UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Department of Arctic and Marine Biology

Seasonal abundance of parasitic Marine Alveolate Group II (MALV II) in an Arctic fjord, Svalbard

Stuart Thomson

Master thesis in Biology - 15th May 2014 – BIO-3950 (60ECTS)

Supervisors – Tove Gabrielsen (UNIS) and Else Nøst Hegseth (UiT)

Abstract

During the last decade, knowledge has been building of the parasitic dinoflagellate group Marine Alveolate Group II (MALV II, within Syndiniales). While environmental cloning and sequencing approaches have indicated a high abundance of MALV II throughout the world's oceans, relatively little is known about their seasonality and the significance of their role in the marine ecosystem. No studies to date have addressed either of these issues in Arctic waters. In this thesis, the relative abundance of MALV II rDNA was studied in Adventfjorden, Svalbard during spring 2012. Specific PCR primers were designed to the MALV II 18S V4 region and TaqMan PCR was used to quantify the relative abundances of MALV II 18S rDNA. 18S copy numbers rose from 400,000 – 500,000 copies / mL in February to a maximum of ~ 900,000 copies / mL in April, before a sudden crash to below 100,000 copies / mL after April 19th from which the population did not recover. The oceanographic conditions within Adventfjorden over the season explained the MALV II abundance pattern. Atlantic water incurred into the fjord immediately prior to the study and was replaced with Arctic water around April 19th. The MALV II populations detected in this study were probably transported into the fjord on the Atlantic currents. Attempts to determine a host species by comparing to phytoplankton and zooplankton abundances were unsuccessful. Further studies are required to investigate seasonality amongst MALV II in more Arctic water dominated systems and to determine host species.

Introduction

1. Overview

The focus of this study is an order of parasitic dinoflagellates, known as Marine Alveolate Group II (MALV II), which are often identified as a dominant group in environmental cloning approaches in worldwide marine environments (Koid et al. 2012; Massana 2011; Massana et al. 2011), including in the Arctic (Lovejoy et al. 2006; Bachy et al. 2011). In spite of their abundance in marine systems, little knowledge has been gained of their seasonality and role in the marine ecosystem (but see Chambouvet et al. 2008). This study aims to quantify populations of MALV II over the course of a spring season in a Svalbard fjord to provide insight into the ecological role of these organisms.

This introduction will first detail the oceanographic conditions of Svalbard's west coast fjords before moving on to the marine ecosystems in these environments. Background concerning the current state of knowledge of MALV II will then be presented, followed by an introduction to the molecular tools that will be used to quantify their populations. Finally, a hypothesis for the work will be outlined.

2. Oceanography

The study location, Adventfjorden, is a small fjord forming part of the larger Isfjorden system, the largest fjord system on the west coast of Spitsbergen (Nilsen et al. 2008). Adventfjorden is an open-ended shallow fjord, approximately 8.3 km long and 3.4 km wide. Two rivers, Longyearelva and Adventelva, supply a freshwater input to the fjord from June to October (Zajączkowski & Włodarska-Kowalczuk 2007). Adventelva is one of the largest rivers in the Svalbard archipelago and therefore provides a very large freshwater input to the fjord (Dobrzyn et al. 2005).

Isfjorden is a sill-less fjord, meaning there is a direct interface between water inside the fjord, and the open ocean waters outside. This potentially allows relatively warm Atlantic water (AW) from the West Spitsbergen Current (WSC) to periodically enter the fjord. In some years, AW from the WSC penetrates into Isfjorden and can fill the main fjord basin (Nilsen et al. 2008). AW entering fjords on the west coast of Spitsbergen has been transformed as a result of mixing with Arctic water (ArW) on the shelf and may actually be described as transformed Arctic water (TAW), with decreased temperature and salinity values as compared to AW (Svendsen et al. 2002). The actual temperature and salinity of this TAW can vary greatly year to year however, depending on the extent of mixing (Svendsen et al. 2002).

Whilst the waters in Isfjorden are affected by water masses and ocean currents, the dynamic effects of the Earths rotation also impacts the hydrography (Cottier et al. 2010). Outflowing glacial melt water, i.e. cold and fresh water, will usually hold to the right hand shore with regards to the direction of the outflowing water (Skogseth et al. 2005). This has been documented in Kongsfjorden, another Svalbard west coast fjord, down to a depth of 20 m (Aliani et al. 2004). Water flowing into fjords will adhere to the same principle and be deflected to the right side of the fjord in the direction of the inflow (Skogseth et al. 2005), as has been observed in Isfjorden itself (Nilsen et al. 2008).

Sedimentation in summer and autumn, and ice cover in winter and spring, are determining factors in the irradiance available for photosynthetic organisms in the pelagic (Wiktor et al. 1998; Zajączkowski et al. 2009). The freshwater volume in Spitsbergen glacial-fed fjords is positively correlated to the amount of sediments suspended in the water (Wiktor et al. 1998). This is of particular relevance in Adventfjorden, which is supplied by one of the largest rivers in Svalbard (Dobrzyn et al. 2005). Ice conditions in the spring also play a large role in determining the types and timings of plankton communities in the pelagic environment. In recent years, fjords on the west coast of Spitsbergen have been remaining open during the winter with increasing regularity (Hegseth & Tverberg 2013).

3. Fjord Ecosystems

Environmental conditions vary greatly throughout the year for the fjords on the west coast of Svalbard; there is 24 hour sunlight in summer and 24 hour darkness in winter; there are periodic AW (TAW) incursions; in winter, sea ice cover may create a 'lid' on the marine ecosystem; and in summer there is a large freshwater input from rivers. Therefore biomass and productivity of phytoplankton fluctuate greatly throughout the course of the year (Aliani et al. 2004). Ice covered latitudes can be characterised by 2 peaks in yearly primary production; firstly an ice algal bloom and following the melting of the ice, a phytoplankton bloom (Sakshaug et al. 2009). In the absence of ice cover there will be only a phytoplankton bloom. Spring blooms on the west coast fjords tend to occur around April / May although the timing can vary considerably from year to year (Hodal et al. 2011; Węsławski 1988; Zajączkowski et al. 2009). The start of the spring bloom is coupled with the increasing irradiance and a stratification of the upper water column which traps nutrients at the surface (Aliani et al. 2004), but also to the inoculum seeding the bloom (Hegseth & Tverberg 2013). The end of the phytoplankton bloom is linked to the exhaustion of nutrients such as silicate and nitrate (Hodal et al. 2011; Wassmann et al. 1999). Fjords on the west coast of Spitsbergen are estimated to have productivity levels comparable to that of the Barents Sea (Hop et al. 2002; Wassmann & Slagstad 1993; Hegseth 1998).

There are numerous protists associated with the spring bloom in Svalbard waters. The diatom *Nitzschia frigida* dominates ice algal assemblages in the Barents Sea (Hegseth 1992). Pelagic spring blooms in fjords around Svalbard are generally dominated by other diatom species and the haptophyte *Phaeocystis pouchetii* (Degerlund & Eilertsen 2009; von Quillfeldt 2005; Hodal et al. 2011). There is a general pattern in Arctic spring blooms that pennate chain forming diatoms like *Fossula arctica* and *Fragilariopsis oceanica* bloom before centric diatoms such as *Chaetoceros* spp. and *Thalassiosira* spp., which bloom later (von Quillfeldt 2005; Hodal et al. 2011). Outside of the bloom period, unidentified flagellates 3 – 10 μm in diameter have been found to dominate phytoplankton

communities in Adventfjorden during the polar night (Wiktor 1999). In Kongsfjorden, dinoflagellate populations have been found to be most abundant during the bloom in April, with plastid-containing species such as *Heterocapsa triquetra*, *Scripsiella trochoidea*, and *Gymnodinium arcticum* most prevalent, although still insignificant compared to diatom numbers (Seuthe et al. 2010). In a study in Disko Bay, Greenland, heterotrophic dinoflagellates were estimated to graze 7-18% of the diatom spring bloom (Levinsen & Nielsen 2002). Figures on primary production for dinoflagellates in Arctic fjords are difficult to find. However a recent poster shows data suggesting that most dinoflagellates in the euphotic zone in the Barents Sea north west of Svalbard are autotrophic (Hegseth poster 2013). Dinoflagellates are therefore likely important both as primary producers and as consumers in the marine arctic food web. There have however, so far, been no studies examining the importance of parasitic dinoflagellates in Arctic waters. The body of literature on parasitic dinoflagellates has grown considerably in the past 20 years, and much of this can be attributed to the economic costs that these parasites inflict upon fisheries (Coats 1999). These fisheries generally occur at lower latitudes, although the range of some affected species can reach lower Arctic latitudes.

4. Syndiniales and Marine Alveolate Group II (MALV II)

There is considerable genetic, lifestyle, trophic, and habitat diversity within Dinophyceae (Alveolata). Dinoflagellates can be autotrophic, heterotrophic, parasitic, symbiotic or a combination of the above. For instance, out of nearly 2400 described dinoflagellate species, roughly half contain a plastid, about 80% are found in the marine environment, 7% are parasitic and 1% are mutualistic symbionts (Gomez 2012).

Marine alveolate groups I and II (MALV I and II, and also referred to as MAG I and II) belong to the order Syndiniales within Dinophyceae (Guillou et al. 2008; Bråte et al. 2012). There are no described phenotypic differences between the two MALV groups, most likely owing to an absence of taxonomically informative morphological characters in combination with their small size (Moon-van der

Staay et al. 2001). Instead, MALV I and MALV II are separated based on the DNA sequences of their small subunit (SSU) RNA (Moreira & López-García 2002). Both MALV I and MALV II are entirely endoparasitic and do not contain a plastid (Miller et al. 2012; Small et al. 2012). Sometimes in the literature, 'Syndiniales' is used to refer to MALV II specifically (Gomez 2012; Lin 2011). To avoid confusion, in this thesis I will use the term 'Syndiniales' when referring to both MALV I and MALV II, and the more specific 'MALV II' when referring solely to MALV II.

Very few species have been identified within Syndiniales to date. Seven species of *Amoebophrya* (Miller et al. 2012), two species of *Hematodinium* (Stentiford & Shields 2005), and at least one species of *Syndinium* (Skovgaard & Saiz 2006) have been described and confirmed to belong to MALV II. *Dubosquella,* known to infect ciliates, is the only genus thus far attributed to MALV I (Harada et al. 2007). *Hematodinium* spp. parasitise crustaceans including commercially important species such as the snow crab *Chionoecetes opilio*, the American blue crab *Callinectes sapidus* and the Norway lobster *Nephrops norvegicus* (Stentiford & Shields 2005). *Amoebophrya* spp. parasitize harmful algal bloom-forming dinoflagellate species and may be a significant factor in their control (Chambouvet et al. 2008). *Syndinium* species are known to parasitize copepods (Skovgaard & Saiz 2006). Radiolarians (Bråte et al. 2012), fish eggs (Skovgaard et al. 2009) and bivalves (Noguchi et al. 2013) are also hosts for undescribed species within Syndiniales.

Most of the current knowledge of life histories of Syndiniales comes from *Amoebophrya* spp. (Cachon and Cachon, 1987; Coats et al. 2002; Miller et al. 2012). Three life stages have been identified in *Amoebophrya* sp.; an infectious dispersal stage called the dinospore; an intracellular growth stage called the trophont; and a short-lived extracellular stage known as the vermiform (Cachon and Cachon 1987; Miller et al. 2012). *Amoebophrya* spp. dinospores are free-swimming, small (2 - 5 μm diameter), biflagellate cells (Miller et al. 2012). The dinospore attaches to a host cell and penetrates the cell membrane. Inside the host, the parasite becomes the trophont, where it increases in size and performs a series of nuclear divisions and flagellar replications. The vermiform is formed

at the end of the infection cycle and is only present in the external environment after the host cell ruptures for a short time. It then divides to form many dinospores, ready to begin the infection cycle anew (Cachon and Cachon 1987; Miller et al. 2012). The vermiform may be able to produce 60 to 400 dinospores (Chambouvet et al. 2008).

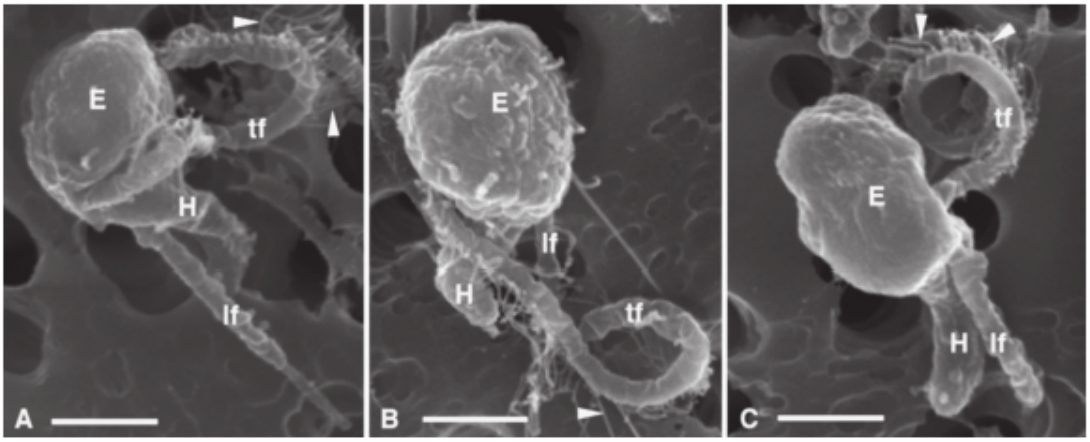


Figure 1. Scanning electron microscope (SEM) images of *Amoebophrya* sp. dinospores from *Akashiwo sanguinea* showing the episome (E), hyposome (H), longitudinal flagellum (lf), transverse flagellum (tf), and flagellar hairs (arrowheads) viewed from several perspectives. **A**. From the left displaying the constriction at the distal end of the longitudinal flagellum and the transverse flagellum emerging into the girdle, which at this point is an indentation of the plasma membrane. **B**. Dorsal showing the continuation of the girdle as a groove formed by the bulbous episome and twist of the narrow hyposome. **C**. From the right. Scale bars = 1 μm.

Figure 1. SEM photos of *Amoebophrya* sp. dinospores, taken from Miller et al 2012

Previously it was thought that protist diversity could be uncovered solely using classical isolation techniques, morphology, and ultrastructure data (Moreira & López-García 2002). However, taxonomic identification of marine microorganisms has been historically difficult, owing to a lack of conspicuous morphological features (Massana et al. 2002). These techniques are also more suited to studies at the individual level. When studying morphologically indistinct single-celled organisms at the population level, molecular tools are invaluable (see Molecular tools section below).

Syndiniales display extensive genetic diversity with at least 44 clades being identified in MALV II to date and 8 in MALV I (Guillou et al. 2008). A wide range of host organisms have been identified to date and clades within Syndiniales have been shown to infect hosts specifically (Chambouvet et al. 2008). Host

specialisation may provide an explanation for the considerable genetic variation within Syndiniales (Chambouvet et al. 2011).

Syndiniales 18S ribosomal DNA (rDNA) sequences frequently dominate marine eukaryote surveys, contributing as highly as 33% (Massana 2011) and 23% (Koid et al. 2012) to the total number of sequences. No DNA sequences corresponding to Syndiniales have been detected from terrestrial or freshwater habitats to date, suggesting the order is solely marine (Guillou et al. 2008). However, Syndiniales environmental DNA sequences have been detected in every marine environment investigated thus far (Guillou et al. 2008). MALV II DNA sequences are found mostly in coastal and oceanic waters, while MALV I sequences have been found in harsh environments such as hydrothermal vents and anoxic waters (Groisillier et al. 2006; Guillou et al. 2008). The high abundance of Syndiniales sequences in the environment makes studying them with molecular tools an attractive possibility.

5. Molecular tools

Molecular tools have been responsible for a revolution in microbial ecology, enabling studies of unprecedented detail in areas such as evolutionary relationships and diversity. The first molecular survey of marine eukaryotes was published just 16 years ago (Rappé et al. 1998). Since then small sub-unit ribosomal RNA (SSU rRNA) studies of eukaryotic diversity have detected new lineages and continue to discover an ever-growing diversity of microbial life (Massana et al. 2002; Moon-van der Staay et al. 2001; Sørensen et al. 2011).

The now commonplace molecular biology technique, polymerase chain reaction (PCR), was invented less than 30 years ago and first reported in the scientific literature in 1985 (Saiki et al. 1985). PCR based techniques have wide ranging applications from clinical diagnostics (Bernard & Wittwer 2002) to phylogenetics (Skovgaard et al. 2009) to biomarker analysis (Kroh et al. 2010) to DNA fingerprinting techniques (Morgante & Olivieri 1993). The technique is used to synthesize specific DNA sequences in a targeted reaction capable of

producing many millions of copies of a starting template molecule. Two short oligonucleotide primers, specific and complementary to a target DNA sequence, bind to the target sequence and enable amplification by a DNA polymerase (Erlich 1989). PCR was later refined with the addition of a third oligonucleotide termed a probe, usually labelled fluorescently, that binds at a site between the two amplification primers and is used to quantify the amount of product produced (Holland et al. 1991). This quantitative technique is now commonly referred to as TaqMan PCR or real-time PCR (RT-PCR). To avoid confusion with another widely used PCR technique, reverse transcription PCR (also referred to as RT-PCR), I will use the term TaqMan when referring to quantitative PCR.

6. Aim of the investigation

Knowledge of the Syndiniales order and MALV II clade has been building since they were first identified at the start of the 2000's. One probable ecosystem role for MALV II is in regulating populations of hosts (Siano et al. 2011) including harmful bloom forming algal species (Chambouvet et al. 2008). Due to the huge diversity and abundances of MALV sequences in all marine environments, they are potentially highly significant ecosystem factors.

I hypothesize that MALV II populations will increase around the spring bloom period in response to high plankton abundances which may act as host populations. In the study I aim to quantify the relative abundance of MALV II organisms in Adventfjorden from February to June 2012 by utilizing a TaqMan PCR approach with MALV II specific PCR primers and a MALV II specific fluorescently labelled probe. I will attempt to link patterns of plankton abundances identified in parallel investigations with MALV II abundances to identify potential host species for MALV II organisms in Adventfjorden. Any changes observed in population size during the study period will help to build the knowledge base and ultimately contribute to the goal of understanding the role that these enigmatic picoeukaryotes have within the Arctic marine ecosystem.

Materials and methods

1. Sampling area

The sampling location (named ISA) was in Adventfjorden at GPS co-ordinates N 78 15 60 E 15 32 05 (Fig. 2). The distance between the ISA station and UNIS is approximately 5.5 km. The depth at the station was around 90 m. Adventfjorden was chosen as the sampling site for this study primarily for logistics reasons. Sampling is often challenging in the Arctic, especially in winter, and the proximity of the fjord to the University Centre in Svalbard (UNIS), ensured sampling could be undertaken at least once a week during the study period.

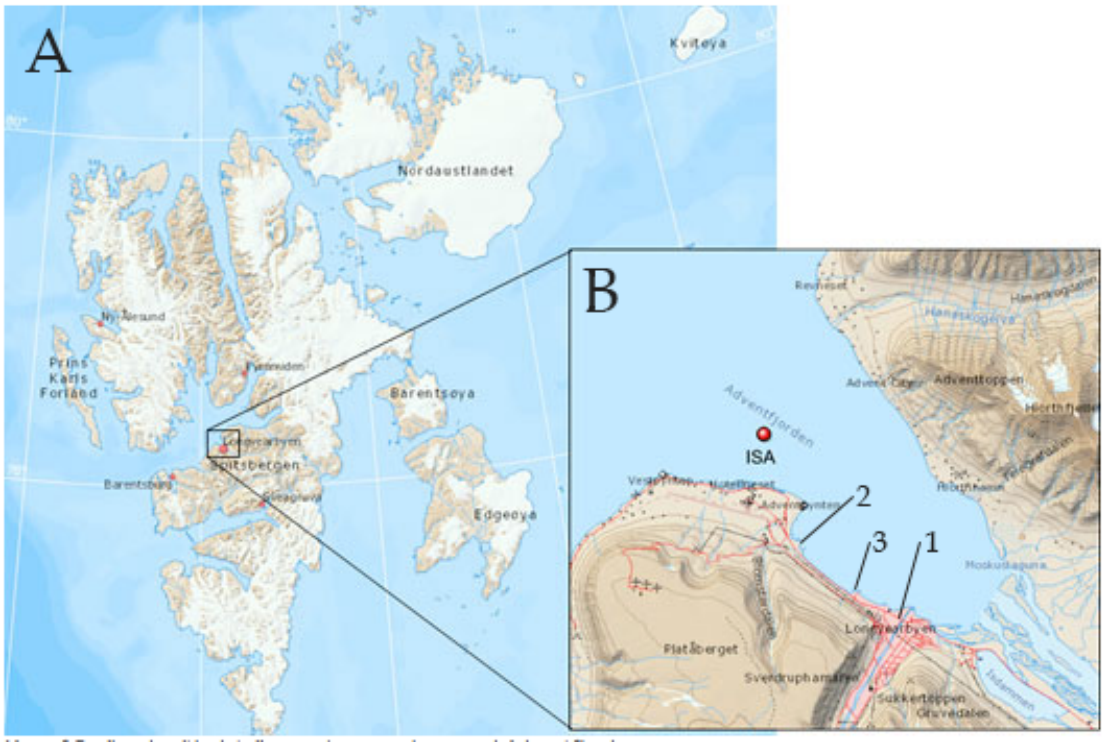


Figure 2 – Location of the ISA sampling station. A – Map of Svalbard. B – Adventforden and the ISA station, 1 – UNIS, 2 – Small boat harbour, 3 – Main harbour.

2. Field sampling

Weekly field sampling was undertaken from February 23rd 2012 until June 21st 2012 utilizing a Rigid Buoyancy Boat (Polarcirkel, Norway) equipped with a custom made winch for lowering and raising equipment into and out of the water. The boat was driven to the sampling area from the small boat harbour (or

the main harbour when ice conditions prevented deployment in the small boat harbour) in Longyearbyen, which took approximately five minutes. The GPS system aboard the boat was used for exact positioning at the ISA station for each sampling.

