Temporal bacterial community dynamics in the coastal northwest Atlantic Ocean

Heba El-Swais

A Thesis
in
the Department
of
Biology

Presented in Partial Fulfillment of the Requirements
of the Degree of Master of Science (Biology) at
Concordia University
Montreal, Quebec, Canada

August 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Heba El-Swais

Entitled: **Temporal bacterial community dynamics in the coastal northwest Atlantic Ocean**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____Jim Grant_____Chair

_____Dylan Fraser_____Examiner

_____Robert Weladji_____Examiner

_____David Walsh_____Supervisor

_____Dan McLaughlin____External Examiner

Approved by _____

Chair of Department or Graduate Program Director

Date: _____ _____

2013 Dean of Faculty

Abstract

Temporal bacterial community dynamics in the coastal northwest Atlantic Ocean

Heba El-Swais

Marine microbial communities are essential drivers of globally important processes, including carbon, nitrogen, iron and sulfur cycles. Because these diverse bacteria play crucial roles, understanding the structure of microbial communities across space and time is important. This will allow better comprehension of the earth's response to global change. In this study, we used 16S rRNA sequencing to investigate the temporal shifts in bacterial community structure in surface communities in a coastal inlet in the Northwest Atlantic Ocean (Bedford Basin). We demonstrated that in the spring the dominant bacterial groups were the Bacteroidetes, particularly *Polaribacter* and *Cytophaga*. The nutrient-poor summer is associated with the alpha-proteobacterial oligotrophic lineage SAR11. Alteromonadales and SAR11 characterize autumn, and winter is strongly associated with Gamma-proteobacterial sulfur oxidizers (GSO) group. We have established that bacterial richness is highest in winter, which is consistent with global latitudinal studies of bacteria. We further demonstrated that similarity between surface bacterial communities varied according to different time-scales. We showed that high-frequency sampling can reveal episodic blooms from the rare bacterial biosphere. In summary, we used a multi-scale study to reveal the temporal dynamics of bacterial community structure in the temperate North Atlantic coastal ocean.

Acknowledgments

We would like to thank William K. Li and Susan E. Craig from the Bedford Institute of Oceanography for access to the archival samples and providing us with all the contextual data used in this master's thesis. We would also like to thank Etienne Yergeau and the NRC Biotechnology Research Institute for sequencing the seasonal dataset.

I would like to thank my committee members for their input and help with this project. And finally, I would like to greatly thank my advisor, Dr. David Walsh, for replying to a girl from the other side of the world and for his extensive mentorship and support, making all this possible.

Table of Contents

List of Figures v
List of Tables vii

1.0 Introduction v
1.1 Marine microbial diversity 3
1.2 Metabolic diversity of marine bacteria 6
1.3 Marine microbial biogeography 7
1.4 Microbial plankton in the Northwest Atlantic Ocean 12
1.5Annual cycle in Bedford Basin 13
1.6 Study Objectives 14
2.0 Methods 15
2.1 Site description and sampling 15
2.2 Molecular methods 16
2.3 16S rRNA gene sequence processing 17
2.4 Statistical and taxonomic analysis 18
3.0 Results 19
3.1 Summary of the 16S rRNA gene sequence datasets 19
3.1.1 Multi-year time-series (2005-2010) 19
3.1.2 Biweekly time-series (2009) 20
3.2 Physico-chemical environment 20
3.3 Phytoplankton patterns during 2005-2010 20
3.4 Taxonomic composition of Bedford Basin bacteria 22
3.5 Seasonal variation in bacterial diversity 24
3.6 Seasonal variation in bacterial community structure 26
3.6.1 Linking community structure to environmental change 28
3.6.2 Stability and succession over a single annual cycle 29
4.0 Discussion 32
4.1 Placing bacterial community structure into Bedford Basin monitoring program 32
4.2 The question of temporal scales in bacterial community dynamics 39
5.0 Conclusions, implications and future directions 41
6.0 Bibliography 42
Figures 52
Tables 68
7.0 Supplementary information 78
7.1 Appendix I. Caveats of the study 78
7.2 Appendix II. Different phylogenetic resolution reveals different processes 79
7.3 Appendix III. Community Beta-diversity using other dissimilarity indexes: 80
7.3.1 Multi-year Betadiversity 80

List of Figures

Figure 1. Map of Bedford Basin

Figure 2. Annual patterns of microbial community change in the surface layer (0-5m) at Bedford Basin
Figure 3. Environmental data from 2005 through 2011
Figure 4. Weekly phytoplankton trends over 2009
Figure 5. Rank-abundance curve**.**
Figure 6. Seasonal average taxonomic composition associated with each season
Figure 7. Seasonal trends of select main groups of bacteria**.**
Figure 8. Seasonal α-diversity (Chao1 index) box plots per season.
Figure 9. Seasonal Community Evenness (Shannon-evenness index) box plots per season.
Figure 10. Community Richness and Evenness over the single (2009) year**.**
Figure 11. Time-decay box plots of Seasonal β-diversity (Bray-Curtis 90%).
Figure 12. NMS ordination based on Bray-Curtis similarities calculated using relative abundance of OTUs clustered at 90% sequences similarity
Figure 13. Time-decay box plots of different time scales within one year based on similarity values (Bray-Curtis 90%).
Figure 14 Similarity values between each two weeks that occurred in 2009.
Figure 15. Taxonomic and environmental trends that occurred week 17-20 and 30-32.
Figure 16. NMS ordination on 2009 biweekly samples based on Bray-Curtis similarities calculated using relative abundance of OTUs clustered at 90% sequence similarity

List of Tables

Table 1: OTUs assigned to each sample in the seasonal (2005-2010)
Table 2: OTUs assigned to each sample in the biweekly (2009)
Table 3: Chemical environmental measurements for seasonal dataset (2005-2010)
Table 4: Chemical environmental measurements for biweekly dataset (2009)
Table 5: Biotic Environmental measurements for seasonal dataset (2005-2010)
Table 6: Biotic Environmental measurements for biweekly dataset (2009)
Table 7: Analysis of similarity between seasons and p value.

1.0 Introduction

The ocean covers 70% of the earth's surface and plays a central role in moderating the earth's climate. The oceans serve as a sink for carbon dioxide (CO_2) and heat from the atmosphere. Therefore, as a result of increasing atmospheric temperature and CO_2 concentration, negative impacts on the ocean ecosystem are expected to occur, and indeed have been observed, such as the rise in sea surface temperature (SST), ocean acidification, and oxygen depletion (Orr et al, 2005, Doney et al, 2012, Doney et al, 2009, Bindoff et al, 2007). With the 0.04°C increase in average SST that has occurred in the past 50 years, secondary changes have been observed such as increased stratification by warming of the upper layer of oceans and by reducing mixing events, hence affecting nutrient availability (Doney et al, 2012, Levitus et al, 2009)

The oceans contain a wide array of microbial plankton including eukaryotes, bacteria, archaea and their associated viruses (Giovannoni and Stingl, 2005, Fuhrman, 2009). In one milliliter of water, there exists on average 10^5-10^6 prokaryotic cells, leading to an estimated 10^{29} prokaryotes in the global ocean (Fuhrman, McCallum, and Davis, 1992, Kirchman, 2012, DeLong, 1992, Kirchman and Mitchel, 2008, Massana et al, 1997, Whitman, Coleman, and Wiebe, 1998). These highly abundant and diverse marine microbial communities are essential drivers of important biogeochemical processes, such as the regulation of global carbon and nitrogen cycles (Giovannoni et al, 1990, Fuhrman, 2009, Falkowski et al, 2008). Understanding the structure of microbial communities and their distribution across space and time is important because this will allow better

comprehension of the ocean's response to global climate change and its effect on global biogeochemical cycles.

Because nearly half of global primary production is carried out by phytoplankton in the oceans, the most direct biological response to global warming has been observed in phytoplankton (Li and Harrison, 2008, Azam and Malfatti, 2007, Kirchman, 2012) For example, In the Arctic ocean, a shift from larger phytoplankton (2-20 um in diameter) to smaller picoplankton (<2 um) has been documented (Li et al, 2009). In the northwest and northeast Atlantic oceans, phytoplankton abundances have increased according to the continuous plankton recorder survey from 1960-2006 (Head and Pepin, 2010, Edwards, Peird, and Planque, 2001). Furthermore, a study of multi-year trends of microbial plankton in the northwest Atlantic Scotian Shelf revealed increased abundances in the spring and decreased abundances in the autumn (Li, Harrison, and Head, 2006). Moreover, changes in phytoplankton communities have been shown to propagate through to changes to the marine bacterial community as bacterial abundances closely follow the phytoplankton trends observed (Li and Harrison, 2008, Li, Harrison, and Head, 2006b).

In the classic food web, phytoplankton are consumed by zooplankton, and zooplankton are eaten by fish and higher trophic levels. Dissolved organic matter (DOM) is produced from each of these trophic levels through lysis of phytoplankton or sloppy feeding of zooplankton. Contrasting to the classic food chain there is a bacterial component known as the microbial loop in which bacteria incorporate DOM from various sources and are then consumed by predators (Azam et al. 1983, Pomeroy et al, 2007). In

this process DOM, which would otherwise be lost without the microbial loop, is recycled back into higher trophic levels (Pomeroy et al, 2007, Azam et al, 1983). In addition, marine bacteria regenerate essential nutrients (*e.g.* nitrogen and phosphorous) that would otherwise be lost in the ecosystem. This regeneration of nutrients allows phytoplankton to further generate organic matter. DOM that is not recycled will sink into the deep ocean interior and be stored. This sedimentation of aggregate organic particles from the euphotic zone to the benthic zone contributes to the large sink of organic matter in the ocean (Doney et al, 2012, Jiao et al, 2010, Levitus et al, 2009). Given the link between phytoplankton and heterotrophic bacteria and the role they play in carbon cycling, monitoring microbial community structure and distribution across time and space is important to understand long-term response to global warming.

1.1 Marine microbial diversity

Much of our knowledge of bacterial community structure and function has been driven by the advent of cultivation-independent molecular approaches. Since Carl Woese first demonstrated the use of ribosomal RNA (rRNA) as a phylogenetic marker to reconstruct the universal tree of life (Woese, 1987), rRNA gene sequence analysis, specifically the small subunit rRNA gene (16S rRNA), has been used for over 30 years to assess bacterial community structure and diversity in many environments, aquatic environments included (Kirchman, 2012, Giovannoni and Stingl, 2005, Kirchman and Mitchel, 2008, Whitman, Coleman, and Wiebe, 1998). The earliest cultivation-independent studies of marine bacterial diversity profoundly altered our perception of these communities. For example, the presence of archaea in the ocean changed the central

paradigm that this group were strictly extremophiles (Fuhrman, 2009, Fuhrman, McCallum, and Davis, 1992, Falkowski et al, 2008, DeLong, 1992, Massana et al, 1997). Cultivation-independent methods also lead to the discovery of the most abundant bacteria in the ocean, the SAR11 bacteria. SAR11 was first discovered in the Sargasso sea in 1990 by extracting total community DNA from a water sample, PCR amplifying, cloning and sequencing 16S rRNA gene sequences (Kirchman, 2012, Giovannoni et al, 1990, Azam and Malfatti, 2007, Giovannoni and Stingl, 2005). That was followed by fluorescence in situ hybridization (FISH) showing the abundance of SAR11 in surface bacterial communities (Morris et al, 2002). A SAR11 representative was first cultured in 2002 and given the name *Candidatus* "Pelagibacter ubique" (Rappé et al, 2002). Based on its abundance and wide distribution, SAR11 is considered one of the most successful marine lineages (Morris et al, 2002). However, the discovery of SAR11 was the tip of the iceberg as it was soon realized that the majority of bacterial lineages had not yet been cultivated (Rappé and Giovannoni, 2003).

In addition to the discovery of novel prokaryotic lineages in the ocean, 16S rRNA analyses also revealed an astounding level of bacterial and archaeal diversity. Estimates of bacterial diversity in the ocean are continually increasing as our technology for analyzing diversity advances. Early fingerprinting methods such as Denaturing Gradient Gel Electrophoresis (DGGE) greatly underestimated bacterial richness as they sampled only around 10-50 of the most abundant taxa (Pedrós-Alió, 2012, Casamayor et al, 2000). Afterwards, traditional clone libraries and Sanger sequencing showed that there were perhaps hundreds of taxa (Acinas et al, 2004, Pommier et al, 2007). Then with the development of next generation sequencing technologies, it became evident that that there

was 1-2 orders of magnitude of diversity unaccounted for in previous studies (Sogin et al, 2006). The development of faster and more affordable technologies, such as the 454 Roche, Illumina, and Ion Torrent platforms, has led to a bloom in marine bacterial community studies (Andersson, Reimann, and Bertilsson, 2009, Caporaso et al, 2011b, Yergeau et al, 2012).

In 1961, G.E. Hutchinson described the high diversity of phytoplankton in the oceans as a paradox because the rules of competitive exclusion do not seem to apply. He named this phenomenon the "paradox of the plankton"(Hutchinson, 1961). Cultivation-independent methods showed that this paradox seemingly applied even more strongly to bacterioplankton. It also became evident that bacterial communities were unevenly distributed and that bacterial richness was dominated by many rare bacterial taxa, coined by Sogin et al (2006) as the "rare biosphere". Rank-abundance curves showed that bacterial richness was composed of a few abundant taxa with a long tail of rare taxa found at usually less than 1% of sequences (Pedrós-Alió, 2006). Furthermore, it became apparent that the number of taxa identified depended on sampling depth (number of sequences), as collector's curve (or rarefaction curve) showed that the as the number of individual sequences increases, the more taxa were found (Pedrós-Alió, 2006, Pedrós-Alió, 2012). Recently, considerable thought has gone into the nature of the rare biosphere (Pedrós-Alió, 2006, Pedrós-Alió, 2012). The rare bacteria may be active and controlled by top-down predation by viruses and grazers, which keep cell abundance low, or they may serve as dormant seed banks which carry out specific functions and only grow to high levels under optimal conditions (Caporaso et al, 2011b, Sjostedt et al, 2012, Lennon

and Jones, 2011). Moreover, the rare biosphere further supported the idea of "paradox of the plankton" (Hutchinson, 1961).

1.2 Metabolic diversity of marine bacteria

The maintenance of high phylogenetic diversity suggests functional and ecological differences among marine taxa, and this is indeed the case. Historically, marine bacteria in the photic zone were mostly thought of as either photoautotrophs that harvest light energy to fix CO_2 to organic matter, or organoheterotrophs that use organic compounds as both energy and carbon source (Kirchman and Mitchel, 2008). However, the metabolic diversity of marine bacteria is much more complex than that. Bacteria can also be mixotrophic (Eiler, 2006). For example, the Roseobacter clade of Alpha-proteobacteria are considered generalists as they are able to break down a wide variety of organic compounds (Brinkhoff, Giebel, and Simon, 2008, Kirchman, 2012) and yet are also found to be capable of phototrophy (Newton et al, 2010). The discovery of the proteorhodopsin gene, encoding a light driven proton pump, in several bacterial groups suggested phototrophy was widespread among "heterotrophic" bacteria. (Béjà et al, 2000). Proteorhodopsin was first discovered in the SAR86 clade, an abundant uncultivated lineage of Gamma-proteobacteria, but has since been found in many marine lineages (Béjà et al, 2000, Dupont et al, 2011, Sabehi et al, 2004) Another group of photoheterotrophic bacteria are the aerobic anoxygenic phototrophic (AAP) bacteria, which require organic carbon, but can also generate energy phototrophically (Cottrell, Ras, and Kirchman, 2010). Marine bacteria are also capable of chemolithotrophy in which

energy is obtained by oxidizing inorganic chemical compounds such as NH_3 (Stahl & la Torre, 2012), H_2S (Zaikova et al. 2010) and H_2 (Anantharaman et al. 2013). In addition, bacteria can switch their metabolic lifestyles depending on the environmental conditions (Eiler, 2006).

1.3 Marine microbial biogeography

The study of microbial biogeography is currently of immense interest in the general field of microbial ecology. Understanding microbial biogeography has been more difficult than that of macro-organisms given the complex diversity of microbes (Prosser et al, 2007). However, certain classic concepts formulated from ecology of macro-organisms seem to hold for the microbial world. For example, a decline in community similarity as the distance increases, better known as the distance-decay relationship, indicates not only that composition is different across locations but also that variation is spatially correlated (Martiny et al, 2006).

Aquatic microbial diversity is, in part, a result of the heterogeneous environmental conditions and nutrients found in the global ocean. For example, salinity has been shown to be a defining "bottom-up" control on communities, as demonstrated by studies along salinity gradients such as those found in the Baltic sea (Herlemann et al, 2011, Lozupone and Knight, 2007, Pontarp, Sjostedt, and Lundberg, 2013). Temperature is another strong bottom-up factor, and in addition to having spatial scale (such as across latitudes) it also has a temporal scale (such as seasonality). A gradient of seasonality and temperature variability have been invoked in explanations for why communities differ at different

latitudes, but are similar between the two poles (Ghiglione et al, 2012). This is due to the similarly extreme seasonality that occurs at both poles (Ghiglione et al, 2012). In a global survey of coastal habitats, bacterial richness increased with increasing latitude indicating a general higher diversity near the equator (Pommier et al, 2007). In contrast, a latitudinal study of both coastal and open-ocean habitats revealed that diversity increases at temperate latitudes (peaking at $\pm 30^{o}$) rather than towards the equator, and that peaks in bacterial richness occurs in the winter season (Ladau et al, 2013).

We can further understand patterns in bacterial diversity by further understanding the different ecosystems found in the ocean. Coastal ecosystems are typically nutrient-rich and highly productive areas. For example, in the spring, primary production from coastal oceans at temperate latitudes is responsible for approximately 19% of net global primary production (Field et al, 1998). It is during this season that larger heterotrophic bacteria capable of degrading biopolymers originating from phytoplankton are known to bloom in response (Teeling et al. 2012). An example of these biopolymer-degrading bacteria associated with spring phytoplankton blooms are the Bacteroidetes, specifically the Flavobacteria (Kirchman, 2002, Teeling et al, 2012). Flavobacteria are early responders to phytoplankton blooms and are capable of degrading high-molecular weight organic matter such as chitin, pectin, and polysaccharides. One of the immediate responders to spring phytoplankton is the specialist *Polaribacter* sp. (Kirchman, 2002, Fernández-Gómez et al, 2013, Alonso et al, 2007). Certain *Polaribacter* strains have the proteorhodopsin gene suggesting a dual lifestyle in which they grow optimally on organic-rich particles, yet when forced to float freely in nutrient poor waters, for example during a season when nutrients are depleted, they can use the proteorhodopsin to obtain

energy from light (Fernández-Gómez et al, 2013). *Cytophaga* is another flavobacterial group which increases in abundance following phytoplankton blooms (Teeling et al, 2012). Like *Polaribacter*, *Cytophaga* is able to degrade cellulose and has gliding motility (Xie et al, 2007). These flavobacterial clades have been shown to be abundant in the North Atlantic ocean during the spring season (Gómez-Pereira et al, 2011).

Another bacterial specialist found in coastal oceans is the OM43 clade of Beta-proteobacteria. OM43 bacteria are known for the degradation of one-carbon (C1) compounds such as methanol, methylamine or formate, but not methane, in surface water (Chistoserdova and Lidstrom, 2013, Chistoserdova, Kalyuzhnaya, and Lidstrom, 2009, Giovannoni et al, 2008). Methanol uptake was found in the Northeast Atlantic (Dixon, Beale, and Nightingale, 2010). Furthermore the OM43 clade was observed along the Oregon coast to be associated with a diatom bloom (Morris et al, 2006).

Contrasting to the coastal nutrient-rich environment, is the oligotrophic open-ocean where nutrients are limited. Studies have shown that the oligotrophic ocean is characterized by lower diversity and a higher relative abundance of the free-living generalist SAR11 bacteria (Zinger et al, 2011, Morris et al, 2002). SAR11 are aerobic, free-living heterotrophic bacteria that prefer oligotrophic environments (Giovannoni and Stingl, 2005, Tripp, 2013). In a study that modeled the global distribution of SAR11, different characteristic SAR11 ecotypes (phylogenetic subclades within this clade that are adapted to specific environmental conditions), inhabited different latitudes correlating with temperature (Brown et al, 2012).

As important as understanding how bacterial communities change at different spatial scales it is also important to understand how bacterial communities change at different time scales. For example, changes in bacterial metabolism and physiology can occur over very short time-scales (hours), say, for example, in response to a phytoplankton bloom (Rinta-Kanto et al, 2011, Howard et al, 2011). At the weekly time-scale, blooms in phytoplankton can alter the ecosystem, and bacterial groups can further bloom in response (Li and Harrison, 2008, Li and Dickie, 2001). Furthermore, annual and decadal time-scales can show changes in phenology and response to long term environmental forcing (Li and Dickie, 2001; Li and Harrison, 2008). Over the long-term (such as multi-annual and decadal time-scales) microbial changes in phenology, abundance and community composition can occur in response to long-term environmental forcing. For example, changes in phytoplankton, bacterial and archaeal communities were observed in the Arctic after the 2007 sea ice minimum (Comeau et al, 2011). Time-decay models, ie how similar communities are over time, are crucial in understanding temporal biogeography. A recent study has shown that similarity between bacterial communities, or beta-diversity, depend on different time-scales (Hatosy et al, 2013).

Recently, several studies have reported on the seasonal variability of microbial communities at distinct locations in the ocean. A number of locations are home to important ocean microbial monitoring programs, including the subtropical Pacific gyre (Hawaiian Ocean Time-series, HOT) (Brown et al, 2009), the subtropical Atlantic gyre (Bermuda Atlantic Time-series, BATS) (Morris et al, 2005), and coastal locations in California, (San Pedro Ocean Time-Series, SPOT) (Fuhrman et al, 2006, Fuhrman and Steele, 2008), the Western English Channel (Western Channel Observatory, WCO)

(Gilbert et al, 2011), and the Northwest Atlantic Ocean (Atlantic Zone Monitoring program, AZMP) (Li and Dickie, 2001, Li, Harrison, and Head, 2006). Because of the differences in locations of these monitoring programs, seasonal bacterial community composition varies between them. In the open-ocean environment of the northwestern Sargasso sea, annual cycles of productivity occur with spring phytoplankton blooms because of mixing (Treusch et al, 2009, Morris et al, 2005). Then productivity decreases in summer and autumn due to extensive nutrient depletion and strong vertical stratification (Treusch et al, 2009, Morris et al, 2005). Monthly data from 1996-2003 showed that the spring bloom in phytoplankton is associated with SAR11, and marine Actinobacteria OM1, while the summer bloom- is characterized by the SAR86 and SAR92 clades (Giovannoni and Vergin, 2012). In comparison, at HOT, located in the tropical Pacific open ocean, bacterial community composition is not affected as much by seasonal change as stratification is stable throughout the year and resemble those found in the oligotrophic summer at BATS (Brown et al, 2009, Giovannoni and Vergin, 2012).

In contrast to the open-ocean is the coastal environment at SPOT, where bacterial communities are strongly associated with season and could even be predicted based on season (Fuhrman et al, 2006). Furthermore, seasonal studies show SAR11 to be associated with the summer (Chow et al, 2013), however in the coastal environment of the western English Channel, which is much further north in latitude, SAR11 is associated with the winter (Gilbert et al, 2011). In these studies samples are typically collected monthly over many years and due to logistical difficulty in periodically collecting samples, long-term time series studies are harder to conduct at high frequency.

1.4 Microbial plankton in the Northwest Atlantic Ocean

There is a strong relationship between seasonality and global change and therefore observing seasonal cycles of microbial plankton can help understand response to long-term environmental forcing (Li and Dickie, 2001, Giovannoni and Vergin, 2012).
In order to observe plankton responses to environmental change, a network of ecosystem monitoring sites in the Northwest Atlantic Ocean is maintained by the Canadian Department of Fisheries and Oceans. This program, known as the Atlantic Zone Monitoring program (AZMP) includes time-series observations at many locations in different regions. Three regions include the Scotian Shelf, the Labrador Sea and the Bedford Basin. These water bodies are influenced by a coastal current system that is part of the subpolar gyre, and each has regional variations (Loder et al, 1998). Phytoplankton patterns have been shown to be annually similar (since 1999) with diatoms occurring in the spring and dinoflagellates occurring in the autumn (Li, Harrison, and Head, 2006). Phytoplankton abundance ranges from a minimum 10^9 cells m^{-3} in winter to maximum of 10^{11} cells m^{-3} at the autumnal equinox (Li and Dickie, 2001, Li, Harrison, and Head, 2006). Phytoplankton biomass on the continental shelf and the Labrador Sea have undergone sustained changes in the spring and autumn, which is consistently accompanied by changes in bacterioplankton abundance that are coherent in the direction of change(Li, Harrison, and Head, 2006b). Further studies using a 14 year time series in Bedford Basin was used to measure anomalies from long term mean conditions in both climate and plankton to look at multiyear changes and observe long term environmental forcing (Li and Harrison, 2008). Anomalies in nutrients and atmospheric signals were shown to propagate into and affect changes in phytoplankton, such as mean cell size, over multi-year scales (Li and Harrison, 2008). Furthermore, anomalies in surface Chlorophyll

a (Chl a) propagated to anomalies in bacterioplankton abundance (Li, Harrison, and Head, 2006b, Li and Harrison, 2008).

