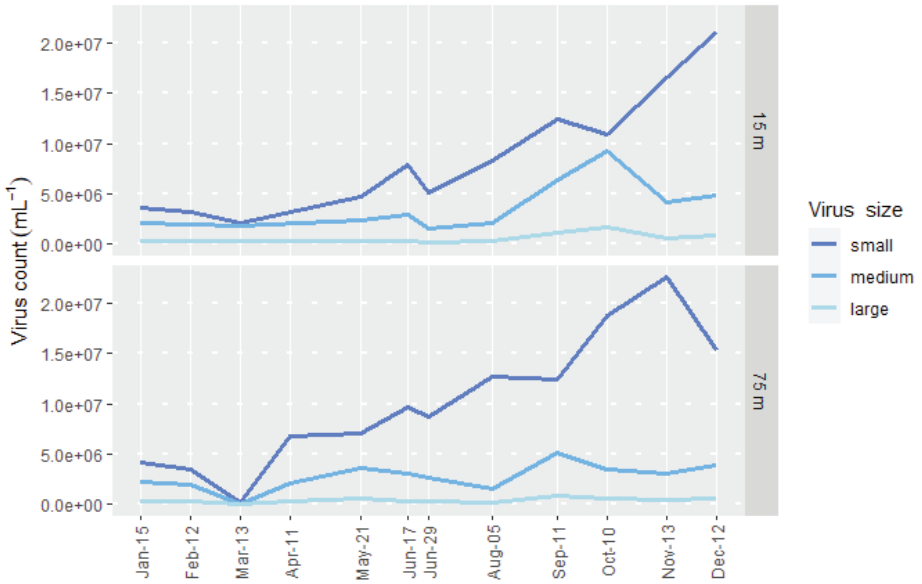


Figure A.1: Virus counts as measured by flow cytometry throughout the year of 2019. Three different categories of virus are shown based on genome size; small, medium and large. Sampling dates are indicated on the x-axis.