CTD

Hydrographical data were measured with a CTD (SAIV A/S Norway). The CTD was lowered on ~ 100 m rope from the side of the boat. Care was taken not to exceed this depth to avoid collision with the sea floor.

Light

Irradiance in the water column was recorded using a Licor LI-192 sensor connected to a Licor LI-1000 data logger at the following depths; just above the water's surface; just below the water's surface; 0.5 m; at every m from 1 – 10 m; and at every 5 m from 10 – 45 m. Irradiance could not be measured at the lowest water sampling depth of 60 m (see next paragraph) because the cable, which transmits the reading from the sensor in the water column to the data logger on the boat, measured only 50 m in length. The Licor was calibrated to record irradiance in water. Irradiance measurements were taken at, or as close to as possible, midday.

Water samples

Water samples were collected using a 10 L capacity Niskin bottle (KC Denmark, Denmark) from depths of 5 m, 15 m, 25 m, and 60 m. Three to four L of water from each depth was immediately passed through a 10 μm mesh (KC Denmark, Denmark) by the force of gravity aboard the boat and stored in four litre containers in the dark for subsequent further filtration in the UNIS labs. 90 ml of water from each depth was collected in 100 mL acid washed (0.3 % HCl overnight) bottles for nutrient analysis. The remaining water was collected in 10 L containers and stored in the dark on the boat for the duration of the sampling.

Depending on the weather conditions, the sampling regime described above took one to two hours to complete. After all samples were taken, the boat was driven back to the harbour. Samples and equipment were taken back to the UNIS labs (approximately five minute drive). Equipment was cleaned with fresh water and samples were kept in the dark in the lab until lab work was begun (usually within one hour). Nutrient samples were immediately stored in a -20 °C freezer.

3. Filtrations and fixations

All filtrations and fixations were carried out at the UNIS labs unless otherwise stated. Latex or nitrile gloves were worn for all procedures.

DNA

Three to four L of 10 µm pre-filtered water from each sampling depth was filtered through individual PVDF hydrophilic 0.45 µm filters (Millipore, Germany) under vacuum. The exact volume of water filtered was recorded. Filters were cut into halves using alcohol cleaned forceps and sterile scalpel blades. Care was taken not to touch the filter with gloved fingers. Filter halves were placed into separate labelled cryo tubes, and flash frozen in liquid nitrogen. The next day, cryo tubes were removed from liquid nitrogen storage and placed in a -80 °C freezer.

Chlorophyll a (Chl a)

For all sampling depths: Water from the 10 L containers was used to filter 400 mL water in triplicate through polycarbonate hydrophillic 10 µm filters (Whatman, United Kingdom) by gravity (for the cells > 10 µm). Another 400 mL water was also filtered in triplicate through 0.7 µm glass microfiber GF/F filters (Whatman) under vacuum (for total Chl *a*). Filters were folded in half twice, wrapped in aluminium foil, and immediately transferred to a liquid nitrogen tank. The next day, cryo tubes were removed from liquid nitrogen storage and placed in a -80 °C freezer.

4. Analyses

Nutrient analysis

Samples were shipped to The Arctic University of Norway (UiT). A Flow Solution IV Analyzer (OI Analytical, United States) was calibrated with reference sea water (Ocean Scientific International Ltd, UK) and used to determine nitrate (NO_3), nitrite (NO_2), phosphate (PO_4), and silicic acid ($Si(OH)_4$) concentrations in the samples. Samples were split in three so triplicate measurements could be recorded.

Chlorophyll a measurements

Chl *a* filters were placed in 10 mL methanol in glass bottles (Assistant, Germany,) for 20 – 24 hours at 4 °C in the dark. The methanol was then aspirated into a 10 ml syringe and forced through a 0.22 μm cellulose acetate filter (Whatman, United Kingdom) into a 10 mL cuvette. Fluorescence of Chl *a* was measured with a 10-AU-005-CE Fluoremeter (Turner Designs, United States) and recorded.

5. Molecular analyses

DNA extraction

DNA was extracted from filter halves in independent procedures to guard against loss of samples due to contamination issues or reaction failure. A negative control, that is a tube treated by the same extraction protocol but in the absence of a filter, was included every time the protocol was carried out. DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Germany) with some modifications to the manufacturers protocol as follows. Filter halves were cut into 4 strips and placed in 2 ml safe lock eppendorf tubes along with 1/3 vial of sterile disruption beads (OPS diagnostics, United States) and 100 µL buffer AP1 from the Qiagen kit. Filters were disrupted for 2 x 60 seconds at 30 Hz with a MM301 Mixer Mill (Retsch, Germany). Subsequently 300 µL buffer AP1 was added to the tube and the tube contents were mixed thoroughly. Next, 400 µL liquid was removed from the tube and retained whilst disruption was repeated

with a fresh 400 µL buffer AP1. All buffer was then aspirated from the disruption tube and combined with the previously retained 400 µL. The rest of the extraction proceeded as per manufacturers kit protocol. DNA was eluted from the column twice with 75 µL kit buffer AE to give a total volume of 150µL. This DNA was stored at -80 °C until use.

Extracted DNA was tested by PCR reaction with eukaryotic primers (short 28SF – GTGTAACAACTCACCTGCCG / short 28SR – GCTACTACCACCAAGATCTG Vestheim and Jarman, 2008). Reactions were comprised of 1 U Dreamtaq polymerase (Thermo Scientific, United States), 1x Dreamtaq buffer, 125 nm primers, 200 nm dNTPs, and 1 µL DNA template in a 20 µl reaction volume. Thermal cycling conditions were as follows; denaturation at 94 °C for 2 minutes, 25 cycles of (denaturation at 94 °C for 10 s, 59 °C annealing for 30 s, 72 °C extension for 30 s), extension at 72 °C for 5 minutes. PCR reactions were run on 2 % agarose gels with 1 x Gel Red (Biotium, United States) for 60 minutes at 100 V. Gels were viewed under UV light to ensure presence of a DNA band for each sample extraction at the expected size, and the absence of a DNA band in negative controls. Digital gel photos were recorded. If a DNA band was observed in the negative control, all samples from the extraction batch were considered to be contaminated with external DNA and were not included in further experiments.

If both halves of a filter were successfully extracted and free from contamination, DNA from the two extractions was combined for a total volume of ~300 µL. This DNA, termed the stock, was stored at -80 °C in a freezer for long term storage. Working aliquots were taken from the stock as needed and stored at -20 °C in a freezer.

Primer design

Adventfjorden MALV II 18S rDNA variable region 4 (V4) sequence alignments from samples collected in Spring 2009 prepared by Nikolaj Sørensen (UNIS master student) were obtained. Three of the most recurring sequences from Sørensen clone libraries (100609/24, 300509/11, and EU817941, see

supplementary data file 'Syndiniales_Nikolaj.aln') were used to design specific primers for MALV II (Sørensen et al. 2011). Two forward primers (STF2, STF3), two reverse primers (STR1, STR2) and one fluorescent probe (STF1) were designed, each matching 100 % to all three mentioned MALV II sequences. Primers were also BLAST searched and determined to be specific for MALV II sequences and contained mismatches for non MALV II organisms. One forward primer (ALV01 from Guillou et al. 2008) and one eukaryotic reverse primer (TAREukREV3 from Stoeck et al. 2010) were also tested.

Standard desalted primers were ordered from Eurofins MWG Operon (Germany) and Biomers.net (Germany). HPLC purified fluorescent probes were ordered from Life Technologies (now acquired by Thermo Scientific, United States) and Integrated DNA Technologies (United States).

Table 1. Names and sequences of PCR primers and probes used in PCR and TaqMan experiments to quantify MALV II populations.

Primer	Sequence
STF1	5'-6-FAM-ATTGCGTTACGACGAACTACTGCG-A-MGB-3'
STF2	5'-TAACGCGTCAGAGGTGGA-3'
STF3	5'-CCTCTTGAAAAACAGCTCTGCG-3'
STR1	5'-ATCCCCTAACTTTCGT-3'
STR2	5'-GTTTATGGTTAAGACTAC-3'
ALV01	5'-AGAGTGTTCACGGCAGGC-3'
TAREukREV3	5'-ACTTTCGTTCTTGATYRA-3'

N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N G C T C G T A G T T G
G A T T T C T G T C G A G G A C A G C C G G T C C A C A C T T T G T G T G A G T A T C T G A T T T G
G C C T T G G C A T C C T C T T G A A A A A C A G C T C T G C G C T T G A T T G C G T G G A C T G
G A G T T C A A G A C T T T T A C T T T G A G G A A A T T A G A G T G T T C A C G G C A G
G C A A A C G C C T T G A A T A T A T T A G C A T G G A A T A A T A A T A T A G G A C C T T T
G T T C T A T T T T G T T G G T T T C T A G A G C G A A G G T A A T G A T T A A T A G G G A
T A A T T G G G G G C A T T C G T A T T A A C G C G T C A G A G G T G G A A C T C T T G G A T T
G C G T T A C G A C G A A C T A C T G C G A A A G C A C T T G C C A A G G A T G T T T T C A T T G A T
C A A G A A C G A A A G T T A G G G G A T C G A A G A C G C T T A G A T A C C G T C G T A G T C T
T A A C C A T A A A C T A T G C C G A C T A G A G A T T G G A G G T C G T T C T T T T T A T G A C T C
C T T C A G C A C C T T A T G A G A A A T C A A A G T C T T T G G G T T C C G G G G G G A G T A T G
G T C G C A A G G C T G A A A C T T A A A G G A A T T G A C G G A A G G G C A C C A C C A G
G A G T G G A G C C T G C G G C T T A A T T T G A C T C A A C A C G G G A A A A C T T A C C A G G T C
C A G A C A T A G T G A G G A T

Figure 3. 18S rDNA V4 region sequence from Sørensen clone library (clone 300509_11_Euk528f - Sørensen et al. 2011) with PCR primer sequence positions detailed. From top to bottom, primer sequences highlighted are; STF3 (black); ALV01 (green); STF2 (red); STF1 (purple); TAREukREV3 (blue); STR1 (yellow); and STR2 (pink). TAREukREV3, STR1, and STR2 are all reverse primers, therefore sequences are complimentary to highlighted positions.

Primer	STF2	ALV01	STF3
TAREukREV3	-3.5	-2.7	-2.8
STR1	-3.1	-3.1	-2.1
STR2	-2	-2.2	-3.2

Structure/Primer	STF2	ALV01	STF3	TAREukREV3	STR1	STR2
duplex	-8.8	-4.5	-3.4	-8.6	-0.6	-1.2
hairpin		0.3	0.6	1.1		

Figure 4 PCR primer secondary structure energies in kcal/mol as calculated by NetPrimer software by Premier Biosoft. Upper table shows interactions between different PCR primers. Lower table shows interactions of PCR primers with themselves.

The values presented in Figure 4 describe the strength of the bonds formed when primers bind to themselves or to other primers. The more negative these values, the more efficiently structures are formed which will prevent binding to the target template molecule. Generally, energies higher than -5 kcal/mol can lead to problems with primer dimerization in PCR reactions. None of the primers show binding to a different primer of high enough energy to warrant concern (Fig. 4 - upper table) . STF2 and TAREukRev3 primers do however form dimers with themselves which could be problematic (Fig. 4 – lower table). ALV01 also displays the ability to efficiently dimerise with itself. ALV01 or STF3 forward primer in combination with either STR1 or STR2 reverse primer are the primer combinations which give the lowest overall energy secondary structures.

Constraints are placed on where the primers can be positioned in the V4 region because the sequence must not only be shared between multiple MALV II clones, but also differ from all other organisms. Therefore the STF2 primer has higher dimerization energy than is optimal.

Primer testing

Primers were tested extensively using PCR and TaqMan PCR. Another quantitative PCR method whereby a fluorescent cyanine dye was used to stain all double stranded DNA products was also tested. The method is less specific than TaqMan PCR as it does not require the use of the fluorescent probe. The primers which gave the highest specificity and most efficient amplification during PCR were used for TaqMan assays of the sample DNA. The TaqMan PCR method was chosen over the fluorescent dye method for quantitative analysis owing to issues of specificity.

Plasmid preparation

A plasmid with a MALV II DNA insert was used to generate standard curves in TaqMan PCRs. The plasmid DNA was obtained from UNIS PhD student Archana Meshram (Meshram, paper in prep.) and consisted of an 871 bp MALV II 18S rDNA fragment (including the V4 region), cloned into the pJET1.2 vector. The insert sequence was 100 % complementary to the primers and probe to be used for amplification of MALV II DNA.

The plasmid DNA needed to be multiplied before use. This is commonly achieved by transforming the plasmid into bacterial cells and taking advantage of the fast growth rate of the cells to increase the amount of plasmid DNA. JM107 cells (Thermo Scientific, United States) from a glycerol stock stored at -80 °C were plated onto an agar plate and grown overnight at 37 °C. The next day, a sterile pipette tip was used to innoculate two ml LB broth with cells from a single colony that had grown on the plate. LB broth cultures were grown overnight at 37 °C with shaking. These cultures were used the next day in conjunction with the TransformAid Bacterial Transformation Kit (Thermo Scientific, United States) to transform JM107 cells with 20 ng plasmid DNA. Positive (pUC19) and

negative (water) controls were included and the procedure was carried out as per manufacturers instructions. Transformed cells were plated on LB-ampicillin (50 µg/µL) agar plates and grown at 37°C overnight. The plasmid confers resistance to ampicillin to the cells, therefore only cells that have taken up the plasmid should survive and be able to grow. Single colonies from these plates were picked with sterile pipette tips the next day and grown in individual NUNC tubes (Thermo Scientific, United States) with 4 mL LB-ampicillin broth overnight at 37°C with shaking. The EZNA plasmid mini prep kit (Omega Bio-Tek, United States) was used to isolate plasmid DNA from cultures according to the manufacturers instructions. The DNA was stored at -80 °C in a freezer until use.

Standard curve preparation

The plasmid was linearised before being used as template in TaqMan PCRs because circular plasmids do not give reliable DNA copy numbers in quantitative PCR (Hou et al. 2010). 1.5 µg plasmid DNA was linearised with 5 U HindIII (Thermo Scientific) for two hours at 37°C, followed by a ten minute enzyme inactivation step at 85 °C. The product was cleaned using the EZNA Cycle Pure Kit (Omega Bio-Tek) according to the manufacturers recommendations. The linearized plasmid DNA concentration was then measured using a Nano Drop 2000 (Thermo Scientific).

Preparations of standard concentrations were then calculated as follows;

Plasmid – 2974 bp
Insert – 871 bp
Plasmid + Insert – 3845 bp

Average mass of one base pair = 1.091×10^{-21} g.

Mass of one plasmid = (Total plasmid length) x (Mass of one base pair)
Mass of one plasmid = $3845 \times (1.091 \times 10^{-21}) = 4.214 \times 10^{-18}$ g

Mass for highest standard concentration (3×10^7 copies) =

(Mass of one plasmid) x (Required copy number) =
$(4.214 \times 10^{-18}) \times (3 \times 10^{7}) = 1.264 \times 10^{-10}$ g

Correction for template volume in TaqMan reaction (2 µL):
$(1.264 \times 10^{-10}) / 2 = 6.32 \times 10^{-11}$ g

Plasmid DNA was diluted from the concentration measured on the Nano Drop to 6.32×10^{-11} g/µL to give 3×10^{7} copies per 2 µL volume. This was used as the highest concentration in the standard curve. Subsequent 10 fold dilutions of this standard gave concentrations of 3×10^{6} copies, 3×10^{5} copies, 3×10^{4} copies, and 3×10^{3} copies. Some experiments were also run with a sixth standard of 300 copies.

TaqMan

TaqMan PCRs were set up on ice in 48 well plates (Applied Biosystems, United States) with 1x TaqMan environmental master mix (Applied Biosystems, United States), 750 nM forward primer (STF2), 750 nM reverse primer (STR2), and 250 nM probe (STF1), 2 µl DNA template, made up to 20 µl final volume with milliQ water. Plates were sealed with MicroAmp 48-well optical adhesive film (Applied Biosystems, United States) and run immediately on a Step One Real-Time PCR machine (Applied Biosystems, United States). Thermal cycling conditions were as follows; 40 cycles of (denaturation at 94 °C for 30 seconds, annealing at 50 °C for 45 seconds, extension at 72 °C for 60 seconds). Fluorescence was read by the machine during the extension phase of every cycle.

Standard curves were run on assay plates using plasmid DNA at the standard copy numbers determined above. The standard curve dilutions were made fresh for every new experiment, from a working solution that had been diluted from the linearised plasmid stock. Five standard concentrations were typically used to generate the standard curve. These standard concentrations were included on the plate in triplicate. Ten fold diluted field campaign samples (ISA samples) were run in triplicate alongside standard curve reactions . Sample copy numbers

were determined from standards on the same plate. As long as the number of copies of DNA in ISA samples was between the lowest and highest standard, the number of copies in the sample could be inferred from the standard curve.

Samples were run in at least 3 independent experiments for reasons of accuracy and precision. If there appeared to be an outlier in the copy numbers determined by the three experiments, or if the three experiments produced inconsistent results, a fourth experiment was run to improve accuracy.

6. Data analysis

TaqMan analysis

Step One software v2.1 (Applied Biosystems, United States) was used to collect data from the PCR machine. cT threshold was set to 0.1 for all assays and checked manually for appropriateness. This value was deemed appropriate in all experiments and its consistency enables comparison of standard curves between experiments. Standard curves were checked visually and then the software-calculated reaction efficiency and correlation coefficient were checked. The efficiency value is expressed as a percentage and represents the efficiency with which template molecules were copied during the PCR. A value of 100 indicates that the number of template molecules was doubled with each cycle during the reaction, ie what it theoretically expected. A lower value indicates a problem with the reaction such as primer dimerization, incorrect thermal cycling conditions etc. The correlation coefficient is expressed as a number with a highest value of 1, and represents how well the data fit to a linear regression line. Variability between replicates and a poorly optimised assay will lower this number. Only experiments which fell into a manufacturer recommended range were included in calculations of MALV II copy numbers. These experiments were required to have an efficiency between 90 and 105% and a correlation coefficient of at least 0.99. If the experiment fitted the criteria, cT values for sample triplicates were checked visually and outliers excluded if necessary. Data were then exported from the Step One software as excel files and further analyzed in excel.

In order to infer the number of copies of MALV II 18S DNA per ml of sea water, several calculations needed to be made from the raw copy number determined from the cT value in the TaqMan experiment. This raw copy number from the TaqMan only gives the number of DNA copies present in the DNA sample loaded in the experiment. Therefore the following calculations were used to calculate the DNA copy number in one mL seawater;

Mean DNA copy number from triplicate reactions = Copy number per 2 µL DNA sample

(Copy number per 2 µL DNA sample) / 2 = Copy number per 1 µL DNA sample

Correction for 1:10 DNA sample dilution: (Copy number per 1 µL DNA sample) x 10

Total DNA copies extracted from filter = (Corrected copy number for 1 µL extracted DNA) x (Total volume of DNA extracted: 150 µL or 300 µL depending on whether either half or the whole filter was extracted successfully)

Number of copies in 1 mL sea water = (Total DNA copy number extracted from sample filter) / (Total volume of water filtered in mL)

An excel sheet (MALV_DATA.xlsx) can be found on the supplementary data CD which includes the copy number per mL sea water calculations for every sample.

Statistics

Statistical tests were carried out by UNIS post doc Marie Davey to determine the significant factors affecting the populations of MALV II. The data was not normally distributed, and was log transformed to allow for improved fitting of linear mixed models (see supplementary data, Statistics chapter). Records from a single sampling date (March 8th) were removed from the analysis due to missing environmental data. Correlation was tested between the log-transformed copy

number and each of the predictor variables of julian date, depth, total Chl *a*, salinity and temperature (Table S1). Stepwise model selection was then conducted by forward selection with variables added in order of descending correlation and with the inclusion of a random effects term representing within sample variation. Successive models were compared by ANOVA. P-values were estimated in the best fitted model using MCMC approximation. Additionally, models were fitted for a subset of 10 samples for which zooplankton abundance data were available (courtesy of UNIS PhD student Eike Stübner). Models were not significantly improved by the addition of zooplankton abundance data (data not shown), therefore this data was not included in the final model.

Results

1. Hydrography

Seawater temperatures, salinities and potential densities throughout the spring of 2012 (Figs 5, 6, and 7 respectively) were extracted from the vertical CTD profiles of the sampling site. The figures were prepared by Eike Stübner.

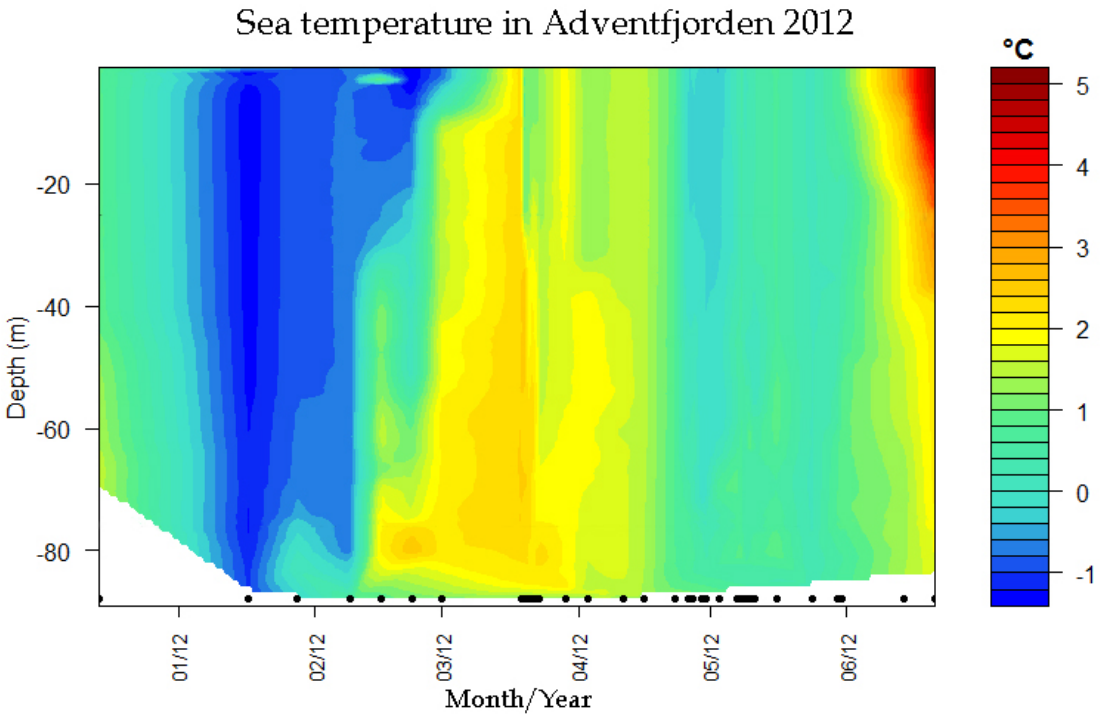


Figure 5 – Seawater temperature profile during Spring 2012 from the Adventfjorden sampling site ISA. Black dots represent CTD profiles taken during sampling from which the graphs are interpolated.