1.5 Annual cycle in Bedford Basin

Bedford Basin, located in Halifax Harbour, Nova Scotia (**Figure 1**), is considered a representative model ecosystem for the Northwest Atlantic (Li and Dickie, 2001). Bedford Basin has been sampled weekly since 1992, and the >20 year weekly time series of phytoplankton shows a repeating annual cycle with strong year-to-year variability. The annual cycle occurring in the Bedford Basin is shown in **Figure 2**. The annual cycle of phytoplankton is characterized by a sustained increase in phytoplankton abundance from winter to summer that is closely coupled with seawater temperature, which increases slowly from week 8-37. The spring bloom peak in Chl *a* occurs when the nutrient-rich water column is stabilized and a thermocline has developed (around week 13). Spring phytoplankton mostly consists of diatom species, characterized by large cell size. In the summer, phytoplankton are maintained, but at lower levels as the nutrients are consumed (Li and Dickie, 2001). Nitrate reaches the detection limit during the summer season. Summer peaks in Chl *a* consist of 50% *Synechococcus*. Cryptophytes bloom in the late spring (week 21) and then fully peak in the late summer coinciding with the other phytoplankton (week 37) having a 10-fold increase from minimum to maximum (Li and Dickie, 2001). The peak in bacterial cell abundance occurs in the early summer (around week 26). In the autumn, as the stability of the water column is slowly eroded, a secondary phytoplankton bloom, consisting mainly of dinoflagellates, occurs (week 38) as a result of nutrients brought upwards (Li and Dickie, 2001). In the winter,

phytoplankton remain at a minimum as the water column is completely mixed and solar radiation is low (Li and Dickie, 2001). The seasonal variation of *Synechococcus* is distinct in that there is a maximum in late summer and minimum in late spring of the following year (Li and Dickie, 2001).

1.6 Study Objectives

In this study, we determined bacterial community composition and diversity over multiple years from Bedford Basin with the objective of situating bacterial community dynamics in with the already well-described annual cycle of phytoplankton dynamics. Over the six years, we used samples that were previously collected at four distinct seasonal time points in 2005-2010. Bacterial communities were assessed from the Spring Equinox representing the beginning of the spring season and early phytoplankton bloom; Summer Solstice representing the longest day length in the summer and the highest amount of bacterial production; the Autumn Equinox representing the beginning of the autumn season and characteristic of the secondary peak in phytoplankton productivity; and the Winter Solstice representing the winter season and characteristic of the shortest day length. In addition, to investigate the stability and succession of bacterial communities over shorter time scales, we performed, a biweekly nested study over a single annual cycle (year 2009).

2.0 Methods

2.1 Site description and sampling

Bedford basin is the inner portion of the Halifax Harbor on the Atlantic Coast of Canada (**Figure 1**). The basin is an estuary with a surface area of 17 km^2 and a maximum depth of 71 m and connects to the adjoining continental shelf through a narrow (400 m) and shallow (20 m) sill (Li and Dickie, 2001). The basin receives fresh water from the Sackville River at an annual average inflow of 5000 $L\,s^{-1}$ (Li and Dickie, 2001). The average circulation is a two-layer structure in which lower density surface water flows into the Atlantic ocean and deeper saline water flows into the basin over the sill (Li and Dickie, 2001). The exchange of shelf and inshore water is caused by alongshore winds driving the Ekman transport (Li and Dickie, 2001).

Seawater samples were collected by Bedford Institute of Oceanography (BIO) researchers weekly from the Compass Buoy station (44^0 41' 30'' N, 63^0 38' 30''W) at 1 m depth, fixed in 1% paraformaldehyde, quick frozen in liquid nitrogen, and subsequently stored at -80 °C (Li and Dickie, 2001, Li, Harrison, and Head, 2006). The conductivity, temperature, and depth (CTD) profiles were recorded in tandem with a rosette of Niskin bottles that was deployed into the water and controlled electronically to open at specific depths to collect seawater (Li and Dickie, 2001, Li, Harrison, and Head, 2006). Methods for physicochemical and biotic measurements are described in detail in Li and Dickie, 2001.

2.2 Molecular methods

One milliliter of the archival seawater samples was pre-filtered through a Whatman GF/D pre-filter (2.7 μm) to remove larger particles and eukaryotes, and microbial biomass was then collected on a GE polycarbonate filter (0.2 μm) using vacuum filtration (-0.2 bar or 100kPa). Filters were washed with 15 ml of MilliQ ultra-pure water to remove any residual formaldehyde preservative. Filters were cut into 8 equal sections and each placed into a single 200 ul tube. Polymerase Chain Reaction (PCR) was used to amplify the hyper-variable V5 region on the 16S ribosomal RNA gene. A previously developed filter PCR approach PCR was performed using universal bacterial 16S PCR primers (Kirchman et al, 2001). The primers used were DW786F (5'-GATTAGATACCCTSGTAG-3') and DW926R (5'-CCGTCAATTCMTTTRAGT-3') modified from Baker, Smith, & Cowan, (2003), to eliminate bias against marine Alpha-proteobacteria. Reactions were performed in 50-μl volumes containing $1/8^{th}$ of a filter immersed in 2.5 μL of 10 mM (conc. final 0.5 μM) each primer, 10 μL of 5X Phire Reaction Buffer (conc. final 1.5mM $MgCL_2$), 1 μL of 10 mM dNTPs (conc. final 0.20 mM of dNTPs) and 1.0 μL of Phire Hot Start II DNA Polymerase (Finnzymes Thermofischer Scientific). Cycling conditions involved an initial 3 min denaturing step at 98°C, followed by 30 cycles of 5 s at 98°C, 5s at 49°C, and 10 s at 72°C, and a final elongation step of 1 min at 72°C. Each reverse primer was barcoded with a specific IonXpress sequence to identify samples. DNA was purified using QIAquick Gel Extraction Kit (QAIGEN). Quantification of PCR product was conducted using Quantifluor dsDNA System (Promega), a fluorescence DNA-binding dye enabling sensitive quantification of small amounts of dsDNA. PCR products were pooled and a total of 3.5×10^7 molecules were used in an emulsion PCR using an Ion OneTouch

Template kit (Life Technologies) and OneTouch and OneTouch ES instruments (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. PCR amplicons from the seasonal dataset (2005-2010) was sequenced on the Ion Torrent PGM using the 314 chip with the Ion sequencing 100 base pair kit at the National Research Council Biotechnology Research Institute in Montreal (NRC BRI). PCR amplicons from the biweekly dataset (2009) was sequenced using the Ion Torrent PGM using a 316 chip, with the Ion Sequencing 200 base pair kit at Concordia University according to the manufacturers protocol.

2.3 16S rRNA gene sequence processing

The V5 sequence data processing was conducted using the open-source Mothur project pipeline (Schloss et al, 2009). Raw sequences were sorted and quality controlled according to size range 80-140bp, and a quality score of 17 or above. Sequences were aligned to the SILVA reference alignment, and we screened out sequences that did not align to over the V5 regions. In the biweekly dataset, because deeper sequencing technology in a new 200 base pair kit, in order to reduce potential sequencing error, sequences that were two base pairs different from each other were pre-clustered together. Then a distance matrix was made on the aligned sequences. The sequences were then clustered together at 90 and 97% identity using the average neighbor algorithm in the seasonal dataset to obtain OTU definitions at 0.03, and 0.10 distances. The sequences from the biweekly dataset were clustered together using the furthest neighbor-joining algorithm (Saitou and Nei, 1987) to obtain OTU definitions also at 0.03 and 0.10 distances. Taxonomic assignment of the final dataset was assigned using the SILVA database and GreenGenes taxonomy using the Wang method (Wang et al, 2007).

2.4 Statistical and taxonomic analysis

Relationships between bacterial communities and environmental parameters were assessed using multivariate Non-metric multidimensional scaling ordination (NMS Ordination) using Bray-Curtis distances implemented in PC-ORD statistical program (McCune and Mefford, 2011). Unweighted Pair group Method with Arithmetic mean (UPGMA) dendograms, and all alpha and beta diversity were measured on normalized dataset were conducted using Mothur. For the non-parametric estimates of species richness and species evenness, Chao1 index(Chao et al, 2000), and Shannon-evenness were used, respectively. To determine dissimilarity between samples two indices were used to compare results: Bray-Curtis (Bray and Curtis, 1957) and Yue and Clayton's theta (ThetaYC) (Yue and Clayton, 2005). To account for community membership, the Jaccard index was also calculated. The Bray-Curtis index accounts for similarity by calculating the minimum times a phylotype is commonly found between two samples and dividing it by the total number of both libraries. ThetaYC index calculates relative abundance of the specific phylotype in each sample before doing comparison, therefore placing more weight on relative abundance than Bray-Curtis. Intra- versus inter-sample diversity comparisons were analyzed using analysis of similarity statistics (ANOSIM)(Clarke, 1993) on Bray-Curtis and thetaYC indices, and attained through the Mothur. Analysis of Similarity (ANOSIM) using Bray-Curtis distances was conducted to note whether there is more significant similarity inter-seasonally or intra-seasonally. ANOSIM results in a value between 0 and 1, when comparing two communities. A zero value indicates that there is as much variation within each community as between them, therefore the lower the number the less distinguishable the communities are from each other in terms of variation. The value 1 indicates that the communities have less variation within the

community than between communities, therefore indicating distinct communities that can be distinguished from each other. Bartlette's test for homogeneity of variance, Kolmogorov-Smirnov test for normality, Analysis of variance (ANOVA), Tukey's multiple comparisons of means, and box plots of taxonomic groups for each season were done using R (R core team, 2013).

3.0 Results

3.1 Summary of the 16S rRNA gene sequence datasets

3.1.1 Multi-year time-series (2005-2010)

To investigate temporal bacterial community dynamics in Bedford Basin, community composition was assessed at four seasonal time points between 2005 and 2010. The seasonal time points include the spring equinox (SE, week 13), summer solstice (SS, week 26), autumn equinox (AE, week 38) and winter solstice (WS, week 51), respectively (**Figure 3**). The V5 hypervariable region of the 16S rRNA gene was analyzed from 23 time points in total. After trimming and pre-clustering, the dataset consisted of 32,810 V5 sequences, which were comprised of 24,314 unique sequences. We generated between 612-1,200 sequences for each time point (**Table 1**). Sequences were clustered at 97% sequence similarity (*i.e.* the species level) which resulted in 12,937 operational taxonomic units (OTUs, here is referred to as phylotypes). Sequences were also clustered at 90% sequence identity (*i.e.* the family level), resulting in 2,196 phylotypes. After correcting for uneven sample size there were 6,283 phylotypes at 97% and 1,160 phylotypes at 90%.

3.1.2 Biweekly time-series (2009)

To determine bacterial community succession at higher temporal frequency over a single year, 26 biweekly time points from 2009 were selected for V5 sequence analysis. After trimming and pre-clustering, we generated 151,870 V5 sequences, consisting of 75,418 unique sequences. Between 3,037 and 10,045 sequences were generated per time point (**Table 2**). Sequences were clustered at 97% sequence similarity, which resulted in 50,943 phylotypes, and sequences clustered at 90% sequence similarity resulted in 14,116 phylotypes. To correct for uneven sample size, sub-sampling the dataset resulted in 14,273 phylotypes at 97%, and 5,730 phylotypes at 90% sequence identity.

3.2 Physico-chemical environment

Weekly records of water temperature, salinity and inorganic nutrients indicated annual differences but showed a consistent pattern of seasonal variation over the 2005-2010 time frame of this study (**Figure 3**). Surface water temperature was lowest in mid-February (**Figure 3a**) and highest in early September. Inorganic nutrients (NO_3^- and PO_4^{3-}) **(Figure 3b and 3c)** were highest in winter, decreased through the spring, and were near or at the detection limit in the summer samples. A table containing all environmental measurements can be found for both the long-term (2005-10) (**table 3**) and the high frequency (2009) time-series **(table 4)**.

3.3 Phytoplankton patterns during 2005-2010

Because of the strong dependence of bacteria on phytoplankton and how bacterial communities can potentially differ in relation to the timing of phytoplankton bloom, it is

important to understand phytoplankton dynamics during 2005-2010. The Chl *a* data showed year-to-year variability in the timing of the spring and autumn phytoplankton blooms (**Figure 3e and 3f**). Over the time course of our study, spring phytoplankton blooms occurred during week 13 (2005, 30.7 mg/m^{-3}), week 10 (2006, 17.3 mg/m^{-3}), week 12 (2007, 16.91 mg/m^{-3}), week 15 (2008, 13.48 mg/m^{-3}; 2009, 22.01 mg/m^{-3}), and week 11 (2010, 18.49 mg/m^{-3}). Based on Chl *a* peaks, the SE-2006, SE-2007, and SE-2010 communities can be considered post-bloom communities, although we predict that these communities will not be the same as they differ in duration after the phytoplankton peak. The SE-2008 and SE-2009 communities are pre-diatom bloom communities, distinguished from the post-diatom bloom communities.

We also observed inter-annual variability in the autumn secondary Chl *a* peak. Over the six year time series the autumn phytoplankton peaks occurred at week 38 (2005, 15.28 mg/m^3), week 41 (2006, 23.75 mg/m^{-3}), week 45 (2007, 15.4 mg/m^{-3}, week 40 (2008, 14.17 mg/m^3), Week 36 (2009, 18.05 mg/m^{-3}), and week 44 (2010, 9.75 mg/m^{-3}). Although the AE-2010 communities were surveyed before the autumn peak in Chl *a*, these communities occurred one week after the secondary peak in small (<20 um) phytoplankton (week 37, $3.15x10^5$ cells/ml). All specific phytoplankton abundances, counted by flow cytometry, for each sample in both datasets can be seen in **Tables 5 and 6.**

To investigate short-term changes in bacteria over the 2009 annual cycle, it is important to understand the yearly phytoplankton trends in 2009, which influence bacterial community structure. Three Chl *a* peaks occurred during 2009: the spring peak

(week 15, 22.01 mg/m^{-3}), a high summer peak (week 27, 27.74 mg/m^{-3}), and a peak in the autumn (week 36, 18.05 mg/m^{-3}). However, because the autumn is characterized by smaller (<20 um), more-diverse chlorophyll-poor phytoplankton, it is important to understand the distribution of different phytoplankton groups (**Figure 4**). There are three characteristic peaks that are observed in total phytoplankton (**Figure 4a**): week 29 ($1.09x10^5$ cells/ml), dominated by small nanophytoplankton (2-10 μm, **Figure 4b**); week 39 ($1.19x10^5$ cells/ml) dominated by large nanophytoplankton (10-20 μm, **Figure 4c**) and week 48 ($1.32x10^5$ cells/ml) dominated by picophytoplankton (<2 μm, **Figure 4d**).

3.4 Taxonomic composition of Bedford Basin bacteria

The major taxonomic groups identified in Bedford Basin bacterial community were typical of the coastal ocean, consisting of Proteobacteria, Bacteroidetes, and marine Actinobacteria (**Figure 6**). Within the Proteobacteria, the classes Alpha- Gamma- and Beta-proteobacteria were present. In the Bacteroidetes, the Flavobacteria clade was the most abundant group (28%). At the SE, the most abundant bacterial phylum, on average, was the Bacteroidetes (45%), which was largely comprised of *Polaribacter* (26%) and *Cytophaga* (14%). In the SS, samples were dominated by Alpha-proteobacteria (43%), comprised mostly by SAR11 (23%). We also observed *Cytophaga* (16%), and the gamma-proteobacteria SAR86 (8%) at the SS. The AE was associated with SAR11 (21%), Rhodobacteriales (17%), and the Actinobacteria OM1 (6%). The WS is associated with the Gamma-proteobacterial Sulfur Oxidizers (GSO) clade (10%) as well as Alteromonadales (10%), Rhodobacteriales (13%) and OM38 (5%).

We then investigated the temporal distribution of the most abundant taxonomic groups in order to explore their annual variability in Bedford Basin (**Figure 7**). *Polaribacter* was on average more abundant during the spring, yet exhibited high year-to-year variability (13% to 53%) (**Figure 7a**). The highest *Polaribacter* abundance was observed in SE-2007. The peak phytoplankton bloom occurred one week prior to SE-2007, hence the observed *Polaribacter* bloom (53%) may be a response to an increase in DOM due to phytoplankton lysis. *Polaribacter* was significantly more abundant at the SE than at the AE (ANOVA, $p<0.001$) and WS ($p<0.01$). The second must abundant Bacteroidetes group was *Cytophaga*. We observed that *Cytophaga* was associated with the SE and sustained through SS (**Figure 7b**). The spring and summer samples were significantly different than the autumn and winter samples with regards to *Cytophaga* abundance ($p<0.01$), possibly indicating that this clade is associated with the spring bloom, but follows in succession after *Polaribacter*.

The SAR11 clade was associated with nutrient-depleted summer waters, being significantly more abundant at the SS than at the SE ($p<0.05$). SAR11 sustained throughout the autumn seasons and was more abundant in the summer/autumn than winter/spring. The gamma-proteobacteria SAR86, was on average more abundant in the SS, but showed inter-annual variability over the multi-year time series. The Rhodobacteriales group was associated with the AE, most likely responding to autumn primary production.

The GSO clade was significantly more abundant in winter than in any other season ($p<0.005$) (**Figure 7F**). The OM43 clade of the Methylophilales, although

seeming to be associated with winter, did not exhibit a clear seasonal trend (**Figure 7G**). OM43 do not associate with a specific season but showed a general increase in relative abundance over the 6-year time-series. This hinted that there are other factors not directly linked to seasonality to which members of this group are responding. Actinobacteria, in general, were in low abundance, comprising less the 5% of the V5 sequences, however in Autumn 2010, the OM1 clade of Actinobacteria comprised 24% of the V5 sequences, signifying a response to a possible environmental perturbation. The single year dataset also contained the same bacterial groups found in the seasonal dataset associated with the same seasons at the biweekly time-scale (Supp Figures 4 and 5**)**.

3.5 Seasonal variation in bacterial diversity

Previous time-series studies in the temperate coastal ocean have documented that bacterial richness is maximal in the winter and minimal in the summer (Gilbert et al, 2011). To test whether we observed a similar pattern we investigated seasonal variation in bacterial diversity in Bedford Basin. Bacterial richness was estimated using the Chao1 index, and bacterial evenness was estimated using the Shannon-evenness index. Both estimates were generated at two phylogenetic scales (at 97% and 90% sequence identity) (**Figure 8**). We estimated that there were between 1,600 and 4,500 bacterial phylotypes (97% identity) and highest richness was indeed observed during winter (WS-2008, 4,467 phylotypes). Although there was variability in richness for all time points, there was no significant difference between any seasons. At lower phylogenetic scale, we estimated bacterial richness to be between 220 and 520 phylotypes. The lowest observed richness was in the summer (SS-2010; 220), and the highest was observed in the autumn (AE-

2009, 520). Moreover, at lower phylogenetic scale there was statistical difference between the estimated richness in summer and the autumn (ANOVA, $p<0.05$). This indicated that in the summer, there was lower bacterial richness at the family level, but similar diversity at the species level (97%).

The general trend in bacterial community evenness indicated that spring communities were the least even, and evenness progressively increased through summer and autumn, peaking in the winter (**Figure 9**). The spring community evenness at 97% was significantly lower than the winter community (ANOVA $p<0.05$), as spring samples were characterized by dominating bacterial groups (blooms) whereas species in the mixed winter were at a more similar relative abundance (**Figure 9a**). There were two outlier time points observed to have dramatically lower evenness values and they both coincided with the bloom of specific bacterial lineages. The lowest evenness score was observed at SE-2007 (0.83), coinciding with the *Polaribacter* bloom at this time. In addition, a low evenness score was also observed in AE-2010 (0.86 at 97%), which was a result of the actinobacterial OM1 bloom. At 90%, evenness showed similar patterns, however, there were no statistical differences, which indicates that evenness is not as variable at the family level than it is at the species level.

Bacterial richness in the biweekly dataset is one order of magnitude higher than the multi-year dataset, however this is due to the larger number of sequences used in the 2009 dataset. Diversity indices are presented for both datasets at 97% and 90% in Supplementary **Tables 1 and 2.** Although the absolute values cannot be directly compared between datasets, the general temporal trends can be compared. We estimated

that there were between 3,200 and 9,200 bacterial phylotypes with the average being 6,200 phylotypes (97% id, **Figure 10a**). The highest richness was observed during autumn (week 42, 9,200 phylotypes). At lower phylogenetic resolution we observe between 1,100 and 2,300 phylotypes, with an average of 1,750 phylotypes (Figure **10b**). The highest richness observed at lower phylogenetic resolution occurred in the winter (week 2, 306 phylotypes). Comparing community richness throughout the single year, there is an observed decrease in richness and evenness both at 97% and 90% between weeks 10-13, coinciding with a bacterial bloom associated with the spring (**Figure 10a-d**). The same trend occurred between week 17 and week 20 showing the bloom in oligotrophic summer-associated bacteria. While bacterial richness varied in the autumn (between week 40-48), evenness was constant at 97% and varied slightly at 90% indicating that the autumn consists of a highly diverse, but relatively even community. Richness and evenness both decrease at week 52 indicating another bloom of bacterial, and this is as a result of an OM38 (alpha-proteobacteria) phylotype bloom (discussed in section 4.1).

3.6 Seasonal variation in bacterial community structure

To observe how bacterial community structure changed over time, we used a number of different community similarity calculators (Bray-Curtis, Jaccard index, ThetaYC). However, for simplicity, and because general results were similar among calculators, I will focus on the Bray-Curtis similarity index calculated at the family level (90% sequence identity). The ThetaYC dissimilarity index is discussed in **Appendix 3**. The average similarity between communities sampled at the same season was 55% (Bray-

Curtis similarity value), with the least variation observed among WS communities (average 58% similarity) and the most variation among AE communities (average 45% similarity). In general, similarity decreased proportionally to the time lag between samples, with the least similar communities being those that were six months apart, particularly between the SE and AE (average 30% similar) (**Figure 11**).

Although there was significant variation among samples collected from the same season, Analysis of Similarity (ANOSIM) showed that community variation within seasons was much less than the variation between seasons (**Table 7**), especially between those that are six months apart (SE-AE, $R=0.85$, $p=0.007$; and SS-WS, $R=0.84$, $p=0.002$). However, ANOSIM values between SS and the AE ($R=0.39$, $p=0.015$) indicated that there is much variation both within and between the SS and AE, making these two seasons less distinguishable from each other. This signified that the summer and autumn are both highly dynamic and variable inter-annually. Although not as pronounced as the SS-AE ANOSIM score, the SE-WS also showed a lower ANOSIM score ($R=0.61$, $p=0.001$), making these two seasons also less distinguishable from each other. This analysis suggests the largest differentiation of communities is between the colder, nutrient rich seasons (SE-WS) and the warmer, nutrient-poor seasons (SS-AE) as each of these pairs are less distinguishable from each other, while between these pairs variation is evident.

3.6.1 Linking community structure to environmental change

Further exploration of bacterial communities using Nonmetric Multidimensional Scaling (NMS) ordination revealed seasonality over the six years, but high inter-annual variability **(Figure 12)**. The colder, nutrient-rich WS and SE samples were separated from the warmer, nutrient-poor SS and AE samples along ordination axis 1 and axis 2. Temperature negatively correlated with axis 1 and axis 2 (r_1= -0.65 and r_2 = -0.68, respectively) with a strong positive association with the AE communities, as expected because, on average, temperature reaches its maxima in early September (week 37) in this coastal system (Li and Dickie, 2001). Nutrients (nitrate, phosphate, silicate) also positively correlated with the winter and spring communities along axis r_1=0.7, 0.77, 0.6) In relation to the phytoplankton communities, the SE/SS samples were separated from the AE/WS samples along axis 2. Because the phytoplankton groups in the autumn are different from those in the spring diatom communities, the different classes of phytoplankton (*Synechococcus*, picophytoplankton (<2μm), small nanophytoplankton (2-10μm), large nanophytoplankton (>10μm)) correlated with the AE samples. Total Ch*l a* did not strongly associate with any particular season, as Chl *a* is highly variable during the year owing to the succession of numerous phytoplankton groups.

As a result of the *Polaribacter* bloom that occurred in SE-2007 one week after the diatom bloom, the SE-2007 sample was an outlier in the NMS ordination (STDEV=3.86). To the same effect, as a result of the OM1 phylotypes bloom, the AE-2010 sample was also considered an outlier (STDEV=2.05). Although not so striking as seen with SE-2007 and AE-2010, inter-annual variability was observed among all the seasons further justifying the need for higher-frequency sampling of each year.

3.6.2 Stability and succession over a single annual cycle

The multi-year time series of Bedford Basin revealed high inter-annual variability in bacterial community diversity and community structure that could be linked to environmental change. To further investigate the stability and succession of bacterial communities over shorter time scales, we analyzed a single year (2009) biweekly dataset. Just as we observed a decrease in similarity as time increased between seasons, we also observe similarity decrease over different time scales (**Figure 13**). Bacterial communities sampled two weeks apart are on average 34% similar (Bray Curtis similarity) to each other. As time increased to a month, community similarity decreased to an average of 30%. Bacterial communities that are three months apart are on average 25% similar and further decreasing to 20% similar at six months. It was expected that bacterial communities would change as time increases. However, we observed that even within two weeks, communities could be just as different as those found 6 months apart, showing the dynamic nature of bacterial communities over short time scales.

