INVESTIGATING BACTERIAL COMMUNITY STRUCTURE OVER TEMPORAL AND SPATIAL SCALES IN THE NORTHWEST ATLANTIC OCEAN

by

Jackie Zorz

Submitted in partial fulfilment of the requirements for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
February 2016

Table of Contents

List of Tables

List of Figures

Abstract

Bacteria are important members of every marine ecosystem and the composition of their communities has implications for global biogeochemical cycling. The Northwest Atlantic Ocean is an ecologically and economically significant region that exhibits wide ranges in physiochemical parameters that vary seasonally. In this thesis the bacterial community structure of two areas within the Northwest Atlantic Ocean, the Scotian Shelf and the Bedford Basin, were analyzed using 16S rRNA gene sequencing. The Scotian Shelf was analyzed spatially over two time points. Environmental parameters of the region, seasonality, and depth were found to heavily influence community structure. In the Bedford Basin, a weekly bacterial time series was established and the first year of data from the deepest samples were analyzed. The deep basin exhibited seasonal patterns with respect to community similarity, diversity, and composition. Overall, these datasets provide novel information regarding community composition and drivers of community shifts in this region.

List of Abbreviations and Symbols Used

16S rRNA	16S Ribosomal RNA gene
amoA	Gene for Ammonia Monooxygenase
anammox	Anaerobic ammonium oxidation
ANOVA	Analysis of Variance
AZMP	Atlantic Zonal Monitoring Program
BBL	Brown's Bank Line
BIO	Bedford Institute of Oceanography
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CCGS	Canadian Coast Guard Ship
CSL	Cabot Strait Line
CTD	Instrument used to measure Conductivity, Temperature and Depth
DGGE	Denaturing Gradient Gel Electrophoresis
DNA	Deoxyribonucleic acid
GULD	Gully Station
FISH	Fluorescence *in situ* hybridization
HL	Halifax Line
HRM	Halifax Regional Municipality
IMR	Integrated Microbiome Resource
LHB	LaHave Basin Line
LL	Louisbourg Line
MLD	Mixed Layer Depth
mvabund	Multivariate analysis of abundance data
NMDS	Non-metric Multidimensional Scaling
NOB	Nitrite Oxidizing Bacteria
O_2	Oxygen
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PF	Passing Filter
PSU	Practical Salinity Units
PyNAST	Python Nearest Alignment Space Termination
QIIME	Quantitative Insights into Microbial Ecology
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic acid
STAB	St. Anne's Bank Line
TB	Thebaud Oil and Gas Platform Station
V6-V8	Variable region 6 to Variable region 8 of 16S rRNA gene

Acknowledgements

First of all, I would like to thank my supervisor Dr. Julie LaRoche. She entrusted me with the task of collecting my own samples and samples for others in our working group, and sent me off to sea on multiple occasions during my Master's degree. The times I spent on the ocean are experiences that I will never forget. Additionally, Dr. Julie LaRoche provided me with guidance and support which assisted me in completing this research. I would also like to thank the other members of the LaRoche working group, Jenni Ratten, Dr. Dhwani Desai, Ian Luddington, Jennifer Tolman, Joerg Behnke, Afrah Alothman, Jude van der Meer, and Hannah Blanchard. I am forever grateful for the support and intellectual discussion from this group which motivated me throughout my time here, and gave me new ideas to pursue in my research. I would also like to extend gratitude to Scott McCain who was always available and eager to answer my incessant statistics questions. For the time I spent on the CCGS *Hudson* I would like to thank the crew who took exceptional care of all the scientists. Additionally, I owe a considerable amount to the Bedford Institute of Oceanography who allowed students to come aboard their annual research cruises, made amendments to their sampling protocols in order to satisfy my obnoxious water requirements, and shared all their physiochemical data with me so that I could have a multifaceted dataset. I also need to thank BIO for sharing data from their weekly sampling series in the Bedford Basin, and allowing members of Dalhousie University to piggyback on their sampling protocol. By extension, I would like to thank the crew from Dalhousie University, headed by Anna Haverstock and Richard Davis, with funding from MEOPAR and CERC, who collected our water samples from the Bedford Basin every week, regardless of weather. I would also like to thank Andre Comeau and Morgan Langille of the Integrated Microbiome Resource working group, who not only prepared our library and ran our samples for sequencing, but also continue to answer questions and provide essential support for the analysis of our sequencing data. Lastly I would like to thank Dalhousie University, the government of Nova Scotia, and the National Science and Engineering Council of Canada for providing the funding that made this research possible.

Chapter 1: Introduction

Understanding marine microbial community structure is paramount to interpreting the biogeochemical cycling of nutrients in the ocean (Arrigo, 2005). The metabolism of countless microbes drives these cycles through such processes as photosynthesis, nitrogen fixation, and nutrient recycling. The ability to characterize members of microbial communities provides insight into the metabolic processes that are occurring in a given ecosystem. Throughout the world's oceans, these processes can be controlled by vastly different groups of organisms that vary widely across time and space (Falkowski *et al.*, 2008). The outcome of oceanic biogeochemical cycling has implications in several issues important to human culture and industry including carbon sequestration, bioremediation, nutrient availability, and marine primary production which has consequences for higher trophic levels (Atlas, 1995; Azam, 1998; Falkowski, 1998; Herbert, 1999). It is thus of significance to study the members of microbial communities and to determine the potential metabolic pathways they possess.

The main issue when describing microbial communities is that most species resist culturing using traditional laboratory based techniques with the result being that the number of sequenced genomes are biased towards cultured representatives (Rappé and Giovannoni, 2003). Additionally, culturing success only decreases with depth in the water column (Simonato *et al.*, 2006), and still misses the majority of environmental microbial diversity. Developing more effective, alternative ways to characterize microbial communities, has been the aim of microbial ecology in the past decades. With the increased efficiency introduced by high throughput sequencing technologies and increased computing power, techniques like 16S rRNA tag sequencing and metagenomics have become common and accessible ways to comprehensively analyze microbial community structure (Tringe and Hugenholtz, 2008). The outcome of these techniques is the ability to sample environmental microbial communities *in situ* which substantially increases the degree of microbial diversity that can be observed (Caporaso, Lauber, *et al.*, 2012; Lynch and Neufeld, 2015). Since the inception of high throughput sequencing, knowledge on the expanse of microbial diversity has exploded, but these techniques have also highlighted the collective scientific ignorance of the diversity of

environmental microbes. Thus these technologies have ushered in a new and exciting era of discovery with respect to environmental microbial ecology.

The consequences of anthropogenic climate change on the world's oceans have only begun to be understood. Ocean warming, acidification, and expanding oxygen minimum zones are all examples of immediate threats to ocean homeostasis (Stramma *et al.*, 2008; Finkel *et al.*, 2010). These chemical and physical processes have been monitored regularly in the ocean, however the effect on biological systems, particularly microbial systems, is greatly understudied. Considering the incredible role marine microbial communities play in ocean functioning and cycling, it is very important to understand how microbial communities will change in a changing ocean and how this will affect overall ocean health. With the introduction of cheaper sequencing methods and *in situ* experiments, this momentous gap in knowledge is beginning to be explored. Many studies have worked towards establishing a "baseline" of microbial community structure across numerous marine environments (Gilbert *et al.*, 2012; Vezzulli *et al.*, 2012; Karl and Church, 2014; Cram *et al.*, 2015; Fuhrman *et al.*, 2015). This "who is there" survey can be conducted as a sampling time series with a predefined temporal frequency, and changes over time in both the physical and biological ocean can be monitored and compared. Ideally, the microbial survey of a particular region should be conducted at a frequency that allows for the shifts in community structure to be resolved. Realistically however, restrictions in funding and site accessibility will limit the frequency of most regional microbial surveys.

The following work aims to establish a baseline of the current microbial community for two economically and ecologically important regions in the Northwest Atlantic, realizing that the marine environment is not static but shifting with major climatic trends and anthropogenic activity. Chapter 2 of this thesis focuses on determining bacterial community structure of the Scotian Shelf region off the south coast of Nova Scotia, Canada. This region is a vast (700km x ~200km) continental shelf that was visited twice, aboard two research cruises in April 2014 and September/October 2014. In addition to the latitudinal and longitudinal sampling dimensions, the vertical water column was also sampled at each site to examine structure across different depths. In total 168 sites were sampled from both cruises of the Scotian Shelf, and this work

represents the first extensive bacterial survey of this Northwest Atlantic region, adding to the existing knowledge on eukaryotic community structure of the area (Dasilva *et al.*, 2014). Bacterial patterns of diversity over the Scotian Shelf across four dimensions; two dimensional space, depth, and time, were analyzed. Additionally, the environmental factors that drive overall community change, and taxa specific spatial distribution were identified.

Building on a 20 year time series carried out by the Bedford Institute of Oceanography (BIO), Chapter 3 of this thesis focuses on expanding the weekly measurements of the Bedford Basin time series to include microbial community structure obtained by 16S rRNA tag sequencing. The Bedford Basin is a coastal basin connected to the Scotian Shelf, situated entirely within the Halifax Regional Municipality (HRM) in Nova Scotia, Canada. The addition of microbial community structure to the time series began in 2014 and weekly samples have been collected over four depths: 1 m, 5 m, 10 m, and 60 m since then. This project compiles the results of the first year of sampling and represents the first analysis from this high resolution microbial time series. Specifically, this chapter will focus on the bacterial patterns of abundance and diversity at 60 m depth, and compare these results with previous studies of the Bedford Basin (El-Swais *et al.*, 2014; Georges *et al.*, 2014). Also this chapter will focus on "spiking" or fleetingly abundant species, which are rare for the majority of the year but become abundant rapidly over the course of one or two weeks.

Environmentally dynamic deep water intrusion events occurred twice during 2014 and their effect on the bacterial community at 60 m depth will also be explored. This time series of microbial community composition initiated by the LaRoche lab is ongoing and samples have been collected since 2014 and will continue to be collected for the foreseeable future, thus increasing the scientific value of the data analyzed in this chapter. Additionally, the weekly sampling structure of this time series allows for the identification of rapid shifts in microbial community structure which have rarely been observed in other microbial time series studies, with some exceptions (El-Swais *et al.*, 2014; Lindh *et al.*, 2015).

In Chapter 4, the general discussion, I compare results from the Scotian Shelf and Bedford Basin analyses, specifically commenting on the similarities and differences of

the most abundant and most dynamic taxa across both datasets. I also discuss the findings of these studies that would be most interesting to explore further into the future.

Chapter 2: Bacterial Community Structure of the Scotian Shelf

2.0 Abstract

An extensive bacterial survey of the biologically productive and diverse Scotian Shelf region in the Northwest Atlantic Ocean has been lacking to date. This study provides the first comprehensive analysis of the bacterial communities of the Scotian Shelf using 16S rRNA sequencing. The region was sampled in the spring and fall of 2014, across multiple depths and transects, thus creating a multidimensional dataset. Overall, only a few bacterial taxa dominated the majority of the sites sampled. Specifically, species from Pelagibacteraceae, Rhodobacteraceae, and Bacteroidetes were most common, although every bacterial taxon exhibited distinct biogeographic patterns seasonally, across the shelf, and throughout the water column. The season and depth at which a sample was taken also had pronounced effects on the environmental variables that correlated with a given site's community diversity. When all sites were taken into consideration, environment was shown to have a greater effect on community composition than geographic distance. Oxygen and temperature were the individual variables that displayed the highest correlation with community similarity over all sites. Season was the largest driver of bacterial community change, followed by depth in the water column, but again different taxa had very distinct responses to these groupings. The extensive dataset produced from this study establishes a baseline of bacterial community composition that can be used to monitor how the bacterial community of this region changes with a changing climate.

2.1 Introduction

A current central aim of microbial ecology is to understand the biogeography of microorganisms, or rather how microorganisms are distributed across space and time and what factors determine this distribution (Hanson *et al.*, 2012). The famous mantra describing the prevalent theory on microbial biogeography was coined by Baas-Becking decades ago: "Everything is everywhere, but the environment selects" (Baas Becking, 1934). This view has been reiterated frequently since, however only recently has the

technology become available to directly test the validity of this statement (Green and Bohannan, 2006). Essentially, because microorganisms are small, have extremely large populations and high turnover rates, it is assumed that their dispersal limitation is negligible and thus differences in species distribution are mainly due to environmental heterogeneity, also referred to as contemporary selection (Green and Bohannan, 2006; Hanson *et al.*, 2012). Historical processes, such as dispersal limitation, result in the formation of a distance decay relationship, defined by a decay in similarity between different communities as a function of the geographic distance separating them (Nekola and White, 1999). The effects of contemporary selection and historical processes are often confounded in distance decay relationships due to the spatial autocorrelation of environmental characteristics with geographic distance. Therefore, if the origin of the variation in microbial community structure is of concern, the environmental effects on distance must be taken into account (Borcard *et al.*, 1992; Peres-Neto and Legendre, 2010). Previous studies have found evidence for distance decay patterns purely related to geographic distance in microbial communities, most notably in severely isolated habitats (dispersal limited) or extreme environments (Papke *et al.*, 2003; Whitaker *et al.*, 2003; Takacs-Vesbach *et al.*, 2008; Pagaling *et al.*, 2009) However, there is a lack of consensus and support for a holistic distinct relationship between geographic distance and microbial community structure in uninterrupted environments with fewer dispersal barriers (Hewson *et al.*, 2006; Schauer *et al.*, 2010; Monier *et al.*, 2014; Zinger *et al.*, 2014; Nguyen and Landfald, 2015). Generally, global studies identify spatial separation as the main driver of community change, microscale studies often only identify environmental effects, and regional studies usually find evidence for both contemporary and historical processes (Martiny *et al.*, 2006; Wang *et al.*, 2015). However, as microbial biogeography is still a relatively new discipline, more research is required, particularly in the marine environment, to determine with confidence the effect of geography and environment on microbial communities.

As contemporary selection is often a predominant determinant of microbial community structure, identifying the specific environmental variables that affect this structure over global and local scales is another main goal of microbial biogeography. Past studies on microorganisms in the marine environment have found that the factors

that affect microbial distribution can vary based on spatial and temporal scales, as well as location (Pommier *et al.*, 2007; Fuhrman *et al.*, 2015). Specific bacterial taxa will respond differently to environmental gradients, and these differences can vary depending on the taxonomic resolution used, further complicating the ability to identify relationships (Gómez-Pereira *et al.*, 2010; Fu *et al.*, 2013). The distribution of some bacterial taxa are known to be cosmopolitan, e.g. SAR11 cluster, or defined by latitude, e.g. cyanobacterial genera of *Prochlorococcus* and *Synechococcus* (Morris *et al.*; Partensky *et al.*, 1999; Ladau *et al.*, 2013), however the distribution of most marine bacterial taxa is unknown despite their important roles in nutrient cycling (Arrigo, 2005; Swan *et al.*, 2011). Thus the study of these underrepresented clades is important in gaining a holistic overview of microbial distribution in the global oceans.

The marine environment is complex and three-dimensional, with variability expected over latitude, longitude, and depth, in addition to seasonal changes (Rusch *et al.*, 2007; Gilbert *et al.*, 2012). Few studies attempt to address these multiple dimensions simultaneously. The majority of the ocean is below the reaches of solar radiation and the processes that shape microbial communities at the surface are unlikely to be the same that shape microbial community structure at depth (Galand *et al.*, 2010; Orcutt *et al.*, 2011). As waters below the photic zone are so extensive, and are associated with important processes such as nutrient recycling and carbon fixation, understanding more about the microbial community that resides there is imperative. In this chapter, the effect of geographic, and environmental distance on microbial community similarity are determined in in the Scotian Shelf region of the North Atlantic Ocean, at differing depths and seasons. Additionally, the specific environmental variables that most influence community structure are identified.

The Scotian Shelf is a dynamic, highly productive region within the Northwest Atlantic Shelves Province (Longhurst 1998) that has been well studied with respect to its physical and chemical characteristics (Hannah *et al.*, 2001; Drinkwater and Gilbert, 2004; Townsend *et al.*, 2005), but is lacking in comprehensive studies of its microbial community structure (Yeung et al. 2010; Dasilva et al. 2014). The Scotian Shelf extends offshore between 120 km and 240 km and is on average 90 m deep but rapidly drops off to below 3000 m at the Scotian Slope. It is 700 km in length and is bound by the

Laurentian Channel on the Northeast and the Northeast Channel and the Gulf of Maine on the Southwest. The region is influenced by shelf and bank-scale forces that are caused by the cold southward current from the Labrador Sea that flows over the shelf and the warm northward flow from the Gulf Stream which can extend its influence as far north as the Scotian Slope (Loder et al, 1998; Hannah et al, 2001). These two main currents result in a general pattern of decreasing temperature and salinity across the Scotian Shelf from the SW to the NE (Drinkwater and Gilbert, 2004). In addition, warmer and more saline offshore waters from the south result in a gradient of increasing salinity and temperature from the inshore region to the offshore region (Drinkwater and Gilbert 2004; Dasilva et al. 2014)

The Scotian Shelf also experiences many seasonal changes and can exhibit a temperature range in surface waters of about 16°C which is one of the widest in the Atlantic Ocean (Drinkwater and Gilbert, 2004). At depths greater than 100-150 m however, there is effectively no change in temperature throughout the year (Drinkwater and Gilbert, 2004). During the seasonal heating of the surface layer the water column becomes stratified and traps a colder intermediate layer beneath the seasonal thermocline and above the warmer and saltier bottom waters that intrude from beyond the shelf edge (Townsend *et al.*, 2005). These factors create a diverse environment with respect to temperature throughout the water column, that in surface waters, can change drastically seasonally but at depth remains fairly constant.

The Scotian Shelf is a unique marine environment that in the spring and to some extent in the fall, harbours favourable conditions for the massive growth of phytoplankton, resulting in phytoplankton blooms and increased productivity in the region. These blooms are vital for supporting higher trophic levels in the region, and consequently many studies have previously observed and measured the timing and strength of phytoplankton blooms on the Scotian Shelf (Song *et al.*, 2011; Zhai *et al.*, 2011; Craig *et al.*, 2014). The Scotian Shelf is of further interest because it is a primary spawning area for many commercial fishes in the region (Brander and Hurley, 1992) and because of the development of offshore oil platforms around Sable Island, making the region both ecologically and economically significant.

Biannually since 1998, the Atlantic Zone Monitoring Program (AZMP) has measured physical, chemical, and biological parameters of the region via research cruises in the spring and fall seasons (Therriault et al. 1998). To date, the majority of the biological microbial work on the Scotian Shelf has been done using methods involving flow cytometry and microscopy (Li *et al.*, 2011). Only two 16S rRNA gene based microbial surveys of the Gully region and the area surrounding the Thebaud oil and gas platform (Yeung *et al.*, 2010, 2011), and one eukaryotic 18S rRNA gene microbial survey of select areas on the Scotian Shelf have been published (Dasilva et al. 2014).

We report here the first bacterial survey that spans the Scotian Shelf throughout the vertical water column in two distinct seasons, spring and early fall. We use 16S rRNA gene sequencing and physiochemical measurements to generate a multifaceted dataset of the bacterial community structure at various stations and depths across the shelf. This dataset provides a tremendous opportunity to explore patterns in bacterial community structure spatially, throughout the water column, and over two seasons, and will undoubtedly be used for other analyses in the future as well as a resource for other studies in the region. The specific goals of this study were first to determine the bacterial community composition and drivers of bacterial diversity across the Scotian Shelf over the spring and fall seasons (section 2.3.2). Secondly, the dataset was used to determine the drivers of regional bacterial community structure, specifically the main geographic and environmental factors driving the biogeography of the region (section 2.3.3). As environmental factors were found to be the strongest driver of community change in the region (section 2.3.3), the response of individual taxa to environmental gradients and other taxa, was then explored, as well as the effect of specific seasonal, depth, and location groupings on discrete taxa and OTUs (section 2.3.4). Lastly, the final goal of this study was to compare the community composition of the Thebaud oil and gas platform to the bacterial communities of surrounding shelf stations (section 2.3.5). In short, this chapter provides a comprehensive investigation into the constituents of the bacterial community of the Scotian Shelf across space, depth, and time, and determines the environmental and geographic factors that drive changes in bacterial community structure in this region.

2.2 Materials and Methods

2.2.1 Sample Collection

Water samples from the Scotian Shelf were collected on two voyages aboard the CCGS *Hudson*; April 4-23, 2014 and September 19 - October 12, 2014. The cruises surveyed various transects beginning near shore and extending across the Scotian Shelf and Scotian Slope, out to the open ocean. At select stations (Figure 2.1), 4 L of water was collected in triplicate from four discrete depths through a 160 µm mesh from 12 L Niskin-type bottles attached to a rosette. The water was then immediately pre-filtered using a vacuum pump through 3µm polycarbonate Isopore filters (Millipore). Depending on the filtering rates during the pre-filtration process, either 2 L or 4 L of the pre-filtered water was then redistributed into 4 L bottles and vacuum filtered through 0.2 µm polycarbonate Isopore filters (Millipore). All samples were collected in triplicate and the filters were stored at -80˚C. Of the four depths chosen for sampling, two depths were always the surface layer and 20 m, and the other two depths varied depending on the depth of the ocean floor at the station. Whenever the station was deep enough, a sample was collected from the oxygen minima of the water column which usually corresponded to 250 or 300 m.

At all depths and all stations sampled for DNA extraction, 1.8 mL of seawater was collected and fixed with 1% paraformaldehyde (Alfa-Aesar). These samples were left at room temperature for an hour and then stored at -80˚C for later flow cytometry analysis.

All sites sampled were grouped into the following discrete categories: Above or below the mixed layer depth (MLD), located near the coast, on the shelf or in the open ocean (Figure 2.1), and whether they were collected during the spring or fall cruise (Figure 2.1). MLD was determined by calculating the minimum depth where the density gradient ($gradient_z(\sigma_t)$) was equal to or exceeded 0.01 kg m^{-4} and verified by visual observation (Johnson *et al.*, 2014). Stations located closest to the shoreline were referred to as coastal stations, shelf stations were firmly located on the Scotian Shelf, and open stations were located off the shelf with a water column depth of greater than 500 m.

All physiochemical data measured aboard the CCGS *Hudson* was collected by BIO and the details of their collection and sampling regime can be found elsewhere (Mitchell *et al.*, 2002).

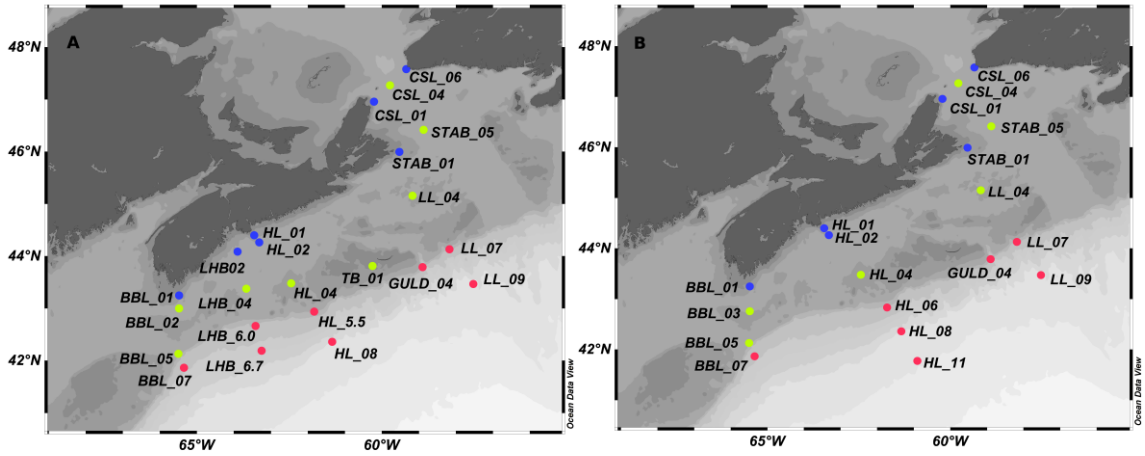


Figure 2.1. Map of stations sampled aboard the CCGS *Hudson* in April (A) and September/October (B) of 2014. Stations are labelled with their transect acronym and station number and coloured according to their location groupings: coast (blue), shelf (green), or open (red).

2.2.2 Flow Cytometry

All flow cytometry samples were analyzed with a BD Accuri Flow Cytometer. In order to measure bacterial concentrations, SYBR (Invitrogen) stain was added to a final concentration of 1:10,000 to a sample of seawater and incubated in the dark for 15 minutes. The SYBR stained sample was run with a threshold of 800 at FL1 and the gates used to determine bacterial counts followed a bacterial gating strategy developed for the Accuri instrument (Gatza *et al.*, 2013; Prest *et al.*, 2013).

2.2.3 DNA Extraction

All DNA was extracted from the 0.2 µm polycarbonate filters using the DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions with some alterations in the cell lysis procedure. 50 µL of lysozyme (5 mg/ml) (Fisher BioReagents) was initially added to each filter sample and each sample was vortexed on high for 30 seconds. Then 400 µL of lysis buffer AP1 (from the DNeasy kit) was added to each sample tube followed by the addition of 45 µL of proteinase K (20 mg/mL) (Fisher BioReagents). The samples were then incubated at 55˚C with shaking for one hour. After

this incubation, 4 µL of RNase A (Qiagen) was added to the samples which were then kept on ice for 10 minutes. From here the extraction followed published protocols except in the last step, 100 µL of DNA was eluted. DNA concentrations and purity were confirmed with a NanoDrop 2000 (Thermo Scientific).

2.2.4 Amplification and Library Preparation

The sample preparation procedure for the Illumina MiSeq sequencer followed that as described in the Microbiome Amplicon Sequencing Workflow (Comeau and Langille, unpublished, also at cgeb-imr.ca/protocols.html). Briefly, each sample of extracted DNA was amplified using dual-indexing Illumina fusion primers that targeted the V6-V8, 438 bp region, of the bacterial 16S rRNA gene (Comeau et al. 2011; Comeau and Langille, unpublished) (Supplementary Table S2). Each sample was amplified in duplicate, with one duplicate containing the original extracted DNA template and the other containing a 1:10 template dilution. The dilution series was done to reduce the effects of PCR bias. The pooled duplicate PCR product quality was verified using an E-gel 96-well high-throughput system (Invitrogen). Library normalization and PCR clean-up was conducted using a SequalPrep 96-well Plate Kit (Invitrogen). After normalization, all samples were pooled together and the final library pool was quantified using Qubit with PicoGreen (Invitrogen).

2.2.5 Illumina MiSeq Sequencing

After quantification, the pooled samples were run on an Illumina MiSeq Sequencer using paired-end 300+300 bp v3 chemistry. Approximately 16,918,000 raw reads were obtained from the Illumina sequencing run, and 15,394,747 of these sequences passed the PF (passing filter) quality control (Supplementary Table S3). The MiSeq on-board software demultiplexed the reads by trimming the barcode sequences from the sequence reads and assigning each read to the appropriate sample. The number of sequences at each step of the clean-up process is compiled in Supplementary Table S3.

2.2.6 QIIME 16S rRNA Data Analysis

All preliminary analysis and processing of 16S sequences followed a QIIME version 1.8.0 (Caporaso, Kuczynski, *et al.*, 2010) pipeline workflow (Langille, github.com/mlangill/microbiome_helper). The program PEAR version 0.9.6 was first used to merge the demultiplexed, paired-end sequences together (J. Zhang *et al.*, 2014). After merging paired ends, sequences less than 400 bp in length or with a quality less than 30 over 90% of bases were discarded. Chimeric sequences were removed using UCHIME (Edgar *et al.*, 2011). OTUs were picked based on 97% sequence similarity using sortmerna (Kopylova et al., 2012) for reference picking and sumaclust (Mercier *et al.*, 2013) for *de novo* OTU picking (i.e. "open-reference" picking, referred to as "New.ReferenceOTU" in results). This process uses the reference Greengenes database version 13.8 (McDonald *et al.*, 2012) for preliminary OTU picking, and then subsamples failed sequences using *de novo* picking. From this point on, only OTUs with two or more sequences (no singletons) that aligned with PyNAST (Caporaso, Bittinger, *et al.*, 2010) were used for further analysis. Additional quality control measures included removing all sequences belonging to Archaea and Chloroplasts from the dataset as the primers used in this study were only meant to amplify bacterial sequences reliably.

The Greengene's database erroneously classifies the family of SAR86 within the class of Gammaproteobacteria as the genus "*Candidatus Portiera*", which is primarily an endosymbiont of the white fly *Bemisia tabaci* (Jiang *et al.*, 2012). All OTUs assigned to the *Candidatus Portiera* classification were reclassified as belonging to the family of SAR86. A BLAST search of these misclassified sequences was also done to support this decision.

In order to compare relative abundance of OTUs between samples, sequences were rarefied to a sequencing depth of 7500 which corresponded to the duplicate sample pair with the lowest sequencing depth. This resulted in the loss of a replicate from the following samples from the spring cruise: HL2_20, LHB6.7_250, HL2_80, LHB6_20, LL7_250, BBL7_20, BBL1_10, STAB05_1, BBL1_1, TB01_40, HL5.5_80, GULD04_20, and LL7_80. The observations from the remaining duplicate samples were combined by averaging using the QIIME script collapse_samples.py.

The taxonomic output provided by QIIME contains OTUs that are identified to varying taxonomic levels. For instance some OTUs can be identified to genus level, while others are only identified to the level of phylum. This creates difficulties when attempting to choose only one taxonomic level to compare across samples: selecting phylum will result in the loss of intricate data patterns and the loss of underlying functional meaning in results, while selecting genus or even family, order, or class, is not possible for all OTUs. For this reason, the taxonomic output was grouped for some analyses according to a series of steps that resulted in ecologically meaningful clades, rather than by taxonomic level. First, all OTUs that represented more than 5% of relative abundance for at least one site in the Scotian Shelf dataset were identified. The highest taxonomic level that these OTUs could be identified to was noted and all other OTUs that could be classified to this level of taxonomic resolution (e.g. the family of Pelagibacteraceae) were grouped. Following this step, all remaining OTUs that did not fall into the first groups were classified to the level of phylum or class (if belonging to Proteobacteria). Table 2.1 summarizes this process and the daughter taxa that are absent from the parent taxon's grouping are listed. The result of this approach is a more manageable and ecologically meaningful dataset of 55 taxa (as compared to 16410 OTUs) that contains information from every sequence identified (Supplementary Table S4).

Table 2.1. Summary of the taxonomy of the OTUs that represented at least 5% of the community composition at one site over the entire Scotian Shelf dataset. Parent taxa are included to show which taxa are omitted from their overall abundance counts.

Taxa representing at least 5%	Level of Taxonomy	Parent Taxa
Nocardioides	Genus	Actinobacteria
Pelagibacteraceae	Family	Alphaproteobacteria
Rhodobacteraceae	Family	Alphaproteobacteria
Methylobacterium	Genus	Alphaproteobacteria
Erythrobacteraceae	Family	Alphaproteobacteria
Pseudoalteromonas	Genus	Alteromonadales/Gammaproteobacteria
Alteromonas	Genus	Alteromonadales/Gammaproteobacteria
Polaribacter	Genus	Bacteroidetes
Ulvibacter	Genus	Bacteroidetes
Prochlorococcus	Genus	Cyanobacteria
Synechococcus	Genus	Cyanobacteria
SAR324	Family	Deltaproteobacteria
Alteromonadales	Order	Gammaproteobacteria
Vibrionaceae	Family	Gammaproteobacteria
Oceanospirillales	Order	Gammaproteobacteria
SAR86	Family	Oceanospirillales/Gammaproteobacteria
Alcanivorax	Genus	Oceanospirillales/Gammaproteobacteria
Octadecabacter	Genus	Rhodobacteraceae/Alphaproteobacteria

2.2.7 Statistical Analysis and Data Visualization

Statistical analyses were done using either base R version 3.2.1 (R Development Core Team, 2015) or the specific R packages described below. Figures, excluding networks and maps, were made using base R, *gplots*, or the *ggplot2* packages in R (Wickham and Chang, 2015). Network diagrams, and network statistics were generated using Cytoscape v 3.2.1 (Shannon *et al.*, 2003). Ocean Data View version 4.*6.5* (Schlitzer, 2015) was used to create images featuring maps and transects of the Scotian Shelf. Density shading using the DIVA gridding algorithm was implemented through Ocean Data View to visualize approximate spatial distribution of environmental variables and the distribution of select taxa (Barth *et al.*, 2010).

The Shannon diversity index was used as a measure of alpha diversity and calculated using the entire dataset through QIIME. NMDS (non-metric multidimensional scaling) analyses were carried out on all sites. The abundance data was Hellinger transformed prior to conducting the NMDS analysis on Bray-Curtis dissimilarities using the *metaMDS* function in R from the package *vegan* (Dixon, 2003). The Hellinger

transformation is recommended for use on species abundance data and gives low weights to variables with low counts and many zeroes (Legendre and Gallagher, 2001; Ramette, 2007; Buttigieg and Ramette, 2014). The specific transformation involves dividing each value in a data matrix by its row sum and taking the square root of the quotient. For the NMDS plot of all sites, only taxa that cumulatively made up 99% of the relative abundance of the dataset were used. The function *envfit* from *vegan* was used to fit environmental vectors onto the NMDS ordination plot for all sites.

For the analysis of biogeography in the Scotian Shelf, abundance, environmental, and geographic distance matrices were created. The data for this analysis was either kept complete to determine overall patterns, or divided by season and depth to identify specific differences in community drivers. For the abundance matrices, Bray-Curtis dissimilarities were calculated after Hellinger transformation. The variables included in the environmental distance matrix were temperature, salinity, chlorophyll, oxygen, nitrite, nitrate, phosphate, silicate, ammonium, Shannon diversity, and bacterial concentration. All environmental variables were standardized (centered and scaled) before the Euclidean distance between sites was calculated. Partial Mantel Tests were then calculated with the *mantel.partial* function from the package *vegan* (Spearman's rank correlation, 10000 permutations) in order to test the significance of the correlation between either geographic distance or environmental distance, and community similarity while controlling for the influence of the other matrix.

To further determine which specific environmental variables best explain the variations in community structure across the Scotian Shelf, Partial Mantel Tests were used as described above but with only one environmental variable contributing to the environmental matrix. The geographic distance matrix was still used as the control for these tests. In this way, the effect of each environmental variable on community structure could be determined over all sites, or compared across seasons and depth.

The package *indicspecies* and the function *multipatt* were used to identify indicator species pertaining to seasons, mixed layer depth, and location on the shelf (De Cáceres *et al.*, 2010). Specifically, the *multipatt* function was used to determine lists of species that were associated to particular groups of sites according to an abundance-based point biserial correlation coefficient. Statistical significance with this method was

determined using permutation tests (9999 permutations). Networks of OTUs were created using OTUs that exhibited significant Spearman's correlations (positive or negative) with other OTUs in the dataset. Only highly correlated OTUs with a *p* value <0.01 and a rho> 0.7 were included in the networks.

The *mvabund* package (Wang *et al.*, 2012) was used to conduct multivariate analysis on abundance data. The *mvabund* package uses generalized linear models fit to each response variable (abundance data) to determine statistical significance of groups of predictor variables (Warton et al., 2012). The method uses a sum of likelihood ratios of each model as a test statistic and 999 permutations of the data to determine the significance of each model term. This method was used to test the significance of the discrete environmental groupings of season, location, and position in relation to the MLD (hereafter just MLD), in influencing microbial community structure. Sequence counts were used to fit generalized linear models with negative binomial distributions, as is common practice for count data (O'Hara and Kotze, 2010). To determine which taxa contributed the most to the differences between the predictor variables, univariate test statistics with adjusted *p* values for multiple testing, using a step down resampling procedure, were calculated for each response variable for each application of *mvabund*.

In order to test the hypothesis that the factors of season, MLD, and location significantly influenced microbial community structure an ANOVA table was constructed using the output of *manyglm* with the function *anova.manyglm*. The model used for this purpose was:

Community ~ Season + MLD + Location

The *mvabund* package was used again to test whether a station in the vicinity of the Thebaud gas field was significantly different than surrounding shelf stations, and if significance was found, to identify the specific taxa that were more abundant in the gas field associated site to those that were in the surrounding areas. The model used for this purpose is below and the function *anova.manyglm* was used to create an ANOVA table of the results.

Community ~ Station.Type

Where Station.Type refers to whether or not the station was associated with the Thebaud gas field.

2.3 Results

2.3.1 Environmental Parameters

During the spring leg of the CCGS *Hudson* cruise in 2014, 6 transects including the Browns Bank Line (BBL), the LaHave Basin Line (LHB), the Halifax Line (HL), the Louisburg Line (LL), the St. Anne's Bank Line (STAB), and the Cabot Strait Line (CSL) were extensively surveyed. In addition, the station GULD04 corresponding to the ecologically rich Gully region, and the Thebaud platform (TB) station (subject of analysis in section 2.3.5) corresponding to a station in close proximity to the Thebaud oil and gas platform were sampled. During the fall cruise in 2014, all of the above transects were sampled with the exception of the LHB transect and the Thebaud platform station (Figure 2.1). Variations in cruise routes led to slight differences in stations sampled in the fall and spring cruises. In total, 42 stations were sampled at 4 depths, resulting in 168 samples from spring (92) and fall (76) cruises with 64 sites shared between the two cruises.

Physical and chemical measurements were conducted by BIO and a thorough overview of the physical and chemical data collected for each station sampled in this study can be found in Supplementary Table S1. As expected and previously reviewed (Townsend *et al.*, 2005), temperatures and salinity increased with increasing distance from shore in both spring and fall cruises (Figure 2.2a-b,f-g). Maximum temperature (13.63°C) in the spring was observed at LHB_6.7, and minimum spring temperature (-1.47°C) was observed at CSL_4. In the fall maximum temperature (23.69°C) was observed at HL_11 and minimum temperature (0.97°C) was again observed at CSL_4. In general, surface temperatures during the fall cruise were much warmer than those measured during the spring cruise, while temperatures at depth generally fell within the same range (5°C-15°C) in both the spring and fall cruises. The majority of transects in both spring and fall cruises exhibited relatively high oxygen concentration throughout the shelf region and developed a layer of comparatively low oxygen concentration situated between 200 and 300 m once off the shelf and into waters of greater depth (Figure 2.3). The average O_2 concentration in the fall was lower (5.4 ml/L; ~231 μmol/kg O_2) than it was in the spring (7.0 ml/L; ~300 μmol/kg O_2) undoubtedly due to the higher temperatures and lower productivity observed.

The maximum chlorophyll concentration measured in the spring of 17.49 mg/m^3 (STAB_01, 20 m) was much higher than the maximum value of 2.13 mg/m^3 (CSL_01, 1 m) measured in the fall. The higher chlorophyll concentrations in the spring (average concentration of 3.97 mg/m^3) correspond with the spring bloom which is dominated by diatoms on the Scotian Shelf (Dasilva et al., 2014). This contrasts with the early fall phytoplankton community (average concentration of 0.39 mg/m^3), which is known to be comprised of picophytoplankton in the North Atlantic (Li and Dickie, 2001; Craig *et al.*, 2014). Bacterial concentrations were much higher in the fall than in the spring with an average of 1,257,148 cells/mL and 700,264 cells/mL respectively (Supplementary Figure S1). There was also about twice as many bacterial cells on average above the MLD than below, with 1,223,901 cells/mL and 660,347 cells/mL respectively. In addition, concentrations of ammonium, nitrite, phosphate, and silicate were all higher on average in the spring as compared to the fall season.

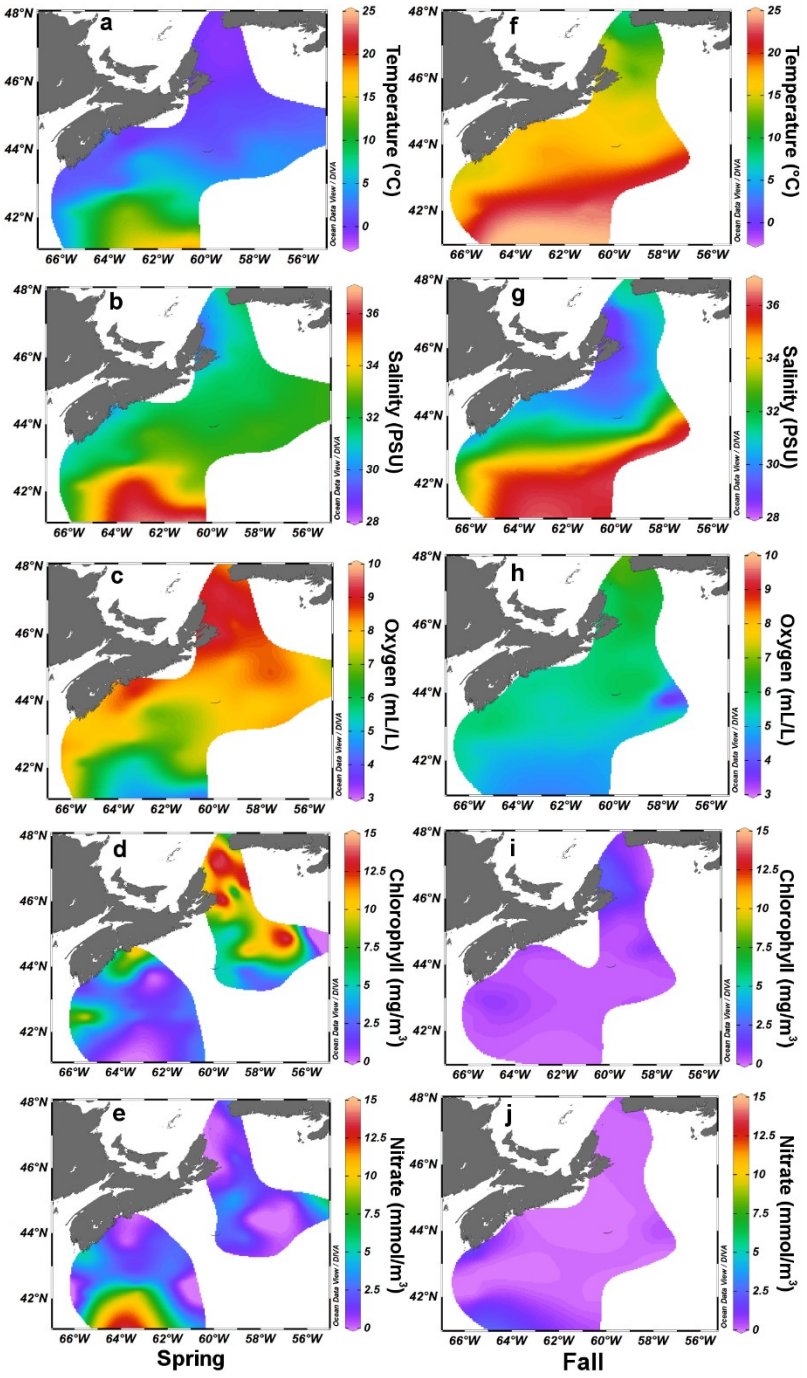


Figure 2.2. Surface plots of data from spring 2014 (a-e) and fall 2014 (f-j) cruises. Environmental variables displayed include: Temperature (a,f), Salinity (b,g), Oxygen concentration in mL/L (c,h), Chlorophyll concentration in mg/m^3 (d,i), and Nitrate concentrations in mmol/m^3 (e,j). Slightly different cruise tracks resulted in slightly different coverage of the Scotian Shelf between spring and fall cruises.

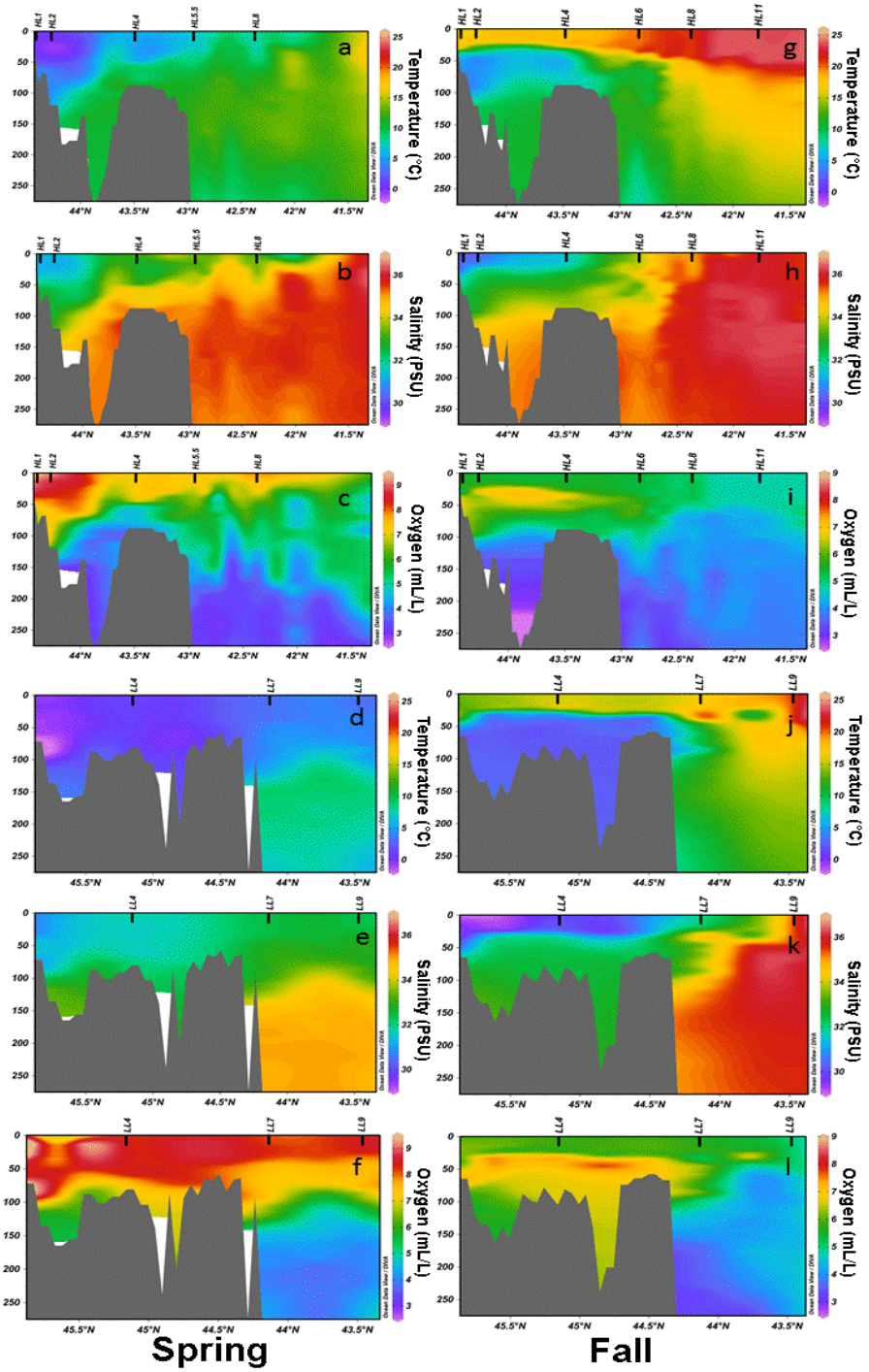


Figure 2.3. Section plots displaying CTD data from Halifax Line (a-c, g-i) and Louisburg Line (d-f, j-l) transects. a-f contain data from the spring cruise, g-l contain data from the fall cruise. Depth in the water column is along the y axis and latitude along the x axis. a, d, g, j show temperature data from CTD casts, b,h,e,k display salinity data from CTD casts, and c,i,f,l, display oxygen concentration (mL/L) data from CTD casts. Stations sampled in this study are labelled above the figure.

2.3.2 Community Composition and Alpha Diversity

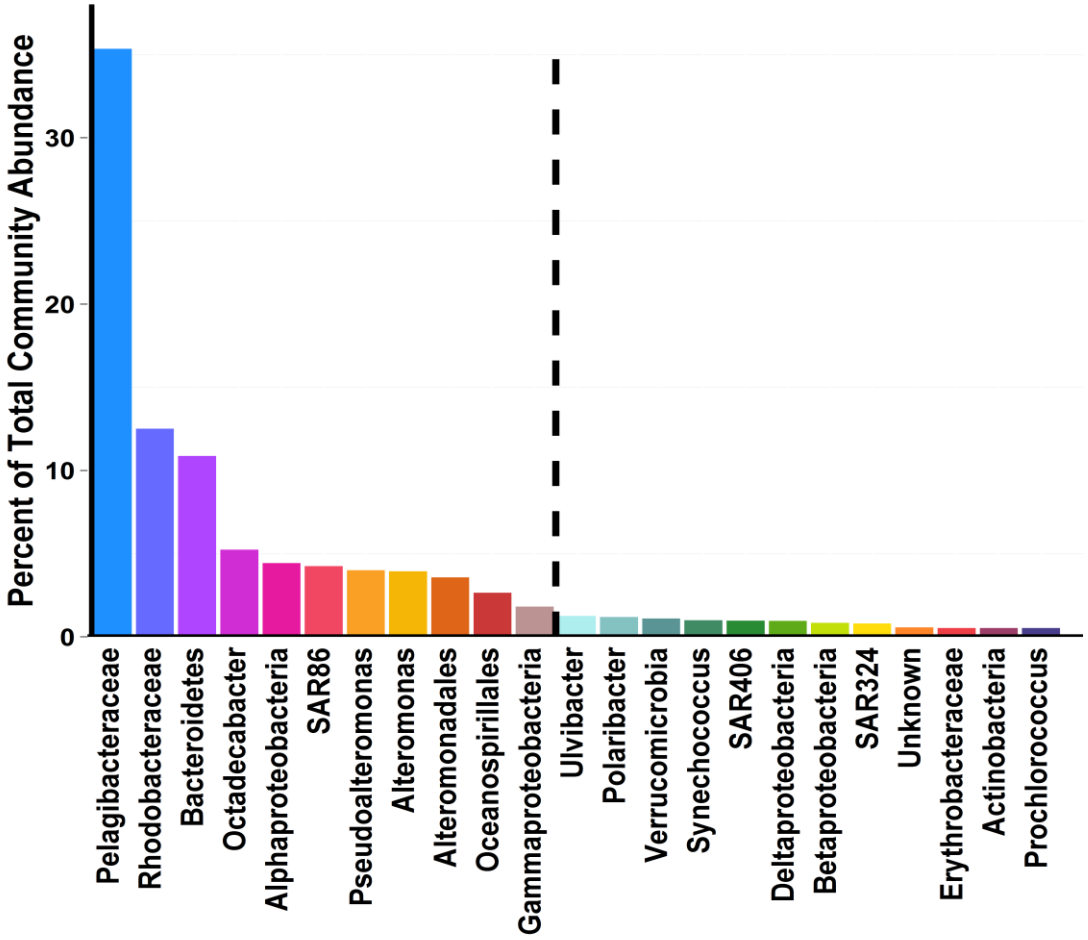


Figure 2.4. Bar chart of percent of total community abundance represented by the taxa that make up 99% of the relative abundance of all sites sampled from both seasons. The dashed line delineates taxa that make up 90% of the relative abundance of the entire dataset.