A well-mixed and cold water column was observed in December 2011 and January 2012. The water column cooled during this period from a high of approximately 1 °C at the beginning of December to a low below -1 °C in mid January. At the very end of January, there was an increased temperature towards the bottom of the water column at around 80 metres depth (~0.2 °C) relative to the rest of the water column (~-0.8 °C). This was the beginning of a warm water incursion into the fjord which can be clearly seen in the second half of February, during which time temperatures reached > 2 °C towards the bottom. In March,

the warm water spread throughout the water column and on March 23rd fairly uniform temperatures of around 2.4 °C from the surface to the bottom were recorded. In the last week of March and throughout April, temperatures fell until a low of approximately -0.4 °C was recorded throughout the upper 40 m in the first week of May. The upper 40 m warmed slightly and the whole water column was well mixed (with respect to temperature, ~0.6 °C) through the remainder of May. In June, warm water of over 4 °C incurred at the surface and had a warming effect on the entire water column. Temperatures of 1.8 °C were recorded at 80 metres depth at this time.

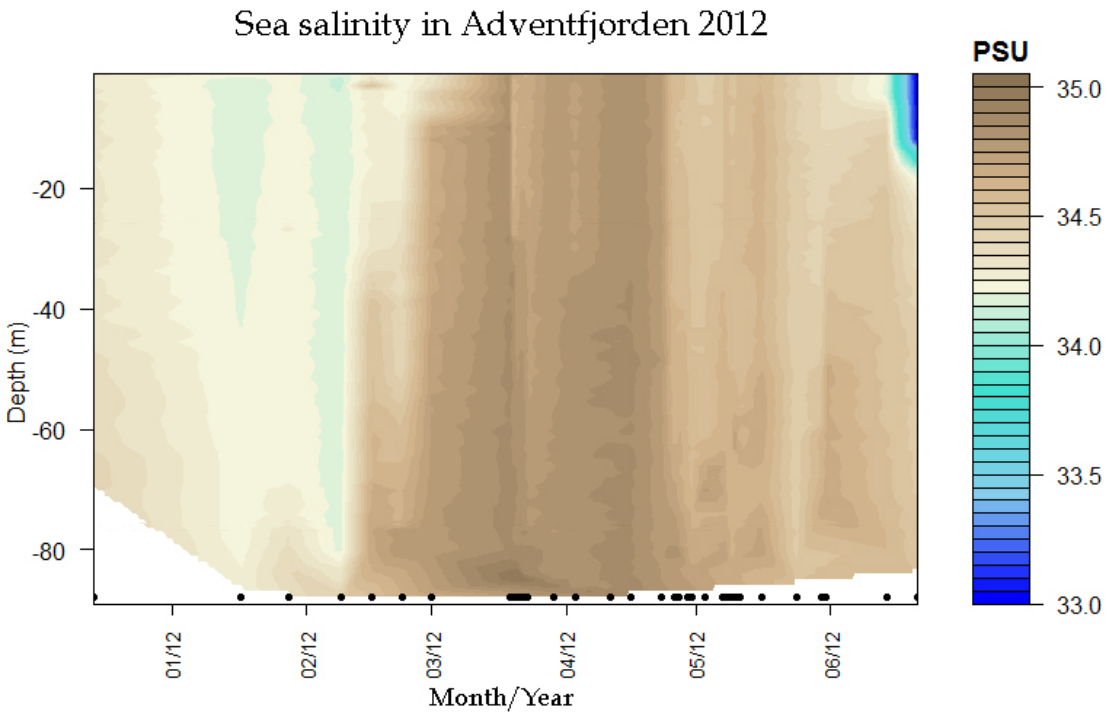


Figure 6 – Seawater salinity profile during Spring 2012 from the Adventfjorden sampling site ISA. Black dots represent CTD profiles taken during sampling from which the graphs are interpolated.

In mid January, a fairly uniform salinity of around 34.2 PSU was observed throughout the water column. Slightly higher salinities of 34.4 PSU were seen at the bottom of the fjord at the very end of January, in keeping with the higher temperatures also seen at this time. This warmer more saline water is representative of the beginning of an incursion by Atlantic water. Throughout the second half of February, this more saline water spread upwards in the water column with salinities of around 34.7 PSU recorded towards the bottom. Through March and April, the Atlantic water incursion led to maximum salinity

values of around 35 PSU, with salinities of 34.8 PSU recorded throughout most of the water column. In May, the salinity dropped slightly to a uniform value of 34.6 PSU in mid May (34.5 PSU in late May) throughout the water column, concurring with the temperature data that the water column is well mixed at this time. In the second half of June, where high surface temperatures were observed in Figure 5, there were lower salinities of around 33 PSU. These temperatures and salinities are consistent with a freshwater input which is expected at this time of year. Large differences in salinity over a short depth range resulted in a stratification at around 15 m depth in the second half of June.

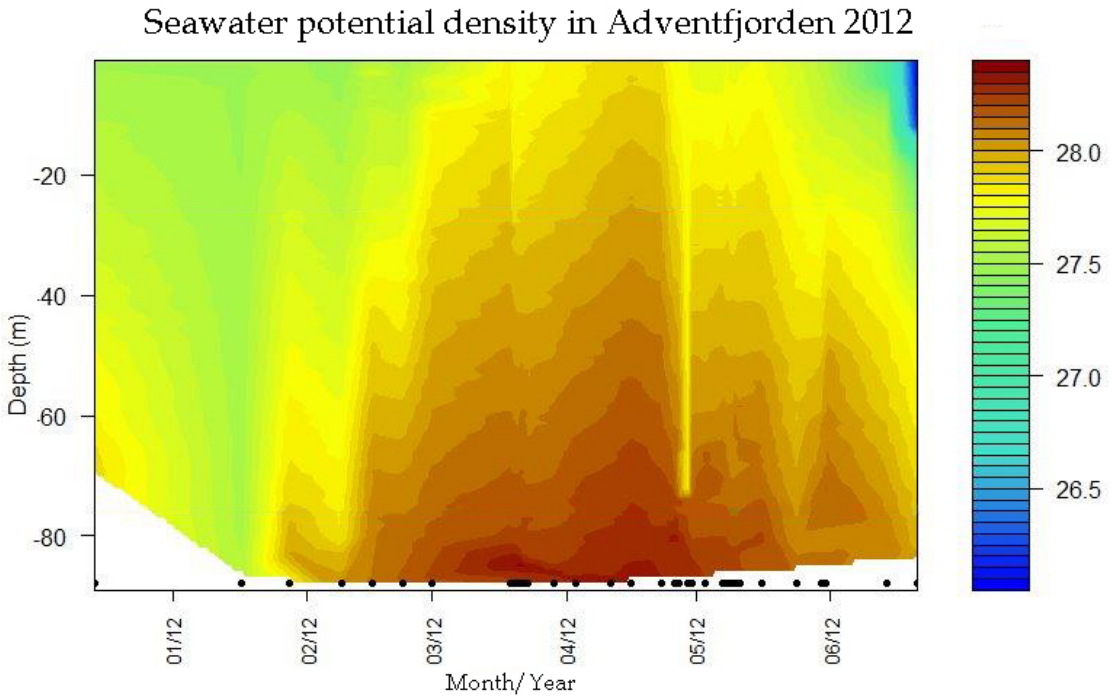


Figure 7 - Seawater density profile during spring 2012 from the Adventfjorden sampling site ISA. Black dots represent CTD profiles taken during sampling from which the graphs are interpolated.

At the beginning of December, slightly higher potential densities were recorded at the bottom of the water column (27.8) than in the surface layers (27.5). By mid January, densities of around 27.5 recorded throughout the water column. At the very end of January, the Atlantic water incursion at the bottom of the fjord resulted in water with increased densities of up to 28. Throughout February this higher density water spread upward into the water column and by March densities at the surface were affected by the incursion at the bottom. In mid-

March, densities at the surface had increased to 27.8, whilst at the bottom, season maximum densities of around 28.5 were recorded. This core of high density water persisted through to mid April at below 80 m depth and densities at the surface were further increased at this time to season maximums of 27.9. In the first half of May, densities at all depths throughout the water column decreased by around 0.1 compared with mid April values. The upper 20 m returned to lower values of around 27.8. Densities further decreased through the second half of May, with surface and bottom values recorded at 27.7 and 28.1 respectively. On June 21st, the lowest salinity values of the season were recorded at the surface and down to about 12 m depth. The potential density value of 26 increased to approximately 26.7 over a depth 1 to 2 metres, resulting in a pycnocline at around 12 m.

2. Light

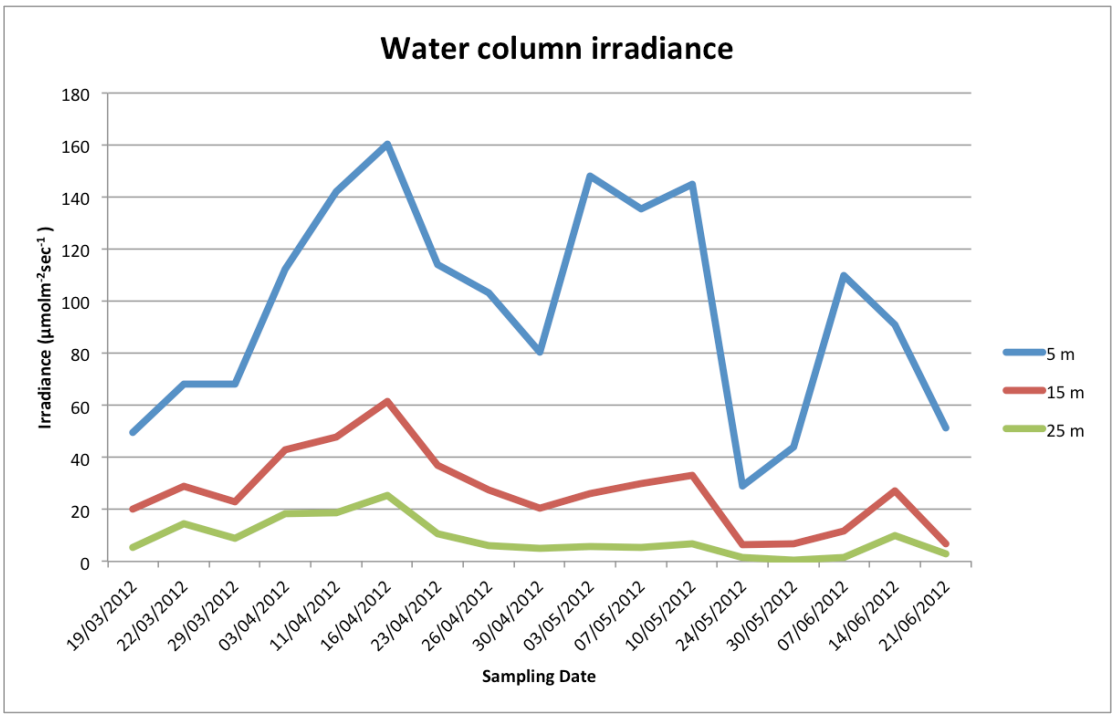


Figure 8 - Irradiance in the water column at 5 m, 15 m, and 25 m depths in spring 2012 at the Adventfjorden sampling site (ISA)

From March 19th until April 16th there was a continuous trend of increasing irradiance with progressing date. On March 19th an irradiance at 5 m depth of around 50 µmolm^{-2}s^{-1} was recorded, which rose to 160 µmolm^{-2}s^{-1} at the same depth on April 16th. On the same dates, the irradiance at 25 m was about 8 and

25 $\mu molm^{-2}s^{-1}$ respectively. This period corresponds to the change in season, with days becoming increasingly bright until April 18th when the midnight sun period begins and then persists to the end of the study period. The highest irradiance values are always seen at 5 m depth because irradiance decreases with depth as a consequence of absorption and scattering of photons by the water column. The irradiance in the water column decreased during the period from April 16th to April 30th. On May 3rd, the 5 m irradiance increased again to over 140 $\mu molm^{-2}s^{-1}$ but a respectively similar increase was not seen at 15 m and 25 m depths. Between May 3rd and May 10th, the irradiance at 5 m depth was stable at around 140 $\mu molm^{-2}s^{-1}$, but at 15 m depth, an increase from 20 to 35 $\mu molm^{-2}s^{-1}$ was recorded. The lowest irradiance of the season was observed on May 24th, with a value of 30 $\mu molm^{-2}s^{-1}$ at 5 m depth. In June, more fluctuation was recorded, with a maximum of 110 $\mu molm^{-2}s^{-1}$ at 5 m being recorded on June 7th, which dropped to 50 $\mu molm^{-2}s^{-1}$ at the same depth 2 weeks later. At 15 m and 25 m, the highest values in June were mismatched with the 5 m recordings, peaking on June 14th with values of 30 and 10 $\mu molm^{-2}s^{-1}$ respectively.

3. Nutrients

Nutrient concentrations are presented from 25 m depth. Nutrient concentrations at the other three sampling depths were largely similar to those recorded at 25 m across the sampling campaign. Six additional data points were available for the creation of graphs from the 25 m depth (the six sampling dates where only water from 25 m was collected). The graphs for 5 m, 15 m and 60 m depths, and the data used to plot them, are available in the supplementary data, Nutrients chapter.

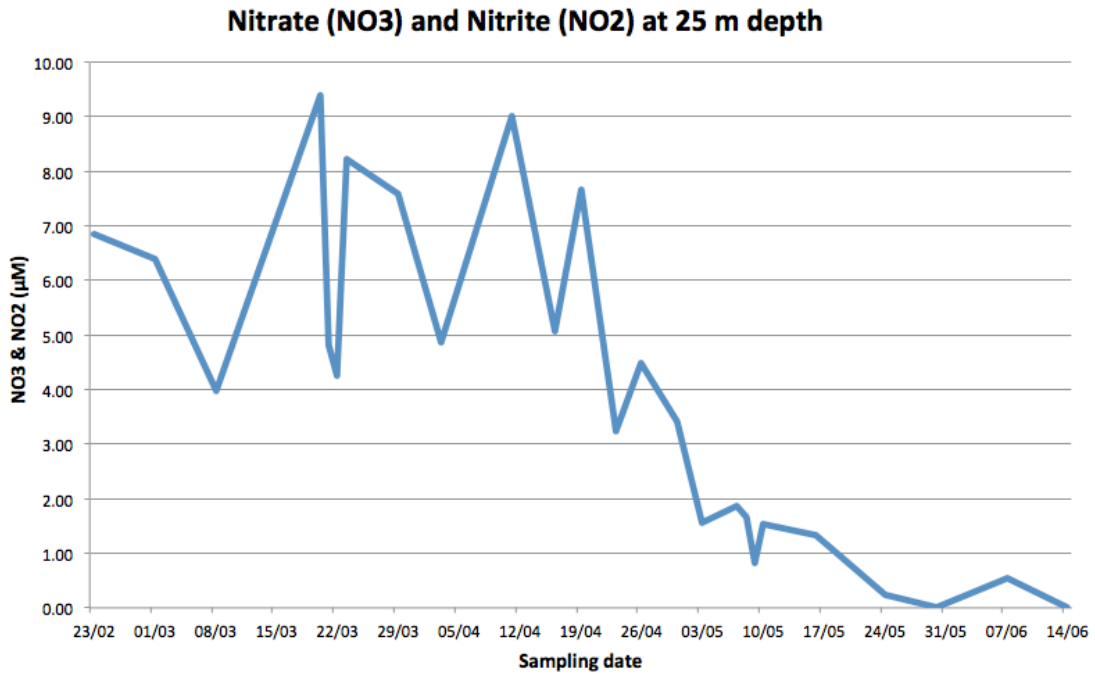


Figure 9 - Nitrate & Nitrite concentrations in spring 2012 at a depth of 25 m at the Adventfjorden sampling site ISA

On February 23rd, the combined nitrate and nitrite concentration was 7 µM, and this decreased to 4 µM by March 8th. After this date and until April 19th, values fluctuated between about 4 and 9 µM, with no discernable pattern to the fluctuation. After April 19th, a decline was observed until the end of the sampling season. Concentrations dropped to below 1 µM on May 10th and total depletion was seen by the end of May. A small increase to 0.5 µM was recorded on June 7th but concentrations returned to zero again on June 21st.

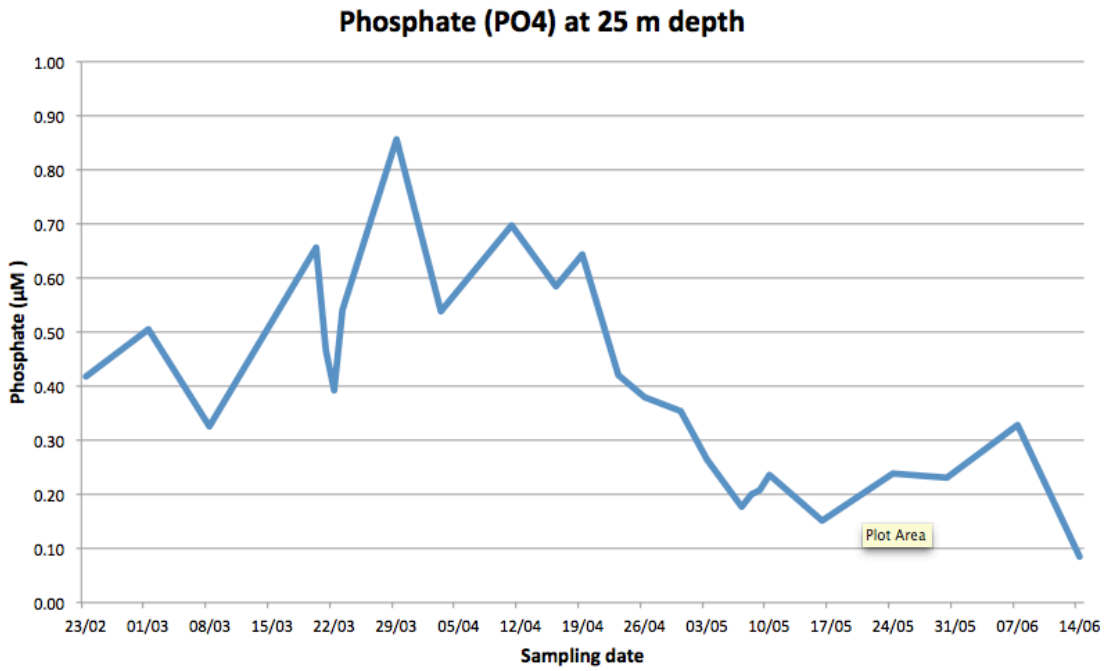


Figure 10 - Phosphate concentrations in spring 2012 at a depth of 25 m at the Adventfjorden sampling site ISA

A phosphate concentration of around 0.4 µM was recorded on February 23rd. A general upward trend was seen for the next month until March 29th, when concentrations reached a maximum of around 0.85 µM. Concentrations then decreased on April 5th but remained stable around 0.55 – 0.7 µM until April 19th. A decline was then observed until May 7th where concentrations were under 0.2 µM. Between May 7th and June 7th, concentrations were stable and rose slightly to a maximum of over 0.3 µM on June 7th. On June 14th, concentrations reached their lowest point at less than 0.1 µM.

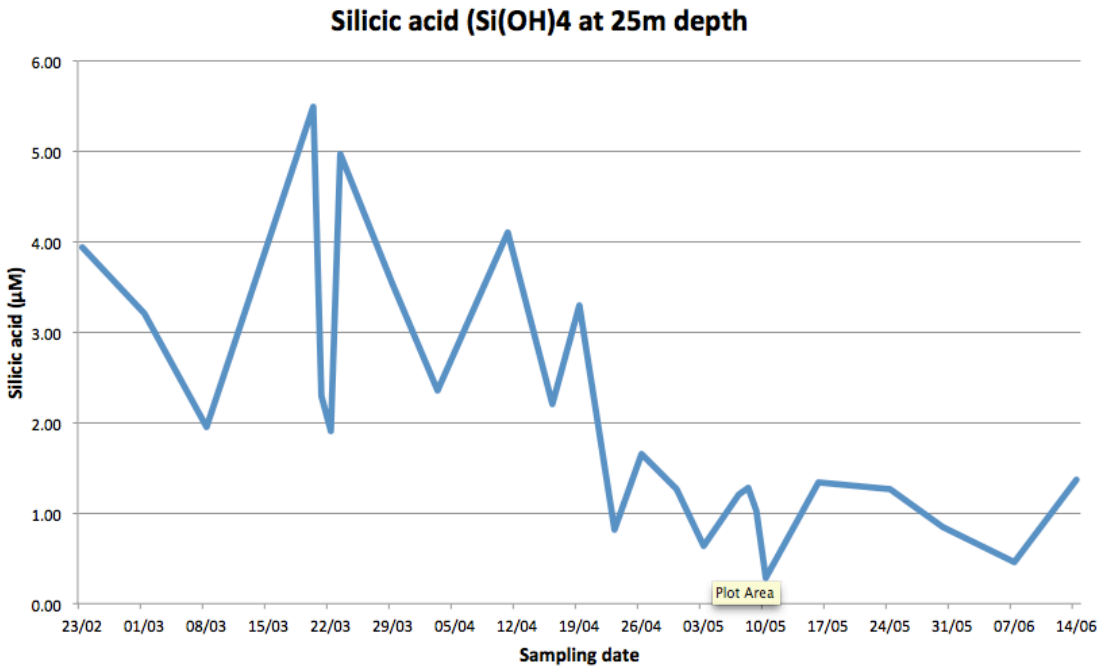


Figure 11 - Silicic acid concentrations in spring 2012 at a depth of 25 m at the Adventfjorden sampling site ISA

Silicic acid concentrations were 4 µM on February 23rd. Large fluctuations between sampling dates were seen throughout the first half of the sampling campaign until April 12th. During this period concentrations ranged from lows of around 2 µM to highs of around 5.5 µM. Despite the fluctuation, a general downward trend was observed between March 20th and May 10th. On May 10th, the lowest concentration of the sampling season was recorded at 0.28 µM. On May 17th, concentrations recovered to over 1 µM. Thereafter and until the end of the sampling on June 21st, concentrations remained relatively stably low with some fluctuation around 1 µM.