We further explored temporal patterns by plotting community similarity values between time points that were two weeks apart (**Figure 14**). The highest similarities were observed in comparisons between winter communities, particularly during week 2-4 (43%) and 6-8 (46%), suggesting winter communities are stable, in addition to being even and diverse (**Figure 10**). In support of the high similarity over these two time periods, the taxonomic compositions were nearly identical between these samples (Supplementary Figure 2 and 3**).** For example, over weeks 6 to 8, bacterial communities were comprised

of Bacteroidetes (30% ±1%), Gamma-proteobacteria (40% ±2%), and Alpha-proteobacteria (23% ±2%).

Contrasting to the relative stability observed in winter, we observed a few periods of rapid succession through the spring and summer. The first instance where communities rapidly changed occurred between weeks 17-20 (21%), which coincided with the dramatic change that occurs annually in association with the end of the spring phytoplankton bloom. From a taxonomic perspective, this dramatic shift was characterized by a transition from a Bacteroidetes dominated community (31% *Cytophaga* and 24 % *Polaribacter*) to one dominated by the oligotrophic summer-associated Alpha-proteobacteria SAR11 (42%) in week 20 (**Figure 15a)**.

The second instance in which community succession was particularly rapid was from weeks 30 to 32 (19%) (**Figure 14**). Because this shift occurred in the middle of summer, it can be categorized as an unusual event that would have otherwise gone undetected if it were not for our high-frequency sampling. Taxonomically, we observed a striking shift that appeared to be opposite to that observed during the post-phytoplankton transition. At lower taxonomic resolution the community change from one dominated by Alpha-proteobacteria (48%) to one dominated by Bacteroidetes (52%). However, these Bacteroidetes were not comprised of typical marine Flavobacteria that were observed previously, but rather a deep-branching group within the Sphingobacteria termed the KSA1 clade (27% of sequences) (**Figure 15b**). Since this unusual Bacteroidetes group was less than 1% of the community at all other times, we can infer that this is recruitment from the rare biosphere in response to some potential environmental perturbation.

Ordination analysis of the 2009 time-series revealed a seasonal cycle similar to that of the multi-year analysis. We observed a general seasonal succession in the community, but as we observed earlier, we also captured specific events that occur during the year. For example, week 20 looks more like the summer season samples, rather than the spring samples, which reflects the shift to the summer-associated SAR11 phylotypes. Also, as a result of the increase of the Bacteroidetes clade, as was discussed earlier, week 32 is considered an outlier (STDEV=2.56). The autumn samples show high variability as weeks 44, 46, and 48 show further association on the NMS from the rest of the autumn samples and from each other. Coinciding with the richness and similarity estimates, these are the weeks that are highly dynamic (variable chao1, and Bray-Curtis similarity), but also highly even (**Figure 10 and 14**).

4.0 Discussion

4.1 Placing bacterial community structure into the Bedford Basin monitoring program

In this study, we provide the first investigation of bacterial community dynamics in the Northwest Atlantic Ocean. There are many studies that have looked at temporal dynamics in other ocean locations (Head & Pepin 2010; Treusch et al. 2009; Rappé et al. 2002; Fuhrman & Steele 2008; Edwards et al. 2001; Gilbert et al. 2011; Giovannoni et al. 1990), however, never has a study been done in the coastal North Atlantic, particularly with such high frequency of sampling. It is important to understand the annual cycle and, for the first time, place the bacterial community structure in the context of the biotic (phytoplankton) and abiotic (physico-chemical) environment. Coinciding with the annual cycle that occurs in the water column, there is an annual shift in the surface bacterial community composition.

In the spring, surface ocean temperature begins to rise, which drives stratification. The stratification that occurs leads to the first phytoplankton bloom characterized by diatoms. Because the spring phytoplankton in the Bedford Basin are the largest of the year (Li and Dickie, 2001), they release complex biopolymers. Therefore, biopolymer specialists, such as the Bacteroidetes are the first to respond. For example, in the biweekly study, the Bacteroidetes abundances increased two-fold from week 13 to week 15 (27% to 47%) closely following the peak in Chl *a* (week 15). We observed the flavobacterial groups *Polaribacter* and *Cytophaga,* which are linked with phytoplankton blooms because of their ability to degrade complex polymers (Alonso et al, 2007, Teeling et al, 2012, Kirchman, 2002). Genome analysis of flavobacteria isolates have shown they

possess genes involved in the breakdown of polysaccharides (carbohydrate-active enzymes, CAzymes), including β-D-fucosidases, which can degrade fucose (Teeling et al, 2012). Fucose was found to be a major constituent of diatom exopolysaccharides (Khodse and Bhosle, 2010). Furthermore, many algal polysaccharides are sulfated, and recent reports showed sulfatases are expressed in *Polaribacter* during phytoplankton blooms (Gómez-Pereira et al, 2011). Although not as abundant, we also observed the gamma-proteobacterial clade SAR92 (dominated by the uncultivated clade ZA2333c, average SE, 8%) associated with the spring phytoplankton bloom. Like the flavobacteria, the gamma-proteobacteria SAR92 clade contains genes for glycoside hydrolases, necessary to break down biopolymers, as was observed by Teeling et al. (2012). Furthermore, both these groups contain the similar transport protein profiles (TonB-dependent transporters) thought to be carbohydrate transporters as well as other substrate transporters (Gómez-Pereira et al, 2011, Noinaj et al, 2010), indicating that they are both utilizing phytoplankton derived carbohydrates (Teeling et al, 2012).

In contrast to a study observing the succession of phytoplankton associated bacterial groups in the North Sea (Teeling et al, 2012), we observed that the gamma-proteobacteria peaked before *Polaribacter* and *Cytophaga* (Figure **15a**). We also observed coupled blooms of *Polaribacter* and *Cytophaga* over the spring phytoplankton bloom, with both peaking two weeks after the Chl *a* peak (week 17; *Polaribacter*, 25%; *Cytophaga*, 30%) (Supplementary Figure 2 and 3). In 2007, the spring Chl *a* peak occurred at week 12, and the bacterial sample was taken from week 13, and although this showed a bloom in *Polaribacter* (58%), this was also the highest peak observed for *Cytophaga* (21%) in the spring. After the phytoplankton bloom, both *Polaribacter* and

Cytophaga decreased by half (7% and 9%, respectively by week 20 in biweekly dataset), and this could be due to grazers or viruses. In Bedford Basin, bacterial and viral abundances increase at sustained rates in the spring and may lead to the rapid crash that is observed in the phytoplankton-associated bacteria (Li and Dickie, 2001).

In the summer, from SS to AE, the thermocline is fully formed and surface nitrate approaches undetectable limits, resulting in a warm, stratified, oligotrophic environment. However, ammonium becomes available as a result of microbial remineralization (Li and Dickie, 2001). Furthermore, virus to bacteria ratios reach their respective maxima at the summer solstice, coupled to peak in bacterial cell counts (Li and Dickie, 2001).

In summer, while bacterial diversity at the "species" level is similar to other seasons, we observed a significant decrease in diversity at the "family" level (**Figure 4a and b**). This is a result of the abundance in the Alphaproteobacteria SAR11 clade, characterized as small, aerobic, free-living heterotrophic generalists. The SAR11 that was observed post-bloom in Teeling *et al*. (2012), showed high expression of ATP-binding cassettes (ABC) and tripartite ATP-independent (TRAP) transporters, which reflects the activity of SAR11 (Giovannoni et al, 2005, Tripp, 2013, Teeling et al, 2012). SAR11 is able to grow under oligotrophic conditions by means of high-affinity ABC and TRAP transporters augmented by energy produced by proteorhodopsin (Teeling et al. 2012; Tripp 2013; Giovannoni 2005). Peaks in SAR11 relative abundances occurred in the SS and the AE, comparable to the same observations seen off the coast of California at SPOT (Fuhrman and Steele, 2008, Chow et al, 2013) and in the Sargasso sea at BATS (Carlson

et al, 2009). This is in contrast to the winter peaks of SAR11 at HOT (Eiler et al, 2009) and the Western English channel (Gilbert et al, 2011).

In general, most phylotypes were rare and only represented by a small fraction of the surface community (**Figure 5**). However, these rare phylotypes may represent a seed bank of diversity and may become abundant when conditions shift, as was observed in the ephemeral growth of the uncultivated Bacteroidetes clade (named KSA1) that occurred mid-summer (week 32, **Figure 16b**). This uncultivated Bacteroidetes KSA1 group was most closely related to the *Balneola* genus within the family Sphingobacteriia, (93% 16S rRNA gene sequence similarity) (Agogué et al, 2005, Urios et al, 2006, Urios et al, 2008). The genus *Balneola* is closely related to other extremophilic Bacteroidetes, and have been implicated in the degradation of toluene and benzene (Li et al, 2012, Urios et al, 2008). Furthermore, *Balneola* were the dominant organisms in enriched toluene media from saline (2-2.5 M NaCl), contaminated sediment off the east China sea (Li et al, 2012). This evident recruitment from the rare biosphere in Bedford Basin could be linked to an environmental pulse of an aromatic hydrocarbon such as benzene.

Another bacterial clade that was observed at the SS and the AE are the alpha-proteobacteria Rhodobacteriales (**Figures 7**). This clade typically comprises 20% of coastal ocean bacterial communities (Buchan, Gonzalez, and Moran, 2005). The Rhodobacteriales are dominated by the Roseobacter group, which are typically wide-spread and found in many marine environments (Buchan Gonzalez, and Moran, 2005). They are metabolically diverse generalists, being able to degrade DMSP to DMS during an algal bloom (Buchan, Gonzalez, and Moran, 2005), and have been associated with

algal blooms in the north Atlantic (Gonzalez et al, 2000). In our study, Rhodobacteriales begins to bloom in the SS and are sustained at higher relative abundance in the AE, potentially responding to algal blooms that occur during the SS and the AE. This is comparable to the Western English channel, in which it was observed that Rhodobacteriales were associated with the summer (Gilbert et al, 2011). They have been found in other systems such as the south Atlantic (at BATS) (Morris, Frazar, and Carlson, 2012) and the Arctic (Fu et al, 2013).

At the AE, surface water temperature is highest, and at this time, total phytoplankton is at a maximum (Li and Dickie, 2001). In the autumn, stratification begins to breakdown due to lower air temperatures and stronger winds, which bring nutrients to the surface (Li and Dickie, 2001). Mixing events lead to a secondary autumn bloom in phytoplankton characterized by different phytoplankton groups such as dinoflagellates, *Synechococcus*, and nanophytoplankton (**Figure 4**) (Li and Dickie, 2001). This season is the time of highest diversity of phytoplankton (Li and Dickie, 2001). In our study, it is during this time that we observed community richness fluctuate; indicating that this season is highly dynamic. Therefore, it might be hypothesized that if there were diverse phytoplankton groups blooming during this season, there were also diverse blooms of bacteria that respond to specific phytoplankton groups. This could explain the lower and more variable similarity values when we compared our autumn communities (**Figure 11**, **14**). Furthermore, we hypothesize, because diverse bacterial groups could be consistently blooming and crashing, causing fluctuation in richness (figure 5), the bacterial community remained even because no bacterial group dominated during this season (further supported by the richness and evenness in weeks 42-48, **Figure 10**).

An example of an intermittent bloom that occurred in the autumn was observed at AE-2010, in which bloom in Actinobacterial OM1 clade (24% of the community) dominated the sample (**Figure 7h**). Studies have shown that marine Actinobacteria OM1 clade are usually present at approximately 5% of surface bacterioplankton communities (Morris, Frazar, and Carlson, 2012). This bacterial clade was also found in the deep chlorophyll maximum at BATS during the spring bloom of picophytoplankton (Treusch et al, 2009, Giovannoni and Vergin, 2012). Our OM1 bloom was observed one week after the peak in picophytoplankton that occurred in week 37 of 2010. Based on the distribution of this clade in previous studies at BATS (Morris, Frazar, and Carlson, 2012, Morris et al, 2005a), and in our study, this marine OM1 Actinobacteria is predicted to be associated with the picophytoplankton communities.

Based on the average 52-week cycle from 1992-2000, microbial abundance is summarized as highest in the summer and lowest in the winter (Li and Dickie, 2001), which is contrary to bacterial diversity trends observed here, being low in the summer and high in the winter (**Figure 8**). A comparative study observing and modeling diversity in the winter and the summer has shown that phylotype diversity peaks in the winter season at temperate and high latitudes in the northern hemisphere (Ladau et al, 2013). This is also consistent with other seasonal studies conducted at temperate and high latitudes as a result of pronounced seasonality (Ghiglione et al, 2012, Gilbert et al, 2011).

The bacterial phylotypes associated with the winter in Bedford Basin belonged to the gamma-proteobacterial GSO clade and the Alteromonadales (**Figure 6**). The GSO

clade is comprised of two major lineages, SUP05 and ARCTIC96BD-19 (Walsh et al, 2009). The ARCTIC96BD-19 clade has been observed in surface communities (Marshal and Morris, 2012), and deep ocean communities (Swan et al, 2011), while the SUP05 clade is considered to be restricted to hydrothermal vents, oxygen minimum zones, and anoxic fjords (Stevens and Ulloa, 2008, Sunamura et al, 2004, Zaikova et al, 2010). Based on a previous genome analysis of ARCTIC96BD-19, this clade shows the potential ability for mixotrophy and sulfur oxidation. Metagenomic analysis of SUP05 showed potential for chemolithoautotrophic metabolism using reduced sulfur compounds (Walsh et al, 2009). Unfortunately the V5 region on the 16S rRNA gene cannot distinguish between SUP05 and ARCTIC96BD-19, however, based on the observed distribution of these clades, it is more likely that the ARCTIC96BD-19 clade is observed in the surface communities. Furthermore, based on a recent metaproteomic analysis of spring and winter surface and deep-water communities in Bedford basin, ARCTIC96BD-19 transporter proteins were found from surface communities (Georges et al, *submitted*).

It is during the winter season that we have observed metabolically diverse bacterial groups. In 2009, we observed an alpha-proteobacterial OM38 phylotype increase that influenced diversity estimates (week 52, **Figure 10**; Supplementary figure 8). Little is known to date about this clade, however it has been found off the coast of North Carolina (Tripp 2013; Rappé et al. 1997). In a metatranscriptomic and metagenomic study of the stratified water at HOT, OM38 was identified in surface communities (Shi et al, 2010). Furthermore, transcripts for aerobic anoxygenic phototrophy were found, thus suggesting that this phylotype could be aerobic anoxygenic occurring in the winter (Shi et al, 2010).

After placing bacteria into the seasonal cycles that occurs in Bedford Basin, we can identify bacterial groups whose abundances do not seem to be as linked to the seasonal cycle. One such example is the beta-proteobacterial specialist OM43 clade. Although this clade showed overall increase in the multi-year dataset, there was no seasonality associated with it. The biweekly dataset further confirmed this OM43 relative abundance increased and declined sporadically, pointing to another factor that is controlling its activity. The OM43 is a member of the Methylophilales group, which are known to uses one carbon compounds, such as methanol as carbon and energy sources (Giovannoni et al, 2008, Chistoserdova, Kalyuzhnaya, and Lidstrom, 2009). These coastal methylotrophs possesses methanol dehydrogenase genes (Giovannoni et al, 2008). Methanol dehydrogenase catalyzes the oxidation of methanol to formaldehyde, which then can be used in assimilatory C1 pathways (Giovannoni et al, 2008). Because Methylophilales has a unique metabolism, it is evident that there is another environmental factor controlling the growth of this clade that is not linked to seasonality, such as the input of methanol in the water column. Methanol may be deposited from the atmosphere (Heikes et al, 2002), or produced by phytoplankton (Milne et al, 1995), but generally little is known about the sources of methanol in the ocean. Furthermore, the OM43 clade has been shown to be associated with a diatom bloom (Morris, Longnecker, and Giovannoni, 2006), however, in our study, we did not observe a strong association with the diatom bloom that occurred in the spring.

4.2 The question of temporal scales in bacterial community dynamics

Changes in microbial community dynamics are often reflections of the functional

capabilities of those communities, and temporal patterns are critical to understanding ecosystem processes (Shade et al, 2013). One of the questions to be addressed is at what time-scales do the greatest microbial community changes occur? Because surveying bacterial communities is a complicated task, current survey efforts may lack the temporal resolution to capture rapid community changes. Depending on the scale of the study, species-time relationships can be explained by ecological or evolutionary processes, or even sampling depth.

We successfully demonstrated that bacterial communities change as time progresses and that β-diversity depends on temporal scale through our distance-decay graphs (**Figures 11 and 14**), which showed that bacterial communities become less similar as time increases with 6 months showing the least similar communities, with exceptions. At SPOT, bacterial surface communities were on average 40.7 ±0.2% similar (Chow et al, 2013), in this study we show bacterial communities are on average 34% similar at the weekly scale. Furthermore, the average similarity between seasons seen at SPOT is 38.4% (Fuhrman et al. 2006; Chow et al. 2013), whereas in our study it is lower (25%). This could be a result of more pronounced seasonality that occurs in Bedford basin in comparison to SPOT. In the Western English Channel only 73% of bacterial communities demonstrated reoccurring patterns at the inter-annual scale, thus implying that episodic events are important in structuring marine communities (Hatosy et al, 2013). These studies do not address episodic events as they sample monthly over many years. Therefore, because different time-scales reveal different processes, high-frequency time-series and deep sequencing can reveal short-lived two-week blooms from the rare biosphere, such as those that we have identified in our study.

5.0 Conclusions, implications and future directions

It is evident in our study, as well as the studies of others, that observing the rate at which microbial communities change can give insight into how microbial communities will be affected in the long-term (Wallenstein and Hall, 2011). As was done in many studies looking at time-decay (Hatosy et al, 2013, Shade et al, 2013), we used dissimilarity values to determine rates of change. Because we have observed microbial communities at high frequency we can begin to determine and model what is considered the standard rate of change of marine bacterial communities. However, we also have seen that microbial communities are highly dynamic, and therefore questions arise such as how to incorporate episodic events as part of a standard rate of change, and does each system (such as BATS and HOT) need its own rate of change or can we determine a universal rate of change for all marine bacterial communities. Using high frequency and long time duration studies, we can attempt to provide a baseline for community changes and for potentially quantifying the processes that drive rapid or latent change in bacterial community structure. In combining datasets from different locations, we can begin modeling global bacterial distribution over time and predict how it will change as was previously done with the cyanobacteria *Prochlorococcus* and *Synechococcus* (Flombaum et al, 2013). With a longer duration of high frequency long-term time series, we can model and predict the impacts of seasonality and how it will change in the long term.

6.0 Bibliography

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, et al. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554.

Agogue H, Joux F, Obernosterer I, Lebaron P. (2005). Resistance of Marine Bacterioneuston to Solar Radiation. Appl and Envir Microbiol 71:5282–5289.

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of Flavobacteria in marine plankton. Envir Microbiol 9:1253–1266.

Anantharaman K, Breier JA, Sheik CS, Dick GJ. (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. PNAS110:330–335.

Andersson AF, Riemann L, Bertilsson S. (2009). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. ISME J 4:171–181.

Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F. (1983). The ecological role of water-column microbes in the sea. Marine ecology progress series. Oldendorf 10:257–263.

Azam F, Malfatti F. (2007). Microbial structuring of marine ecosystems. Nat Rev Micro 5:782–791.

Baker GC, Smith JJ, Cowan DA. (2003). Review and re-analysis of domain-specific 16S primers. J of Microb Meth 55:541–555.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A. (2000). Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. Science 289:1902–1906.

Bindoff NL, Willebrand J, Artale V, Cazenave A, Gregory JM, Gulev S, et al. (2007). Observations: oceanic climate change and sea level. In: Climate Change 2007: the Physical Science Basis. Cambridge University Press: Cambridge UK, p. 386-487

Bray JR, Curtis JT. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecolog monographs 27:325–349.

Brinkhoff T, Giebel H-A, Simon M. (2008). Diversity, ecology, and genomics of the Roseobacter clade: a short overview. Arch Microbiol 189:531–539.

Brown MV, Lauro FM, DeMaere MZ, Muir Les, Wilkins D, Thomas T, et al. (2012). Global biogeography of SAR11 marine bacteria. Mol Sys Biol 8:1–13.

Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM, et al. (2009). Microbial community structure in the North Pacific ocean. ISME J 3:1374–1386.

Buchan A, Gonzalez JM, Moran MA. (2005). Overview of the Marine Roseobacter Lineage. AEM 71:5665–5677.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. (2011a). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS U.S.A. 108:4516.

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA. (2011b). The Western English Channel contains a persistent microbial seed bank. ISME J 6:1089–1093.

Carlson CA, Morris RM, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. ISME J 3:283–295.

Casamayor EO, Schäfer H, Bañeras L, Pedrós-Alió C, Muyzer G. (2000). Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. AEM 66:499–508.

Chao A, Hwang WH, Chen YC, Kuo CY. (2000). Estimating the number of shared species in two communities. Statistica Sinica 10:227–246.

Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. (2009). The Expanding World of Methylotrophic Metabolism. Annu. Rev. Microbiol. 63:477–499.

Chistoserdova L, Lidstrom ME. (2013). Aerobic Methylotrophic Prokaryotes. In: Springer Berlin Heidelberg: Berlin, Heidelberg, pp. 267–285.

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

Clarke KR. (1993). Non- parametric multivariate analyses of changes in community structure. Austral Ecol 18:117–143.

Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. (2011). Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLoS ONE 6:e27492.

Cottrell MT, Ras J, Kirchman DL. (2010). Bacteriochlorophyll and community structure of aerobic anoxygenic phototrophic bacteria in a particle-rich estuary. ISME J 4:945–954.

DeLong EF. (1992). Archaea in coastal marine environments. PNAS U.S.A. 89:5685–5689.

Dixon, J. L., Beale, R., & Nightingale, P. D. (2010). Microbial methanol uptake in northeast Atlantic waters. ISME J *5*(4), 704–716.

Doney SC, Fabry VJ, Feely RA, Kleypas JA. (2009). Ocean Acidification: The Other

CO_2 Problem. Annu. Rev. Mar. Sci. 1:169–192.

Doney SC, Ruckelshaus M, Emmett Duffy J, Barry JP, Chan F, English CA, et al. (2012). Climate Change Impacts on Marine Ecosystems. Annu. Rev. Mar. Sci. 4:11–37.

Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Richter RA, Valas R, et al. (2011). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J 6:1186–1199.

Edwards M, Reid P, Planque B. (2001). Long-term and regional variability of phytoplankton biomass in the Northeast Atlantic (1960–1995). ICES J of Mar Sci 58:39–49.

Eiler A. (2006). Evidence for the Ubiquity of Mixotrophic Bacteria in the Upper Ocean: Implications and Consequences. AEM 72:7431–7437.

Eiler A, Hayakawa DH, Church MJ, Karl DM, Rappé MS. (2009). Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre. Environ Microbiol 11:2291–2300.

Falkowski PG, Fenchel T, DeLong EF. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science 320:1034–1039.

Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, et al. (2013). Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J 7:1026–1037.

Field C, Behrenfeld M, Randerson J, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281:237–240.

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. PNAS 110:9824-9829.

Fu Y, Keats KF, Rivkin RB, Lang AS. (2013). Water mass and depth determine the distribution and diversity of Rhodobacteralesin an Arctic marine system. FEMS Microbiol Ecol:1-13

Fuhrman JA. (2009). Microbial community structure and its functional implications. Nature 459:193–199.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. PNAS 103:13104–13109.

Fuhrman JA, McCallum K, Davis AA. (1992). Novel major archaebacterial group from marine plankton. Nature 356:148–149.

Fuhrman JA, Steele J. (2008). Community structure of marine bacterioplankton: patterns,

networks, and relationships to function. Aquat. Microb. Ecol. 53:69–81.

Ghiglione J-F, Galand PE, Pommier T, Pedrós-Alió C, Maas E, Bakker K, et al. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. PNAS 109:17633-17638

Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Ben Temperton, et al. (2011). Defining seasonal marine microbial community dynamics. The ISME Journal 1–11.

Giovannoni SJ, Tripp JH, Givan S, Vergin KL, Podar M, Baptista D, et al. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. Science 309:1242–1245.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. (1990). Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63.

Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, et al. (2008). The small genome of an abundant coastal ocean methylotroph. Environ Microbiol 10:1771–1782.

Giovannoni SJ, Stingl U. (2005). Molecular diversity and ecology of microbial plankton. Nature 437:343–348.

Giovannoni SJ, Vergin KL. (2012). Seasonality in Ocean Microbial Communities. Science 335:671–676.

Gonzalez JM, Simo R, Massana R, Covert JS, Casamayor EO, Pedros-Alio C, et al. (2000). Bacterial Community Structure Associated with a Dimethylsulfoniopropionate-Producing North Atlantic Algal Bloom. AEM 66:4237–4246.

Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J, et al. (2011). Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. Environ Microbiol:1-15.

Hatosy SM, Martiny J, Sachdeva R, Steele J. (2013). Beta-diversity of marine bacteria depends on temporal scale. Ecolog Soc Am:1-31.