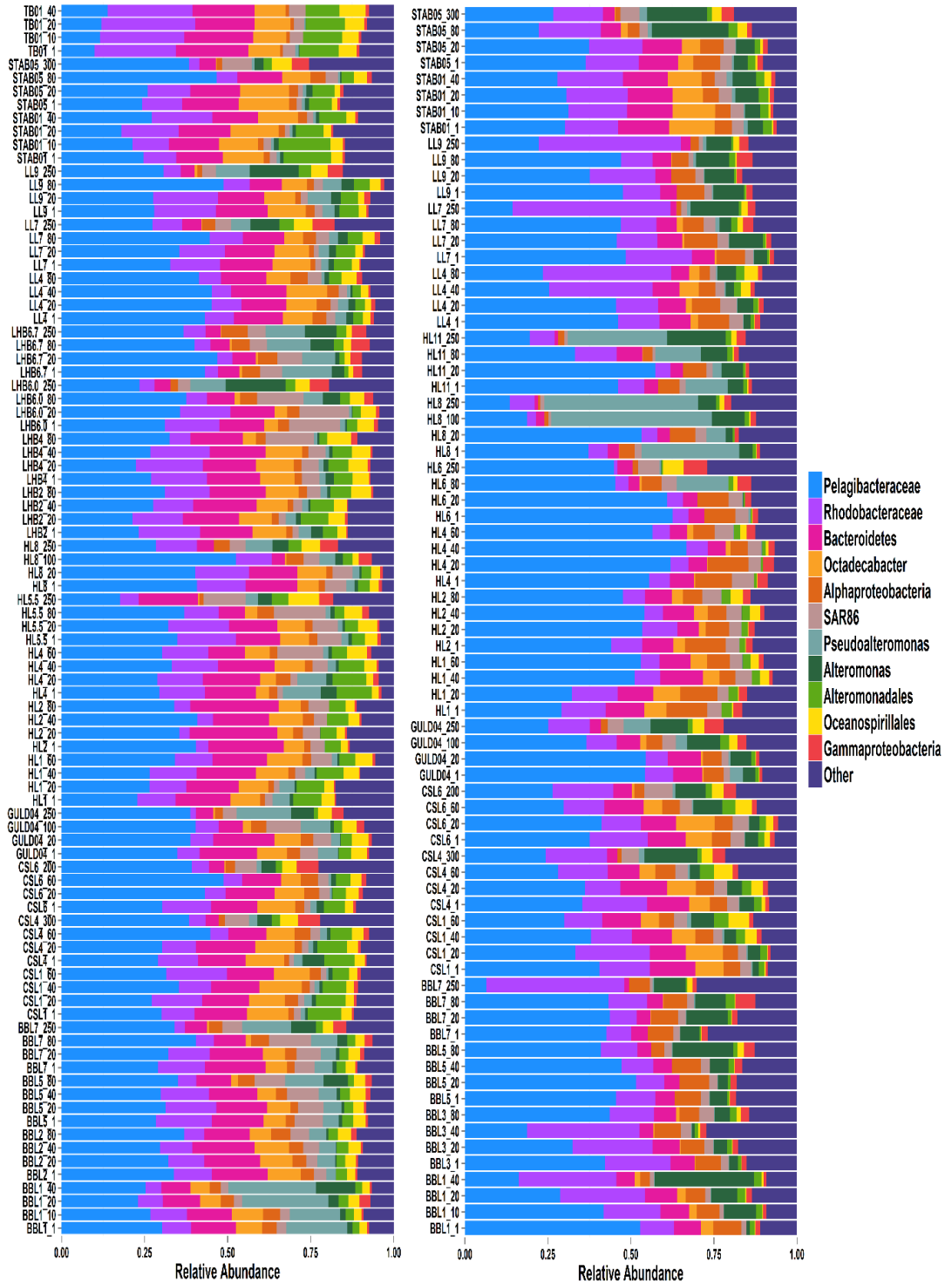


Figure 2.5. Stacked bar charts displaying the distribution of the most abundant taxa that represent on average, 90% of the relative abundance across all sites sampled, all remaining taxa are identified as "other" in this figure. The names of the sites are along the y axis and denoted by transect name, followed by station number, followed by depth in the water column. Spring stations are on the left, fall stations are on the right.

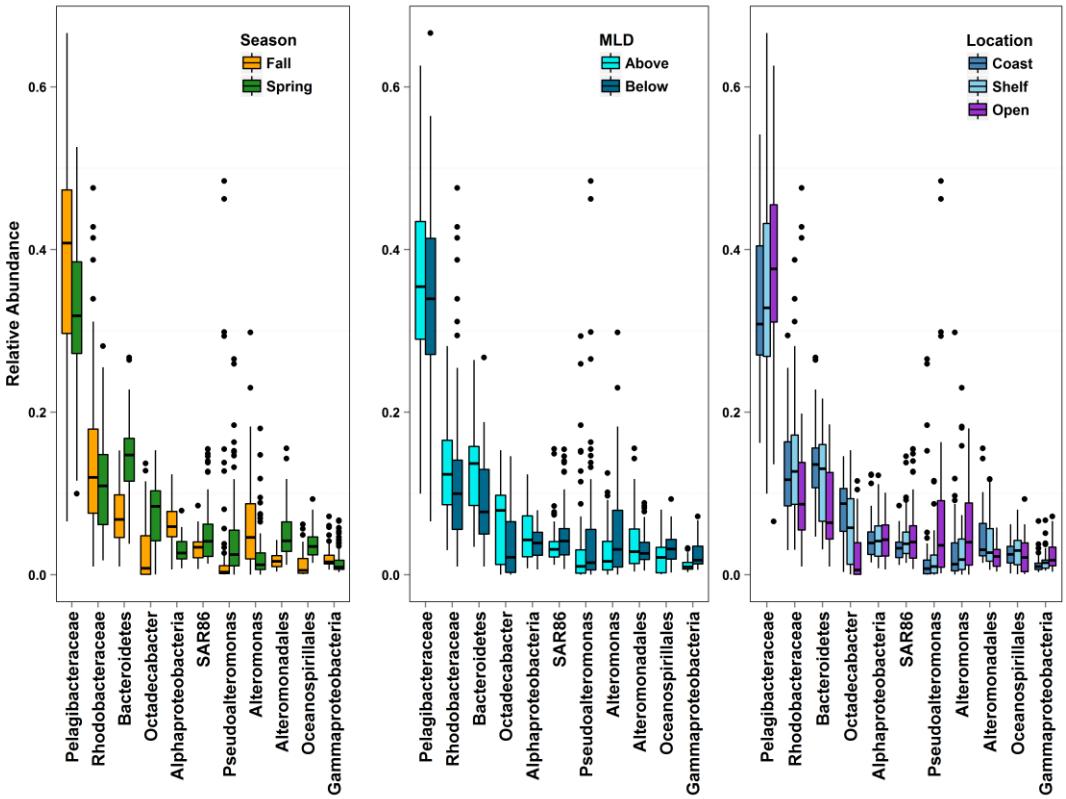


Figure 2.6. Boxplots showing the distribution of relative abundance of the ten most abundant bacterial taxa that make up on average 90% of total community relative abundance split into seasonal, depth, and location groups.

Before rarefaction of the raw data there were 9,953,481 sequence reads and 27,428 OTUs identified (Supplementary Table S3). The raw data however had large variations in the number of sequences per sample, a usual outcome of modern sequencing technologies, which leads to confounding and non-comparable alpha and beta diversity metrics (Gotelli and Colwell, 2001). Rarefaction of all samples to 7500 sequences was done to allow for comparisons of alpha and beta diversity between sites. After rarefaction, there were 1,260,000 total sequence reads and 16,410 OTUs identified.

Overall only 11 clades of bacteria were responsible for 90% of the relative abundance across stations (Figure 2.5). The most abundant group, Pelagibacteraceae, accounted for on average 35.3% of community abundance at every site. Locally however it ranged from 6.5% of community abundance at fall station BBL7, 250 m depth, to 66.6% at fall station HL4, 40 m. The next most abundant groups, Rhodobacteraceae and

Bacteroidetes, also exhibited wide ranges in abundance across sites. Rhodobacteraceae accounted for on average 12.5% of community abundance, but ranged from 1% at fall station HL6, 250 m, to 48% at fall station LL7, 250 m. Bacteroidetes ranged from 1% at Fall HL11, 250 m depth to 27% at spring, HL2 80 m and accounted for 11% on average. At the OTU level, the five most abundant OTUs were from Pelagibacteraceae (OTU 637092), Pelagibacteraceae (OTU 307744), Rhodobacteraceae (OTU 569286), *Pseudoalteromonas* (OTU 827726), and *Octadecabacter* (OTU 804483) and accounted for on average 13%, 6%, 4%, 3%, and 2% of all sequences respectively. OTU 569286 corresponding to the family of Rhodobacteraceae had the highest abundance at any one site, accounting for 44% of community abundance at Fall LL7 250 m. As the Rhodobacteraceae family was at its highest relative abundance contribution of 48% at this site, this one OTU was responsible for the majority of this peak.

In order of greatest relative abundance, the ubiquitous bacterial taxa that were present to some measureable extent at every station and depth sampled in the Scotian Shelf were Pelagibacteraceae, Rhodobacteraceae, Bacteroidetes, *Octadecabacter*, Alphaproteobacteria, SAR86, Alteromonadales, Oceanospirillales, Gammaproteobacteria, Verrucomicrobia, SAR406 (marine group A), Deltaproteobacteria, Betaproteobacteria, unclassified OTUs, and Actinobacteria. Although these clades can be considered "cosmopolitan" in their absolute distribution, there were substantial differences in the relative abundance of these taxa at various sites (Figure 2.6) suggesting distinct, taxa-specific, biogeographical patterns. Conversely, when looking at the level of OTU, only 5 OTUs (637092, 307744, 556042, 311349, 317958) were present in every sample, all of which belonged to Pelagibacteraceae.

Of additional interest are those taxa and OTUs that were generally rare across the shelf but at select stations would become dominant in the community. The OTUs that had the highest maximum abundance as compared to their average abundance across the shelf were Vibrionaceae OTU 230384 (average = 0.08%, max = 6.9%, Fall, HL8, 250 m), *Pseudoalteromonas* OTU 145002 (average = 0.19%, max = 13.7%, Fall, HL8 100 m), and *Nocardioides* OTU 257075 (average = 0.08%, max = 5.34%, Fall BBL7, 250 m).

While 16S sequencing of the V6-V8 region has the potential to return members of the microbial community to a specific level like genus, many of the OTUs returned by

our analysis could only be identified to higher taxonomic levels. For example, bacterial OTUs that could only be classified to the level of Alphaproteobacteria were responsible for 2.6% of sequences per sample on average, and 0.6% of sequences on average in each sample could not be classified to any taxonomic level at all. These poorly classified OTUs demonstrate that there are still many bacterial genera, as well as orders, families, classes and even phyla that have yet to be discovered and characterized.

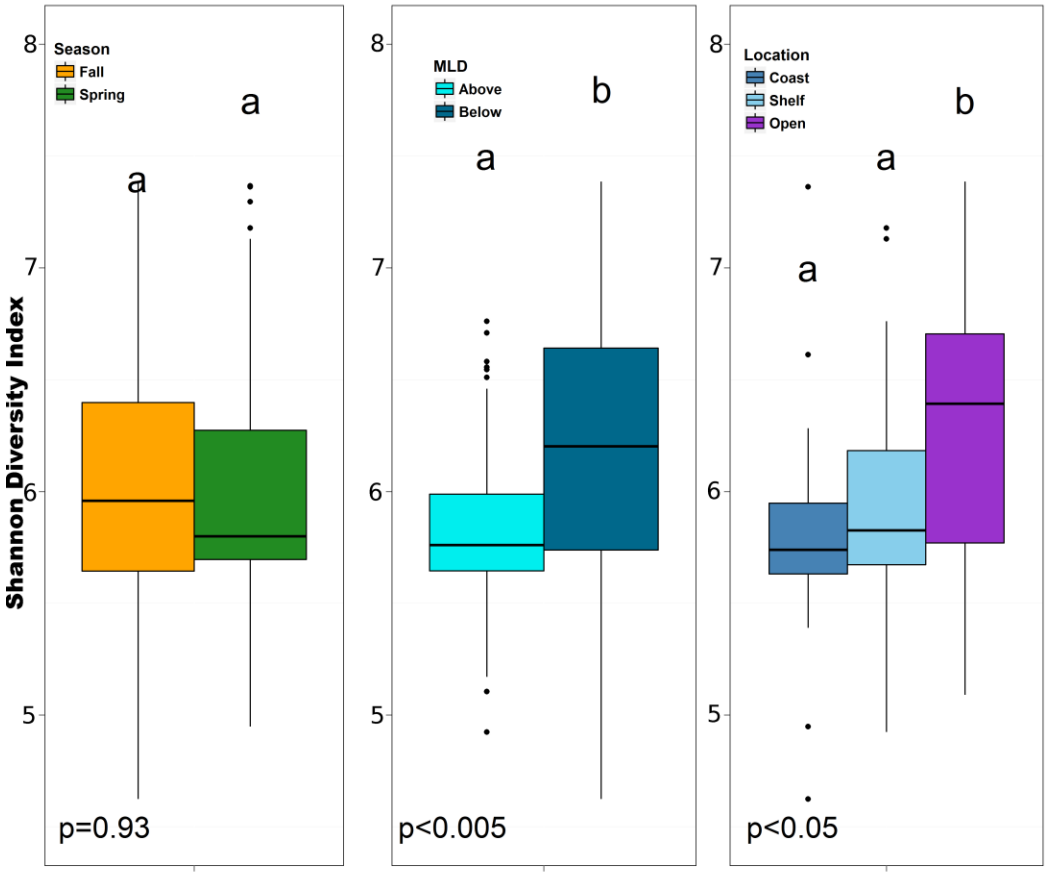


Figure 2.7. Boxplots of distributions of Shannon diversity indices of stations divided into different environmental groups. The season grouping showed no significant difference between fall and spring (Mann-Whitney U test, W = 3466, $p = 0.93$), there was a significant difference between above and below the mixed layer (Mann-Whitney U test, W = 2214, $p < 0.005$), Location groups were found to be significantly different (Kruskal-Wallis, KW Chi squared = 17.996, df = 2, $p < 0.001$), and a post-hoc Dunn test was carried out to determine statistical differences between location groups. No significant differences between coast and shelf were found ($p = 0.19$), but both coast and shelf were significantly different than the open ocean samples ($p < 0.01$ for both).

Figure 2.7 shows the distribution of the alpha diversity metric, Shannon diversity, across sites divided into their main environmental groupings. The significance of each difference in distribution was calculated using the non-parametric tests, Mann-Whitney (2 groups) and Kruskal-Wallis (3 groups). The diversity of the communities is generally greater in communities found below the mixed layer than above the mixed layer, and in communities found in the open ocean region compared to coastal and shelf regions. There was no significant difference in the distributions of alpha diversity based on the seasonal grouping. Additionally, the samples from above the mixed layer have a smaller range of alpha diversity values compared to the wide range of alpha diversity values exhibited by the samples collected from below the mixed layer. The coastal and shelf regions have relatively similar composition of abundant taxa as shown by Figure 2.6, and so it is unsurprising to see similar alpha diversity distributions between the two groups here. Similar to the MLD grouping, the location grouping shows that the samples belonging to the open ocean exhibit a broader range in alpha diversity than those samples collected from the shelf and coastal areas.

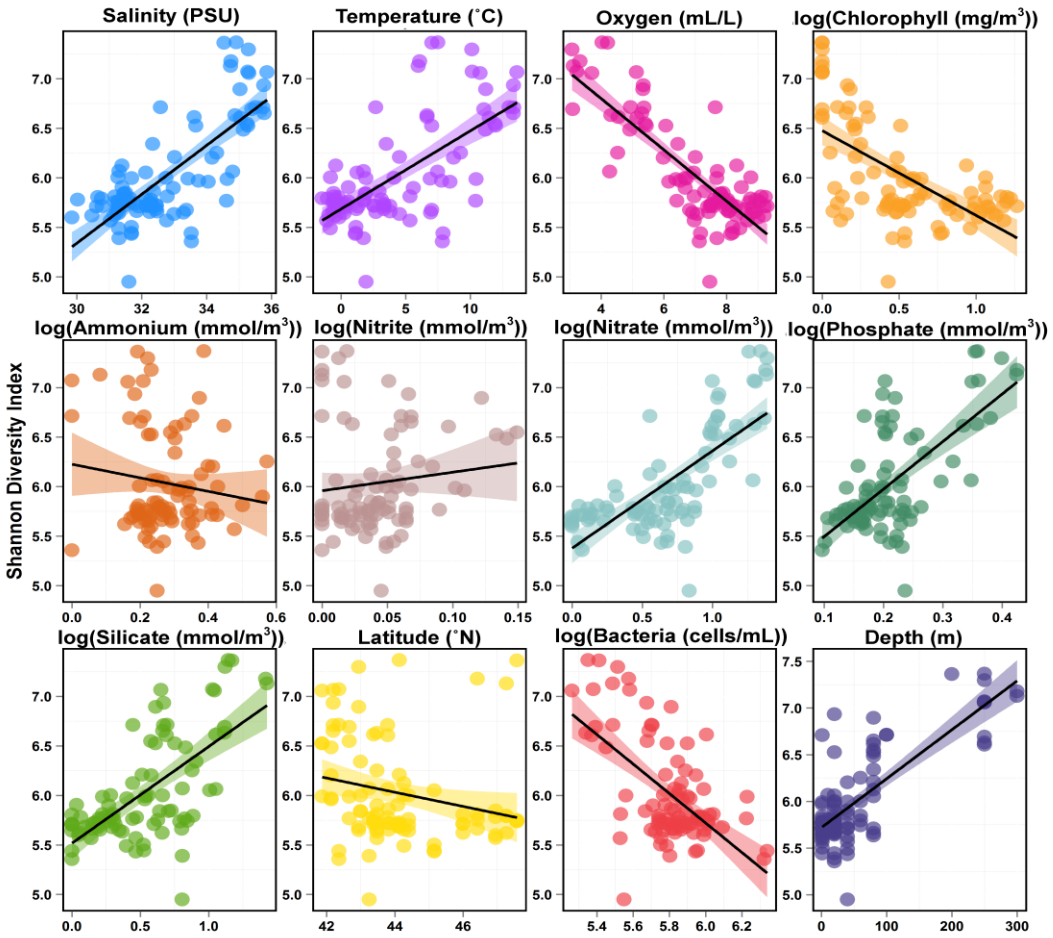


Figure 2.8. Linear regressions of the Shannon alpha diversity metric against various environmental variables using spring samples. The shaded area of each plot represents the 95% confidence interval for the regression. Chlorophyll, ammonium, nitrite, nitrate, phosphate, silicate, and bacteria were all transformed by log+1 to establish a more normal distribution of data. Table 2.2 reports significance, adjusted R^2 and slope for each regression.

Table 2.2. Parameters from the linear regressions of environmental variables and the Shannon diversity metric for stations grouped by season. Environmental variables in bold showed a significant ($p < 0.05$) linear relationship with alpha diversity.

	Spring			Fall		
Variable	**Slope**	***p* value**	**Adjusted R^2**	**Slope**	***p* value**	**Adjusted R^2**
Salinity	0.25	**2e-16**	0.57	0.11	**0.0005**	0.14
Temperature	0.79	**5e-13**	0.44	0.014	0.2	0.0093
Oxygen	-0.26	**2e-16**	0.68	-0.12	0.06	0.033
Log Chlorophyll	-0.86	**3e-11**	0.38	-0.85	0.1	0.017
Log Ammonium	-0.68	0.2	0.0066	-0.24	0.7	-0.012
Log Nitrite	1.9	0.3	0.002	1.4	0.6	-0.0095
Log Nitrate	0.98	**3e-16**	0.54	0.12	0.4	-0.0025
Log Phosphate	4.8	**2e-12**	0.43	-0.0095	1	-0.014
Log Silicate	0.98	**7e-12**	0.42	0.20	0.3	0.0012
Log Bacteria	-1.5	**4e-10**	0.34	-0.069	0.8	-0.012
Latitude	-0.072	**0.03**	0.041	-0.055	0.1	0.017
Depth	0.0052	**2e-16**	0.61	0.00035	0.7	-0.011

In spring samples, most of the environmental parameters measured displayed a significant relationship with alpha diversity at the site (Figure 2.8, Table 2.2) however this pattern was not evident in fall samples (Supplementary Figure S2, Table 2.2). This demonstrates that alpha diversity, which encompasses species richness and species evenness, is driven by different factors seasonally. In the spring, species richness increases significantly with temperature, salinity, depth, and nutrients (nitrate, phosphate and silicate) and decreases significantly with oxygen, chlorophyll, bacteria, and latitude. Fall samples only show a significant positive relationship between salinity and alpha diversity although it is much less pronounced than the same relationship in the spring. Other than salinity, no environmental variables exhibited a significant relationship with alpha diversity in the fall.

The ten sites with the highest diversity were all found in the spring, with the exception of HL6 250 m from the fall, which had the highest alpha diversity of any site. Additionally, the nine most diverse sites were all from 200 m and below, agreeing with the highly significant positive relationship found between depth and alpha diversity in the spring. Sites that were diverse in the spring, were not necessarily as diverse in the fall. The most striking example of this was station LL7 at 250 m depth which was one of the most diverse stations in the spring with a Shannon diversity measure of 7.37 but was the least diverse station in the entire dataset in the fall, with a Shannon diversity index of

only 4.37. This further illustrates that diversity and the factors that drive it are strongly influenced by season.

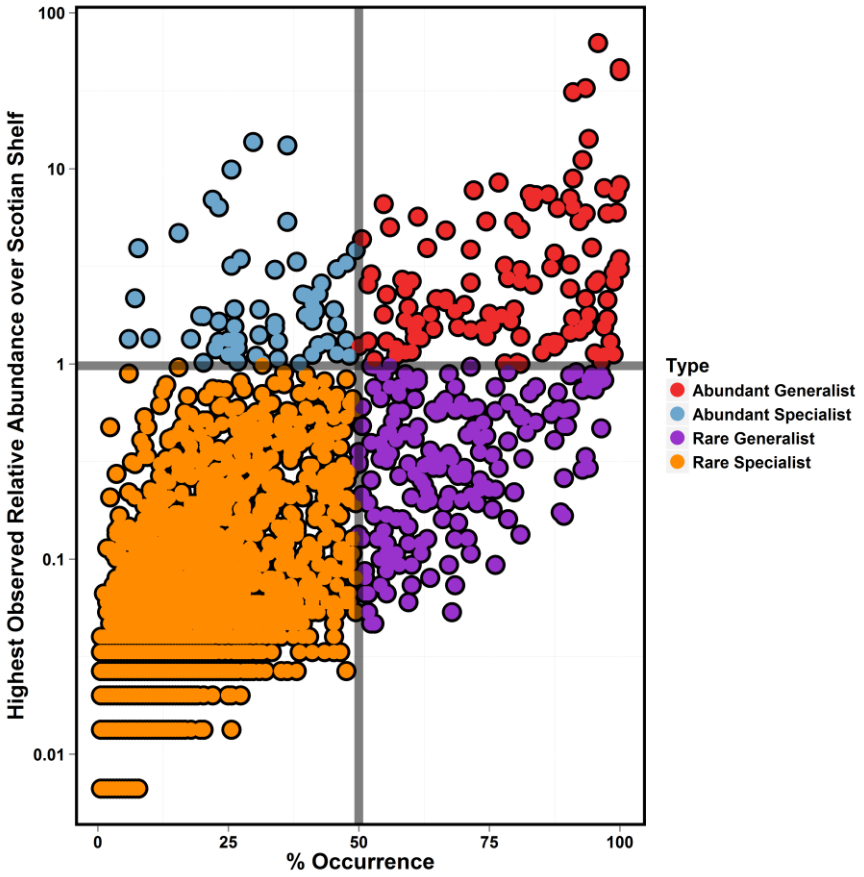


Figure 2.9. OTUs plotted with percent of sites present on the x-axis, and highest observed relative abundance in dataset on the y-axis (log scale). In this dataset, a Generalist is defined as an OTU that is present in at least 50% of all samples, whereas a Specialist is defined as an OTU that is present in less than 50% of all samples. An OTU is abundant when it reaches at least 1% of the relative community abundance in at least one sample in the dataset. Rare OTUs do not meet this criterion.

The terms used in Figure 2.9 were loosely based off of classifications described by Barberan et al. (2012). To classify a specialist, Barberan et al. (2012) used the criterion of occurrence in less than 6% (10 of 151 samples) of samples and a relative abundance of greater than 2%. Generalists included all OTUs that were present in 53% (80 of 151 samples) of samples, regardless of abundance (Barberán *et al.*, 2012). The terms specialist and generalist were kept in this analysis but the criteria were expanded in order to include all OTUs in the dataset and allow for percentwise reproducibility with other datasets. Also abundant and rare were added as descriptors for the maximum

abundance of OTUs in the dataset. 1% was used as a cut-off for this criterion as it has previously been applied as a generic limit in other studies both as a technological restraint of some older sequencing techniques and as a limit for inclusion in statistical analyses (Forth *et al.*, 2014; Lynch and Neufeld, 2015)

With these classifications, the vast majority of OTUs in this dataset (16058 OTUs - 97.85%), were assigned to Rare Specialists occurring in less than 50% of the samples and never reaching quantitative dominance in the community. Rare Generalists accounted for 178 OTUs (1.08%), Abundant Generalists represented 114 OTUs (0.69%) and with 60 OTUs (0.37%) Abundant Specialists were the smallest group. The distribution of OTUs into these categories demonstrates the diversity and numerical dominance of the species associated with the "rare biosphere".

2.3.3 Beta Diversity, Biogeography and Patterns in Spatial Distribution

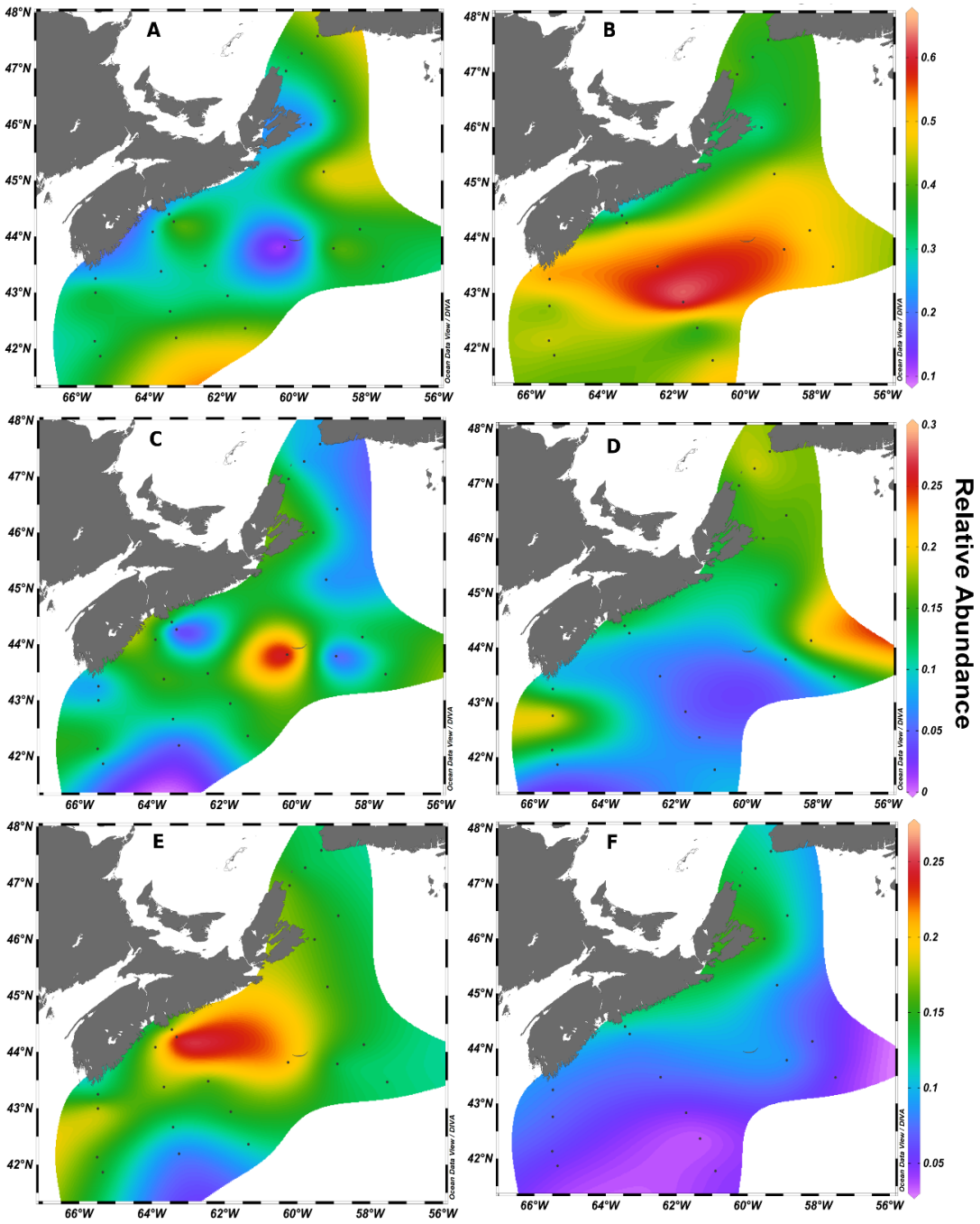


Figure 2.10. Surface plots from the 1 m depth, of taxa over the Scotian Shelf in spring (A,C,E) and fall (B,D,F). Pelagibacteraceae (A,B), Rhodobacteraceae (C,D), and Bacteroidetes (E,F) groups are displayed. Note that colour scales are fixed between seasons of the same taxa but vary between taxa. Black dots mark surface sites.

Figure 2.10 shows the distribution of the three most abundant taxa groups across the surface of the Scotian Shelf. There are striking differences in all groups between seasons. Pelagibacteraceae for instance is relatively more abundant in the fall than in the spring, whereas Rhodobacteraceae and Bacteroidetes are both relatively more abundant in the spring samples. Additionally, Pelagibacteraceae and Rhodobacteraceae appear to have roughly reversed distributions between their spring and fall samples. In the spring, Pelagibacteraceae is less abundant on the shelf area and increases its relative abundance in the northeast and southwest corners of the study sites. In the fall however, the group is relatively more concentrated on the shelf than in the outlying areas. The opposite trend is true for Rhodobacteraceae, which is more abundant on the shelf region in the spring and in the outlying areas in the fall. Generally, at the surface, particularly in the spring it appears that Rhodobacteraceae and Pelagibacteraceae exhibit complementary spatial distribution patterns.

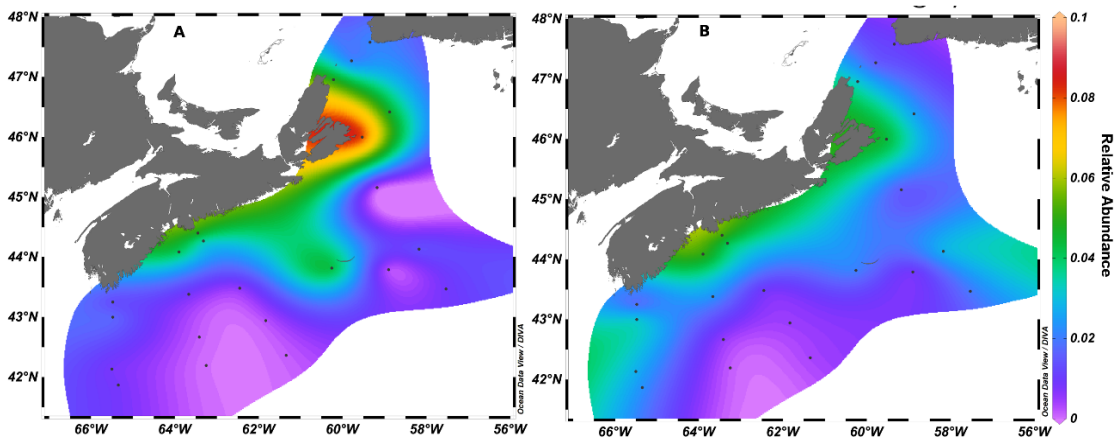


Figure 2.11. *Polaribacter* (A) and *Ulvibacter* (B) distributions in relative abundance across the surface of the Scotian Shelf in spring. Black dots mark surface sites.

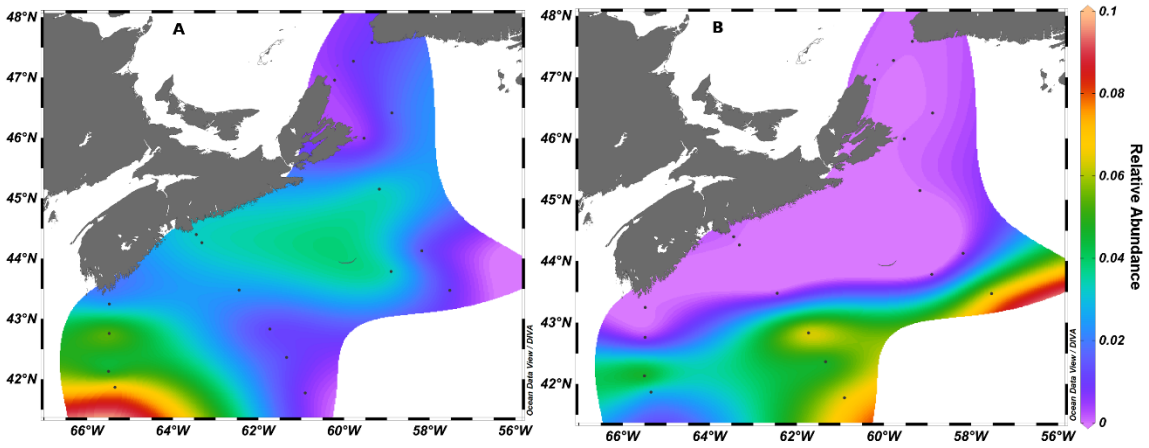


Figure 2.12. *Synechococcus* (A) and *Prochlorococcus* (B) distributions in relative abundance across the surface of the Scotian Shelf in fall. Black dots mark surface sites.

Figures 2.11 and 2.12 show two genera that are characteristically abundant in either the spring (Figure 2.11) or the fall (Figure 2.12). Both *Polaribacter* and *Ulvibacter* are members of the Bacteroidetes phylum and although both groups are distinctly more abundant near the coast in the spring, they have decidedly different distributions over the shelf. *Polaribacter* is more abundant mid shelf near Sable Island, whereas *Ulvibacter* is relatively more abundant in the southwest shelf region and off the western slope. *Synechococcus* and *Prochlorococcus* are both oligotrophic specialists that are more common in this dataset in the low nutrient fall conditions. Even though these groups are closely related, there is evident differences in their spatial distribution. In the fall, *Prochlorococcus* is essentially absent on the shelf region and only gains abundance off the slope in the open ocean. *Synechococcus* conversely is present to some extent on the shelf.

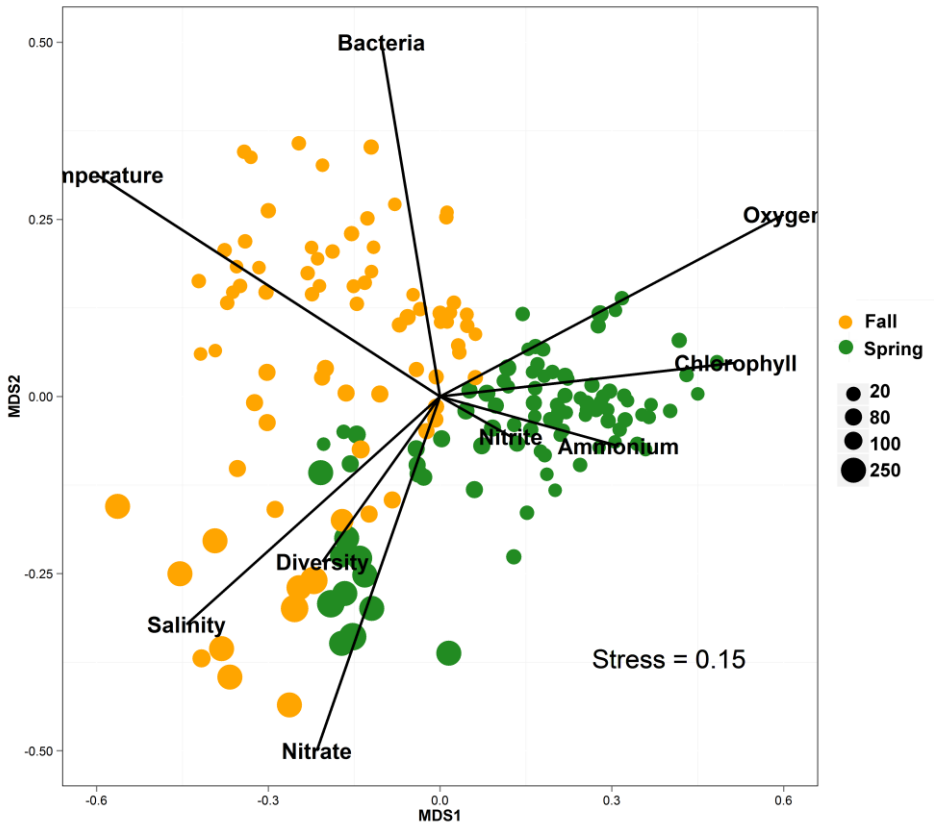


Figure 2.13. NMDS plot of Hellinger transformed abundance data from all sites using the Bray-Curtis dissimilarity matrix. Sites are coloured based on season and sized relative to depth in the water column. The top most abundant taxa that make up on average 99% of the relative abundance of the entire dataset were used for this analysis. Environmental vectors were fit onto the ordination using the function *envfit*.

Figure 2.13 shows the NMDS ordination of all sites in the dataset where the closeness of the sites in the ordination reflects their community similarity. In this analysis no separated clusters are formed but a gradient of varying community similarity across sites is evident. Sites coloured based on season show that there is a strong distinction between fall and spring samples. This distinction is slightly more apparent with surface samples than it is with deep water samples. Depth of the samples in the water column also appears to distinguish communities from one another. The spring surface communities are situated in closer proximity to each other than the fall surface communities which cover more area of the ordination, suggesting that fall samples have larger variance in community structure between sites than spring samples. Additionally, vectors were fit onto the ordination to summarize the environmental trends over the

sampling sites. From this analysis temperature and high bacteria concentrations were associated with fall sites, whereas oxygen, chlorophyll, nitrite, and ammonium were associated with spring sites. The samples from deeper sites were associated with high concentrations of nitrate, high salinity and high bacterial diversity.

Table 2.3. Partial Mantel tests using Spearman's correlation between matrices of all environmental variables, geographic distances, and individual environmental variable gradients with the matrix of community similarity performed on all sites collectively. Individual variables arranged in descending order of greatest correlation with abundance

Distance Matrix	**Partial Mantel (Spearman's)[a]**	***p* value***
Environment (all)	**0.5954**	**0.0001**
Geography (all)	-0.0231	0.8
Oxygen	**0.5846**	**0.0001**
Temperature	**0.4949**	**0.0001**
Phosphate	**0.4513**	**0.0001**
Diversity	**0.3868**	**0.0001**
Nitrate	**0.3616**	**0.0001**
Salinity	**0.3367**	**0.0001**
Bacteria	**0.2673**	**0.0001**
Silicate	**0.25**	**0.0001**
Chlorophyll	**0.175**	**0.0001**
Ammonium	**0.06067**	**0.04**
Longitude	-0.00111	0.5
Latitude	-0.01745	0.7
Nitrite	-0.02371	0.8

*significance $p<0.05$

[a] Partial Mantel tests: 10000 permutations testing for abundance matrix correlation against either environmental or geographic distance matrix, while controlling for the alternate matrix

Table 2.4. Partial Mantel Tests using Spearman's correlation between matrices of all environmental variables, geographic distances, and individual environmental variable gradients, with the matrix of community similarity split into season and depth groups. Spearman's correlation values are displayed with corresponding *p* value in brackets. Significant correlations ($p < 0.05$) are in bold.

	Partial Mantel (Spearman's)[a]				
Spring	***1 m***	***20 m***	***80 m***	***250 m***	***All Depths***
Environment(all)	**0.53(0.0001)**	**0.43(0.0009)**	**0.77(0.0001)**	**0.40(0.04)**	**0.68(0.0001)**
Geography(all)	-0.077(0.8)	-0.038(0.7)	0.18(0.1)	0.09(0.3)	0.016(0.3)
Oxygen	**0.50(0.0001)**	**0.37(0.0009)**	**0.63(0.0001)**	0.22(0.09)	**0.64(0.0001)**
Temperature	**0.46(0.0001)**	**0.45(0.0001)**	**0.57(0.0001)**	0.24(0.06)	**0.45(0.0001)**
Phosphate	0.064(0.3)	0.11(0.2)	0.031(0.4)	**0.50(0.005)**	**0.43(0.0001)**
Diversity	0.21(0.07)	**0.39(0.004)**	**0.32(0.01)**	0.024(0.4)	**0.69(0.0001)**
Nitrate	**0.50(0.0001)**	**0.37(0.0009)**	**0.50(0.0004)**	**0.37(0.02)**	**0.68(0.0001)**
Salinity	**0.40(0.001)**	**0.38(0.0004)**	**0.65(0.0001)**	0.37(0.06)	**0.55(0.0001)**
Bacteria	0.19(0.09)	0.18(0.1)	**0.27(0.05)**	**0.45(0.005)**	**0.29(0.0001)**
Silicate	0.19(0.07)	**0.23(0.05)**	0.053(0.3)	**0.34(0.05)**	**0.52(0.0001)**
Chlorophyll	**0.30(0.005)**	0.17(0.09)	**0.45(0.004)**	NA	**0.18(0.001)**
Ammonium	-0.025(0.5)	0.010(0.4)	0.21(0.06)	-0.36(0.9)	0.03(0.3)
Longitude	-0.10(0.8)	-0.082(0.8)	-0.36(0.9)	-0.18(0.8)	-0.086(0.9)
Latitude	0.11(0.2)	0.10(0.2)	0.18(0.1)	0.26(0.1)	**0.08(0.05)**
Nitrite	**0.32(0.0006)**	**0.25(0.03)**	**0.28(0.03)**	0.32(0.07)	**0.23(0.0001)**
Fall	***1 m***	***20 m***	***80 m***	***250 m***	***All Depths***
Environment(all)	**0.25(0.02)**	**0.38(0.0001)**	**0.48(0.002)**	-0.14(0.7)	**0.53(0.0001)**
Geography(all)	0.14(0.06)	**0.22(0.01)**	-0.13(0.8)	0.29(0.07)	**0.11(0.0005)**
Oxygen	**0.29(0.007)**	**0.32(0.0003)**	**0.44(0.008)**	-0.27(0.9)	**0.53(0.0001)**
Temperature	**0.38(0.0006)**	**0.42(0.0001)**	**0.60(0.0001)**	0.028(0.4)	**0.28(0.0001)**
Phosphate	**0.37(0.0008)**	**0.50(0.0001)**	**0.48(0.004)**	-0.13(0.7)	**0.56(0.0001)**
Diversity	**0.33(0.001)**	**0.36(0.0007)**	**0.47(0.01)**	**0.56(0.001)**	**0.25(0.0001)**
Nitrate	-0.18(0.9)	0.015(0.4)	0.16(0.1)	-0.064(0.6)	**0.58(0.0001)**
Salinity	**0.66(0.0001)**	**0.57(0.001)**	**0.53(0.0008)**	-0.22(0.8)	**0.36(0.0001)**
Bacteria	-0.23(0.9)	-0.075(0.8)	-0.09(0.6)	-0.27(0.9)	**0.30(0.0001)**
Silicate	-0.27(1)	-0.14(0.9)	**0.42(0.009)**	-0.27(0.9)	**0.44(0.0001)**
Chlorophyll	0.14(0.1)	**0.32(0.001)**	**0.46(0.001)**	NA	0.067(0.1)
Ammonium	0.06(0.3)	**0.18(0.04)**	-0.12(0.7)	-0.24(0.8)	-0.061(0.8)
Longitude	-0.23(0.9)	-0.28(0.9)	-0.24(0.9)	**0.50(0.007)**	-0.072(0.93)
Latitude	**0.28(0.005)**	**0.33(0.0001)**	**0.33(0.04)**	-0.28(0.9)	**0.11(0.008)**
Nitrite	0.18(0.09)	**0.18(0.03)**	0.03(0.4)	0.17(0.1)	-0.030(0.67)

[a] Partial Mantel tests: 10000 permutations testing for community similarity matrix correlation with either environmental or geographic distance matrix, while controlling for the alternate matrix

Partial Mantel tests were calculated between community dissimilarity matrices and either the entire environmental distance matrix, the geographic distance matrix, or distance matrices generated from individual variables. Partial Mantel tests were conducted to determine the significance of either environment or geography in shaping

community abundance while controlling for the effect of spatial autocorrelation between environmental parameters and geographic distance. This analysis was conducted on all sites (Table 2.3), all spring or all fall sites (Table 2.4), and sites divided into discrete season and depth categories (Table 2.4). Overall, the entire environmental matrix showed the strongest correlation with community similarity, while comparatively, geographic distance didn't have a significant correlation with community similarity. Of the specific environmental variables, oxygen concentration showed the highest correlation with community similarity over all sites, followed by temperature, and phosphate concentration. Longitude, latitude, nitrite, and ammonium were the only factors that did not show a highly significant correlation with community similarity over all sites.

To investigate the relationship of environment and geography with community similarity further, the sites were subdivided based on season and discrete depth categories. When analyzing all sites and depths belonging to spring, environmental distance showed a very high and significant correlation with community dissimilarity, while the correlation between geographic distance and community dissimilarity was insignificant when controlling for environmental variations. For the fall samples, both environmental distance and geographic distance showed significant relationships with community dissimilarity, although environmental distance had a much higher correlation. Gradients of nitrate, phosphate and oxygen concentration were all highly significantly correlated with community dissimilarity in the fall, while bacterial diversity, nitrate, oxygen and silicate were highly significantly correlated with community dissimilarity in the spring.

Subdividing the sites even further to discrete depth categories displayed a wide range of subtle trends and relationships that were hidden when all sites and depths were observed together. Community dissimilarity of all discrete categories were significantly correlated with environmental distance except for samples from fall 250 m. Contrastingly, only fall 20 m communities were significantly correlated with geographic distance when environmental distance was taken into account. The fall 250 m communities were particularly interesting because they were not significantly correlated with environment, geography, or any individual variables except for bacterial diversity and longitude. Even more intriguing was that this group was the only category to be

correlated with longitude at all. The samples from spring 250 m also had fairly unique environmental correlations when comparing with other depths from spring. Specifically, oxygen and temperature were highly significantly correlated with bacterial community similarity for spring samples from 1 m, 20 m, and 80 m, but not for 250 m. The strongest relationship observed at spring 250 m was with phosphate, a variable that was not correlated with community similarity at 1 m, 20 m, or 80 m spring samples. Overall, in addition to providing insights into potential drivers of community structure over the Scotian Shelf, this analysis shows that the factors associated with bacterial community structure vary depending on the season and depths under consideration.

2.3.4 Multivariate Analysis of Abundance Data and Correlation Analyses

As identified in section 2.3.3, environmental distance had a greater effect on biogeography in the region than geographic distance. Additionally, the season and depth at which a sample was taken had an effect on the factors that correlated significantly with community change within those groups. This section aims to further analyze the response of bacterial communities to environmental gradients and groupings (such as season, depth, and location on the shelf) by determining how different bacterial taxa and OTUs respond to environmental factors and to each other.

Multivariate analysis of abundance was conducted through use of the *mvabund* package. The model, Community ~ Season+MLD+Location, was used to determine which group effects significantly influenced community composition. Season refers to the season the sample was taken (spring or fall), MLD refers to whether the sample was taken above or below that station's mixed layer depth (not the depth of the mixed layer itself), and location refers to the location of the site along the shelf, i.e. coastal, shelf (mid-shelf), or open ocean. Univariate models specific to each taxon were generated to determine which taxa contributed most to the changes in community structure within a given model term.

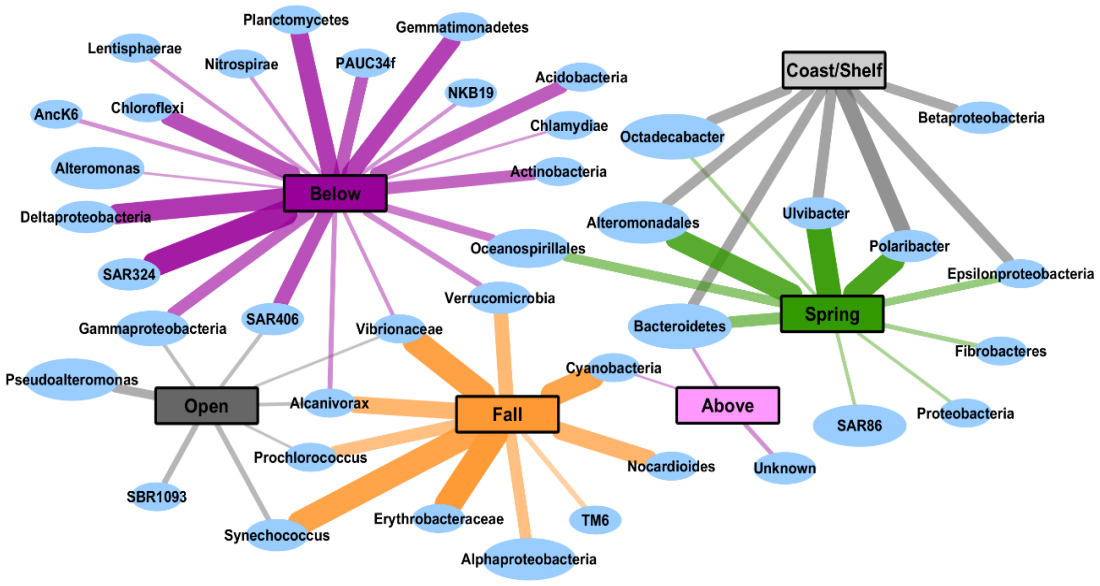


Figure 2.14. Results of a multivariate implementation of generalized linear models from the *mvabund* package in R performed to determine significance of environmental groupings. Season, MLD, and Location effects were all shown to be significant with respect to community composition. Taxa that had statistically significant (MGLM-ANOVA with p adjusted < 0.05) contributions to group effects are included in this network. The taxa that have greater contributions to specific group effects have thicker edges. Average abundance of the taxon across all sites is related to the size of the taxon's node. Colour of the edge corresponds to the group that the taxa was significantly associated with. Above and Below refer to above and below the mixed layer depth, and Open refers to the open ocean.