4. Chlorophyll *a*

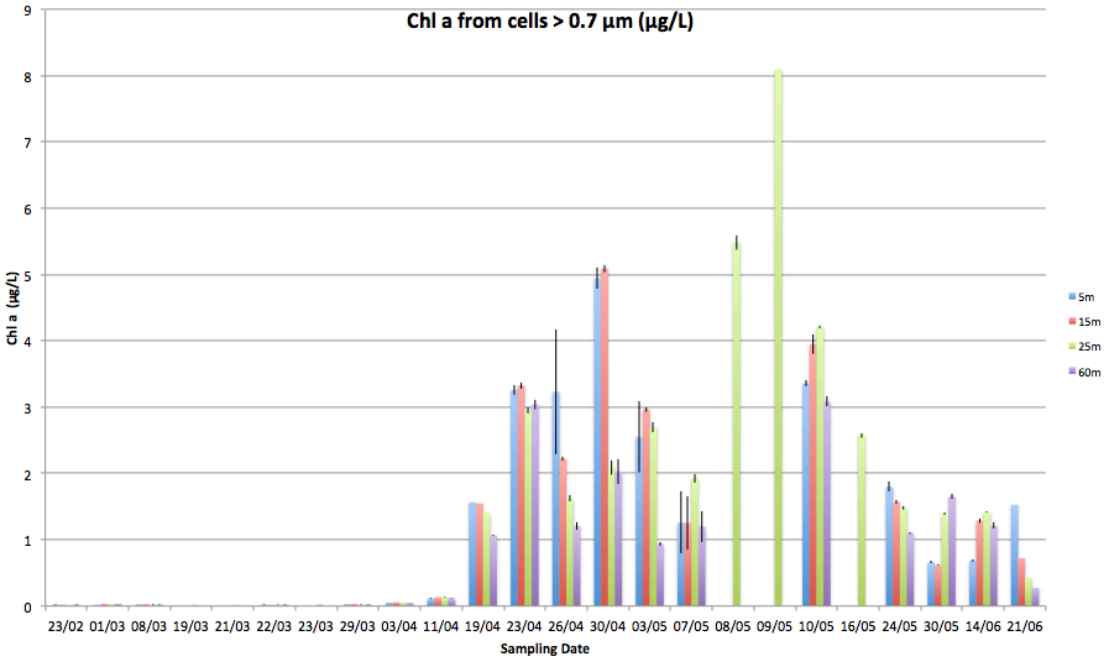


Figure 12 - Total chlorophyll *a* concentrations from the ISA station plotted at the 4 sampling depths against time. Error bars represent SEM from measurements of 3 separate filters.

At all four study depths, total Chl *a* concentrations were very low at around 0.01 µg/L from the start of the study period until April 11th. On April 19th concentrations increased to 1 - 1.5 µg/L throughout the water column to at least 60 m depth. A maximum of approximately 5 µg/L in the surface layers at 5 m and 15 m was observed on April 30th. At this time, the concentrations at 25 m and 60 m depths remained around 2 µg/L. A decline was observed in the two shallowest sampling depths through the first week of May, with concentrations returning to just above 1 µg/L on May 7th. The 25 m depth concentration stayed in a range of just under 2 to just under 3 µg/L at this time, with 60 m depth concentrations remaining even lower at 1 – 2 µg/L. On May 8th and May 9th sampling was only carried out at 25 m depth so there are no data points for 5 m, 15 m, and 60 m depths. Concentrations at 25 m increased markedly between May 7th and May 9th culminating in the maximum concentration observed during the season (8 µg/L) on May 9th. Fluoresence data which was recorded by the CTD (data not shown) indicated a large increase between May 7th and May 8th values at all study depths which remained at similar levels between May 8th and May 9th. The trend recorded by the CTD pointed toward an increase after May 7th which may have brought the 5 m, 15 m, and 60 m Chl *a* into the region of those determined from

the 25 m filters. Values of around 3 – 4 µg/L were recorded across all sampling depths on May 10th. Again on May 16th, samples were only taken from 25 m and the concentration was approximately 2.5 µg/L. This downward trend was continued to the next sampling date on May 24th, where concentrations between 1 and 2 µg/L were recorded at all sampling depths. Concentrations then remained low at under 2 µg/L throughout the rest of the sampling season at all study depths until June 21st. However some differences between study depths were still observed. On May 30th concentrations at 25 m and 60 m depths were over twice those recorded at 5 m and 15 m depths. On June 21st, concentrations were highest at 5 m depth and decreased with depth.

	23/02	01/03	08/03	19/03	21/03	22/03	23/03	29/03	03/04	11/04	19/04	23/04
5m	21	45	40			66		42	50	51	72	71
15m	33	32	40			67		47	43	45	69	63
25m	31	40	46	48	43	38	39	59	45	46	66	70
60m	44	45	30			38		62	50	47	36	76

	26/04	30/04	03/05	07/05	08/05	09/05	10/05	16/05	24/05	30/05	14/06	21/06
5m	57	85	97	140			76		47	23	7	8
15m		85	69	113			82		52	30	18	13
25m		67	77	65	47	67	82	79	61	48	23	11
60m		68	75	98			82		61	60	16	18

Figure 13 - Chl *a* detected in cells >10 µm expressed as a percentage of Chl *a* detected in cells >0.7 µm in spring 2012 at the Adventfjorden sampling site ISA. Darker shading indicates higher percentages. A value of 50 indicates equal Chl *a* biomass in cells >10 µm and <10 µm.

From February 23rd until March 8th, roughly 60 – 80% of the total Chl *a* biomass was contributed by cells <10 µm (small cells), at all four study depths. This trend largely continued until March 23rd, although on March 22nd cells >10 µm (large cells) were of higher biomass (67% total biomass) than small cells in the two shallower sampling depths, 5 m and 15 m. Small cells were still more dominant at 25 m and 60 m depths. Between March 29th and April 11th, large cells approached and sometimes exceeded an equal contribution to Chl *a* biomass as that of the small cells. On both April 19th and April 23rd, large cells generally became more dominant throughout the water column to 60 m and contributed around 70% to the total Chl *a* biomass. From April 30th until May 7th the large

cells became totally dominant in the surface waters at 5 m and 15 m depths making up from 69 % to 100 % of the total Chl *a* biomass. Note that on May 7th, at 5 m and 15 m depths, values exceeded 100 % for the contribution of larger cells. This is likely due to variation in density of organisms in the sampled water. Separate subsamples were taken from sampled water to measure each size fraction. The contribution from larger cells at 25 m and 60 m depths at these times is less than at the surface but is still around 70 % of the total. On May 8th and May 9th, where only the 25 m depth was sampled, larger cells made up 47 % and 67 % of the totals respectively. On May 10th, larger cells made up around 80 % of the total Chl *a* biomass throughout the water column to 60 m depth. This was also seen on May 16th at 25 m. Thereafter, the contribution of large cells decreased continually with progressing date until the end of the study period. By June 14th, small cells dominated once again, accounting for 77 – 93 % of total Chl *a* biomass. This dominance continued into the last sampling date, June 21st.

5. Genetic experiments

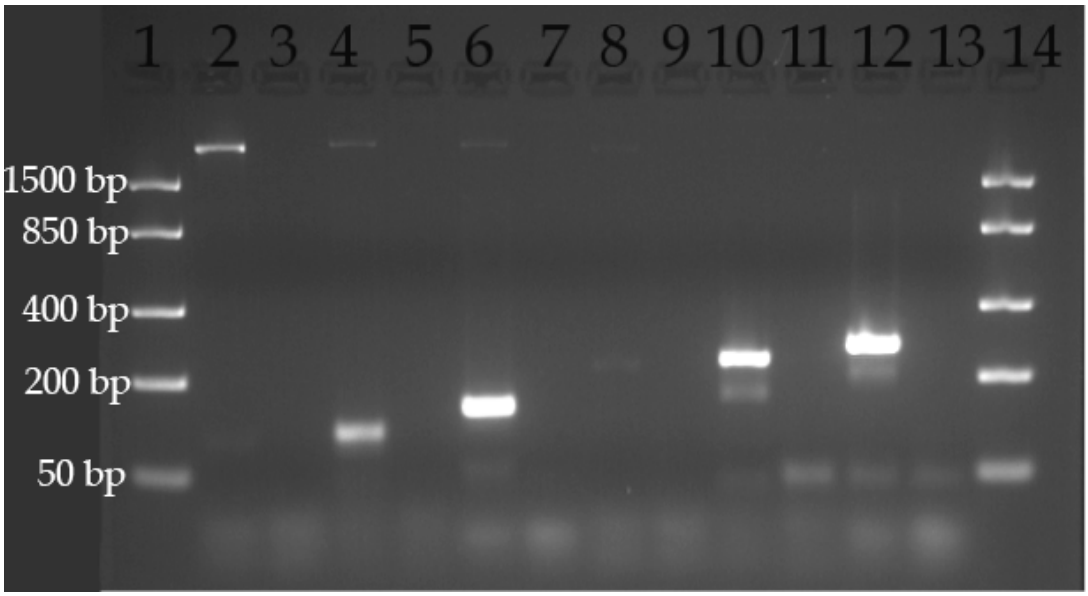


Figure 14 - Digital image of agarose gel loaded with MALV II PCR products after electrophoresis. Lanes filled with PCR product or ladder as follows. Note '+' means positive control (19/04/2012 25 m depth ISA sample) and '-' means negative control (milliQ water). 1 – FastRuler low range DNA ladder (Thermo Scientific); 2 – STF2 TAREUKRev3 + (expected product 93 bp); 3 – STF2 TAREUKRev3 -; 4 – STF2 STR1 + (expected product 101 bp); 5 – STF2 STR1 -; 6 – STF2 STR2 + (expected product 140 bp); 7 – STF2 STR2 -; 8 – ALV01 TAREUKRev3 + (expected product 221 bp); 9 – ALV01 TAREUKRev3 -; 10 – ALV01 STR1 + (expected product 229 bp); 11 – ALV01 STR1 -; 12 – ALV01 STR2 + (expected product 268 bp); 13 – ALV01 STR2 –; 14 – FastRuler low range DNA ladder (Thermo Scientific).

DNA bands were seen at the expected size for all primer combinations tested. The combination of primers STF2 and TAREUKRev3 (lane 2) and of primers ALV01 TAREUKRev3 (lane 8) gave very weak amplification products. The STR2 primer (lanes 6 and 12) resulted in stronger amplification products than the STR1 primer (lanes 4 and 10) when used in combination with both forward primers, STF2 and ALV01. The primer combinations which gave the strongest amplification products were STF2 with STR2 (lane 6) and ALV01 with STR2 (lane 12). The STF3 forward primer did not produce a DNA band in combination with any of the three reverse primers (data not shown).

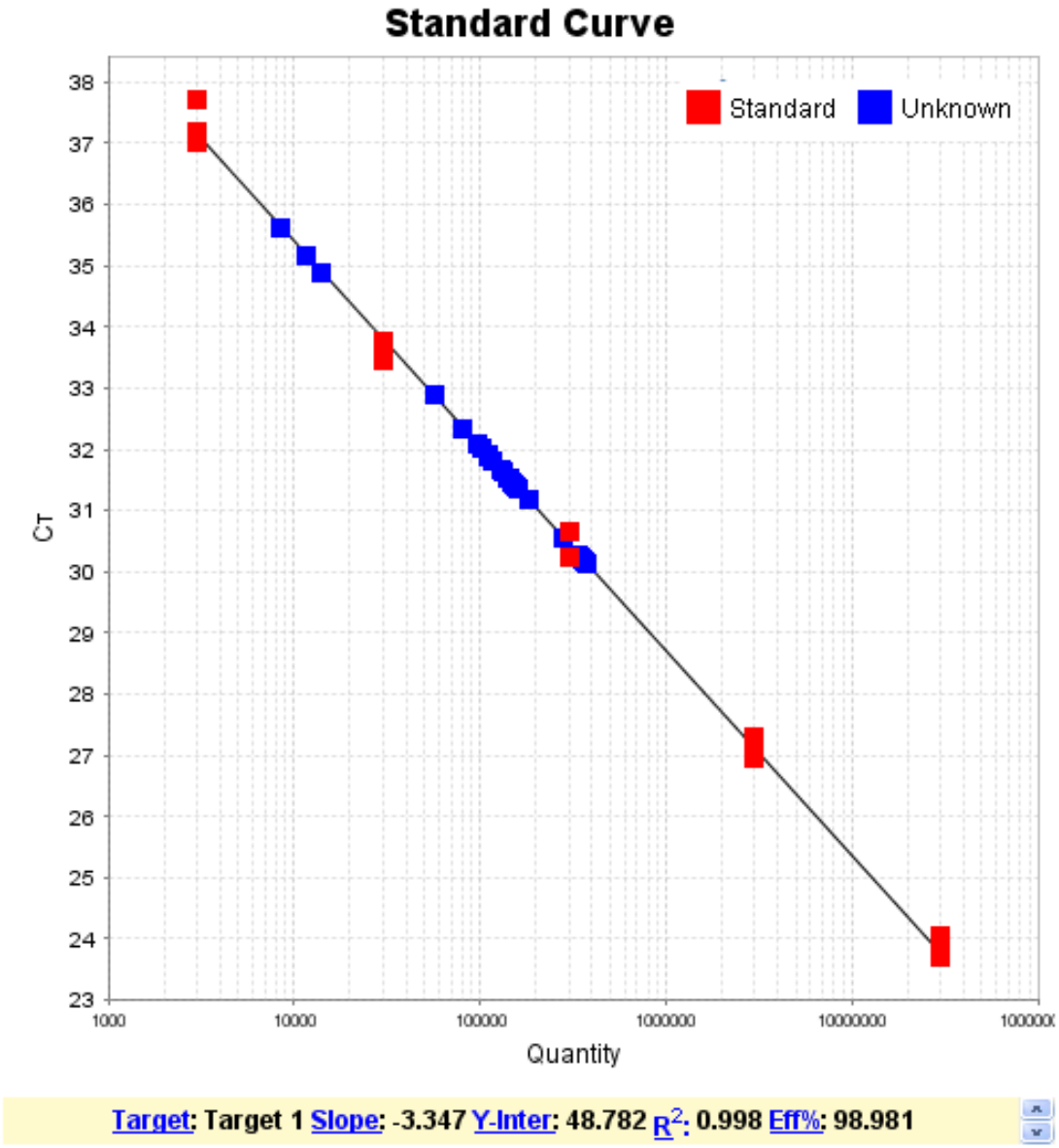


Figure 15 – MALV II plasmid standard curve (red squares in triplicate for each point on the curve) generated from a TaqMan experiment (experiment ID – 130712-2) using Applied Biosystems StepOne software. The blue squares represent samples, which must fit inside the range of the standard curve in order to extrapolate copy numbers.

Figure 15 shows an example of an ideal standard curve generated from a TaqMan experiment (a reaction efficiency of 98.98 % and a correlation coefficient (R^2) of 0.998 can be seen at the bottom right of the figure).
A total of 51 TaqMan PCR experiments were carried out on samples in optimised conditions as described in the TaqMan section of the Materials & Methods chapter. Of these, data from 25 experiments (including the one shown in Fig. 15) fitted the acception criteria (90 – 105% efficiency and correlation coefficient >0.99). 92% of the accepted experiments achieved a correlation coefficient of

0.995 or higher and 68% scored efficiency over 95%. Only 2 experiments exceeded 100% efficiency with values of 100.2% and 104.1%. Every sample from the Adventfjorden field campaign was represented in at least three individual experiments which fitted the acceptance criteria, and data from these experimental replicates were averaged and are represented in the graph below.

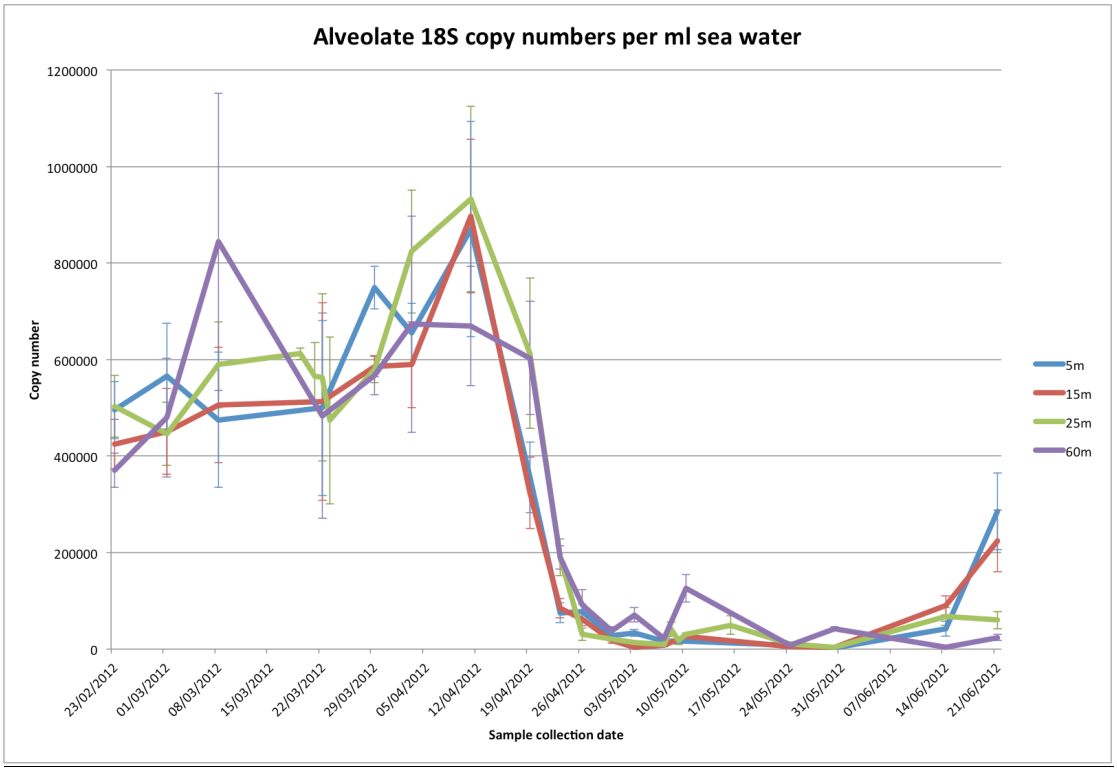


Figure 16 – MALV II 18S rDNA copy numbers from the Adventfjorden sampling site ISA during spring 2012. Data are plotted from TaqMan experiments with efficiency between 90 and 105% and with a correlation coefficient higher than 0.99. Error bars are SEMs (n=3-4).

At the beginning of the sampling season, on February 23rd, MALV II 18S rDNA copy numbers were around 400,000 – 500, 000 per ml of sea water at all four study depths. On March 8th, the copy number at 60 m depth spiked to over 800,000 although this was associated with a high standard error. Aside from this spike, copy numbers were relatively stable between 400,000 and 600,000 for the first month of sampling until March 22nd. After this, an upward trend was observed at all four sampling depths, culminating in maximum values of around 900,000 for 5 m, 15 m, and 25 m, the highest of the season, on April 12th. The increase between April 5th and April 12th seen at these depths was not observed at 60 m depth. Between April 12th and April 26th, copy numbers plummeted at all

4 sampling depths to around 100,000 at 5m, 15 m and 60 m, and under 40,000 at 25 m. Copy numbers at all sampling depths remained in a similar range and generally less than 100,000 copies per ml throughout May and for the first two weeks of June. A brief recovery was observed on May 10th at 60 m depth where copy number increased to over 150,000 but this number decreased again at the next sampling date. On the last sampling date, June 21st a large increase in copy number was recorded at the two shallowest sampling depths with values climbing to over 200,000. This was recorded at both 5 m and 15 m but not at 25 m and 60 m.

7. Statistics

The best fitted linear mixed model included sample as a random effect and the fixed effects of date, Chl *a* and temperature, as well as the interactions between temperature and date, salinity and Chl *a*, salinity and date, and temperature, salinity and date. All terms were found to be significant by Markov Chain Monte Carlo (MCMC) estimates (Table 2). A table with the other models which were tested can be found in the supplementary data, Statistics chapter.

Table 2. MCMC estimated p-values for the fixed effects of the best-fitted model to data generated from the Adventfjorden sampling site ISA during spring 2012.