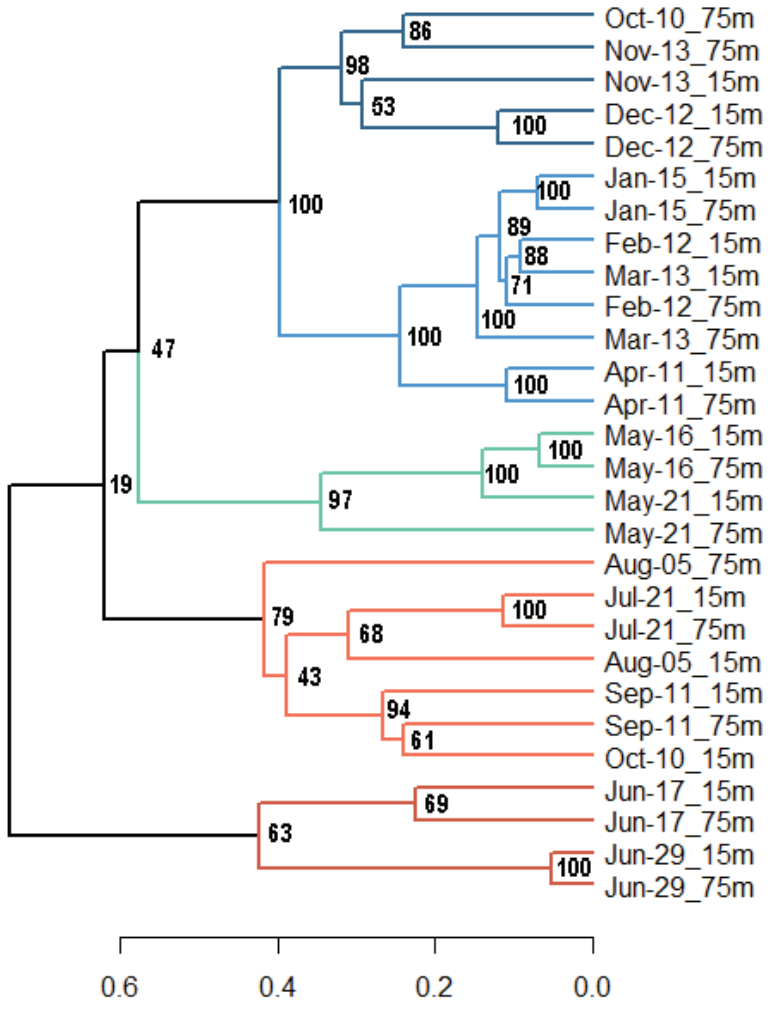


Figure A.2: Dissimilarity of samples based on community 16S rDNA metabarcoding. The cluster dendrogram is based on Bray-Curtis distance between the different samples **after OTU processing**. The distance is indicated by the lower axis. The values shown at