Head EJH, Pepin P. (2010). Spatial and inter-decadal variability in plankton abundance and composition in the Northwest Atlantic (1958-2006). J Plankton Res 32:1633–1648.

Heikes BG, Chang W, Pilson MEQ, Swift E, Singh HB, Guenther A, et al. (2002). Atmospheric methanol budget and ocean implication. Global Biogeochem. Cyc 16:80–1–80–13.

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5:1571–1579.

Holmes B. (2006). The Genera Flavobacterium, Sphingobacterium and Weeksella. In:The

Prokaryotes. Springer New York: New York, NY, pp. 539–548.

Howard EC, Sun S, Reisch CR, del Valle DA, Burgmann H, Kiene RP, et al. (2011). Changes in Dimethylsulfoniopropionate Demethylase Gene Assemblages in Response to an Induced Phytoplankton Bloom. AEM 77:524–531.

Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol 12:1889–1898.

Hutchinson GE. (1961). The paradox of the plankton. The Am Nat 95:137–145.

Jiao N, Herndl GJ, Hansell DA, Benner R, Kattner G, Wilhelm SW, et al. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nat Rev Micro 8:593–599.

Khodse VB, Bhosle NB. (2010). Differences in carbohydrate profiles in batch culture grown planktonic and biofilm cells of Amphora rostrataWm. Sm. Biofouling 26:527–537.

Kirchman DL. (2012). Processes in Microbial Ecology. Oxford University Press.

Kirchman DL, Yu L, Fuchs B, & Amann R. (2001). Structure of bacterial communities in aquatic systems as revealed by filter PCR. Aquat Microb Ecol, 26 (1), 13–22.
Kirchman DL. (2002). The ecology of Cytophaga–Flavobacteria in aquatic environments. FEMS Microbiol Ecol 39:91–100.

Kirchman DL, Mitchell R. (2008). Microbial Ecology of the Oceans (Wiley Series in Ecological and Applied Microbiology). 2nd ed. Wiley-Liss.

Konstantinidis KT, Tiedje JM. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microb 10:504–509.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12:118–123.

Ladau J, Sharpton TJ, Finucane MM, Jospin G, Kembel SW, Dwyer JOA, et al. (2013). Global marine bacterial diversity peaks at high latitudes in winter. ISME J 1–9.

Lau E, Fisher MC, Steudler PA, Cavanaugh CM. (2013). The Methanol Dehydrogenase Gene, mxaF, as a Functional and Phylogenetic Marker for Proteobacterial Methanotrophs in Natural Environments. PLoS ONE 8:e56993.

Lennon JT, Jones SE. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Micro 9:119–130.

Levitus S, Antonov JI, Boyer TP, Locarnini RA, Garcia HE, Mishonov AV. (2009). Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. Geophys. Res. Lett. 36:1-5.

Li H, Zhang Q, Wang X-L, Ma X-Y, Lin K-F, Liu Y-D, et al. (2012). Biodegradation of

benzene homologues in contaminated sediment of the East China Sea. Bioresource Technology 124:129–136.

Li WKW. (2006). Coherent Sign Switching in Multiyear Trends of Microbial Plankton. Science 311:1157–1160.

Li WKW. (2009). From cytometry to macroecology: a quarter century quest in microbial oceanography. Aquat. Microb. Ecol. 57:239–251.

Li WKW, Dickie P. (2001). Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. Cytometry 44:236–246.

Li WKW, Harrison WG. (2008). Propagation of an atmospheric climate signal to phytoplankton in a small marine basin. Limnol. Oceangr. 53:1734–1745.

Li WKW, Harrison WG, Head EJH. (2006). Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. Proceedings of the Royal Society B: Biological Sciences 273:1953–1960.

Li WKW, McLaughlin FA, Lovejoy C, Carmack EC. (2009). Smallest Algae Thrive As the Arctic Ocean Freshens. Science 326:539–539.

Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. PNAS U.S.A. 104:11436–11440.

Marshall KT, Morris RM. (2012). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. 7:452–455.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. (2006). Microbial biogeography: putting microorganisms on the map. Nat Rev Micro 4:102–112.

Massana R, Murray AE, Preston CM, DeLong EF. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. AEM 63:50–56.

McCune, B. and M. J. Mefford. 2011. PC-ORD. Multivariate Analysis of Ecological Data. Version 6. MjM Software, Gleneden Beach, Oregon, U.S.A.

Milne PJ, Riemer DD, Zika RG, Brand LE. (1995). Measurement of vertical distribution of isoprene in surface seawater, its chemical fate, and its emission from several phytoplankton monocultures. Mar Chem 48:237–244.

Morris RM, Frazar CD, Carlson CA. (2012). Basin-scale patterns in the abundance of SAR11 subclades, marine Actinobacteria (OM1), members of the Roseobacter clade and OCS116 in the South Atlantic. Environ Microbiol 14:1133–1144.

Morris RM, Longnecker K, Giovannoni SJ. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. Environ Microbiol 8:1361–1370.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810.

Morris RM, Vergin K, Cho J-C, Rappé MS, Carlson C, Giovannoni SJ. (2005). Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnol. Oceangr. 50:1687–1696.

Munn C. (2003). Marine Microbiology. Taylor & Francis.

Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, et al. (2010). Genome characteristics of a generalist marine bacterial lineage. ISME 4:784–798. http://www.nature.com/ismej/journal/v4/n6/full/ismej2009150a.html.

Noinaj, N., Guillier, M., Barnard, T. J., & Buchanan, S. K. (2010). TonB-Dependent Transporters: Regulation, Structure, and Function. Ann Rev Microbiol, 64(1), 43–60.

Orr JC, Fabry VJ, Aumont O, Bopp L, Doney SC, Feely RA, et al. (2005). Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437:681–686.

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? Trends Microb 14:257–263.

Pedrós-Alió C. (2012). The Rare Bacterial Biosphere. Annu. Rev. Marine. Sci. 4:449–466.

Pomeroy LR, Williams PJ, Azam, F, Hobbie JE. (2007). The microbial loop. Oceanography 20:28-33.

Pommier T, Canbäck B, Rienmann L, Bostrom KH, Simu K, Lundberg P, et al. (2007). Global patterns of diversity and community structure in marine bacterioplankton. Molec Ecol16:867–880.

Pontarp M, Sjostedt J, Lundberg P. (2013). Experimentally induced habitat filtering in marine bacterial communities. Mar. Ecol. Prog. Ser. 477:77–U406.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, et al. (2007). The role of ecological theory in microbial ecology. Nat Rev Micro 5:384–392.

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418:630–633.

Rappé MS, Giovannoni SJ. (2003). The uncultured microbial majority. Annu. Rev. Microbiol. 57:369–394.

Rappé MS, Kemp PF, Giovannoni SJ. (1997). Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol. Oceangr. 42:811–826.

Reeder J, Knight R. (2009). The 'rare biosphere': a reality check. Nature Meth 6:636–637.

Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. (2011). Bacterial community transcription patterns during a marine phytoplankton bloom. Environ Microbiol 14:228–239.

Sabehi G, Béjà O, Suzuki MT, Preston CM, DeLong EF. (2004). Different SAR86 subgroups harbour divergent proteorhodopsins. Environ Microbiol 6:903–910.

Saitou N, Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molec Biol and Evol 4:406–425.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. AEM 75:7537–7541.

Shade A, Caporaso JG, Handelsman J, Knight R, Fierer N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. ISME J 7:1493–1506.

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, et al. (2012). Fundamentals of microbial community resistance and resilience. Front Microbiol 3:1-19.

Shi Y, Tyson GW, Eppley JM, DeLong EF. (2010). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J 5:999–1013.

Sjostedt J, Koch-Schmidt P, Pontarp M, Canbäck B, Tunlid A, Lundberg P, et al. (2012). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. AEM 78:1361–1369.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". PNAS U.S.A. 103:12115–12120.

Stahl DA, la Torre de JR. (2012). Physiology and Diversity of Ammonia-Oxidizing Archaea. Annu. Rev. Microbiol. 66:83–101.

Stevens H, Ulloa O. (2008). Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. Environ Microbiol 10:1244–1259.

Sunamura M, Higashi Y, Miyako C, Ishibashi J-I, Maruyama A. (2004). Two bacteria phylotypes are predominant in the Suiyo seamount hydrothermal plume. AEM 70:1190–1198.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, et al. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. Science 333:1296–1300.

Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. (2012). Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. Science 336:608–611.

Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, et al. (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. ISME J 3:1148–1163.

Tripp HJ. (2013). The unique metabolism of SAR11 aquatic bacteria. J Microbiol. 51:147–153.

Urios L, Agogué H, Lesongeur F, Stackebrandt E, Lebaron P. (2006). Balneola vulgaris gen. nov., sp. nov., a member of the phylum Bacteroidetes from the north-western Mediterranean Sea. Int. J. Syst. Evol. Microbiol. 56:1883–1887.

Urios L, Intertaglia L, Lesongeur F, Lebaron P. (2008). Balneola alkaliphila sp. nov., a marine bacterium isolated from the Mediterranean Sea. Int. J. Syst. Evol. Microbiol. 58:1288–1291.

Wallenstein MD, Hall EK. (2011). A trait-based framework for predicting when and where microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry 109:35–47.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, et al. (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. Science 326:578–582.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. AEM 73:5261–5267.

Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: the unseen majority. PNAS U.S.A. 95:6578–6583.

Woes CR. (1987). Bacterial evolution. Microbiology and Molecular Biology Reviews. 51(2):210-225

Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P, et al. (2007). Genome sequence of the cellulolytic gliding bacterium Cytophaga hutchinsonii. AEM 73:3536–3546.

Yergeau E, Lawrence JR, Sanschagrin S, Waiser MJ, Korber DR, Greer CW. (2012). Next-Generation Sequencing of Microbial Communities in the Athabasca River and Its Tributaries in Relation to Oil Sands Mining Activities. AEM 78:7626–7637.

Yue JC, Clayton MK. (2005). A similarity measure based on species proportions.

Commun in Stat-Theor and Meth 34:2123–2131.

Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Environ Microbiol 12:172–191.

Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, et al. (2011). Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLoS ONE 6:e24570.

Figures

Figure 1. Map of Bedford Basin. Geographic location and bathymetry of Bedford Basin, located in Halifax Harbour, Nova Scotia. The red circle marks the location of the Compass Buoy station (44^0 41' 30'' N, 63^0 38' 30''W).

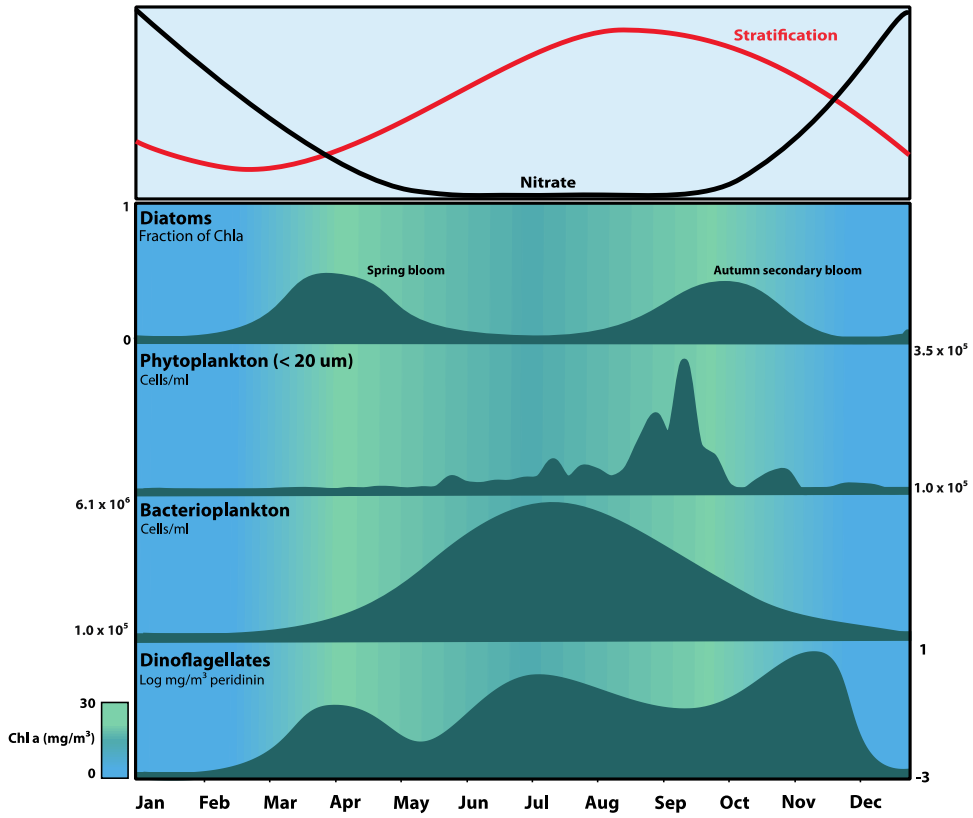


Figure 2. Annual patterns of microbial community change in the surface layer (0-5 m) at Bedford Basin. *Top panel*: chemical environment during the year with nitrate representing nutrient patterns and stratification driven by temperature. Nitrate is highest at times of least stratification as mixing causes upwelling of nitrate. As stratification increases nitrate levels reach almost undetectable levels during the summer and regenerate as stratification degrades again in the autumn. *Bottom panel:* annual biotic cycle in Bedford Basin. Blue to green background gradient indicates levels of Chlorophyll *a* (Chl a) with blue representing lowest level and green representing highest level (in mg/m-3). Peaks in Chl *a* occur during Spring (week 13) and Autumn (week 38). Diatom blooms occur in the Spring (week 13) and secondarily in the Autumn (week 38) calculated as fractions of total Chl *a*. Panel 2 shows peak in phytoplankton counted by flow cytometry (<20um) in cells/ml. This group includes Synechococcus. Panel 3 shows annual bacterial cell production, which peak in the summer months (week 26) counted by flow cytometry (cells/ml). Panel 4 indicated annual patterns of dinoflagellates, which characteristically are found from Spring (week 13) until Autumn (week 38) (measured in log mg/m-3 of peridinin pigment characteristic of dinoflagellates). Figure was made based on data taken from Bedford Institute of Oceanography, courtesy of William K. Li.

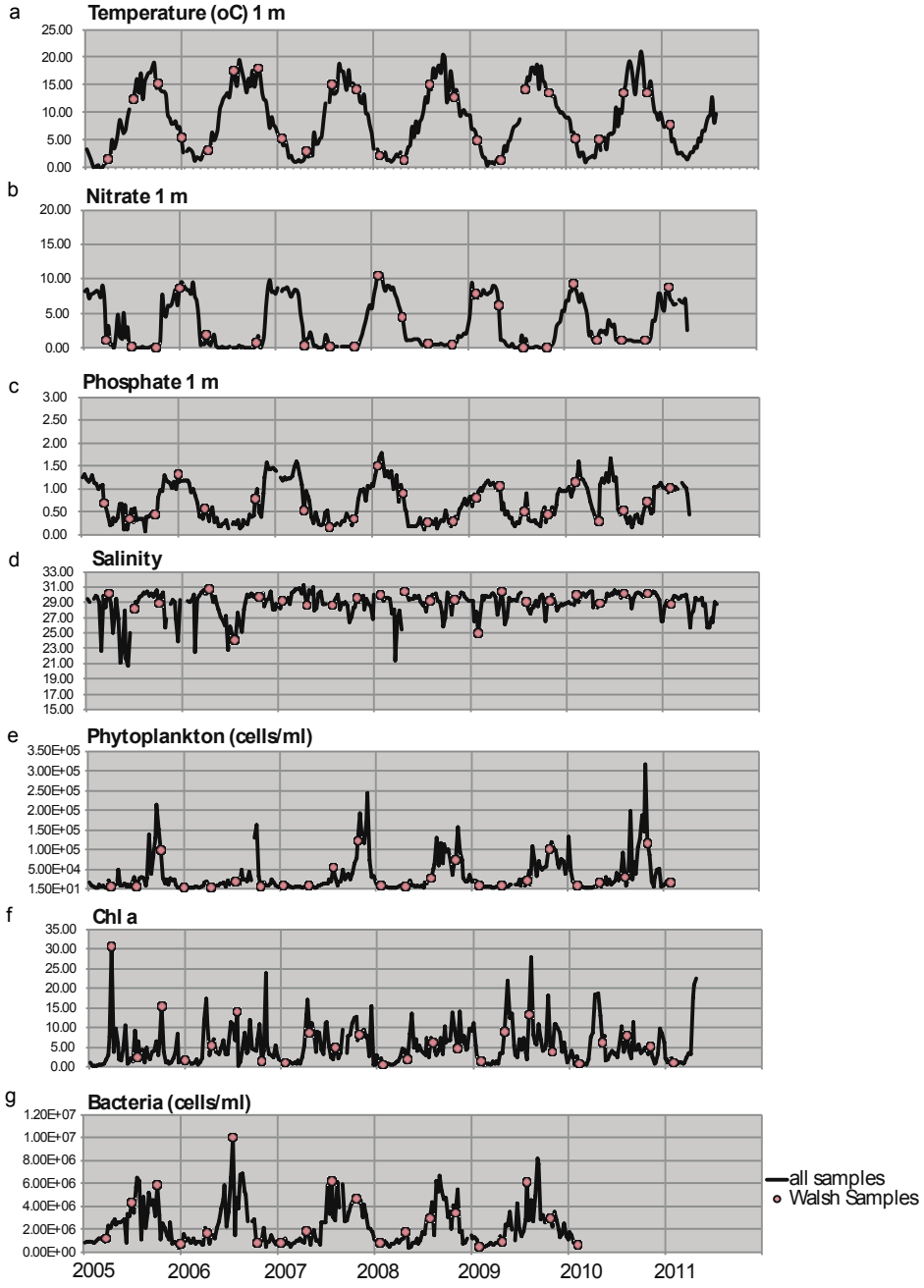


Figure 3. Environmental data from 2005 through 2011. Panel a) shows the temperature patterns with temperature peaking typically in the late summer. Panel b-c show nitrate and phosphate (representing nutrients) which peak typically during the winter. Panel d) shows salinity. Panel e) shows total phytoplankton abundance (enumerated by flow cytometry). Panel e) shows chlorophyll a trends throughout the time series and there is inter-annual variability observed. The circles indicate the samples used for bacterial community composition in relation to the weekly measurements in the

environment.

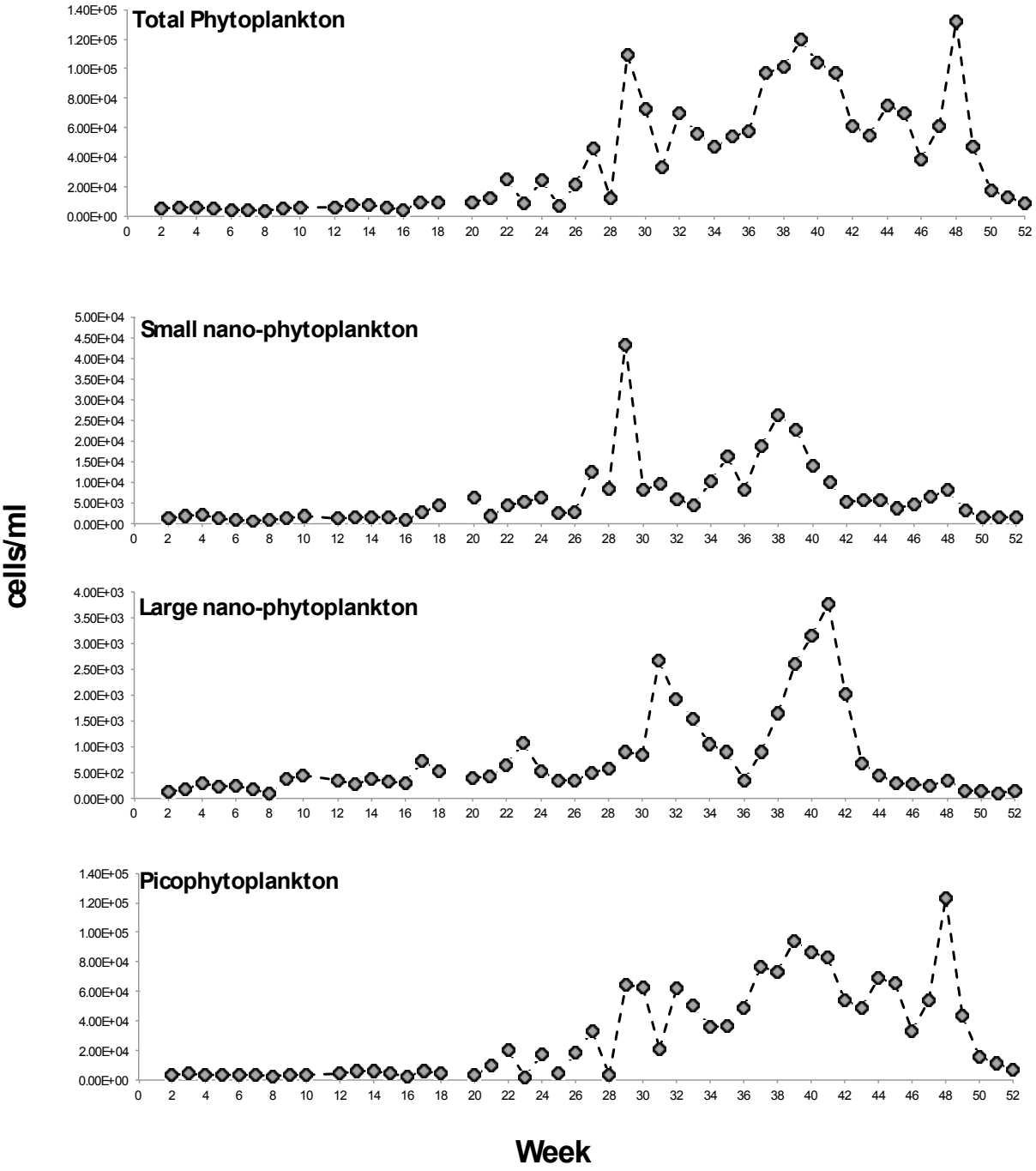


Figure 4. Weekly phytoplankton trends over 2009. Phytoplankton groups were split into different groups based on size as in Li and Dickie, 2001. Panel a) shows the total phytoplankton, b) Small nanophytoplankton (2-10µm), c) Large nanophytoplankton (10-20 µm), and d) Picophytoplankton (<2µm).

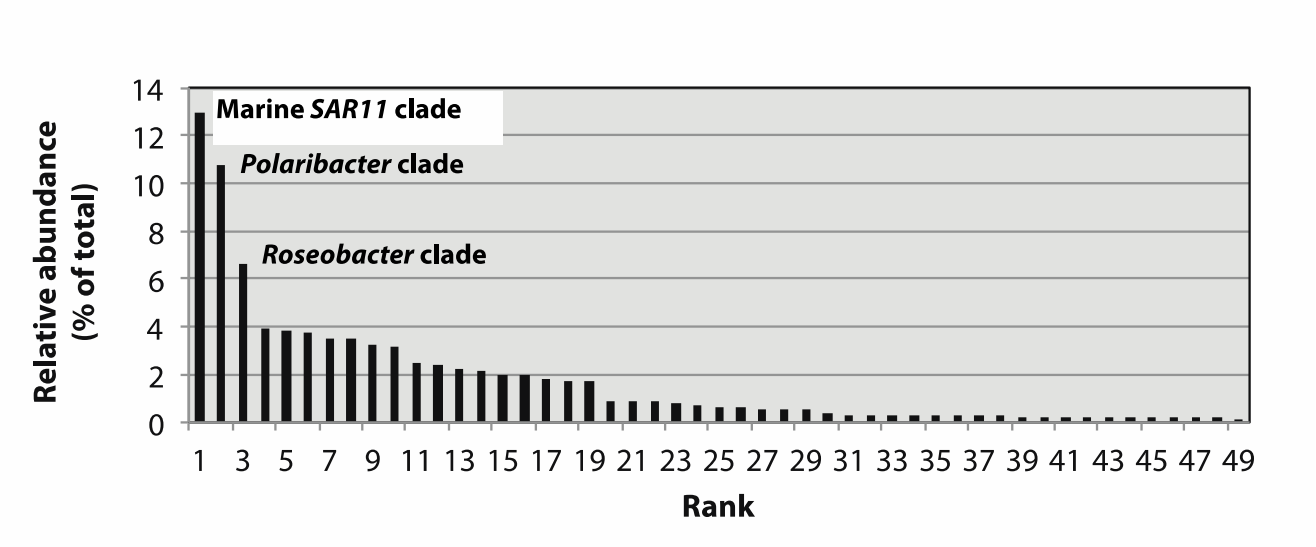


Figure 5. Rank-abundance curve. Rank abundance curve showing that subsampled dataset consists of three abundant clades, marine SAR11, Polaribacter, and Roseobacter clades. Followed by a long list of rare taxa.