	Estimate	Std.Error	t.value	Pr(>|t|)
(Intercept)	17.586742	0.391589	44.91	0
date	-2.43532	0.345307	-7.05	0
Chl_GFF	37.301099	17.726688	2.1	0.0367
temp	-1.473349	0.330486	-4.46	0
date:temp	0.368513	0.066631	5.53	0
Chl_GFF:sal	-1.080797	0.512188	-2.11	0.0362
date:sal	0.068978	0.009993	6.9	0
date:temp:sal	-0.010171	0.001918	-5.3	0

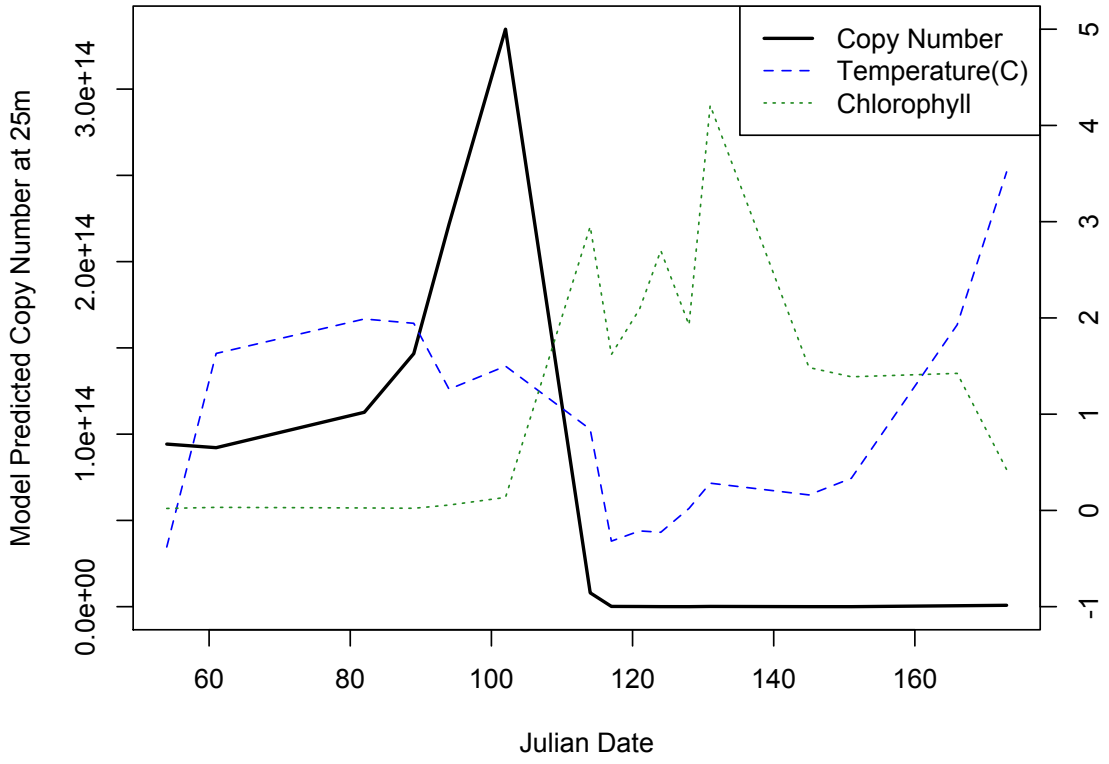


Figure 17. Model predicted MALV II copy number at 25 m depth from the Adventfjorden sampling site ISA during spring 2012 as a function of julian date, co-plotted with temperature and chlorophyll measurements.

The trend in MALV II copy numbers predicted by the best fitted linear mixed model identifies a strong trend related to date that is independent of the depth of sampling (see Fig. 17 for predicted values at 25m). The main trend identified in the model predictions includes a gradual increase for the first two months of sampling, followed by a crash at the end of April, and the absence of a recovery in copy numbers for the rest of the sampling season.

It is important to note that there are no true replicates among water samples collected during the sampling campaign. A single sample from each date at each depth was collected. This means that the date term in the model cannot be strictly interpreted as date as variation between samples collected on the same date is not addressed. Pseudo-replicates are included to overcome this in the statistical analysis (TaqMan experiment replicates n=3-4). Therefore, whilst the

model suggests the tested interactions are significant, the interactions cannot be statistically conclusively interpreted this way.

Discussion

1. Physical characteristics

In combination, the hydrographic data revealed a fairly typical Arctic fjord winter scenario in December 2011 and January 2012, with a well-mixed and cold water column. At the end of January however, an Atlantic water influx at the bottom of the fjord began, the influence of which persisted until around the middle of April as evidenced by the higher temperatures and salinities throughout the water column until this time (Figs 5 and 6). The rapid temperature and salinity decline recorded around April 19th is likely explained by a shift in the front between Arctic and Atlantic water present in the fjord system at this time. Before the shift, the sampling location was likely on the Atlantic water dominated side of the front, whereas afterward it was on the Arctic water dominated side. South westerly and north westerly winds recorded continuously over 16th – 17th April may have contributed in moving the front eastwards into the mouth of Adventfjorden around this time (weather data from met.no recorded at Svalbard airport, just a few km from sampling site - not shown). Predominating wind conditions throughout most of April were south easterly. In the second half of June, surface warming and river input from Adventelva and Longyearelva resulted in increased temperatures and decreased salinities at the surface. The increase in temperature warmed the whole water column whilst the low salinities only impacted the upper 20 metres. A pycnocline was also seen at around 12 m as density changed rapidly over a short depth range.

These physical changes which occurred throughout the sampling period, can be expected to have a large impact on the phytoplankton communities living in Adventfjorden. There were no signs of stratification in the water column as Chl *a* concentrations increased on April 19th, although density did increase with depth. Stratification is generally considered an important factor in the initiation of spring blooms (Loeng 1991; Strass and Nöthig 1996), although blooms have been observed to develop in Svalbard fjords (Hodal et al. 2011) and northern

Norwegian fjords (Hegseth et al. 1995) with very weak or no stratification of the water column. Arctic diatom growth rates have been shown to increase with increasing irradiance up to 50 – 70 $\mu molm^{-2}s^{-1}$ and are quite independent of irradiance above these levels (Gilstad & Sakshaug 1990). Irradiances within this 50 - 70 $\mu molm^{-2}s^{-1}$ range were recorded at 5 m depth from the very first measurements on March 19th, although at 15 m depth, irradiances did not reach these levels until the increase in Chl *a* concentrations was observed on April 19th. This suggests that penetration of light into the upper water layers may be a factor in overall diatom growth levels. However and more significantly, longer day length has been shown to increase diatom growth rates far in excess of the growth rates reached when irradiance becomes a limiting factor and has been implicated as a likely factor in the regulation of the onset of the phytoplankton bloom (Gilstad & Sakshaug 1990; Eilertsen et al. 1990). Therefore it is more likely that the increasing day lengths for the first month of the study contributed more to increasing growth rates in diatom species than maximal irradiances. The reduction in irradiance at 5 m depth during the second half of April may be explained by local weather conditions. Cloud cover was generally high during this time with the last four sampling dates in April recording average cloud cover of at least 4.7 octas (weather data from met.no - not shown). Alternatively or perhaps in combination, the development of the phytoplankton bloom may have altered of the inherent optical properties (IOP) of the water resulting in increased absorption and scattering of light, and thus reducing the irradiance at depth (McKee & Cunningham 2006). On April 30th, the peak observed in Chl *a* biomass in the surface layers at 5 m and 15 m depths may be explained by the irradiance in the water column. Hydrographic conditions were very similar and there were no large differences in nutrient availability between sampling depths at this time point. However the irradiance had reached a low of 80 $\mu molm^{-2}s^{-1}$ at 5 m depth on April 30th, which would lead to light becoming a limiting factor for diatom growth deeper in the water column. This may have made the difference between the higher Chl *a* biomasses recorded at 15 m and above and the lower Chl *a* biomasses recorded at 25 m and below. The reduction in Chl *a* biomass that occurred during the first week of May may have been due to the large drop in irradiance in combination with reducing nutrient concentrations, especially

silicic acid, acting as limiting factors for growth. The large increase in Chl *a* biomass between May 7th and May 9th could then be explained by the hydrographic conditions, as temperature and salinity both rise around this period. An increased amount of Atlantic water present at the sampling site after May 7th had the potential to bring more blooming cells from Isfjorden. The irradiance at 5 m depth had returned to a much higher level of around 140 $\mu molm^{-2}s^{-1}$ and an upward trend in the irradiance at 15 m was also being recorded over this period, allowing potentially higher growth rates deeper into the water column than in the preceding week. Fluorescence data from the CTDs taken on these sampling dates (not shown) does not suggest that Chl *a* concentrations were only high at 25 m depth and below which may have represented a sinking of photosynthetic cells out of the euphotic zone. Rather it would seem the comparable Chl *a* concentrations at all sampling depths between May 7th and May 9th represents phytoplankton growth in the water column due to the improved light climate and transportation of cells from Isfjorden into the sampling area. The decline in nutrient availability, particularly silicic acid, was probably responsible for the end of the growth phase of the phytoplankton bloom on May 10th and the subsequent decline in Chl *a* concentrations through the rest of May. The sinking of photosynthetic cells (see 30th May Chl *a* values) may also have been a factor. The irradiance also decreased from 140 $\mu molm^{-2}s^{-1}$ on May 10th to just 30 $\mu molm^{-2}s^{-1}$ at 5 m depth on May 24th which may have contributed to Chl *a* biomass decline. The irradiance decrease is explained by weather conditions over the period. From May 7th – 13th, 2.4 octas average daily cloud cover was recorded compared with an average of 5.9 octas between May 14th and May 27th (data from met.no). The low Chl *a* biomasses recorded in the first half of June were likely due to persistently low nutrient concentrations acting as a limiting factor for growth of phytoplankton. The increased Chl *a* values observed in surface waters (5 m and 15 m) on June 21st may be due to freshwater input creating a pycnocline that trapped photosynthetic cells near the surface.

2. Phytoplankton bloom species composition

Parallel investigations in Poland carried out by PhD student Anna Kubiszyn (Kubiszyn in prep.) have studied phytoplankton species abundances using subsamples of seawater collected during this study. These investigations have identified the main species dominating in terms of numbers during the bloom and these are as expected in accordance with the literature (*Thalassiosira* spp*., Chaetoceros* spp*., Phaeocystis pouchetii, Fragilariopsis* spp.).

Thalassiosira spp. were the first cells to peak during the bloom reaching a maximum of nearly 6 x 10^5 cells/L on April 23rd, shortly followed three days later by a peak in *Fragilariopsis* spp. at around 5 x 10^5 cells/L. This is in contrast to many studies which have found *Fragilariopsis* spp. to bloom before *Thalassiosira* spp. in Arctic waters (von Quillfeldt 2005; Hodal et al. 2011). The bloom in *Thalassiosira* however, did last longer than in *Fragilariopsis*, persisting with relatively stable abundances until May 7th. *Chaetoceros* spp. bloomed later in the period, with numbers only starting to increase on April 30th, before reaching peak abundances of 6 x 10^5 cells/L on May 7th. This was in accordance with previous studies in Arctic waters which have found *Chaetoceros* spp. to generally bloom later than *Thalassiosira* and *Fragilariopsis* spp (von Quillfeldt 2005; Hodal et al. 2011). Despite these blooms of diatom cells, *Phaeocystis pouchetii* was the numerically dominant species during the entire bloom period from except for the first three days. *Phaeocystis pouchetii* also peaked later in the bloom on May 7th with counts of 1.7 x 10^6 cells/L.

3. MALV II abundances

MALV II DNA copy numbers were generally similar throughout the water column (ie at all four sampling depths) at the same time points. This is indicative of good mixing of the water column and is evidenced by the vertically uniform temperature and salinity plots at most time points across the study period (Figs 5, 6). The first data point for MALV II DNA copy number is at the February 23rd time point, which occurs after the Atlantic water incursion has begun and the

water column has warmed compared to mid January values (Fig. 5). Unfortunately, at the time of writing, I do not have MALV II abundance values from mid January, when a more typical Arctic fjord scenario was observed. This makes it difficult to say whether MALV II populations in the fjord at the end of February are from Arctic water masses or if they were transported into the fjord with the Atlantic water mass. However, the general upward trend in abundance of MALV II populations from the beginning of the study until April 11th does coincide with an increasing contribution of Atlantic water at the sampling site, suggesting that these populations are carried in on the Atlantic current. The decline in MALV II 18S rDNA copy number from April 19th co-occurred with the rapid decrease in temperature at the sampling site. The water temperature decline was not compatible with cooling by atmospheric processes, where surface water is first cooled by heat exchange with the atmosphere, and subsequently convective mixing occurs to reduce the temperature through the water column (Bjerknes 1964). During this study, a simultaneous and rapid cooling of the whole water column was observed, in conjunction with falling salinity values, which is indicative of a change in the characteristics of the water mass present at the sampling site, i.e. a change from an Atlantic water dominated site to an Arctic water dominated site. It is therefore likely that a front between Atlantic water and Arctic water present in the Isfjorden system, moved over the sampling location and led to a large change in hydrographic data collected from the station in the second half of April. As the Arctic water moved into the sampling site, the Atlantic water moved away and carried with it the MALV II population. The influence of Atlantic water then remained diminished for the rest of the study period, and MALV II rDNA copy numbers also remained stable and low. Water samples have been taken in weeks prior to February 23rd and determining MALV II 18S rDNA copy numbers from these samples should be able to confirm that MALV II populations arrived as a result of the Atlantic water incursion. If this is the case, my results imply that Arctic water, at this time of year, contains lower abundances of MALV II organisms than Atlantic water. However, there are several studies which have identified MALV II in Arctic waters previously and their contribution in clone libraries has been comparable to studies further south. For example, in a study in the Arctic ocean close to the

north pole in April, around 20% of phylotypes from clone libraries belonged to the MALV II group (Bachy et al. 2011), and in a spring study of two Svalbard fjords (including Adventfjorden), 27% of clones from libraries belonged to MALV II (Sørensen et al. 2011). Diverse MALV II DNA sequences have previously been discovered from both the Canadian and the Norwegian Arctic from the mixed upper water layers during summer (Lovejoy et al. 2006), as well as from the deep in the coastal Canadian Arctic throughout the year with peak abundances during winter (Terrado et al. 2009). Therefore it does not necessarily hold that MALV II abundances are generally lower in the Arctic than in the Atlantic. One possible scenario is that the high MALV II populations detected in this study may have more recently come into contact with a host species than the local Svalbard population, and thus we may be detecting large quantities of DNA due to dinospore production further south, which have subsequently been carried northward by the ocean currents. It would be interesting to compare the MALV II DNA abundances in this study to abundances detectable using RNA as a starting material. It is well documented that community compositions vary between DNA and RNA based approaches, owing largely to RNA based methods selecting more in favour of the active community (Koid et al. 2012; Not et al. 2009). MALV II relative abundances are routinely much lower in RNA based community compositions, contributing a mere fraction to the total abundances compared with DNA based community compositions in several studies (Terrado et al. 2011; Not et al. 2009; Koid et al. 2012). This could be indicative that whilst MALV II organisms produce large quantities of dinospores that are detected in DNA based studies, many of these do not go on to infect new hosts as evidenced by much lower abundances in RNA based studies. It would be extremely interesting to compare whether the same seasonal trend observed in this study would be replicated with an RNA based approach. For instance, if the high MALV II populations between February and April were mainly inactive dinospores carried on Atlantic currents, an RNA based study may not show a crash in April as seen here. Water samples from 25 m depth were collected during the sampling period and RNA was extracted for use in other projects at UNIS, so there may be possibility for this study.

In addition to the temperature and salinity data suggesting a change in the water mass present at the sampling site in April, an inverse correlation can be seen between MALV II 18S rDNA copy numbers and Chl *a* concentrations during this period. The linear mixed model also showed Chl *a* concentrations to be a significant variable in explaining the pattern of MALV II abundances observed during the study. Water masses originating in different geographical areas contain different phytoplankton species compositions and abundances from one another, for example when Atlantic water intrudes into the Arctic, the non-Arctic species *Emiliania huxleyi* can be found dominating in blooms (Hegseth & Sundfjord 2008). It is possible that as Atlantic water, and with it the MALV II organisms, was flushed from the fjord in April, the Arctic water replacing it contained a different phytoplankton composition which was in a more advanced stage of bloom development. This is supported by the bloom species abundance data discussed in the previous section as all the dominating species identified are typically found in Arctic waters. If this is the case, then the statistical significance of the Chl *a* data is explained by the movement of the water masses rather than a direct and causal relationship between Chl *a* concentrations and MALV II abundances.

4. Potential hosts for MALV II

The phytoplankton abundance data generated by Anna Kubiszyn did not reveal any patterns of phytoplankton species abundances during the study period that may serve as an explanation for the MALV II population changes seen in this study. *Gymnodinium sanguineum,* has been shown to be parasitized by the MALV II species *Amoebophyra ceratii* at lower latitudes (Coats, 1994; Goss & Coats, 1999). Dinoflagellates of the genus *Gymnodinium* were present at concentrations of around 1 x 10^3 cells/ml at the end of February and beginning of March. At the height of the MALV II population on April 11th, they were present in the same concentration. If MALV II are able to parasitize these dinoflagellates, they may have acted to keep their populations low in the beginning of the study period, as during the MALV II population crash, their numbers increase by nearly 2 orders of magnitude to a maximum concentration of 9 x 10^4 cells/ml on April 30th.

However this change in abundance is more likely explained in terms of differences in *Gymnodinium* abundances between the Atlantic and Arctic water masses combined with the fact that cell numbers typically increase around this time in accordance with the onset of the spring bloom (Tiselius & Kuylenstierna 1996).

To my knowledge, there are no papers describing MALV II organisms utilizing diatoms as a host. This could be due to the silica cell wall (frustule) possessed by diatoms, acting as a physical barrier to entry. It has been shown for the MALV II *Amoebophyra* sp. by transmission electron microscopy (TEM), that infection of an athecate dinoflagellate host begins with attachment to the cell membrane (Miller et al. 2012). This method of cell entry may be more difficult to achieve in diatoms with a silica frustule. This could serve as a potential explanation for why MALV II abundances do not increase as the diatoms of *Thalassiosira* and *Fragilariopsis* are in bloom.

The number of dinospores released by Arctic MALV II species is unknown. In the context of hundreds of spores being released from a single host, the decline in DNA abundance observed in this study would not seem to be particularly biologically significant. This could be suggestive that a lower number of dinospores are released by Arctic MALV II species compared to *Amoebophyra* spp. (60 – 400 dinospores estimated by Chambouvet et al. 2008), or perhaps more likely, that the MALV II populations identified in this study were not producing dinospores at all. This could be due to several reasons such as the lack of a present host population or environmental factors associated with being transported northward from the Atlantic to the Arctic. In the period before the bloom, MALV II abundances increased but only by approximately two-fold, (more likely explained by the increasing contribution of Atlantic water to the sampling site over this period than by parasite-host interactions). No decline in any phytoplankton or zooplankton species was observed. Therefore, at least among species counted by Anna (phytoplankton) and Eike (zooplankton), there appears not to be an obvious relationship in abundances between MALV II species and potential hosts.

However, if MALV II organisms were successfully parasitizing a host or hosts during this study, a host may be discovered within the nanoflagellates, which were neglected here. There is also the possibility that the host is not found amongst the planktonic community at all, and is free from the influence of being transported by ocean currents, for example one or several crab species, many of which are already known to serve as hosts (Stentiford & Shields 2005). If the MALV II populations quantified are indeed of Atlantic origin, then a physical separation of host and parasite would explain why we appear not to observe parasitism at these latitudes. The possibility exists that the MALV II rDNA detected in this study does not come from living organisms and is merely present in the environment. *Amoebophrya* sp. dinospores have been shown to remain infective generally for between 2 – 5 days (Coats et al. 2002). However after this infectious period has elapsed, it is unknown how long DNA from these dinospores can remain in the environment before being broken down.

Single cell PCR techniques have been used to successfully amplify MALV II DNA from within a host cell (Bråte et al. 2012). Single cell PCR has also been successfully demonstrated on cells preserved in Lugol (Auinger et al. 2008). Therefore the future possibility exists to identify a host species experimentally utilizing Lugol preserved phytoplankton samples.

5. Experimental considerations

Sampling

The number of organisms in an environment can vary in space and time. This study addressed temporal variation in organism numbers in an Arctic fjord environment. Spatial controls should have been included in the sampling program to account for any spatial effects on populations (Underwood 1991). This may have taken the form of replicates taken at the same sampling position or numerous samples taken from a wider area in the course of a single sampling day. We have assumed spatial variation to be non significant due to water mixing in the fjord, an assumption which is bolstered by vertical hydrographic profiles

(figs 5 and 6) and analyses demonstrating MALV II abundances are not statistically different between depths at the same time points (the linear mixed model was not improved with the addition of sampling depth – see supplementary data, Statistics chapter).

DNA extraction

After water filtration, filters were cut in half and stored at -80 °C before DNA extraction. This was designed to prevent total loss of the sample in the event of a problem with DNA extraction such as protocol failure or contamination. After successful DNA extraction of both filter halves, DNA was combined. However, several filter halves had suffered from contamination with unknown eukaryotic DNA during extraction as evidenced from negative controls run on agarose gels (data not shown). Therefore for some samples, only half the filter was extracted successfully and used in subsequent quantitative PCR. This has the potential to skew results for the following two reasons; filters are cut by scalpel blade manually and may not be cut precisely into halves; distribution of cells across the surface of the filter may not be equal.

It is worth noting that eukaryotic contamination may not have had a detrimental effect on the TaqMan PCRs performed in this study due to their specificity. However, sample DNA is used for a variety of other projects, some of which would be adversely impacted by contamination. Therefore DNA extractions from separate filter halves were not combined in cases with contamination issues.

Genetics

The challenge when picking a gene to amplify one group of species but not another is to choose a gene with the correct level of variability. If the gene is too conserved, amplification of unwanted species may occur. However if variability is too great, it may become impossible to design primers to encompass all species of interest. In this study the 18S gene was used. It was difficult to design an adequate primer set for the study, with severe limitations on where primers could be placed on the 18S V4 region due to variability issues described above. The primers used in this study did prove to be specific. After the conclusion of

lab work for this thesis, Illumina sequencing of PCR products amplified during the study was carried out and returned MALV II sequences in approximately 90% of cases from a total of approximately 3.4 million sequences (data not shown). Therefore the 18S V4 region can be used effectively for this type of study in MALV II.