the nodes are the bootstrap probability value in percent as calculated using the *pvclust* function in the **pvclust** package for R with number of bootstraps equal to 10,000.

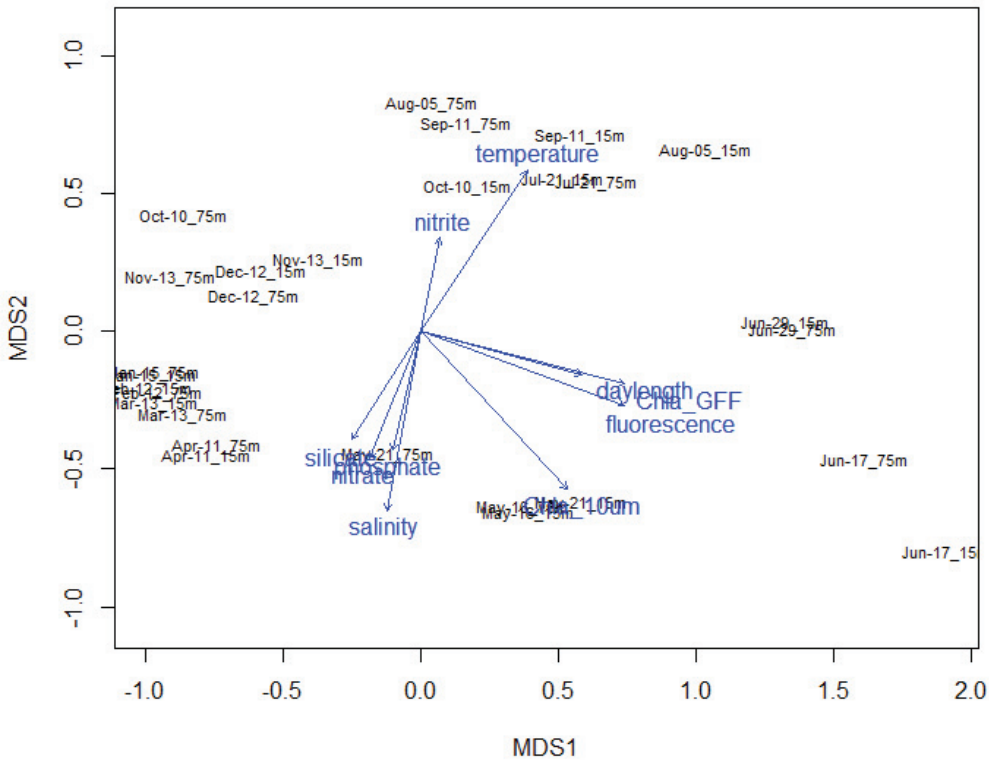
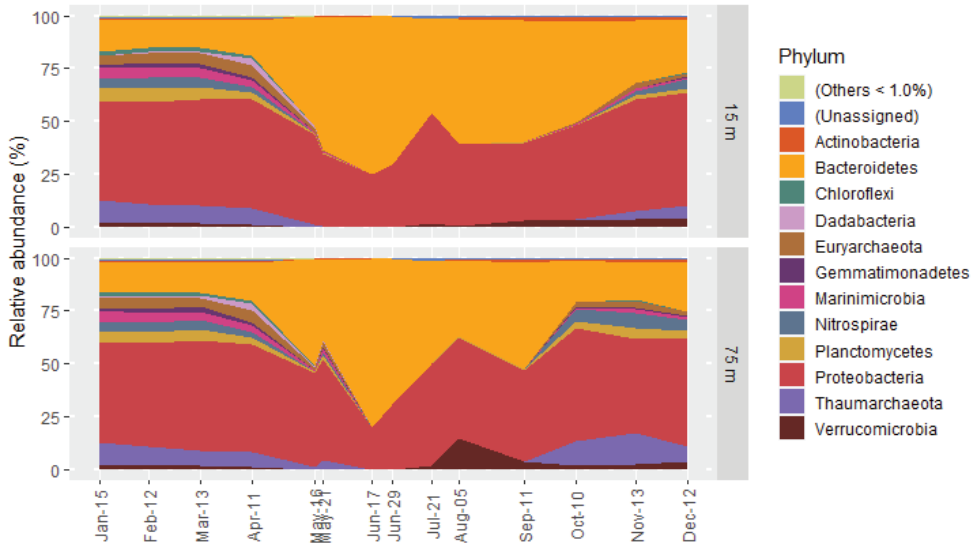
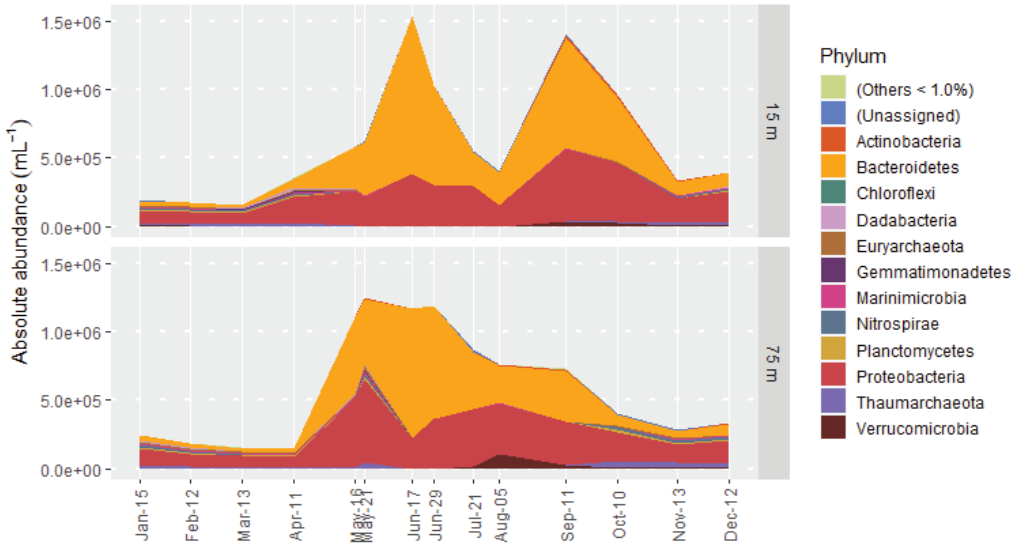