Season, MLD, and Location were all found to be significant with respect to driving changes in bacterial community composition on the Scotian Shelf (Supplementary Table S5). The results of this analysis are summarized in Table 2.6 and discussed in section 2.4.4.1, however it is clear from Figure 2.14 that there is a diverse and complex response at the level of bacterial taxa to seasonality, depth, and location on the Scotian Shelf. Overall, season was shown to contribute the strongest to community composition, followed by MLD, and then location, which had a weaker but still significant effect (Supplementary Table S5).

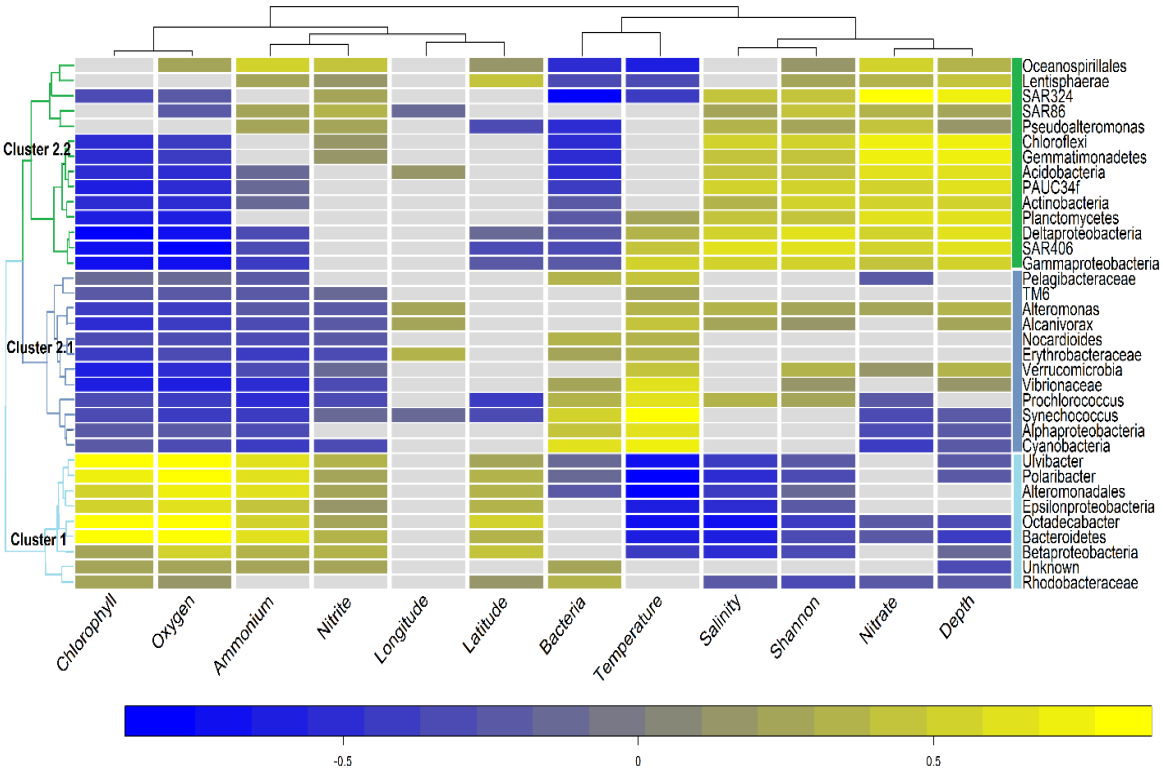


Figure 2.15. Heat map of correlation (Spearman's rank correlation) matrix between environmental variables and taxa. Only significant correlations (p value < 0.05) are coloured. Main clusters are labelled and their corresponding taxa are marked, light blue for cluster 1, dark blue for cluster 2.1, and green for cluster 2.2. The variable Bacteria refers to bacterial abundance, and the variable Shannon refers to bacterial diversity as measured by the Shannon diversity index.

Figure 2.15, displays the significant relationships between various environmental variables and taxa. The taxa and environmental variables are also grouped using hierarchical clustering into distinct clusters with similar trends in environmental relationships. Spearman correlations, as a non-parametric technique, were used in this analysis to account for the non-normal distributions of both abundance and environmental data. Depth was included in this heat map to show how it relates to other variables as its effect on community structure was already tested as MLD above. Latitude roughly relates to location and was included for the same reason as depth. The variables chlorophyll, oxygen, temperature, ammonium, and Shannon diversity exhibit strong relationships with many taxa. Longitude noticeably had few significant relationships with taxa, suggesting that precise east, west, location on the shelf has little effect on the abundance of most taxa. Taxa are clustered into two main overarching groups that mainly

coincide with the taxa's response to chlorophyll and oxygen, and to a lesser extent with the taxa's relationship to temperature, salinity, nutrients and depth. The OTUs belonging to cluster 1 include *Ulvibacter, Polaribacter*, Alteromonadales, Epsilonproteobacteria, *Octadecabacter*, Bacteroidetes, Betaproteobacteria, Unassigned, and Rhodobacteraceae. They are generally positively correlated with chlorophyll, oxygen and ammonium, and negatively associated with temperature, salinity and alpha diversity. The remaining taxa, with some exceptions, are negatively associated with chlorophyll and oxygen and positively associated with temperature, salinity, alpha diversity, nutrients, and depth. This larger cluster is further subdivided into two clusters, cluster 2.1 and cluster 2.2, that appear to split mainly on their response to nitrite, nitrate, and bacteria.

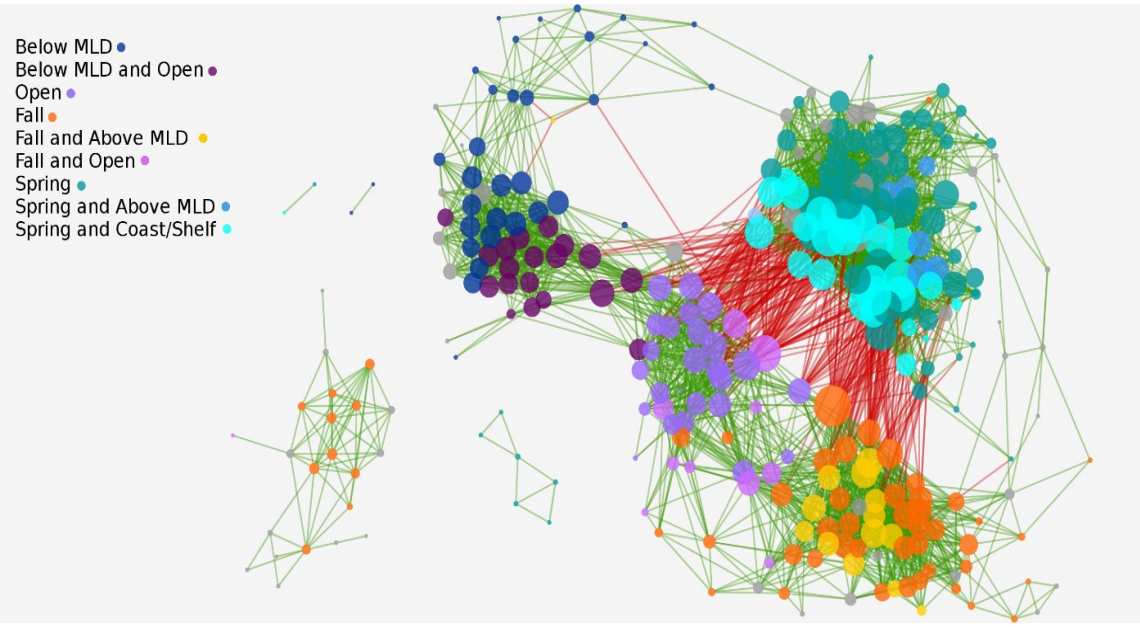


Figure 2.16. Correlation network of OTUs (Spearman's rho >0.7, *p* value <0.01). Nodes are coloured when their respective OTU is a strongly significant indicator species for a particular group (*p* value <0.0005, point biserial correlation coefficient > 0.3). Size of node is correlated to the degree (number of connections) of the OTU. Red edges denote negative correlations and green edges denote positive correlations.

352 OTUs were included in this correlation analysis and they were represented by at least 336 sequences across the shelf, or an average community abundance of 0.027%. The limit of at least 336 sequences in the dataset per OTU was chosen as a cut off because it was 2x the number of sites, thus providing a representative number of sequences with which to infer species interactions. 307 OTUs showed a significantly strong correlation (rho >0.7) with at least one other OTU and were included as nodes in

the network. This left only 45 OTUs from the original 352 that weren't highly significantly spatially correlated with any other OTUs. Out of a possible 61,776 connections, 4014 connections were made with the stringent significance and correlation filters. The average degree (node connectivity) of a node in the network was 26.15, with a range of 1 connection to 81 connections. This shows that a large number of OTUs present in the Scotian Shelf are interacting via some means with many other OTUs in their vicinity. The clustering coefficient of this network was 0.657, and the network diameter, or the longest, shortest path between any two connected nodes in the network was 8. There were 5 connected components, but most of the nodes were distributed in the one main component, and two of the components were only represented by one connection between two OTUs. The network density was 0.085, and the network heterogeneity was 0.745. The average path length, which corresponds to the average number of steps along the shortest paths for all possible pairs of network nodes, was 2.91. Additionally, the layout of these networks aims to highlight the clustered structure of the graphs, of which most OTUs fall into one of the four main clusters in the main component.

Indicator species analysis was used to identify OTUs that were preferentially abundant in specific season, mixed layer depth and location groups. The OTUs that were strong indicator species (point biserial correlation coefficient > 0.3, p value <0.0005) are coloured in Figure 2.16 and show the habitat tendencies of the main clusters of nodes. For instance, the largest cluster is dominated by OTUs that are preferentially associated with the spring, and coastal spring samples. Two main smaller clusters are negatively associated with this main cluster and are made up of OTUs found preferentially in the fall, or in the open ocean. The fourth cluster from the main component has a few negative associations with the spring cluster, and has many positive interactions with the open cluster. This final cluster is made up of mainly deep water associated OTUs. The second largest connected component also contains OTUs preferentially found in the fall.

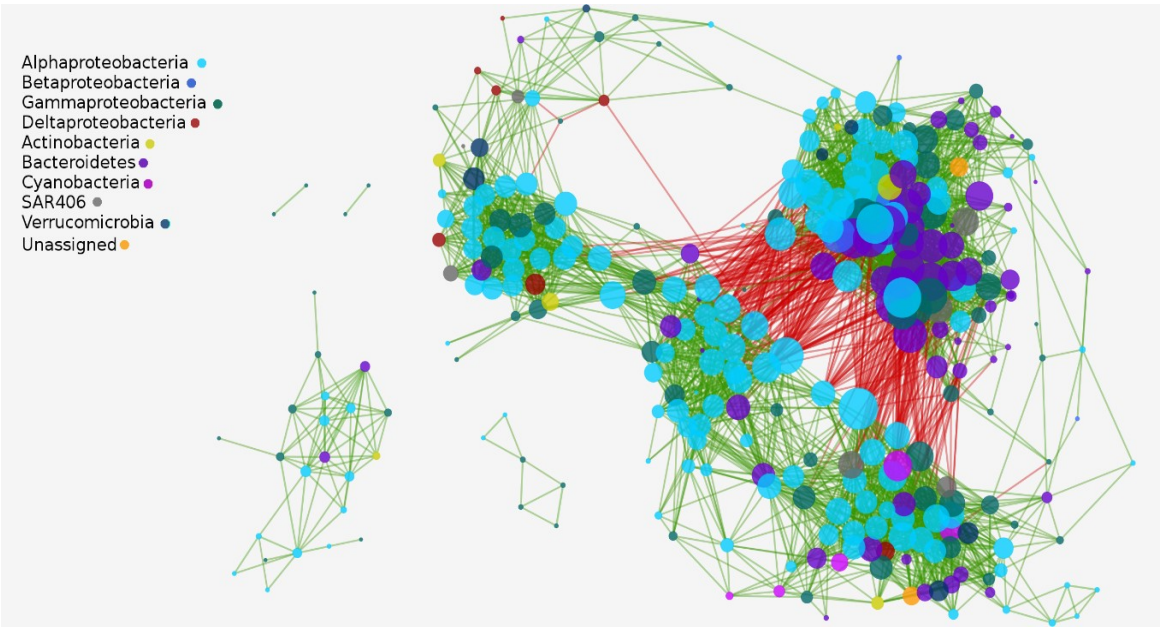


Figure 2.17. Same network as Figure 2.16 but with nodes coloured on taxonomy (class or phylum of OTU) rather than niche preference. Correlation network of OTUs (Spearman's rho >0.7, *p* value <0.01). Size of node is correlated to the degree (number of connections) of the OTU. Red edges denote negative correlations and green edges denote positive correlations.

Figure 2.17 shows the same network as Figure 2.16, but with nodes coloured based on class or phyla of the OTU. There are many interactions between both OTUs of the same phylum and OTUs of different phyla. Alphaproteobacteria was the most abundant group of the dataset and were found in every cluster. Bacteroidetes, conversely was also relatively abundant but was generally constrained to the largest cluster, which as identified by the indicator species analysis, was primarily made up of spring associated bacteria. Previous analyses in this section have also associated the Bacteroidetes taxa with spring and coastal spring samples.

In this network, the OTUs with the highest degree of connectivity were Pelagibacteraceae (OTU 3156804) with 81 significant connections, *Octadecabacter* (OTU 272142) with 77 significant connections, Flavobacteriaceae (OTU 659486) and Flavobacteriaceae (OTU 691928) with 76 significant connections each and Flammeovirgaceae (OTU 919715) and Alphaproteobacteria (OTU 733552) with 75 significant connections each. Although these were the OTUs with the highest degree of connectivity, none of these OTUs had an average percent abundance greater than 1%.

By normalizing node degree to average abundance, the OTUs with the highest disproportionate contribution, or rather the highest number of connections per sequence count, were identified. The ten OTUs with the highest disproportionate contribution were SAR406 phylum *ZA3312c* genus (OTU 229118), Rhodospirillaceae (OTU 891587), *Pseudoruegeria* (OTU 828048), Actinobacteria phylum SC3-41family (OTU 275365), Piscirickettsiaceae (OTU 547139), Bacteroidetes phylum NS9 family (OTU 620696), Gammaproteobacteria class *HTCC* genus (OTU 275776), *Flavobacterium* (OTU 308318), Pelagibacteraceae (New Reference OTU36), and Alphaproteobacteria (OTU 127587). On average, these OTUs accounted for only 0.03% of the relative abundance of their community and never reached above 1% abundance at any site. Despite their low abundance, these highly connected OTUs were correlated with an average of 42 other OTUs, or 13.7% of the OTUs in the network. These results highlight the importance of these rare OTUs in interacting with other species in their communities.

When comparing Figures 2.16 and 2.17, it is apparent that the distribution of the OTUs within the main clusters was better explained by the environmental group to which each OTU was preferentially abundant, rather than the OTU's taxonomy at the level tested.

2.3.5 Thebaud Platform Station – Comprehensive analysis of the bacterial community surrounding an offshore oil and gas platform

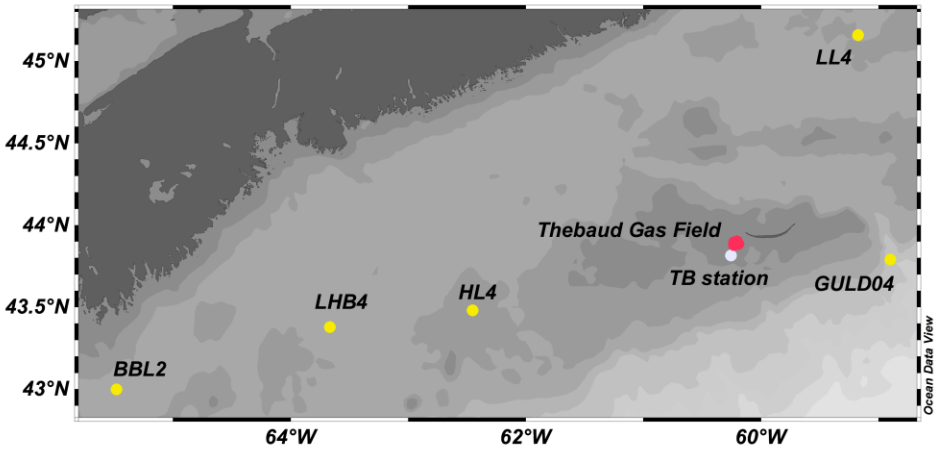


Figure 2.18. Map of the region surrounding the Thebaud gas field. The stations in yellow were included in the analysis for comparison to the TB station (white), which is situated closest to the gas field (red).

This final section focuses on the area surrounding the Thebaud gas field within the Scotian Shelf, and aims to determine how an anthropogenic, industrial activity may be affecting the bacterial community in the area. The Thebaud gas field is an offshore oil and gas field located 9 km southwest of Sable Island. The field consists of one discovery well and three other wells. It also contains an adjacent processing complex that collects gas from the Thebaud wells, as well as the North Triumph and Venture wells. During the spring cruise aboard the CCGS *Hudson* a station near the Thebaud gas field, hereafter TB station, was sampled (Figure 2.18). A generalized linear model approach similar to the one described in section 2.3.4 was used to determine if there was any statistical differences between the bacterial communities located at TB station and those located in similar positions along the shelf but not in proximity to an oil and gas platform. The model: Community ~ Station Type was used with a negative binomial distribution, where station type refers to whether the station was in the proximity of the oil and gas platform or not.

The results of this analysis confirmed that there was a significant difference between the TB station and other stations along the shelf (Supplementary Table S6). Further analysis highlighted the main OTUs that contributed significantly to the differences between groups. Figure 2.19 shows the relative abundance of the statistically significant contributing OTUs. The five OTUs that were preferentially found in the TB station compared to the rest of the shelf stations could be assigned to the family of Colwelliaceae, the family of Oceanospirillaceae, the genus of *Polaribacter*, the family of Flavobacteriaceae, and the family of Rhodobacteraceae. Groups that were significantly more abundant in other shelf stations were mainly from the family of Pelagibacteraceae, with the exception of one OTU from the genus of A*lteromonas*, and one OTU that could not be classified to any level.

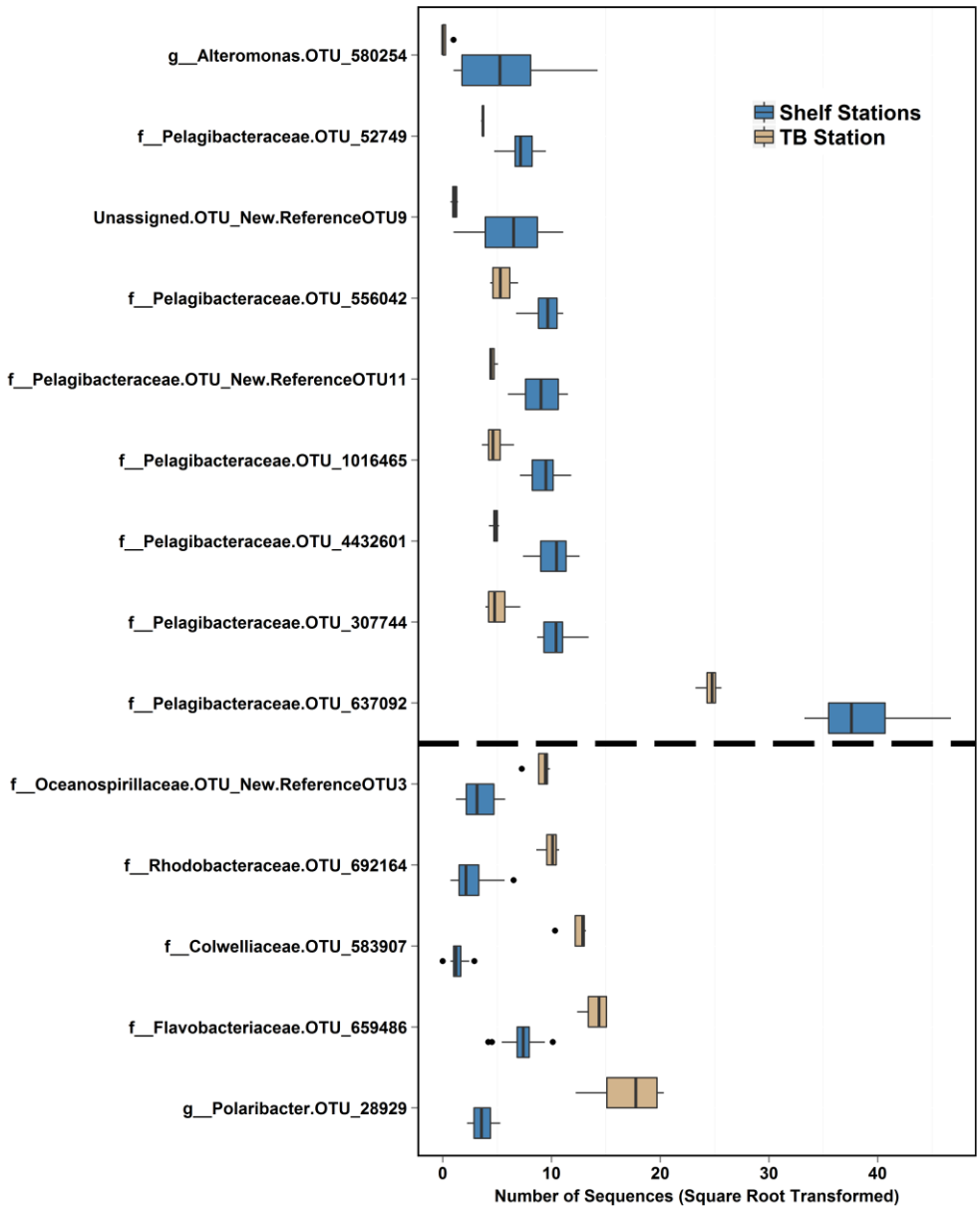


Figure 2.19. Relative abundances of bacterial OTUs that were significantly (MGLM-ANOVA with adjusted p value < 0.05) different between the TB platform station and the other shelf stations. The first letter of the OTU's name indicates the taxonomic level that the OTU could be identified to (f = family, g = genus), and is followed by the phylogenetic name and the OTU number. The dashed line divides the taxa into those that are negatively associated with the TB station (above) and those that are positively associated with the TB station (below). The x-axis has been square root transformed to account for large differences in sequence counts between OTUs.

2.4 Discussion

2.4.1 Assumptions and Cautions when interpreting 16S rRNA sequencing data

Although the ability to analyze microbial community structure *in situ* through the sequencing of marker genes is a tremendous advance in the study of microbial ecology, the method does come with some caveats. In particular the choice of marker gene (e.g. 16S vs functional gene), the DNA extraction process, and the PCR process all can contribute to biases in the resulting analysis, especially when comparing relative abundances of microorganisms (Hong *et al.*, 2009; Kembel *et al.*, 2012; Cruaud *et al.*, 2014). With respect to 16S rRNA genes and the PCR process, the choice of primers and the choice of variable regions can significantly affect the estimates of OTU richness and microbial diversity analysis (Hong *et al.*, 2009). Cruaud et al. (2014) found that the variable region amplified can have a dramatic effect on the taxonomic identification of the results. One way to alleviate this specific issue is to sequence a larger region, which is becoming increasingly viable with decreasing sequencing costs. In Cruaud et al. (2014), they used primers that amplified a region around 300 bp while the primers used in this study amplified a region over 400 bp in length. This appreciably increases the amount of information available for assigning accurate taxonomic identification. Additionally, the number of 16S rRNA gene copies in a microbial genome can vary widely across taxa which can distort the relationship between observed gene counts and actual organismal counts in samples (Kembel *et al.*, 2012). As knowledge of organismal 16S rRNA gene copy number is dependent on available complete genome information, a greater number of genomes need to be sequenced in order to address this problem completely (Kembel *et al.*, 2012). Even after the sequencing process, during the analysis of the resulting data there are many steps which can introduce variability between methods. These include the database used to assign taxonomy (Poretsky *et al.*, 2014), the decision to standardize, rarefy, or normalize the data and the resulting sequencing depth (Legendre and Gallagher, 2001; Gihring *et al.*, 2012; McMurdie and Holmes, 2014; Zinger *et al.*, 2014), the statistical methods used (Warton et al. 2012), and even the resolution of the assigned taxonomic level (Green and Bohannan, 2006; Hanson *et al.*, 2012). The variability in choices made at each of these steps can reduce the comparability of ecological results between studies. Thus it is extremely important in studies of microbial ecology to be

transparent and report accurately what methods were followed and what statistical and analytical tools were used. These caveats remind readers to be aware of potential downfalls and areas to improve upon with high throughput sequencing methods. Overall however, the introduction of high throughput sequencing has resulted in a tremendous surge in our understanding of the diversity and components of environmental microbial communities.

2.4.2 Bacterial Composition of Scotian Shelf

The first goal of this study was to identify the major bacterial taxa and OTUs present on the Scotian Shelf and to determine the factors that correlated with bacterial diversity. Overall, there were many taxa and OTUs present on the Scotian Shelf, however only a few were consistently dominant. The most abundant bacterial group in this dataset was the family Pelagibacteraceae. This group contains members of the SAR11 clade which are postulated to be the most abundant organisms on the planet, consistently identifiable as dominant community members in many marine bacterial surveys (Morris et al. 2002; Gilbert et al. 2012; Cram et al. 2015; Rusch et al. 2007). Similarly, the family of Rhodobacteraceae and phylum of Bacteroidetes (containing the order of Flavobacteriales) were dominant taxa at most sites sampled and have also been identified as dominant marine community members in other studies (Gilbert et al. 2012; El-Swais et al. 2014; Xiaomin et al. 2015).

Sampling at two different time points allows for the comparison of the bacterial community structure of the Scotian Shelf over space and time. Immediately, distinct seasonal differences in relative abundances of dominant taxa were apparent which are most likely in response to the seasonal oceanographic patterns of the Scotian Shelf (Figures 2.5 and 2.6). Spring in the Scotian Shelf is associated with the spring bloom, a massive primary production event caused by the initialization of thermal stratification and the lengthening of daytime sunlight (Sverdrup, 1953; Zhai *et al.*, 2011). After a winter of strong mixing events, stratification makes subsurface nutrients available to surface photosynthetic eukaryotes that now have enough solar radiation to grow rapidly. The phytoplankton bloom on the Scotian Shelf has been studied previously, and is mainly made up of diatom species (Li *et al.*, 2006; Craig *et al.*, 2014; Dasilva *et al.*, 2014).

These high nutrient and high chlorophyll conditions favour bacterial copiotrophs or "opportunists" that can reproduce quickly given the right conditions (Fuhrman *et al.*, 2015). After an initial bloom of growth and productivity, surface nutrients are depleted and thermal stratification becomes so strong that no mixing occurs between the nutrient poor surface layer and the nutrient rich deep layer. This is how the shelf remains throughout the summer and into the early fall, the next sampling time point. In these early fall conditions, nutrient concentrations are generally low above the mixed layer and thus the environment favours minimalistic "scavenger" taxa (Giovannoni *et al.*, 2005, 2014; Partensky and Garczarek, 2010). These groups are able to survive on exceptionally low nutrient concentrations and often have extremely small genomes and cell sizes. Depending on the environmental conditions, a second, generally smaller, fall bloom can occur on the Scotian Shelf (Song *et al.*, 2010). The most noticeable seasonal difference concerning abundant taxa is the relative increase of Bacteroidetes, and related genera like *Polaribacter* and *Ulvibacter* in the spring, and the relative increase of Pelagibacteraceae and the cyanobacterial genera *Prochlorococcus* and *Synechococcus* in the fall. Members of the Bacteroidetes phylum, specifically from the order of Flavobacteriales, are known to be associated with phytoplankton blooms and upwelling regions suggesting the group prefers high nutrient environments and has the characteristics of a copiotroph (Alonso-Sáez *et al.*, 2007; Schattenhofer *et al.*, 2009; Gómez-Pereira *et al.*, 2010). Contrastingly, Pelagibacteraceae, *Prochlorococcus*, and *Synechococcus* are known to be associated with oligotrophic conditions (Rocap *et al.*, 2003; Giovannoni *et al.*, 2005).

Additionally, sampling throughout the water column adds another dimension to this dataset which can also be analyzed with respect to seasonal variation. Similar to season, distinct patterns are apparent regarding community structure at deep stations as compared to surface stations. In addition to light attenuation, there are also well-defined temperature, salinity, oxygen, and nutrient gradients throughout the water column that, can affect bacterial community composition. The more intricate differences corresponding to rarer taxa are summarized in section 2.4.4 and Table 2.6, but noticeable differences include that deeper stations have more of the classes of Deltaproteobacteria and Gammaproteobacteria than surface stations. This trend has also been noted in previous studies and is probably due to the versatile metabolic pathways of these groups

which can take advantage of a wide range of electron acceptors and donors, independent of light requirements (Arístegui *et al.*, 2009; Orcutt *et al.*, 2011).

The final sampling dimension investigated in this dataset is that of latitude pertaining to relative location along the shelf. Previous studies at a global scale have found correlations between the distribution of certain taxa and latitude (Schattenhofer *et al.*, 2009), as well as a general negative trend of marine bacterial diversity with increasing latitude (Pommier *et al.*, 2007; Fuhrman *et al.*, 2008). Stations were divided into three categories: coast, shelf, and open which corresponded to their position with respect to the continental shelf. Coastal stations are closest to land and expected to be more influenced by freshwater runoff and anthropogenic interactions. Shelf stations are expected to be influenced by a shallower water column and the Labrador Current flowing south. The open water stations are off the shelf and characterized by a water column depth of greater than 500 m. These open stations are also expected to be influenced by the Gulf Stream flowing north. In general, the coastal and shelf stations appeared similar in bacterial composition (Figure 2.6) and alpha diversity (Figure 2.7), while the open stations are more distinct. Specifically, there are more representatives of Rhodobacteraceae and Bacteroidetes on the shelf and coast stations than in the open ocean.

Also noticeable in this dataset were large spikes of specific bacterial taxa. For instance, the genus *Pseudoalteromonas* was generally absent or low in abundance over most of the shelf, except for when it would spike to near dominance at select stations – Spring BBL1, Spring LHB6.7, Fall HL8, and Fall HL11. *Pseudoalteromonas* in the class of Gammaproteobacteria is known for its association with phytoplankton blooms and for its anti-fouling compounds (Lovejoy *et al.*, 1998; Holmström and Kjelleberg, 1999). Other studies have attributed spikes in *Pseudoalteromonas* abundance to phytoplankton blooms (Lau *et al.*, 2013), however there was no direct correlation (positive or negative) between *Pseudoalteromonas* abundance and chlorophyll concentration in our dataset. Since we could not measure chlorophyll concentration throughout the temporal period of the *Pseudoalteromonas* spike we cannot rule out the possibility of *Pseudoalteromonas* spiking directly pre or post bloom (Lovejoy et al. 1998).

The Shannon diversity index takes into account both species richness and species evenness, and returns a value that indicates the "entropy" of the system, i.e. high values correspond to a more "chaotic" or biodiverse and evenly mixed community. An interesting pattern emerged when analyzing the correlation of seasonal alpha diversity with environmental variables (Figure 2.8, Table 2.2). Mainly, only salinity was found to have a significant relationship with alpha diversity when considering sites sampled in the fall. However, salinity, temperature, oxygen, chlorophyll, nitrate, silicate, phosphate, bacteria, latitude, and depth all had significant relationships with alpha diversity when considering sites sampled in the spring (Table 2.2). Salinity, temperature, nitrate, phosphate, silicate and depth all exhibited significantly positive relationships with diversity, whereas oxygen, chlorophyll, latitude and bacteria concentration had significantly negative relationships with diversity. At a local scale, Gilbert et al. (2012) observed a cyclical pattern in species richness with a peak in winter months at the L4 site in the English Channel. The time series study in the North Atlantic Bedford Basin by El Swais et al. (2014) observed a similar cyclical pattern in alpha diversity with a peak in winter months and a relative low in summer months. Studies by El Swais et al. (2014) and Gilbert et al. (2012) showed that spring and fall months displayed comparable magnitudes of bacterial diversity, which is similar to our findings (Figure 2.7). However, neither study investigated the direct effect environmental variations have on alpha diversity.

Other previous studies have shown a trend with latitude and temperature and diversity at a global scale (Pommier *et al.*, 2007; Fuhrman *et al.*, 2008). Additionally, a regional study by Wang et al. (2015) from August in the East China Sea found many significant correlations with diversity and environmental variables. Specifically, they found that the presence of suspended particles explained the largest amount of variation in alpha diversity. In addition they found that Shannon diversity was positively correlated with latitude, and nutrients, and negatively with longitude, salinity, pH and dissolved oxygen. The relationships identified in our study are in agreement with some of the relationships identified in the East China Sea study. In particular, the negative relationship between oxygen concentrations and bacterial diversity, and the positive relationship between nutrient concentrations and bacterial diversity were the same.

However, we observed opposite trends to Wang et al. (2015) with respect to latitude and salinity, and we did not observe a significant relationship between longitude and diversity. The lack of agreement in these results suggest regional differences in factors that drive diversity in the marine environment. The main variability between these two studies is likely linked to the lack of deep samples in the East China Sea study and the heightened anthropogenic influences on the East China Sea study as compared to the Scotian Shelf. An alternate study by Walsh et al. (2015), looked at bacterial diversity over the water column in three locations in the Pacific Ocean. This study found that bacterial diversity mirrored oxygen concentrations throughout the water column, with diversity peaking in the pelagic oxygen minimum zone. Both of these studies provided singular "snapshots" of the microbial community at one time and thus did not investigate how correlations between diversity and environmental factors change over time. In this present study, two such "snapshots" were compared in order to incorporate a temporal dimension to analyses of bacterial diversity.

Ladau et al. (2013) investigated through modeling, the effect seasonality has on global estimates of diversity and found a distinct pattern that contradicts the negative correlation of latitude with diversity found in our study, and by the studies from Pommier et al. (2007) and Fuhrman et al. (2008). Specifically, Ladau et al. (2013) found that the winter season in the polar and temperate regions of the ocean corresponded to the greatest bacterial diversity and the summer season of the same region corresponded to the lowest bacterial diversity estimates. In other words, the magnitude of the relationship between latitude and bacterial diversity in the temperate ocean changes depending on the season when the samples were taken. Thus, this study by Ladau et al. (2013) demonstrates that season can have a profound impact on the factors that shape community diversity, as was found in our study.

The apparent contradiction between the seasonal responses of alpha diversity to environmental variables in bacterial communities could be explained by a number of other factors. Most realistically, the bacterial diversity in the spring is well defined by the environmental variables measured. However, the bacterial diversity of the fall communities is mainly influenced by alternate factors that were not measured, or focused on in this study. These factors could include physical parameters like water masses and

currents, chemical parameters like micronutrients (iron, B12), sulfur, hydrocarbons, and dissolved organic matter, or biological parameters like predation, viruses, host availability, and eukaryotic abundance (Lima-Mendez *et al.*, 2015).

2.4.3 Beta Diversity, Biogeography and Patterns in Spatial Distribution

The second goal of this study was to identify patterns of biogeography and their main drivers across the Scotian Shelf, qualitatively by observing individual taxa, and quantitatively by using overall community similarity (beta diversity). By traversing along numerous transects, the Scotian Shelf area was well covered by sampling sites in this study, allowing for the recreation of a rough "snapshot" of the distribution of different bacterial taxa over the shelf. These surface plots in Figures 2.10, 2.11 and 2.12, provide an overview of how these groups vary across the shelf at the surface and allow for comparisons across seasons and between taxa. Striking seasonal differences in distributions between taxa are apparent. Specifically, the oligotrophic clade Pelagibacteraceae is relatively more abundant in the fall than in the spring, whereas the phytoplankton bloom associated Rhodobacteraceae and Bacteroidetes groups are more abundant in the spring.

When looking specifically at closely related genera like *Polaribacter* and *Ulvibacter* slight differences in their distributions are apparent. Both *Polaribacter* and *Ulvibacter* have been associated with spring blooms (Klindworth *et al.*, 2014), and were much more abundant during the spring in the Scotian Shelf. Both groups are also generally more abundant near the coast where phytoplankton blooms are strongest on the shelf at the time of sampling (Zhai *et al.*, 2011), but they have decidedly different distributions over the shelf. Thus, there could be slight differences in their preferred ecological niches that cause this spatial separation (Xing *et al.*, 2015).

Defined spatial separation between closely related taxa is even more apparent when looking at the distributions of Cyanobacterial genera *Prochlorococcus* and *Synechococcus* in the fall season. Globally, *Prochlorococcus* is most abundant in the oligotrophic tropical regions of the ocean where nutrients are most scarce and its small size and energy requirements offer a significant advantage (Partensky and Garczarek, 2010). The range of *Synechococcus* is similar but extends more towards coastal regions

and towards the poles, as *Synechococcus* is more tolerant of colder conditions than *Prochlorococcus* (Partensky *et al.*, 1999). In this study, *Synechococcus* was found to a small extent on the shelf, whereas *Prochlorococcus* was completely absent. *Prochlorococcus* gained abundance directly off the shelf in the open ocean, while *Synechococcus* became abundant in the lower southwest corner of the study area, which is still in close proximity to the coast of the United States. DaSilva et al. (2013) conducted a eukaryotic survey of the Scotian Shelf during fall and spring cruises, which was supplemented with flow cytometry measurements on *Synechococcus* and *Prochlorococcus*. They also found a dramatic increase in *Synechococcus* and *Prochlorococcus* numbers in the fall as compared to the spring and suggested that the influence of the warm Gulf Stream waters was creating a suitable habitat for these picophytoplankton, particularly *Prochlorococcus*, on the shelf region (Dasilva *et al.*, 2014). The influence of the Gulf Stream is the most probable explanation for the distribution of *Prochlorococcus* in this study as well, as the genus was mainly found off the shelf, in the lower latitudes (Figure 2.12), where temperatures were higher and nutrients were lower (Figure 2.2f,j).

Beta diversity is the degree of community differentiation between sites and accounts for the ratio between local and regional diversity (Whittaker, 1960; Jost, 2007). In this study, beta diversity was used to visualize the separation of sites based on bacterial community similarity on a two dimensional plane through the NMDS analysis (Figure 2.13), as well as to determine which environmental or geographic factors best correlated with community similarity (Tables 2.3 and 2.4). Figure 2.13 shows an NMDS ordination using the Bray-Curtis measure of beta diversity of all sites sampled, coloured on seasonal groupings and sized on depth. No clear clusters appear in this ordination, however there is a definite separation of sites based on season, and a gradient across the second axis of sites based on depth. Although a divide between spring and fall samples at depth is still apparent, it seems to be less defined than the split in the shallower sites. This suggests a general decrease in seasonality at depth which has been described previously (Hatosy *et al.*, 2013; Fuhrman *et al.*, 2015).

The central goal of biogeography is to determine the distribution of organisms in space and time, and to identify the underlying mechanisms that control or shape this

diversity (Hanson *et al.*, 2012). Environmental factors are often spatially autocorrelated with geographic distance, meaning that sites in close geographic vicinity will generally have similar environmental conditions, and thus effects distinctly due to historical processes can often be confounded by environmental similarity. In this study we attempted to disentangle the effects of environmental distance from geographic distance using Partial Mantel tests that controlled for one matrix while testing for the effects of the other on community structure. Correlations were calculated for all sites, sites per season, and sites categorized by season and depth in order to observe how these relationships changed over different seasons and depths. The results summarized in Table 2.3, indicate that the environmental variation between sites has a larger influence on community structure than geographic distance has on community structure. Our results support the cosmopolitan "environment selects" theory and are in line with other marine microbial studies from different locations (Nguyen and Landfald, 2015; Wang *et al.*, 2015). The results suggest that the regional scale of this study coupled with the fairly continuous marine environment limits the degree of dispersal limitation, thus reducing the effect of historical processes in shaping overall bacterial community structure (Martiny *et al.*, 2006).

Although overall environmental distance is undoubtedly a greater driver of bacterial community assemblage than geographic distance, there are subtle differences over the water column and between spring and fall samples (Table 2.4). The relationship between environment and community seems to be slightly stronger in the spring samples than in the fall samples and stronger in the upper three depths of both seasons than in their respective 250 m depths. Additionally, the relationship between community dissimilarity and geographic distance is actually significant (albeit with a low correlation value) with all fall samples. As the geographic relationship isn't nearly as strong as the environmental relationship in the fall samples it is clear that contemporary selection is still the main process driving bacterial community structure during this season. However, the significant relationship between geography and community may suggest that there are some historical processes shaping bacterial structure over the Scotian Shelf, which act more strongly in the fall than in the spring. The water properties of the fall stations sampled were much more heterogeneous than the water properties of the spring samples,

as shown by their TS diagrams (Supplementary Figures S3 and S4), suggesting multiple distinct water masses in the fall and few in the spring. Previous studies have identified distinct microbial communities in different water masses, suggesting that water mass has the potential to restrict bacterial dispersion to some extent (Agogué *et al.*, 2012; Monier *et al.*, 2014).

Oxygen was the single environmental variable that best correlated with changes in community similarity over all sites. Oxygen is an extremely important molecule in defining bacterial niches, as it is absolutely necessary for obligate aerobes, fatally toxic for obligate anaerobes, and of varying importance to microbes with intermediate phenotypes (Wright *et al.*, 2012). The presence or absence of oxygen in the marine environment can dramatically change the type of microorganisms and the types of metabolic processes occurring in that region (Ganesh *et al.*, 2014; Hawley *et al.*, 2014), and although the waters of the Scotian Shelf do not reach anoxic conditions, the oxygen concentrations can decrease substantially at depth as compared to the surface waters (Figure 2.3 c,i,f,l), and in the fall as compared to the spring (Figure 2.2 c,h). The results of the Partial Mantel tests suggest that even the relatively small changes in oxygen concentration over the Scotian Shelf had a strong impact on the bacterial community of the region. Temperature was another strongly correlated variable with community structure in this study, and has also been found as a strong driver of bacterial community structure locally in the Bedford Basin (El-Swais *et al.*, 2014), for the eukaryotic community on the Scotian Shelf (Dasilva *et al.*, 2014), as well as globally (Rusch *et al.*, 2007; Ladau *et al.*, 2013; Sunagawa *et al.*, 2015). The identification of oxygen and temperature as highly correlated with community structure could have many implications for the future of bacterial communities in the region as oxygen levels are predicted to decrease and temperatures are predicted to rise with climate change (Stramma *et al.*, 2008; Finkel *et al.*, 2010; Wright *et al.*, 2012).

In addition to analyzing all sites together, sites were analyzed based on specific season, or season and depth categories. These analyses highlighted main differences in drivers between seasons and over depths, showcasing the complexities underlying community assemblage at different spatial and temporal scales. While many trends became apparent with this analysis a few main results will be discussed now. Firstly a

few variables, (e.g. nitrate in the fall) were significant for all sites in one season but not for any particular depth from the fall. This would suggest that nitrate is a main factor driving community change across the depths in the fall, but is not associated with community structure across any one depth for that season, i.e. is not associated with community structure for any horizontal segment within the water column.

Additionally, the fall 250 m samples behaved unusually as the community dissimilarity of these sites was not correlated significantly with environmental or geographic distance matrices. The fall 250 m sites however were significantly correlated with bacterial diversity and longitude, but no other individual variables. Bacterial diversity was also strongly associated with community for all spring sites collectively and all other individual fall depths. This result suggests biotic factors, such as the relationships between certain taxa, and the functional and taxonomic variety of species in a community, may have a stronger role in structuring communities than other abiotic factors. Recent studies have found that biotic interactions are stronger in marine microbial communities than previously thought (Lima-Mendez *et al.*, 2015). The fall 250 m sites were the only group show a response to longitude. This was interesting as the group's relationship with geographic distance was not quite significant. This would suggest that at 250 m in the fall, the bacterial community is structured strictly from east to west but not north to south. Most of the 250 m stations were off the shelf, either south of the shelf or east of the shelf within the Cabot Strait. The Gulf Stream would most likely heavily influence sites south of the shelf as the northern edge of the Gulf Stream (shelf-slope front) is known to migrate annually, locating to its most northward position in the fall (Zhai *et al.*, 2011). The Gulf Stream, which can span up to 100 km, moves rapidly east in the North Atlantic, becoming cooler and fresher in the process (Affholder and Valiron, 2001). Thus the position of the Gulf Stream in the fall could create a strong water mass that could potentially be influencing community structure at 250 m depth.

The sites of spring 250 m also behave quite differently with respect to the upper depths. In particular, the effect of phosphate is strong at 250 m but not in the upper three depths in the spring. This suggests that the gradient of this nutrient strongly affects the community at depth but not at the surface. Additionally, oxygen, temperature, and salinity were all strongly correlated with community in both the fall and spring upper

depths but not at the 250 m depths. This would suggest that the gradients of these variables at depth are not large enough to elicit a response in the bacterial community. Lastly, there was a noticeable difference in the response to particular environmental variables from the three upper depths in the spring as compared to the fall. Specifically, for 1 m, 20 m, and 80 m in the spring, nitrate was strongly correlated with community similarity, however phosphate and silicate were not. In contrast, for 1 m, 20 m, and 80 m in the fall, phosphate was strongly correlated with community similarity but nitrate and silicate were not. This would suggest that there are possible shifts in the limiting nutrients that affect bacterial community composition between the two seasons sampled.

As the results of the Partial Mantel tests conclusively show that environmental variables drive community structure changes to a greater degree than geography, the next goal of this study was to identify the responses of individual taxa and OTUs to the observed variations in environmental drivers.

2.4.4 Multivariate Analysis of Abundance and Correlation Analyses

Table 2.5. Description of significant correlations associated with at least 2/3 of the taxa in each cluster from Figure 2.15.

Cluster	Positive Correlations	Negative Correlations
1	Chlorophyll Oxygen Ammonium Nitrite Latitude	Temperature Salinity Diversity Depth
2.1	Bacterial Abundance Temperature	Chlorophyll Oxygen Ammonium Nitrite
2.2	Salinity Diversity Nitrate Depth	Chlorophyll Oxygen Bacterial Abundance

Table 2.6. Descriptions of select taxa that are significantly affected by group and environmental explanatory variables. Taxa are ordered on heat map cluster group (Table 2.5). Note that characteristics included are general and are in no way meant to describe every bacterial species of the associated clade.