The number of copies of the 18S gene differs between species and is unknown in MALV II organisms. In addition, the copy number almost certainly varies even among organisms within MALV II due to the high genetic diversity described (Guillou et al. 2008). Therefore a cell number estimate cannot be calculated from the gene copy number present in the water samples. To further complicate matters, different life stages of Syndiniales species may have different gene copy numbers, which may skew cell number estimates if these estimates are being determined from a single value of gene copy number per cell (Miller et al. 2012). There are issues with gene copy number estimation from an experimental standpoint as well as a biological one. Environmental cT values from TaqMan PCR can only be translated into copy numbers if the sample assay efficiency is equivalent to that of the plasmid standard curve. Inhibitors present in DNA extracted from environmental samples may mean that the assumption of equal efficiencies is incorrect (Schriewer et al. 2011). However in this study a sample derived standard curve was run as a control and resulted in similar efficiency to the plasmid standard curve under the same experimental conditions (data not shown).

TaqMan PCRs

The biggest experimental issue faced in the study was in the generation of the standard curve in quantitative PCRs. After standard dilutions were made, experiments were run immediately and reaction efficiencies of over 90 % were generally observed. However, when the same standard dilutions were used repeatedly, reaction efficiencies dropped in each subsequent experiment. By only the third use, efficiencies that dropped to 60 % were not uncommon. This problem was encountered whether the standards where frozen at -20 °C between experiments, or kept on ice for three hours on the bench between runs.

All reaction components were replaced with fresh components and MilliQ water used to make the dilutions was replaced with molecular biology grade sterile water. This still did not result in an improvement in efficiencies. If standard dilutions were made up fresh from the stock immediately prior to each experiment, consistently high efficiencies (90-100 %) were seen.

Although at first glance, this may appear to be a simple DNA degradation problem, this would seem not to be the case. The cT value for the standard of highest concentration did not show signs of increase in subsequent experiments with the same standard dilutions (ie the standard curve begun from the same point but the slope changed between experiments). The stock from which the standards were made up was made up with milliQ water itself, and from the same source as the standard dilutions. This stock was thawed and refrozen many times during the course of the study. cT values in standard curves did not show a general increase over the course of the study. Both of these observations imply that DNA degradation is not the factor causing the phenomenon. However, perhaps the more dilute the DNA becomes, the more susceptible it is to degradation, although the mechanism by which this may occur is unknown.

TaqMan assays are highly sensitive. Less than ten copies of norovirus genome have been detected in contaminated shellfish samples (Jothikumar et al. 2005). It has also been shown that less than one cell of *Pfisteria piscida,* a dinoflagellate species which can cause harmful algal blooms, can be detected by TaqMan PCR targeting the 18S gene (Bowers et al. 2000). The same study showed that the same detection level was achieved if the target DNA was pure versus if it was extracted in the presence of extraneous cells (i.e. a simulation of environmental DNA). Another study demonstrates that ten plasmid copies of 5.8S rDNA from *Alexandrium minutum*, another HAB forming dinoflagellate species, can be detected by TaqMan PCR (Galluzzi et al. 2004). In this study a sensitivity of 3,000 plasmid copies was observed. This represents a sensitivity at least three orders of magnitude lower than what can generally be found described in the literature. One possible explanation for this lies with the primer design. It was difficult to design primers that did not dimerise with themselves in a region specific only to

MALV II DNA sequences. Therefore, primers formed higher energy dimers than is ideal (see Fig. 4 and supplementary data, PCR primer dimers chapter), and this may act to reduce the overall sensitivity of the experiment.

6. Conclusion

I have shown in this study that the abundance of MALV II DNA copies in the marine environment of Adventfjorden drops considerably at a time coinciding with the onset of the phytoplankton spring bloom. The decline in 18S rDNA copy number from the pre-bloom to the bloom period represents a statistically significant change most likely explained by the change in hydrographical conditions at the sampling site in April, where an Atlantic water dominated sampling site gave way to an Arctic water dominated site. MALV II organisms were likely washed out of the fjord with the Atlantic water, although the study should be expanded with samples taken prior to February 23rd to provide confirmation that MALV II rDNA did arrive on the Atlantic currents. No potential host species were successfully identified from phytoplankton or zooplankton species counts. It is possible that a host species may be found amongst the nanoflagellates or is perhaps a larger organism that is more independent from ocean currents. Whilst knowledge of the MALV II group has expanded significantly, especially in the last decade, it is clear that further studies are required to elucidate the role and significance of MALV II in the Arctic marine ecosystem. These might include further studies of seasonality in the autumn / winter period, RNA based studies looking at the biological activity level within populations, and host identification studies with the possibility of utilizing the single-cell PCR technique.

Acknowledgements

I would like to thank my supervisor Tove for her guidance and assistance throughout my master. I would like to thank my supervisor Else for her contributions to the writing of the thesis and the new angles she provided me to look at the data. Thanks to Ragnheid Skogseth for help in interpreting the hydrographic data. A big thanks to Anna Kubiszyn and Eike Stübner for providing me with their data to help me with the interpretation of my own. Also thank you Marie Davey for the statistical analysis and comments on the thesis. Last but not least, thanks to Miriam for those long joyful hours of sampling, filtering and DNA extracting.

References

Aliani, S. et al. 2004. Multidisciplinary investigations in the marine environment of the inner Kongsfiord, Svalbard islands (September 2000 and 2001). *Chemistry and Ecology*, 20 (Sup1), pp.S19–S28.

Auinger, B.M. et al. 2008. Improved methodology for identification of protists and microalgae from plankton samples preserved in Lugol's iodine solution: combining microscopic analysis with single-cell PCR. *Applied and environmental microbiology*, 74(8), pp.2505–10.

Bachy, C. et al. 2011. Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the north pole at the end of the polar night. *Frontiers in microbiology*, 2, 106.

Bernard, P.S. & Wittwer, C.T. 2002. Real-time PCR technology for cancer diagnostics. *Clinical chemistry*, 48(8), pp.1178–85.

Bjerknes, J. 1964. Advances in Geophysics 10(1) pp.8-13.

Bowers, H.A. et al. 2000. Development of Real-Time PCR assays for rapid detection of *Pfiesteria piscicida* and related dinoflagellates. *Applied and environmental microbiology*, 66(11), pp.4641–4648.

Bråte, J. et al. 2012. Radiolaria associated with large diversity of marine alveolates. *Protist*, 163(5), pp.767–77.

Cachon, J and Cachon, M. 1987. Parasitic dinoflagellates. In *The biology of dinoflagellates*. pp. 571 – 610.

Chambouvet, A. et al. 2008. Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science*, 322(5905), pp.1254–7.

Coats D.W. 1994. Occurrence of the parasitic dinoflagellate *Amoebophrya ceratii* in Chesapeake Bay populations of *Gymnodinium sanguineum*. *Journal of Eukaryotic Microbiology*, 41(6), pp.586–593.

Coats, D.W. 1999. Parasitic life styles of marine dinoflagellates. *The Journal of Eukaryotic Microbiology*, 46(4), pp.402–409.

Coats, D.W. et al. 2002. Parasitism of photosynthetic dinoflagellates by three strains of Amoebophrya (Dinophyta): Parasite survival, infectivity, generation time and host specificity. *Journal of Phycology*, 528, pp.520–528.

Cottier, F.R. et al. 2010. Arctic fjords: a review of the oceanographic environment and dominant physical processes. *Geological Society, London, Special Publications*, 344(1), pp.35–50.

Degerlund, M. & Eilertsen, H.C. 2009. Main species characteristics of phytoplankton spring blooms in NE Atlantic and Arctic waters (68–80° N). *Estuaries and Coasts*, 33(2), pp.242–269.

Dobrzyn, P. et. al. 2005. Sedimentation of chlorophylls in an Arctic fjord under freshwater discharge. *Hydrobiologia*, 532(1-3), pp.1–8.

Eilertsen, H.C. et. al. 1990. Photoperiodic control of diatom spore growth; a theory to explain the onset of phytoplankton blooms. *Marine ecology progress series*, 116(1), 303-307

Erlich, H.A. 1989. Polymerase chain reaction. *Journal of Clinical Immunology*, 9(6), pp.437-447.

Galluzzi, L. et al., 2004. Development of a Real-Time PCR Assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate). *Applied and environmental microbiology*, 70(2), pp.1199–1206.

Gilstad, M. & Sakshaug, E. 1990. Growth rates of ten diatom species from the Barents Sea at different irradiances and day lengths. *Marine Ecology Progress Series*, 64, pp.169–173.

Gomez, F.G. 2012. A quantitative review of the lifestyle , habitat and trophic diversity of dinoflagellates (Dinoflagellata , Alveolata). *Systematics and Biodiversity,* 10(3), pp.267–275.

Goss, S.H. & Coats, D.W. 1999. The Phylogenetic Position of Amoebophrya sp. Infecting Gymnodinium sanguineum. , 46(2), pp.194–197.

Groisillier, A. et al. 2006. Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquatic Microbial Ecology*, 42, pp.277–291.

Guillou, L et al. 2008. Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental microbiology*, 10(12), pp.3349–65.

Harada, A. et. al. 2007. Species of the parasitic genus Duboscquella are members of the enigmatic marine alveolate group I. *Protist*, 158(3), pp.337–47.

Hegseth, E.N. 1992. Sub-ice algal assemblages of the Barents Sea : Species composition , chemical composition , and growth rates. *Polar Biology*, 12(5) pp.485–496.

Hegseth, E.N. & Sundfjord, A. 2008. Intrusion and blooming of Atlantic phytoplankton species in the high Arctic. *Journal of Marine Systems*, 74(1-2), pp.108–119.

Hegseth, E.N. et. al. 1995. Phytoplankton in fjords and coastal waters of northern Norway : environmental conditions and dynamics of the spring bloom. *Ecology of fjords and coastal waters,* (1995) pp.45-72

Hegseth, E.N. & Tverberg, V. 2013. Effect of Atlantic water inflow on timing of the phytoplankton spring bloom in a high Arctic fjord (Kongsfjorden, Svalbard). *Journal of Marine Systems*, 113-114, pp.94–105.

Hegseth, E.N. 1998. Primary production of the northern Barents Sea. , 17(2), pp.113–123.

Hodal, H. et al. 2011. Spring bloom dynamics in Kongsfjorden, Svalbard: nutrients, phytoplankton, protozoans and primary production. *Polar Biology*, 35(2), pp.191–203.

Holland, P.M. et al. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), pp.7276–80.

Hop, H. et al. 2002. The marine ecosystem of Kongsfjorden, Svalbard. *Polar Research*, 21(1), pp.167–208.

Hou, Y. et al. 2010. Serious overestimation in quantitative PCR by circular (supercoiled) plasmid standard: microalgal pcna as the model gene. *PloS one*, 5(3), e9545.

Jothikumar, N. et al. 2005. Rapid and sensitive detection of noroviruses by using TaqMan-based one-step reverse transcription-PCR assays and application to naturally contaminated shellfish samples. *Applied and environmental microbiology*, 71(4), pp.1870–1875.

Koid, A. et al. 2012. Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. *Applied and environmental microbiology*, 78(11), pp.3958–65.

Kroh, E.M. et al. 2010. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods*, 50(4), pp.298–301.

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

Lin, S. 2011. Genomic understanding of dinoflagellates. *Research in microbiology*, 162(6), pp.551–69.

Loeng, H. 1991. Features of the physical oceanographic conditions of the Barents Sea. *Polar Research*, 10(1), pp.5–18.

Lovejoy, C. et. al. 2006. Diversity and distribution of marine microbial eukaryotes in the Arctic ocean and adjacent seas. *Applied and environmental microbiology* 72(5) pp.3085-3095.

Massana, R. 2011. Eukaryotic picoplankton in surface oceans. *Annual review of microbiology*, 65, pp.91–110.

Massana, R. et al. 2011. Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian ocean. *The ISME journal*, 5(2), pp.184–95.

Massana, R. et al. 2002. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Applied and environmental microbiology*, 68(9), pp.4554–4558.

McKee, D. & Cunningham, A. 2006. Identification and characterisation of two optical water types in the Irish Sea from in situ inherent optical properties and seawater constituents. *Estuarine, Coastal and Shelf Science*, 68(1-2), pp.305–316.

Miller, J.J. et.al. 2012. Ultrastructure of *Amoebophrya* sp. and its changes during the course of infection. *Protist*, 163(5), pp.720–45.

Moon-van der Staay, S.Y. et. al. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 409(6820), pp.607–10.

Moreira, D. & López-García, P. 2002. The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in microbiology*, 10(1), pp.31–8.

Morgante, M. & Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. *The Plant journal : for cell and molecular biology*, 3(1), pp.175–82.

Nilsen, F. et al. 2008. Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Svalbard. *Continental Shelf Research*, 28(14), pp.1838–1853.

Noguchi, F. et al. 2013. A novel alveolate in bivalves with chemosynthetic bacteria inhabiting deep-sea methane seeps. *The Journal of eukaryotic microbiology*, 60(2), pp.158–65.

Not, F. et al. 2009. New insights into the diversity of marine picoeukaryotes. *PloS one*, 4(9), e7143.

Rappé, M.S. et al. 1998. Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from

the Pacific and Atlantic coasts of the United States. *Applied and environmental microbiology*, 64(1), pp.294–303.

Saiki, R. et. al. 2011. Enzymatic amplification of beta-globulin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science*, 230(4732), pp.1350-1354.

Sakshaug, E. et. al. 2009. Ecosystem Barents Sea. pp.

Schriewer, A. et. al. 2011. Improving qPCR efficiency in environmental samples by selective removal of humic acids with DAX-8. *Journal of microbiological methods*, 85(1), pp.16–21.

Seuthe, L. et. al. 2010. Microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): II. Ciliates and dinoflagellates. *Polar Biology*, 34(5), pp.751–766.

Siano, R. et al. 2011. Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences*, 8(2), pp.267–278.

Skogseth, R. et. al. 2005. Watermass transformations in Storfjorden. *Continental Shelf Research*, 25(5-6), pp.667–695.

Skovgaard, A. et.al. 2009. Identifying the lethal fish egg parasite *Ichthyodinium chabelardi* as a member of Marine Alveolate Group I. *Environmental microbiology*, 11(8), pp.2030–41.

Skovgaard, A. & Saiz, E. 2006. Seasonal occurrence and role of protistan parasites in coastal marine zooplankton. *Marine Ecology Progress Series* 327(37), pp.37–49.

Small, H.J. et al. 2012. Morphological and molecular characterization of *Hematodinium perezi* (Dinophyceae: Syndiniales), a dinoflagellate parasite of the harbour crab, *Liocarcinus depurator*. *The Journal of eukaryotic microbiology*, 59(1), pp.54–66.

Sørensen, N. et. al. 2011. Molecular diversity and temporal variation of picoeukaryotes in two Arctic fjords, Svalbard. *Polar Biology*, 35(4), pp.519–533.

Stentiford, G.D. & Shields, J.D. 2005. A review of the parasitic dinoflagellates *Hematodinium* species and *Hematodinium*-like infections in marine crustaceans. *Diseases of aquatic organisms*, 66(1), pp.47–70.

Stoeck, T. et al. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular ecology*, 19 (Sup1), pp.21–31.

Strass, V.H. & Nöthig, E. 1996. Seasonal shifts in ice edge phytoplankton blooms in the Barents Sea related to the water column stability. *Polar Biology*, 16(6) pp.409–422.

Svendsen, H. et al. 2002. The physical environment of Kongsfjorden – Krossfjorden, an Arctic fjord system in Svalbard. *Polar Research* 21(1), pp.133–166.

Terrado, R. et. al. 2009. Mesopelagic protists: diversity and succession in a coastal Arctic ecosystem. *Aquatic Microbial Ecology*, 56, pp.25–39.

Terrado, R. et al. 2011. Protist community composition during spring in an Arctic flaw lead polynya. *Polar Biology*, 34(12), pp.1901–1914.

Tiselius, P. & Kuylenstierna, B. 1996. Growth and decline of a diatom spring bloom phytoplankton species composition, formation of marine snow and the role of heterotrophic dinoflagellates. *Journal of Plankton Research*, 18(2), pp.133–155.

Underwood, A.J., 1991. Beyond BACI : Experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research*, 42(5), pp.569–587.

Vestheim, H. & Jarman, S.N. 2008. Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs. *Frontiers in zoology*, 5(1), 12.

von Quillfeldt, C. 2005. Common diatom species in Arctic spring blooms: Their distribution and abundance. *Botanica Marina*, 43(6), pp.499–516.

Wassman, P. & Slagstad, D. 1993. Seasonal and annual dynamics of particulate carbon flux in the Barents Sea. *Polar Biology*, 13(6), pp.363-372.

Wassmann, P. et al. 1999. Spring bloom development in the marginal ice zone and the central Barents sea. *Marine Ecology*, 20(3-4), pp.321–346.

Węsławski, J. 1988. Seasonality in an Arctic fjord ecosystem: Hornsund, Spitsbergen. *Polar Research*, 6(2) pp.185-189

Wiktor, J. 1999. Early spring microplankton development under fast ice covered fjords of Svalbard , Arctic *. *Oceanologia*, 41(1), pp.51–72.

Wiktor, J. et al. 1998. Phytoplankton and suspensions in relation to the freshwater in Arctic coastal marine ecosystems. *Polish Polar Research*, 19(3-4), pp.219–234.

Zajączkowski, M. et al. 2009. Vertical flux of particulate matter in an Arctic fjord: the case of lack of the sea-ice cover in Adventfjorden 2006–2007. *Polar Biology*, 33(2), pp.223–239.

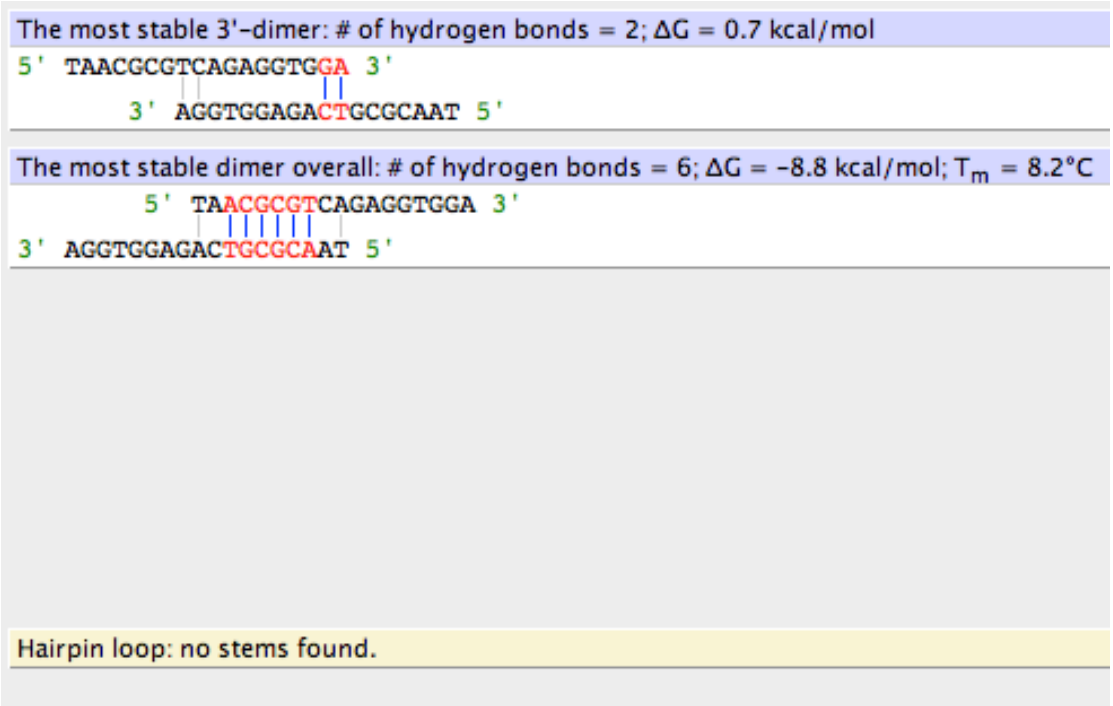
Zajączkowski, M. & Włodarska-Kowalczuk, M. 2007. Dynamic sedimentary environments of an Arctic glacier-fed river estuary (Adventfjorden, Svalbard). I. Flux, deposition, and sediment dynamics. *Estuarine, Coastal and Shelf Science*, 74(1-2), pp.285–296.

Supplementary data

1. PCR primer dimers

The data below shows the PCR primer sequences and highlights where bonds can form between primers leading to primer dimers.