Figure A.3: NMDS ordination plot of the different samples with environmental factors fitted on top. The ordination was based on the Bray-Curtis distance between the different samples **after OTU processing**. The length and orientation of the arrows indicate the strength and direction of each factors role relative to the NMDS plot. Made using the *monoMDS* and *envfit* (999 permutations) functions of **vegan**.

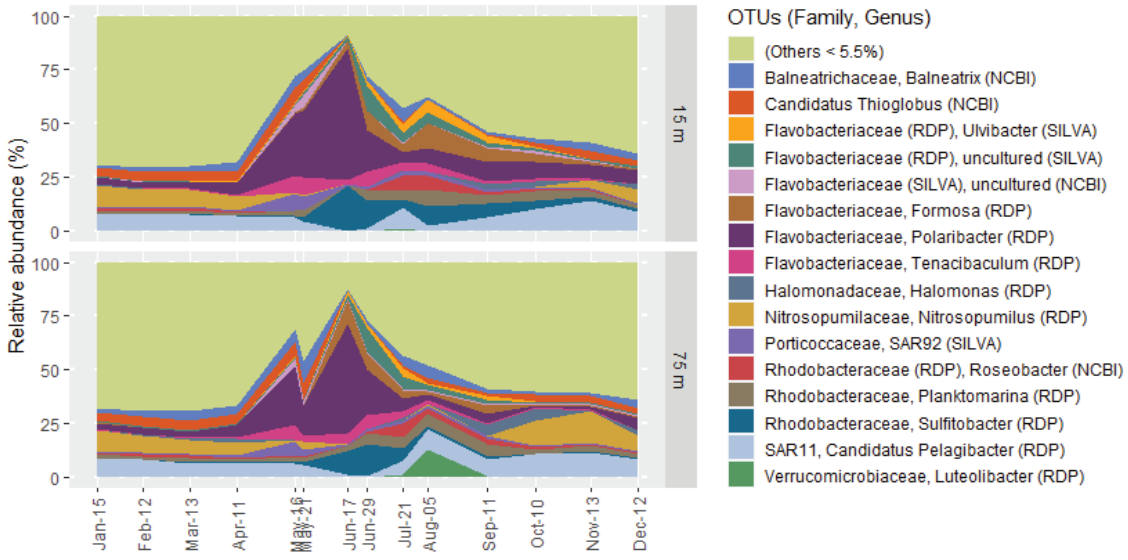


(a) Phylum relative abundance

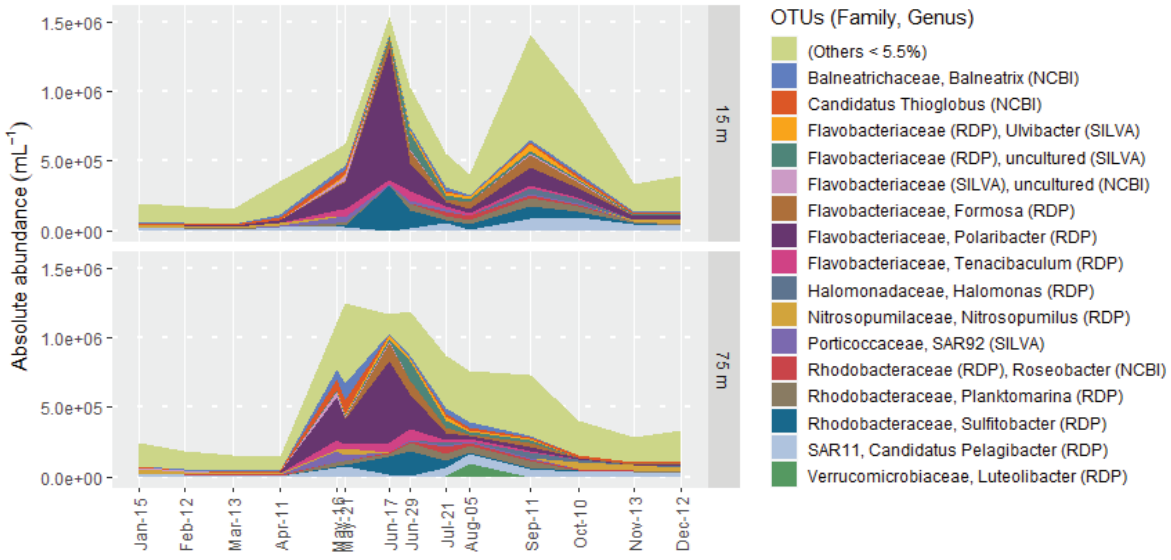


(b) Phylum absolute abundance

Figure A.4: Phyla composition of samples taken throughout 2019. Relative **(a)** and absolute **(b)** abundances are shown. The taxonomic assignments were made **based on the OTU pipeline** and SILVA reference database. The date of each sample is indicated along the x-axis, the areas in between are filled based on linear interpolation for illustrational purposes. For certain dates and depths the cell counts were unavailable and are then also based on linear interpolation (Figure 12). Phyla making up less than 1% of all the samples were pooled together as "Others".