Clade	Characteristics	Significance
Alteromonadales	-Characteristic of early spring bloom in Bedford Basin (El-Swais *et al.*, 2014) -Found to utilize diatom-derived organic carbon (Sarmento and Gasol, 2012)	Cluster 1 Season(Spring) Location(Coast/Shelf)
Bacteroidetes	-Dominant class of Bacteroidetes in marine picoplankton is Flavobacteriales (Gómez-Pereira *et al.*, 2010) -Distribution linked to cold waters, phytoplankton blooms (Abell and Bowman, 2005), photic zone (Schattenhofer *et al.*, 2009) and upwelling systems (Alonso-Sáez *et al.*, 2007) suggesting an inclination for highly productive regions (Gómez-Pereira *et al.*, 2010) -Some clades dominate during phytoplankton blooms and some dominate before (Riemann *et al.*, 2000) -Some are also associated physically with algal cells (Gómez-Pereira *et al.*, 2010) -Potential ability to degrade polymers released by phytoplankton (Kirchman, 2002) -Large diversity in phylum, however ability to degrade polymers is common among members (González *et al.*, 2008)	Cluster 1 Season(Spring) MLD(Above) Location(Coast/Shelf)
Betaproteobacteria	-Most of the known bacteria that perform ammonium oxidation are from this class, specifically *Nitrosospira* and *Nitrosomonas* (Li *et al.*, 2012) -Usually rare in marine environment but have been found to bloom to community dominance sporadically (Alonso-Sáez *et al.*, 2014)	Cluster 1 Location(Coast/Shelf)
Epsilonproteobacteria	-Deep branching ubiquitous class with predominantly autotrophic or mixotrophic and metabolically diverse members (Campbell *et al.*, 2006) -Some marine pelagic members have the ability for dark carbon dioxide fixation, sulfur oxidation and chemolithoautotrophic denitrification (Grote *et al.*, 2007; Bruckner *et al.*, 2013) -Potentially associated with phytoplankton bloom sedimentation (Sakami *et al.*, 2015)	Cluster 1 Season(Spring) Location(Coast/Shelf)
Octadecabacter	-Member of the ubiquitous Roseobacter clade -Psychrophilic, display a bipolar distribution, and make up a large portion of polar sea ice communities (Vollmers *et al.*, 2013) -Horizontal gene transfer suggested to have a large influence on diversity of clade (Newton *et al.*, 2010)	Cluster 1 Season(Spring) Location(Coast/Shelf)
Polaribacter	-First isolated from Antarctic sea ice (Gosink *et al.*, 1998), but also found in temperate marine environments (Xing *et al.*, 2015) -Found associated in the late stages of phytoplankton blooms (Teeling *et al.*, 2012; El-Swais *et al.*, 2014) -Also dominate particle/algae associated bacterial community, and contain genes for attachment to particle surfaces and polymer degradation (González *et al.*, 2008; Teeling *et al.*, 2012) -Some members contain proteorhodopsin, perhaps as a mechanisms to survive in oligotrophic ocean surface devoid of particles (González *et al.*, 2008)	Cluster 1 Season(Spring) Location(Coast/Shelf)

Clade	Characteristics	Significance
Rhodobacteraceae	-Widespread, abundant, and extremely metabolically versatile (Fu *et al.*, 2013) -Also contain genes for gene transfer agents (GTAs) that transfer cellular genomic DNA between cells, suggests possible role in promoting genetic exchange in microbial communities (Lang and Beatty, 2007) -Ecotype differentiation among species, one in particular was found to be negatively correlated with salinity and positively with oxygen (Fu *et al.*, 2013)	Cluster 1
Ulvibacter	-Characteristically associated with spring bloom conditions (El-Swais *et al.*, 2014) -Responded quickly to phytoplankton bloom in succession time series study (Teeling *et al.*, 2012)	Cluster 1 Season(Spring) Location(Coast/Shelf)
Alcanivorax	-Mainly associated with hydrocarbon degradation (Schneiker *et al.*, 2006) -Fairly ubiquitous in distribution with oligotrophic lifestyle (Schneiker *et al.*, 2006)	Cluster 2.1 Season(Fall) MLD(Below) Location(Open)
Alteromonas	-Globally distributed, aerobic bacteria found in both surface and deep waters (Math *et al.*, 2012) -Suggested copiotroph as previous studies have found *Alteromonas* associated with heterotrophic blooms (McCarren *et al.*, 2010) -Some members are able to metabolize aromatic hydrocarbons (Math *et al.*, 2012)	Cluster 2.1 MLD(Below)
Erythrobacteraceae	-Belong to the order of Sphingomonadales -Most contain bacteriochlorophyll a, and various carotenoids -Isolated from diverse environments including seawater, marine sediment, marine invertebrates, cold-seep sediment and desert sand (Tonon *et al.*, 2014) -Some members are aerobic anoxygenic phototrophic bacteria (Wu *et al.*, 2012; Wang *et al.*, 2014)	Cluster 2.1 Season(Fall)
Nocardioides	-Members isolated from marine sediment from wide variety of locations (Dastager *et al.*, 2009; D.-F. Zhang *et al.*, 2014; Deng *et al.*, 2015) -Generally aerobic (Dastager *et al.*, 2009; Deng *et al.*, 2015)	Cluster 2.1 Season(Fall)
Pelagibacteraceae	-Earth's most abundant organism (Morris et al. 2002) -Heterotrophic and previously linked to euphotic zones -Also referred to as SAR11 clade -Extremely oligotrophic, with a tiny cell size and genome (Giovannoni *et al.*, 2005) -Contains proteorhodopsin (Giovannoni *et al.*, 2005)	Cluster 2.1
Prochlorococcus	-Highly studied and very abundant group of cyanobacteria -Thrive in oligotrophic conditions due to small cell size (Partensky *et al.*, 1999) -Can reach depths of 150 m in the water column but does not generally grow in temperatures below 15°C -In depth analysis of distribution has been reviewed elsewhere (Flombaum *et al.*, 2013)	Cluster 2.1 Season(Fall) Location(Open)
Synechococcus	-Highly studied and very abundant group of cyanobacteria -Thrive in oligotrophic conditions due to small cell size (Partensky *et al.*, 1999) -Doesn't extend quite as far deep as *Prochlorococcus* but has wider geographic distribution including colder and more coastal regions -In depth analysis of distribution has been reviewed elsewhere (Flombaum *et al.*, 2013)	Cluster 2.1 Season(Fall) Location(Open)

Clade	Characteristics	Significance
TM6	-Candidate phylum found in diverse habitats -Little is known, preliminary genome analysis suggests certain members are Gram-negative, and facultative anaerobes (McLean *et al.*, 2013) -Small genome size suggest potential for symbiosis (McLean *et al.*, 2013)	Cluster 2.1 Season(Fall)
Verrucomicrobia	-Nearly ubiquitous in marine environment (Freitas *et al.*, 2012) -Nearly all isolates can grow chemoheterotrophically on many organic carbon compounds (Yoon *et al.*, 2007; Yoon, 2011) -Grow slowly in culture and many isolates have a cell diameter of approximately 1µm (Yoon *et al.*, 2007) -distinct ecotypes within the phylum that occupy separate niches and respond differently to environmental gradients (Freitas *et al.*, 2012)	Cluster 2.1 Season(Fall) MLD(Below)
Vibrionaceae	-Chemoorganotrophic bacteria, some members cause infections in humans and fish -Can be associated with host organisms (Turner *et al.*, 2009) -*Vibrio* genus found to increase over decades in response to increasing sea surface temperatures in the North Sea (Vezzulli *et al.*, 2012)	Cluster 2.1 Season(Fall) MLD(Below) Location(Open)
Acidobacteria	-Widely distributed phylum in environmental samples, but particularly abundant in soils and sediments (Ward *et al.*, 2009) -Most are aerobic heterotrophs (Ward *et al.*, 2009)	Cluster 2.2 MLD(Below)
Actinobacteria	-Primarily soil bacteria but also found in aquatic environments -One of the largest and most diverse bacterial phyla -Gram positive with a generally high GC content (Ventura *et al.*, 2007) -Play important role in recycling and produce bioactive compounds with pharmaceutical applications (Valliappan et al. 2014) -Can be associated with marine organisms (Valliappan et al. 2014)	Cluster 2.2 MLD(Below)
Chloroflexi	-Found to be relatively abundant at very deep samples, greater than 500 m to as deep as 4000 m in both Pacific and Atlantic Oceans (Morris *et al.*, 2004) -Involved in carbon cycling: organohalide respiration, respiration of sugars, fermentation, CO_2 fixation and acetogenesis (Hug *et al.*, 2013)	Cluster 2.2 MLD(Below) Location(Open)
Deltaproteobacteria	-Metabolically diverse class that includes sulfate-reducing bacteria, and iron reducing bacteria -Some members involved in various steps of the nitrogen cycle (Wright *et al.*, 2012)	Cluster 2.2 MLD(Below)
Gammaproteobacteria	-Diverse clade that is richer in genera than all bacterial phyla except Firmicutes (Williams *et al.*, 2010) -Members display a wide variety of oxygen tolerance, temperature tolerance and trophism (Williams *et al.*, 2010)	Cluster 2.2 MLD(Below) Location(Open)
Gemmatimonadetes	-Contains one order, one family and one genus -Motile, aerobic, and chemoorganotrophic -Mainly associated with soil, but some members are marine sediment associated (DeBruyn *et al.*, 2011)	Cluster 2.2 MLD(Below)
Lentisphaerae	-Designated in 2004 (Cho *et al.*, 2004) -Isolated from a variety of environments -Some strains produce exopolysaccharides that have been linked to increased pressure and cold tolerance (Thrash *et al.*, 2010)	Cluster 2.2 MLD(Below)

Clade	Characteristics	Significance
Oceanospirillales	-Most members are halophilic or halotolerant (Brenner *et al.*, 2005) -Aerobic, microaerophilic or facultative anaerobes and chemoorganotrophs (Brenner *et al.*, 2005) -Some form symbiotic relationships with marine hosts (Verna *et al.*, 2010) -Contains psychropiezophilic members which reside in high pressure, low temperature environments (Cao *et al.*, 2014)	Cluster 2.2 Season(Spring) MLD(Below)
PAUC34	-Candidate phyla, some members associated with marine sponges (Hentschel *et al.*, 2002; Jeong *et al.*, 2014)	Cluster 2.2 MLD(Below)
Planctomycetes	-Some members can perform anammox – anaerobic oxidation of ammonium (Strous et al. 1999) -Ubiquitous and found in a diverse range of habitats (Wagner and Horn, 2006)	Cluster 2.2 MLD(Below)
Pseudoalteromonas	-Frequently found in association with eukaryotic hosts in marine environment (Holmström and Kjelleberg, 1999) -Produce a range of anti-fouling compounds (Lovejoy *et al.*, 1998; Bowman, 2007)	Cluster 2.2 Location(Open)
SAR324	-Deep branching phylogenetic clade -Ubiquitous in ocean with potential capacities for sulfur oxidation, carbon fixation, C1 utilization, hydrocarbon degradation, and heterotrophy (Sheik *et al.*, 2014) -Among the most abundant taxa in oxygen minimum zones, with abundance correlating with low oxygen conditions (Wright *et al.*, 2012)	Cluster 2.2 MLD(Below)
SAR406	-Also referred to as Marine Group A and has no cultured representatives (Wright *et al.*, 2014) -Ubiquitous in dark ocean, but most prevalent and diverse in interior oceanic regions with distinct oxyclines (Zaikova et al. 2010; Wright et al. 2012) -Contain genes relating to O_2 deficiency and sulfur-based metabolism suggesting role of some groups in sulfur cycle (Wright *et al.*, 2014)	Cluster 2.2 MLD(Below) Location(Open)
SAR86	-Ubiquitous in the global ocean but most abundant in surface ocean following onset of thermal stratification (Treusch *et al.*, 2009) -Aerobic chemoheterotroph with potential for proteorhodopsin-based ATP generation (Dupont *et al.*, 2012) -Metabolic streamlining with a distinct carbon compound specialization (Dupont *et al.*, 2012) -Identified as an indicator species for autumn season in North Atlantic Bedford Basin (El-Swais *et al.*, 2014)	Cluster 2.2 Season(Spring)

Table 2.6 summarizes the results from the multivariate analysis of abundance data. The significance of environmental variables and groupings with respect to specific bacterial taxa are shown along with a few of the taxon's defining characteristics. Although some bacterial taxa, such as *Synechococcus* and *Prochlorococcus*, have been well studied in both laboratory and natural settings (Partensky *et al.*, 1999; Rocap *et al.*, 2003; Partensky and Garczarek, 2010), many taxa are only defined by sequences belonging to uncultured representatives and little is known about their distribution, their metabolism, or their roles in biogeochemical cycling (Epstein, 2013; Youssef *et al.*,

2015). Without knowledge of the main metabolic functions and characteristics of each group it is difficult to link environmental preferences to purpose in the community.

Distance based approaches to multivariate analysis such as SIMPER and ANOSIM have been widely used in microbial ecology studies (Rees *et al.*, 2004; Ramette, 2007). However, since their introduction, it has been known that these methods can easily confound dispersion and location effects between groups and make implicit assumptions that abundance data rarely fits (Anderson, 2001, 2006; Warton *et al.*, 2012). Consequently, when testing for significant group effects and identifying the species that contribute significantly to those effects, the most variable groups and species (dispersion) will often be identified by these methods instead of those with actual differences in their means (location) (Warton *et al.*, 2012). In this study, to avoid these issues, an approach utilizing generalized linear models was used. This was done to test the strength of significance of the three main environmental groupings, season, MLD, and location in driving community structure, and to determine which individual taxa were most responsible for the community shifts introduced by each of these groupings. The results of this analysis are shown in Figure 2.14, compiled in Table 2.6 and discussed in section 2.4.4.1. A heat map was then created (Figure 2.15) to visualize significant non-parametric correlations between taxa and environmental variables, the results of which are compiled in Table 2.5 and Table 2.6, and discussed in section 2.4.4.2. Lastly, the co-occurrence patterns of OTUs with an average abundance of roughly 0.027% across all sites were visualized using networks which are shown in Figures 2.16 and 2.17. This analysis was conducted to determine potential biological interactions shaping community structure, and was supplemented with indicator species analysis to determine the niche preferences of the OTUs in the network. These results are discussed in section 2.4.4.3.

2.4.4.1. Group Effects: Season, MLD and Location

All three group effects tested: Season, MLD, and Location contributed significantly to bacterial composition. Season had the strongest effect with the highest sum of likelihood ratios, followed by MLD and location. This suggests that seasonality in the Scotian Shelf and the environmental shifts that are attributed to changing seasons drive the majority of the bacterial community structure in the region. The taxa that exhibited the most significant effects with respect to community shifts brought about by

season in order of decreasing univariate deviance scores include Erythrobacteraceae (fall), *Ulvibacter* (spring), Vibrionaceae (fall), *Synechococcus* (fall), *Polaribacter* (spring), Cyanobacteria (fall), and Alteromonadales (spring). The ecological niche preferences of the minimalist cyanobacterial clades are well known and in agreement with previous studies, they were found to be significantly associated with the fall season and the accompanying oligotrophic conditions (Partensky and Garczarek, 2010; Flombaum *et al.*, 2013). The presence of these genera in the Scotian Shelf region highlights the extent of stratification and the oligotrophic conditions in the early fall. In addition to the trends in cyanobacterial seasonal fluxes that support existing literature, other less studied taxonomic groups were identified as key seasonal taxa. The largest contributor was the family of Erythrobacteraceae which contains the genus *Erythrobacter*. Erythrobacteraceae are a relatively understudied group with respect to their marine distribution, although members of the group seem to be fairly ubiquitous and have been isolated from the South China Sea (Y. Zhang *et al.*, 2014), the Northwest Pacific (Sato-Takabe *et al.*, 2012), and the west Mediterranean Sea (Denner *et al.*, 2002), in addition to being associated with seaweed and cyanobacterial mats (Yurkov *et al.*, 1994; Wang *et al.*, 2014). Members of this group include aerobic anoxygenic bacteria containing bacteriochlorophyll *a*, and thus they are considered important members of the carbon cycle (Wang *et al.*, 2014). The taxa ranged from being undetectable to having a relative abundance of almost 6%. On average, in the spring they made up 0.04% of the bacterial community, and in the fall they made up on average 1% of the bacterial community, a 25 fold change.

Taxa more abundant in spring samples are likely associated with either phytoplankton directly via symbiosis and degradation, or indirectly by favouring the productive conditions that lead to their initial blooms (Teeling *et al.*, 2012; Georges *et al.*, 2014). In this analysis, the genera *Ulvibacter* and *Polaribacter* from the Bacteroidetes phylum, and the order Alteromonadales were strongly associated with spring samples. The average relative abundance of *Polaribacter* in the spring was 2.1%, but in the fall it was only 0.09%, a 22 fold difference. *Ulvibacter* experienced similar dynamics with an average abundance of 0.05% in the fall, and an average abundance of 2.2% in the spring. Alteromonadales was one of the more abundant taxa overall, but was three times as

abundant in spring samples with an average of 5% of the community as it was in the fall with an average of 1.7%. A previous study by El Swais et al (2014), in this region associated all three of these taxa with spring bloom conditions. Specifically, they found *Ulvibacter* and *Polaribacter* associated with post bloom conditions and Alteromonadales associated with early bloom conditions, however this dataset is not designed to analyze the fine dynamics of the spring bloom bacterial communities.

SAR324 and other Deltaproteobacteria expressed the strongest community shift with respect to depth, with both groups being dramatically more abundant below the mixed layer than above. Specifically, SAR324 was 0.1% of community above the MLD and 1.5% of the community below, while other Deltaproteobacteria were 0.4% of the community above the MLD and 1.5% of the community below the MLD. SAR324, and other members of Deltaproteobacteria are known to have a wide range of metabolic capacities, including sulfur oxidation, carbon fixation, C1 utilization, and heterotrophy (Sheik *et al.*, 2014). Below the mixed layer, in an environment untouched by sunlight and with limited surface derived carbon sources, relying on a solely heterotrophic lifestyle is difficult to sustain (Arístegui *et al.*, 2009; Sheik *et al.*, 2014). In order to survive at these depths many bacterial taxa, like SAR324 and other Deltaproteobacteria members, have gained an advantage by evolving a diverse and flexible metabolic arsenal (Orcutt *et al.*, 2011; Swan *et al.*, 2011). Alternate groups with large contributions to the bacterial community shift corresponding to depth include Planctomycetes and SAR406 (or Marine Group A), which have members that participate in anammox processes and sulfur cycling in low oxygen environments (Francis *et al.*, 2007; Wright *et al.*, 2014). Additionally, groups that are found more frequently below the mixed layer may be associated with marine sediment, such as Acidobacteria and Gemmatimonadetes (Ward *et al.*, 2009; DeBruyn *et al.*, 2011; Walsh *et al.*, 2015). Further transcriptomic studies on these environments would provide information on the role of these normally sediment associated taxa in the marine water column, specifically if they are active members of the community or if they are dormant, or inactive and resuspended by chance (Nemergut *et al.*, 2013). Only three taxa were significantly, preferentially found above the mixed layer. These taxa included Cyanobacteria, Bacteroidetes, and those OTUs that could not be assigned to any level of taxonomy. The dichotomy in the number of taxa that are

associated with below and above the MLD suggests that the conditions in the deep ocean select for a very distinct yet diverse bacterial community.

Location effects, although still significant, were less pronounced than the previous two groups. There was little difference between the coast and shelf stations so these two groupings were analyzed together. The conditions associated with the coast and shelf stations are more influenced by the cold Labrador Sea Current and can thus be expected to contain different bacterial communities than the open stations which are more influenced by the warm Gulf Stream. Additionally, areas on the Scotian Shelf, and particularly in the northeastern Scotian Shelf have been shown to experience a shorter, but more pronounced and productive spring bloom as compared to areas off the shelf (Zhai *et al.*, 2011). From the results of this analysis there is a definite link of spring associated taxa with coastal and shelf stations and fall associated taxa with open water stations. Taxa associated with the Scotian Shelf and the spring season are further expected to be involved with the spring bloom and associated with highly productive conditions. Alteromonadales, Bacteroidetes, *Ulvibacter, Polaribacter, Octadecabacter,* and Epsilonproteobacteria were all preferentially abundant in the spring and coastal/shelf stations. While *Prochlorococcus*, *Synechococcus*, *Alcanivorax*, and Vibrionaceae were all found to be preferentially abundant off the shelf in the fall.

The advantage of using a generalized linear model approach to determine significantly contributing taxa is the ability to detect significant differences in the composition of rare taxa such as Erythrobacteraceae and Epsilonproteobacteria which otherwise might be overlooked in other distance based approaches. Identifying and observing rare taxa is of increasing interest due to their extreme diversity, their roles in microbial communities, and their ability to act as markers for community shifts (Sogin *et al.*, 2006; Bachy and Worden, 2014; Shade *et al.*, 2014). Also of note in this analysis are that the most abundant groups, Pelagibacteraceae and Rhodobacteraceae were not significantly associated with any one group. This gives further support to their ubiquitous dispersal across the marine environment, however it is apparent that the groups still have preferred ecological niches and that there are still spatial patterns in their distribution (Figure 2.10, 2.11, and 2.12). Undoubtedly, more pronounced ecological niches would become apparent when looking at these groups at the taxonomic resolution of OTU.

2.4.4.2 Responses to Environmental Gradients

The heat map (Figure 2.15) displays significant correlations between environmental variables and bacterial taxa, and from this figure, it is obvious that not all environmental variables strongly affected all bacterial taxa. Chlorophyll and oxygen had strong relationships with many taxa, further supporting the importance of phytoplankton blooms and the associated productivity for bacterial taxa on the Scotian Shelf, as well as the importance of oxygen for the metabolic diversity of bacteria in the region. The majority of associations of taxa with chlorophyll and oxygen were negative in nature (Figure 2.15), and a taxon's response to these variables appeared to be one of the main indicators of cluster formation in the taxa-environment correlation analysis. Cluster 1, which responded positively to chlorophyll and oxygen, included *Ulvibacter*, *Polaribacter*, Alteromonadales, Epsilonproteobacteria, Octadecabacter, Bacteroidetes, Unassigned bacteria and Rhodobacteraceae. The first six of these eight taxa were also significantly associated with spring and coastal conditions, suggesting this cluster is preferentially found in the productive spring environment. In the correlation analysis, these groups were significantly positively correlated with chlorophyll, oxygen, ammonium, nitrite, latitude, and negatively associated with temperature, salinity, diversity, nutrients to some extent, and depth. The response of this cluster to bacterial concentration appeared to further divide the cluster into Rhodobacteraceae and Unknown bacteria, and the remaining six taxa that were significantly associated with coastal spring conditions.

The second cluster, composed of clusters 2.1 and 2.2, was larger and responded negatively to chlorophyll and oxygen. This group could further be broken down into taxa that were generally positively correlated with bacterial abundance (cluster 2.1) and taxa that were negatively correlated with bacterial abundance (cluster 2.2). Members of cluster 2.2 were also generally positively associated with salinity, diversity, nitrate and depth, whereas members of cluster 2.1 had mixed reactions to these variables. Only *Pseudoalteromonas* and SAR86 of cluster 2.2 were not found to be significantly associated with sites below the MLD, suggesting that in general this cluster contains deep water dwelling organisms. Of cluster 2.1, only Pelagibacteraceae and *Alteromonas* weren't significantly associated with the fall season, however both these taxa were on

average more abundant in the fall than the spring (Figure 2.6). This would suggest that cluster 2.1 contains taxa that are preferentially found in the fall. The divisions of these clusters then could be generally summarized based on a particular taxon's seasonal preference: spring (cluster 1) or fall (cluster 2.1), or based on their preference for deeper waters (cluster 2.2).

A broad taxonomic resolution was chosen for the majority of analyses in this study to give a general overview of spatial trends across the many samples over the shelf. This broad taxonomic resolution however cannot account for the intricate patterns and relationships that are undoubtedly associated with OTUs. Because of this, certain taxonomic groupings containing OTUs with distinctly different ecological niches and responses to environmental variables can confound or conceal environmental effects. More research is needed to consistently return precise and accurate taxonomic annotations of OTUs so that high quality, high resolution comparisons between studies can be made, and to determine how each of these OTUs responds to environmental variables. Marine bacterial community structure is a complex mix of intricate microscale associations and macroscale processes and a cohesive understanding at every scale is needed to generate an accurate model of the global ocean's microbiome.

2.4.4.3 Co-occurrence and Exclusion Patterns of Bacteria in the Scotian Shelf

Network analysis has been used as a means to visualize co-occurrence or co-abundance patterns in macroecological studies as well as microecological studies (Chaffron *et al.*, 2010; Faust and Raes, 2012; Berry and Widder, 2014; Lima-Mendez *et al.*, 2015). The observation of non-random co-occurrence patterns suggests species interactions are shaping community structure together with abiotic forces. Observing bacterial co-occurrence patterns allows for the visualization of clusters of bacteria that are often detected together or those that are mutually exclusive, which can provide information on a number of biological principles. Although networks are useful in visualizing intricate patterns between species they can often be difficult to interpret without having causal evidence for the functionality of the observed interaction (Faust and Raes, 2012). For instance, a positive interaction between two taxa could be due to a number of ecological reasons including syntrophy, co-aggregation in biofilms, co-

colonization, or niche overlap. Alternatively, negative relationships may result from amensalism, non-overlapping niches, prey-predator relationships, or competition, among other reasons (Faust and Raes, 2012). Additionally, relationships between taxa could lag in time and thus be difficult to recognize in spatial studies. As microbial communities are incredibly complex and difficult to observe in real time, uncovering causal evidence for species interactions can be extremely challenging, and often requires parallel techniques such as FISH (Fluorescence in situ hybridization) microscopy (Lima-Mendez *et al.*, 2015). Despite these caveats, microbial network analyses can be used to decipher underlying patterns in the data as well as intricate interactions between species on a finite scale.

In this study, we identified highly significant pairwise co-occurrence and mutual exclusion relationships between taxa. The resulting network was coloured based on either taxonomy or habitat preference through indicator species analysis. Four core clusters formed in the main network component and these clusters could quite easily be explained by the indicator species characterization of their member OTUs. OTUs preferentially found in the spring or spring and coast were mutually exclusive from OTUs preferentially found in the fall and in the open ocean. OTUs found in the deep also clustered together and formed positive interactions with OTUs found in the open ocean. The high degree of connectivity in this network (on average nodes were connected to 26.1 other nodes) suggests that there are specific groups of OTUs that are found consistently together (or apart). These groups may share the same ecological niche or directly rely on each other for metabolites. Further experiments would have to be carried out in order to discern the nature of these interactions.

Network topology describes the physical connective properties of the network. In this study, the average path length was 2.91 and the clustering coefficient was 0.657, suggesting a very highly connected network, i.e. many significant associations between the OTUs in the region. These values are on par with other bacterial network analysis studies. Specifically, a recent network analysis using taxa across many domains of life conducted by the TARA expeditions had an average clustering coefficient of 0.229 and an average path length of 3.43 (Lima-Mendez *et al.*, 2015). A human microbiome study reported an average path length of 3, and a cluster coefficient of 0.1 (Faust *et al.*, 2012),

and a global network analysis across an extensive array of habitat sites had a clustering coefficient of 0.501 and an average path length of 6.30 (Chaffron *et al.*, 2010). Our analysis reported a higher value for clustering coefficient and a lower average path length suggesting a higher degree of connectivity amongst species in the Scotian Shelf. The higher degree of connectivity of OTUs in the Scotian Shelf could be explained by the fact that the Scotian Shelf is a more uninterrupted environment than the body, and as a region it covers a smaller distance and habitat range than the other two global network studies. Networks with small average shortest path lengths are also known as 'small-world' networks and communities with small average path lengths are proposed to be better able to respond to perturbations in the environment (Watts and Strogatz, 1998; Zhou *et al.*, 2010). Also the network heterogeneity for this region was high (0.745) suggesting the network has a greater tendency to contain hub nodes, or species with a high degree of connectivity. These hub species could be "keystone species", and thus have critical roles in the functioning of their ecosystem perhaps through predation, essential metabolite production, or nutrient cycling (Dong and Horvath, 2007; Faust and Raes, 2012). Additionally, many of these highly connected, potential hub species had a low overall abundance suggesting a disproportionate relationship between abundance and community interaction or involvement (Shade *et al.*, 2014; Lynch and Neufeld, 2015). More research on the function of these rare OTUs in their community and the nature of their many interactions with other species is needed.

The underlying distribution of nodes in this network relates to niche preference rather than phylogenetic relatedness, with the main clusters dividing samples based on spring, fall, fall and open ocean, and deep water niche preferences (Figure 2.16). This patterning of samples, taxa and OTUs based on their association with the spring, the fall, or the deep has been identified many times in this study (Figures 2.13, 2.14, and 2.15), and further supports the main findings that seasonality and depth are the strongest factors affecting community structure.

2.4.5 Thebaud Platform Station – Comprehensive analysis of the bacterial community surrounding an offshore oil and gas platform

The final analysis of this chapter was preformed to determine how the bacterial community surrounding the Thebaud Platform (TB) Station differs from other similar regions along the shelf. The TB Station is a gas field located approximately 9 km southwest of the western end of Sable Island. The field was discovered in 1972 and has been delineated with three additional wells. The TB field produces gas for the Sable Offshore Energy Project's Tier 1 phase. Additionally, the project's central processing complex is located adjacent to the TB well jacket and collects gas from two other fields, the Venture and the North Triumph which are connected to the complex by subsea gathering pipelines. During the spring cruise aboard the *Hudson,* a sample was taken from a site labelled "TB" about 10 km southwest of the platform. This sample was chosen for further analysis at a higher taxonomic resolution to identify any key bacterial genera that were found at this oil platform site preferentially to other similar sites along the shelf region. A similar generalized linear model approach to that described in section 2.4.4.1 was used to determine statistical differences between the TB station and other shelf stations in the vicinity (Figure 2.19). The OTUs that showed the strongest association to the TB station belonged to the taxa *Polaribacter* with 0.1% relative abundance on the shelf and 4% relative abundance at the TB station; Flavobacteriaceae with 0.7% relative abundance on the shelf and 2.6% relative abundance at the TB station, Colwelliaceae with 0.03% relative abundance on the shelf and 2% relative abundance at the TB station, Rhodobacteraceae with 0.1% relative abundance on the shelf and 1% relative abundance at the TB station, and Oceanospirillales with 0.2% relative abundance on the shelf and 1% relative abundance at the TB station. These groups have been shown previously to be associated with oil spills, to have oil degrading properties and specifically, were discovered to be enriched during the recent Horizon oil spill disaster in the Gulf of Mexico (Brakstad *et al.*, 2008; Vila *et al.*, 2010; Yeung *et al.*, 2011; Redmond and Valentine, 2012; Rivers *et al.*, 2013; Joye *et al.*, 2014; Mason *et al.*, 2014).

The results of this analysis suggest that there is an effect of the offshore oil platform on the bacterial community even at a distance of 10 km. Yeung et al. (2011) studied the produced water of the Thebaud platform in addition to the sediment surrounding it at a small spatial scale using DGGE analysis. Yeung et al. (2011) concluded that there was a difference in the bacterial community structure of the

sediment near the platform, but not of the seawater surrounding the platform. Specifically, in the produced water they found a bacterial community mainly made up of the genera of *Acinetobacter* and *Geobacillus* from the taxa of Gammaproteobacteria and Firmicutes respectively, and they found unique sequences belonging to *Sulfurovum* and *Arcobacteria* of the Epsilonproteobacteria group in sediment close to the platform. Our study uses a sequencing technique that reveals more information about the rarer bacterial OTUs in the dataset, which could potentially explain why we found a significant difference between the bacterial community of the sea water surrounding the TB station and the bacterial community of other stations on the shelf, while the study by Yeung et al. (2011) did not. Regardless, our findings suggest that more research is needed to truly understand the effect that offshore drilling has on marine bacterial communities in the vicinity of these offshore oil platforms and to determine the spatial extent of this effect, especially as oil exploration in the area continues to develop.

2.4.6 Conclusions

This study provides the first extensive bacterial community survey of the Scotian Shelf in the Northwest Atlantic Ocean. In addition to collecting information on this region, this study also integrates the dimensions of space, depth and time and attempts to holistically determine the factors driving changes in microbial community structure and biodiversity.

The most abundant taxa in the Scotian Shelf include the groups of Pelagibacteraceae, Rhodobacteraceae, and Bacteroidetes, although these groups varied widely in dominance across sites. Also, community diversity in the spring was found to significantly correlate with temperature, salinity, chlorophyll, oxygen, nutrients, bacterial concentration, latitude, and depth, but only a significant trend with salinity was found in the fall. Overall, the effect of environmental distance at this regional scale had a stronger influence on community structure than geographic distance. This supports a theory of contemporary selection acting on a generally cosmopolitan distribution of taxa, in which environmental gradients establish the conditions to create partitioning of taxonomic groups into their respective ecological niches. The strength of the relationship between the environmental variables measured and the bacterial community seems to be slightly stronger in the spring than in the fall and in shallower waters compared to deeper waters.

With all sites considered, oxygen had the greatest correlation with community similarity, followed by temperature and phosphate. Community similarity in spring and fall sites separately showed varying responses to environmental gradients with nitrate, phosphate and oxygen strongly correlated with community similarity in the fall, and bacterial diversity, nitrate, and oxygen highly significantly correlated with community in the spring. When observing sites further separated based on discrete depth and season categories, the response to environmental gradients was even more variable, with the 250 m depths from each season generally responding to different variables than the upper three depths.

The main results from this study demonstrate that alpha diversity, beta diversity, and composition of bacterial communities on the Scotian Shelf are significantly affected by seasonality, depth, and to a lesser extent, location. Based on bacterial community similarity, sites clustered into core fall, spring, and deep water clusters. This pattern was also apparent in describing the main clusters of OTU co-occurrence networks and the main clusters of bacterial taxa assembled on their responses to environmental variables. Although season and depth could explain the central bacterial assemblages, there were more subtle responses of specific bacterial taxa to each of the environmental variables tested that resulted in different patterns within each assemblage. Furthermore, this work provides insight into the distribution and niche preferences of many understudied OTUs and even understudied phyla of marine bacteria, and provides a baseline measurement of the bacterial community with which to compare future studies of the region.

Chapter 3: Investigating Bacterial Communities through Frequent Sampling of Deep Waters in the Bedford Basin, Canada

3.0 Abstract

The majority of bacterial time series studies sample sites monthly or quarterly, and since bacteria are known to have very high turnover rates these time series may miss out on higher resolution bacterial dynamics. In this work we expanded upon a weekly time series in the Bedford Basin by introducing bacterial community measurements using 16S rRNA sequencing. This chapter focused on analyzing the first year of data from the bottom depth (60 m) sampled. Overall, we found that the 60 m depth experienced seasonal patterns over the year with respect to bacterial composition, diversity, and community similarity. In particular, bacterial diversity peaked in the late fall and early winter, and samples taken 6 months apart showed the greatest community dissimilarity. Also, silicate and oxygen concentrations were the most strongly correlated environmental variables to community similarity over the entire year. Alphaproteobacteria, specifically Rhodobacteraceae and Pelagibacteraceae, were the most common bacterial taxa, and exhibited distinct seasonal patterns with Rhodobacteraceae most common in the spring and summer, and Pelagibacteraceae most common in the fall and early winter. Weekly sampling highlighted the extent of bacterial dynamics in the Bedford Basin, with many normally rare bacterial OTUs spiking in abundance during certain weeks. The genus *Pseudoalteromonas* showed the most extreme example of this spiking trend, and became the most dominant taxa in the community for a one week period. Two main intrusion events were observed during the year of sampling and the bacterial responses to these abrupt environmental changes showed distinct partitioning between pre and post intrusion communities. Overall, this weekly time series emphasized the need for more studies on highly dynamic, conditionally rare bacterial communities, in addition to initiating the beginnings of a bacterial community time series that will continue to produce high quality information on the bacterial structure of the Bedford Basin into the future.

3.1 Introduction

Oceanic bacterial communities are extremely diverse and dynamic, and their composition can have dramatic effects on both the physical environment and the local biological system. Changes in environmental conditions such as fluctuations in temperature, light intensity, and nutrient concentrations can significantly alter marine bacterial community composition (Fuhrman *et al.*, 2015). In regions that experience profound seasonal changes the bacterial community can exhibit characteristic shifts and cycles in structure and abundant taxa (El-Swais et al., 2014; Gilbert et al., 2012) Through continuous monitoring, a "baseline" of microbial community structure can be established (Fuhrman *et al.*, 2006; Karl and Church, 2014). By establishing this baseline, deviations from the norm can be identified through continuing observations and the sources of anomalies can be explored. After defining the characteristic bacterial community structure throughout the year, anomalies in abundance of certain bacterial taxa can be identified and the resulting implications can be assessed. Long term time series are integral for investigating the perturbations in microbial community structure that can be expected from short term weather events and long term climate change (Vezzulli *et al.*, 2012; Karl and Church, 2014).

Coastal regions of the world's oceans are responsible for approximately 19% of oceanic net primary productivity and are thus regions of intense interest both economically and environmentally (Field et al, 1998). In addition, 44% of the global population lives within 150 km of the coast which only further adds to the importance of this region (UN Atlas of the Oceans). The ongoing study site for this research is the Bedford Basin, which is a large (17 km^2) enclosed bay forming the northwestern end of the Halifax Harbour. The basin is 71 m at its deepest point and is connected to the adjoining Scotian Shelf through a narrow (400 m) and shallow sill (20 m). The Bedford Basin becomes thermally stratified periodically throughout the year and due to the isolation of deep waters caused by the sill, it can experience hypoxic conditions at depth (Figure 3.1) (Punshon and Moore, 2004; Fader and Miller, 2008). Interestingly, N_2O production has been measured in the Bedford Basin at times of extreme hypoxia suggesting dynamic metabolic changes within the microbial community in response to environmental fluctuations (Punshon and Moore, 2004). In addition, the basin exhibits

strong seasonal variations in temperature, productivity and nutrient concentrations (Li and Harrison, 2008). It has also been shown to mirror the marine seasonal and intra-annual patterns that the coastal shelf water exhibits (Li *et al.*, 2006) and due to its close proximity to BIO and Dalhousie University it is very accessible. Since 1991, BIO has extensively monitored the Bedford Basin looking specifically at factors affecting the plankton ecosystem from the Compass Buoy station (44° 41' 37" N, 63° 38' 25" W) (Li and Dickie, 2001). Due to its geographic location and the extensive monitoring program, the Bedford Basin provides an excellent platform for the study of variations in microbial community structure over time, both seasonally and annually.

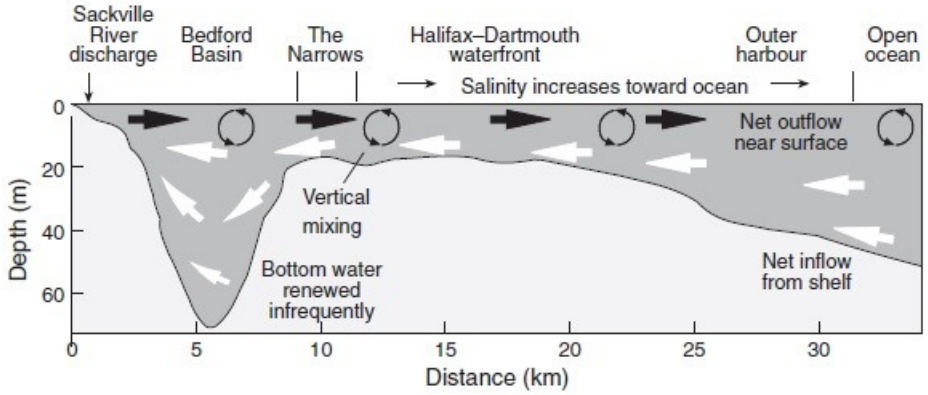


Figure 3.1. From Fader and Miller 2008. A simplified cross-sectional depiction of water circulation in Halifax Harbour from Bedford Basin to the harbour mouth.

In this work we expand upon the Bedford Basin time series, conducted for the past 20 years by BIO, by introducing microbiologically focused weekly sampling. From the weekly samples we collect, we use next generation high throughput sequencing of the 16S rRNA gene to determine bacterial community structure. Recently, a historical six year time series was conducted by El Swais et al (2014) in the Bedford Basin using fixed 1 mL samples from 5 m depth. One year in this time series was sampled biweekly and the other 5 years were sampled quarterly throughout the year. The current study adds to and increases the resolution of this dataset through the collection of fresh samples each week from four depths spanning the water column, and through the use of larger volumes of water that will better represent the microbial community and allow for higher taxonomic resolution. The analyses presented in this chapter will focus primarily on the bacterial community dynamics at 60 m depth. Specifically we want to observe how community

diversity and similarity change over time and with environmental gradients, in addition to identifying taxa and species specific patterns in temporal distribution.

3.2 Materials and Methods

3.2.1 Sample Collection

Water samples were collected weekly by BIO from the Compass Buoy Station (44° 41' 37" N, 63° 38' 25" W) in the Bedford Basin from January 15th 2014 until December 17th 2014. 500 mL of water from four depths (1 m, 5 m, 10 m, and 60 m) were filtered each week onto 0.2 µm Isopore filters (Millipore). These filters were then flash frozen and stored at -80˚C until needed. In total, 49 weeks were sampled which resulted in 196 samples across all four depths.

Water was also collected each week and at each depth for cell counts using flow cytometry analysis. The water was fixed with paraformaldehyde (Alfa-Aesar) to a final concentration of 1%, left to incubate at room temperature for an hour, and then immediately flash frozen and stored at -80°C until analysis. Prior to analysis, samples were thawed, stained with SYBR dye (Invitrogen) to a final concentration of 1:10,000, and allowed to incubate at room temperature for twenty minutes. The Accuri C6 Flow Cytometer (BD Biosciences) was used for flow cytometry analyses. A fluorescence threshold of 800 was set for the FL1 filter for bacterial counts, and gates designed to be used with the Accuri Flow Cytometer were obtained from the manufacturer (Gatza *et al.*, 2013).

Weather (wind, rain, and cloud coverage) data for the year of 2014 was obtained from an online weather mapping source (Wind Guru, 2014). All weather measurements taken over a day were averaged and then all days from the week prior to the sampling date were averaged. The stratification index was calculated as specified by BIO (Johnson *et al.*, 2014):

$$Strat_{Ind} = (\sigma_{t\text{-}60} - \sigma_{t\text{-}zmin})/(60\text{-}z_{min})$$

Where $\sigma_{t\text{-}60}$ and $\sigma_{t\text{-}zmin}$ are density values (σ_t) at 60 m and z_{min}. z_{min} corresponds to the average of the three upper depths measured in this study, 1 m, 5 m, and 10 m.

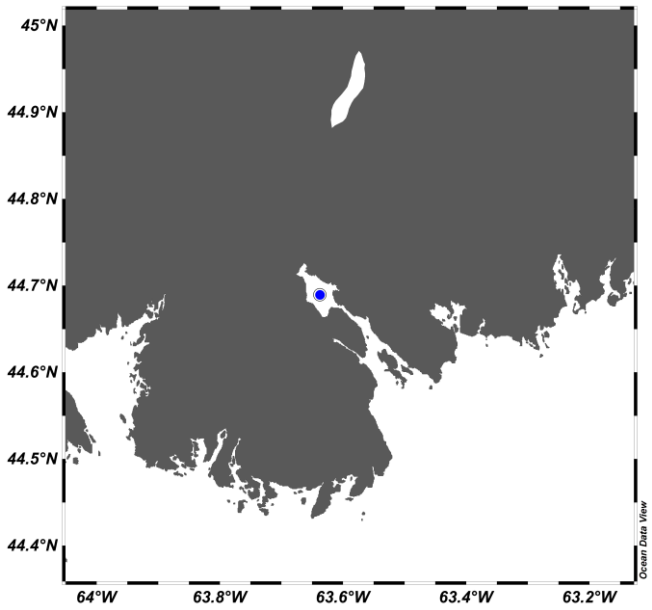


Figure 3.2. Compass Buoy sampling station (blue dot) in relation to the Halifax Peninsula and the Atlantic Ocean.

3.2.2 DNA Extraction

All DNA was extracted from the 0.2 µm polycarbonate filters using the DNeasy Plant Mini Kit from Qiagen according to the manufacturer's instructions with some alterations in the cell lysis procedure. 50 µL of lysozyme (5 mg/ml) (Fisher BioReagents) was initially added to each filter sample and each sample was vortexed on high for 30 seconds. Then 400 µL of lysis buffer AP1 (Qiagen) was added to each sample tube followed by the addition of 45 µL of proteinase K (20 mg/mL) (Fisher BioReagents). The samples were then incubated at 55˚C with shaking for one hour. After this incubation, 4 µL of Rnase A was added to the samples which were then kept on ice for 10 minutes. From here the extraction followed the instruction manual and in the last step, 100 µL of DNA was eluted. DNA concentrations and purity were checked with a NanoDrop 2000 (Thermo Scientific).

3.2.3 Amplification and Library Preparation

The sample preparation procedure for the MiSeq sequencer (Illumina) is the same as described in the Microbiome Amplicon Sequencing Workflow (Comeau and Langille, unpublished, also at cgeb-imr.ca/protocols.html). Briefly, each sample of extracted DNA

was amplified using dual-indexing Illumina fusion primers that targeted the V6-V8, 438 bp region, of the bacterial 16S rRNA gene (Comeau et al. 2011; Comeau and Langille, unpublished) (Supplementary Table S2). Each sample was amplified in duplicate, with one duplicate containing the original extracted DNA template and the other containing a 1:10 template dilution. The dilution series was done to reduce the effects of PCR bias. The pooled duplicate PCR product quality was verified using an E-gel 96-well high-throughput system (Invitrogen). Library normalization and PCR clean-up was conducted using a SequalPrep 96-well Plate Kit (Invitrogen). After normalization, all samples were pooled together and the final library pool was quantified using Qubit with PicoGreen (Invitrogen).

3.2.4 Illumina MiSeq Sequencing

After quantification, the pooled samples were run on an Illumina MiSeq Sequencer using paired-end 300+300 bp v3 chemistry. Approximately 14,304,947 sequences were obtained, of which 12,565,285 passed the quality filter. The MiSeq on-board software demultiplexed the reads by trimming the barcode sequences from the sequence reads and assigning each read to the appropriate sample. The number of sequences and quality control information at each step of the clean-up process can be seen in Supplementary Table S3.

3.2.5 QIIME 16S rRNA Data Analysis

All preliminary analyses and processing of 16S sequences followed a QIIME version 1.8.0 (Caporaso, Kuczynski, *et al.*, 2010) pipeline workflow (Langille, github.com/mlangill/microbiome_helper). The program PEAR version 0.9.6 was first used to merge the demultiplexed, paired-end sequences together (J. Zhang *et al.*, 2014). After merging paired ends, sequences shorter than 400 bp in length or with a quality less than 30 at over 90% of the bases were discarded. Chimeric sequences were removed using UCHIME (Edgar *et al.*, 2011). OTUs were picked based on 97% sequence similarity using sortmerna (Kopylova et al., 2012) for reference picking and SUMACLUST (Mercier *et al.*, 2013) for *de novo* OTU picking (i.e. “open-reference”

picking, referred to as "New.ReferenceOTU" in results). This process uses the reference Greengenes database version 13.8 (McDonald *et al.*, 2012) for preliminary OTU picking and then subsamples failed sequences using *de novo* picking. From this point on, only OTUs with 2 or more sequences (no singletons) that aligned with PyNAST (Caporaso, Bittinger, *et al.*, 2010) were used for further analysis. Additional quality control measures included removing all sequences belonging to Archaea and chloroplasts from the dataset as the primers used in this study only amplify bacterial sequences reliably.

In order to compare relative abundances of OTUs between samples and depths, sequences were rarefied to a sequencing depth of 9900 which corresponded to the sample across all depths with the lowest amount of sequences. Depending on the purpose of the analysis, i.e. whether it was done to observe fine scale or broad scale patterns, OTUs or bacterial family or class were used as the taxonomic level for resolution.

3.2.6 Statistical Analysis and Data Visualization

The temporal breakdown for season was based on the traditional solstice and equinox classifications, and this translated to weeks 3-12 for winter, weeks 13-25 for spring, weeks 26-38 for summer, and weeks 39-51 for fall.

Network diagrams were created using Cytoscape v 3.2.1 (Shannon *et al.*, 2003), and all other figures were made using the *ggplot2* package in R (Wickham and Chang, 2015), and Ocean Data View version 4.*6.5* (Schlitzer, 2015). All statistical analyses were conducted using R version 3.2.1 (R Development Core Team, 2015). A Loess regression smoother was added to a few temporal plots in order to highlight the general trends in the data. The principal mechanism of a Loess regression is to fit simple models to localized subsets of data, negating the need for a global function for the entire data set, and instead creating local functions to fit segments of data (Jacoby, 2000). Loess regressions are especially helpful when dealing with non-linear relationships between variables (e.g. annual cycles). Shannon diversity was calculated in QIIME, and Bray-Curtis beta diversity was calculated between samples using the R package *vegan* (Dixon, 2003). NMDS ordination was also conducted using the R package *vegan* and all abundance data was transformed using a Hellinger transformation prior to ordination. Mantel and Partial Mantel tests were conducted to determine the factors driving community similarity in the

Bedford Basin. A Bray-Curtis dissimilarity matrix was generated using Hellinger transformed OTU abundance data from all OTUs that reached above 1% community abundance at one point during the year. Environmental matrices were made using Euclidean distances of the scaled variables temperature, salinity, oxygen, chlorophyll, nitrate, phosphate, silicate, nitrite, ammonium, bacterial diversity, bacterial concentration, wind, rain, and day length. Each of these variables was also transformed into an individual matrix and separately analyzed in individual tests with the abundance matrix. A temporal matrix was also generated using weeks between samples. Partial Mantel tests were conducted controlling for the temporal matrix, and were compared to standard Mantel tests to determine the effect time had on each of the variables with respect to its effect on community similarity.

Indicator species for each season, and combinations of seasons, were identified using the *multipatt* function from the *indicspecies* package in R (De Cáceres *et al.*, 2010). Correlation networks were generated using only highly significant correlations of species that represented on average over 0.1% of community composition at 60 m. High significance was defined as a Spearman's rho >0.7 with a *p* value less than 0.01. Environmental variables were also included. The method of node placement used for network figures aimed to highlight clusters of closely correlated OTUs and variables.

To identify pulsing or conditionally rare OTUs, the criterion of a change of 5% of total community abundance from one week to the next was used. Additionally, the largest weekly percent change perceived had to be at least 5x the average abundance of that OTU throughout the year. As an example, if an OTU was 2% of relative abundance one week, 7% of relative community abundance the following week, and had an average abundance across all samples at that depth of 0.5% it would be classified as a conditionally rare OTU. To identify OTUs that changed rapidly in response to the two deep water intrusion events the criterion of either a 5% absolute change or a 2% absolute change in community relative abundance from the weeks surrounding the intrusion was used for the first and second intrusions respectively. A lower criterion for the second intrusion event was used because the bacterial dynamics around this intrusion were not as extreme.

3.3 Results

3.3.1 Environmental Parameters

The Bedford Basin displays a great degree of seasonality in physiochemical factors throughout the water column (Figure 3.3). The temperature in the surface waters ranged from 0.29°C in February to 19.2°C in August. At 60 m depth the temperature range was less extreme but still exhibited seasonal changes, with a minimum temperature of 1.35°C in the spring and early summer season and a maximum of 5.25°C in the late fall and early winter. Salinity remained fairly constant within the basin, with some small intrusions of fresher waters at depth in the spring. Additionally there were numerous spikes in fresh water at the surface. Oxygen concentration at depth was generally much lower than oxygen concentration at the surface, and showed high seasonal variability. The minimum values at 60 m reached as low as 1.09 mL/L (~47 µmol/kg) at the very beginning and end of the year and the maximum values reached 6.7 mL/L (287 µmol/kg) at the winter to spring transition. Bacterial counts in the basin varied greatly throughout the year with maximum concentrations around 10 million cells/mL in the surface waters in late August and early September. On average there were about 2.4 million cells/mL in the upper three depths (1, 5, and 10 m) and 1.6 million cells/mL at 60 m.

At 60 m there were two unusual events characterized by abrupt changes in temperature and oxygen levels. The first dynamic period occurred from weeks 10-11 (March 5-12) and triggered the first phytoplankton bloom of the year at the surface as well as a dramatic change in bacterial composition throughout the water column. At 60 m depth, the transition between these two weeks brought a temperature reduction of 1.89°C, an increase in oxygen concentration from 3.6 mL/L (~154 µmol/kg) to 6.32 mL/L (~271 µmol/kg) and a reduction in nitrate from 15.26 mmol/m^3 to 9.15 mmol/m^3. Chlorophyll concentrations at depth also reached their highest levels throughout the year of 1.7 mg/m^3 at week 12, and spiked dramatically from 0.11 mg/m^3 at week 10 to 0.83 mg/m^3 at week 11. In the surface waters, chlorophyll went from 3.82 mg/m^3 in week 10 to 17.67 mg/m^3 in week 11, with a corresponding dramatic decrease in nitrate from 7.81 mmol/m^3 to 1.89 mmol/m^3. A second event from weeks 27-29 (June 30th - July 14th, 2014) saw a depression in oxygen levels in the summer that reduced concentrations to 2.88 mL/L

(~123 µmol/kg) at week 28 from 4.16 mL/L (~178 µmol/kg) at week 26, which promptly returned back up to 6.11 mL/L (~262 µmol/kg) oxygen at week 29. Temperature during this time went from 1.61°C to 3.86°C, nitrate dropped from 9.37 mmol/m^3 to 2.95 mmol/m^3, with a corresponding decrease in ammonium from 12.84 mmol/m^3 to 3.96 mmol/m^3. Chlorophyll concentrations at depth also fell dramatically at week 29 and 30, from a small peak above 1 mg/m^3 for weeks 25-28. The bacterial dynamics in response to these two deep water mixing events will be analyzed further in section 3.3.5.