STF2;



ALV01;

The most stable 3'-dimer: # of hydrogen bonds = 4; ΔG = -4.5 kcal/mol

```
5' AGAGTGTTCACGGCAGGC 3'
              || ||
           3' CGGACGGCACTTGTGAGA 5'
```

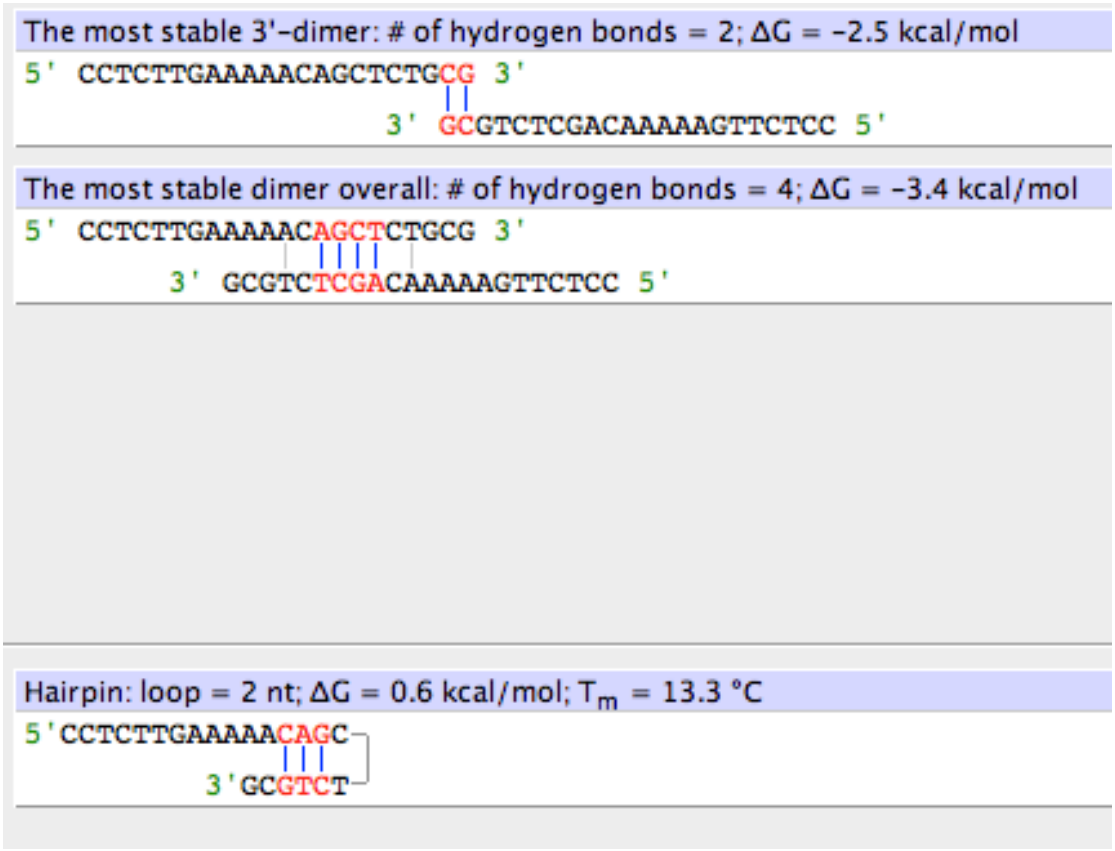
The most stable dimer overall: # of hydrogen bonds = 4; ΔG = -4.5 kcal/mol

```
5' AGAGTGTTCACGGCAGGC 3'
              || ||
           3' CGGACGGCACTTGTGAGA 5'
```

Hairpin: loop = 2 nt; ΔG = 0.3 kcal/mol; T_m = 20.1 °C

```
     5'AGAGTGT
        |||
3'CGGACGGCACT
```

STF3;


The most stable 3'-dimer: # of hydrogen bonds = 2; ΔG = -2.5 kcal/mol
5' CCTCTTGAAAAACAGCTCTGCG 3'
3' GCGTCTCGACAAAAAGTTCTCC 5'
The most stable dimer overall: # of hydrogen bonds = 4; ΔG = -3.4 kcal/mol
5' CCTCTTGAAAAACAGCTCTGCG 3'
3' GCGTCTCGACAAAAAGTTCTCC 5'
Hairpin: loop = 2 nt; ΔG = 0.6 kcal/mol; Tm = 13.3 °C
5'CCTCTTGAAAAACAGC
3'GCGTCT

TAREukREV3;

The most stable 3'-dimer: # of hydrogen bonds = 8; ΔG = -8.6 kcal/mol; T_m = 8.0°C

```
5' ACTTTCGTTCTTGATCAA 3'
           ||||||||
      3' AACTAGTTCTTGCTTTCA 5'
```

The most stable dimer overall: # of hydrogen bonds = 8; ΔG = -8.6 kcal/mol; T_m = 8.0°C

```
5' ACTTTCGTTCTTGATCAA 3'
           ||||||||
      3' AACTAGTTCTTGCTTTCA 5'
```

Hairpin: loop = 2 nt; ΔG = 1.1 kcal/mol; T_m = 6.3 °C

```
5'ACTTTCGTTCTTGA
          |||
     3'AACT
```

STR1;

The most stable 3'-dimer: # of hydrogen bonds = 4; ΔG = -0.6 kcal/mol

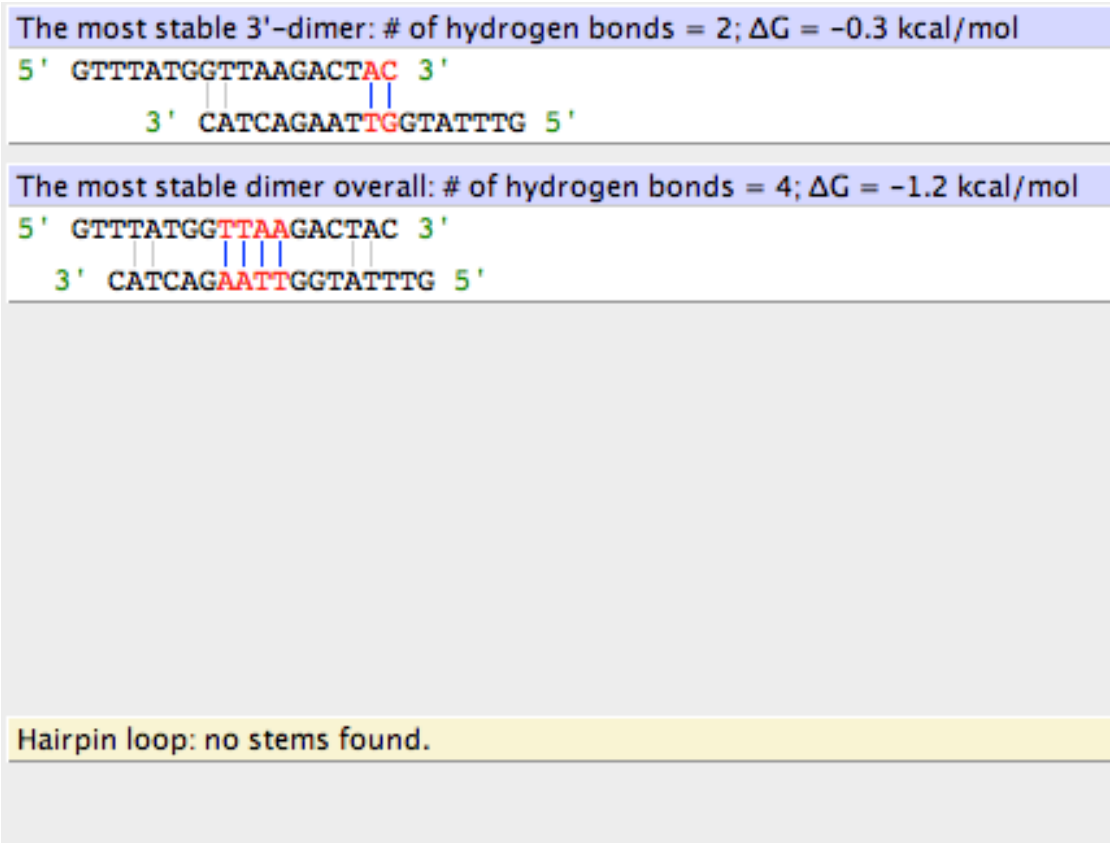
```
5' ATCCCCTAACTTTCGT 3'
          ||    ||
     3' TGCTTTCAATCCCCTA 5'
```

The most stable dimer overall: # of hydrogen bonds = 4; ΔG = -0.6 kcal/mol

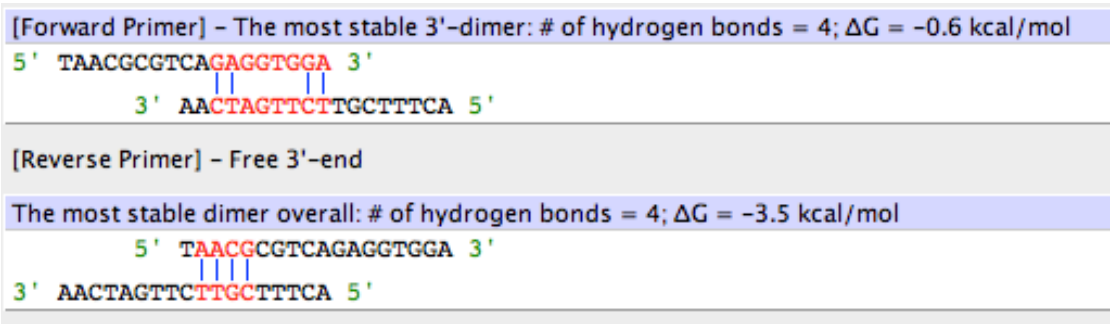
```
5' ATCCCCTAACTTTCGT 3'
          ||    ||
     3' TGCTTTCAATCCCCTA 5'
```

Hairpin loop: no stems found.

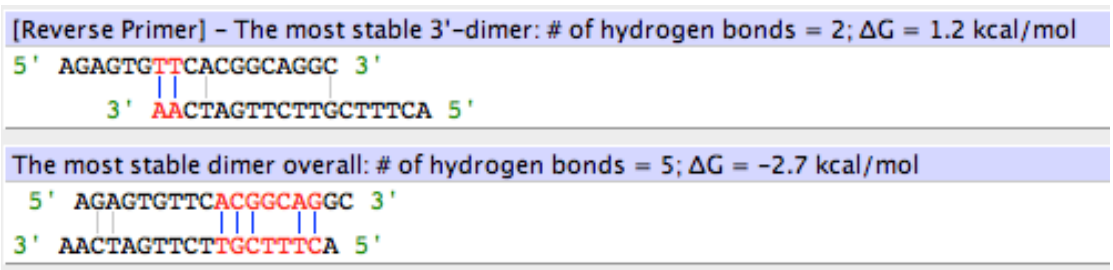
STR2;



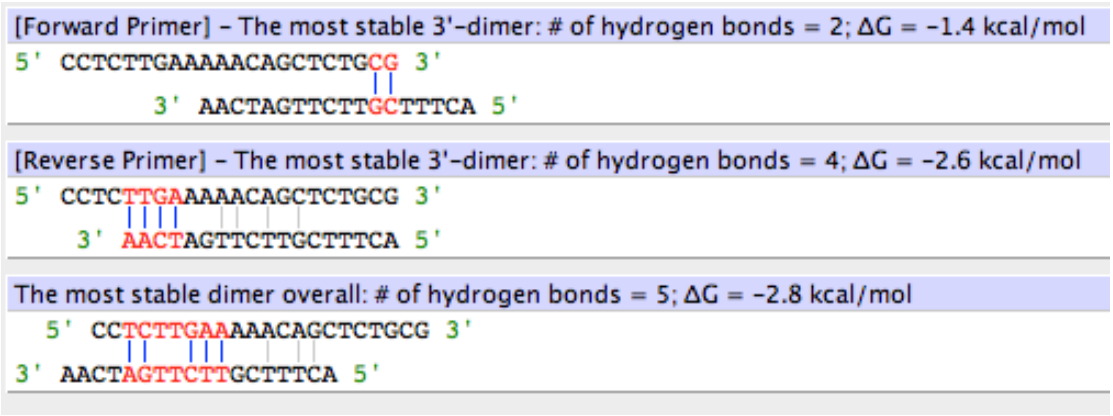
STF2 – TAREukREV3



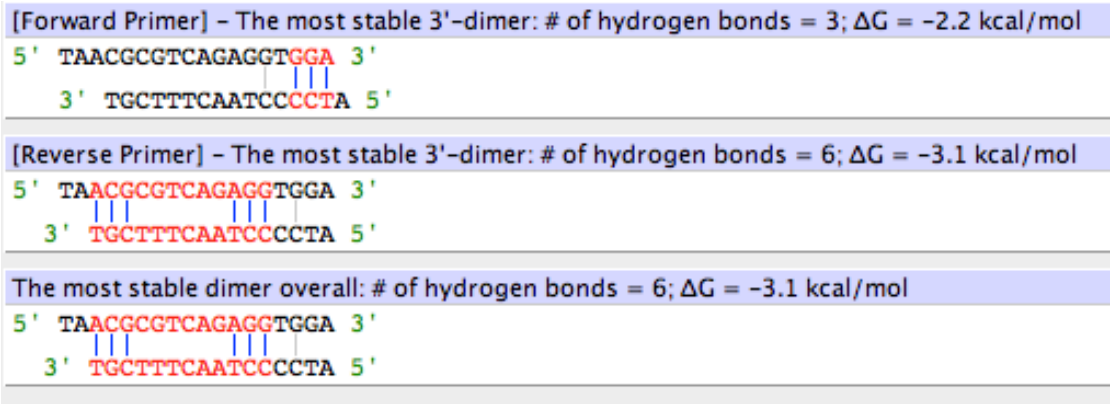
ALV01 – TAREukREV3



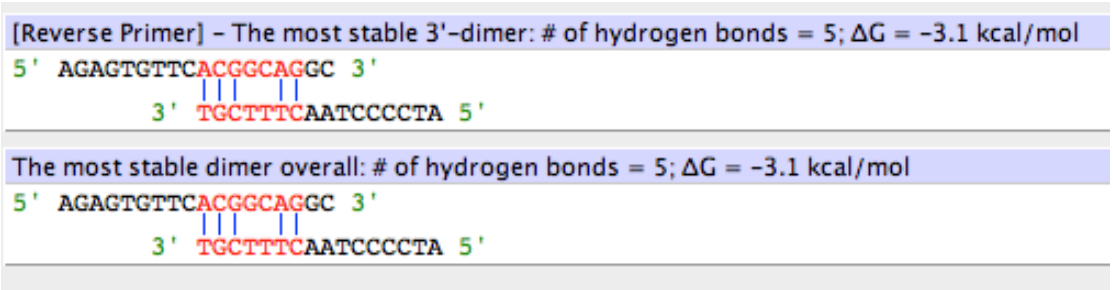
STF3 – TAREukREV3



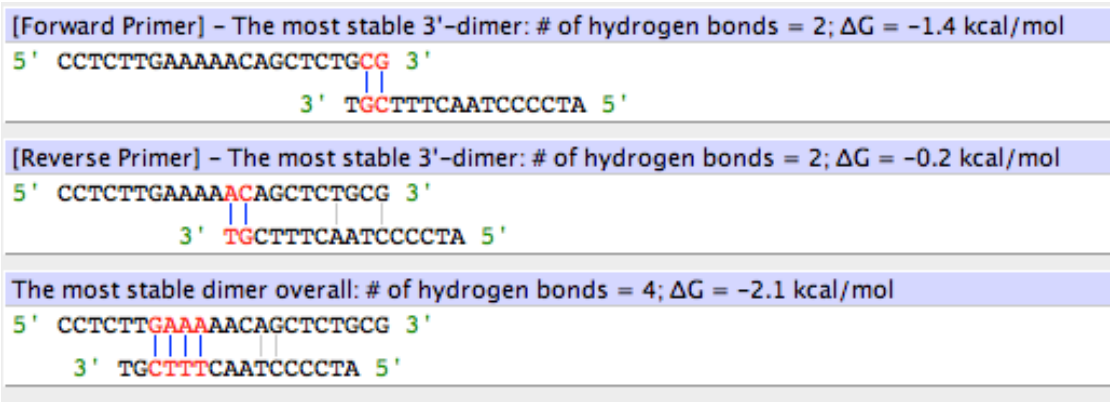
STF2-STR1



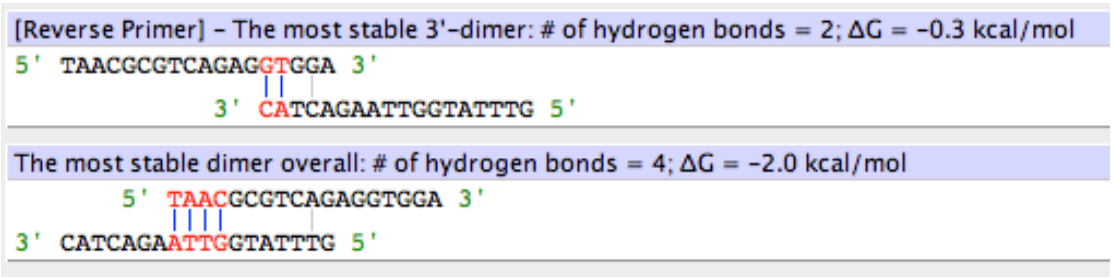
ALV01-STR1



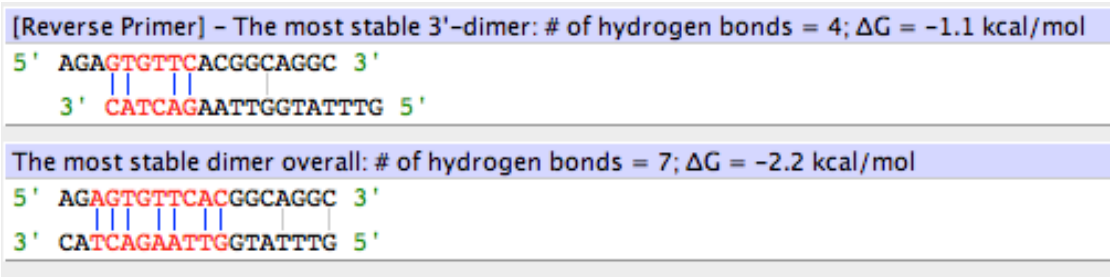
STF3 – STR1



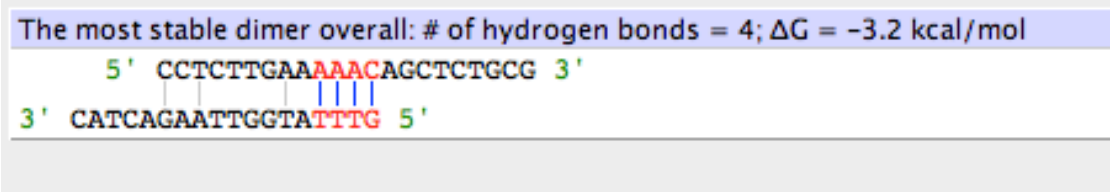
STF2 – STR2



ALV01 – STR2



STF3 – STR2



2. Nutrients

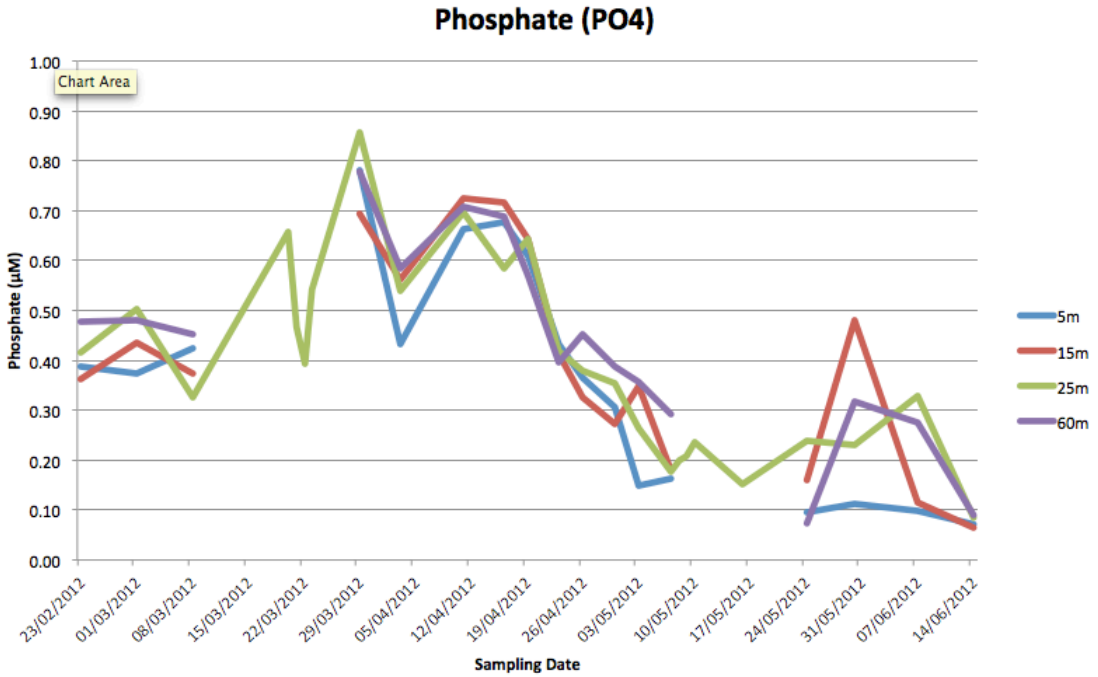


Figure S1 - Phosphate concentrations in spring 2012 at all sampling depths at the Adventfjorden sampling site ISA

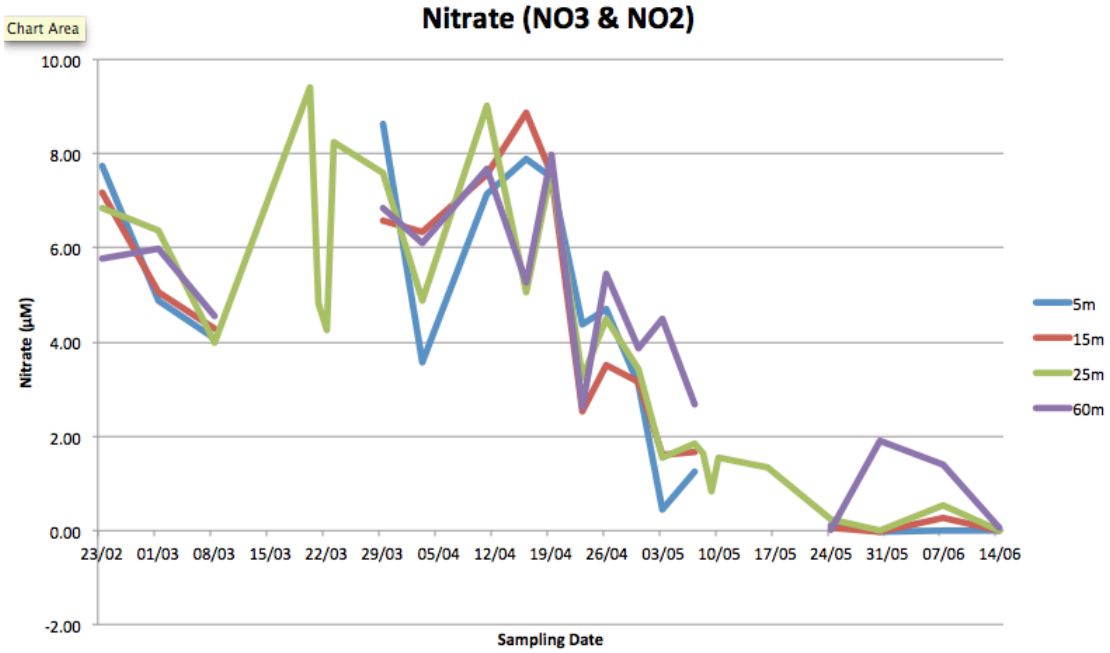


Figure S2 – Nitrate and nitrite concentrations in spring 2012 at all sampling depths at the Adventfjorden sampling site ISA

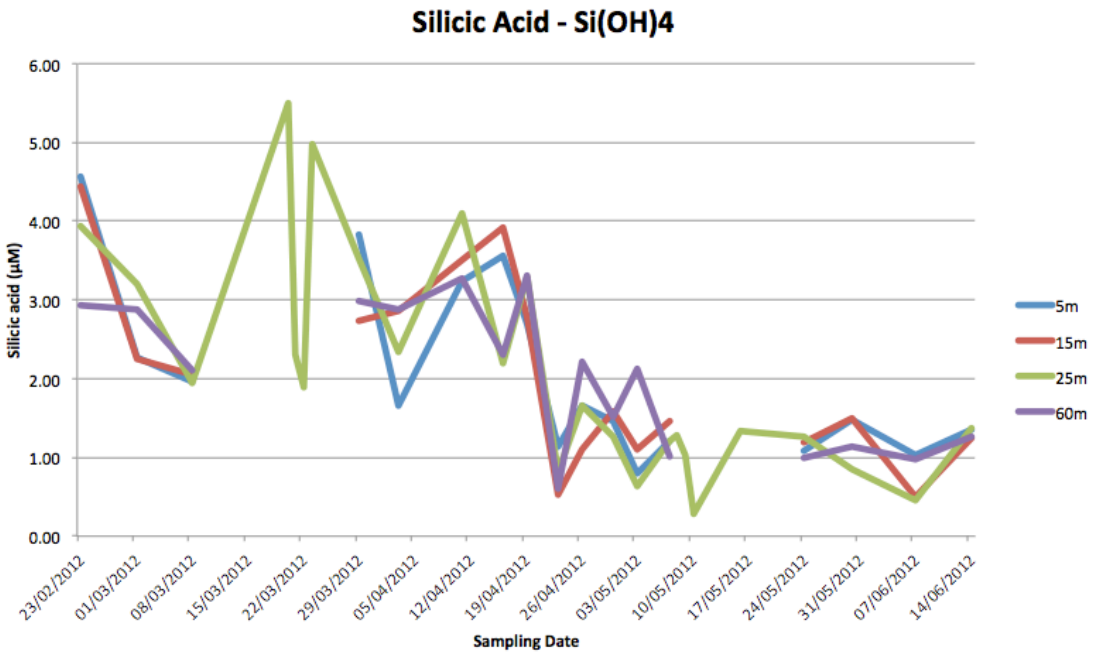


Figure S3 – Silicic acid concentrations in spring 2012 at all sampling depths at the Adventfjorden sampling site ISA

3. Statistics

Table S1. Table of fitted models. The best fitted model is in bold.