(a) OTU relative abundance



(b) OTU absolute abundance

Figure A.5: Abundances of most numerous **OTUs** throughout 2019. Both the relative abundance **(a)** and absolute abundance based on scaling by flow cytometry **(b)** are shown. The dates of each sample are indicated along the x-axis, the areas in between are filled based on linear interpolation for illustrational purposes. For certain dates and depths the cell counts were unavailable and are then also based on linear interpolation (Figure 12). OTUs with a relative abundance less than 5.5% for all samples were pooled into "Others".

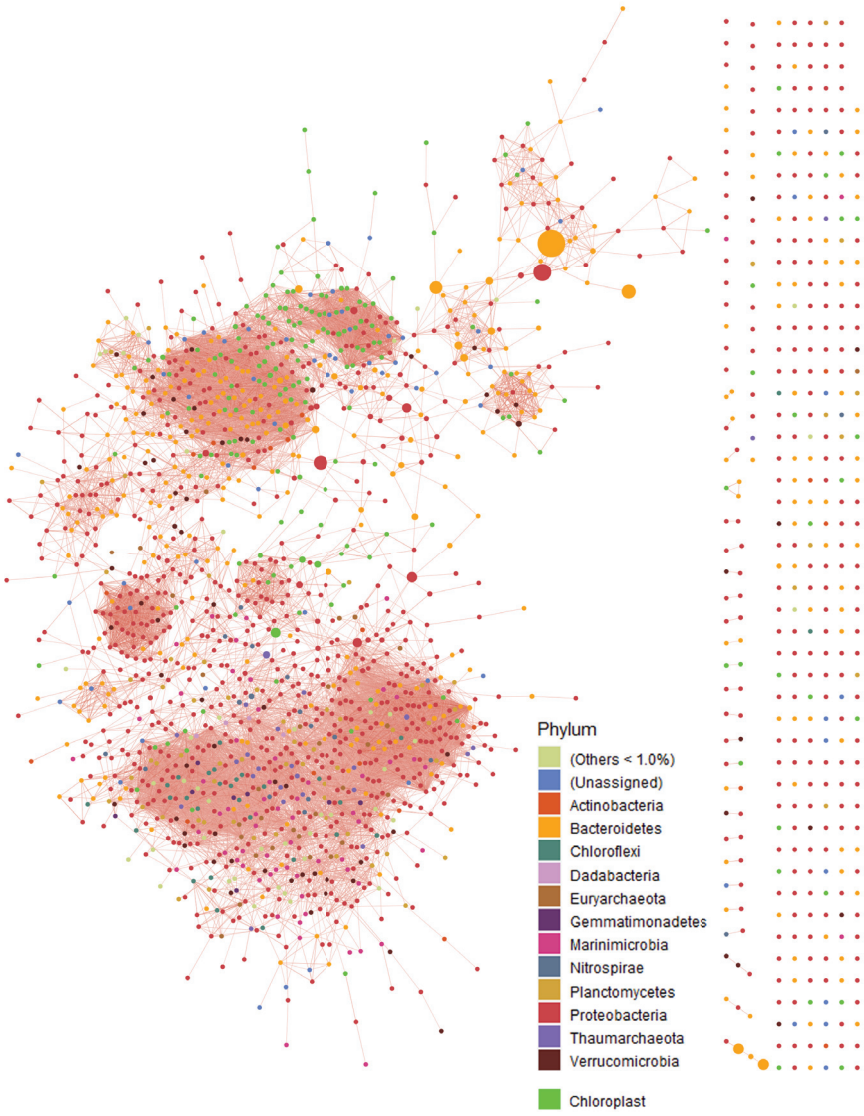


Figure A.6: Correlation analysis of the IsA community **using absolute abundances**. The nodes represent individual zOTUs with sizes scaled to the mean absolute abundance across all samples. Nodes are coloured according to their phylum (similarly to Figure 16a and 16b). zOTUs originating from Chloroplasts were also included. The edges represents the Pearson correlation between the zOTUs across all samples using their absolute abundance. A cut off value of 0.85 was used (no correlations were below -0.85 and thus no negative correlations are shown). Highly connected groups are indicated by arrows and numbers (groups 1-4). *Cytoscape v3.8.0* was used for visualization.

NTNU
Norwegian University of
Science and Technology

UNIS
The University Centre in Svalbard