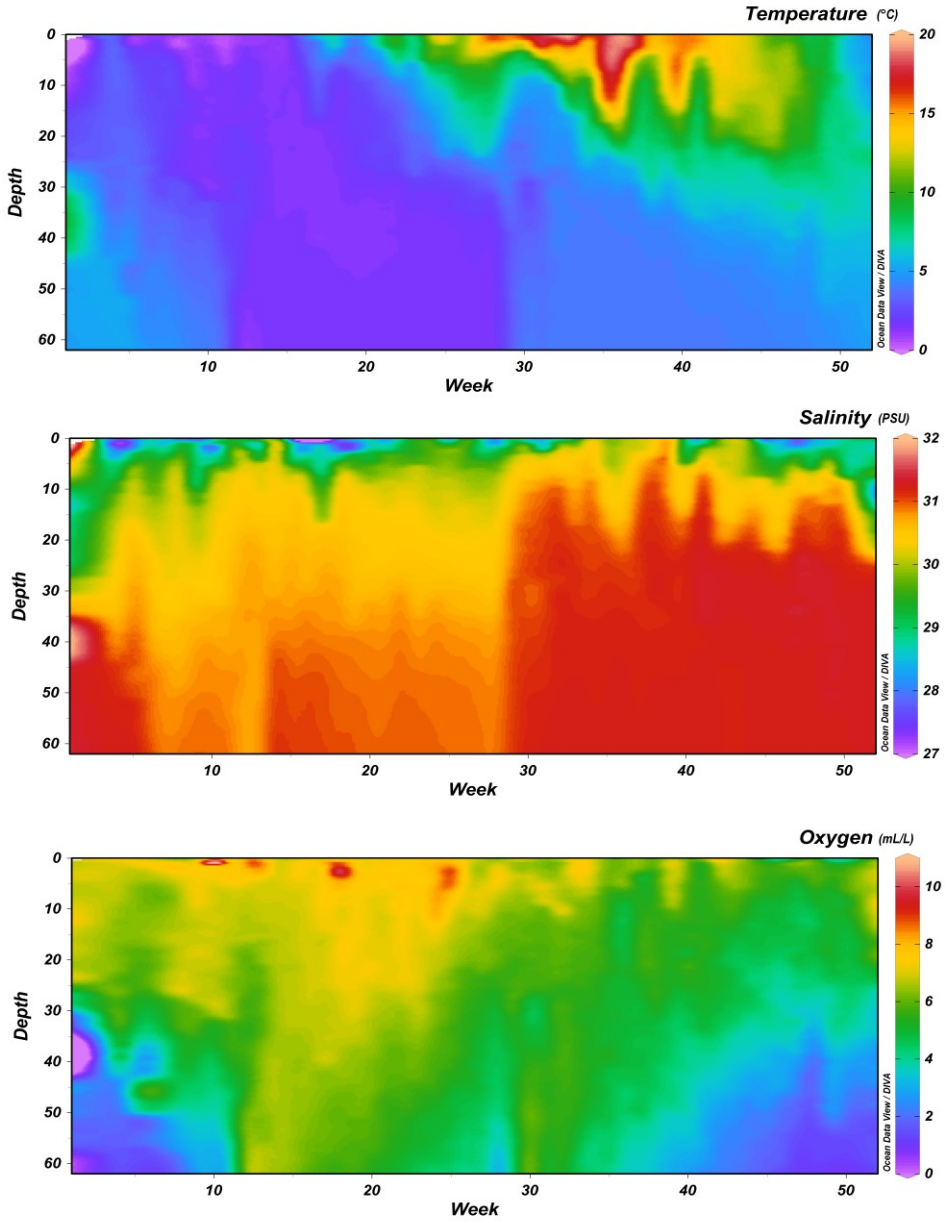


Figure 3.3. Salinity (PSU), Temperature (°C), and Oxygen (mL/L) throughout the year in the Bedford Basin. CTD data courtesy of BIO.

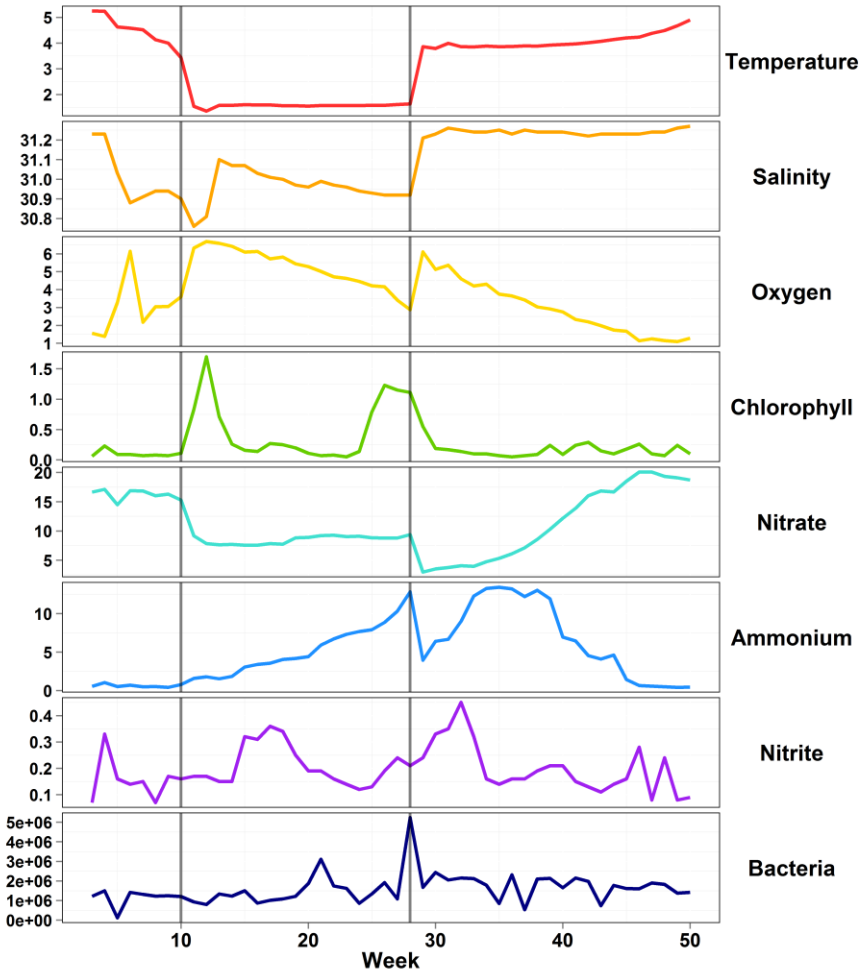


Figure 3.4. Environmental parameters at 60 m depth over the course of a year. Temperature in °C, salinity in PSU, oxygen in mL/L, chlorophyll in mg/m^3, nitrate, ammonium, and nitrite in mmol/m^3, and bacteria in cells/mL. Vertical lines represent intrusion events to be analyzed further in section 3.3.5.

Figure 3.4 shows the weekly variations in environmental parameters at 60 m depth in the Bedford Basin over the year. The vertical lines on Figure 3.4 represent the timing of the two intrusion events described above, which are mirrored in the temporal patterns of the environmental variables, especially temperature. After the second intrusion event, the bottom water does not undergo mixing for the rest of the year. This resulted in a gradual decrease of oxygen to almost hypoxic levels, a corresponding gradual increase in nitrate, and a decrease in ammonium, while temperature, salinity and chlorophyll remain unchanged.

3.3.2 Bacterial Community Analysis at 60 m – Alpha Diversity

When analyzing all depths, 11,294 OTUs were identified after rarefaction of all samples to 9900 sequences. Three OTUs, 637092, 696544, and 645011 were identified to some extent in every sample taken. 637092 is a member of the family of Pelagibacteraceae, and 696544 and 645011 are from the family of Rhodobacteraceae (the genus of *Phaeobacter,* and an unknown genus respectively). The remainder of this chapter will focus on the bacterial community at 60 m depth.

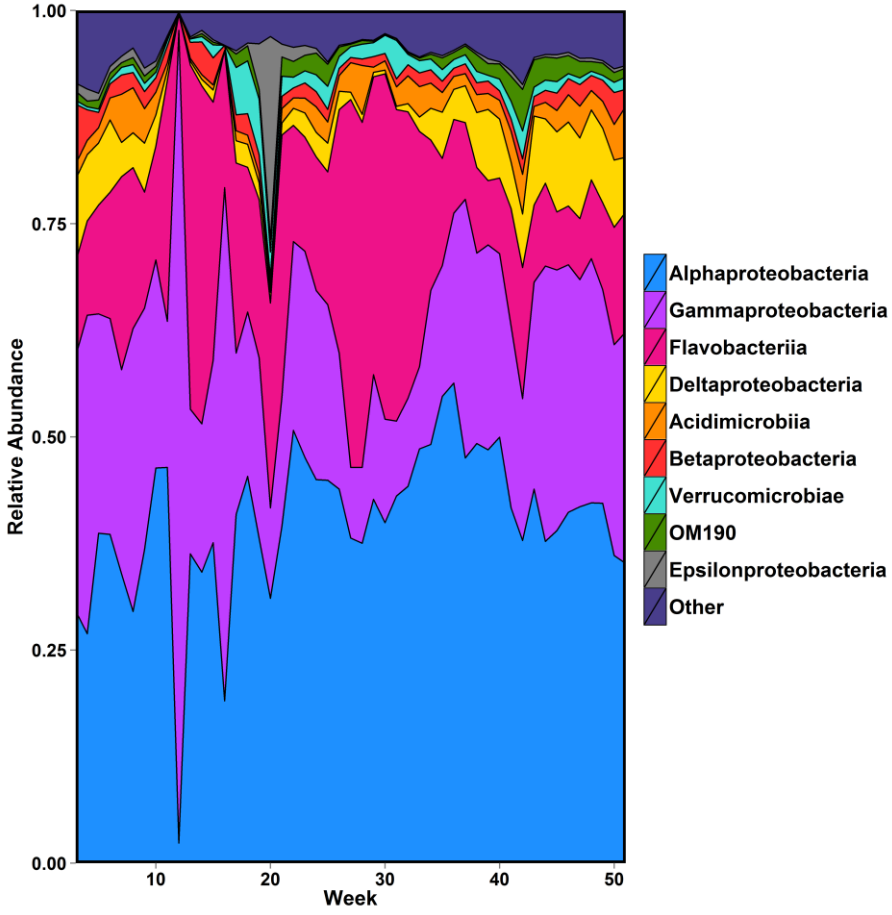


Figure 3.5. Abundance of bacterial taxa at the taxonomic level of class over the year at 60 m depth. The 9 bacterial classes shown in this figure represent on average 95% of the community abundance with all remaining classes pooled into the "Other" group.

The relative abundances of bacterial taxa grouped to the level of class over the year at 60 m are shown in Figure 3.5. Alphaproteobacteria represented the most abundant class, with an average of 40.2% of the annual community.

Gammaproteobacteria made up the next largest class with 23.6% of sequences, then Flavobacteriia with 18.9% of sequences. Together these three classes made up more than 80% of community abundance, and only 9 classes of bacteria were required to make up 95% of community abundance. The remaining bacterial classes ("Other") made up on average less than 5% of the annual community however some groups peaked to a greater degree of dominance at certain times of the year.

At 60 m there were 10 OTUs present to some extent every week sampled, out of the 8783 OTUs identified at this depth. These OTUs accounted for on average between 0.2% and 12.4% of average relative abundance. The OTUs (637092, 645011, 774258, 696544, 148145, 630330, 703006, 630967, 540364, and 316226) represented the families of Pelagibacteraceae, Rhodobacteraceae, Flavobacteriaceae, Halomonadaceae (order of Oceanospirillales), OM60 (order of Alteromonadales), SAR86, and HTCC2188 (order of Alteromonadales). The OTUs corresponding to the families of OM60 (630330), SAR86 (540364), and Rhodobacteraceae (316226) didn't reach above 1% relative abundance at 60 m depth all year, suggesting they are generalists that favour a low abundance lifestyle. A ubiquitous OTU (637092) from the Pelagibacteraceae family was the most abundant OTU at 60 m, representing on average 12.4% of sequences each week, with a minimum abundance of 0.01% community sequences on April 16th and a maximum abundance of 32.2% of community sequences on September 3rd. All other OTUs at 60 m depth represented less than 5% of relative abundance on average. Only 11 OTUs averaged more than 1% of relative abundance at this depth throughout the year. These OTUs (637092, 804483, 645011, 583880, 804449, 827726, 272142, 659486, 28929, 814290, 535135) are members of the families of Pelagibacteraceae, Rhodobacteraceae, SUP05, Pseudoalteromonadaceae (order Alteromonadales), Flavobacteriaceae, ZA3409c (order Acidimicrobiales), and Alteromonadaceae. At 60 m, the OTU 827726 corresponding to the *Pseudoalteromonas* genus showed the highest peak of abundance of any OTU at 63.8% on March 19th. The OTU also peaked to 30.2% of community relative abundance on April 16th but was generally rare for the rest of the year (0.9% on average). The top 244 OTUs accounted for 85% of the average relative abundance across samples, and only the top 26 OTUs were needed to account for 50% of the average relative abundance. 112 OTUs reached higher than 1% relative abundance in any sample at 60 m, meaning that

the remaining 8671 OTUs never reached an abundance greater than 1% of their community. This highlights the diversity and sheer number of rare OTUs, which are often overlooked when considered singly, but contribute significantly to the diversity of a community overall. Rare taxa are addressed further in section 3.3.4.

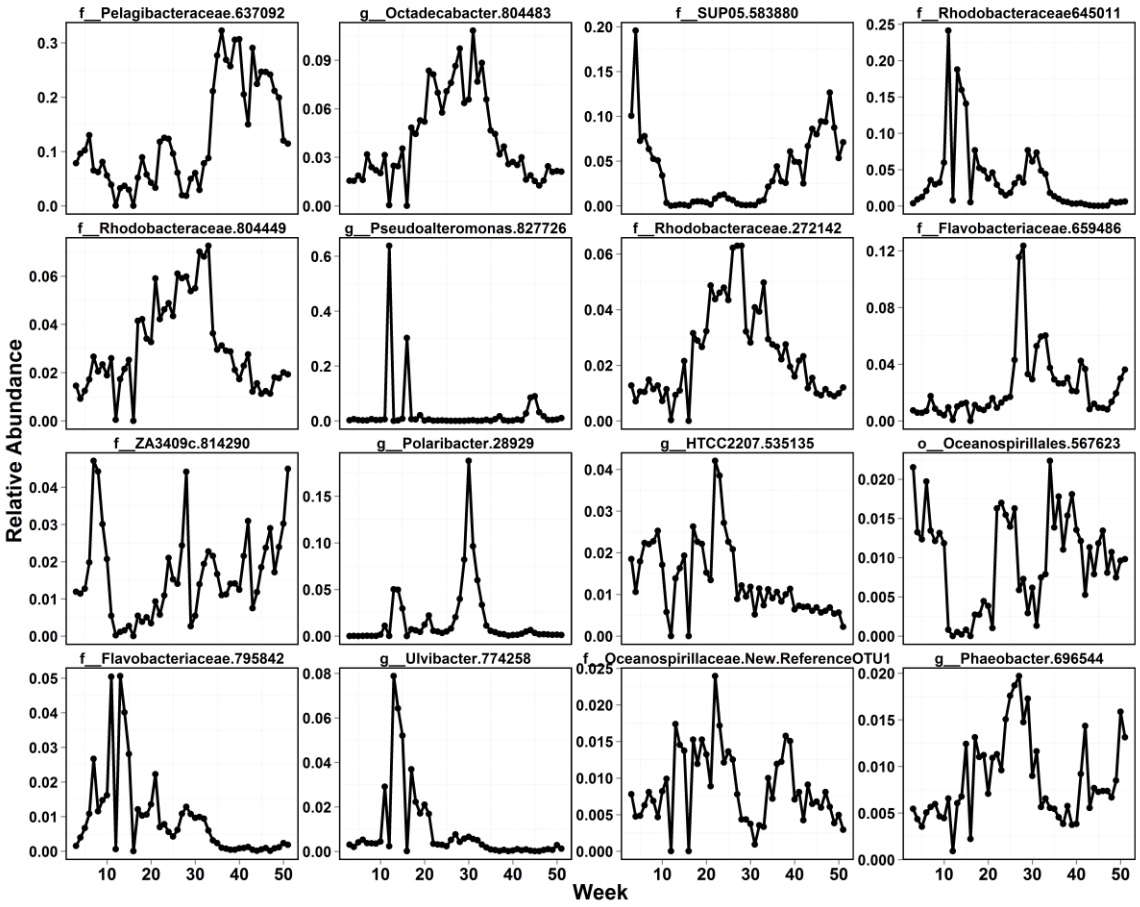


Figure 3.6. The 16 most abundant OTUs at 60 m, together contributing to 43% of the relative abundance at this depth. OTUs in the grid (from left to right) are ordered on their average relative abundance over 60 m. Note: Y scales are not fixed. The letter in front of the taxonomic name indicates the level of taxonomy that the OTU could be identified to; e.g. o = Order, f = Family and g = Genus.

Figure 3.6 shows the weekly dynamics of the most abundant OTUs at 60 m. Evident from these graphs is that there are extreme weekly variations in community composition for every OTU. Additionally, there are incongruities in the temporal patterns of relative abundance of OTUs that represent the same family, e.g. Rhodobacteraceae (OTUs 545011 and 272142) or Flavobacteriaceae (OTUs 659486 and795482).

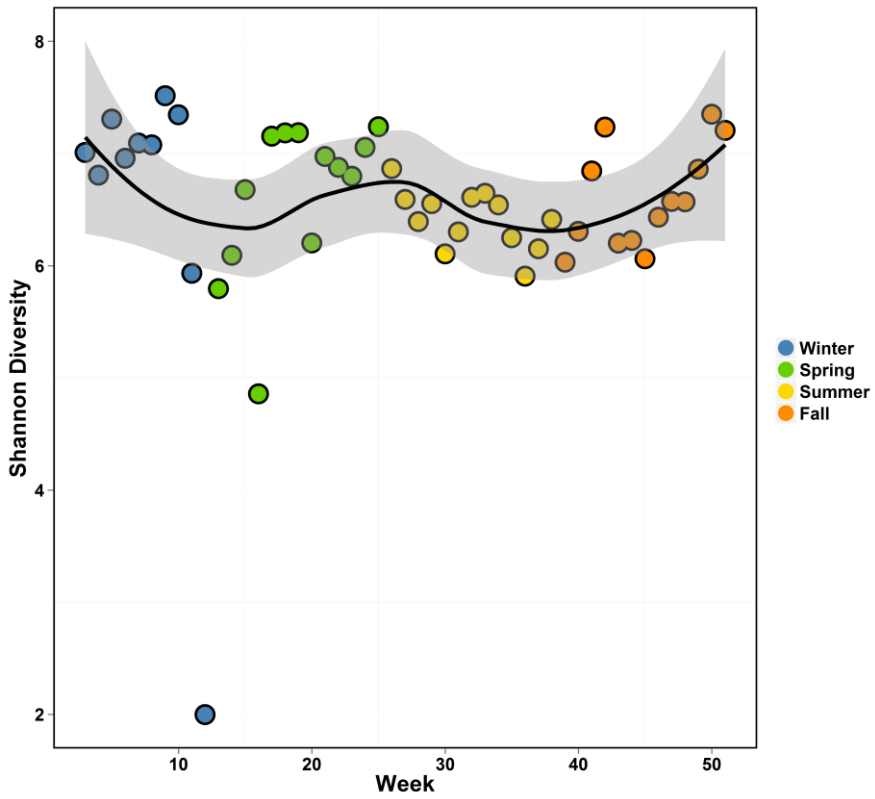


Figure 3.7. Diversity at 60 m over the year as measured by the Shannon Diversity Index. A Loess smoother was included to show general trends in bacterial diversity over the year.

The bacterial diversity of the Bedford Basin at 60 m depth as depicted by Figure 3.7 shows apparent seasonal patterns. Generally, there is higher diversity associated with early winter and late fall samples and lower diversity seen in summer and early fall samples. There appears to be a significant drop in diversity in the late winter and early spring that can be attributed to the extreme spike in dominance of a few opportunistic OTUs, which lowers both species richness and evenness. After this initial drop, community diversity recovers again in late spring before it decreases during summer and early fall. The most diverse sample in the entire data set was from February 26th 2014 at 60 m and had a Shannon diversity index of 7.51. The least diverse sample at 60 m was from March 19th 2014 and coincided with the significant bloom of a few *Pseudoalteromonas* OTUs (Figure 3.6). This sample only had a Shannon diversity index of 2. Overall, the 60 m depth was more diverse throughout the year than the three surface

depths, with an annual average of 5.79 for 1, 5, and 10 m samples and an annual average of 6.59 for the 60 m samples.

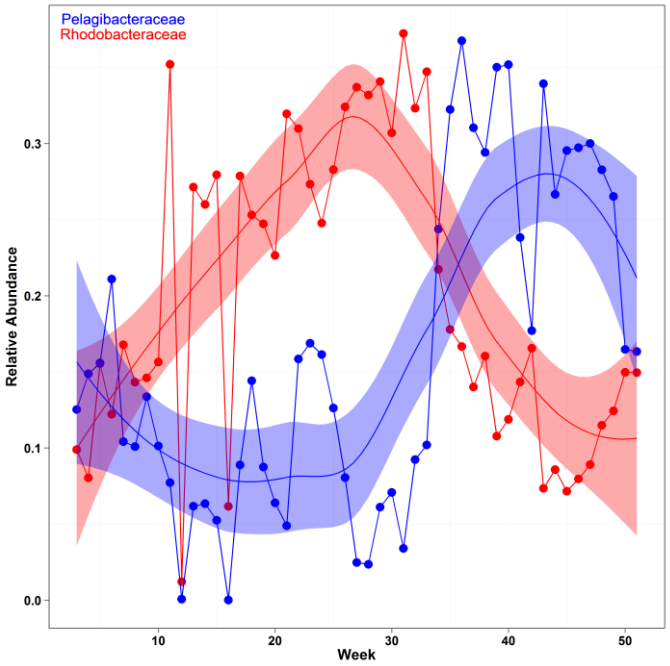


Figure 3.8. Seasonal dynamics of the most abundant families, Pelagibacteraceae and Rhodobacteraceae, throughout the year at 60 m depth with a Loess smoother overlaid to display overall trends.

Seasonal patterns were also apparent at depth with respect to specific taxa. For Figure 3.8, all OTUs within the Rhodobacteraceae and Pelagibacteraceae families were summed in order to show the overall annual pattern in abundance. The families of Rhodobacteraceae and Pelagibacteraceae were the most abundant groups over the year at 60 m depth with Rhodobacteraceae accounting for on average 19.9% of community abundance and Pelagibacteraceae accounting for an average of 16.1% of community abundance. Specifically, Rhodobacteraceae begins to dominate the community in the early spring and peaks in abundance during the summer. Pelagibacteraceae comes to dominate the community in fall and early winter.

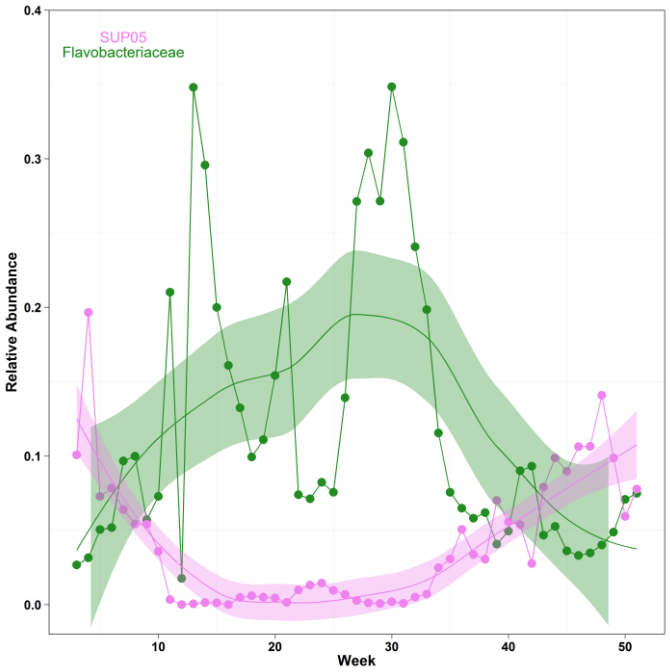


Figure 3.9. Seasonal dynamics of the abundant families of Flavobacteriaceae and SUP05 throughout the year at 60 m depth with a Loess smoother overlaid to display overall trends.

Flavobacteriaceae and SUP05 were the next most abundant families at 60 m in the basin with 12.1% and 4.1% of bacterial community abundance respectively. In this case, SUP05, although significantly less abundant over the year as a whole is more abundant than OTUs belonging to the Flavobacteriaceae family in late fall and early winter. Flavobacteriaceae becomes dominant in the spring and summer weeks, in a similar general pattern to the Rhodobacteraceae family. It differs from Rhodobacteraceae in that the family experiences two main peaks during its most abundant times in the spring and summer.

3.3.3 Bacterial Community Analysis at 60 m – Beta Diversity

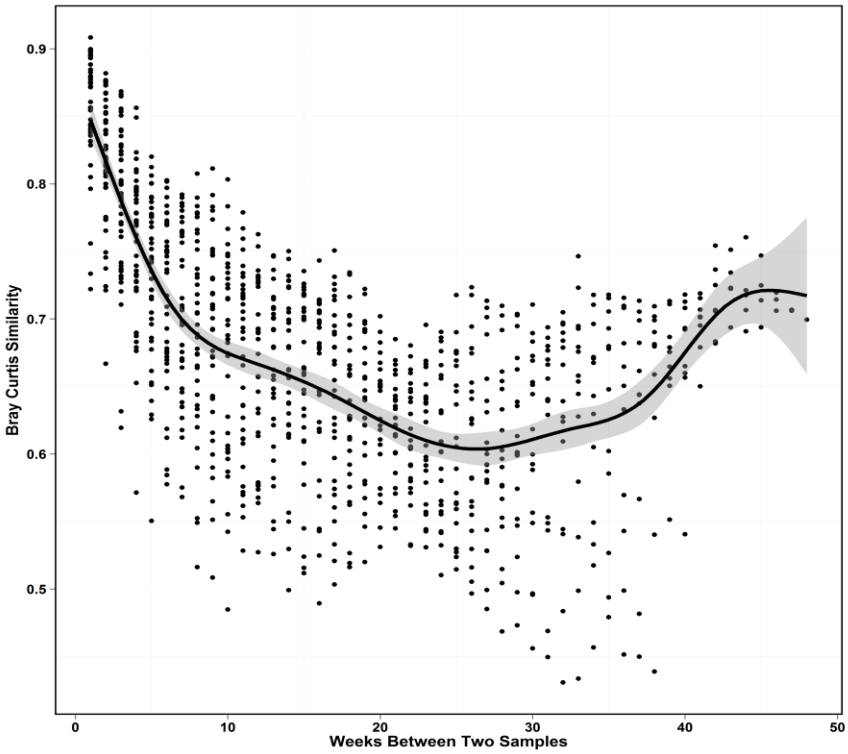


Figure 3.10. Pairwise Bray-Curtis Similarity as a function of temporal distance between two sites. The samples with the extreme peaks in *Pseudoalteromonas* are excluded from this figure (weeks 12 and 16). A Loess smoother is overlaid on the points to show the general trend.

The general trend identified in Figure 3.10 shows that there is an initial steep decrease in community similarity as samples increase in temporal separation. However this decrease plateaued after approximately half a year (26 weeks) of separation, (i.e. samples from January and samples from June). After this point, sample similarity begins to increase again, but doesn't quite reach the similarity levels of minimally separated samples. As this dataset only includes samples from one year, the number of samples that were separated by just short of a year were few and only correspond to those taken in January and December 2014. However, even with only a year of data, it is evident that seasonality influences the bacterial community composition at the 60 m depth in the Bedford Basin.

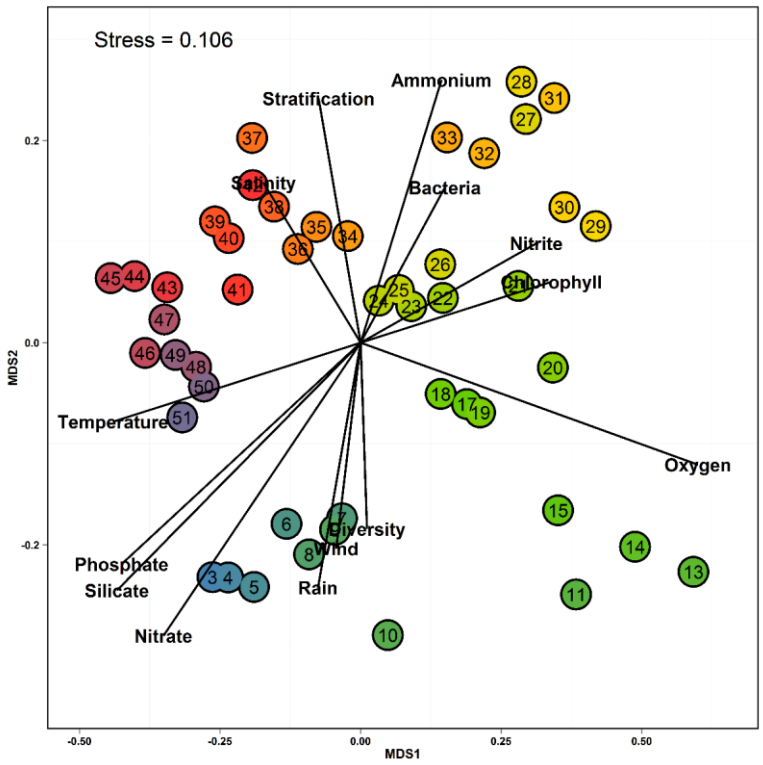


Figure 3.11. NMDS plot using Bray-Curtis dissimilarity with 60 m samples and significant ($p < 0.05$) environmental variables fit to the ordination. Analysis included all OTUs that reached at least 1% of community abundance during one week of sampling. All abundance data was Hellinger transformed before ordination, and weeks 12 and 16 were taken out due to their extreme bacterial communities and low diversity (sections 3.3.5.1 and 3.3.4). Colours and numbers on points correspond to the week of the year when the sample was taken.

The seasonal patterning of samples shown in Figure 3.10 is again evident here in Figure 3.11. Primarily, temporally close samples tend to cluster closer together in ordination space. Additionally, sampling times that are half a year apart are generally the furthest apart in ordination space and the winter samples from January 2014 (e.g. samples 3 and 4), are in relatively close proximity to the late fall samples from December 2014 (e.g. samples 50 and 51). These factors create a circular structure of samples loosely arranged by the numerical weeks of the year. In addition to this general seasonality, there are multiple weeks that cluster closer together in comparison to other samples (e.g. 3, 4, 5 or 17, 18, 19), showing fluctuating rates of community change within the year. Environmental vectors were fit to the site ordination and aim to capture the greatest gradient of environmental change over the sites. The length of the vector relates to the

strength of the environmental gradient and the significance of the relationship, and right angle projections from a sample point to a vector will estimate the value of that variable at that sampling point. In general, higher temperatures are found in late fall and early winter samples, and high oxygen is found in spring samples and a few summer samples. Higher salinity and a stratified water column are associated with late summer and early fall samples and ammonium and bacterial concentrations are associated with summer samples. High nutrient concentrations as well as increased rain and wind are associated with winter samples.

Table 3.1. Results of Mantel and Partial Mantel tests on the matrix of community similarity with matrices of environmental variables and temporal separation at 60 m depth. Correlations calculated using Spearman's correlation and significant correlations are in bold (p value <0.05). Variables ordered on the strength of their Partial Mantel correlations.

Variable	Mantel	Partial Mantel
All Environment	**0.670(0.0001)**	**0.571(0.0001)**
Silicate	**0.596(0.0001)**	**0.513(0.0001)**
Oxygen	**0.604(0.0001)**	**0.499(0.0001)**
Day Length	**0.555(0.0001)**	**0.447(0.0001)**
Phosphate	**0.553(0.0001)**	**0.437(0.0001)**
Time (weeks)	**0.563(0.0001)**	**0.405(0.0001)**
Nitrate	**0.483(0.0001)**	**0.357(0.0001)**
Temperature	**0.498(0.0001)**	**0.336(0.0001)**
Chlorophyll	**0.165(0.005)**	**0.255(0.0001)**
Wind	**0.235(0.0001)**	**0.195(0.0005)**
Nitrite	**0.107(0.03)**	**0.106(0.04)**
Bacterial Diversity	**0.164(0.0009)**	**0.091(0.02)**
Salinity	**0.314(0.0002)**	**0.077(0.05)**
Rain	**0.153(0.003)**	**0.077(0.04)**
Ammonium	**0.171(0.001)**	0.068(0.07)
Bacteria	0.064(0.1)	0.024(0.3)

The results of the various Mantel and Partial Mantel tests performed are compiled in Table 3.1. The results of this analysis show that there was high correlation between bacterial community similarity and the similarity in most environmental parameters. Additionally, time had a highly significant relationship even when accounting for environmental factors. The correlation between community similarity and environmental similarity decreased for most environmental variables when time was taken into account.

Distances between samples based on all environmental variables showed the greatest correlation with community dissimilarity. Silicate was the single variable that had the highest correlation with community similarity when taking temporal separation into account (Partial Mantel), and oxygen was the single variable that had the highest correlation with community similarity regardless of temporal separation (Mantel). Day length and phosphate were all also highly significantly correlated with community similarity.

3.3.4 Dynamics of Individual OTUs – Rare and Indicator Species at 60 m

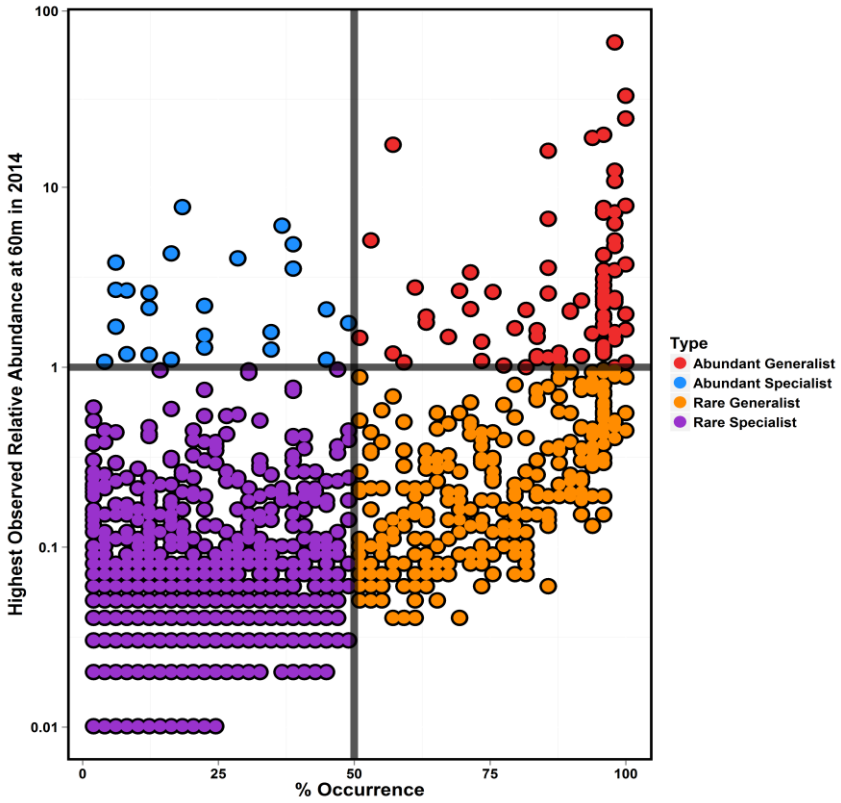


Figure 3.12. OTUs observed at 60 m in the Bedford Basin in 2014 with the percent of samples where the OTU was observed on the x-axis and the highest percent of relative abundance observed for that taxa on the y-axis. Points are coloured based on occurrence and abundance patterns. "Abundant Generalists" are present in over 50% of samples and reach at least 1% of the relative community for at least one sample. "Abundant Specialists" reach at least 1% of the relative community at one sample but occur in less than 50% of samples. "Rare Specialists" are present in over 50% of the samples but never reach higher than 1% of a sample's relative abundance, and "Rare Generalists" are present in less than 50% of the samples and never reach above 1% of the relative community abundance.

Figure 3.12 displays all the OTUs present at 60 m depth positioned on their percent occurrence over the year and their highest maximum relative abundance at any week in 2014. As described above, there were 88 (1%) Abundant Generalist OTUs, 24 (0.3%) Abundant Specialist OTUs, 293 (3.3%) Rare Generalists, and 8378 (95.4%) Rare Specialists. The vast majority of OTUs were detectable in less than 50% of the samples and never reached above 1% of any weekly community. The Rare Generalists were relatively common in this dataset especially when compared to the Abundant Generalists.

Table 3.2. OTUs that fit the "Abundant Specialist" classification where they occur in less than 50% of samples at 60 m but become conditionally abundant as they reach over 1% of the community at least once during the year. The specific OTU's percent occurrence in the samples and the max percent that the OTU reached during the year are also included.

Class/Phylum	OTU	Percent Occurrence	Max Percent
Gammaproteobacteria	*Pseudoalteromonas*;40554	18.4	7.7
Epsilonproteobacteria	*Arcobacter*;New.ReferenceOTU56	36.7	6.1
Gammaproteobacteria	Gammaproteobacteria;814067	38.8	4.8
Alphaproteobacteria	*Sphingopyxis*;577679	16.3	4.3
Bacteroidetes	*Polaribacter*;81397	28.6	4.0
Betaproteobacteria	Burkholderiales;808372	6.1	3.8
Gammaproteobacteria	Colwelliaceae;New.ReferenceOTU24	38.8	3.5
Bacteroidetes	Cyclobacteriaceae;106281	6.1	2.7
Bacteroidetes	*Flavobacterium*;626021	8.2	2.7
Gammaproteobacteria	*Stenotrophomonas*;591708	12.2	2.6
Deltaproteobacteria	Myxococcales;New.ReferenceOTU96	22.4	2.2
Alphaproteobacteria	*Thalassospira*;114216	12.2	2.1
Betaproteobacteria	*Gallionella*;1628999	44.9	2.1
Gammaproteobacteria	Legionellaceae;100883	49.0	1.8
Bacteroidetes	*Flavobacterium*;New.ReferenceOTU131	6.1	1.7
Gammaproteobacteria	Alteromonadaceae;656996	34.7	1.6
Bacteroidetes	*Leeuwenhoekiella;*636147	22.4	1.5
Alphaproteobacteria	*Parvibaculum*;147243	22.4	1.3
Actinobacteria	*Gordonia*;New.ReferenceOTU38	34.7	1.3
Alphaproteobacteria	*Thalassospira*;112782	8.2	1.2
Alphaproteobacteria	Rhodobacteraceae;724648	12.2	1.2
Bacteroidetes	*Maribacter*;4371451	16.3	1.1
Gammaproteobacteria	Thiohalorhabdales;4406969	44.9	1.1
Alphaproteobacteria	*Sphingomonas*;347846	4.1	1.1

The Abundant Specialists were the least common group and Table 3.2 summarizes the 24 OTUs that fit this classification. These OTUs are perhaps the most intriguing group because they were either absent or undetectable for the majority of the year, but were able to "bloom" to relatively high abundance periodically. The most pronounced Abundant Specialist OTU was from the genus *Pseudoalteromonas* (40554) and was only present in 18.4% of samples and reached a max percentage of 7.7%. Seven other OTUs from the class of Gammaproteobacteria fit this classification, showing that this class has a proportionately high number of species that are able to bloom in this manner. Most other OTUs came from the class of Alphaproteobacteria and the phylum of Bacteroidetes.

Table 3.3. Results from Indicator Species Analysis on OTUs that represent at least 0.1% on average of annual community relative abundance. R.g stat is the correlation index used to measure the strength of the association between the OTU and the group of samples.

Season	*Phylum/Class*	*Classification*	*R.g Stat*	*Significance*
Fall	**Actinobacteria**	Acidimicrobiales.237721	0.658	***
	Alphaproteobacteria	Pelagibacteraceae.637092	0.612	***
		Rhodobacteraceae.30068	0.592	***
		Octadecabacter.New.ReferenceOTU6	0.468	**
	Bacteroidetes	Saprospiraceae.New.ReferenceOTU8	0.740	***
	Deltaproteobaceria	*Nitrospina*.936371	0.777	***
		OM27.New.ReferenceOTU45	0.667	***
		Bacteriovorax.New.ReferenceOTU42	0.460	**
	Gammaproteobacteria	Ectothiorhodospiraceae.New.ReferenceOTU17	0.805	***
		Legionellaceae.100883	0.529	***
		Halomonadaceae.554937	0.427	*
	Planctomycetes	CL500-15.813173	0.606	***
	Verrucomicrobia	Verrucomicrobiaceae.544574	0.564	***
Spring	**Bacteroidetes**	*Ulvibacter*.543487	0.739	***
		Ulvibacter.774258	0.571	***
		Polaribacter.81397	0.464	**
	Epsilonproteobacteria	*Arcobacter*.New.Reference.OTU56	0.333	*
		Arcobacter.4301711	0.333	*
	Gammaproteobacteria	Oceanospirillaceae.970344	0.618	***
		Oceanospirillaceae.New.Reference.OTU1	0.596	***
		SAR86.540364	0.424	*
	Verrucomicrobia	Verrucomicrobiaceae.3603504	0.623	***
Summer	**Alphaproteobacteria**	Rhodobacteraceae.692164	0.684	***
		Rhodobacteraceae.730637	0.429	*
	Bacteroidetes	Saprospiraceae.242559	0.754	***
		Flavobacterium.4454631	0.713	***
		Flavobaceriaceae.659486	0.686	***
		Flavobacteriaceae.311600	0.536	***
		Polaribacter.28929	0.491	**
	Deltaproteobacteria	Deltaproteobacteria.New.ReferenceOTU26	0.538	**
Winter	**Alphaproteobacteria**	*Sphingopyxis*.577679	0.322	***
	Bacteroidetes	NS9.831425	0.768	***
		Flavobacteriales.146541	0.550	***
		Cryomorphaceae.590041	0.528	**
	Gammaproteobacteria	Gammaproteobacteria.814067	0.371	*
		Pseudoalteromonas.40554	0.281	*
Fall + Spring	**Bacteroidetes**	Flavobacteriales.2684406	0.452	**
	Planctomycetes	Agg27.828348	0.680	***
Fall + Summer	**Alphaproteobacteria**	Rhodospirillaceae.535304	0.729	***
		Rhodobacteraceae.645555	0.637	***
		Pelagibacteraceae.787847	0.583	***

Season	*Phylum/Class*	*Classification*	*R.g Stat*	*Significance*
Fall + Summer		Rhodobacteraceae.700215	0.540	***
	Bacteroidetes	Flavobacteriales.86113	0.697	***
		Flavobacteriaceae.279006	0.651	***
		NS11-12.164181	0.388	*
Fall + Winter	**Alphaproteobacteria**	Pelagibacteraceae.307744	0.688	***
		Pelagibacteraceae.317958	0.532	***
		Rhodobacteraceae.534690	0.530	***
		Pelagibacteraceae.556042	0.505	**
		Pelagibacteraceae.311349	0.481	**
	Bacteroidetes	Flammeovirgaceae.919715	0.480	**
	Betaproteobacteria	*Gallionella*.1628999	0.616	***
	Deltaproteobacteria	SAR324.827371	0.748	***
		Nitrospina.New.Reference.OTU7	0.734	***
		JTB38.111722	0.651	***
		Nitrospina.549675	0.734	***
		SAR324.834526	0.627	***
	Gammaproteobacteria	SUP05.583880	0.702	***
		Methylococcales.532630	0.678	***
	Gemmatimonadetes	Gemm-2.148154	0.670	***
	SAR406	SGSH944.542423	0.518	**
Spring + Summer	**Alphaproteobacteria**	*Octadecabacter*.804483	0.727	***
		Rhodobacteraceae.272142	0.703	***
		Rhodobacteraceae.804449	0.696	***
		Rhodobacteraceae.785501	0.679	***
		Kiloniellaceae.New.Reference.OTU14	0.652	***
		Rhodobacteraceae.714708	0.627	***
		Rhodobacteraceae.703006	0.497	**
	Bacteroidetes	Flavobacteriaceae.1000656	0.691	***
		Flavobacterium.689465	0.675	***
		Flavobacteriaceae.691928	0.669	***
		Flavobacteriales.587244	0.667	***
		Flavobacteriales.537283	0.603	***
		Flavobacteriales.512681	0.596	***
		NS9.590564	0.477	**
		Cryomorphaceae.138877	0.471	**
		Cryomorphaceae.666579	0.416	*
		Cryomorphaceae.686381	0.405	*
	Betaproteobacteria	Methylophilaceae.320198	0.447	**
		Methylophilaceae.614504	0.432	**
	Gammaproteobacteria	SAR86.685508	0.663	***
		HTCC2207.613465	0.624	***
		OM60.630330	0.585	***
	Verrucomicrobia	Verrucomicrobiaceae.564853	0.467	**

Season	*Phylum/Class*	*Classification*	*R.g Stat*	*Significance*
Spring* + *Winter	**Alphaproteobacteria**	Rhodobacteraceae.45295	0.604	***
		Alphaproteobacteria.800299	0.603	***
		Rhodobacteraceae.316226	0.586	***
		Hyphomonas.2847581	0.457	**
	Bacteroidetes	Flavobacteriaceae.795842	0.485	**
		Flavobacteriales.613128	0.371	*
		Flavobacteriaceae.238821	0.365	*
	Betaproteobacteria	Comamonadaceae.616920	0.472	**
	Gammaproteobacteria	*HTCC2207*.535135	0.584	***
		Oceanospirillales.1712364	0.566	***
		HTCC.630967	0.433	*
Summer* + *Winter	**Bacteroidetes**	*Crocinitomix*.172051	0.562	***
Fall* + *Spring	**Actinobacteria**	Actinomycetales.759855	0.412	*
+ Summer	**Alphaproteobacteria**	*Phaeobacter*.696544	0.471	**
	Bacteroidetes	*Fluviicola*.228555	0.492	**
	Planctomycetes	Pirellulaceae.149272	0.406	*
Fall* + *Summer	**Actinobacteria**	ZA3409c.814290	0.470	**
+ Winter				
	Gammaproteobacteria	SAR86.533354	0.396	*
Fall* + *Spring	**Alphaproteobacteria**	Pelagibacteraceae.52749	0.573	***
+ Winter		Pelagibacteraceae.830091	0.566	***
		Pelagibacteraceae.352844	0.561	***
		Pelagibacteraceae.1016465	0.544	***
		Pelagibacteraceae.4432601	0.484	***
		Pelagibacteraceae.New.Reference.OTU22	0.395	*
	Bacteroidetes	Flavobacteriales.3751007	0.402	*
	Gammaproteobacteria	*HTCC*.571659	0.651	***
		*HTCC*275776	0.520	**
		SAR86.549897	0.480	**
***Spring* +**	**Actinobacteria**	SC3-41.275365	0.529	**
***Summer* +**	**Alphaproteobacteria**	Rhodobacteraceae.645011	0.386	*
Winter	**Bacteroidetes**	NS9.New.Reference.OTU2	0.588	***

p value Significance level: $p<0.0001$***, $0.0001<p<0.01$**, $0.01<p<0.05$*

Table 3.3 summarizes the results of the indicator species analysis that was used to determine OTUs that are preferentially abundant during different seasons or combinations of seasons. Of the 149 OTUs used in this analysis, 115 showed significant relationships between relative abundance and season. Of these 115 OTUs, 36 were preferentially abundant in one season, 60 were preferentially abundant in two seasons, and 19 OTUs were preferentially abundant in a combination of three seasons. For the

seasons of fall, spring, summer, and winter, there were 13, 9, 8, and 6 OTUs that were identified as indicator species respectively. The indicator species analysis is mutually exclusive, an OTU can only be assigned to the grouping in which it has the highest correlation value.

The strongest indicator OTU of any one season, was New Reference OTU17 from Ectothiorhodospiraceae, a family of generally anaerobic purple sulfur bacteria in the class of Gammaproteobacteria, which was strongly associated with the fall season. In the spring, OTUs from the Bacteroidetes phylum, e.g. *Ulvibacter* and *Polaribacter*, and from the Verrucomicrobia phylum were indicators. In the summer, the species with the largest indicator species statistic were from the family of Saprospiraceae (OTU242559) and the genus of *Flavobacterium* (OTU4454631). The winter had the fewest indicator species, and an OTU from the family of NS9 (831425) in the phylum of Bacteroidetes was the strongest indicator.

Most of the indicator species that were preferentially found spanning two seasons, were either found in spring and summer (23 OTUs), or fall and winter (16 OTUs). In the spring and summer, most indicator OTUs were from the phylum of Bacteroidetes and the class of Alphaproteobacteria. In the fall and winter most of the OTUs were from Deltaproteobacteria, and Alphaproteobateria, but there was also representatives from less characterized phyla like SAR406 and Gemmatimonadetes, as well as Gammaproteobacteria and Betaproteobacteria. OTUs that were indicator species for three seasons, had generally weaker significance values as compared to the strong indicator species for one or two seasons. These taxa, including many OTUs from Pelagibacteraceae, and one from Rhodobacteraceae (645011) could be considered more ubiquitous or generalist species that have a wider range of potential habitats than indicators species of only one or two seasons.

This analysis provides a large amount of data on many bacterial species that are considered rare and often excluded from literature analyses. Discovering the ecological niches of these rare species will help to determine their function in their communities, including interactions with other taxa and roles in biogeochemical cycling.

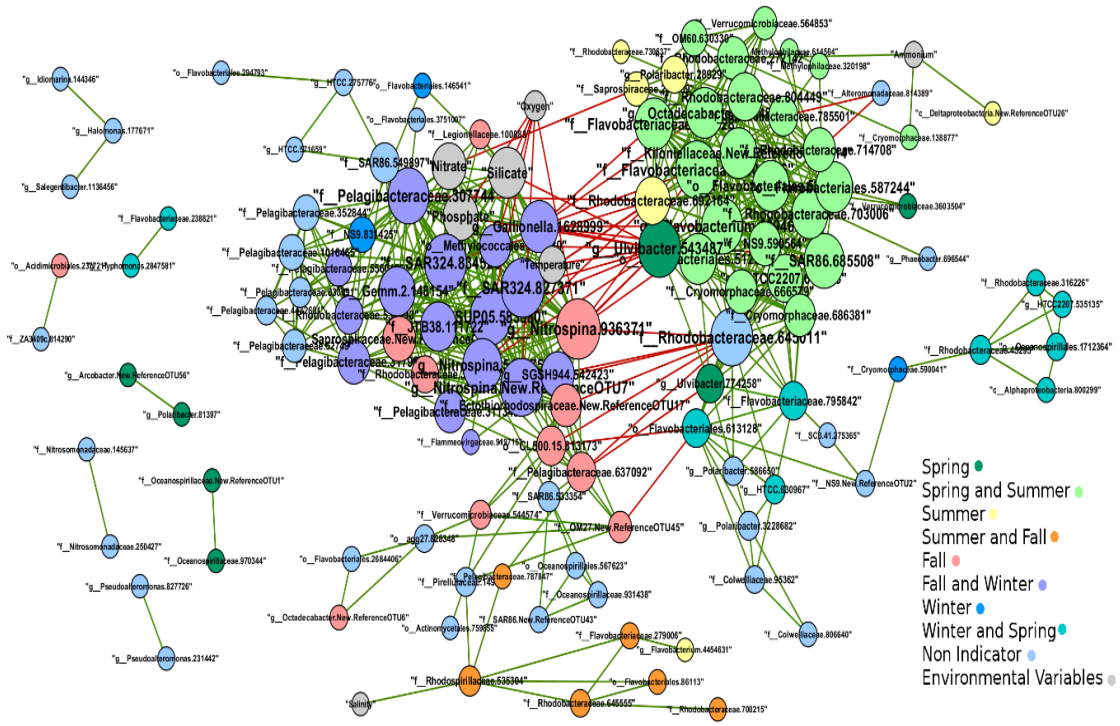


Figure 3.13. Spearman correlation network (*p* values <0.01, rho > 0.7). Nodes are coloured based on indicator species analysis for season. Only OTUs that were indicator species for one season or two seasons were coloured. Red edges indicate a negative relationship, green edges indicate a positive relationship. Node size represents degree of each OTU, or number of connections.