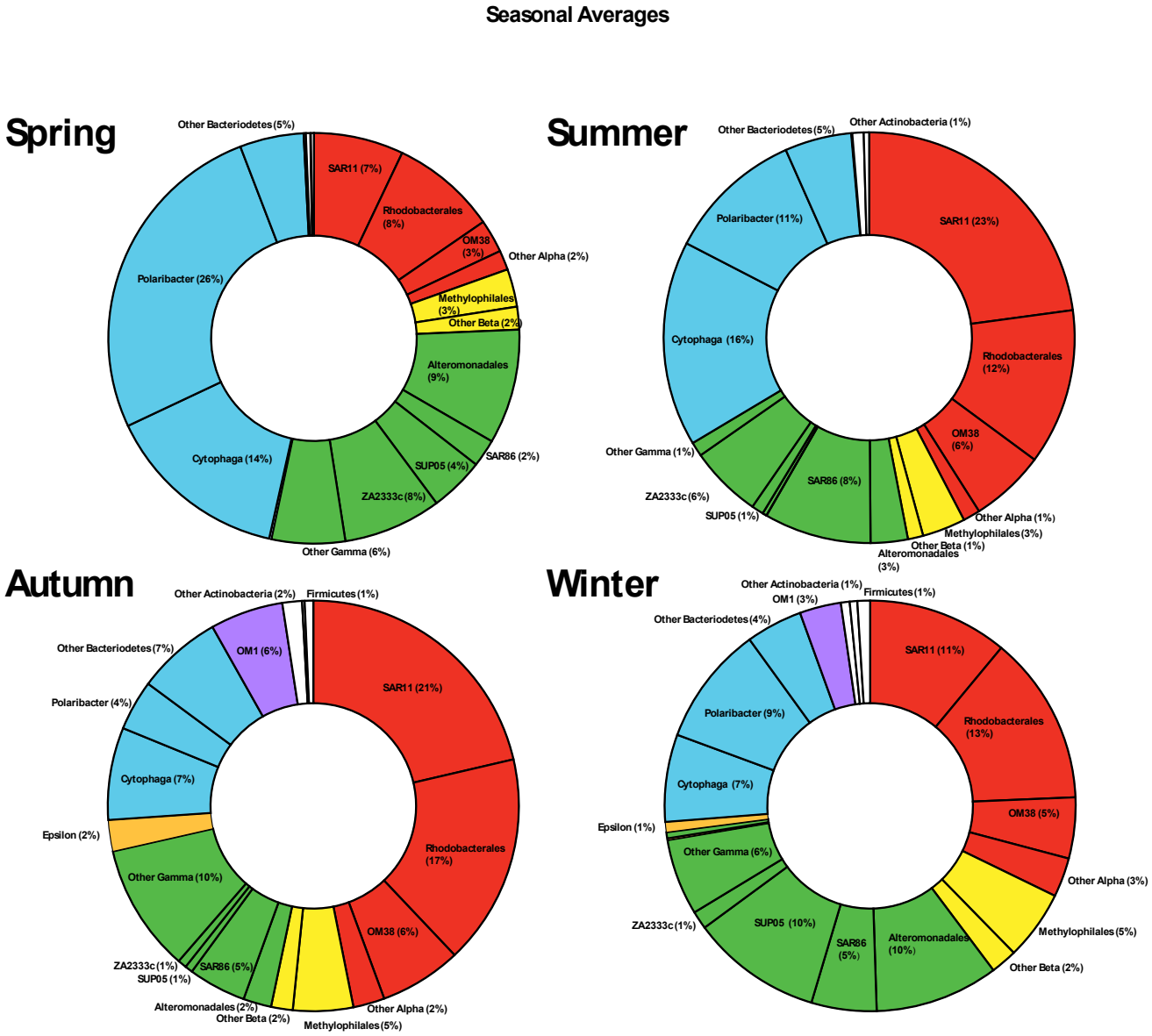


Figure 6. Seasonal average taxonomic composition associated with each season. The Summer bacterial communities comprise of SAR11 (23%) Rhodobacteriales(12%), OM38 (6%), SAR86 (6%) Za2333c (6%) Cytophaga (16%) and Polaribacter (11%). The Spring bacterial communities on average comprise of Polaribacter (26%), SAR11 (11%), Alteromonadales (9%) and Cytophaga (14%). In the Autumn, bacterial communities are comprised of SAR11 (21%), Rhodobacteriales (17%), OM38 (6%), and OM1 (6%). The Winter bacterial communities on average are composed of SAR1 (11%) Rhodobacteriales (13%) Methylophilales (5%), GSO clade (10%), Polaribacter (9%).

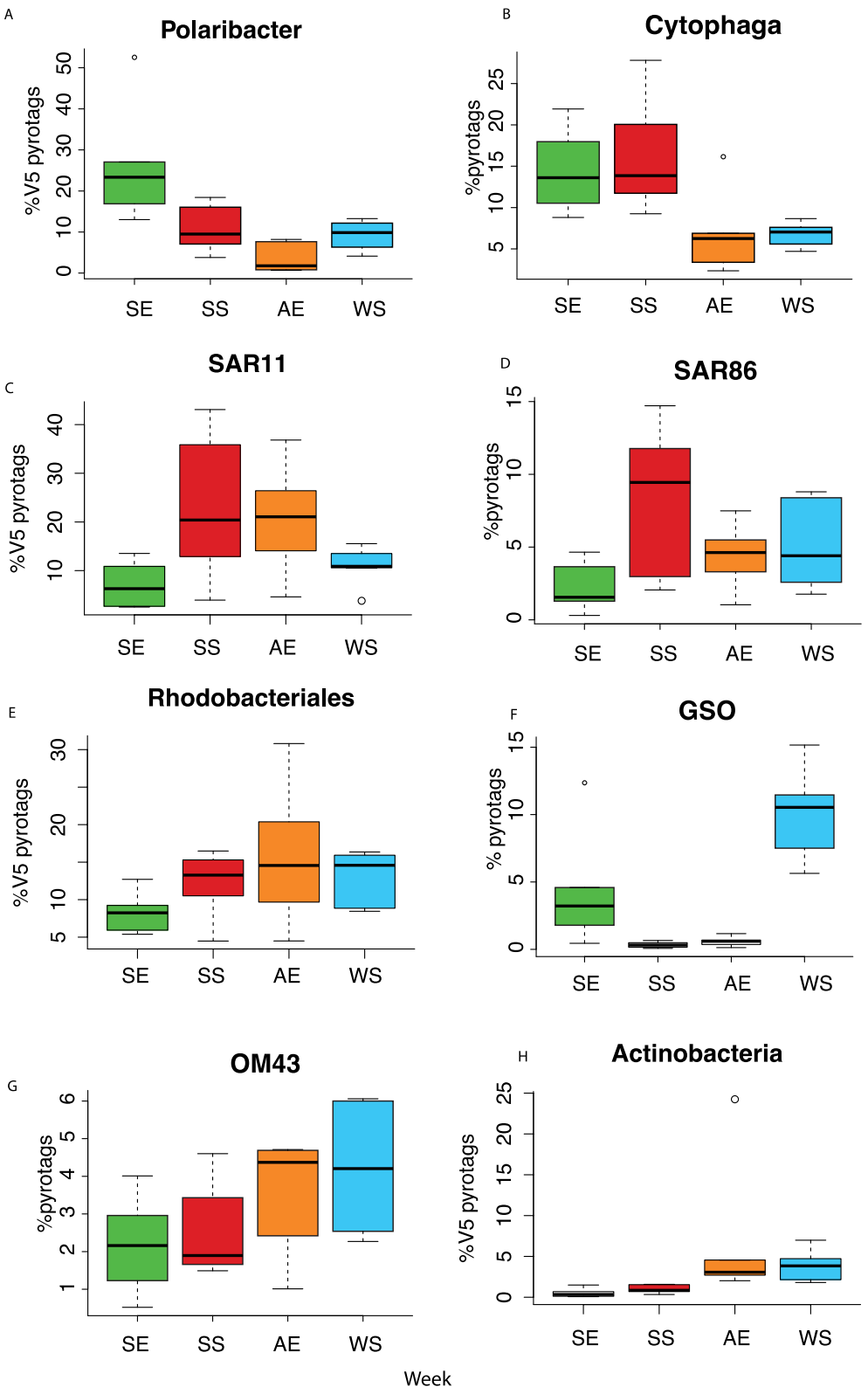


Figure 7. Seasonal trends of select main groups of bacteria. Seasonal trends of six main bacterial taxa (7a Polaribacter 7b Cytophaga, 7c SAR11, 7d SAR86, 7e Rhodobacteriales, 7f GSO, 7g OM43, and 7h Actinobacteria). Lines indicate Mean.

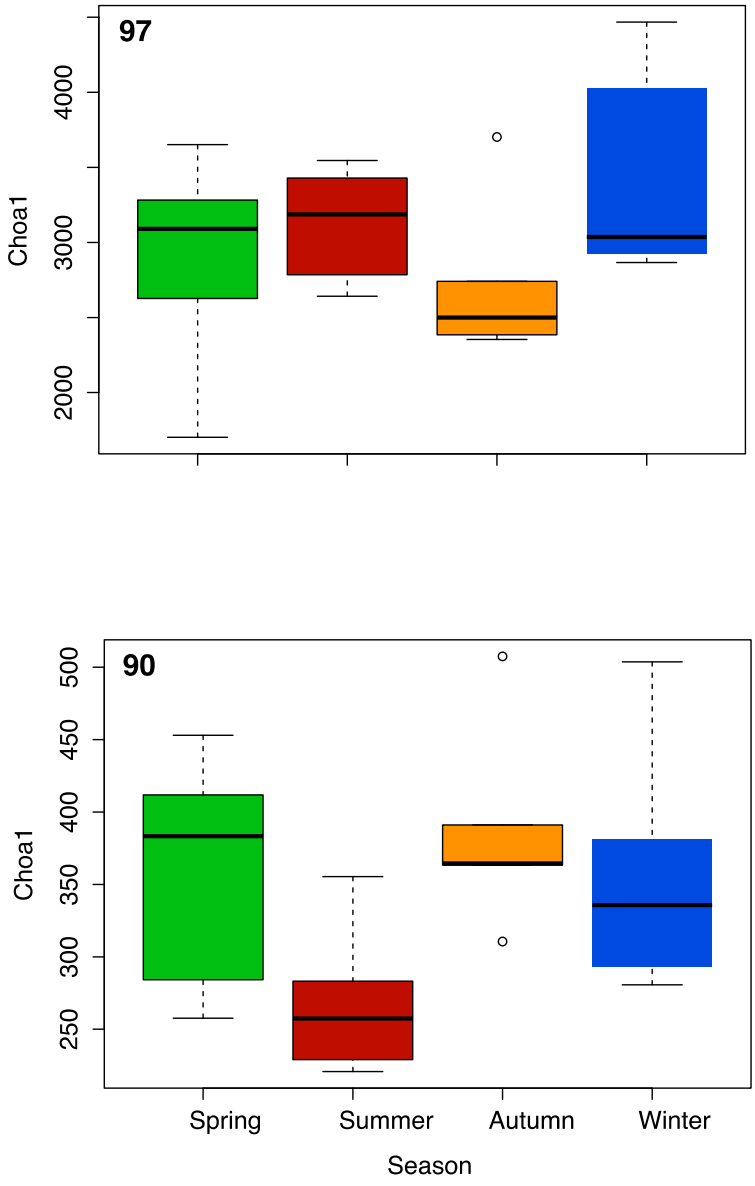


Figure 8. Seasonal α-diversity (Chao1 index) box plots per season. Green represents Spring samples, red represents Summer samples, orange represents Autumn samples, and blue represents Winter samples.(A) Seasonal Chao1 indices at 97% sequence similarity. (B) Seasonal Chao1 indices per season at 90% sequence similarity

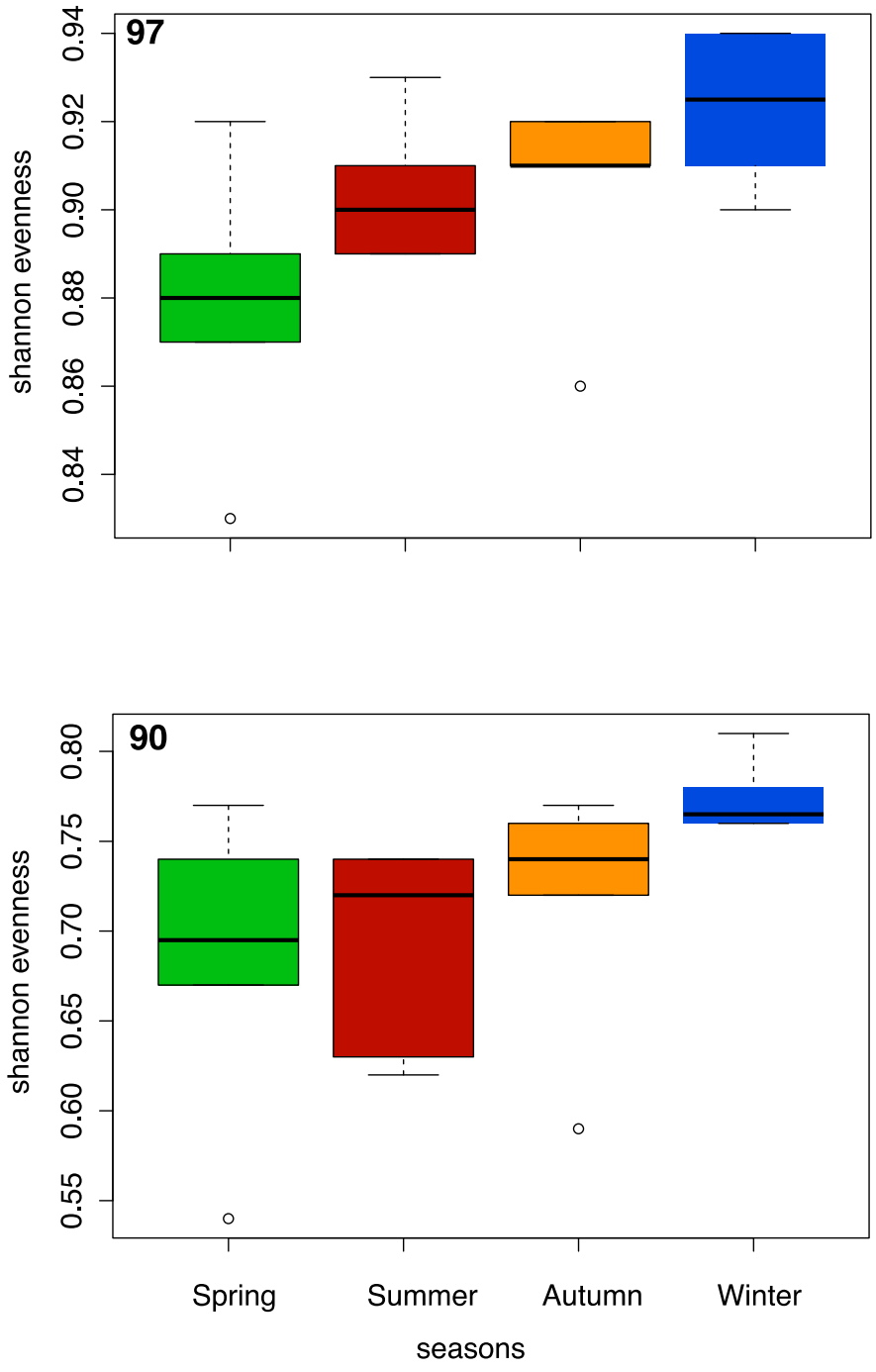


Figure 9. Seasonal Community Evenness (Shannon-evenness index) box plots per season. Green represents Spring samples, red represents Summer samples, orange represents Autumn samples, and blue represents Winter samples. (A) Seasonal Chao1 indices at 97% sequence similarity. (B) Seasonal Chao1 indices per season at 90% sequence similarity.

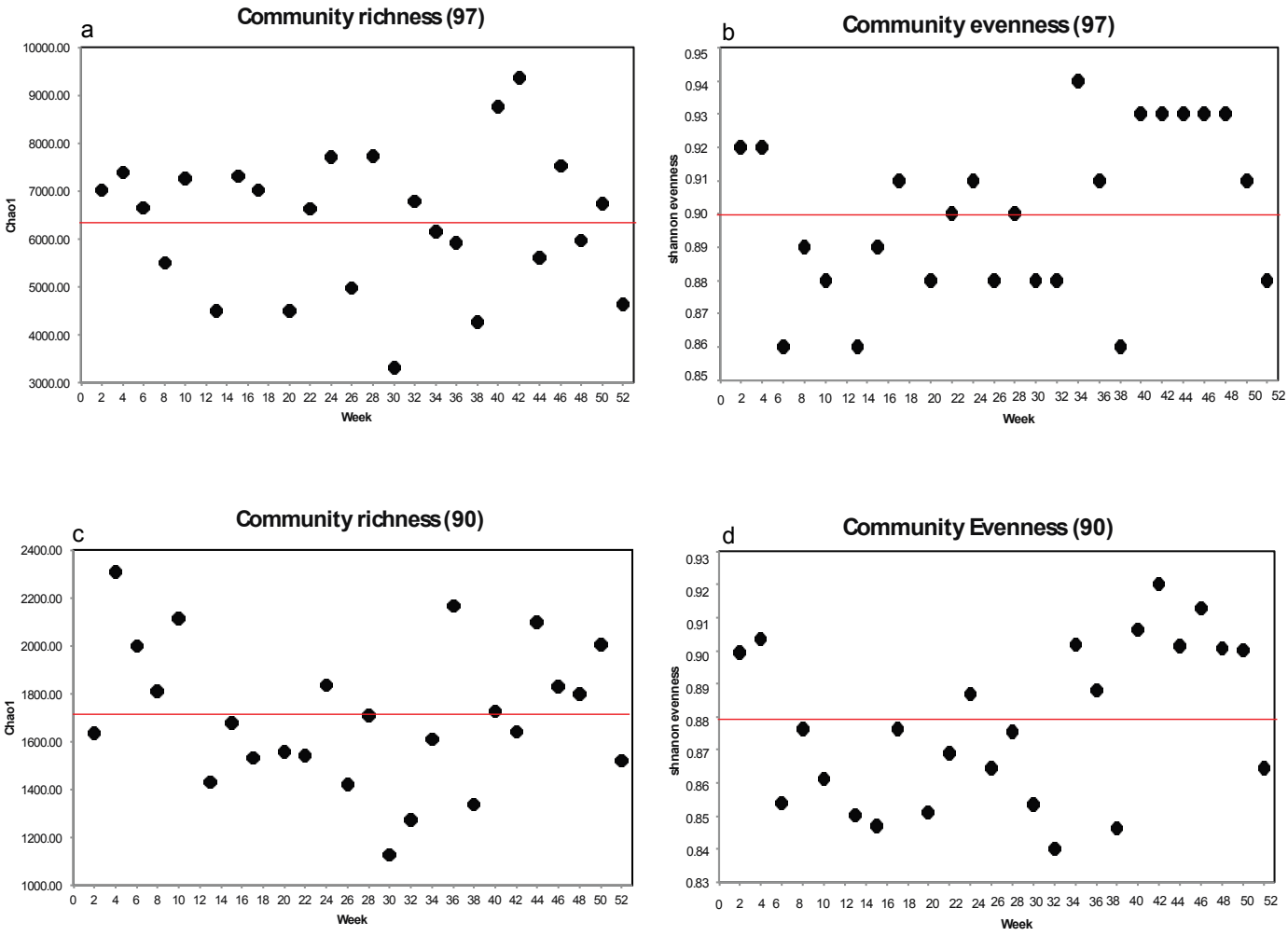


Figure 10. Community Richness and Evenness over the single (2009) year. (A) Richness (Choa1) estimates every two weeks over one year at 97% (B) Evenness (shannon-evenness) estimates every two weeks over one year at 97%. (C) Richness (Choa1) estimates every two weeks over one year 90% sequence similarity (D) Evenness (shannon-evenness) estimates every two weeks over one year at 90%.

90

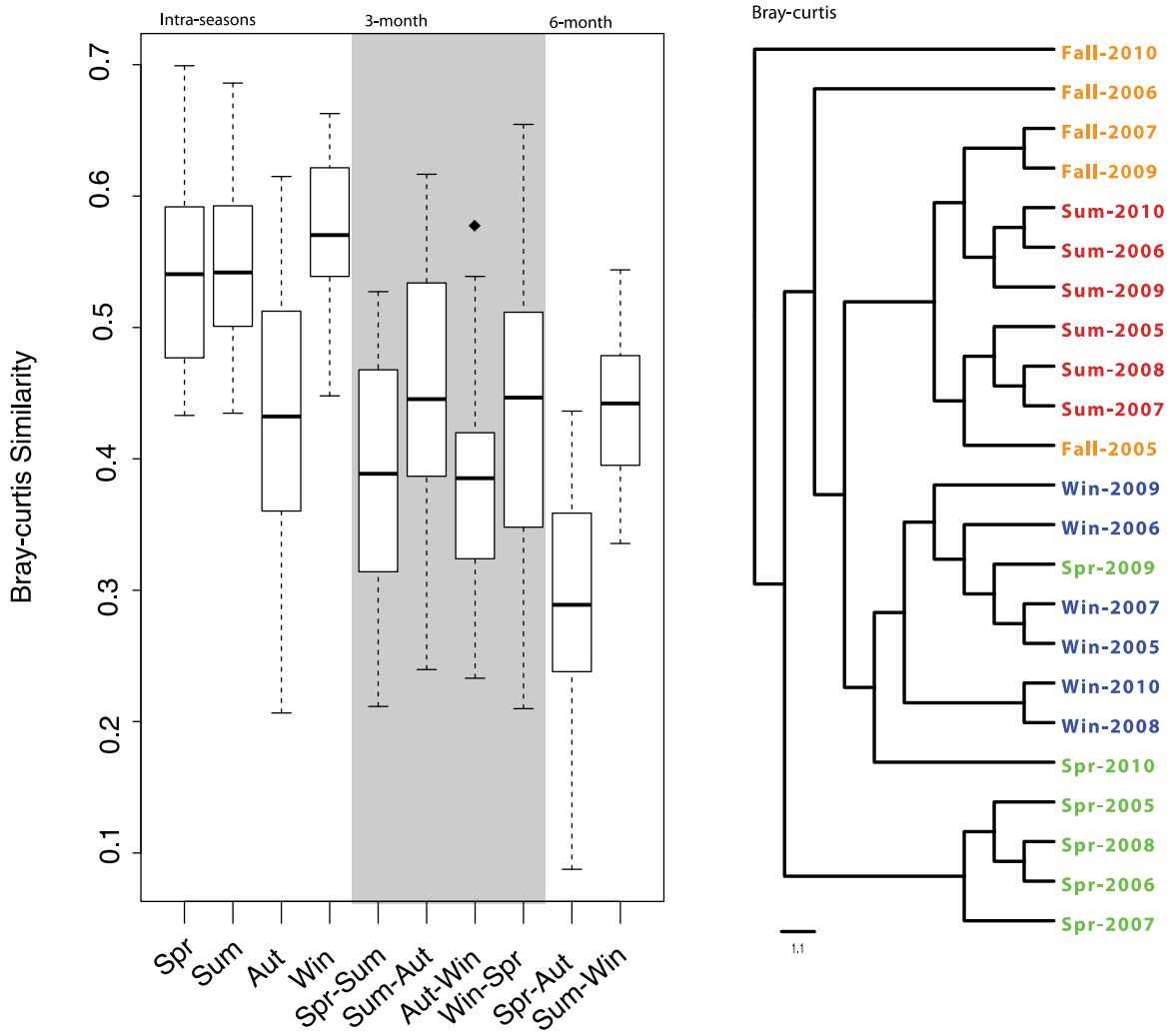


Figure 11. Time-decay box plots of Seasonal β-diversity (Bray-Curtis 90%). (A) Bray-curtis similarity index showing intra-season similarity inter-seasonally in first four box plots. In gray box, similarity values between seasons that are 3 months apart, and the last two box plots show similarity values between seasons that are 6 months apart. (B) UPGMA dendograms of samples clustered together based on Bray-Curtis values indicate a partitioning of samples between cold-nutrient rich samples and warm nutrient poor samples.

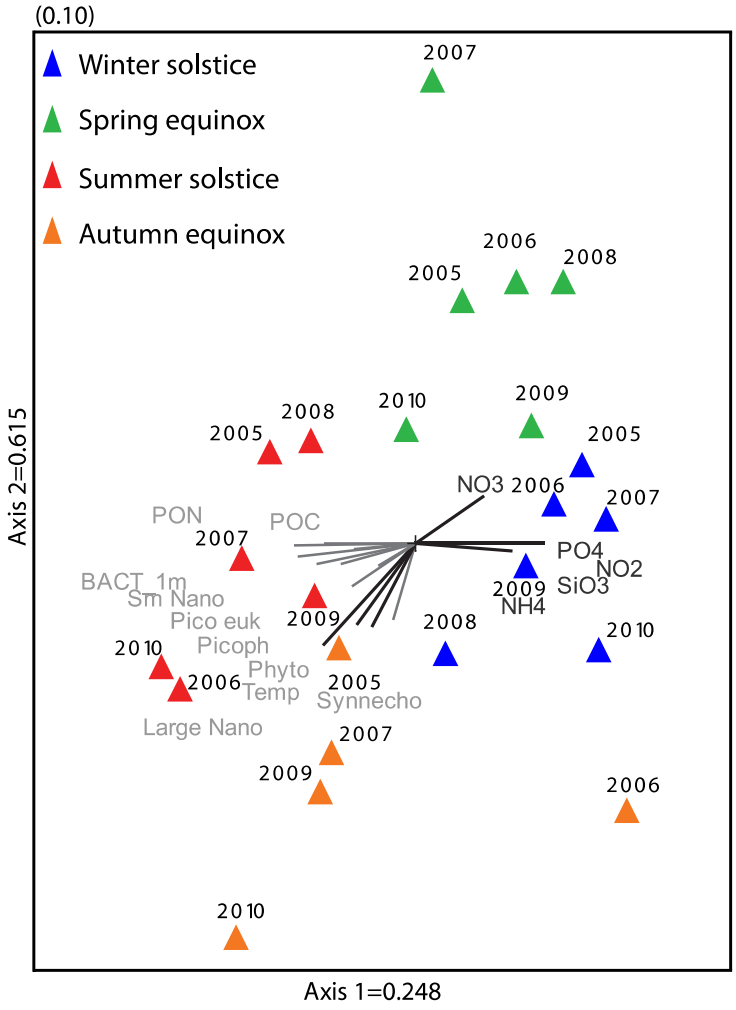


Figure 12. NMS ordination based on Bray-Curtis similarities calculated using relative abundance of OTUs clustered at 90% sequences similarity (2D stress: 36.5) Colors indicate seasonal samples: blue indicates winter solstice samples (Week 52); green indicate spring equinox samples (Week 13); red indicates summer solstice (week 26), Orange indicates Autumn equinox samples. Numbers indicate years from which each sample was taken. Vectors point to direction of samples in which environmental factors relate. X-axis (r=0.248) divides high temperature, high bacterial and phytoplankton production (left) water from nutrient-rich cold-water samples (right).