Intercept	Df	AIC	BIC	logLik
Intercept+RandomEffect	3	513.38	523.24	-253.69
Date	4	473.75	486.9	-232.87
date+Chl_GFF	5	465.17	481.61	-227.59
date+Chl_GFF+temp	6	424.15	443.88	-206.07
date+Chl_GFF+temp+salinity	7	425.65	448.66	-205.82
date+Chl_GFF+temp+depth	7	426.13	449.15	-206.07
date+Chl_GFF+temp+depth:temp	7	425.33	448.35	-205.67
date+Chl_GFF+temp+depth:sal	7	426.13	449.15	-206.07
date+Chl_GFF+temp+depth:Chl_GFF	7	423.1	446.12	-204.55
date+Chl_GFF+temp+temp:salinity	7	426.2	449.22	-206.1
date+Chl_GFF+temp+temp:Chl_GFF	7	424.22	447.24	-205.11
date+Chl_GFF+temp+temp:date	7	421.44	444.45	-203.72
date+Chl_GFF+temp+temp:date+sal:Chl_GFF	8	418.43	444.74	-201.22
date+Chl_GFF+temp+temp:date+sal:Chl_GFF+sal:date	9	407.22	436.81	-194.61
date+Chl_GFF+temp+temp:date+sal:Chl_GFF+sal:date+sal:Chl_GFF	10	408.34	441.22	-194.17
date+Chl_GFF+temp+temp:date+sal:Chl_GFF+sal:date+temp:sal:date	10	383.35	416.23	-181.67
date+Chl_GFF+temp+temp:date+sal:Chl_GFF+sal:date+temp:sal:date+temp:sal:Chl_GFF	11	383.41	419.58	-180.71

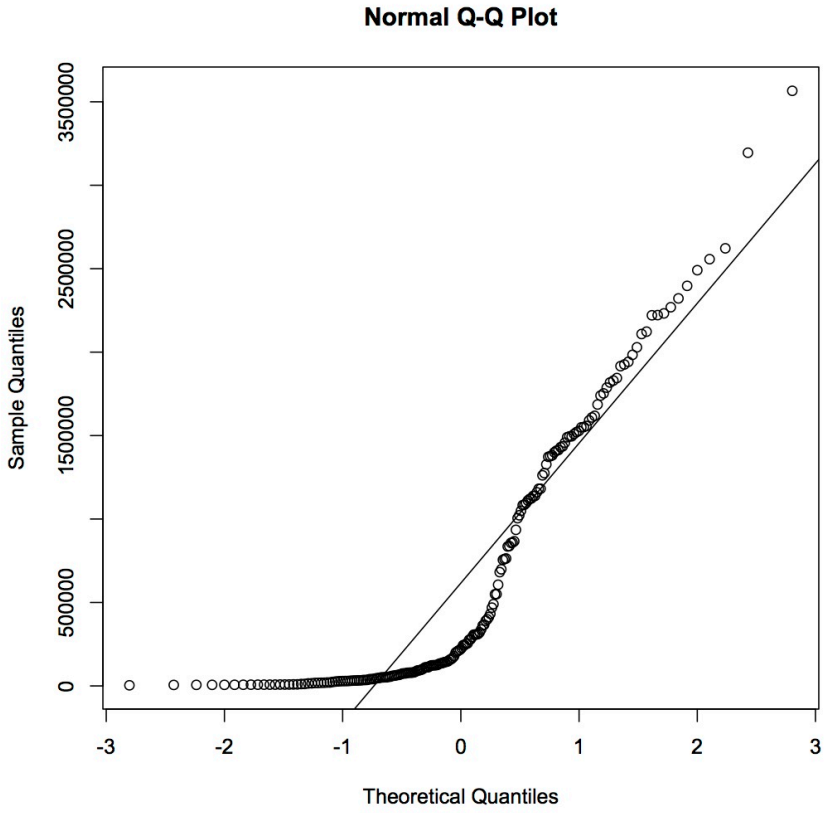


Figure S4 - Quantile-quantile plot of the raw MALV II rDNA copy number data generated from TaqMan experiment averages (see supplementary CD for this data).

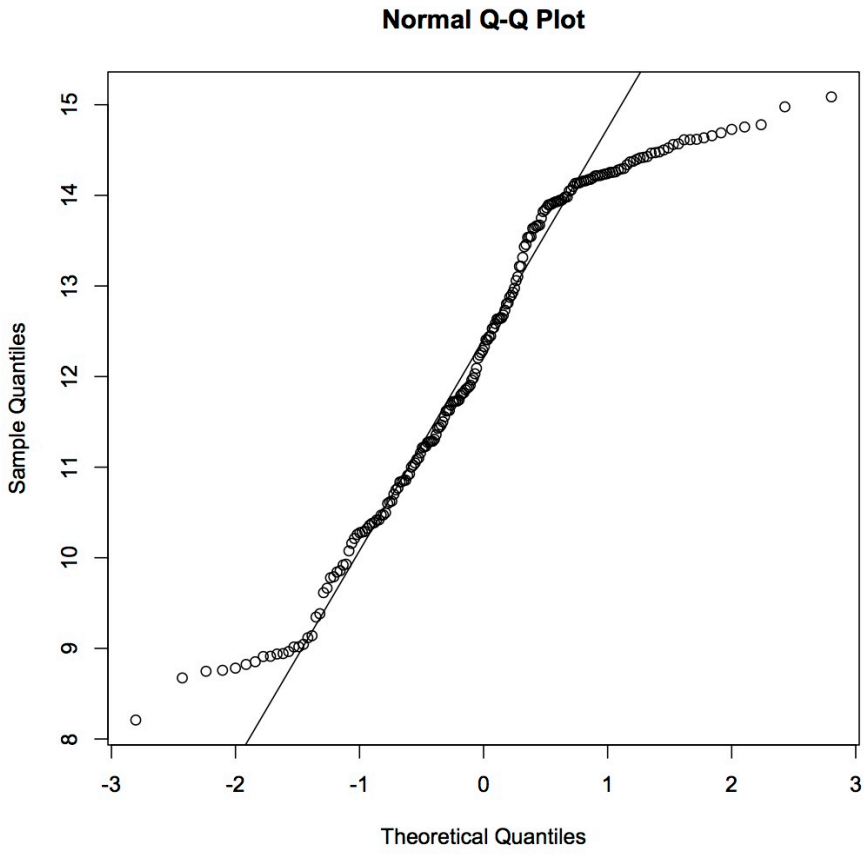


Figure S5 - Quantile-quantile plot of the log transformed MALV II rDNA copy number data generated from TaqMan experiment averages.

Table S2 – Correlations of variables in the model with log MALV II copy numbers

	Correlation with log(MALV_Copy_Nr)
Date	-0.67746020
Chlorophyll A	-0.52790840
Temperature	0.36974780
Salinity	0.25174700
Depth	0.03467517

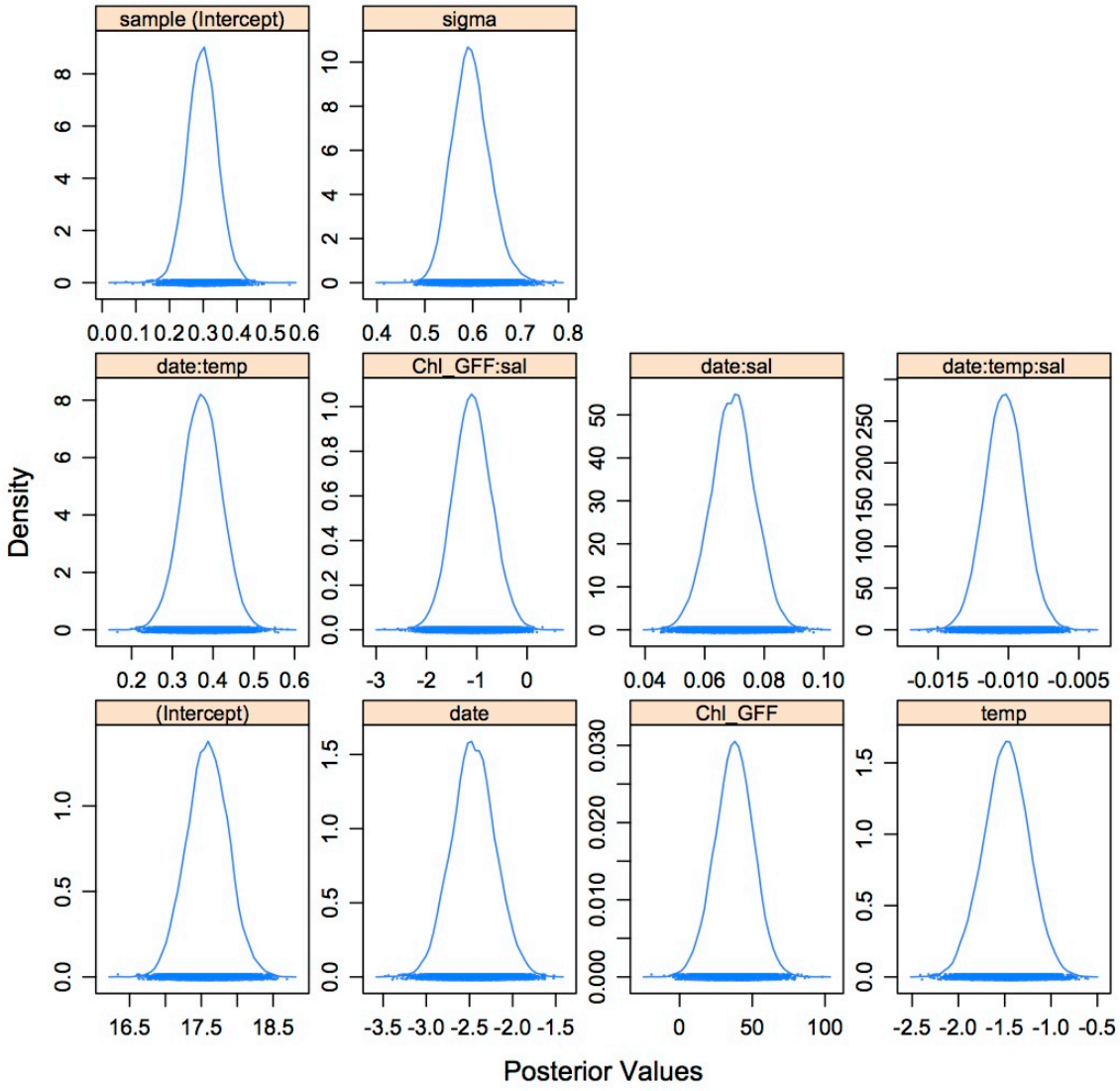


Figure S6 MCMC estimates of p-values of the best fitted model

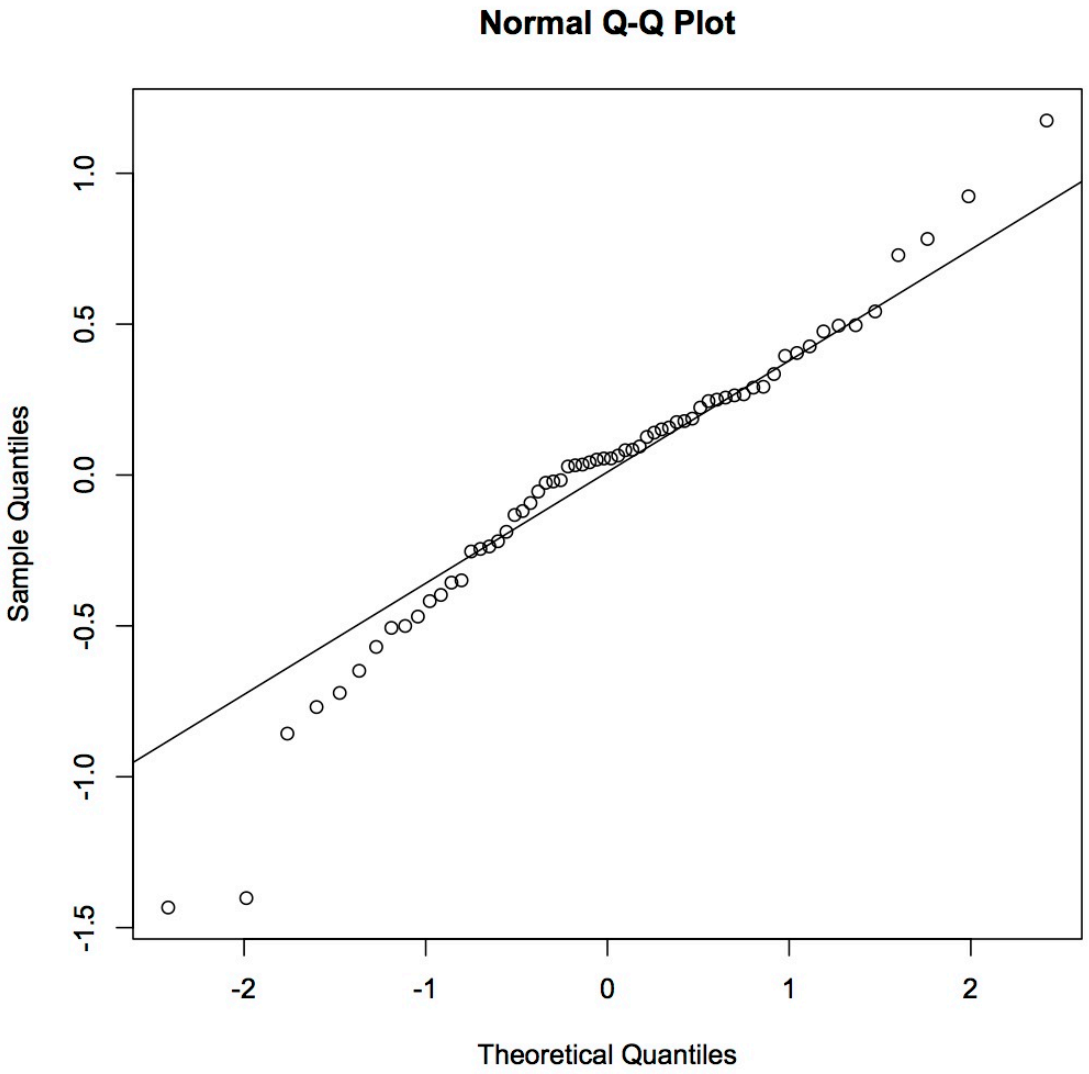


Figure S7. Quantiles plot of random effects

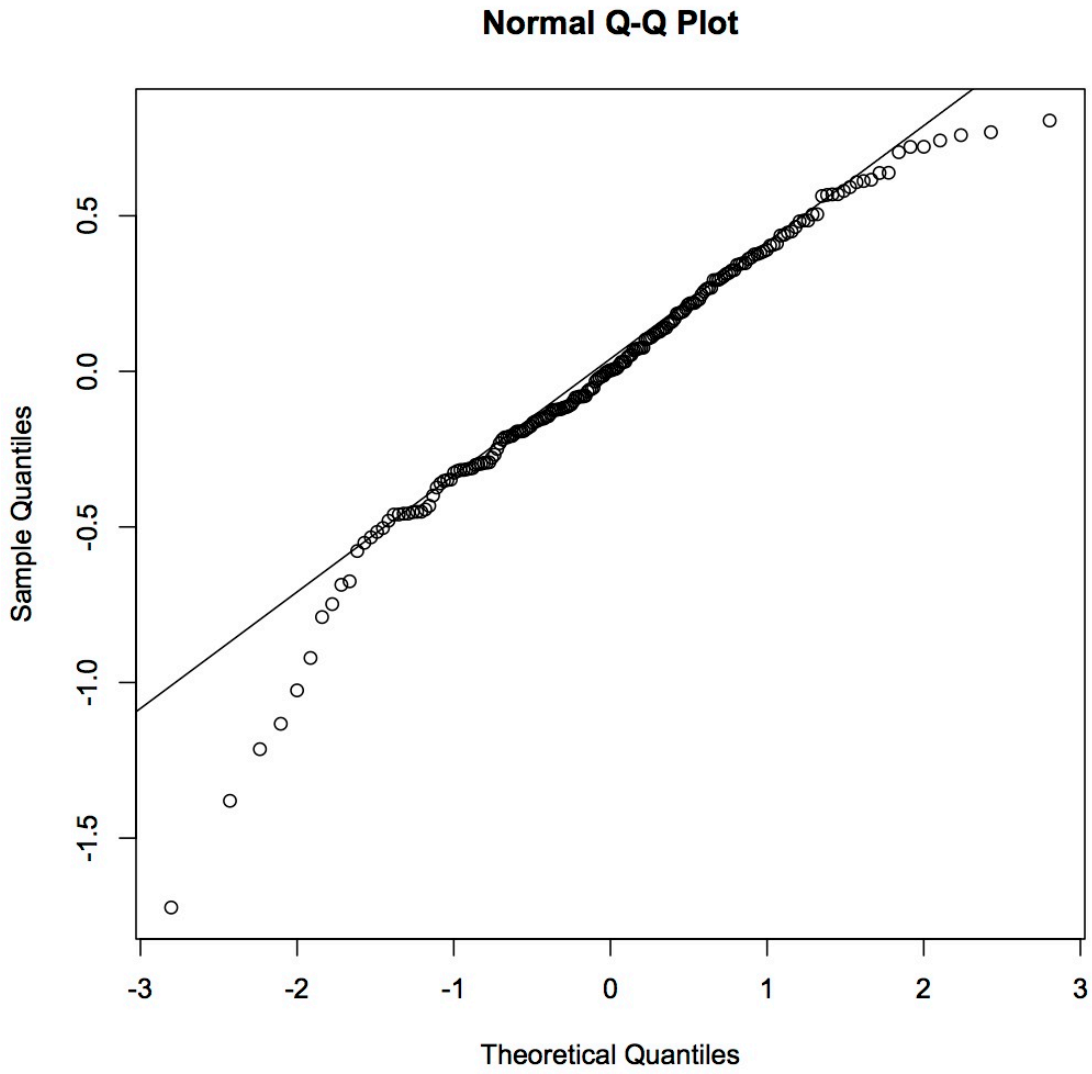


Figure S8 Quantiles plot of residuals of best fitted model

4. TaqMan

Table 1 – Assay efficiency and correlation coefficient values for standard curves in TaqMan experiments included in calculations of MALV II abundances

Experiment ID	Efficiency	Correlation
130803-2	96.35	0.998
130803	92.79	0.999
130802-2	96.12	0.998
130801	93.44	0.996
130724	91.44	0.998
130509	91.82	0.996
130514	94.23	0.996
130516	96.20	0.998
130518	95.70	0.996
130520	97.70	0.997
130522	94.84	0.998
130522-2	95.63	0.990
130702	96.03	0.995
130702-2	96.12	0.998
130711	90.70	0.999
130711-2	100.21	0.998
130712	104.08	0.998
130712-2	98.98	0.998
130715	99.28	0.997
130718	99.28	0.998
130717	98.12	0.996
130722-2	98.00	0.994
130723-3	99.84	0.998
130804	94.74	0.997
130805	96.48	0.997