Figure 3.13 displays the significant correlations between OTUs in network form, where the layout is meant to highlight clusters of co-occurring OTUs. Additionally, OTUs were coloured based on seasonal preference as identified by the indicator species analysis (Table 3.2). The correlated OTUs in this network are divided into two main clusters of co-occurrence. The members of these two main clusters were negatively correlated with each other, or in other words, not often observed together. In the first large cluster there are OTUs that were indicator species for fall, winter, and fall and winter. In the second large cluster there are mainly OTUs that were indicator species for both spring and summer, or spring and summer separately. In addition to the main interacting clusters which include the majority of species, there are some species that are correlated with only a few other OTUs or environmental variables.

OTUs that were the most connected included *Nitrospina* (OTU 936371) with 28 connections, and SAR324 (827371) with 27 connections. The *Nitrospina* OTU was an

indicator species for fall and the SAR324 OTU was an indicator species for both fall and winter seasons. Both OTUs never reached above 1% of community abundance at any point during the year and were thus chronically rare species. Other OTUs from Flavobacteriaceae (1000656), Rhodobacteraceae (645011), SAR324 (834526), *Nitrospina* (New Reference OTU7), and *Ulvibacter* (543487) were also all highly connected with 26 significant connections. These OTUs were indicators for spring and summer, three seasons (spring, summer and winter), fall and winter, fall and winter, and spring, respectively. Overall, the network was fairly highly connected with the average OTU or environmental variable having 9.5 highly significant relationships and an average path length of 3.5. With 21 significant relationships, silicate was the most connected environmental variable, including positive relationships with nitrate, phosphate, and temperature, and a negative relationship with oxygen.

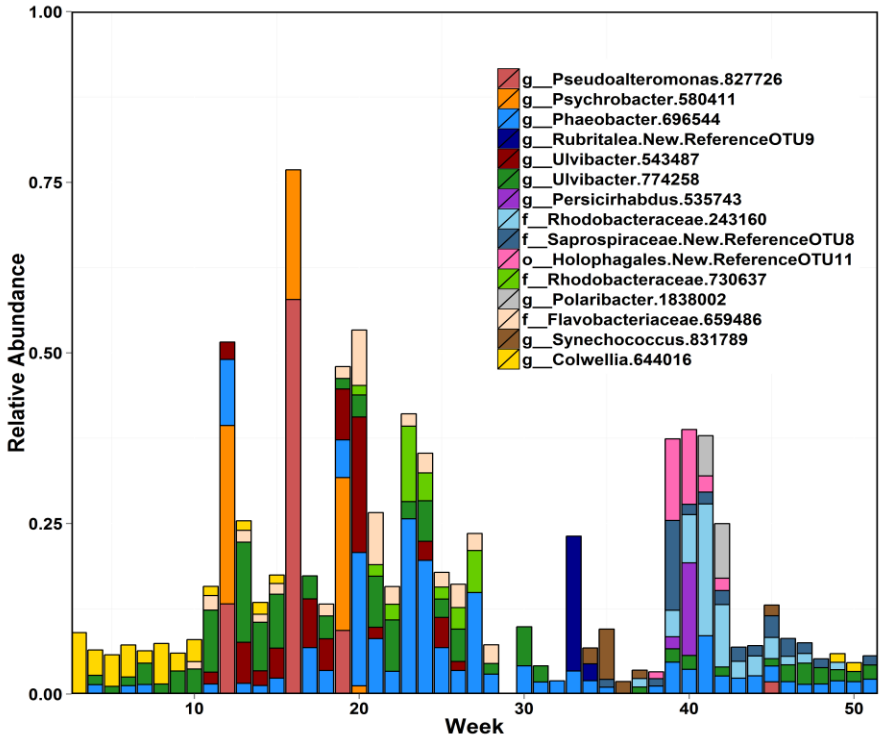


Figure 3.14. Stacked bar chart of "spiking" bacterial OTUs over the year at 5 m. Only OTUs from 5 m that meet the following criteria are included in this figure. Firstly, the absolute value for the weekly spike (change from the week before) is greater than 5x the average abundance of that taxa. Additionally the max value for the spike has to be a relative change of 0.05 in the community relative abundance from the week before. In this figure, all relative community abundances less than 1% at any given sample are excluded to reduce overcrowding in the figure.

Figure 3.14 shows the spiking bacterial community at 5 m. *Pseudoalteromonas* (OTU 827726) had the largest spike of any OTU at 5 m, as it went from 0.42% of bacterial abundance on April 9th (week 15) to 57.8% of bacterial abundance on April 16th (week 16), and then immediately returned back down to 0.3% of bacterial abundance the following week. This *Pseudoalteromonas* OTU was relatively abundant at 5 m depth two other times through the year, reaching 13.2% of community abundance on March 19th (week 12), and 9.3% of community abundance on May 6th (week 19). Other than that it remained below 1% of community abundance all other weeks sampled except for a small peak to 1.8% on November 5th (week 45). Other dramatic spiking bacterial OTUs include *Psychrobacter* (OTU 580411), and *Rubritalea* (New Reference OTU9). *Psychrobacter* spiked to abundance the same three weeks that *Pseudoalteromonas* (OTU 827726) did and reached 26.1% (week 12), 19% (week 16), and 22.4% (week 19) of relative community abundance. *Rubritalea* spiked in abundance only on week 33 (August 13th) where it went from 0.02% of community abundance the week before, to 19.8% of community abundance, and then back down to 2.5% of community abundance the week after. *Synechococcus* (OTU 831789), *Polaribacter* (OTU 1838002), and *Persicirhabdus* (OTU 535743) also only peaked in abundance over one or two weeks before becoming undetectable again. Other OTUs, like those from *Phaeobacter* (OTU 696544), and *Ulvibacter* (OTUs 543487 and 774258) gradually became abundant over the course of a few weeks, peaked in abundance one week during this time, and then gradually became less abundant again.

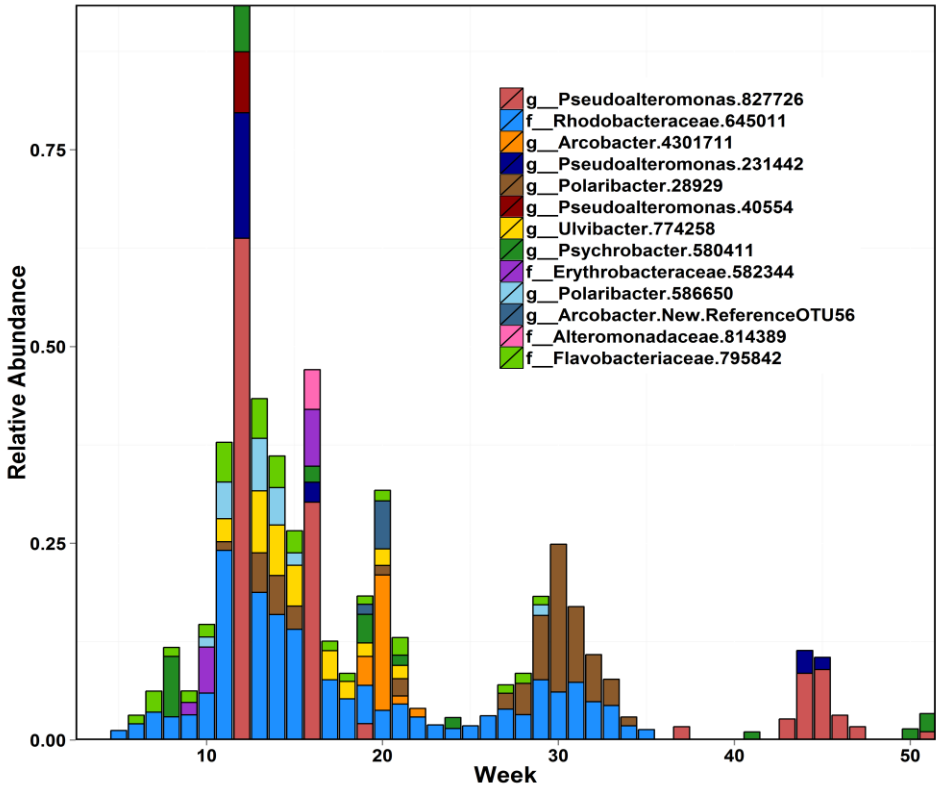


Figure 3.15. Stacked bar chart of "spiking" bacterial OTUs over the year at 60 m. Only OTUs from 60 m that meet the following criteria are included in this figure. Firstly, the weekly spike (change from the week before) is greater than 5x the average abundance of that taxa. Additionally the max value for the spike has to be a relative change of 0.05 in the community relative abundance from the week before. In this figure, all relative community abundances less than 1% at any given sample are excluded to reduce overcrowding.

Figure 3.15 displays the temporal patterns of bacterial taxa that fit the "spiking" criteria. In this dataset (both 5 m and 60 m) there are generally two distinct observable patterns in spiking bacteria. Some of the OTUs spike for only a week, and are essentially absent from the community the weeks prior to and post the spike event. The other temporal pattern of spiking OTUs is a more gradual increase in relative abundance in the community over time. These OTUs may be relatively abundant for many weeks, sometimes with fluctuations, and nearly absent for others. As an example, the relatively abundant OTU from Rhodobacteraceae (645011) is on average 3.6% of the community but stays below 1% for all weeks of the year after week 35. It reaches its maximum abundance of 24% at week 11 and then begins to gradually decrease in abundance again.

At 60 m, the most dramatic spike of any OTU is that of *Pseudoalteromonas* (827726) at week 12 to 64% of the bacterial community, and then again at week 16 to 30% of the bacterial community. The OTU is mainly absent again until it reappears to just under 10% for a few weeks in fall. The *Pseudoalteromonas* OTUs 231442, and 40554 also spike substantially at week 12, and *Pseudoalteromonas* 231442 was present with *Pseudoalteromonas* 827726 at three other weeks. The OTU from the genus *Psychrobacter* (580411) spiked at various times during the year, including with the *Pseudoalteromonas* OTUs on week 12, where it reached 6% of the community abundance. It also reached 8% of the community abundance at week 8, and as it averages 0.7% abundance throughout the year this change in abundance marks a more than 10x increase. There was also a dramatic spike of *Arcobacter* (4301711) at week 20, where the OTU represented 17% of the community. On average *Arcobacter* represented less than 0.5% at 60 m, meaning that this spike brought the OTU to 34x its normal values. Another *Arcobacter* OTU (New Reference OTU56) also spiked at week 20 to 6% of the community, and since it was on average only 0.2% of the community this corresponded to a 30x increase. As described above, many OTUs seemed to respond to the same triggers and spiked in unison during the same weeks. In addition to those already mentioned, Alteromonadaceae (814389) and Erythrobacteraceae (582344) spiked together on week 16, and OTUs from *Ulvibacter* (774258), *Polaribacter* (586650) and Flavobacteriaceae (795842) gained abundance together over weeks 11, and 13-15. These weeks coincide with the first intrusion event that will be discussed more in the next section. Lastly many OTUs from the same genus show similar temporal patterns (e.g. OTUs from *Pseudoalteromonas* or OTUs from *Arcobacter*), however the OTUs from other genera, such as *Polaribacter*, show very distinct temporal patterns of abundance. *Polaribacter* (586650) is abundant in the early spring, whereas *Polaribacter* (28929) becomes most abundant in the summer. This suggests that these two phylogenetically similar lineages exhibit different preferences for ecological niches and thus could be considered different ecotypes (Xing *et al.*, 2015).

Figures 3.14 and 3.15 show strikingly different patterns in spiking bacterial OTUs. The most dynamic OTU at both depths is *Pseudoalteromonas* (OTU 827726). At 60 m the OTU reaches 64% of community abundance at week 12, and 30% of

community abundance at week 16. At 5 m depth the OTU is 15% of community abundance at week 12, and 60% of community abundance at week 16. This shows a shift in the vertical water column of *Pseudoalteromonas* community domination. At 60 m depth, 2 other *Pseudoalteromonas* OTUs meet the spiking criterion, however at 5 m depth only the dominant OTU 827726 is abundant during the spikes. Also, other than *Pseudoalteromonas* (OTU 827726), *Ulvibacter* (OTU 774258), and *Psychrobacter* (OTU 580411), there were very different OTUs classified as conditionally rare taxa between 5 m and 60 m samples.

3.3.5 Intrusion Events

3.3.5.1 First Intrusion – March 5th to March 12th 2014 (weeks 10-11)

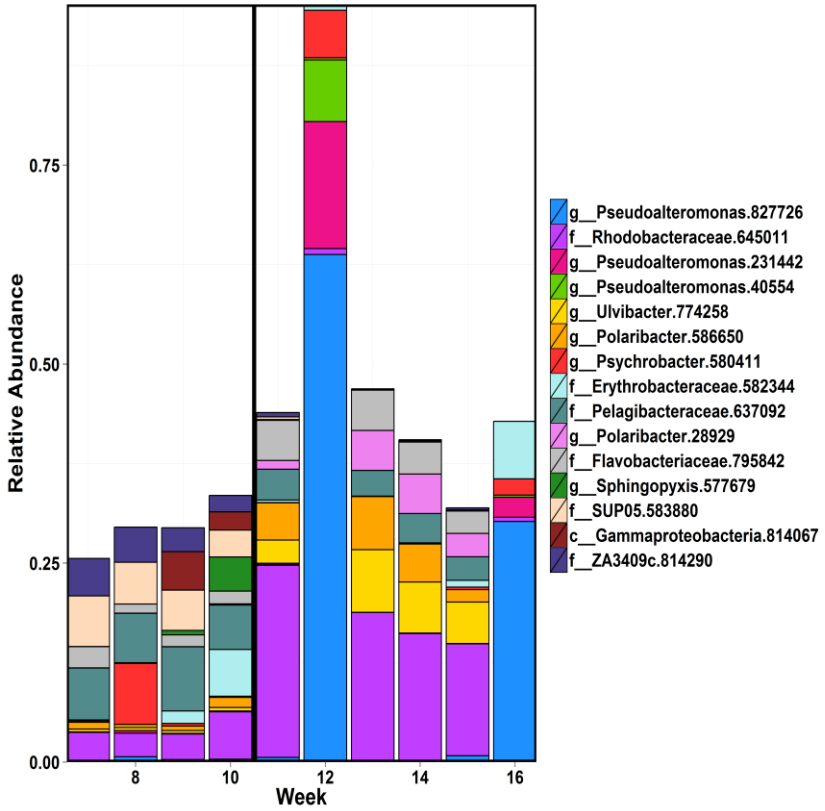


Figure 3.16. Bar charts showing the OTUs that exhibited a change of at least 0.05 in community relative abundance from the period surrounding the deep water intrusion event between March 5th and March 12th 2014. The intrusion event is demarcated by the black line.

Figure 3.16 shows the OTUs of the bacterial community that experienced a change in relative abundance of greater than 5% over the course of the weeks preceding and following the March intrusion event. The most obvious change in community structure actually happens two weeks after the intrusion event (black line) on March 19th or week 12. At this time, the three habitually rare OTUs from *Pseudoalteromonas*: 8277726, 231442, and 40554, become dominant members of the community making up 64%, 16%, and 7.7% of the relative abundance respectively. The OTU *Psychrobacter* 580411 also peaks at this time to 6% of relative abundance. A similar, but less dramatic spike in *Pseudoalteromonas* OTUs 8277726, 231442, *Psychrobacter* OTU 580411, and *Erythrobacteraceae* OTU 582344 occurred four weeks later on April 16th week 16. These spike events are incredibly interesting as the bacterial OTUs that surge to community dominance are nearly negligible in abundance the week preceding and the week following their spike. Aside from these spikes the community appears to shift with respect to certain OTUs. The OTUs that strongly decrease in abundance post intrusion include SUP05 OTU 583880, and ZA3049c OTU 814290 which were relatively abundant for the four weeks preceding the intrusion. The OTUs from Gammaproteobacteria (814067), Erythrobacteraceae (582334), and *Sphingopyxis* (577679) became relatively abundant the two weeks preceding the intrusion then became rare again the week post intrusion. The OTUs that generally increased in abundance post intrusion (disregarding week 12) include Rhodobacteraceae OTU 645011, *Ulvibacter* OTU 774258, *Polaribacter* OTU 586650, *Polaribacter* OTU 28929, and Flavobacteriaceae OTU 795842.

3.3.5.2 Second Intrusion – July 8th to July 15th (weeks 28-29)

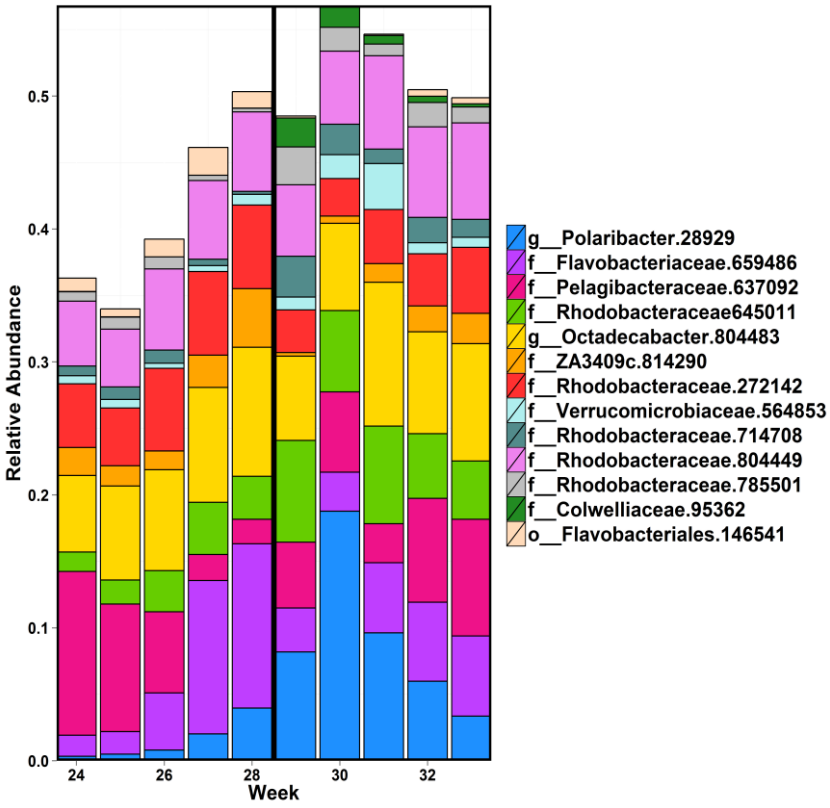


Figure 3.17. Bar charts showing the OTUs that exhibited a change of at least 0.02 in community relative abundance from the period surrounding the intrusion event between July 8th and July 15th 2014. The intrusion event is demarcated by the black line.

At 60 m, the second intrusion event created less dramatic changes in relative abundance of OTUs than the first intrusion event. The OTUs that increased the most post intrusion include OTUs from *Polaribacter* (28929), Rhodobacteraceae (645011, 714708, and 785501), and Colwelliaceae (95362). Flavobacteriaceae OTU (659486), Flavobacteriales OTU (146541), ZA3409c OTU (814290) and Rhodobacteraceae OTU (272142) all decrease following the intrusion event by at least a factor of two. Even though these OTUs decrease immediately following the intrusion, they all begin to regain pre-intrusion abundances after two to three weeks. By this same pattern, the OTUs that gained abundance after the intrusion event start gradually decreasing two to three weeks later as well. This suggests an ability of the bacterial community to return to status quo after a dramatic perturbation event.

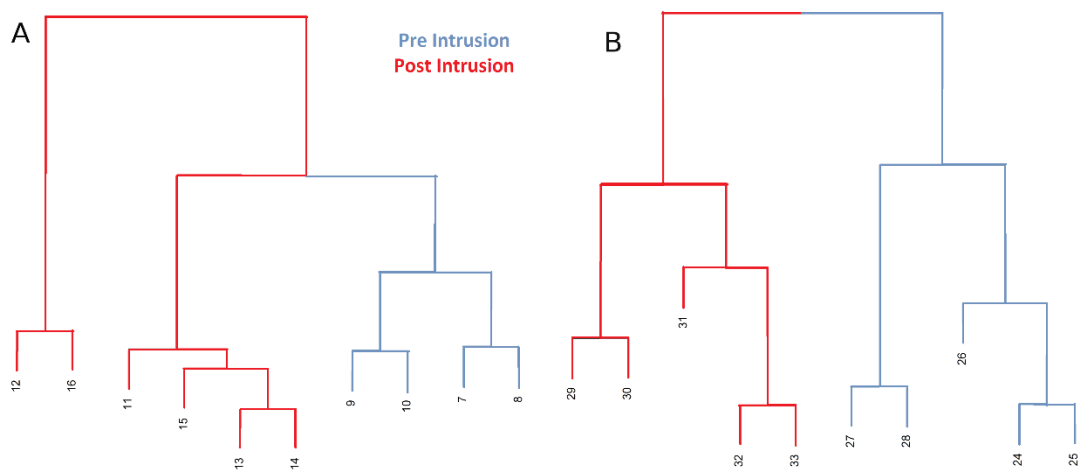


Figure 3.18. Dendrograms of weeks surrounding the first intrusion (A) and the second intrusion (B) clustered according to Bray-Curtis dissimilarity. All abundance data was Hellinger transformed prior to clustering analysis. Branches leading to pre intrusion weeks are shown in blue and branches leading to post intrusion weeks are shown in red.

The dendrograms above further illustrate the grouping of the samples based on bacterial community similarity surrounding the intrusion events. Figure 3.18A shows the first intrusion where there is a clear outlier group containing weeks 12 and 16 which are the samples with the extreme *Pseudoalteromonas* peaks, and then two main groups made up of pre-intrusion samples from weeks 7-10, and post-intrusion samples from weeks 11, 13-15. In Figure 3.18B, the samples are also cleanly divided into pre and post intrusion events. Samples 24-28 from before the mixing event are in one main cluster, and samples 29-33 after the mixing event are in the other. Looking further into the post intrusion cluster, samples immediately after the intrusion from weeks 29 and 30 are grouped together while the later samples from weeks 31, 32 and 33 are grouped on a different branch.

In general the dynamic bacterial OTUs that change rapidly in response to these intrusion events are quite different from the first to the second intrusion. However, there are a select few OTUs that responded dramatically to both events either by gaining in abundance or decreasing in abundance after the intrusion. Dynamic OTUs found during both events that increased post intrusion include *Polaribacter* (28929), and Rhodobacteraceae (645011). The OTU that decreased after both events was ZA3409c (814290) from the phylum of Actinobacteria. The consistent response of these OTUs to

these somewhat similar mixing events would be something of interest to continue to study in the future.

3.4 Discussion

3.4.1 Establishing a Microbial Time Series in the Bedford Basin through use of High Throughput Sequencing

The climate is changing, ultimately affecting the ocean environment. Many studies have addressed the physical and chemical parameters associated with this change, however the impact that climate change will have on biological systems, particularly microbial communities, is not well understood. For this reason, it is imperative to research biological responses to a changing marine environment. Perhaps one of the best ways to observe this change over time on the microbial community is to take advantage of established time series that are meant to sample the same location at a defined temporal frequency over many years (Fuhrman *et al.*, 2015). Many such time series have been established (Chow et al. 2013; J. a Cram et al. 2015; J. a Fuhrman et al. 2006; Karl and Church 2014; Vezzulli et al. 2012; Beman, Steele, and Fuhrman 2011), and most sample at monthly or quarterly intervals. The average generation time of marine bacteria is approximately a day in surface waters, resulting in whole community turnover that is estimated to take in the range of less than a day to about a week (Fuhrman *et al.*, 2015). These rapid turnover rates would suggest that monthly sampling intervals are not frequent enough to observe rapid community changes including short-lived microbial dynamics. Another advantage of increasing sampling frequency is that the variation in the abundances of microbial groups over weeks in a month or season can be averaged, or standardized in some way in order to identify and distinguish between annual norms and one-off anomalies. In this work we report the first findings from a weekly microbial community structure time series that was introduced to the Bedford Basin sampling regime in 2014 as an addition to the established 20 year Bedford Basin time series.

The Bedford Basin serves as an ideal model for determining how changing seasons and therefore changing environmental parameters shape bacterial community structure. For twenty years the Bedford Basin has been used as a study site for a plankton monitoring program that has been observing physical, chemical and biological (flow

cytometry based) patterns over the last two decades (Li and Harrison, 2008). Because of its location, the basin is easily accessible by boat and can be sampled weekly, and previous studies (Li *et al.*, 2006) have shown that the basin acts as a good model for the rest of the Atlantic shelf region. Additionally, as the Bedford Basin is part of an active harbour, there is opportunity here to study anthropogenic influences, as well as the bacterial community response to any sort of industry or weather triggered perturbation. As multiple depths are sampled, this time series will allow for integration of the effects of seasonality and climate over the vertical water column. Multiple domains of life will also be monitored including eukaryotes and Archaea in addition to the bacteria studied in this chapter. This will allow for the integration and identification of cross-domain associations and interactions, an emerging area of study that has proved of interest in other areas of the world's oceans (Lima-Mendez *et al.*, 2015). Another added benefit of the Bedford Basin is that it becomes seasonally anoxic at depth creating a "natural laboratory" that can be used to monitor how bacterial communities respond to decreasing oxygen conditions and sudden oxygenation events. This has many implications as the warming ocean is expected to create expanding oxygen minimum zones (Stramma *et al.*, 2008), in addition to increased anthropogenic derived dead zones along heavily industrialized coastal areas (Diaz and Rosenberg, 2008). As has been observed in this study, the deep waters of the basin are often flushed through shelf water intrusions and deep mixing events which can bring in dramatically different water masses from one week to the next. The results of these rapid shifts in physiochemical water properties can also be studied in the future. Overall, the introduction of this Bedford Basin microbial time series allows for many high quality experiments and observations at present and into the future that can help address the question of how the microbial community will respond to a changing climate.

3.4.2 Environmental Variation in the Bedford Basin in 2014

With respect to the previous 20 years of monitoring at the Bedford Basin, the year 2014 was on average, warmer, fresher, and more stratified than normal. These trends are consistent with the observations from the last three years and suggest a period of warming that is congruent with global estimates of warming (Bindoff *et al.*, 2007).

Additionally, chlorophyll was lower than normal which agrees with the generally negative relationship between chlorophyll and stratification. Nitrate and silicate levels were average, but phosphate levels were lower than previous years, which was mainly attributed to the operation of the Halifax wastewater treatment facility and has been the trend since its reopening, in years post 2010 (from unpublished BIO 2014 AZMP report).

3.4.3 Bacterial Community Analysis at 60 m – Alpha and Beta Diversity

Community structure in the Bedford Basin changed dramatically seasonally as well as throughout the water column. This chapter specifically addressed the seasonal changes in microbial community structure at 60 m depth in the Bedford Basin. Numerous time series investigating the seasonal effects of (sub)surface bacterial communities have found distinct, predictable patterns in microbial communities that repeat yearly (Fuhrman *et al.*, 2006; Gilbert *et al.*, 2012; Chow *et al.*, 2013; El-Swais *et al.*, 2014). However, fewer studies have focused on the effect of seasonality at depths below the surface (J. a Cram et al. 2015). The study by Cram et al. (2015), analyzing data from a times series off the coast of California, found evidence for seasonality at 890 m, corresponding to the depth above the sea floor, and associated this seasonality to changes in sinking particles originating from the surface. The study however, did not find seasonal patterns in bacterial communities at intermediate depths (150 m and 500 m). At 60 m in this study, the beginning of seasonal patterns in bacterial community similarity were evident as were seasonal patterns in bacterial diversity (Figures 3.7, 3.10, and 3.11). The results of seasonal influences on community alpha and beta diversity at 60 m are similar to the results from the Bedford Basin study of 5 m small volume water samples by El Swais et al (2014). Specifically, in the 5 m study, the highest alpha diversity metrics were found in the late fall and early winter as was observed at 60 m in our study, suggesting a propagation of bacterial diversity patterns throughout the water column. Other temporal studies have also observed the highest bacterial diversity in winter months in temperate latitudes (Gilbert *et al.*, 2012; Ladau *et al.*, 2013; Cram *et al.*, 2015). Additionally, the 5 m Bedford Basin study (El-Swais *et al*., 2014) observed large variability with respect to species richness and evenness in spring samples, as we also witnessed at 60 m. In our study the extremely low bacterial diversity metrics in late winter and early spring samples

can be attributed to the few distinct bacterial taxa that become dominant in response to phytoplankton blooms, most likely those taxa in association with phytoplankton and able to degrade polysaccharides. When analyzing temporal patterns of community similarity over the year at 60 m, seasonal patterns were also apparent. This is again similar to what was identified for the previous 5 m study, as our samples were most dissimilar leading up to and at approximately 6 months of separation, and generally became increasingly similar after 6 months of temporal separation (El-Swais *et al.*, 2014).

Community change at multiple levels of taxonomy, from OTUs to class showed rapid changes in relative abundance of taxa at a weekly scale (Figures 3.5 and 3.6), and also broader scale temporal changes over the year (Figures 3.8 and 3.9). Only a few studies have looked at annual bacterial changes at this temporal resolution (El-Swais *et al.*, 2014; Lindh *et al.*, 2015) suggesting that many time series studies might be missing the finer scale dynamics in bacterial community structure. The most abundant families of bacteria at 60 m in the Bedford Basin were Rhodobacteraceae, Pelagibacteraceae, Flavobacteriaceae and SUP05. Members of the orders Rhodobacteriales and Rickettsiales (which contain the families of Rhodobacteraceae and Pelagibacteraceae respectively) were found to be the most abundant groups in the six year English Channel time series (Gilbert *et al.*, 2012). In the English Channel, these two groups exhibited temporal patterns in community dominance that were similar to what was observed at 60 m in the Bedford Basin, with Rhodobacteraceae becoming more abundant in the spring and summer and Pelagibacteraceae becoming more abundant in the fall and early winter. Members of both Rhodobacteraceae and Pelagibacteraceae were also very abundant at 5 m in the previous Bedford Basin time series, however the groups in that study didn't show the same distinct temporal patterns seen here and in the English Channel. This discrepancy could be the result of inter-annual changes in temporal patterns of select taxa, differences across the water column in temporal patterns, or differences in assignment of taxonomy between studies. The 5 m Bedford Basin study used a different variable region (V5 region) and a lower taxonomic cut-off (90%) for OTU (phylotype) identification, so differences in taxonomic assignment between these two studies is highly probable (El-Swais *et al.*, 2014). In general, both groups are ubiquitous, however Pelagibacteraceae, and the affiliated SAR11 cluster, are associated with oligotrophic

conditions (Giovannoni *et al.*, 2005), whereas members of Rhodobacteraceae are generally more associated with productive conditions (Klindworth *et al.*, 2014).

Flavobacteriaceae increased in abundance over the spring and summer seasons, and have previously been associated with phytoplankton blooms (Teeling *et al.*, 2012; El-Swais *et al.*, 2014; Klindworth *et al.*, 2014). In this dataset, Flavobacteriaceae specifically spiked at depth at the same time as the two chlorophyll peaks around weeks 13 and 28 which further support this group's association with phytoplankton. SUP05 is a relatively newly discovered clade of sulfur oxidizing Gammaproteobacteria, that has been suggested to have roles in a variety of different nutrient cycles including carbon, nitrogen, and hydrogen in addition to sulfur (Sunamura *et al.*, 2004; Walsh *et al.*, 2009; Anantharaman *et al.*, 2013). Additionally, the SUP05 clade is consistently found in environments with low oxygen concentrations (Walsh *et al.*, 2009; Wright *et al.*, 2012; Glaubitz *et al.*, 2013). In this study, SUP05 was highly correlated with low oxygen levels, and it became abundant in fall and winter when oxygen levels were lowest. A proteomics study of two time points in the Bedford Basin was conducted previously, and SUP05 was found to be relatively abundant and functionally active in January at 60 m, which corresponds to our findings (Georges *et al.*, 2014). The SUP05 clade has also previously been detected at 5 m in the Bedford Basin in the winter and late fall (El-Swais *et al.*, 2014).

Mantel tests were used to determine the environmental factors that were most correlated with community similarity over the year. Environmental distance between samples better explained community similarity patterns compared to temporal separation between samples (Table 3.1), although both had a significant effect. Silicate differences between samples had a highly significant correlation with bacterial community dissimilarity. As silicate is necessary for diatom frustule production, the strong association between silicate and community similarity suggests that the presence of this nutrient and by extension, diatoms, are important drivers of community composition at 60 m. As diatoms are the main phytoplankton group associated with the spring bloom in the Bedford Basin (Li *et al.*, 2006), this would also suggest that the surface spring bloom largely influences bacterial community structure at depth. Oxygen concentrations were also highly correlated with community similarity. The presence or absence of oxygen is

known to dictate the types of bacteria and biogeochemical cycles that can occur in a given environment (Lam and Kuypers, 2011; Wright *et al.*, 2012). This concept will be discussed more in the following section (section 3.4.4), as low oxygen concentrations were a characteristic feature of fall and early winter. Day length was another highly correlated variable, and can be considered a continuous proxy for seasonality, as early winter and late fall have similar day lengths. In Figure 3.11 samples are ordinated in a circular, seasonal pattern, and the results of this particular Mantel test further supports the strong effect of seasonality at 60 m depth. In the 5 m study of the Bedford Basin by El Swais et al. (2014), they found temperature to have the strongest effect on bacterial community. In our study temperature did have a significant correlation with community similarity, and there were many other variables that were more strongly correlated. This suggests that there are differences in the environmental drivers of bacterial community between the 5 m near surface and the 60 m bottom waters of the Bedford Basin.

3.4.4 Dynamics of Individual OTUs – Rare and Indicator Species at 60 m

The introduction of 16S high throughput sequencing has increased our ability to study rare bacterial taxa and the effects they have on their ecosystems (Bachy and Worden, 2014; Lynch and Neufeld, 2015). The majority of species in this study were part of the rare biosphere, found in less than 50% of samples and never represented by more than 1% of sample sequences. A point to consider regarding the rare biosphere is that "rare" in microbial terms is relative. To put this into perspective, a very rare OTU in this analysis, say only one sequence per 10,000 sequences (roughly 1 in 10,000 cells) will still represent 10^5 cells in a cubic meter, and thus may still strongly influence global biogeochemical cycles (Fuhrman, 2009). The rare biosphere serves as a functional reservoir or a "microbial seed bank" that contains a wide range of species that are available to respond to disturbance events (Caporaso, Paszkiewicz, *et al.*, 2012). Exceptionally diverse, rare microbiota have been considered strategic adaptations to temporally variable environments and contribute to the resilience of functional stability in the marine bacterial community (Lynch and Neufeld, 2015).

The indicator species analysis highlighted specific OTUs that were preferentially abundant during certain times of the year, and as such act as markers for certain seasons

or combinations of seasons (Dufrene and Legendre, 1997; De Cáceres and Legendre, 2009; De Cáceres *et al.*, 2010). As all OTUs with greater than 0.1% average abundance were included in this analysis, both rare and abundant taxa were identified as indicator species. This provides further support that both rare and abundant taxa respond to environmental changes (Sogin *et al.*, 2006; Campbell *et al.*, 2011; Lynch and Neufeld, 2015). This analysis, in addition to providing insight into bacterial temporal dynamics, can also be used in concert with the physical and chemical characteristics of each season to provide evidence for the main biogeochemical cycles occurring and suggest bacterial candidates that may be performing these functions. For instance, many indicator species were identified for the fall season, which could generally be associated with a steady decrease in oxygen, an increase in nitrate, and a decrease in ammonium (Figure 3.4). *Nitrospina* is proposed to be the most abundant NOB (nitrite oxidizing bacteria) in the oceans, oxidizing nitrite and converting it to bioavailable nitrate, facilitating the second step of the two-step nitrification process (Lücker *et al.*, 2013). The chemical signature of the fall season suggests that nitrification is occurring as there is a gradual decrease in ammonium and a corresponding build-up of nitrate (Beman *et al.*, 2010). Additionally, fall and winter seasons can also be characterized by their low oxygen conditions, resulting from stagnant deep water and the respiration of organic matter over time (Platt and Conover, 1975). An OTU from the family of Ectothiorhodospiraceae, was also a strong indicator species for the fall season. This family is known to contain anaerobic purple sulfur bacteria, potentially involved in the sulfur cycle (Imhoff, 2006). Also SUP05 discussed above was an indicator species for both fall and winter, when oxygen concentrations were low. Previous proteomics studies in the basin found SUP05 proteins capable of oxidizing reduced sulfur compounds like thiosulfate and elemental sulfur, but not sulfide, which is generally thought to be absent as a sulfur source in the basin (Georges *et al.*, 2014). Additionally, OTUs belonging to the groups of Methylococcales and SAR324 were also indicator species for fall and winter. Both of these groups have been associated with C1 metabolism (Bowman, 2006; Swan *et al.*, 2011) suggesting a possible C1 cycle in the basin. Georges et al. (2014) also found proteins associated with C1 metabolism in the basin, further supporting this hypothesis.

Spring and summer indicator species included many OTUs from Bacteroidetes and Alphaproteobacteria, for example *Ulvibacter*, *Polaribacter*, Flavobacteriaceae, and Rhodobacteraceae. As discussed previously, these groups generally coincide with species known to degrade phytoplankton and phytoplankton derived polymers. The Bedford Basin is known to experience large phytoplankton blooms at the surface (Li and Dickie, 2001), and although the chlorophyll concentrations at 60 m were consistently quite low, the discovery of known phytoplankton degraders at depth would suggest that the deep bacterial community is heavily influenced by the surface bloom. Increased respiration by heterotrophic bacteria is probably responsible for the gradual oxygen drawdown observed at 60 m consistently seen after these bloom events (Figure 3.4).

A network was generated to observe co-occurrence patterns of the OTUs included in the indicator species analysis. From the structure of the network (Figure 3.13) it is apparent that OTUs cluster on the basis of seasonal preference. The network contained two main co-occurring clusters that were negatively correlated with each other. The slightly larger cluster contained OTUs that were classified as indicator species for fall, fall and winter, or winter, and were negatively associated with oxygen, and positively associated with nutrients and temperature. The second cluster was mutually exclusive from the fall and winter cluster and contained OTUs that were primarily indicator species for spring, spring and summer, and summer. There were many associations on average per OTU suggesting a very highly connected bacterial community. Representatives of *Nitrospina* and SAR324 were the most highly connected OTUs in the network, and interestingly, both *Nitrospina* and SAR324 were also highly connected members of a network compiled for a review of bacterial communities in OMZs (Wright *et al.*, 2012). A potential reason for the many interactions of these groups could be that their metabolism is tightly coupled to the products of other bacteria. For example, *Nitrospina* is a nitrite oxidizer and requires nitrite (product of ammonia oxidizers) and produces nitrate, while SAR324 metabolism is diverse and could potentially include C1 metabolites and sulfur cycling (Lam and Kuypers, 2011; Swan *et al.*, 2011; Wright *et al.*, 2012). Also, SAR324 has been previously identified as a bacterial group that can be particle associated, and thus found in association with the microscale oxyclines that form in these micro environments (Swan *et al.*, 2011). Interactions between the micro

ecosystems that form on these particles could be another potential explanation for the many connections of the SAR324 group. There were also OTUs that were only significantly correlated with only a few other OTUs or environmental variables. These OTUs could also have distinctive roles in their community, and their correlations with particular bacteria could be caused by phenomena other than similar niche preferences. One particular example of an exclusive interaction that would be interesting to explore in future experiments is the positive relationship of ammonium with New Reference OTU26 from the class of Deltaproteobacteria. The strong association of this poorly defined OTU with ammonium potentially suggests that this OTU has an important biogeochemical role, specifically in the nitrogen cycle.

The culmination of the results from the indicator species analysis would suggest that the deep waters of the Bedford Basin have the potential to be metabolically diverse, and that the different metabolic processes could include actively occurring nitrogen, sulfur and carbon cycles. As the Bedford Basin is such an accessible study site, future research could take advantage of this metabolically diverse, and conceivably interdependent community.

Conditionally rare taxa (CRT), "spiking" or "pulse" taxa have recently been defined as organisms that are usually very rare within a community but can become prevalent under certain conditions (Shade *et al.*, 2014; Lindh *et al.*, 2015; Shade and Gilbert, 2015). Identification of these CRTs has been hindered by their fleeting spikes in abundance and rapid return to the obscurity of low abundance, suggesting a need for studies of higher temporal resolution. In this study, weekly samples were analyzed, resulting in higher temporal resolution than most previous time series studies. In this work we identified a number of spiking OTUs, many of which dominated the community for periods of only one week.

Pseudoalteromonas (OTU 827726) perhaps best illustrates the extreme potential for community dominance by CRTs. This particular OTU underwent dramatic escalations in abundance from 0.6% to 64% in the course of one week, then immediately decreased in abundance to 0.02% by the next week. The fact that this observation of transient community dominance was only apparent for a week suggests that current studies may be missing very interesting bacterial dynamics. Additionally, the short time scale of this

change makes the case for even more frequent sampling efforts, particularly in times where blooms are to be expected, in order to monitor how these select species go from near obscurity to dominance so quickly. Increases in temporal sampling effort will obviously be more expensive, but with substantially lowered sequencing costs, these types of high temporal resolution studies should become more common.

Pseudoalteromonas has previously been of interest for its algicidal properties (Lovejoy *et al.*, 1998; Bowman, 2007). Interestingly both large 60 m spikes coincided with spring phytoplankton bloom events in the surface waters, where surface chlorophyll reached some of its highest levels all year. Also the *Pseudoalteromonas* blooms propagated throughout the water column as evidenced by the similar temporal dynamics of the OTU at 5 m (Figure 3.14). Overall, the bacterial dynamics observed post spring blooms in the basin suggest that there is a fairly characteristic response of bacterial groups like *Polaribacter*, *Ulvibacter*, Flavobacteriales, and Rhodobacteraceae (Teeling *et al.*, 2012; Klindworth *et al.*, 2014) as shown by the indicator analysis and some conditionally rare OTUs, but also the potential for an extremely dynamic, short-lived response of a few bacterial OTUs mostly belonging to the *Pseudoalteromonas* genus. The fact that the study by El Swais et al (2014) did not observe these spikes could be the result of their lower sampling frequency (biweekly), or perhaps that these dramatic spikes in abundance do not occur every year. More research needs to be conducted in order to understand the exciting dynamics of this potentially very active bacterial community member and its subsequent effect on the rest of the bacterial community.

Other prominent CRTs at 60 m included *Arcobacter*, *Psychrobacter* and *Erythrobacteraceae*, which respectively spiked to 17%, 7.6%, and 7.2% of relative community abundance, essentially over one week. Due to the nature of most time series studies, these pulsing bacterial taxa are only beginning to be identified and studied. In the weekly time series of the surface Baltic Sea community by Lindh et al (2015), no OTUs from the aforementioned taxa were identified as "pulsing" OTUs. Instead they found pulsing representatives from *Sulfurimonas*, *Synechococcus*, Verrucomicrobiaceae, Flavobacteriaceae as well as other groups. The biweekly study of the Bedford Basin identified two groups of bacteria that exhibited rare blooms, one from Sphingobacteria and one from OM1 (El-Swais *et al.*, 2014). The Sphingobacteria group was thought to

have bloomed in response to an anthropogenic environmental disturbance potentially involving some sort of hydrocarbon substance, and the OM1 group was thought to have bloomed in response to a phytoplankton bloom. Studying these transiently abundant groups can provide information on how bacterial communities respond to environmental change (both natural and anthropogenic), as well as provide information on how these groups can grow so rapidly in addition to why they disappear so quickly afterwards (viruses, predation, or an alternate factor). The wide diversity of bacterial taxa apparently capable of a "boom and bust" lifestyle and the lack of previous studies with the ability to identify these organisms suggests that these conditionally rare taxa should be a very interesting emerging area of research in the future (Shade and Gilbert, 2015).

3.4.5 Intrusion Events

The Bedford Basin is separated from the Scotian Shelf by a long and narrow sill that usually acts as a barrier to mixing of Shelf waters with deep bottom Basin waters (Figure 3.1). Occasionally however, significant storm events, particularly those with alongshore winds, result in upwelling of shelf water into the basin due to Ekman transport, which can dramatically change the characteristics of Bedford Basin water at depth rapidly (Platt *et al.*, 1972; Burt *et al.*, 2013). Burt et al. (2013) followed the physiochemical dynamics of one such intrusion event in October of 2010. Specifically, they found that the intrusion event resulted in abrupt changes in the physical and chemical characteristics of the deep-water column and brought oxygen-saturated shelf surface waters into the deep basin, which prevented the deep waters of the basin from reaching their characteristic low oxygen states. During the course of the 2014 sampling reported here, two similar shelf water intrusions, or deep mixing events, were identified. One event occurred in late winter from weeks 10-11 (March 5-12, 2014) and resulted in cooler, more oxygenated waters entering and mixing with the deep basin water. In this first instance, the abrupt change in water conditions at depth in the basin could also be attributed to a general basin mixing event that resulted in surface water nutrient renewal and the start of the surface spring bloom. It is difficult to pinpoint the exact origin of the new water mass as information from the shelf from this time is not available. The second event, occurred in early summer from weeks 28-29 (July 8-July 15, 2014) and resulted in

warmer, saltier, and more oxygenated waters entering the deep basin. Hurricane Arthur hit Nova Scotia July 5th and brought strong winds and rain to the Halifax region. Although the intrusion and the hurricane don't line up exactly, it is probable that the aftermath of this storm caused the dramatic shelf water intrusion and deep water mixing event.

For each intrusion event, the bacterial communities from the weeks preceding the intrusion were more similar to each other than the bacterial communities from the weeks post intrusion (Figure 3.18). This demonstrates that these deep mixing events had a profound effect on the bacterial community. Few marine *in situ* studies on bacterial community change in response to a dramatic disturbance in the natural environment have been conducted, and the majority of these few have focused on oil spills (Redmond and Valentine, 2012; Rivers *et al.*, 2013). Within this time series, we were able to monitor high resolution changes in bacterial community in response to natural, dynamic water mass shifts that were in part characterized by steep increases in oxygen concentrations. A recent study by Forth et al (2014) experimentally manipulated the oxygen content in an anoxic inlet in the Baltic Sea by pumping oxygenated surface sea water in mass quantities into the anoxic bottom layer until the desired levels of oxygen were reached. In this study, we have the ability to monitor how a naturally occurring version of this experiment affects the bacterial community, consequently using the Bedford Basin as a "natural laboratory".

In general, the Bedford Basin bottom waters do not become as anoxic as the Baltic Sea, but during 2014, oxygen levels were low enough (1.1 mL/L; ~47 μmol/kg) to allow for known microaerophilic bacteria like SUP05 to gain abundance. In Forth's study, OTUs from SUP05, *Arcobacter*, *Sulfurimonas*, *Desulfocapsa*, *Chlorobium,* Spirochaetes and Flavobacteriaceae were most abundant in the sulfidic and anoxic pre oxygenation conditions, while members of SAR11 (Pelagibacteraceae) and SAR86 increased in abundance post oxygenation (Forth *et al.*, 2014). In that study however, the techniques used for sequencing and OTU identification only allowed detection of OTUs that were greater than 1% of the community and the sampling times were few and relatively infrequent. With the weekly Bedford Basin time series, the more intricate patterns in community change in response to abrupt environmental disturbances can be

detected, and as these intrusion and mixing events occur annually, these changes can be monitored into the future as well.

In the Baltic Sea (Forth *et al.*, 2014), *Arcobacter* was a key anoxic bacterium. Marine *Acrobacter* are known as a microaerophilic sulfide oxidizing bacteria that are often found in the vicinity of oxic-anoxic interfaces (Sievert *et al.*, 2007). In our study, *Arcobacter* was not associated with low oxygen conditions, but was instead identified as one of the CRTs that spiked in abundance in late spring, a period of time that was well oxygenated. The discrepancies between oxygen concentrations of the *Arcobacter* habitat in these two studies suggests either that there are potential ecotypes of this genus that respond to different environmental conditions, or that *Arcobacter* can be particle associated, and thus potentially situated in a microaerophilic environment (Alldredge and Silver, 1988). Additionally, *Sulfurimonas* wasn't found to any appreciable level in this study, suggesting there is a fundamental difference between the dominant bacterial communities of sulfide containing and sulfide lacking low oxygen environments.

After the oxygenation event in the Baltic Sea (Forth *et al.*, 2014), SAR11 and SAR86 which are both known oligotrophs (Dupont *et al.*, 2012), increased dramatically. In the Bedford Basin OTUs associated with phytoplankton degradation (e.g. *Polaribacter* and *Ulvibacter*) became much more abundant after the first mixing event. As this event coincided with the dramatic spring bloom in the surface waters this would suggest that the mixing event provided a source of phytoplankton rich, oxygenated waters to the deep. Two weeks after this first mixing event is when the large *Pseudoalteromonas* bloom occurred, suggesting these OTUs were also responding to the conditions established by the intrusion. The second intrusion event resulted in a less dramatic community change, mainly made up of increasing *Polaribacter*, Rhodobacteraceae and Colwelliaceae OTUs. Also a few weeks after the second intrusion the community appeared to slowly revert to a more pre-intrusion like structure, signifying that the community was beginning to re-stabilize. The larger change in community structure in the first event compared to the second event could be explained by the lower initial oxygen concentrations prior to the first event. Overall, both these environmental perturbations provide a high quality opportunity to study bacterial community dynamics in response to changing conditions. Specifically, understanding bacterial responses to changes in oxygen and temperature

will become of increasing importance to the global ocean as temperatures are expected to rise and oxygen concentrations are expected to decrease in the future (Bindoff *et al.*, 2007; Stramma *et al.*, 2008).

3.4.6 Conclusion

This work provides part of the first analyses from the Bedford Basin microbial community structure time series. This specific addition to the established Bedford Basin time series was introduced in 2014 and since then, we have received weekly samples from multiple depths spanning both surface and deep layers in the Bedford Basin. This particular chapter focused on the 60 m depth throughout the year. Seasonal patterns in individual taxa, diversity, and community similarity were identified at depth. The single most abundant OTU at 60 m depth belonged to the family of Pelagibacteraceae. However, at the family level, the collective OTUs that belong to the family of Rhodobacteraceae made up the greatest proportion of community abundance, followed by the Pelagibacteraceae family of OTUs. These two families exhibited strong seasonal patterns with Rhodobacteraceae dominating the community in spring and summer weeks and Pelagibacteraceae dominating the community in fall. Many metabolically diverse taxa were discovered in the basin at depth, particularly in association with the low oxygen conditions of the fall and winter. The presence of these taxa suggest that there are potentially active sulfur, carbon, nitrogen, and C1 carbon cycles occurring at depth at various times of the year. Additionally, multiple extreme, but brief spikes of normally rare taxa were observed in the basin due to this study's high temporal resolution. The most dramatic incidence was from the genus *Pseudoalteromonas* in late winter and early spring where three OTUS from the normally rare taxa dominated the community and combined, accounted for over 85% of relative abundance. Two large mixing events resulted in rapid changes in environmental parameters in the deep basin which brought about dramatic shifts of the bacterial community composition. Overall, this work provides the foundation for many more high quality studies in the Bedford Basin and has identified multiple exciting new avenues to explore with high temporal resolution sampling.