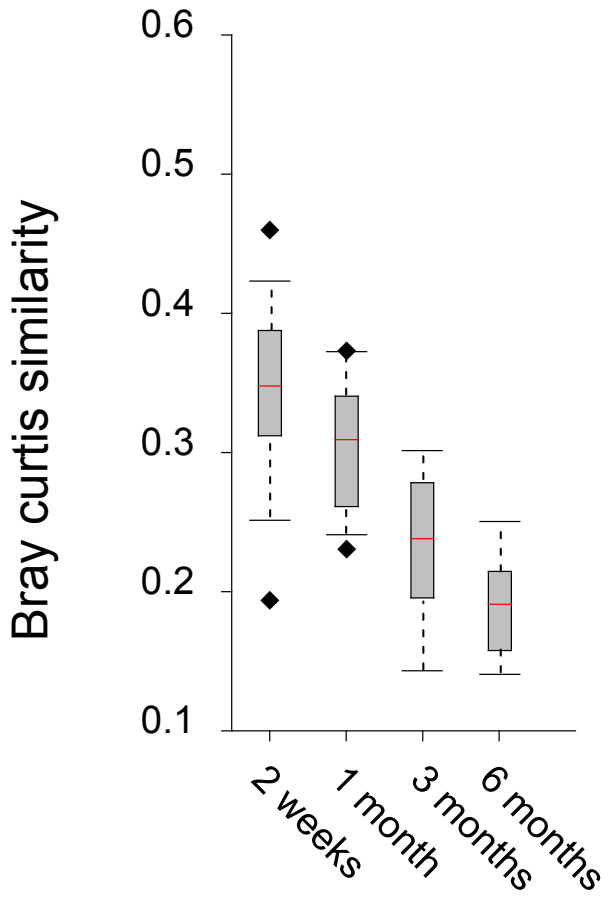


Figure 13. Time-decay box plots of different time scales within one year based on similarity values (Bray-Curtis 90%). Red lines indicate mean.

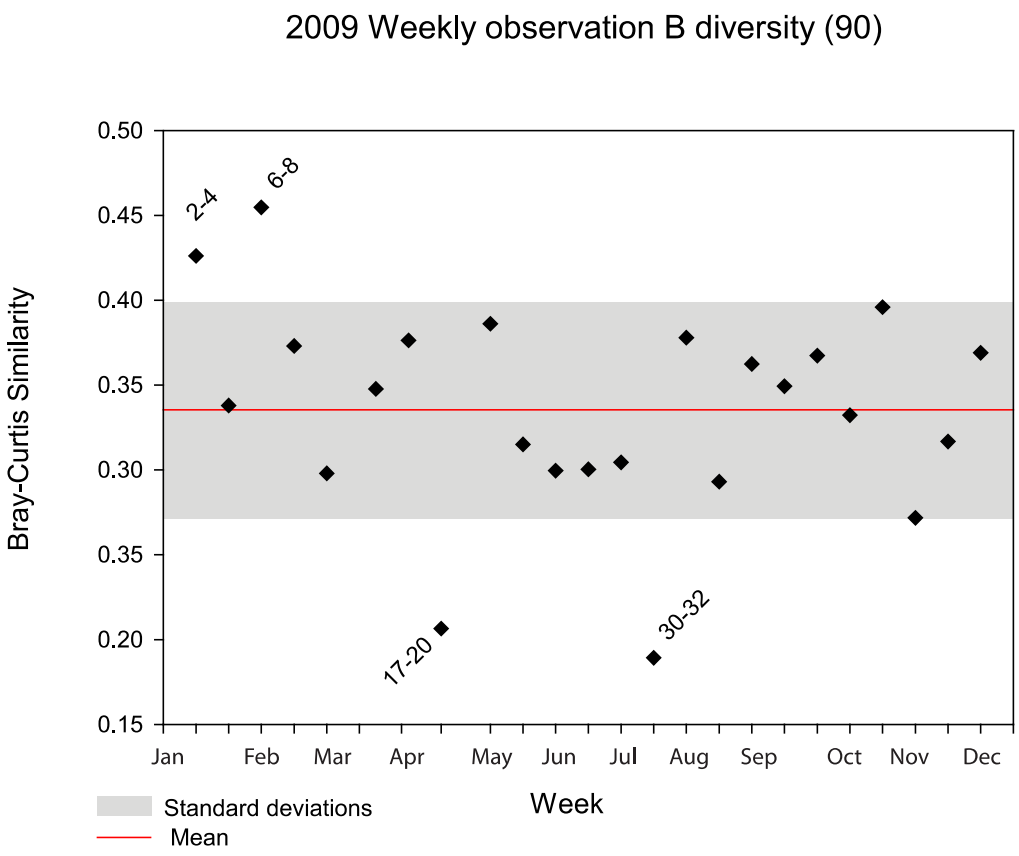


Figure 14. Similarity values between each two weeks that occurred in 2009. The Red line indicates the mean and the grey box shows ± standard deviation. Highest similarity observed in Winter samples (2-4 and 6-8). Lowest similarity values occurred between spring and summer (17-20) and in the summer (30-32).

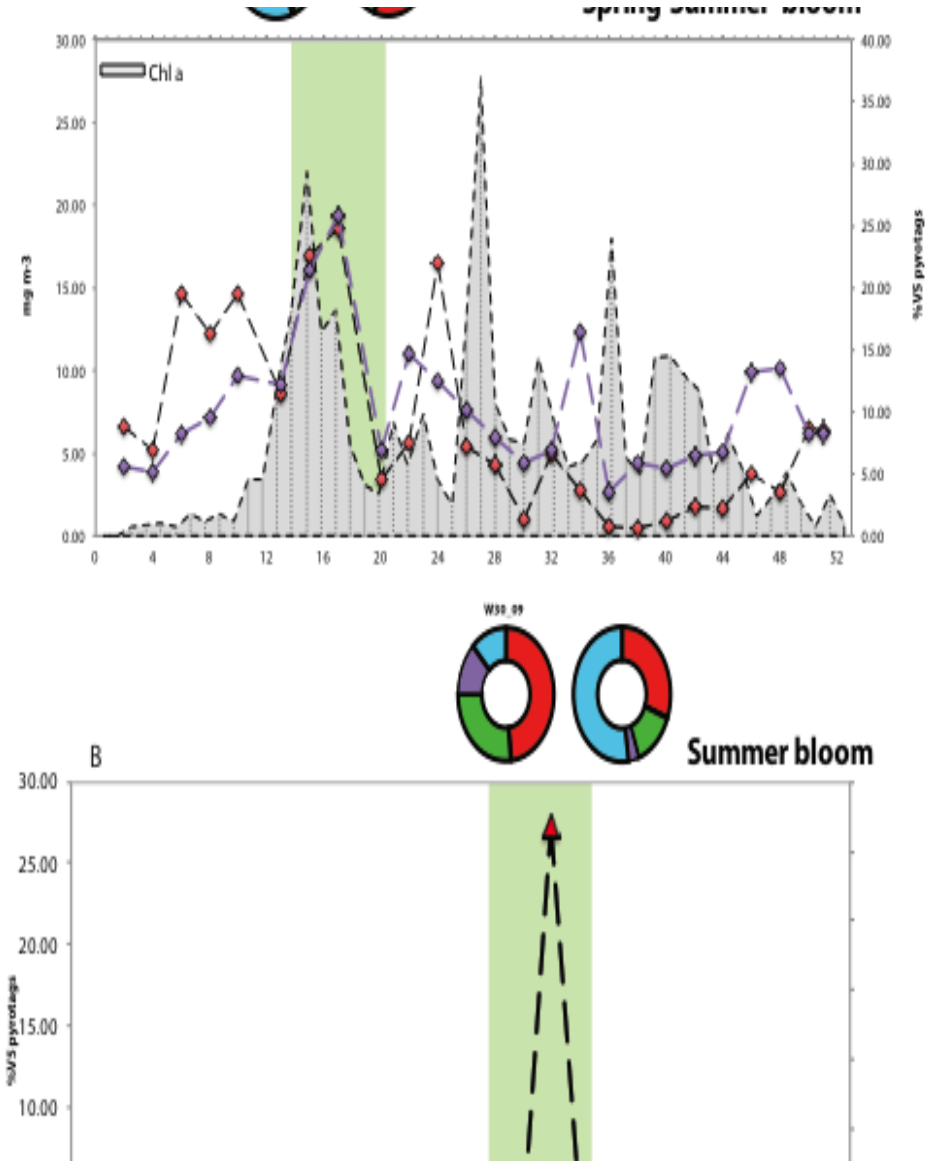


Figure 15. Taxonomic and environmental trends that occurred week 17-20 and 30-32. The first shift **(A)** from post-diatom bloom associated *Bacteroidetes* (Blue, 55%; 31% *Cytophaga* and 24 % *Polaribacter*) in week 17 to summer associated *Alpha-proteobacteria* (Red, 42% *SAR11*) in week 20, as observed in the donut plots, demonstrated the seasonal shift that occurs between the spring and summer seasons each year. The shaded grey area shows the chlorophyll trends in 2009, and the green box indicates week 17-20. The *Bacteroidetes* characterizing this shift are Polaribacter (red diamonds) and Cytophaga (purple diamonds) is a decrease in both these phylotypes from 20% of the sequences in week 17 to 5% seen in week 20 and a corresponding increase in SAR11. Another rapid shift in bacterial community composition was observed between weeks 30-32 (0.80) outlined in **(B)**. During these two weeks there was another shift in bacterial community composition in the summer. Week 30 was characterized by 48% *Alpha-proteobacteria* (20% SAR11, 15% Rhodobacteriales, and 13% OM38) and 12% *Bacteroidetes*, and shifted in week 32 to

30% Alpha-proteobacteria and 52% Bacteroidetes (characterized by 27% uncultivated KSA1, 16% *cytophaga* and 6% *Polaribacter*) as seen in the donut plots. The red triangles show the uncultivated Bacteroidetes bloom relative abundance over the 2009 year, with the bloom that occurred at week 32.

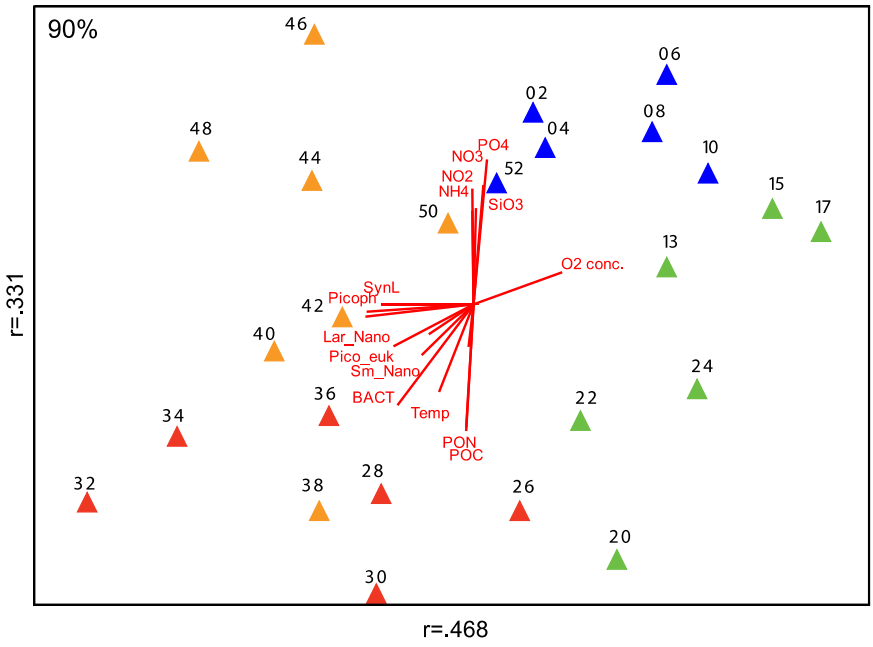


Figure 16. NMS ordination on 2009 biweekly samples based on Bray-Curtis similarities calculated using relative abundance of OTUs clustered at 90% sequences similarity (2D stress: 36.5) Colors indicate seasonal samples: blue indicates winter samples; green indicate spring samples; red indicates summer samples; Orange indicates Autumn samples. Numbers indicate the week from which the sample was taken from in 2009. Vectors point to direction of samples in which environmental factors relate. Y-axis (r=0.331) divide high temperature, high bacterial and phytoplankton production water (right) from nutrient-rich cold water samples (left).

Tables

Table 1: OTUs assigned to each sample in the seasonal (2005-2010) in full dataset, subsampled dataset, after removal of sequences observed once, and once removed, subsampled. OTUS below calculated at different clustering distances.

				Sequences assigned to samples			Subsampled			Rare-removed (n=1)						Rare-sampled (n=1)		
Week	*New label*	*Sample*	*Clones*	*Unique*	*97*	*90*	*Unique*	*OTU (97)*	*OTUs (90)*	*Unique total reads*	*Unique*	*Total reads (97)*	*OTUs (97)*	*Total reads (90)*	*OTUs (90)*	*Unique*	*OTUs (97)*	*OTUs (90)*
38	W38_05	Fall-2005	1585	1358	830	211	556	383	120	437	212	1094	339	1489	141	102	144	73
38	W38_06	Fall-2006	1740	1531	1007	261	554	409	139	358	166	1076	343	1585	135	90	146	77
38	W38_07	Fall-2007	1406	1241	808	188	568	391	104	335	178	940	342	1318	117	101	156	67
38	W38_09	Fall-2009	1030	935	612	153	558	364	104	246	154	689	271	970	102	100	137	73
38	W38_10	Fall-2010	1002	840	539	145	530	347	108	279	120	690	227	940	93	74	122	64
13	W13_05	Spr-2005	1482	1266	851	181	529	377	99	403	191	999	83	1389	115	101	137	63
13	W13_06	Spr-2006	1629	1341	822	158	532	348	80	474	191	1160	368	1546	90	86	127	46
13	W13_07	Spr-2007	1336	1068	636	106	511	314	65	402	146	954	353	1278	59	77	109	34
13	W13_08	Spr-2008	1419	1171	750	173	531	352	93	434	189	988	254	1336	99	90	123	64
13	W13_09	Spr-2009	1481	1258	778	157	538	346	82	438	218	1054	352	1408	105	106	140	53
13	W13_10	Spr-2010	1397	1232	805	170	556	381	97	339	175	903	311	1314	108	101	136	61
26	W26_05	Sum-2005	1266	1097	710	160	547	376	111	348	184	857	301	1191	108	106	136	62
26	W26_06	Sum-2006	605	551	379	90	551	379	90	**152**	98	**388**	162	**571**	64	98	132	56
26	W26_07	Sum-2007	1478	1235	747	146	537	337	76	443	205	1064	333	1407	85	103	134	50
26	W26_08	Sum-2008	1499	1309	818	180	541	361	91	393	207	1019	338	1408	107	98	141	50
26	W26_09	Sum-2009	843	791	572	123	569	424	99	180	129	515	244	787	80	110	151	60
26	W26_10	Sum-2010	1427	1211	783	148	543	378	80	364	157	984	340	1355	99	88	139	54
51	W51_05	Win-2005	970	827	600	174	535	397	120	293	152	645	275	892	110	95	147	76
51	W51_06	Win-2006	3461	2865	1774	408	561	391	136	923	347	2335	648	3223	221	104	148	82
51	W51_07	Win-2007	1062	950	670	196	557	405	136	254	144	671	279	973	126	104	152	84
51	W51_08	Win-2008	1932	1751	1153	237	567	393	109	387	214	1254	476	1818	154	104	159	77
51	W51_10	Win-2010	1501	1361	915	262	576	403	134	333	175	963	334	1364	141	109	153	87
52	W52_09	win-2009	1261	1151	784	216	578	416	141	281	199	811	149	1163	210	98	155	90
Total			**32810**	**24314**	**12937**	**2196**	**10938**	**6283**	**1160**									

Table 2: OTUs assigned to each sample in the biweekly (2009) in full dataset, subsampled dataset, after removal of sequences observed once, and once removed, subsampled. OTUS below calculated at different clustering distances.

			Sequences assigned to samples			**Subsampled**			**Rare-removed (n=1)**			**Rare-subsampled (n=1)**		
Week	*Sample*	*Clones*	*Unique*	*97*	*90*	*Unique*	*OTU (97)*	*OTUs (90)*	*Unique*	*OTUS (97)*	*OTUs (90)*	*Unique*	*OTUs (97)*	*OTUs (90)*
2	W02_09	6473	3553	3383	1902	803	784	583	287	1164	1413	131	328	326
4	W04_09	6687	3764	3592	2030	799	780	613	295	1163	1464	147	344	481
6	W06_09	8397	4128	3954	2160	745	733	570	300	1256	1518	128	300	504
8	W08_09	6646	3376	3220	1804	754	742	566	271	1134	1359	118	331	449
10	W10_09	7136	3668	3534	2025	731	718	557	245	1060	1445	113	307	463
13	W13_09	5325	2140	2008	1223	590	563	425	261	884	970	115	267	460
15	W15_09	7868	3866	3697	1985	715	704	531	274	1139	1406	120	291	386
17	W17_09	6671	3634	3429	1839	787	769	557	250	1082	1271	100	329	448
20	W20_09	4631	2133	1991	1099	670	637	444	238	801	879	131	304	458
22	W22_09	6915	3759	3531	1830	820	790	573	282	1160	1331	134	329	393
24	W24_09	4397	2445	2331	1357	771	754	551	219	821	1048	131	324	447
26	W26_09	3770	1882	1787	1072	722	693	509	232	739	826	126	312	455
28	W28_09	8187	4286	4034	2017	795	771	574	330	1244	1393	152	331	415
30	W30_09	3405	1607	1487	896	691	649	478	229	637	724	147	284	464
32	W32_09	6996	3615	3306	1646	752	720	506	273	963	1136	130	282	385
34	W34_09	6914	3957	3716	1902	810	784	584	283	1132	1307	145	351	405
36	W36-09	5727	3121	2961	1682	786	767	581	255	1013	1206	128	319	473
38	W38_09	5963	2449	2264	1264	669	627	443	313	925	976	160	270	482
40	W40_09	8615	4670	4365	2289	839	808	603	384	1402	1600	163	350	371
42	W42_09	5060	3048	2927	1713	852	830	631	260	1006	1296	153	359	493
44	W44_09	6966	3828	3607	2017	801	780	600	304	1206	1472	153	344	559
46	W46_09	2668	1791	1722	1184	877	847	655	143	554	902	106	354	482
48	W48_09	1224	843	818	615	843	818	615	102	368	507	102	368	561
50	W50_09	10044	5376	5046	2601	799	781	617	396	1673	1846	161	350	507
52	W52_09	5185	2404	2244	1385	682	654	504	300	954	1090	135	326	520
	Total	**151870**	**75418**	**50943**	**14116**	**17421**	**14273**	**5730**						

Table 3: Chemical Environmental measurements for Seasonal dataset (2005-2010) In analyses missing data was corrected with average measurement for that week over 20 years.

	Temperature (° C)	*Salinity (psu)*	*Photosyntheticall y active radiation (uE/m2/s)*	*Fluorescence mg/m3*	*Oxygen sat (%)*	*Oxygen conc. (mL L-1)*	*Seawater density (kg m-3)*	*Turner chlorophyll (mg m-3)*	*Nitrate + Nitrite (mmol m-3)*	*Silicate (mmol m-3)*	*Phosphate (mmol m-3)*	*Nitrite (mmol m-3)*	*Ammonium (mmol m-3)*
W13_05	1.47	30.12	88.74	n/a	0.84	6.74	24.10	30.70	1.07	1.89	0.70	0.11	0.43
W13_06	3.01	30.67	889.14	n/a	0.84	6.47	24.43	5.37	1.78	1.35	0.58	0.12	1.06
W13_07	2.90	28.62	75.60	n/a	1.04	8.10	22.81	8.72	0.27	1.27	0.53	0.10	0.70
W13_08	1.28	30.42	79.18	6.93	0.99	7.98	24.35	1.93	4.41	5.66	0.90	0.24	1.83
W13_09	1.39	30.43	115.17	7.25	1.04	8.35	24.35	9.04	6.12	5.15	1.06	0.31	0.90
W13_10	4.93	28.90	39.34	14.18	1.18	8.79	22.85	6.14	1.11	0.45	0.31	0.14	1.65
W26_05	12.41	28.26	904.97	n/a	0.76	4.74	21.29	2.48	0.09	1.27	0.35	0.09	0.52
W26_06	17.50	24.00	133.74	n/a	0.86	4.98	16.98	14.12	0.00	0.00	0.00	0.00	0.00
W26_07	15.05	28.63	772.07	n/a	1.44	8.55	21.05	4.90	0.08	2.05	0.17	0.07	0.30
W26_08	15.14	29.15	157.01	9.56	1.25	7.39	21.44	6.31	0.54	0.29	0.27	0.12	1.25
W26_09	14.12	29.06	93.84	24.34	1.19	7.16	21.58	13.22	0.02	1.76	0.52	0.16	1.11
W26_10	13.61	30.19	185.25	21.56	1.32	7.97	22.55	7.85	1.07	2.15	0.53	0.13	0.54
W38_05	15.23	28.94	63.33	n/a	0.75	4.43	21.26	15.29	0.02	0.75	0.45	0.07	0.53
W38_06	17.91	29.79	21.27	n/a	0.77	4.28	21.31	1.49	0.71	2.27	0.79	0.16	5.01
W38_07	14.04	29.60	355.35	n/a	1.02	6.14	22.01	8.02	0.13	0.52	0.34	0.08	0.21
W38_09	13.48	29.21	114.84	22.60	1.12	6.81	21.82	3.75	0.04	3.37	0.44	0.13	0.90
W38_10	13.50	30.16	85.18	33.16	1.23	7.45	22.55	5.20	1.01	1.37	0.73	0.12	1.00
W51_05	5.36	n/a	n/a	n/a	n/a	n/a	n/a	1.72	8.58	10.47	1.32	0.31	2.76
W51_06	5.24	29.26	67.44	n/a	0.69	5.10	23.10	0.99	n/a	n/a	n/a	n/a	n/a
W51_07	2.13	29.96	51.99	n/a	0.66	5.19	23.93	0.55	10.55	13.27	1.51	0.26	5.02
W51_08	4.93	25.02	23.38	5.95	1.07	8.14	19.78	1.45	7.91	16.36	0.81	0.27	4.90
W51_10	7.81	28.81	43.44	8.46	1.10	7.60	22.45	1.13	8.84	15.44	1.02	0.46	8.45
W52_09	5.15	30.02	41.48	3.31	0.93	6.81	23.72	0.77	9.23	11.92	1.15	0.38	3.01

Table 4: Chemical Environmental measurements for biweekly dataset (2009) In analyses missing data was corrected with average measurement for that week over 20 years.

	Temperature (°C)	*Salinity (psu)*	*Photosynthetically active radiation (uE/m2/s)*	*Fluorescence mg/m3*	*Oxygen saturation (percent)*	*Oxygen concentration (mL L-1)*	*Seawater density (kg m-3)*	*Turner chlorophyll (mg m-3)*	*Nitrate+Nitrite concentration (mmol m-3)*	*Silicate (mmol m-3)*	*Phosphate (mmol m-3)*	*Nitrite (mmol m-3)*	*Ammonium (mmol m-3)*
W02_09	2.74	29.94	91.41	2.58	1.02	7.93	23.87	0.61	7.67	10.95	1.03	0.34	4.46
W04_09	1.70	29.92	59.42	4.33	0.99	7.93	23.93	0.79	8.40	13.00	1.06	0.39	10.36
W06_09	1.03	29.72	147.85	3.34	1.00	8.15	23.80	1.36	7.80	10.71	1.19	0.35	5.44
W08_09	0.92	30.22	88.27	2.49	1.04	8.49	24.20	1.36	8.09	9.86	1.12	0.31	3.59
W10_09	1.45	27.38	332.06	4.60	1.02	8.35	21.91	3.41	8.92	11.62	1.03	0.32	3.89
W13_09	1.39	30.43	115.17	7.25	1.04	8.35	24.35	9.04	6.12	5.15	1.06	0.31	0.90
W15_09	3.88	26.09	125.13	27.68	1.10	8.57	20.72	22.01	1.02	1.90	0.69	0.19	1.27
W17_09	4.56	29.54	134.29	15.84	1.24	9.28	23.39	13.65	0.39	0.38	0.52	0.20	0.72
W22_09	7.84	30.07	378.47	7.05	1.10	7.58	23.42	4.27	0.03	0.37	0.29	0.11	0.58
W24_09	n/a	n/a	n/a	n/a	n/a	n/a	n/a	3.39	0.25	0.35	0.43	0.12	1.06
W26_09	14.12	29.06	93.84	24.34	1.19	7.16	21.58	13.22	0.02	1.76	0.52	0.16	1.11
W28_09	15.82	27.49	417.22	12.90	1.18	6.93	20.02	8.02	0.00	0.14	0.32	0.11	1.79
W30_09	18.09	28.36	31.73	18.03	1.06	5.90	20.17	5.63	0.05	1.79	0.30	0.13	1.05
W32_09	18.60	28.49	377.84	12.00	0.82	4.55	20.15	7.00	0.20	0.12	0.26	0.15	1.15
W34_09	18.37	29.98	850.24	8.03	1.10	6.07	21.34	4.52	0.00	0.03	0.18	0.11	1.02
W36_09	15.54	26.48	47.60	27.91	1.18	7.04	19.30	18.05	0.00	9.88	0.60	0.13	1.06
W38_09	13.48	29.21	114.84	22.60	1.12	6.81	21.82	3.75	0.04	3.37	0.44	0.13	0.90
W40_09	12.97	29.89	75.57	17.84	1.05	6.42	22.44	10.92	0.22	0.59	0.60	0.15	3.03
W42_09	11.14	27.95	47.00	14.07	0.98	6.31	21.27	8.79	1.30	3.69	0.65	0.32	4.63
W44_09	9.88	28.76	152.04	8.11	1.04	6.87	22.11	6.14	3.53	8.12	1.01	0.55	5.07
W46_09	9.45	28.73	101.12	4.87	0.98	6.56	22.15	1.19	5.38	10.62	0.94	0.80	8.86
W48_09	9.03	29.03	66.55	11.75	1.01	6.83	22.44	3.92	6.84	6.92	0.98	0.46	4.16
W50_09	6.20	n/a	139.75	2.40	1.26	10.93	11.04	0.57	8.25	11.74	0.85	0.44	4.84
W52_09	5.15	30.02	41.48	3.31	0.93	6.81	23.72	0.77	9.23	11.92	1.15	0.38	3.01

Table 5: Biotic Environmental measurements for seasonal dataset (2005-2010)

	Bacterial abundance (cells mL-1)	*Particulate organic carbon (mg m-3)*	*Particulate organic nitrogen (mg m-3)*	*Synechococcus at 5m (cells mL-1)*	*Cryptophytes at 5m (cells mL-1)*	*Picoeukaryotes at 5m (cells mL-1)*	*Phytoplankton abundance at 5m (cells mL-1)*	*Picophytoplankton abundance at 5m(cells mL-1)*	*Small Nanophytoplankton abundance (cells mL-1)*	*Large Nanophytoplankton abundance (cells mL-1)*
W13_05	1219574	918	172	210	648	2932	6272	3142	2377	754
W13_06	1708118	394	65	117	14	1621	3014	1739	1164	112
W13_07	1858501	1273	177	337	663	1366	8266	1703	5649	914
W13_08	1784091	511	95	253	2087	2087	6567	2340	3599	628
W13_09	910454	567	95	428	99	5623	7801	6051	1464	286
W13_10	1578937	607	91	238	269	9395	15933	9633	5747	554
W26_05	4321098	549	84	12	65	4255	5805	4267	1464	74
W26_06	10016815	755	138	1478	369	6664	19746	8142	11372	231
W26_07	6252407	2485	264	249	58	50669	55135	50918	3764	452
W26_08	2999892	881	163	65	1041	6244	27920	6309	21416	195
W26_09	6148231	949	177	204	202	18309	21530	18513	2670	347
W26_10	3287311	1053	184	490	684	13528	30975	14018	16324	633
W38_05	5852386	1505	181	51259	122	29534	98041	80793	17003	244
W38_06	792048	326	50	2956	71	2029	6086	4984	1017	85
W38_07	4614018	1045	180	81686	625	30316	121413	112002	7804	1607
W38_09	2970274	1100	182	42250	2235	30804	100906	73054	26216	1636
W38_10	2049926	979	168	88121	2736	18072	116731	106193	8457	2081
W51_05	769488	217	29	1556	201	1180	3592	2736	800	55
W51_06	857749	234	40	2628	266	2928	6901	5556	1297	48
W51_07	857101	318	49	1643	107	5527	8648	7171	1388	90
W51_08	493842	265	33	3681	193	4767	9402	8448	938	15
W51_10	1312952	153	18	6106	342	7667	15435	13773	1549	113
W52_09	608831	175	28	4008	171	3409	9028	7416	1464	147

Table 6: Biotic Environmental measurements for biweekly dataset (2009)

	Bacterial abundance (cells mL-1)	*Particulate organic carbon (mg m-3)*	*Particulate organic nitrogen (mg m-3)*	*Synechococcus abundance at 5m (cells mL-1)*	*Cryptophyte abundance at 5m (cells mL-1)*	*Picoeukaryote abundance at 5m (cells mL-1)*	*Phytoplankton abundance at 5m (cells mL-1)*	*Pico-phytoplankton abundance at 5m(cells mL-1)*	*Small Nano-phytoplankton abundance (cells mL-1)*	*Large Nano-phytoplankton abundance (cells mL-1)*
W02_09	447900.00	192.00	23.00	1325	73.00	2128.00	4861.00	3453.00	1289.00	119.00
W04_09	544996.00	269.00	36.00	885	206.00	2475.00	5873.00	3360.00	2213.00	301.00
W06_09	817276.00	244.00	27.00	565	46.00	2961.00	4673.00	3526.00	906.00	242.00
W08_09	610101.00	309.00	36.00	401	36.00	2483.00	3802.00	2884.00	807.00	111.00
W10_09	859386.00	611.00	78.00	236	184.00	3370.00	5978.00	3606.00	1914.00	458.00
W13_09	910454	567	95	428	99	5623	7801	6051	1464	286
W15_09	2385671.00	766.00	130.00	341	305.00	3667.00	5960.00	4008.00	1626.00	327.00
W17_09	3554478.00	834.00	139.00	290	270.00	6129.00	10018.00	6420.00	2864.00	734.00
W22_09	2037332.00	1125.00	183.00	194	375.00	19933.00	25086.00	20127.00	4318.00	641.00
W24_09	1885614.00	551.00	93.00	67	95.00	17202.00	24173.00	17269.00	6390.00	514.00
W26_09	6148231	949	177	204	202	18309	21530	18513	2670	347
W28_09	4517524.00	920.00	148.00	345	2128.00	2820.00	12120.00	3165.00	8386.00	569.00
W30_09	4941558	1289	192	13584	2152	49676	72173	63260	8068	845
W32_09	7708901.00	910.00	146.00	30853	238.00	31226.00	70075.00	62078.00	6077.00	1920.00
W34_09	3210491.00	846.00	136.00	8895	575.00	27279.00	47456.00	36173.00	10234.00	1049.00
W36_09	2892221.00	606.00	93.00	19937	462.00	28850.00	57318.00	48787.00	8172.00	359.00
W38_09	2970274	1100	182	42250	2235	30804	100906	73054	26216	1636
W40_09	3494995.00	887.00	139.00	66074	1242.00	20256.00	103517.00	86329.00	14038.00	3150.00
W42_09	1957683.00	619.00	115.00	37845	736.00	16135.00	61356.00	53980.00	5345.00	2031.00
W44_09	1612064.00	451.00	75.00	42248	534.00	26522.00	74744.00	68770.00	5532.00	442.00
W46_09	1359225.00	331.00	52.00	25131	274.00	8195.00	38291.00	33325.00	4699.00	266.00
W48_09	1771508.00	508.00	76.00	77830	438.00	45233.00	131653.00	123063.00	8247.00	343.00
W50_09	798778.00	338.00	44.00	12025	252.00	3937.00	17672.00	15962.00	1565.00	145.00
W52_09	608831	175	28	4008	171	3409	9028	7416	1464	147

Table 7: Analysis of similarity between seasons and p value. Asterix (*) indicate significant value (p<0.005) as determined by **Bray–Curtis** similarity coefficient conducted at 97% sequence similarity

	Spring		**Summer**		**Autumn**	
Summer	0.837	0.002*				
Autumn	0.850	0.007*	0.3971	0.015		
Winter	0.615	0.001*	0.845	0.002*	0.641	<0.001*

7.0 Supplementary information

7.1 Appendix I. Caveats of the study

It is evident that there is logistical difficulty in conducting long-term time series analysis, especially over the duration of this study. For a graduate student to replicate the duration of this study, the duration of a PhD would be needed for sampling with an extra year for data analysis taking an estimated seven years. This is why spatial scale studies outnumber temporal scale studies, especially in marine environments. However, the use of archival samples as an innovative method to conduct time-series analysis can be done for bacterial communities as has been demonstrated in this study.

Along with any molecular technique, there were a few problems. Firstly, because there are 10^6 bacterial cells in 1 ml of seawater (Fuhrman, 2009, Kirchman and Mitchel, 2008), filters had to be cut into 8 pieces, and filter-PCR was conducted on $1/8^{th}$ of the filter. While we believe this is an adequate representation of community diversity based on volumes tested in Kirchman et al. (2001), which showed PCR product on as little as 250 μl of sample, there may be underestimation of true diversity.

In scientific studies, observations are as robust as the replicates in which they are observed. In this study, we are aware that there are no replicates for each sample, and this is due to the material constraint of the method. We decided it was more valuable to sequence more samples at different time points rather than replicate samples in order to conduct future experimentation with the material available to us from eight years ago, such as quantitative PCR of a diagnostic gene (such as methanol dehydrogenase). However, we did address the issue of replicates in that we processed three filters from the

same sample (week 13 in 2006) in order to observe how this affects diversity estimates and relative abundance in species. We found that the same taxonomic groups are found in all three samples with the same relative abundance (±1.5%). Beta diversity estimates indicate that these samples are on average 95% similar to each other (Bray-Curtis similarity values).

7.2 Appendix II. Different phylogenetic resolution reveals different processes

In using the 16S approach to determine community structure in Bedford we can see how different observations can be made from different sequence similarity. Richness and diversity estimates differed in this study depending on whether we were looking at the species level, (97%) or the family level (90%). These levels were determined by comparison of the 16S percent identity and percent genes shared between two organisms (Konstantinidis and Tiedje, 2007). However, at the family level (90%) the same ANOSIM between the spring and winter show that each season can be distinguished from the other. This difference indicates that between the spring and the winter, families of bacteria change however there is still high microdiversity in each season (table 17). Another instance in which different sequence similarity portrays different result is seen when comparing the biweekly one-year NMS (2009) for the species and family levels (figure 13). Contrary to the multi-year low-frequency study, the NMS ordination at the species level (97%) and the family level (90%) revealed different patterns (**Supp. Figure 8**). For example, at the species, week 52 is considered an outlier (STDEV=2.9), but at the family level it is not. This indicated that although the OM38 clade is close in relative abundance during the winter (weeks 52, 2, and 4), at the species level there is one phylotype of OM38, which blooms in week 52 (5%.) This result coincides with the

decrease in alpha diversity observed at week 52 (figure 10). On the other hand, at the family level (90%) week 32 was considered an outlier sample (mentioned earlier). This confirmed that it was not one ecotype of a clade, or a particular species adapted to an environmental condition that bloomed, but rather a dramatic shift in phyla, as a result of what was addressed earlier.

7.3 Appendix III. Community Beta-diversity using other dissimilarity indexes:

7.3.1 Multi-year Betadiversity

The UPGMA of ThetaYC and Jaccard index show similar results as seen in supplementary figure**s** 6 and 7. UPGMA of Thetayc dissimilarities at 97%, the winter samples cluster closer to the spring samples except two spring samples (2007 and 2010) which cluster closer to the summer and fall samples as they occur post diatom bloom (supplementary figure 6b). To compare community membership, the UPGMA of Jaccard distances at 97% show that each season has associated community members, see **figure 8c.** . The Jaccard distance at 90% shows two clusters of communities with two exceptions, spring 2007, which closely resembles the Autumn communities, and Autumn 2006, which closely resembles the winter communities. Tables containing dissimilarity values between each season for the three different indices for 97% and 90% can be seen in **tables 7-12**.

ANOSIM using thetaYC and Jaccard

ANOSIM of ThetaYC (at 97%) indices statistically significant dissimilarity between Spring and autumn (r=0.84), Spring and Summer (r=0.811), summer and winter

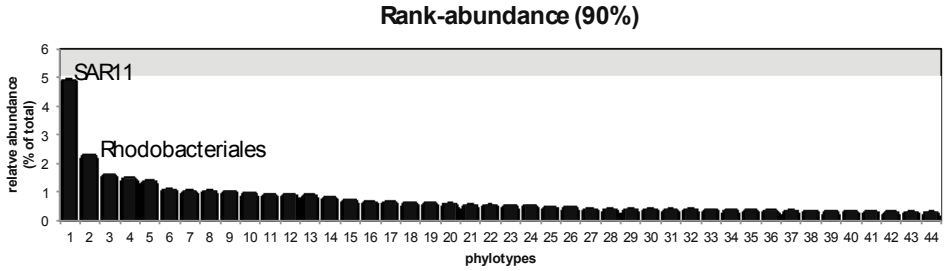
(r=0.78), and autumn and winter (r=0.645). However using ThetaYC, spring and winter samples were less distinguishable from each other (r=0.351, p=0.351) and the variation within the summer and autumn seasons could not be distinguished from the variation between these two seasons (r=0.261, p=0.024).

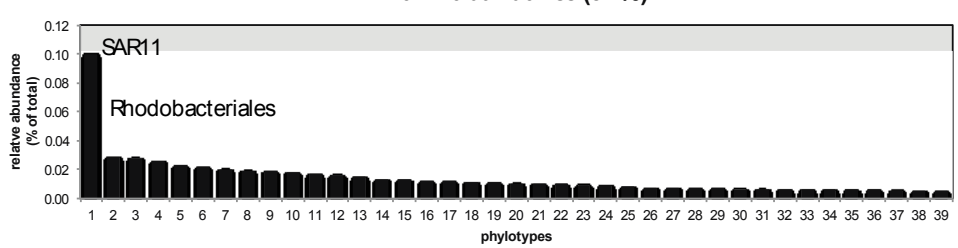
ANOSIM of ThetaYC at 90% showed the same results, as the bray-curtis indicating that at the diversity at family level is same. ANOSIM on the community membership using the Jaccard index also showed the same result at both 97% and 90%; all seasons are statistically different each other with the more similar samples being spring and winter (r=0.74 at 97% and r=0.49 at 90%) and summer and autumn (r=0.6 at 97% and 0.54 at 90%). Tables summarizing results from ANOSIM between seasons can be seen in supplementary tables 6-10**.**

This result is consistent whether the index used places more weight on relative abundance (thetaYC) or on species counts (Bray-Curtis). Looking at ANOSIMs with more weight on relative abundance, we observe that at the species level the spring and winter seasons are just as diverse with the season as between these two seasons (Tables 13,14).

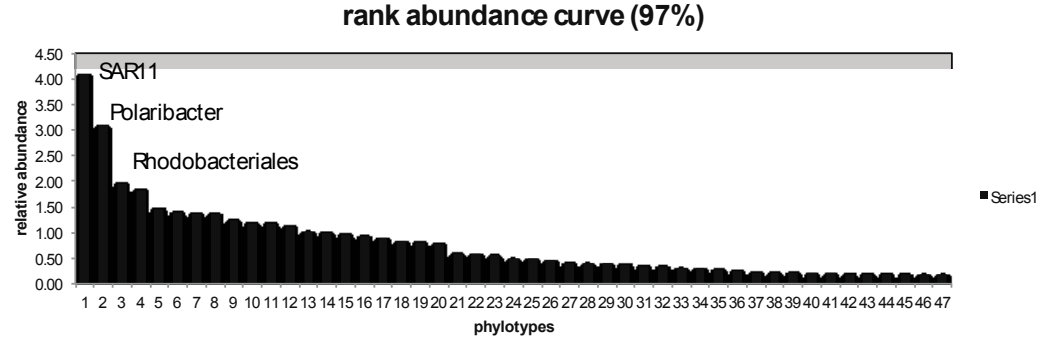
Supplementary figures:
Figure S1. Rank-abundance curves for all datasets.

biweekly (2009)

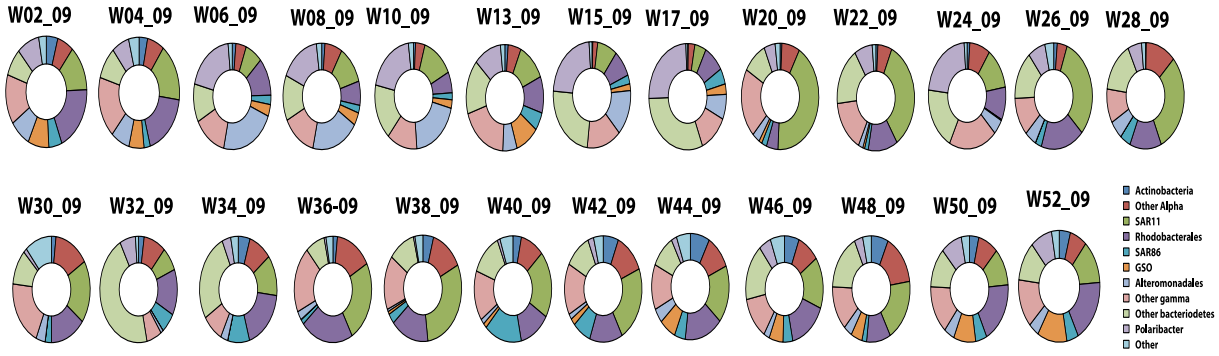




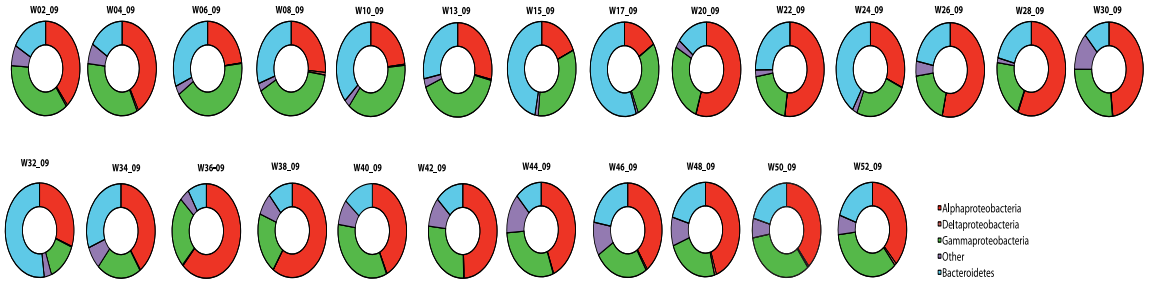
5-year seasonal dataset



S2. Taxonomic composition of each sample occuring in 2009 at the class level.

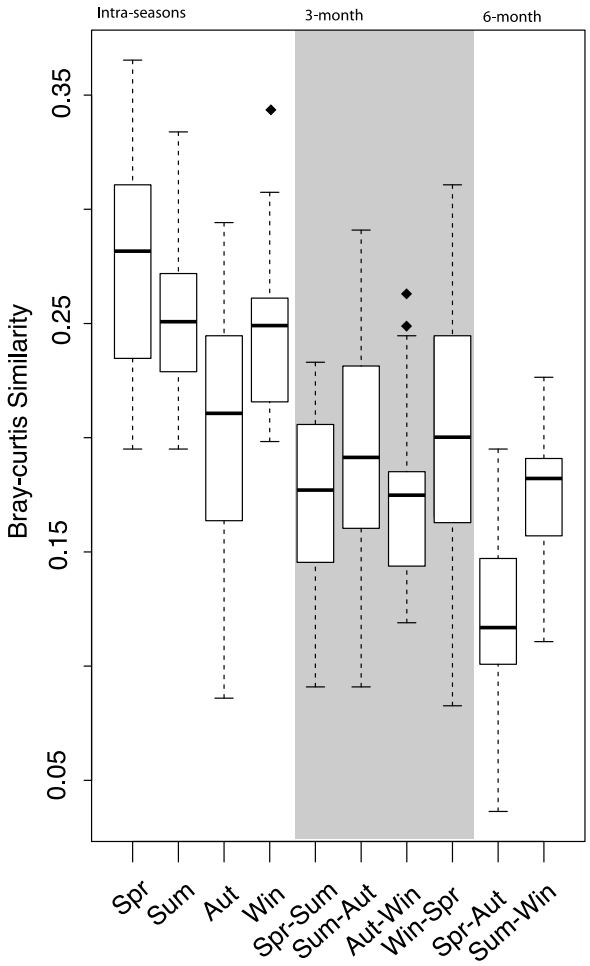


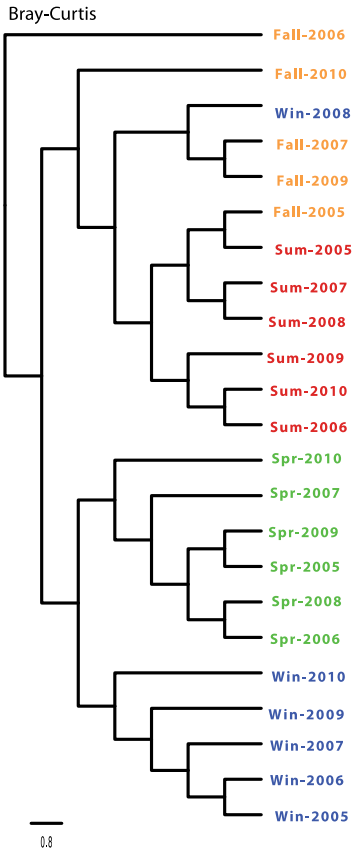
S3 Taxonomic composition of each sample occurring in 2009 at the phylum level.

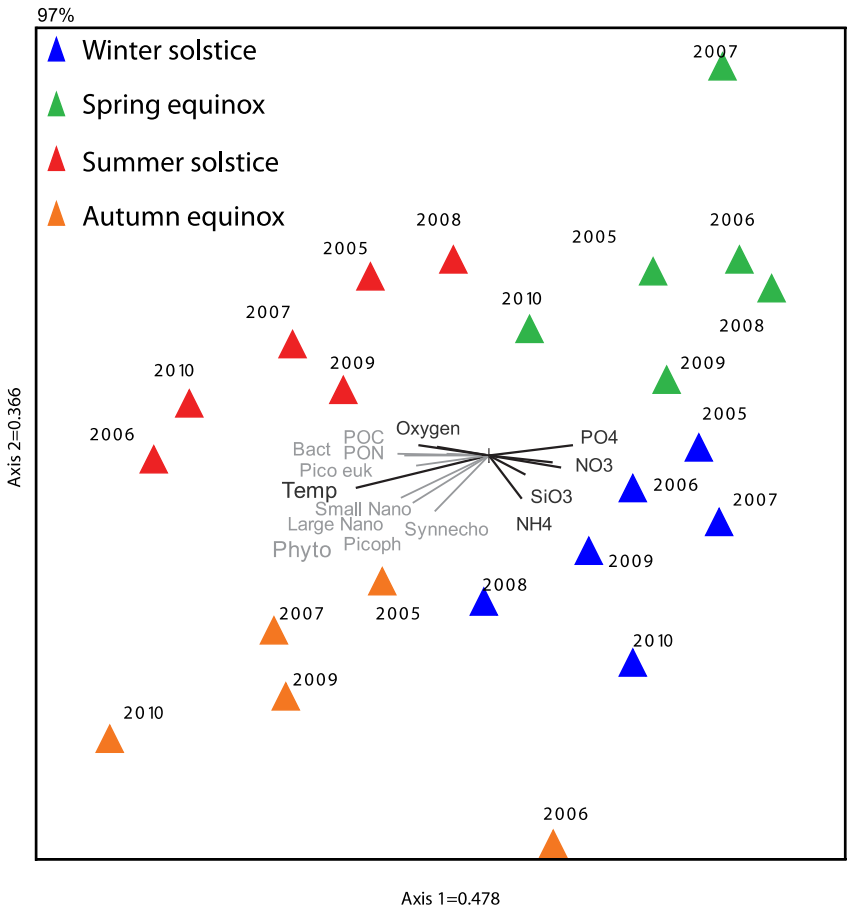


S4. Distance decay using Bray-curtis similarity at 97% sequence similiarty.