Chapter 4: Conclusion

4.1 Comparing Scotian Shelf and Bedford Basin

Both chapters 2 and 3 in this study analyzed bacterial community structure of the Northwest Atlantic Ocean but with different focuses. Chapter 2 of this study investigated the bacterial communities of the Scotian Shelf across space, time, and depth. Chapter 3 of this study focused on one site and mainly one depth, but examined bacterial community structure at a very high temporal resolution made possible by weekly sampling. Although the analysis from these chapters examined bacterial community structure with ultimately different aims and sampling strategies, they both produced results that can be compared in many regards.

The dominant bacterial taxa on the Scotian Shelf and in the Bedford Basin were Rhodobacteraceae and Pelagibacteraceae. On the Scotian Shelf, Pelagibacteraceae were more common than Rhodobacteraceae, but the opposite was generally true in the Bedford Basin when considering annual averages. The most dominant OTU from both datasets was Pelagibacteraceae OTU 637092, and in both datasets it represented just over 12.5% of the total community which demonstrates some continuity of bacterial composition between the basin and the shelf.

The highest spiking OTU in the Scotian Shelf, belonging to the family of Rhodobacteraceae (OTU 569286), spiked to 44% of the relative community, and was on average 4% of the bacterial community across all sites. At 60 m in the Bedford Basin however, the same OTU never reached above 1% and only represented 0.1% on average of the community abundance. Although OTUs from Rhodobacteraceae were very common in both datasets, there were noteworthy variations in the particular OTUs that were abundant at any given site or time which suggests that even in adjacent regions, species of particular taxa will behave and respond differently to the local environment.

There were also dramatic spikes in both datasets of the generally rare genus *Pseudoalteromonas* at select stations and times of year. High chlorophyll concentrations and phytoplankton blooms generally coincided with *Pseudoalteromonas* dominance in the Bedford Basin, whereas *Pseudoalteromonas* spikes seemed to be responding to more variable environmental circumstances in the Scotian Shelf. In particular,

Pseudoalteromonas on the Scotian Shelf was significantly associated with the open ocean, and was quite strongly negatively associated with bacterial abundance, possibly due to the production of anti-fouling compounds (Lovejoy *et al.*, 1998; Bowman, 2007). The main spikes of *Pseudoalteromonas* on the Scotian Shelf were at open ocean stations HL11 (all depths, strongest at 250 m), HL8 (all depths, strongest at 250 m and 100 m) in the fall, and the genus was comparatively more abundant in the spring, with the largest spikes at the coastal BBL1 station. The fact that *Pseudoalteromonas* spiked in the open ocean, near the coast, and in the Bedford Basin, throughout the water column in each instance, suggests that *Pseudoalteromonas* is quite widely distributed and able to bloom in many marine environments. The main OTUs (those that reached above 1% relative community abundance) responsible for the spikes in the Bedford Basin and the Scotian Shelf are compiled in Table 4.1. The Scotian Shelf *Pseudoalteromonas* spiking cohort contained many more OTUs than the Bedford Basin, and interestingly didn't include the relatively abundant OTU from the Bedford Basin, OTU 231442. The range in *Pseudoalteromonas* OTUs suggests possible ecotypes and variations in spiking triggers within the *Pseudoalteromonas* genus that would be very interesting to investigate further.

Table 4.1. Comparison of the main *Pseudoalteromonas* OTUs (those that reached above 1% relative community abundance) across Scotian Shelf and Bedford Basin datasets. OTUs are arranged based on decreasing max abundance and OTUs shared between the two datasets are shown in bold.

Scotian Shelf OTUs	**Max Abundance**	**Bedford Basin OTUs**	**Max Abundance**
827726	**25.9%**	**827726**	**63.8%**
145002	13.7%	231442	15.9%
New ReferenceOTU12	13.2%	**40554**	**7.7%**
8965	6.3%		
526324	4.7%		
New ReferenceOTU27905	3.4%		
40554	**2.6%**		
1931738	1.9%		
935060	1.7%		
561543	1.2%		

The bacterial communities of both the Scotian Shelf and the Bedford Basin were heavily influenced by season and seasonal variations in environmental variables

suggesting a common driver of bacterial change in the temperate Northwest Atlantic region. In the Scotian Shelf, season was the strongest driver of community change and whether a sample was taken during the spring or fall cruise strongly affected the bacterial community's response to different environmental variables. Sites from the same seasons were more similar with respect to bacterial communities (Figure 2.13), there were many bacterial taxa and OTUs that were preferentially more abundant in either the spring or the fall (Figures 2.14 and 2.16), and the conditions affecting alpha and beta diversity were very different in the spring compared to the fall (Tables 2.2 and 2.4). Seasonal controls on phytoplankton and eukaryote concentrations and composition have been identified previously on the Scotian Shelf (Li *et al.*, 2006; Dasilva *et al.*, 2014), and this work provides evidence that this seasonal signal is propagated to the bacterial community. In the Bedford Basin, weekly dynamics in bacterial composition were commonplace, however smoothed trends showed definite seasonal preferences of taxa and OTUs (Figures 3.8 and 3.9). Additionally, both bacterial diversity (Figure 3.7) and community similarity (Figure 3.10) showed seasonal patterns at 60 m, which is the first time this has been shown for the bacterial community in the Bedford Basin at this depth, and is in agreement with what had previously been observed at 5 m (El-Swais *et al.*, 2014).

Although the Scotian Shelf and the Bedford Basin are geographically close and connected (Shan and Sheng, 2012), there are distinct environmental differences between the two regions. The main difference over the course of the study was that the Bedford Basin experienced extremely low oxygen concentrations which sometimes reached near anoxic concentrations at depth (~1 mL/L; ~43 µmol/kg) but the Scotian Shelf remained comparably oxygenated throughout, even at the depths corresponding to the local oxygen minima (~3 mL/L; ~129 µmol/kg). Oxygen concentrations are very important for determining the types of metabolic lifestyles that are able to flourish, and thus extreme variations in oxygen can result in very diverse bacterial communities and biogeochemical processes (Lam and Kuypers, 2011). This was particularly evident in the Bedford Basin where taxa like SUP05, a known anoxic bacterium (Walsh *et al.*, 2009; Georges *et al.*, 2014), became relatively abundant during the hypoxic fall and early winter conditions. When analyzing the comparatively oxygenated Scotian Shelf, SUP05 was all but absent. In addition, in both studies, similarity of oxygen concentrations between samples was

highly significantly correlated with bacterial community similarity between samples (Table 2.3, Table 3.1), providing further evidence of its importance in shaping bacterial community composition.

In both studies, the presence of nutrients (nitrate, phosphate and silicate) were significantly correlated with community structure (Table 2.3, Table 3.1). Additionally, phytoplankton blooms and the corresponding increase in available organic substrate, dramatically shifted bacterial community assemblages in the spring in both the Scotian Shelf and the Bedford Basin. In both regions, select taxa, including members from Rhodobacteraceae, *Ulvibacter*, *Polaribacter*, and other Bacteroidetes became relatively abundant in response to these phytoplankton blooms, a shift in bacterial community structure that has been observed in other studies of temperate oceans (Teeling *et al.*, 2012; Klindworth *et al.*, 2014), including the Bedford Basin at 5 m (El-Swais *et al.*, 2014).

4.2 Future Directions

4.2.1 Future Directions Scotian Shelf

The study of the bacterial community on the Scotian Shelf produced a very rich dataset with extensive 16S rRNA sequencing data accompanied by corresponding environmental information of the region. This data can be used as a baseline for countless future studies and as a future resource for testing new bioinformatics techniques. Additionally, the 3 µm pre-filters of this sampling effort have yet to be analyzed and could provide an interesting comparison of what is assumed to be free-living bacteria (less than 3 µm) and host or particle associated bacteria (between 3 µm and 160 µm). Both of these size fractions could be analyzed for Archaeal and eukaryotic content as well, to broaden the biological breadth of this dataset and provide opportunity for identifying associations between different taxonomic domains. Additionally, DNA remains from the original extraction and can be used for further analysis with qPCR of specific functional genes or to improve the quantitative aspect of the study through use of taxa or OTU specific primers.

This study provides a thorough baseline of the bacterial community of the Scotian Shelf with which to compare samples from future years to monitor how the area changes over time, particularly in response to climate change and to other anthropogenic disturbances. Future studies of this type can also strengthen the links discovered in this study between environmental variables and taxonomic distribution, as well as provide information on inter versus intra-annual variation in the region. Additionally transcriptomics and proteomics analyses can be conducted to compare abundance data with data derived from metabolically active community components, and explore the relationship between members of the dormant and active microbial communities. As this study addressed the question of "who is there", these follow up studies can provide more insight into "what they are doing" in their environment.

Future directions with this dataset could also include 'metabolic inference' analysis with the PICRUSt (Langille *et al.*, 2013) or Paprica programs (Bowman and Ducklow, 2015). These programs take 16S marker gene phylogenetic information and predict community metabolic function, essentially bypassing the transcriptomics and proteomics steps which can still be technically difficult and costly. This allows for a more holistic view of the metabolic operational capacity of the bacterial community and will produce better inferences on biogeochemical cycling in the region (Xu *et al.*, 2014). Additionally, these programs can roughly account for variations in 16S rRNA gene copy number which should provide more accurate relative abundance estimates.

4.2.2 Future Directions Bedford Basin

The second part of this study established a weekly biological time series in the Bedford Basin using 16S rRNA sequencing. Already, seasonality and short-lived bacterial dynamics have been observed at all depths. As this time series continues into the future, it will be very interesting to study both fine scale and broad scale changes in bacterial communities in this dynamic North Atlantic basin. Specific future directions include analyzing the eukaryotic community of the Bedford Basin through 18S rRNA gene sequencing. Additionally, samples have been collected monthly from each depth, to characterize the transcriptome of the microbial community in the basin. Using these results in concert with the 16S rRNA gene data already collected will provide further

information on the functional capacity of the bacterial community in the basin over the year. Also as DNA from the 2014 samples is still remaining after 16S rRNA gene sequencing, many other experiments requiring DNA can be performed, including qPCR of functional genes of interest (e.g. *amoA*) and metagenomics studies of dominant community members (e.g. *Pseudoalteromonas*). Many bacterial OTUs were not identified in the Greengenes database (i.e. New Reference OTUs), and could thus be potentially new species. These OTUs would also be interesting to research further, particularly those with strong associations to other taxa or environmental variables, such as the unclassified Deltaproteobacteria New ReferenceOTU26 that was strongly associated with ammonium.

Furthermore, periodic deep water mixing events in the Bedford Basin caused dramatic environmental changes at depth, in particular a rapid oxygenation of the bottom water. With knowledge of when these events are likely to occur (e.g. after strong storms), higher frequency sampling around these periods can be conducted, providing a higher resolution time series that can detail the fine-scale dynamics of a shifting bacterial community. The Bedford Basin is also heavily trafficked by ships and is surrounded by the HRM which means that there are many opportunities for anthropogenic stressors. The effects of anthropogenic influence on the Bedford Basin will be interesting to continue to monitor in the future, especially in the case of known spill events regarding wastewater, industrial runoff, or hydrocarbons.

Overall, this time series establishes a protocol that will allow for the continuation of high quality, high resolution biological samples in the Bedford Basin that can be used to study an extensive range of topics from microbial interactions, to short term community responses to anthropogenic inputs, to long term community responses to climate change.

References

Abell, G.C.J. and Bowman, J.P. (2005) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol. Ecol.* **51**: 265–77.

Affholder, M. and Valiron, F. (2001) Descriptive Physical Oceanography CRC Press.

Agogué, H., Lamy, D., Neal, P.R., Sogin, M.L., and Herndl, G.J. (2012) Water mass-specificuty of bacterila communities in the North Atlantic revealed by massively parallel sequencing. *Mol. Ecol.* **20**: 258–274.

Alldredge, A.L. and Silver, M.W. (1988) Characteristics, dynamics and significance of marine snow. *Prog. Oceanogr.* **20**: 41–82.

Alonso-Sáez, L., Arístegui, J., Pinhassi, J., Gómez-Consarnau, L., González, J.M., Vaqué, D., et al. (2007) Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean.

Alonso-Sáez, L., Zeder, M., Harding, T., Pernthaler, J., Lovejoy, C., Bertilsson, S., and Pedrós-Alió, C. (2014) Winter bloom of a rare betaproteobacterium in the Arctic Ocean. *Front. Microbiol.* **5**: 425.

Anantharaman, K., Breier, J. a, Sheik, C.S., and Dick, G.J. (2013) Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci U S A* **110**: 330–335.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**: 32–46.

Anderson, M.J. (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**: 245–253.

Arístegui, J., Gasol, J.M., Duarte, C.M., and Herndld, G.J. (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol. Oceanogr.* **54**: 1501–1529.

Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. *Nature* **437**: 349–355.

Atlas, R.M. (1995) Petroleum biodegradation and oil spill bioremediation. *Mar. Pollut. Bull.* **31**: 178–182.

Azam, F. (1998) OCEANOGRAPHY: Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science (80-.).* **280**: 694–696.

Baas Becking, L. (1934) Geobiologie of inleiding tot de milieukunde W.P. Van Stockum & Zoon, Den Haag.

Bachy, C. and Worden, A.Z. (2014) Microbial ecology: finding structure in the rare biosphere. *Curr. Biol.* **24**: R315–7.

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**: 343–351.

Barth, a., Alvera-Azcárate, a., Troupin, C., Ouberdous, M., and Beckers, J.M. (2010) A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA). *Adv. Geosci.* **28**: 29–37.

Beman, J.M., Chow, C., King, A.L., Feng, Y., and Fuhrman, J. a (2010) Global declines in oceanic nitri fi cation rates as a consequence of ocean acidi fi cation.

Beman, J.M., Steele, J.A., and Fuhrman, J.A. (2011) Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. *ISME J.* **5**: 1077–1085.

Berry, D. and Widder, S. (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**: 1–14.

Bindoff, N., Willebrand, J., Artale, V., Cazenave, A., Gregory, J., Gulev, S., et al. (2007) Observations: Oceanic Climate Change and Sea Level.

Borcard, D., Legendre, P., and Drapeau, P. (1992) Partialling out the Spatial Component of Ecological Variation. *Ecology* **73**: 1045.

Bowman, J. (2006) The Methanotrophs — The Families Methylococcaceae and Methylocystaceae. *The Prokaryotes* **5**: 266–289.

Bowman, J.P. (2007) Bioactive compound synthetic capacity and ecological significance of marine bacterial genus Pseudoalteromonas. *Mar. Drugs* **5**: 220–241.

Bowman, J.S. and Ducklow, H.W. (2015) Microbial Communities Can Be Described by Metabolic Structure: A General Framework and Application to a Seasonally Variable, Depth-Stratified Microbial Community from the Coastal West Antarctic Peninsula. *PLoS One* **10**: e0135868.

Brakstad, O.G., Nonstad, I., Faksness, L.G., and Brandvik, P.J. (2008) Responses of microbial communities in Arctic sea ice after contamination by crude petroleum oil. *Microb. Ecol.* **55**: 540–552.

Brander, K. and Hurley, P.C.F. (1992) Distribution of Early-Stage Atlantic Cod (Gadus morhua), Haddock (Melanogrammus aeglefinus), and Witch Flounder (Glyptocephalus cynoglossus) Eggs on the Scotian Shelf: A Reappraisal of Evidence on the Coupling of Cod Spawning and Plankton Production. *Can. J. Fish. Aquat. Sci.* **49**: 238–251.

Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Boone, D.R., De Vos, P., et al. eds. (2005) Bergey's Manual® of Systematic Bacteriology Springer US, Boston, MA.

Bruckner, C.G., Mammitzsch, K., Jost, G., Wendt, J., Labrenz, M., and Jürgens, K. (2013) Chemolithoautotrophic denitrification of epsilonproteobacteria in marine pelagic redox gradients. *Environ. Microbiol.* **15**: 1505–13.

Burt, W.J., Thomas, H., Fennel, K., and Horne, E. (2013) Sediment-water column fluxes of carbon, oxygen and nutrients in Bedford Basin, Nova Scotia, inferred from 224Ra measurements. *Biogeosciences* **10**: 53–66.

Buttigieg, P.L. and Ramette, A. (2014) A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.* **90**: 543–550.

De Cáceres, M. and Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566–74.

De Cáceres, M., Legendre, P., and Moretti, M. (2010) Improving indicator species analysis by combining groups of sites. *Oikos* **119**: 1674–1684.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile ε-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4**: 458–468.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 12776–12781.

Cao, Y., Chastain, R.A., Eloe, E.A., Nogi, Y., Kato, C., and Bartlett, D.H. (2014) Novel psychropiezophilic Oceanospirillales species Profundimonas piezophila gen. nov., sp. nov., isolated from the deep-sea environment of the Puerto Rico trench. *Appl. Environ. Microbiol.* **80**: 54–60.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–7.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) correspondEnce QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Publ. Gr.* **7**: 335–336.

Caporaso, J.G., Lauber, C.L., Walters, W. a, Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**: 1621–1624.

Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R., and Gilbert, J. a (2012) The Western English Channel contains a persistent microbial seed bank. *ISME J.* **6**: 1089–1093.

Chaffron, S., Rehrauer, H., Pernthaler, J., and Mering, C. (2010) A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **2010**: 947–959.

Cho, J.-C., Vergin, K.L., Morris, R.M., and Giovannoni, S.J. (2004) Lentisphaera araneosa gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. *Environ. Microbiol.* **6**: 611–621.

Chow, C.-E.T., Sachdeva, R., Cram, J. a, Steele, J. a, Needham, D.M., Patel, A., et al. (2013) Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *ISME J.* **7**: 1–15.

Comeau, A.M. and Langille, M.G.I. Microbiome Amplicon Sequencing Workflow.

Comeau, A.M., Li, W.K.W., Tremblay, J.É., Carmack, E.C., and Lovejoy, C. (2011) Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One* **6.**:

Craig, S.E., Thomas, H., Jones, C.T., Li, W.K.W., Greenan, B.J.W., Shadwick, E.H., and Burt, W.J. (2014) The effect of seasonality in phytoplankton community composition on CO2 uptake on the Scotian Shelf. *J. Mar. Syst.*

Cram, J.A., Chow, C.-E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A., and Fuhrman, J.A. (2015) Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* **9**: 563–580.

Cruaud, P., Vigneron, A., Lucchetti-Miganeh, C., Ciron, P.E., Godfroy, A., and Cambon-Bonavita, M.A. (2014) Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl. Environ. Microbiol.* **80**: 4626–4639.

Dasilva, C.R., Li, W.K.W., and Lovejoy, C. (2014) Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean. *J. Plankton Res.* **36**: 344–363.

Dastager, S.G., Lee, J.-C., Ju, Y.-J., Park, D.-J., and Kim, C.-J. (2009) Nocardioides sediminis sp. nov., isolated from a sediment sample. *Int. J. Syst. Evol. Microbiol.* **59**: 280–4.

DeBruyn, J.M., Nixon, L.T., Fawaz, M.N., Johnson, a. M., and Radosevich, M. (2011) Global Biogeography and Quantitative Seasonal Dynamics of Gemmatimonadetes in Soil. *Appl. Environ. Microbiol.* **77**: 6295–6300.

Deng, S., Chang, X., Zhang, Y., Ren, L., Jiang, F., Qu, Z., and Peng, F. (2015) Nocardioides antarcticus sp. nov., isolated from marine sediment. *Int. J. Syst. Evol. Microbiol.* **65**: 2615–2621.

Denner, E.B.M., Vybiral, D., Koblízek, M., Kämpfer, P., Busse, H.-J., and Velimirov, B. (2002) Erythrobacter citreus sp. nov., a yellow-pigmented bacterium that lacks bacteriochlorophyll a, isolated from the western Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* **52**: 1655–61.

Diaz, R.J. and Rosenberg, R. (2008) Spreading dead zones and consequences for marine ecosystems. *Science* **321**: 926–929.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**: 927–930.

Dong, J. and Horvath, S. (2007) Understanding network concepts in modules. *BMC Syst. Biol.* **1**: 24.

Drinkwater, K.F. and Gilbert, D. (2004) Hydrographic variability in the waters of the Gulf of St. Lawrence, the Scotian Shelf and the eastern Gulf of Maine (NAFO Subarea 4) during 1991-2000. *J. Northwest Atl. Fish. Sci.* **34**: 85–101.

Dufrene, M. and Legendre, P. (1997) Species Assamblages and Indicator Species: The Need for a Flexible Asymetrical Approach. *Ecol. Monogr.* **67**: 345–366.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.-J., Alexander Richter, R., Valas, R., et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**: 1186–1199.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–200.

El-Swais, H., Dunn, K. a., Bielawski, J.P., Li, W.K.W., and Walsh, D. a. (2014) Seasonal assemblages and short-lived blooms in coastal north-west Atlantic Ocean bacterioplankton. *Environ. Microbiol.* **2424**: n/a–n/a.

Epstein, S. (2013) The phenomenon of microbial uncultivability. *Curr. Opin. Microbiol.* **16**: 636–642.

Fader, G.B.J. and Miller, R.O. (2008) Surficial geology, Halifax Harbour, Nova Scotia.

Falkowski, P.G. (1998) Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science (80-.).* **281**: 200–206.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science (80-.).* **320**: 1034–1039.

Faust, K. and Raes, J. (2012) Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**: 538–550.

Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. (2012) Microbial Co-occurrence Relationships in the Human Microbiome. *PLoS Comput. Biol.* **8**: e1002606.

Finkel, Z. V., Beardall, J., Flynn, K.J., Quigg, A., Rees, T.A. V, and Raven, J. a. (2010) Phytoplankton in a changing world: Cell size and elemental stoichiometry. *J. Plankton Res.* **32**: 119–137.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., et al. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 9824–9.

Forth, M., Liljebladh, B., Stigebrandt, A., Hall, P.O.J., and Treusch, A.H. (2014) Effects of ecological engineered oxygenation on the bacterial community structure in an anoxic fjord in western Sweden. *ISME J.* **1973**: 1–14.

Francis, C.A., Beman, J.M., and Kuypers, M.M.M. (2007) New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J.* **1**: 19–27.

Freitas, S., Hatosy, S., Fuhrman, J. a, Huse, S.M., Mark Welch, D.B., Sogin, M.L., and Martiny, A.C. (2012) Global distribution and diversity of marine Verrucomicrobia. *ISME J.* **6**: 1499–1505.

Fu, Y., Keats, K.F., Rivkin, R.B., and Lang, A.S. (2013) Water mass and depth determine the distribution and diversity of Rhodobacterales in an Arctic marine system. *FEMS Microbiol. Ecol.* **84**: 564–576.

Fuhrman, J. a, Hewson, I., Schwalbach, M.S., Steele, J. a, Brown, M. V, and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 13104–13109.

Fuhrman, J. a., Cram, J. a., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**: 133–146.

Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature* **459**: 193–199.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M. V, Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 7774–7778.

Galand, P.E., Potvin, M., Casamayor, E.O., and Lovejoy, C. (2010) Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J.* **4**: 564–576.

Ganesh, S., Parris, D.J., DeLong, E.F., and Stewart, F.J. (2014) Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* **8**: 187–211.

Gatza, E., Hammes, F., and Prest, E. (2013) Assessing Water Quality with the BD Accuri C6 Flow Cytometer. *BD Biosci.* 1–11.

Georges, A. a, El-Swais, H., Craig, S.E., Li, W.K., and Walsh, D. a (2014) Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME J.* 1–13.

Gihring, T.M., Green, S.J., and Schadt, C.W. (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ. Microbiol.* **14**: 285–290.

Gilbert, J. a, Steele, J. a, Caporaso, J.G., Steinbruck, L., Reeder, J., Temperton, B., et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J.* **6**: 298–308.

Giovannoni, S.J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J.* **8**: 1553–1565.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–5.

Glaubitz, S., Kießlich, K., Meeske, C., Labrenz, M., and Jürgens, K. (2013) SUP05 Dominates the gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central baltic and black seas. *Appl. Environ. Microbiol.* **79**: 2767–2776.

Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E.E., and Amann, R. (2010) Distinct flavobacterial communities in contrasting water masses of the north Atlantic Ocean. *ISME J.* **4**: 472–487.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., et al. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proc. Natl. Acad. Sci. U. S. A.* **105**: 8724–8729.

Gosink, J.J., Woese, C.R., and Staley, J.T. (1998) Polaribacter gen. nov., with three new species, P. irgensii sp. nov., P. franzmannii sp. nov. and P. filamentus sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of 'Flectobacillus glomera. *Int. J. Syst. Bacteriol.* **48 Pt 1**: 223–35.

Gotelli, N.J. and Colwell, R.K. (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**: 379–391.

Green, J. and Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**: 501–507.

Grote, J., Labrenz, M., Pfeiffer, B., Jost, G., and Jürgens, K. (2007) Quantitative distributions of Epsilonproteobactena and a Sulfurimonas subgroup in pelagic redoxclines of the central Baltic Sea. *Appl. Environ. Microbiol.* **73**: 7155–7161.

Hannah, C.G., Shore, J. a., Loder, J.W., and Naimie, C.E. (2001) Seasonal Circulation on the Western and Central Scotian Shelf*. *J. Phys. Oceanogr.* **31**: 591–615.

Hanson, C. a., Fuhrman, J. a., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**: 1–10.

Hatosy, S.M., Martiny, J.B.H., Sachdeva, R., Steele, J., Fuhrman, J.A., and Martiny, A.C. (2013) Beta diversity of marine bacteria depends on temporal scale. *Ecology* **94**: 1898–1904.

Hawley, a. K., Brewer, H.M., Norbeck, a. D., Pa a-Toli, L., and Hallam, S.J. (2014) Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci.* **111**: 11395–11400.

Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J., and Moore, B.S. (2002) Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans. *Appl. Environ. Microbiol.* **68**: 4431–4440.

Herbert, R.A. (1999) Nitrogen cycling in coastal marine ecosystems. *FEMS Microbiol. Rev.* **23**: 563–590.

Hewson, I., Steele, J. a, Capone, D.G., and Fuhrman, J. a (2006) Temporal and spatial scales of oligotrophic surface water bacterioplankton assemblage variation. *Mar. Ecol. Prog. Ser.* **311**: 67–77.

Holmström, C. and Kjelleberg, S. (1999) Marine Speudoalteromonas species are associated with higter organisms and produce biologically active extracellular agents. *FEMS Microb. Ecol.* **30**: 285–293.

Hong, S., Bunge, J., Leslin, C., Jeon, S., and Epstein, S.S. (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* **3**: 1365–1373.

Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., et al. (2013) Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.

Imhoff, J. (2006) The family Ectothiorhodospiraceae. *The prokaryotes* 874–886.

Jacoby, W.G. (2000) Loess: A nonparametric, graphical tool for depicting relationships between variables. *Elect. Stud.* **19**: 577–613.

Jeong, I.H., Kim, K.H., Lee, H.S., and Park, J.S. (2014) Analysis of bacterial diversity in sponges collected from chuuk and kosrae islands in micronesia. *J. Microbiol.* **52**: 20–26.

Jiang, Z.F., Xia, F., Johnson, K.W., Bartom, E., Tuteja, J.H., Stevens, R., et al. (2012) Genome sequences of the primary endosymbiont "Candidatus portiera aleyrodidarum" in the whitefly Bemisia tabaci B and Q biotypes. *J. Bacteriol.* **194**: 6678–6679.

Johnson, C., Li, W., Head, E., Casault, B., and Spry, J. (2014) Optical , chemical , and biological oceanographic conditions on the Scotian Shelf and in the eastern Gulf of Maine in 2013.

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427–2439.

Joye, S.B., Teske, A.P., and Kostka, J.E. (2014) Microbial Dynamics Following the Macondo Oil Well Blowout across Gulf of Mexico Environments. *Bioscience* **64**: 766–777.

Karl, D.M. and Church, M.J. (2014) Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* **12**: 699–713.

Kembel, S.W., Wu, M., Eisen, J. a., and Green, J.L. (2012) Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Comput. Biol.* **8**: e1002743.

Kirchman, D.L. (2002) The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**: 91–100.

Klindworth, A., Mann, A.J., Huang, S., Wichels, A., Quast, C., Waldmann, J., et al. (2014) Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Mar. Genomics* **18 Pt B**: 185–92.

Kopylova, E., Noé, L., and Touzet, H. (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Ladau, J., Sharpton, T.J., Finucane, M.M., Jospin, G., Kembel, S.W., O'Dwyer, J., et al. (2013) Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**: 1669–77.

Lam, P. and Kuypers, M.M.M. (2011) Microbial nitrogen cycling processes in oxygen minimum zones. *Ann. Rev. Mar. Sci.* **3**: 317–345.

Lang, A.S. and Beatty, J.T. (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol.* **15**: 54–62.

Langille, M.G.I. mlangill/microbiome_helper.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J. a, et al. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**: 814–821.

Lau, S.C.K., Zhang, R., Brodie, E.L., Piceno, Y.M., Andersen, G., and Liu, W.-T. (2013) Biogeography of bacterioplankton in the tropical seawaters of Singapore. *FEMS Microbiol. Ecol.* **84**: 259–269.

Legendre, P. and Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.

Li, J., Lin, W., Luo, J., and Tian, G. (2012) [Community structure of microbes involved in nitrification and denitrification in typical shrimp-farming water]. *Wei Sheng Wu Xue Bao* **52**: 478–88.

Li, W.K. and Dickie, P.M. (2001) Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* **44**: 236–46.

Li, W.K.W., Andersen, R. a., Gifford, D.J., Incze, L.S., Martin, J.L., Pilskaln, C.H., et al. (2011) Planktonic microbes in the Gulf of Maine area. *PLoS One* **6.**:

Li, W.K.W. and Harrison, W.G. (2008) Propagation of an atmospheric climate signal to phytoplankton in a small marine basin. *Limnol. Oceanogr.* **53**: 1734–1745.

Li, W.K.W., Harrison, W.G., and Head, E.J.H. (2006) Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. *Proc. Biol. Sci.* **273**: 1953–1960.

Lima-mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015) Determinants of community structure in the global plankton interactome. **348**: 1–10.

Lindh, M. V., Sjöstedt, J., Andersson, A.F., Baltar, F., Hugerth, L.W., Lundin, D., et al. (2015) Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ. Microbiol.* **17**: n/a–n/a.

Lovejoy, C., Bowman, J.P., and Hallegraeff, G.M. (1998) Algicidal effects of a novel marine *Pseudoalteromonas* isolate (class *Proteobacteria*, Gamma subdivision) on harmful algal bloom species of the genera *Chattonella, Gymnodinium*, and *Heterosigma*. *Appl. Environ. Microbiol.* **64**: 2806–2813.

Lücker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013) The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front. Microbiol.* **4**: 1–19.

Lynch, M.D.J. and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**: 217–229.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J. a, Green, J.L., et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**: 102–112.

Mason, O.U., Han, J., Woyke, T., and Jansson, J.K. (2014) Single-cell genomics reveals features of a Colwellia species that was dominant during the Deepwater Horizon oil spill. *Front Microbiol* **5**: 332.

Math, R.K., Jin, H.M., Kim, J.M., Hahn, Y., Park, W., Madsen, E.L., and Jeon, C.O. (2012) Comparative Genomics Reveals Adaptation by Alteromonas sp. SN2 to Marine Tidal-Flat Conditions: Cold Tolerance and Aromatic Hydrocarbon Metabolism. *PLoS One* **7**: e35784.

McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., et al. (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 16420–7.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**: 610–8.

McLean, J.S., Lombardo, M.-J., Badger, J.H., Edlund, A., Novotny, M., Yee-Greenbaum, J., et al. (2013) Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl. Acad. Sci.* **110**: E2390–E2399.

McMurdie, P.J. and Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10.**:

Mercier, C., Boyer, F., Bonin, A., and Coissac, E. (2013) SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. *Programs Abstr. SeqBio 2013 Work.* **Abstract**: 27–29.

Mitchell, M., Harrsion, G., Pauley, K., Gagne, A., Maillet, G., and Strain, P. (2002) Atlantic Zonal Monitoring Program Sampling Protocol. *Can. Tech. Rep. Hydrogr. Ocean Sci* 1–23.

Monier, A., Comte, J., Babin, M., Forest, A., Matsuoka, A., and Lovejoy, C. (2014) Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME J.* **9**: 990–1002.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–10.

Morris, R.M., Rappé, M.S., Urbach, E., Connon, S. a, and Rappe, M.S. (2004) Prevalence of the Chloroflexi -Related SAR202 Bacterioplankton Cluster throughout the Mesopelagic Zone and Deep Ocean Prevalence of the Chloroflexi -Related SAR202 Bacterioplankton Cluster throughout the Mesopelagic Zone and Deep Ocean †. **70**: 2836–2842.

Nekola, J.C. and White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**: 867–878.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F., et al. (2013) Patterns and Processes of Microbial Community Assembly. *Microbiol. Mol. Biol. Rev.* **77**: 342–356.

Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E., et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME J.* **4**: 784–98.

Nguyen, T.T. and Landfald, B. (2015) Polar front associated variation in prokaryotic community structure in Arctic shelf seafloor. *Front. Microbiol.* **6**: 1–12.

O'Hara, R.B. and Kotze, D.J. (2010) Do not log-transform count data. *Methods Ecol. Evol.* **1**: 118–122.

Orcutt, B.N., Sylvan, J.B., Knab, N.J., and Edwards, K.J. (2011) Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol. Mol. Biol. Rev.* **75**: 361–422.

Pagaling, E., Wang, H., Venables, M., Wallace, A., Grant, W.D., Cowan, D.A., et al. (2009) Microbial Biogeography of Six Salt Lakes in Inner Mongolia, China, and a Salt Lake in Argentina. *Appl. Environ. Microbiol.* **75**: 5750–5760.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003) Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**: 650–9.

Partensky, F. and Garczarek, L. (2010) *Prochlorococcus* : Advantages and Limits of Minimalism. *Ann. Rev. Mar. Sci.* **2**: 305–331.

Partensky, F., Hess, W.R., and Vaulot, D. (1999) Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**: 106–127.

Peres-Neto, P.R. and Legendre, P. (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Glob. Ecol. Biogeogr.* **19**: 174–184.

Platt, T. and Conover, R.. (1975) The ecology of St. Margaret's Bay and other inlets on the Atlantic coast of Nova Scotia. *R. Soc. Canada* 249–259.

Platt, T., Prakash, A., and Irwin, B. (1972) Phytoplankton nutrients and flushing of inlets on the coast of Nova Scotia. *Natur Can*.

Pommier, T., Canback, B., Riemann, L., Bostrom, K.H., Simu, K., Lundberg, P., et al. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* **16**: 867–880.

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014) Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS One* **9**: e93827.

Prest, E.I., Hammes, F., Kötzsch, S., van Loosdrecht, M.C.M., and Vrouwenvelder, J.S. (2013) Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res.* **47**: 7131–7142.

Punshon, S. and Moore, R.M. (2004) Nitrous oxide production and consumption in a eutrophic coastal embayment. *Mar. Chem.* **91**: 37–51.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing.

Ramette, A. (2007) Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62**: 142–160.

Rappé, M.S. and Giovannoni, S.J. (2003) The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**: 369–94.

Redmond, M. and Valentine, D. (2012) Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc. Natl. ...* **109**: 20292–20297.

Rees, G.N., Baldwin, D.S., Watson, G.O., Perryman, S., and Nielsen, D.L. (2004) Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. *Antonie Van Leeuwenhoek* **86**: 339–347.

Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of Bacterial Community Composition and Activity during a Mesocosm Diatom Bloom. *Appl. Environ. Microbiol.* **66**: 578–587.

Rivers, A.R., Sharma, S., Tringe, S.G., Martin, J., Joye, S.B., and Moran, M.A. (2013) Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *ISME J.* **7**: 2315–29.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. a, et al. (2003) Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**: 0398–0431.

Sakami, T., Watanabe, T., Kakehi, S., Taniuchi, Y., and Kuwata, A. (2015) Spatial variation of bacterial community composition at the expiry of spring phytoplankton bloom in Sendai Bay, Japan. *Gene*.

Sarmento, H. and Gasol, J.M. (2012) Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ. Microbiol.* **14**: 2348–60.

Sato-Takabe, Y., Hamasaki, K., and Suzuki, K. (2012) Photosynthetic characteristics of marine aerobic anoxygenic phototrophic bacteria Roseobacter and Erythrobacter strains. *Arch. Microbiol.* **194**: 331–41.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M. V, Tarran, G.A., and Pernthaler, J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ. Microbiol.* **11**: 2078–93.

Schauer, R., Bienhold, C., Ramette, A., and Harder, J. (2010) Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. *ISME J.* **4**: 159–170.

Schlitzer, R. (2015) Ocean Data View.

Schneiker, S., dos Santos, V.A.M., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., et al. (2006) Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis. *Nat. Biotechnol.* **24**: 997–1004.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J. a (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* **5**: e01371–14.

Shade, A.L. and Gilbert, J. a (2015) Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends Microbiol.* **23**: 335–340.

Shan, S. and Sheng, J. (2012) Examination of circulation, flushing time and dispersion in Halifax Harbour of Nova Scotia. *Water Qual. Res. J. Canada* **47**: 353–374.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–504.

Sheik, C.S., Jain, S., and Dick, G.J. (2014) Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ. Microbiol.* **16**: 304–17.

Sievert, S.M., Wieringa, E.B.A., Wirsen, C.O., and Taylor, C.D. (2007) Growth and mechanism of filamentous-sulfur formation by Candidatus Arcobacter sulfidicus in opposing oxygen-sulfide gradients. *Environ. Microbiol.* **9**: 271–6.

Simonato, F., Campanaro, S., Lauro, F.M., Vezzi, A., D'Angelo, M., Vitulo, N., et al. (2006) Piezophilic adaptation: a genomic point of view. *J. Biotechnol.* **126**: 11–25.

Sogin, M.L., Sogin, M.L., Morrison, H.G., Morrison, H.G., Huber, J. a, Huber, J. a, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **103**: 12115–20.

Song, H., Ji, R., Stock, C., Kearney, K., and Wang, Z. (2011) Interannual variability in phytoplankton blooms and plankton productivity over the Nova Scotian Shelf and in the Gulf of Maine. *Mar. Ecol. Prog. Ser.* **426**: 105–118.

Song, H., Ji, R., Stock, C., and Wang, Z. (2010) Phenology of phytoplankton blooms in the Nova Scotian Shelf-Gulf of Maine region: Remote sensing and modeling analysis. *J. Plankton Res.* **32**: 1485–1499.

Stramma, L., Johnson, G.C., Sprintall, J., and Mohrholz, V. (2008) Expanding oxygen-minimum zones in the tropical oceans. *Science* **320**: 655–658.

Strous, M., Kuenen, J.G., and Jetten, M.S.M. (1999) Key Physiology of Anaerobic Ammonium Oxidation Key Physiology of Anaerobic Ammonium Oxidation. *Appl. Environ. Microbiol.* **65**: 0–3.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function of the global ocean microbiome. **348**: 1–10.

Sunamura, M., Higashi, Y., Miyako, C., Ishibashi, J.I., and Maruyama, A. (2004) Two Bacteria Phylotypes Are Predominant in the Suiyo Seamount Hydrothermal Plume. *Appl. Environ. Microbiol.* **70**: 1190–1198.

Sverdrup, H. (1953) On conditions for the vernal blooming of phytoplankton. *J. du Cons.* **18**: 287–295.

Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.

Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O., and Reysenbach, A.-L. (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ. Microbiol.* **10**: 1681–9.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., et al. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–11.

Thrash, J.C., Cho, J.C., Vergin, K.L., Morris, R.M., and Giovannoni, S.J. (2010) Genome Sequence of Lentisphaera araneosa HTCC2155T, the Type Species of the Order Lentisphaerales in the Phylum Lentisphaerae. *J. Bacteriol.* **192**: 2938–2939.

Tonon, L., Moreira, A.P., and Thompson, F. (2014) The Prokaryotes.

Townsend, D., Thomas, A., Mayer, L., Thomas, M., and Quinlan, J. (2005) Oceanography of the Northwest Atlantic Continental Shelf (1,W). *Sea Vol. 14*.

Treusch, A.H., Vergin, K.L., Finlay, L.A., Donatz, M.G., Burton, R.M., Carlson, C.A., and Giovannoni, S.J. (2009) Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J.* **3**: 1148–63.

Tringe, S.G. and Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* **11**: 442–6.

Turner, J.W., Good, B., Cole, D., and Lipp, E.K. (2009) Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* **3**: 1082–92.

UN Atlas of the Oceans

Valliappan, K., Sun, W., and Li, Z. (2014) Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl. Microbiol. Biotechnol.* **98**: 7365–7377.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and van Sinderen, D. (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.* **71**: 495–548.

Verna, C., Ramette, A., Wiklund, H., Dahlgren, T.G., Glover, A.G., Gaill, F., and Dubilier, N. (2010) High symbiont diversity in the bone-eating worm Osedax mucofloris from shallow whale-falls in the North Atlantic. *Environ. Microbiol.* **12**: 2355–70.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P.C., Colwell, R.R., Höfle, M.G., and Pruzzo, C. (2012) Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J.* **6**: 21–30.

Vila, J., María Nieto, J., Mertens, J., Springael, D., and Grifoll, M. (2010) Microbial community structure of a heavy fuel oil-degrading marine consortium: linking microbial dynamics with polycyclic aromatic hydrocarbon utilization. *FEMS Microbiol. Ecol.* **73**: 349–62.

Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., et al. (2013) Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS One* **8**: e63422.

Wagner, M. and Horn, M. (2006) The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.* **17**: 241–9.

Walsh, D. a, Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., et al. (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.

Walsh, E.A., Kirkpatrick, J.B., Rutherford, S.D., Smith, D.C., Sogin, M., and D'Hondt, S. (2015) Bacterial diversity and community composition from seasurface to subseafloor. *ISME J.* 1–11.

Wang, K., Ye, X., Chen, H., Zhao, Q., Hu, C., He, J., et al. (2015) Bacterial biogeography in the coastal waters of northern Zhejiang, East China Sea is highly controlled by spatially structured environmental gradients. *Environ. Microbiol.* n/a–n/a.

Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012) Mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* **3**: 471–474.

Wang, Y., Zhang, R., Zheng, Q., and Jiao, N. (2014) Draft Genome Sequences of Two Marine Phototrophic Bacteria, Erythrobacter longus Strain DSM 6997 and Erythrobacter litoralis Strain DSM 8509. *Genome Announc.* **2**: e00677–14.

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., et al. (2009) Three Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These Microorganisms in Soils. *Appl. Environ. Microbiol.* **75**: 2046–2056.

Warton, D.I., Wright, S.T., and Wang, Y. (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* **3**: 89–101.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of "small-world" networks. *Nature* **393**: 440–2.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Hyperthermophilic Archaea. *Science (80-.).* **301**: 2002–2004.

Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California.

Wickham, H. and Chang, W. (2015) ggplot2 An Implementation of the Grammar of Graphics. *Compr. R Arch. Netw.*

Williams, K.P., Gillespie, J.J., Sobral, B.W.S., Nordberg, E.K., Snyder, E.E., Shallom, J.M., and Dickerman, A.W. (2010) Phylogeny of gammaproteobacteria. *J. Bacteriol.* **192**: 2305–2314.

Wind Guru (2014).

Wright, J.J., Konwar, K.M., and Hallam, S.J. (2012) Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **10**: 381–394.

Wright, J.J., Mewis, K., Hanson, N.W., Konwar, K.M., Maas, K.R., and Hallam, S.J. (2014) Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *ISME J.* **8**: 455–468.

Wu, H. -x., Lai, P.Y., Lee, O.O., Zhou, X. -j., Miao, L., Wang, H., and Qian, P.-Y. (2012) Erythrobacter pelagi sp. nov., a member of the family Erythrobacteraceae isolated from the Red Sea. *Int. J. Syst. Evol. Microbiol.* **62**: 1348–1353.

Xiaomin, X., Wang, G., and Hongbin, L. (2015) Dynamics of the bacterial and archaeal communities in the northern South China Sea revealed by 454 pyrosequencing of the 16S rRNA gene. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **117**: 97–107.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T., et al. (2015) Niches of two polysaccharide-degrading Polaribacter isolates from the North Sea during a spring diatom bloom. *ISME J.* **9**: 1410–1422.

Xu, Z., Malmer, D., Langille, M.G.I., Way, S.F., and Knight, R. (2014) Which is more important for classifying microbial communities: who's there or what they can do? *ISME J.* **8**: 2357–9.

Yeung, C.W., Law, B. a., Milligan, T.G., Lee, K., Whyte, L.G., and Greer, C.W. (2011) Analysis of bacterial diversity and metals in produced water, seawater and sediments from an offshore oil and gas production platform. *Mar. Pollut. Bull.* **62**: 2095–2105.

Yeung, C.W., Lee, K., Whyte, L.G., and Greer, C.W. (2010) Microbial community characterization of the Gully: a marine protected area. *Can. J. Microbiol.* **56**: 421–431.

Yoon, J. (2011) Phylogenetic studies on the bacterial phylum "Verrucomicrobia." *Jscc-Home.Jp* 61–65.

Yoon, J., Yasumoto-Hirose, M., Katsuta, A., Sekiguchi, H., Matsuda, S., Kasai, H., and Yokota, A. (2007) Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum "Verrucomicrobia" isolated from seawater in Japan. *Int. J. Syst. Evol. Microbiol.* **57**: 959–63.

Youssef, N.H., Couger, M.B., McCully, A.L., Criado, A.E.G., and Elshahed, M.S. (2015) Assessing the global phylum level diversity within the bacterial domain: A review. *J. Adv. Res.* **6**: 269–82.

Yurkov, V., Stackebrandt, E., Holmes, A., Fuerst, J.A., Hugenholtz, P., Golecki, J., et al. (1994) Phylogenetic positions of novel aerobic, bacteriochlorophyll a-containing bacteria and description of Roseococcus thiosulfatophilus gen. nov., sp. nov., Erythromicrobium ramosum gen. nov., sp. nov., and Erythrobacter litoralis sp. nov. *Int. J. Syst. Bacteriol.* **44**: 427–34.

Zaikova, E., Walsh, D. a., Stilwell, C.P., Mohn, W.W., Tortell, P.D., and Hallam, S.J. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ. Microbiol.* **12**: 172–191.

Zhai, L., Platt, T., Tang, C., Sathyendranath, S., and Hernández Walls, R. (2011) Phytoplankton phenology on the Scotian Shelf. *ICES J. Mar. Sci.* **68**: 781–791.

Zhang, D.-F., Zhong, J.-M., Zhang, X.-M., Jiang, Z., Zhou, E.-M., Tian, X.-P., et al. (2014) Nocardioides nanhaiensis sp. nov., an actinobacterium isolated from a marine sediment sample. *Int. J. Syst. Evol. Microbiol.* **64**: 2718–2722.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, a. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614–620.

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G.J. (2014) Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* **23**: 2260–74.

Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., and Zhi, X. (2010) Functional molecular ecological networks. *MBio* **1.**:

Zinger, L., Boetius, a., and Ramette, a. (2014) Bacterial taxa-area and distance-decay relationships in marine environments. *Mol. Ecol.* **23**: 954–964.

Appendix 1: Supplementary Information

Supplementary Table S1. Environmental parameters measured at each station sampled during the CCGS *Hudson* 2014 spring and fall cruises.

Station (Season_Station_Depth)	Latitude	Longitude	Date	Salinity (psu)	Temperature (°C)	O_2 (ml/L)	Chl (mg/m^3)	NH_3 (mmol/m^3)	NO_2 (mmol/m^3)	NO_2NO_3 (mmol/m^3)	PO_4 (mmol/m^3)	SiO_4 (mmol/m^3)	Bacteria/ mL	Location	MLD
Fall_BBL1_1	43.2498	-65.4805	2014-10-07	31.0657	15.2666	5.7103	0.550044	0.4	0	1.46	0.194	1	1660636	Coast	above
Fall_BBL1_10	43.2498	-65.4805	2014-10-07	31.0655	15.2525	5.6955	0.56576	0.51	0	0.35	0.175	1.04	1614000	Coast	above
Fall_BBL1_20	43.2498	-65.4805	2014-10-07	31.1252	14.5257	5.6956	0.400747	0.69	0.055	0.46	0.256	1.3	1774909	Coast	below
Fall_BBL1_40	43.2498	-65.4805	2014-10-07	31.8783	9.5497	6.0375	0.271093	0.81	0.15	2.4	0.519	3.15	1342600	Coast	below
Fall_BBL3_1	42.76038	-65.483	2014-10-07	32.999	16.8028	5.5793	0.491111	0.3	0.03	0	0.182	1.04	2436111	Shelf	above
Fall_BBL3_20	42.76038	-65.483	2014-10-07	33.269	16.6766	5.5032	1.108464	0.4	0.055	0.13	0.216	1.2	2070182	Shelf	above
Fall_BBL3_40	42.76038	-65.483	2014-10-07	33.8156	13.5852	5.0986	0.298595	0.46	0.19	5.37	0.546	3.45	1577933	Shelf	below
Fall_BBL3_80	42.76038	-65.483	2014-10-07	34.0281	12.4639	4.7999	0.126816	0.38	0.15	7.86	0.697	5.22	1207300	Shelf	below
Fall_BBL5_1	42.13194	-65.499	2014-10-06	34.4555	19.594	5.1976	0.3536	0.37	0	0	0.096	1.12	1923250	Shelf	above
Fall_BBL5_20	42.13194	-65.499	2014-10-06	34.446	19.5775	5.1806	0.357529	0.35	0	0	0.102	1.14	2131545	Shelf	above
Fall_BBL5_40	42.13194	-65.499	2014-10-06	34.4361	19.4854	5.1789	0.455751	0.36	0	0	0.114	1.16	1853167	Shelf	above
Fall_BBL5_80	42.13194	-65.499	2014-10-06	35.3296	14.7067	4.1052	0.115435	0.51	0	8.31	0.607	4.22	775323	Shelf	below
Fall_BBL7_1	41.86732	-65.3485	2014-10-05	34.6694	21.0192	5.1049	0.271093	0.34	0	1.13	0.097	0.96	1669615	Open	above
Fall_BBL7_20	41.86732	-65.3485	2014-10-05	34.6726	21.0304	5.0966	0.271093	0.37	0	0	0.098	1.18	1643077	Open	above
Fall_BBL7_80	41.86732	-65.3485	2014-10-05	34.6621	15.6133	4.9024	0.172871	0.32	0.145	4.46	0.409	2.96	1289294	Open	below
Fall_BBL7_250	41.86732	-65.3485	2014-10-05	35.4453	11.3972	3.1567	0	0.36	0	21.37	1.355	10.72	281840	Open	below
Fall_CSL1_1	46.95989	-60.2176	2014-09-30	28.9121	14.3423	5.8912	2.129417	0.63	0.02	0	0.198	0.56	2618800	Coast	above
Fall_CSL1_20	46.95989	-60.2176	2014-09-30	29.0422	14.0164	5.8445	1.254314	1.54	0.05	0.1	0.25	0.74	2800200	Coast	above
Fall_CSL1_40	46.95989	-60.2176	2014-09-30	31.1337	3.3707	6.2939	0.149298	3.51	0.3	6.34	1.017	10.9	1081133	Coast	below
Fall_CSL1_60	46.95989	-60.2176	2014-09-30	31.9723	1.1585	6.4419	0.099176	2.97	0.22	8.31	1.132	11.72	762080	Coast	below
Fall_CSL4_1	47.27129	-59.7816	2014-09-30	30.9807	9.3775	6.7561	1.662695	1	0	0	0.312	0.8	1587417	Shelf	above
Fall_CSL4_20	47.27129	-59.7816	2014-09-30	31.3586	5.8942	7.2802	1.429335	0.74	0.12	1.49	0.55	1.92	1237750	Shelf	above
Fall_CSL4_60	47.27129	-59.7816	2014-09-30	32.3379	0.9678	7.0504	0.032517	0.42	0.115	7.38	0.905	5.36	669567	Shelf	below
Fall_CSL4_300	47.27129	-59.7816	2014-09-30	34.7908	6.0108	3.2121	0	0.27	0	22.05	1.652	21.74	439167	Shelf	below
Fall_CSL6_1	47.58322	-59.3438	2014-10-01	31.0874	9.2433	6.6342	0.962613	0.36	0	0.2	0.325	1.13	1717273	Coast	above
Fall_CSL6_20	47.58322	-59.3438	2014-10-01	31.3717	7.1718	6.9422	0.977198	0.39	0	0.71	0.403	1.43	1382000	Coast	below
Fall_CSL6_60	47.58322	-59.3438	2014-10-01	32.2639	1.3724	7.2895	0.110557	0.98	0.14	4.65	0.76	3.64	631360	Coast	below
Fall_CSL6_200	47.58322	-59.3438	2014-10-01	34.0466	5.4967	4.4822	0	0.32	0	15.66	1.281	15.14	541119	Coast	below
Fall_GULD04_1	43.78912	-58.9003	2014-09-27	31.3435	16.5866	5.7132	0.216237	0.17	0	0	0.135	0.36	1137444	Open	above
Fall_GULD04_20	43.78912	-58.9003	2014-09-27	31.4415	16.3936	5.7248	0.255378	0	0	0	0.149	0.38	1016150	Open	above
Fall_GULD04_100	43.78912	-58.9003	2014-09-27	33.6884	6.3971	5.6464	0.026013	0.44	0	8.74	0.903	6.58	607921	Open	below
Fall_GULD04_250	43.78912	-58.9003	2014-09-27	35.2933	10.8165	3.5663	0	0.4	0	19.22	1.281	10.24	362400	Open	below
Fall_HL1_1	44.40011	-63.4499	2014-09-19	30.5347	17.2006	5.5165	0.546115	0.41	0	0	0.182	1.14	1259385	Coast	above
Fall_HL1_20	44.40011	-63.4499	2014-09-19	30.5384	17.2247	5.5098	0.498969	0.36	0	0	0.172	0.7	1285769	Coast	above
Fall_HL1_40	44.40011	-63.4499	2014-09-19	31.4847	6.9407	6.3884	0.333955	1.49	0.2	2.76	0.642	3.64	903696	Coast	below
Fall_HL1_60	44.40011	-63.4499	2014-09-19	32.259	4.7504	6.371	0.138196	0.45	0.21	4.58	0.742	4.05	614882	Coast	below
Fall_HL11_1	41.77543	-60.9054	2014-09-25	36.0212	23.6731	4.7573	0.156081	0.73	0	0	0.014	0.85	1204412	Open	above
Fall_HL11_20	41.77543	-60.9054	2014-09-25	36.0259	23.6875	4.7806	0.159332	0.27	0	0	0.025	0.72	1113588	Open	above

Station (Season_Station_Depth)	Latitude	Longitude	Date	Salinity (psu)	Temperature (°C)	O_2 (ml/L)	Chl (mg/m^3)	NH_3 ($mmol/m^3$)	NO_2 ($mmol/m^3$)	NO_2NO_3 ($mmol/m^3$)	PO_4 ($mmol/m^3$)	SiO_4 ($mmol/m^3$)	Bacteria/ mL	Location	MLD
Fall_HL11_80	41.77543	-60.9054	2014-09-25	36.3764	19.231	4.1707	0.1768	0.35	0.08	4.27	0.256	2.13	503286	Open	below
Fall_HL11_250	41.77543	-60.9054	2014-09-25	35.8161	14.0721	3.738	0	0.36	0.015	13.97	0.866	5.65	226160	Open	below
Fall_HL2_1	44.26594	-63.3187	2014-09-20	30.5707	17.6059	5.4616	0.234121	0.43	0	0	0.1	0.46	1346909	Coast	above
Fall_HL2_20	44.26594	-63.3187	2014-09-20	31.2246	16.9764	5.6272	0.408604	0.46	0	0	0.19	0.46	1166692	Coast	below
Fall_HL2_40	44.26594	-63.3187	2014-09-20	32.3743	4.8336	6.289	0.139822	0.38	0.135	4.68	0.736	4.54	788704	Coast	below
Fall_HL2_80	44.26594	-63.3187	2014-09-20	32.8258	4.4029	5.6141	0.042272	0.42	0	8.3	0.975	8.86	641152	Coast	below
Fall_HL4_1	43.48059	-62.4494	2014-09-20	31.6264	18.1601	5.3839	0.188598	0.39	0	0	0.07	0.44	1586857	Shelf	above
Fall_HL4_20	43.48059	-62.4494	2014-09-20	31.9581	18.3993	5.3764	0.24225	0.32	0.015	0	0.072	0.55	2283556	Shelf	above
Fall_HL4_40	43.48059	-62.4494	2014-09-20	32.8467	7.7145	6.7302	1.020953	0.38	0.04	0	0.375	1.75	1773455	Shelf	below
Fall_HL4_60	43.48059	-62.4494	2014-09-20	33.1817	6.1628	5.812	0.118686	0.33	0.06	5.14	0.752	4.24	840577	Shelf	below
Fall_HL6_1	42.83422	-61.7319	2014-09-21	33.9336	19.9698	5.274	0.231804	0.28	0.015	0	0.031	0.7	1809643	Open	above
Fall_HL6_20	42.83422	-61.7319	2014-09-21	34.1071	20.1636	5.2478	0.227875	0.28	0.03	0	0.041	0.62	1906462	Open	above
Fall_HL6_80	42.83422	-61.7319	2014-09-21	34.2325	10.9593	5.0292	0.055279	0.25	0.04	7.32	0.673	4.36	773393	Open	below
Fall_HL6_250	42.83422	-61.7319	2014-09-21	35.1402	8.2685	3.4569	0	0.3	0.035	24.04	1.59	14.66	239240	Open	below
Fall_HL8_1	42.36301	-61.3413	2014-09-23	35.0337	20.9704	5.1894	0.223947	0.3	0.03	0	0.032	0.64	1762769	Open	above
Fall_HL8_20	42.36301	-61.3413	2014-09-23	35.0311	20.9733	5.185	0.235733	0.3	0.015	0	0.041	0.6	1687308	Open	above
Fall_HL8_100	42.36301	-61.3413	2014-09-23	35.7791	15.1736	4.1661	0.050401	0.25	0.045	9.25	0.589	3.27	362820	Open	below
Fall_HL8_250	42.36301	-61.3413	2014-09-23	35.524	12.0326	3.4035	0	0.26	0.045	19.49	1.227	9.14	258540	Open	below
Fall_LL4_1	45.1522	-59.1735	2014-10-01	30.1823	15.8879	5.6817	0.38896	0.26	0	0	0.14	0.2	2322500	Shelf	above
Fall_LL4_20	45.1522	-59.1735	2014-10-01	30.2758	15.7491	5.6828	0.471466	0.34	0	0	0.154	0.19	2224375	Shelf	above
Fall_LL4_40	45.1522	-59.1735	2014-10-01	32.0182	3.0362	7.0997	0.193475	1.3	0.18	3.82	0.715	2.96	798583	Shelf	below
Fall_LL4_80	45.1522	-59.1735	2014-10-01	32.4774	2.2236	6.3354	0.035769	0.29	0	8.27	0.948	8.55	667567	Shelf	below
Fall_LL7_1	44.13037	-58.1822	2014-10-02	32.1513	17.0334	5.5704	0.373244	0.33	0	0.26	0.114	0.35	886091	Open	above
Fall_LL7_20	44.13037	-58.1822	2014-10-02	32.2417	17.1318	5.5746	0.385031	0.55	0	0	0.121	0.35	921650	Open	above
Fall_LL7_80	44.13037	-58.1822	2014-10-02	33.8493	8.7338	5.6268	0.151203	0.46	0.12	5.14	0.613	3.69	779731	Open	below
Fall_LL7_250	44.13037	-58.1822	2014-10-02	35.2563	9.9283	3.1211	0	0.29	0	22.92	1.464	12.63	268040	Open	below
Fall_LL9_1	43.47384	-57.5264	2014-10-03	35.5118	20.9292	5.0609	0.239662	0.34	0.065	0	0.01	0.49	1188278	Open	above
Fall_LL9_20	43.47384	-57.5264	2014-10-03	35.509	20.9375	5.0535	0.239662	0.3	0	0	0.014	0.51	1462000	Open	above
Fall_LL9_80	43.47384	-57.5264	2014-10-03	36.0247	16.9733	4.3619	0.152829	0.27	0.13	4.26	0.294	1.77	738885	Open	below
Fall_LL9_250	43.47384	-57.5264	2014-10-03	35.6457	13.3397	4.1697	0	0.3	0.03	13	0.805	5.26	328560	Open	below
Fall_STAB01_1	45.99619	-59.5334	2014-09-29	29.1231	14.1255	5.979	1.67728	0.39	0.06	0	0.195	0.64	1923375	Coast	above
Fall_STAB01_10	45.99619	-59.5334	2014-09-29	29.1311	14.1119	5.9632	1.67728	0.52	0.065	0	0.201	0.68	1978714	Coast	above
Fall_STAB01_20	45.99619	-59.5334	2014-09-29	29.3535	13.6116	5.8684	0.962613	0.84	0.08	0	0.242	0.99	1686500	Coast	above
Fall_STAB01_40	45.99619	-59.5334	2014-09-29	30.1705	11.4609	6.1095	0.495892	1.45	0.135	1.04	0.41	2.19	1450400	Coast	below
Fall_STAB05_1	46.41611	-58.8835	2014-09-30	30.4089	13.2826	6.0734	0.685497	0.32	0	0	0.178	0.26	2574286	Shelf	above
Fall_STAB05_20	46.41611	-58.8835	2014-09-30	30.8063	10.6789	6.2336	0.860518	0.43	0.045	0.26	0.264	0.62	1950556	Shelf	above
Fall_STAB05_80	46.41611	-58.8835	2014-09-30	32.6796	1.5888	6.5184	0.035769	0.68	0	9.59	0.995	7.9	599793	Shelf	below
Fall_STAB05_300	46.41611	-58.8835	2014-09-30	34.8324	5.7629	2.9943	0	0.51	0.04	22.44	1.807	32.15	611526	Shelf	below
Spring_BBL1_1	43.24955	-65.4803	2014-04-10	31.3037	1.7595	7.70302	1.717948	0.84	0.115	5.22	0.704	5.65	552250	Coast	above
Spring_BBL1_10	43.24955	-65.4803	2014-04-10	31.3023	1.7494	7.71466	1.902014	0.78	0.12	5.24	0.701	5.42	675939	Coast	above
Spring_BBL1_20	43.24955	-65.4803	2014-04-10	31.3044	1.758	7.69858	1.886676	0.78	0.12	5.42	0.704	5.42	635000	Coast	above
Spring_BBL1_40	43.24955	-65.4803	2014-04-10	31.6102	1.9476	7.46799	1.687271	0.78	0.11	5.84	0.724	5.41	355200	Coast	below
Spring_BBL2_1	42.99971	-65.4813	2014-04-10	31.4568	1.9027	7.81527	3.650527	0.74	0.105	3.88	0.613	3.38	464881	Shelf	above
Spring_BBL2_20	42.99971	-65.4813	2014-04-10	31.4597	1.9072	7.80445	3.353392	0.78	0.105	3.8	0.595	4.09	782107	Shelf	above
Spring_BBL2_40	42.99971	-65.4813	2014-04-10	31.6946	1.935	7.69526	2.928912	1.21	0.12	3.58	0.574	2.61	363900	Shelf	below
Spring_BBL2_80	42.99971	-65.4813	2014-04-10	32.3427	3.5052	6.44931	0.78228	1.01	0.135	7.96	0.858	7.1	701867	Shelf	below

Station (Season_Station_Depth)	Latitude	Longitude	Date	Salinity (psu)	Temperature (°C)	O_2 (ml/L)	Chl (mg/m^3)	NH_3 ($mmol/m^3$)	NO_2 ($mmol/m^3$)	NO_2NO_3 ($mmol/m^3$)	PO_4 ($mmol/m^3$)	SiO_4 ($mmol/m^3$)	Bacteria/ mL	Location	MLD
Spring_BBL5_1	42.13297	-65.5007	2014-04-11	33.7562	7.7956	6.52381	2.886464	1.06	0.17	4.33	0.482	1.52	821895	Shelf	above
Spring_BBL5_20	42.13297	-65.5007	2014-04-11	34.0541	8.4364	6.16164	2.377518	1.18	0.285	6.66	0.572	2.42	744975	Shelf	above
Spring_BBL5_40	42.13297	-65.5007	2014-04-11	34.1705	8.7185	5.99162	2.070741	1.59	0.215	7.42	0.61	2.8	885758	Shelf	below
Spring_BBL5_80	42.13297	-65.5007	2014-04-11	35.1335	11.2047	4.92163	0.638258	1.01	0.385	12.22	0.764	5.78	282940	Shelf	below
Spring_BBL7_1	41.86618	-65.3494	2014-04-11	33.4313	6.9141	6.72595	3.395839	0.93	0.14	5.66	0.587	2.45	797909	Open	above
Spring_BBL7_20	41.86618	-65.3494	2014-04-11	33.6725	7.0312	6.06592	2.243354	0.71	0.13	9.5	0.806	4.65	792758	Open	below
Spring_BBL7_80	41.86618	-65.3494	2014-04-11	35.2723	11.9765	5.44577	0.598214	0.68	0.36	8.72	0.599	3.08	613660	Open	below
Spring_BBL7_250	41.86618	-65.3494	2014-04-11	35.3018	10.7695	3.70819	0	0.46	0.04	19.23	1.235	10.01	183360	Open	below
Spring_CSL1_1	46.95817	-60.2164	2014-04-19	29.8531	-0.134	9.00334	9.352116	0.71	0.065	0	0.378	0.12	976900	Coast	above
Spring_CSL1_20	46.95817	-60.2164	2014-04-19	30.025	-0.3695	8.9441	10.56443	0.55	0	0.3	0.383	0.16	969700	Coast	below
Spring_CSL1_40	46.95817	-60.2164	2014-04-19	31.0419	-1.0923	8.6663	14.20136	0.68	0.055	1.74	0.569	1.94	567850	Coast	below
Spring_CSL1_60	46.95817	-60.2164	2014-04-19	31.5146	-0.5657	7.89551	8.312992	1.21	0.09	4.46	0.716	4.32	671886	Coast	below
Spring_CSL4_1	47.27035	-59.7838	2014-04-19	30.4696	-0.4311	9.1999	12.46949	0.43	0	0	0.455	0.15	862800	Shelf	above
Spring_CSL4_20	47.27035	-59.7838	2014-04-19	30.7666	-1.0314	8.77318	15.76005	0.8	0.03	0.76	0.49	0.94	916955	Shelf	below
Spring_CSL4_60	47.27035	-59.7838	2014-04-19	31.6363	-1.4653	7.88109	0.812958	0.8	0.125	6.48	0.834	6.7	553024	Shelf	below
Spring_CSL4_300	47.27035	-59.7838	2014-04-19	34.742	5.9857	3.11555	0	0.21	0	23.21	1.658	25.67	307000	Shelf	below
Spring_CSL6_1	47.57928	-59.3418	2014-04-20	31.2601	-0.6031	8.72429	8.832554	0.69	0.085	1.91	0.57	3.24	927393	Coast	above
Spring_CSL6_20	47.57928	-59.3418	2014-04-20	31.7717	-0.2715	8.13776	2.208791	0.94	0.125	5.03	0.718	5.16	741324	Coast	above
Spring_CSL6_60	47.57928	-59.3418	2014-04-20	32.2978	0.3371	7.69513	0.259862	0.88	0.16	6.44	0.768	5.96	557372	Coast	below
Spring_CSL6_200	47.57928	-59.3418	2014-04-20	34.5386	7.0486	4.21311	0	0.56	0	17.23	1.262	13.75	258160	Coast	below
Spring_GULD04_1	43.79047	-58.9008	2014-04-07	32.2813	1.7715	8.03425	3.1836	0.92	0.16	2.5	0.51	1.97	711460	Open	above
Spring_GULD04_20	43.79047	-58.9008	2014-04-07	34.8062	7.4904	4.27571	2.178113	0.72	0.025	18.42	1.228	10.27	978280	Open	above
Spring_GULD04_100	43.79047	-58.9008	2014-04-07	32.5811	2.6922	7.64272	0.935668	1.25	0.15	2.58	0.471	1.79	307580	Open	below
Spring_GULD04_250	43.79047	-58.9008	2014-04-07	35.045	11.4902	5.14703	0	1.02	0.25	8.93	0.628	3.9	235960	Open	below
Spring_HL1_1	44.40017	-63.4503	2014-04-13	31.29	0.303	8.82382	11.08399	0.52	0.03	0	0.408	0	691962	Coast	above
Spring_HL1_20	44.40017	-63.4503	2014-04-13	31.2944	0.2447	8.81393	11.77674	0.58	0.035	0	0.42	0	583226	Coast	above
Spring_HL1_40	44.40017	-63.4503	2014-04-13	31.6159	0.566	8.03	5.730479	1.24	0.085	2.94	0.622	1.78	617897	Coast	above
Spring_HL1_60	44.40017	-63.4503	2014-04-13	32.7485	4.1155	6.67795	0.705586	1.68	0.1	5.19	0.69	3.6	697548	Coast	below
Spring_HL2_1	44.26671	-63.3168	2014-04-05	31.2337	-0.5783	8.91353	9.698491	0.68	0.05	0.69	0.46	0.6	735100	Coast	above
Spring_HL2_20	44.26671	-63.3168	2014-04-05	31.269	-0.8022	8.77515	10.56443	0.58	0.065	1.02	0.498	0.92	587240	Coast	above
Spring_HL2_40	44.26671	-63.3168	2014-04-05	31.2926	-0.8244	8.43239	8.192463	0.64	0.08	2.22	0.569	1.5	610860	Coast	above
Spring_HL2_80	44.26671	-63.3168	2014-04-05	31.3972	-0.4019	8.32811	7.683087	1.4	0.09	2.6	0.588	1.78	733720	Coast	below
Spring_HL4_1	43.48023	-62.4515	2014-04-14	32.7717	5.3551	6.9751	0.319129	2	0.09	2.13	0.481	0.38	338490	Shelf	above
Spring_HL4_20	43.48023	-62.4515	2014-04-14	32.7916	5.3574	6.97143	0.355601	2.17	0.1	2.38	0.497	0.68	339714	Shelf	above
Spring_HL4_40	43.48023	-62.4515	2014-04-14	32.9768	5.6101	6.68725	0.20865	2.63	0.13	3.18	0.527	0.86	513488	Shelf	below
Spring_HL4_60	43.48023	-62.4515	2014-04-14	34.3104	9.2627	4.53295	0.127652	2.74	0.185	12.24	1.077	6.56	774212	Shelf	below
Spring_HL5.5_1	42.94015	-61.8308	2014-04-14	33.2898	6.942	6.91921	2.024725	0.97	0.16	2.94	0.458	1.21	984069	Open	above
Spring_HL5.5_20	42.94015	-61.8308	2014-04-14	33.3766	7.0203	6.82997	1.518544	0.95	0.16	2.6	0.45	0.96	1238650	Open	above
Spring_HL5.5_80	42.94015	-61.8308	2014-04-14	35.0301	11.2201	5.3398	0.524284	1.37	0.325	9.88	0.659	3.1	645111	Open	below
Spring_HL5.5_250	42.94015	-61.8308	2014-04-14	35.2932	10.1418	3.07775	0	0.67	0.03	23.47	1.508	12.21	327920	Open	below
Spring_HL8_1	42.36328	-61.3449	2014-04-15	33.5302	7.9107	7.17381	2.454212	0.69	0.02	0.1	0.263	0	2183692	Open	above
Spring_HL8_20	42.36328	-61.3449	2014-04-15	33.5337	7.859	7.13824	2.438873	0	0	0.18	0.249	0	2105929	Open	above
Spring_HL8_100	42.36328	-61.3449	2014-04-15	35.5083	13.0148	4.93963	0.259862	0	0	9.92	0.641	3.96	502060	Open	below
Spring_HL8_250	42.36328	-61.3449	2014-04-15	35.2709	10.1657	3.23363	0	0	0	18.76	1.293	9.71	240060	Open	below
Spring_LHB2_1	44.0864	-63.9026	2014-04-13	31.2328	1.2014	8.57424	8.659367	0.69	0.04	0.34	0.406	0.25	719926	Coast	above
Spring_LHB2_20	44.0864	-63.9026	2014-04-13	31.4017	0.5785	8.47668	10.56443	1.05	0.06	1.08	0.482	0.08	599000	Coast	above
Spring_LHB2_40	44.0864	-63.9026	2014-04-13	31.4357	0.2993	8.29616	10.9108	1.58	0.1	1.8	0.554	0.56	608793	Coast	above

Station (Season_Station_Depth)	Latitude	Longitude	Date	Salinity (psu)	Temperature (°C)	O_2 (ml/L)	Chl (mg/m^3)	NH_3 ($mmol/m^3$)	NO_2 ($mmol/m^3$)	NO_2NO_3 ($mmol/m^3$)	PO_4 ($mmol/m^3$)	SiO_4 ($mmol/m^3$)	Bacteria/ mL	Location	MLD
Spring_LHB2_80	44.0864	-63.9026	2014-04-13	31.9081	1.3256	7.58492	2.300824	1.67	0.09	4.08	0.686	2.98	546722	Coast	below
Spring_LHB4_1	43.37902	-63.6671	2014-04-13	32.1467	3.1324	7.62919	1.840659	1.57	0.07	1.72	0.441	0.76	911333	Shelf	above
Spring_LHB4_20	43.37902	-63.6671	2014-04-13	32.2368	3.3381	7.52721	1.656593	1.18	0.07	1.83	0.444	0.84	1030870	Shelf	above
Spring_LHB4_40	43.37902	-63.6671	2014-04-13	32.2681	3.2974	7.40083	1.625915	1.37	0.07	2.14	0.476	1.07	1030625	Shelf	below
Spring_LHB4_80	43.37902	-63.6671	2014-04-13	33.6185	6.6741	4.55288	0.997024	1.8	0.15	13.8	1.162	10.3	1011710	Shelf	below
Spring_LHB6.0_1	42.6657	-63.415	2014-04-12	34.6162	10.4308	6.41549	2.801568	0.65	0.23	3.72	0.363	0.58	1682526	Open	above
Spring_LHB6.0_20	42.6657	-63.415	2014-04-12	34.6351	10.4658	6.39112	2.730311	1.1	0.265	3.87	0.372	0.68	1695211	Open	above
Spring_LHB6.0_80	42.6657	-63.415	2014-04-12	35.2438	11.9178	5.40916	0.656494	0.95	0.41	8.59	0.594	2.78	475660	Open	below
Spring_LHB6.0_250	42.6657	-63.415	2014-04-12	35.3955	11.0641	3.09595	0	0.48	0.04	22.45	1.401	12.12	246600	Open	below
Spring_LHB6.7_1	42.19298	-63.2525	2014-04-12	35.7367	13.3807	5.37231	0.492371	0.63	0.17	9.98	0.58	3.68	507760	Open	above
Spring_LHB6.7_20	42.19298	-63.2525	2014-04-12	35.7616	13.3687	5.34843	0.460458	0.53	0.165	10.1	0.581	3.68	473020	Open	below
Spring_LHB6.7_80	42.19298	-63.2525	2014-04-12	35.7796	13.353	5.29871	0.433104	0.59	0.17	9.46	0.577	3.62	488340	Open	below
Spring_LHB6.7_250	42.19298	-63.2525	2014-04-12	35.8587	13.6265	5.12679	0	0.62	0.125	9	0.595	3.48	381220	Open	below
Spring_LL4_1	45.15797	-59.175	2014-04-18	31.6872	1.1636	8.20017	4.669279	1.26	0.15	3.8	0.67	2.34	565651	Shelf	above
Spring_LL4_20	45.15797	-59.175	2014-04-18	31.6875	1.14	8.2026	5.051311	0.86	0.135	3.76	0.664	2.36	902440	Shelf	above
Spring_LL4_40	45.15797	-59.175	2014-04-18	31.6876	1.1158	8.13963	4.881519	1.35	0.12	3.32	0.624	1.94	881320	Shelf	above
Spring_LL4_80	45.15797	-59.175	2014-04-18	32.1286	1.2516	6.95419	1.165751	0.85	0.145	8.52	0.982	9.2	715939	Shelf	below
Spring_LL7_1	44.13224	-58.1749	2014-04-18	32.445	3.3385	8.03007	5.390895	0.56	0.035	0.36	0.342	0.48	699324	Open	above
Spring_LL7_20	44.13224	-58.1749	2014-04-18	32.4458	3.3322	8.0032	5.688031	0.49	0.035	0	0.277	0.54	780571	Open	above
Spring_LL7_80	44.13224	-58.1749	2014-04-18	33.0148	4.5277	7.00902	0.644231	1.51	0.135	3.14	0.437	2.27	601943	Open	below
Spring_LL7_250	44.13224	-58.1749	2014-04-18	34.9151	7.4957	4.03484	0	1.44	0.045	21.02	1.282	12.93	224360	Open	below
Spring_LL9_1	43.46898	-57.5314	2014-04-17	32.4566	3.7852	8.18478	2.419536	0.58	0	0.42	0.322	0.42	646805	Open	above
Spring_LL9_20	43.46898	-57.5314	2014-04-17	32.4709	3.809	8.06916	2.674224	0.64	0	0.3	0.318	0.41	667125	Open	above
Spring_LL9_80	43.46898	-57.5314	2014-04-17	32.9963	4.5282	7.06902	0.442222	1.26	0.09	2.47	0.47	1.33	528209	Open	below
Spring_LL9_250	43.46898	-57.5314	2014-04-17	34.8727	6.6692	4.3207	0	1.14	0.055	18.94	1.277	12.04	218600	Open	below
Spring_STAB01_1	45.99865	-59.5289	2014-04-20	30.6089	-0.1209	9.20533	14.02817	0.85	0.055	0.1	0.398	0	837370	Coast	above
Spring_STAB01_10	45.99865	-59.5289	2014-04-20	30.6478	-0.2069	9.23902	14.89411	0.84	0.05	0.12	0.416	0	1051889	Coast	above
Spring_STAB01_20	45.99865	-59.5289	2014-04-20	30.7836	-0.3912	9.29269	17.49192	0.7	0	0.2	0.431	0	817875	Coast	below
Spring_STAB01_40	45.99865	-59.5289	2014-04-20	31.1482	-0.94	8.30507	6.536991	1.2	0.085	3.52	0.694	3.58	677625	Coast	below
Spring_STAB05_1	46.41959	-58.8754	2014-04-20	31.1768	-0.5074	9.09734	13.50861	0.57	0	0.52	0.473	1	1103300	Shelf	above
Spring_STAB05_20	46.41959	-58.8754	2014-04-20	31.1833	-0.5049	9.05928	14.89411	0.56	0	0.8	0.496	1.02	898750	Shelf	above
Spring_STAB05_80	46.41959	-58.8754	2014-04-20	31.8466	-0.632	8.08623	0.177801	0.84	0.145	5.65	0.758	5.35	652381	Shelf	below
Spring_STAB05_300	46.41959	-58.8754	2014-04-20	34.7512	6.0958	3.32954	0	0.71	0	23	1.662	25.12	376600	Shelf	below
Spring_TB01_1	43.81491	-60.2549	2014-04-06	32.149	2.5464	7.5183	2.388633						593380	Shelf	above
Spring_TB01_10	43.81491	-60.2549	2014-04-06	32.148	2.5426	7.48093	2.149002						775560	Shelf	above
Spring_TB01_20	43.81491	-60.2549	2014-04-06	32.1488	2.5428	7.48945	2.061444						500560	Shelf	above
Spring_TB01_40	43.81491	-60.2549	2014-04-06	32.3657	2.9587	7.30062	1.698925						648420	Shelf	above

Supplementary Table S2. List of dual-indexing Illumina fusion primers used for Multiplexing samples for Illumina Sequencing

Left arm P5 adapter	i5 name	i5	Right arm P5 adapter	Forward Primer	Final name	Fusion sequence (5'-3')
AATGATACGGCGACC ACCGAGATCTACAC	S502	CTCTCT AT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S503	TATCCT CT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S505	GTAAG GAG	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S506	ACTGCA TA	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S507	AAGGA GTA	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S508	CTAAGC CT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S510	CGTCTA AT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S511	TCTCTC CG	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S513	TCGACT AG	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S513	AATGATACGGCGACCACCGAGATCTACACTCGACTAGTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S515	TTCTAG CT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S515	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S516	CCTAGA GT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S516	AATGATACGGCGACCACCGAGATCTACACCCTAGAGTTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S517	GCGTA AGA	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S517	AATGATACGGCGACCACCGAGATCTACACGCGTAAGATCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S518	CTATTA AG	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S518	AATGATACGGCGACCACCGAGATCTACACCTATTAAGTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S520	AAGGC TAT	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S520	AATGATACGGCGACCACCGAGATCTACACAAGGCTATTCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S521	GAGCCT TA	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S521	AATGATACGGCGACCACCGAGATCTACACGAGCCTTATCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC
AATGATACGGCGACC ACCGAGATCTACAC	S522	TTATGC GA	TCGTCGGCAGCGTCAGATGT GTATAAGAGACAG	ACGCGHNRAA CCTTACC	B969F-S522	AATGATACGGCGACCACCGAGATCTACACTTATGCGATCGTCGGCAGCGTCAG ATGTGTATAAGAGACAGACGCGHNRAACCTTACC

Left arm P7 adapter	i7 name	i7	Right arm P7 adapter	Reverse Primer	Final name	Fusion sequence (5'-3')
CAAGCAGAAGACGGC ATACGAGAT	N701	TCGCCT TA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R-N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N702	CTAGTA CG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R-N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N703	TTCTGC CT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R-N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N704	GCTCAG GA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R-N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N705	AGGAG TCC	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R-N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA

Left arm P7 adapter	i7 name	i7	Right arm P7 adapter	Reverse Primer	Final name	Fusion sequence (5'-3')
CAAGCAGAAGACGGC ATACGAGAT	N706	CATGCC TA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N707	GTAGA GAG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N710	CAGCCT CG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N711	TGCCTC TT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N712	TCCTCT AC	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N714	TCATGA GC	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N714	CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N715	CCTGAG AT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N715	CAAGCAGAAGACGGCATACGAGATCCTGAGATGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N716	TAGCG AGT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N716	CAAGCAGAAGACGGCATACGAGATTAGCGAGTGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N718	GTAGCT CC	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N718	CAAGCAGAAGACGGCATACGAGATGTAGCTCCGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N719	TACTAC GC	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N719	CAAGCAGAAGACGGCATACGAGATTACTACGCGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N720	AGGCTC CG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N720	CAAGCAGAAGACGGCATACGAGATAGGCTCCGGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N721	GCAGC GTA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N721	CAAGCAGAAGACGGCATACGAGATGCAGCGTAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N722	CTGCGC AT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N722	CAAGCAGAAGACGGCATACGAGATCTGCGCATGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N723	GAGCG CTA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N723	CAAGCAGAAGACGGCATACGAGATGAGCGCTAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N724	CGCTCA GT	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N724	CAAGCAGAAGACGGCATACGAGATCGCTCAGTGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N726	GTCTTA GG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N726	CAAGCAGAAGACGGCATACGAGATGTCTTAGGGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N727	ACTGAT CG	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N727	CAAGCAGAAGACGGCATACGAGATACTGATCGGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N728	TAGCTG CA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N728	CAAGCAGAAGACGGCATACGAGATTAGCTGCAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA
CAAGCAGAAGACGGC ATACGAGAT	N729	GACGTC GA	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG	ACGGGCRGTG WGTRCAA	BA1406R -N729	CAAGCAGAAGACGGCATACGAGATGACGTCGAGTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGACGGGCRGTGWGTRCAA

Columns in red were from Nextera XT v2

Primer sequences from Comeau et al. (2011)

Supplementary Table S3. Data from each step of quality control process after sequencing for Scotian Shelf and Bedford Basin datasets

Raw Data	Scotian Shelf	Bedford Basin[a]
Reads	16,917,600	14,304,947
Reads Passing Filter (PF)	15,394,747	12,565,285
Post Chimera Checking	10,536,766	8,443,943
OTU analysis		
Post OTU picking		
Number of Sequences	10,148,091	7,978,549
Number of OTUs	28,036	23,515
Number of New Reference OTUs	21,507	16,105
Removal of Archaea and Chloroplasts		
Number of Sequences	9,953,481	7,744,556
Number of OTUs	27,428	23,056
Post Rarefaction		
Number of Sequences	1,260,000	2,415,600
Number of OTUs	16,410	19,340

[a]Bedford Basin data includes information from all four depths sampled (1 m , 5 m , 10 m, and 60 m)

Supplementary Table S4. Full list of taxa used for Scotian Shelf analysis

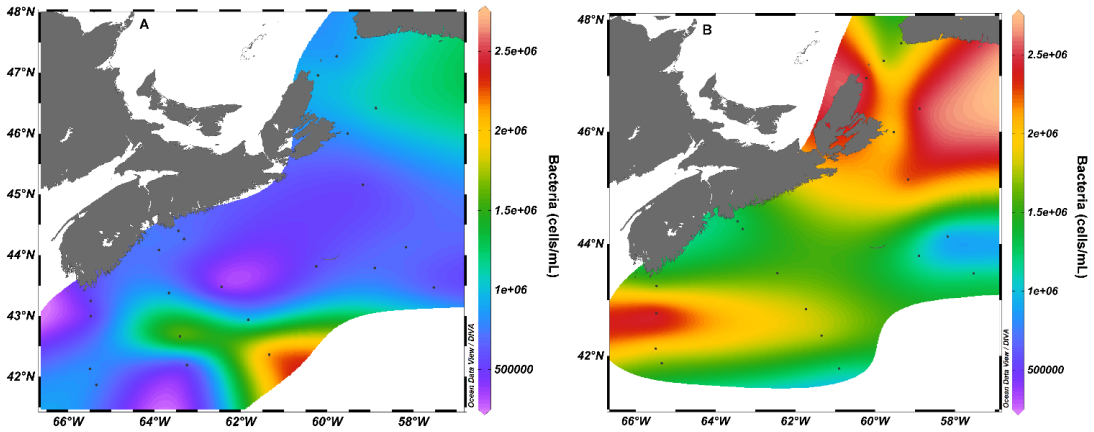
Taxa	Level of Taxonomy	Parent Taxa
Acidobacteria	Phylum	Acidobacteria
Actinobacteria	Phylum	Actinobacteria
Nocardioides	Genus	Actinobacteria
Erythrobacteraceae	Family	Alphaproteobacteria
Methylobacterium	Genus	Alphaproteobacteria
Pelagibacteraceae	Family	Alphaproteobacteria
Rhodobacteraceae	Family	Alphaproteobacteria
Octadecabacter	Genus	Alphaproteobacteria/ Rhodobacteraceae
Alphaproteobacteria	Class	Alphaproteobacteria/Proteobacteria
AncK6	Phylum	AncK6
Bacteroidetes	Phylum	Bacteroidetes
Polaribacter	Genus	Bacteroidetes
Ulvibacter	Genus	Bacteroidetes
Betaproteobacteria	Class	Betaproteobacteria/Proteobacteria
Chlamydiae	Phylum	Chlamydiae
Chlorobi	Phylum	Chlorobi
Chloroflexi	Phylum	Chloroflexi
Cyanobacteria	Phylum	Cyanobacteria
Prochlorococcus	Genus	Cyanobacteria
Synechococcus	Genus	Cyanobacteria
SAR324	Family	Deltaproteobacteria
Deltaproteobacteria	Class	Deltaproteobacteria/Proteobacteria
Epsilonproteobacteria	Class	Epsilonproteobacteria/Proteobacteria
Fibrobacteres	Phylum	Fibrobacteres
Firmicutes	Phylum	Firmicutes
Fusobacteria	Phylum	Fusobacteria
Alteromonadales	Order	Gammaproteobacteria
Oceanospirillales	Order	Gammaproteobacteria
Vibrionaceae	Family	Gammaproteobacteria
Alteromonas	Genus	Gammaproteobacteria/ Alteromonadales
Pseudoalteromonas	Genus	Gammaproteobacteria/ Alteromonadales
Alcanivorax	Genus	Gammaproteobacteria/ Oceanospirillales
SAR86	Family	Gammaproteobacteria/ Oceanospirillales/
Gammaproteobacteria	Class	Gammaproteobacteria/Proteobacteria
Gemmatimonadetes	Phylum	Gemmatimonadetes
Lentisphaerae	Phylum	Lentisphaerae
Unknown	NA	NA
Nitrospirae	Phylum	Nitrospirae
NKB19	Phylum	NKB19
OD1	Phylum	OD1
OP3	Phylum	OP3
PAUC34f	Phylum	PAUC34f
Planctomycetes	Phylum	Planctomycetes
Proteobacteria	Phylum	Proteobacteria
SBR1093	Phylum	SBR1093
Spirochaetes	Phylum	Spirochaetes
Tenericutes	Phylum	Tenericutes
Thermi	Phylum	Thermi
TM6	Phylum	TM6
TM7	Phylum	TM7
Verrucomicrobia	Phylum	Verrucomicrobia
WPS-2	Phylum	WPS-2
WS3	Phylum	WS3
ZB3	Phylum	ZB3

Supplementary Table S5. Results of mvabund Analysis of Deviance Table using the model: community ~ Season + MLD + Location
With a negative binomial distribution and 999 permutations of the data to test for significance

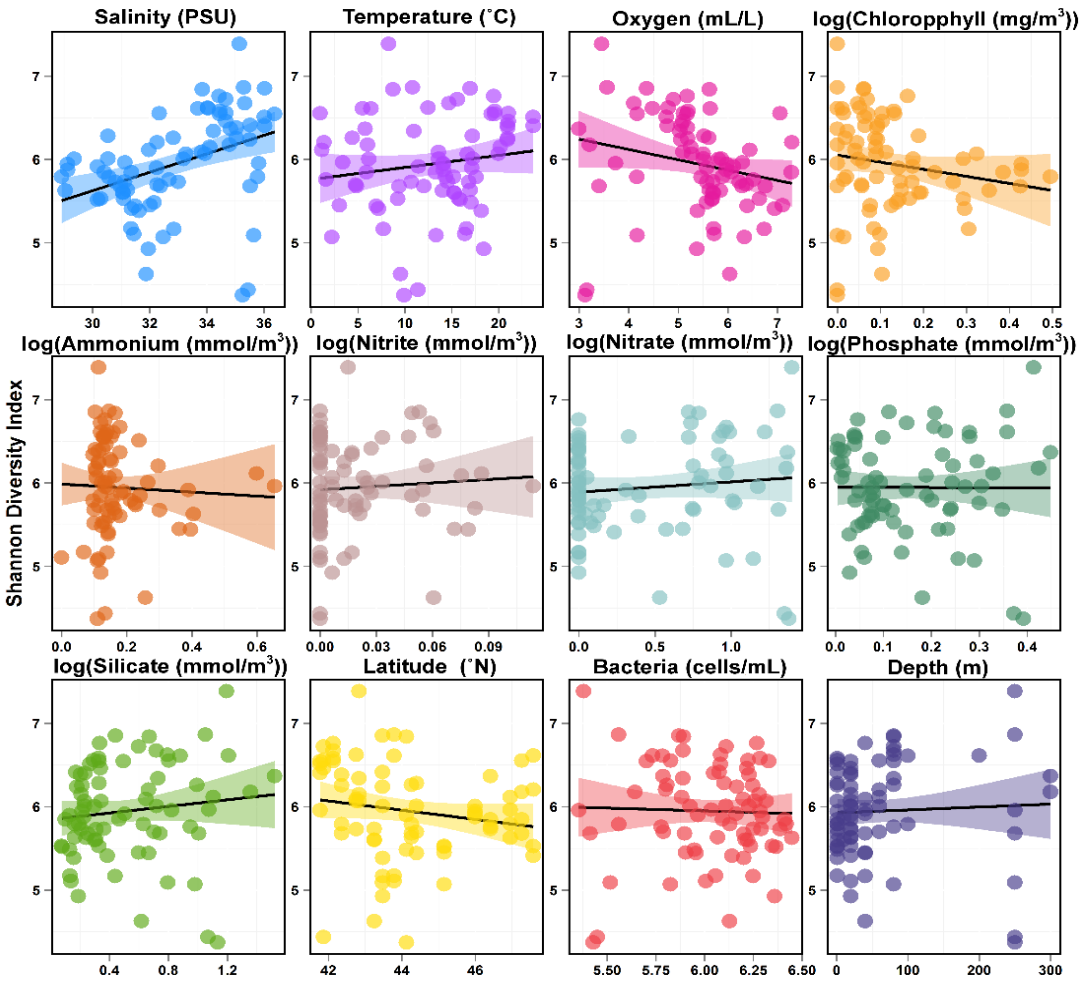
Variable	Deviance Score	*P* value
Season	1800.6	0.001
MLD	1346.7	0.001
Location	840.5	0.001

Supplementary Table S6. Results of mvabund Analysis of Deviance Table on stations on the shelf or near the Thebaud oil and gas platform.
The model: community ~ StationType
Where StationType refers to whether the site was near the Thebaud platform or not, was fit using a negative binomial distribution, with 999 permutations of the data to test for significance

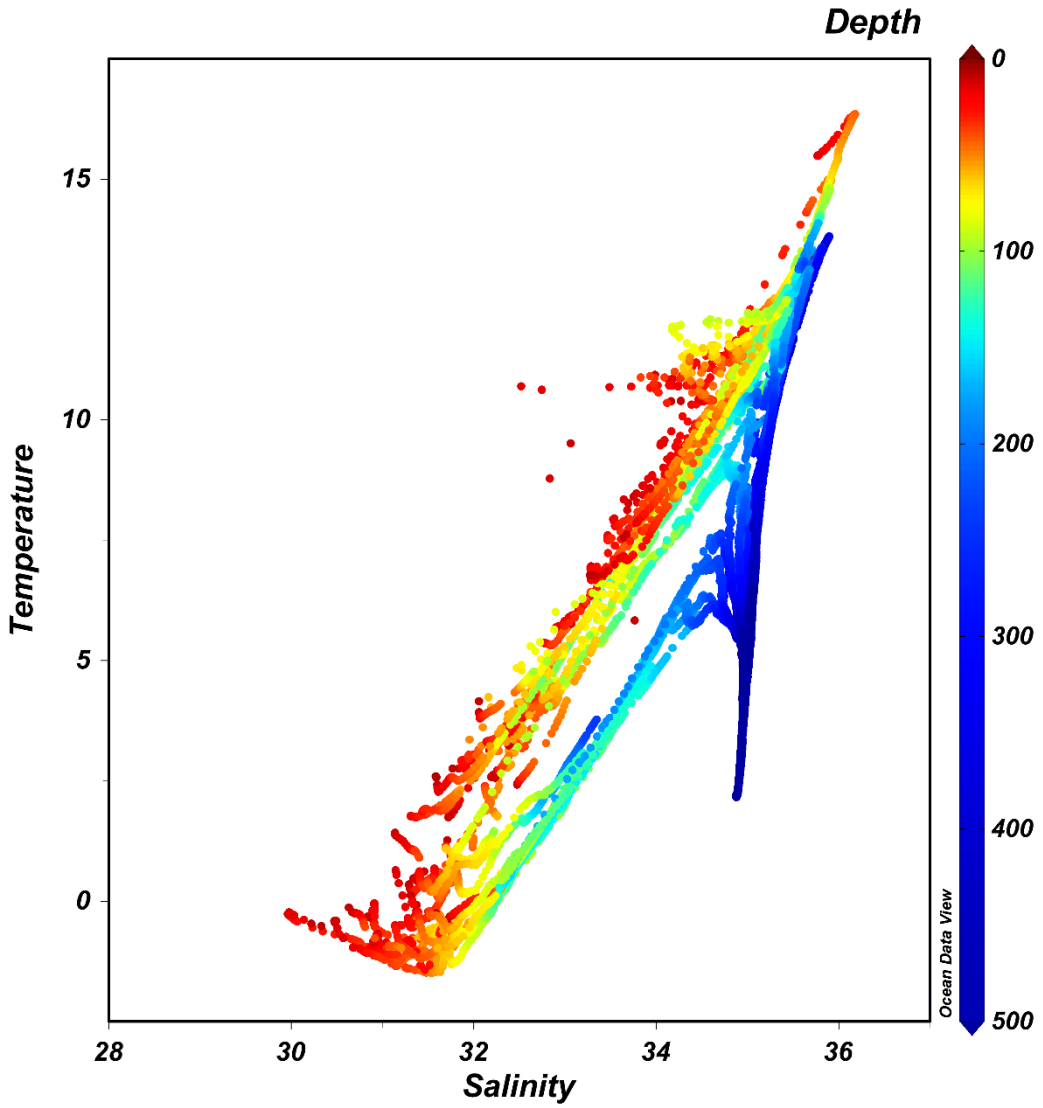
Variable	Deviance Score	*P* value
StationType	1443.6	0.001



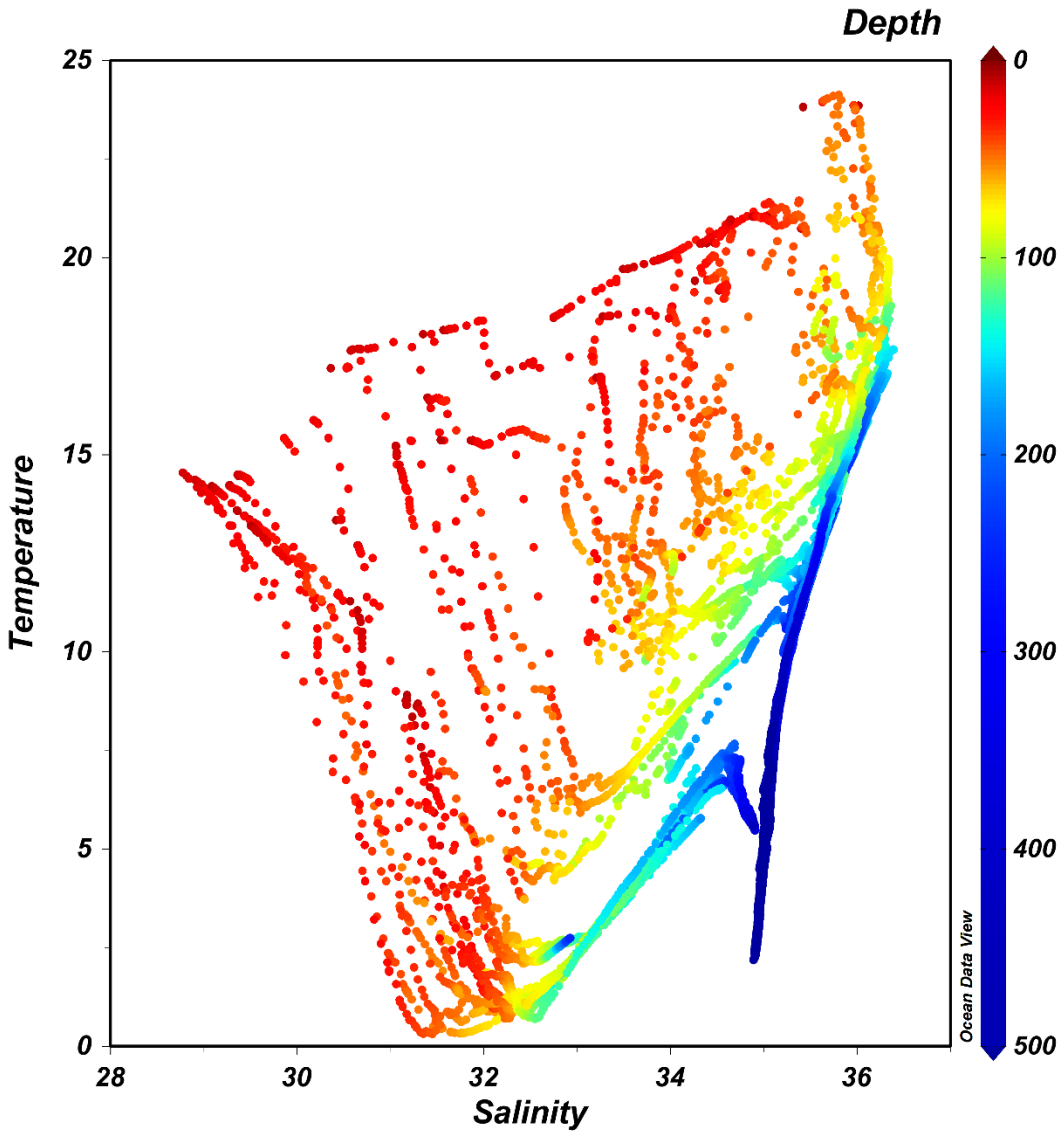
Supplementary Figure S1. Spring (A), and fall (B) bacterial concentrations at 1 m depth on the Scotian Shelf. Bacterial concentration are shown in cells/mL



Supplementary Figure S2. Linear regressions of the Shannon alpha diversity metric against various environmental variables using fall samples. The shaded area of each plot represents the 95% confidence interval for the regression. Chlorophyll, ammonia, nitrite, nitrate, phosphate, silicate, and bacteria were all transformed by log+1 to establish a more normal distribution of data. Table 2.2 reports significance, adjusted R^2 and slope for each regression.



Supplementary Figure S3. Temperature Salinity (TS) diagram of the top 500 m from sites sampled during the spring cruise



Supplementary Figure S4. Temperature Salinity (TS) diagram of the top 500 m from sites sampled during the fall cruise