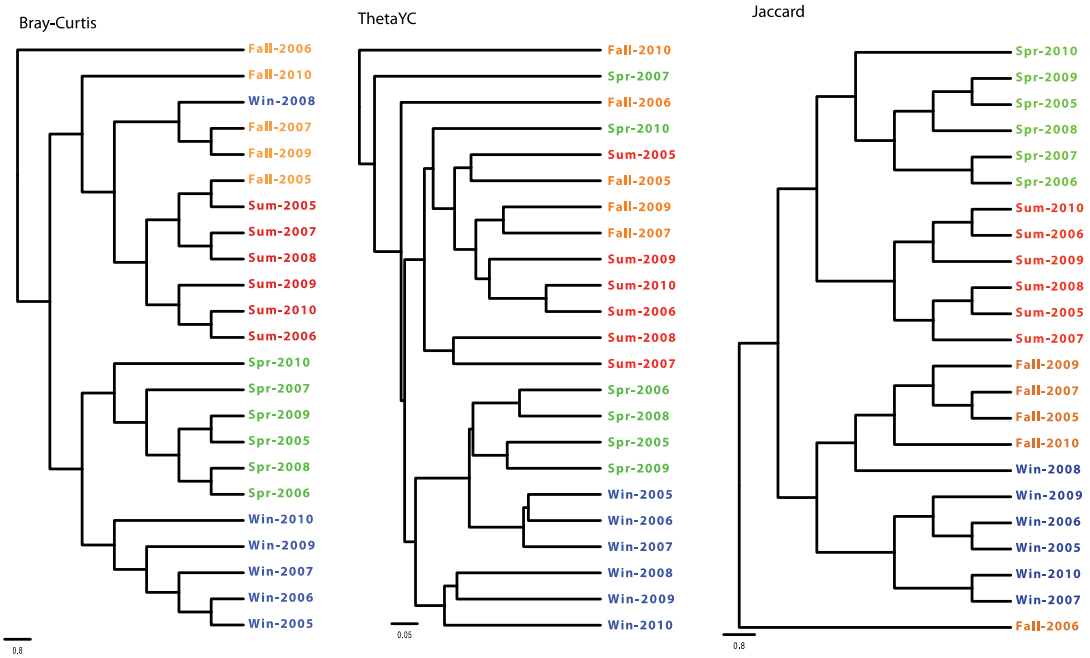
97



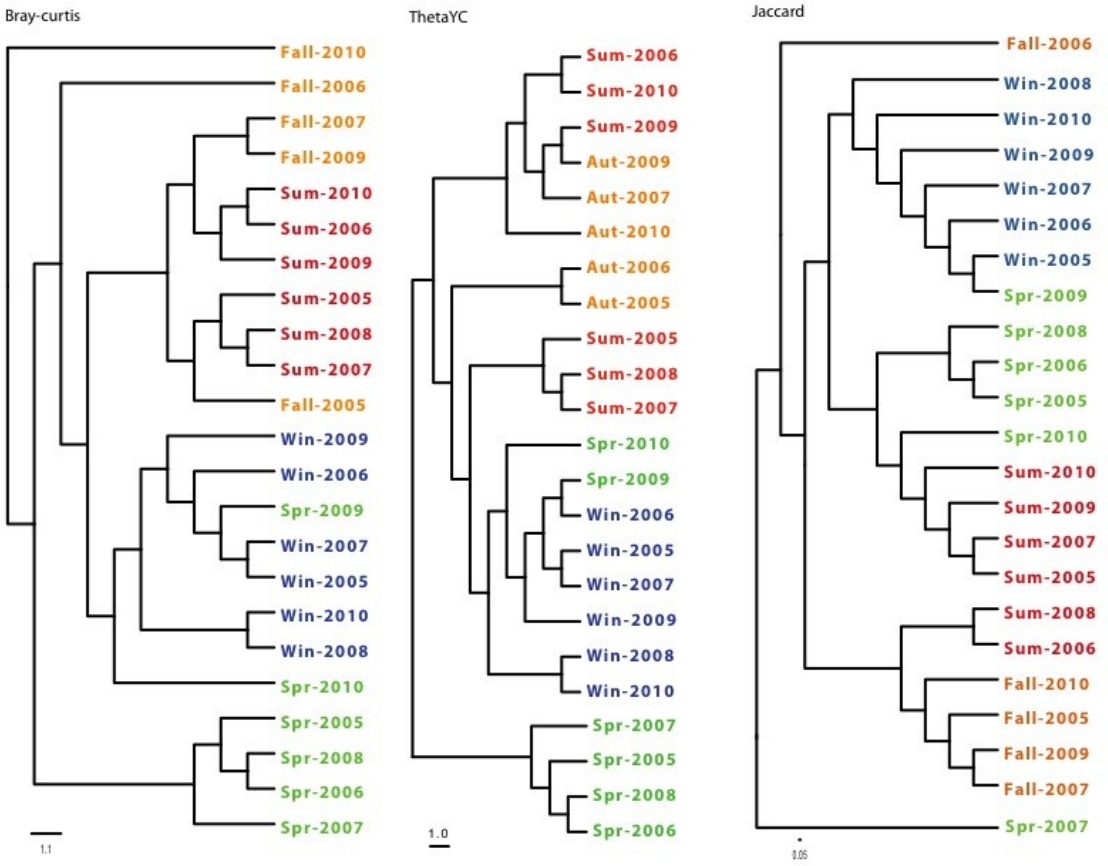




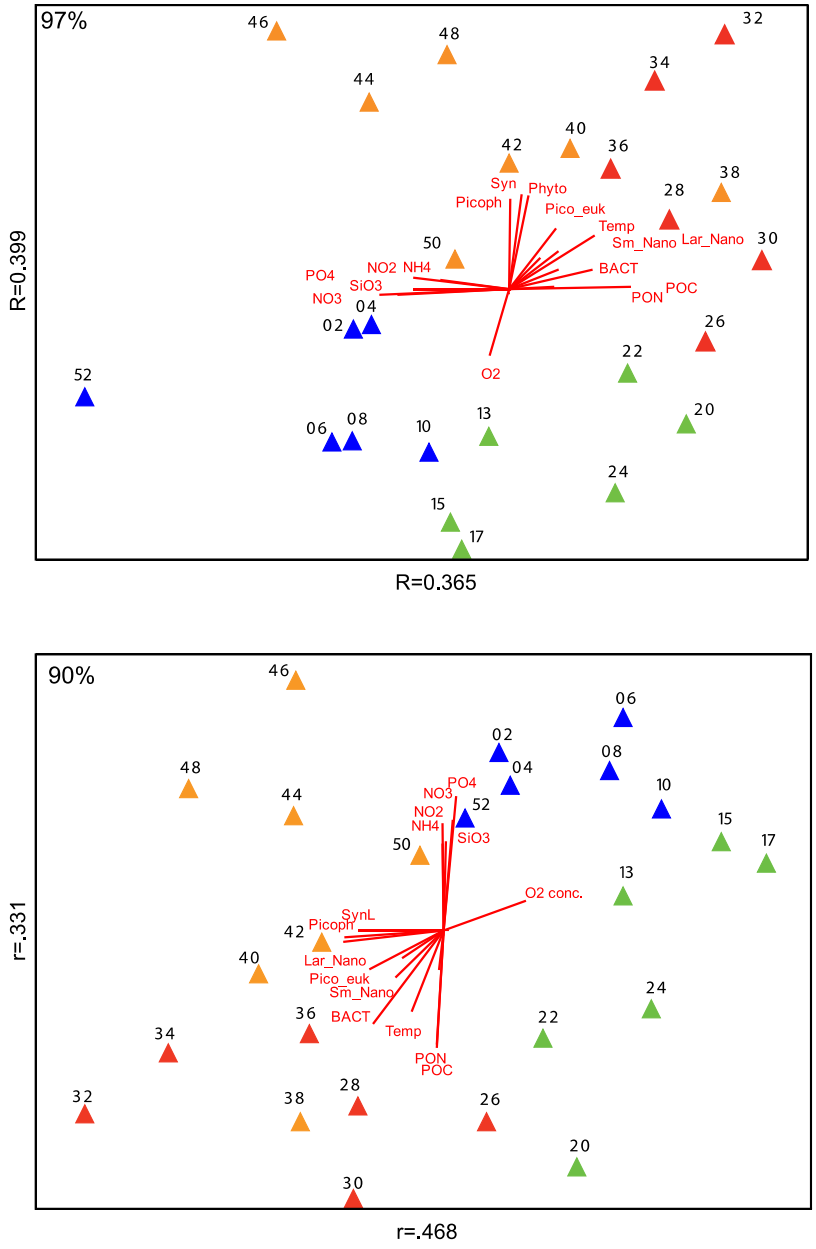
S5. NMS ordination seasonal dataset (2005-2010) at 97% sequence similiarity.



S6. UPGMA dendograms of Bray-curtis, thetaYC and Jaccard index at 97% sequence similiarity.



S7. UPGMA dendograms of Bray-curtis, thetaYC and Jaccard index at 97% sequence similiarity



S8. NMS ordinations for the biweekly study at both 97% and 90%.

Supplementary tables:

Supplementary Table 1 : Alpha diversity estimates for seasonal study (2005-2010) with various calculators at two cluster distances (97 and 90). Bold rows indicate averages for the season.

	97						90					
	CHAO1	**ACE**	**SHANNON**	**SHANNON-EVENESS**	**SIMPSON**	**SIMPSON-EVENESS**	**CHAO1**	**ACE**	**SHANNON**	**SHANNON-EVENESS**	**SIMPSON**	**SIMPSON-EVENESS**
Fall-2005	3702.31	11892.60	5.39	0.91	0.01	0.31	391.07	530.96	3.73	0.77	0.05	0.16
Fall-2006	2741.12	7054.68	5.52	0.92	0.01	0.35	363.40	493.73	3.62	0.74	0.07	0.10
Fall-2007	2384.37	6467.76	5.48	0.92	0.01	0.28	364.50	372.80	3.62	0.76	0.06	0.14
Fall-2009	2499.31	5279.74	5.43	0.91	0.01	0.22	507.50	596.72	3.34	0.72	0.09	0.11
Fall-2010	2353.43	5277.13	5.02	0.86	0.03	0.10	310.57	562.18	2.75	0.59	0.18	0.05
Fall	**2736.11**	**5674.88**	**5.31**	**0.90**	**0.02**	**0.20**	**387.41**	**511.27**	**3.41**	**0.72**	**0.09**	**0.11**
Spr-2005	2627.00	11587.99	5.28	0.89	0.01	0.22	381.75	1075.96	3.24	0.71	0.08	0.13
Spr-2006	3026.06	14448.69	5.09	0.87	0.02	0.16	453.00	772.35	3.05	0.68	0.09	0.13
Spr-2007	1702.22	4003.66	4.77	0.83	0.03	0.09	284.20	375.39	2.29	0.54	0.26	0.06
Spr-2008	3282.65	8712.22	5.10	0.87	0.02	0.13	411.86	765.04	3.05	0.67	0.11	0.10
Spr-2009	3155.16	12853.09	5.29	0.89	0.01	0.23	385.00	571.02	3.39	0.74	0.06	0.17
Spr-2010	3651.24	10504.64	5.46	0.92	0.01	0.34	257.60	312.78	3.55	0.77	0.05	0.21
Spring	**2907.39**	**10351.71**	**5.16**	**0.88**	**0.02**	**0.20**	**362.23**	**645.42**	**3.09**	**0.69**	**0.11**	**0.13**
Sum-2005	3545.88	9636.37	5.37	0.91	0.01	0.28	229.00	375.38	3.39	0.74	0.06	0.17
Sum-2006	3027.14	8337.01	5.29	0.89	0.02	0.15	250.91	606.66	2.85	0.63	0.14	0.08
Sum-2007	2641.89	7869.98	5.21	0.89	0.01	0.23	283.17	457.12	3.14	0.72	0.07	0.17
Sum-2008	2785.29	6260.58	5.41	0.91	0.01	0.32	355.43	589.25	3.37	0.74	0.06	0.18
Sum-2009	3346.88	7909.41	5.65	0.93	0.01	0.30	264.00	398.06	3.32	0.72	0.08	0.11
Sum-2010	3428.56	8034.54	5.27	0.89	0.02	0.14	220.80	284.21	2.73	0.62	0.18	0.07
Summer	**3129.27**	**8007.98**	**5.37**	**0.90**	**0.01**	**0.24**	**267.22**	**451.78**	**3.13**	**0.69**	**0.10**	**0.13**
Win-2005	4026.44	11696.27	5.36	0.90	0.01	0.23	381.46	549.69	3.66	0.76	0.05	0.15
Win-2006	2923.95	10920.69	5.39	0.91	0.01	0.25	293.32	545.48	3.75	0.77	0.04	0.18
Win-2007	3132.96	10356.34	5.52	0.92	0.01	0.30	326.35	387.58	3.70	0.76	0.05	0.14
Win-2008	4467.47	16853.49	5.73	0.94	0.00	0.49	280.67	555.90	3.68	0.78	0.04	0.20
Win-2010	2940.04	11054.17	5.64	0.94	0.01	0.44	345.13	613.47	4.06	0.81	0.03	0.21
Win-2009	2865.81	5892.11	5.57	0.93	0.01	0.34	503.71	1084.27	3.75	0.76	0.05	0.14
Winter	**3392.78**	**11128.85**	**5.53**	**0.92**	**0.01**	**0.34**	**355.11**	**622.73**	**3.77**	**0.77**	**0.05**	**0.17**

Supplementary Table 2: Alpha diversity estimates for biweekly study (2009) with various calculators at two cluster distances (97 and 90). Bold rows indicate averages for the season

	97						90					
	CHAO1	**ACE**	**SHANNON**	**SHANNON-EVENESS**	**SIMPSON**	**SIMPSON-EVENESS**	**CHAO1**	**ACE**	**SHANNON**	**SHANNON-EVENESS**	**SIMPSON**	**SIMPSON-EVENESS**
W02_09	7028.28	28033.83	6.12	0.92	0.01	0.25	1637.33	2633.59	5.70	0.90	0.01	0.24
W04_09	7377.24	31143.54	6.18	0.92	0.00	0.25	2306.67	4466.48	5.79	0.90	0.01	0.25
W06_09	6644.00	27645.84	5.63	0.86	0.02	0.08	2000.40	4742.72	5.42	0.85	0.02	0.09
W08_09	5484.13	21724.08	5.86	0.89	0.01	0.11	1809.63	4221.25	5.56	0.88	0.01	0.13
W10_09	7245.50	25802.88	5.86	0.88	0.01	0.10	2112.68	4323.69	5.48	0.86	0.02	0.10
W13_09	4504.10	12017.43	5.44	0.86	0.01	0.15	1430.54	2399.58	5.19	0.85	0.01	0.17
W15_09	7303.03	26341.66	5.88	0.89	0.01	0.16	1678.00	4754.70	5.28	0.85	0.01	0.13
W17_09	7014.26	19055.69	6.02	0.91	0.01	0.18	1531.05	2940.18	5.54	0.88	0.01	0.17
W20_09	4509.90	10938.86	5.69	0.88	0.02	0.09	1556.96	2915.33	5.25	0.85	0.02	0.10
W22_09	6624.58	14388.10	5.97	0.90	0.01	0.12	1543.54	2764.56	5.49	0.87	0.02	0.11
W24_09	7698.92	21256.46	6.09	0.91	0.01	0.21	1834.60	4687.12	5.64	0.89	0.01	0.19
W26_09	4959.31	15200.04	5.79	0.88	0.01	0.11	1419.55	2737.51	5.37	0.86	0.02	0.13
W28_09	7740.18	19781.19	5.96	0.90	0.01	0.10	1711.19	3664.27	5.53	0.88	0.02	0.12
W30_09	3310.88	8630.70	5.73	0.88	0.01	0.16	1125.40	2189.72	5.21	0.85	0.01	0.16
W32_09	6779.00	16003.47	5.78	0.88	0.02	0.08	1273.10	2651.09	5.22	0.84	0.03	0.07
W34_09	6164.63	19329.85	6.28	0.94	0.00	0.38	1612.00	3827.23	5.73	0.90	0.01	0.28
W36-09	5916.33	23259.88	6.02	0.91	0.01	0.13	2167.92	3798.38	5.69	0.89	0.01	0.16
W38_09	4252.00	10360.51	5.56	0.86	0.02	0.09	1338.54	2036.23	5.16	0.85	0.02	0.11
W40_09	8765.03	25865.28	6.23	0.93	0.00	0.28	1723.82	3382.03	5.80	0.91	0.01	0.22
W42_09	9350.29	30007.66	6.30	0.93	0.00	0.26	1640.12	3027.71	5.92	0.92	0.01	0.29
W44_09	5613.02	18037.50	6.18	0.93	0.00	0.26	2097.88	4362.26	5.77	0.90	0.01	0.25
W46_09	7514.16	26860.02	6.27	0.93	0.00	0.26	1830.11	3313.19	5.91	0.91	0.01	0.23
W48_09	5980.42	18657.17	6.22	0.93	0.01	0.24	1796.97	3756.06	5.78	0.90	0.01	0.20
W50_09	6719.38	22448.88	6.04	0.91	0.01	0.21	2005.44	4121.84	5.76	0.90	0.01	0.25
W52_09	4627.42	10158.53	5.77	0.88	0.01	0.15	1521.31	2595.87	5.38	0.86	0.01	0.16

S. Table 3: Environmental parameters and their correlations with the axis observed on the NMS ordination for seasonal dataset (2005-2010). (Calculated for 97% and OTU 90% sequence similarity)

	97		90	
Axis:	*1*	*2*	*1*	*2*
Temperature	-0.82	-0.41	-0.654	-0.689
Salinity	0.34	0.05	0.328	0.174
PAR	-0.13	0.35	-0.295	0.189
Fluorescence	-0.66	-0.05	-0.68	-0.32
Oxygen saturation	-0.47	0.11	-0.54	-0.17
Oxygen concentration	-0.04	0.31	-0.20	0.20
Seawater density	0.66	0.24	0.57	0.47
Turner chlorophyll	-0.16	0.30	-0.32	0.18
Nitrate+Nitrite	0.57	-0.20	0.70	0.08
Silicate	0.43	-0.32	0.60	-0.06
Phosphate	0.61	-0.25	0.77	0.07
Nitrite	0.43	-0.18	0.56	0.04
Ammonium	0.41	-0.47	0.67	-0.18
Bacterial abundance	-0.68	0.09	-0.75	-0.25
Particulate organic carbon	-0.51	0.21	-0.66	-0.06
Particulate organic nitrogen	-0.60	0.23	-0.76	-0.09
Synechococcus abundance at 5m	-0.53	-0.54	-0.33	-0.60
Cryptophyte abundance	-0.29	-0.18	-0.26	-0.24
Picoeukaryote	-0.62	-0.23	-0.55	-0.45
Phytoplankton	-0.67	-0.47	-0.52	-0.62
Picophytoplankton	-0.63	-0.50	-0.45	-0.62
Small Nanophytoplankton	-0.55	-0.10	-0.59	-0.31
Large Nanophytoplankton	-0.47	-0.21	-0.42	-0.33

S. Table 4: Environmental parameters and their correlations with the axis observed on the NMS ordination for biweekly dataset (2009). (Calculated for 97% and OTU 90% sequence similarity)

	97		90	
Axis:	*1*	*2*	*1*	*2*
Temperature	0.74	-0.38	-0.74	0.50
Salinity	-0.16	0.03	0.13	-0.03
PAR	0.39	-0.06	-0.49	-0.05
Fluorescence	0.39	0.00	-0.37	0.15
Oxygen saturation	0.20	0.36	-0.19	-0.30
Oxygen concentration	-0.45	0.54	0.46	-0.61
Seawater density	-0.29	0.19	0.26	-0.20
Turner chlorophyll	0.23	0.18	-0.25	-0.17
Nitrate+Nitrite	-0.74	-0.02	0.77	-0.07
Silicate	-0.62	-0.15	0.67	0.02
Phosphate	-0.82	-0.05	0.86	-0.09
Nitrite	-0.70	-0.49	0.74	0.28
Ammonium	-0.58	-0.40	0.60	0.20
Bacterial abundance	0.68	-0.16	-0.72	0.27
Particulate organic carbon	0.72	0.10	-0.70	0.14
Particulate organic nitrogen	0.72	0.10	-0.71	0.11
Synechococcus abundance at 5m	-0.10	-0.72	0.05	0.78
Cryptophyte abundance	0.55	-0.19	-0.45	0.43
Picoeukaryote	0.32	-0.47	-0.32	0.74
Phytoplankton	0.15	-0.67	-0.17	0.83
Picophytoplankton	0.07	-0.69	-0.11	0.84
Small Nanophytoplankton	0.47	-0.35	-0.46	0.50
Large Nanophytoplankton	0.47	-0.29	-0.51	0.33

S. Table 5. Beta diversity estimates using several parameters at 90 and 97 % sequence similarity

Comparison		90%					97%				
Week1	*Week 2*	*Bray-Curtis*	*sharedsobs*	*sharedchao*	*jclass*	*thetayc*	*Bray-Curtis*	*sharedsobs*	*sharedchao*	*jclass*	*thetayc*
2	4	0.57	175	458.05	0.83	0.31	0.76	66	251.58	0.96	0.46
4	6	0.66	159	482.89	0.84	0.79	0.84	53	216.33	0.96	0.87
6	8	0.54	173	515.99	0.82	0.16	0.69	70	327.09	0.95	0.18
8	10	0.62	160	403.20	0.84	0.72	0.81	58	95.43	0.96	0.91
10	13	0.70	125	356.93	0.86	0.78	0.86	41	129.61	0.97	0.89
13	15	0.65	121	354.50	0.86	0.57	0.79	41	132.20	0.97	0.73
15	17	0.62	153	394.16	0.83	0.56	0.80	57	188.10	0.96	0.68
17	20	0.79	95	248.98	0.90	0.83	0.88	46	184.77	0.97	0.85
20	22	0.61	160	668.15	0.82	0.22	0.77	54	182.48	0.96	0.28
22	24	0.68	151	402.59	0.85	0.61	0.85	47	153.08	0.97	0.71
24	26	0.69	139	396.17	0.85	0.55	0.86	37	87.65	0.97	0.76
26	28	0.69	131	402.33	0.86	0.36	0.83	42	120.04	0.97	0.34
28	30	0.69	122	300.55	0.86	0.54	0.84	46	171.00	0.97	0.62
30	32	0.80	99	186.70	0.88	0.92	0.90	31	94.83	0.98	0.93
32	34	0.62	163	263.91	0.82	0.72	0.78	70	164.60	0.95	0.75
34	36	0.70	151	406.46	0.85	0.62	0.83	69	177.51	0.95	0.69
36	38	0.63	153	356.31	0.83	0.44	0.78	55	344.43	0.96	0.36
38	40	0.64	148	472.10	0.84	0.53	0.78	68	244.57	0.95	0.63
40	42	0.63	186	436.24	0.82	0.45	0.82	70	346.65	0.96	0.57
42	44	0.66	153	345.99	0.86	0.70	0.85	63	251.83	0.96	0.87
44	46	0.60	164	497.04	0.85	0.38	0.77	66	208.47	0.96	0.43
46	48	0.72	164	430.74	0.85	0.80	0.89	47	160.82	0.97	0.89

48	50	0.68	177	516.17	0.83	0.62	0.84	51	181.81	0.97	0.70
50	52	0.63	181	544.67	0.80	0.52	0.78	68	306.80	0.95	0.57
Average		**0.66**	**150.13**	**410.03**	**0.84**	**0.57**	**0.82**	**54.83**	**196.74**	**0.96**	**0.65**
Stdev		**0.06**	**24.11**	**104.27**	**0.02**	**0.19**	**0.05**	**11.98**	**76.68**	**0.01**	**0.22**

ANOSIMS (97%)

S.Table 6: Analysis of similiarity between seasons and p value. Asterix (*) indicate significant value (p<0.005) as determined by **Bray–Curtis** similarity coefficient conducted at 97% sequence similarity.

	Spring		**Summer**		**Autumn**	
Summer	0.879	0.002*				
Autumn	0.892	0.001*	**0.484**	0.006*		
Winter	**0.599**	0.002*	0.926	0.004*	0.682	0.005*

S. table 7: Analysis of similiarity between seasons and p value. Asterix (*) indicate significant value (p<0.005) determined by **Theta-YC** similarity coefficient conducted at 97% sequence similarity.

	Spring		**Summer**		**Autumn**	
Summer	0.811	0.003*				
Autumn	0.84	0.006*	0.261	0.024		
Winter	0.351	0.016	0.787	<0.001*	0.645	0.001*

Supplementary table 8: Analysis of similiarity between seasons and p value. Asterix (*) indicate significant value (p<0.005) as determined by **JCLASS** similarity coefficient conducted at 97% sequence similarity.

	Spring		**Summer**		**Autumn**	
Summer	0.822	0.001*				
Autumn	0.949	0.001*	0.602	0.003*		
Winter	0.744	<0.001*	0.8	0.002*	0.516	0.009

Anosims 90%

S. Table 9: Percent similarity values within and between groups ±standard deviation (ANOSIM: P<0.05) as determined by Theta-YC similarity coefficient conducted at 90% sequence similarity

	Spring		**Summer**		**Autumn**	
Summer	0.672	0.004*				
Autumn	0.856	0.004*	0.12	0.14		
Winter	0.537	0.001*	0.722	0.001*	0.637	0.002*

S. Table 10: Percent similarity values within and between groups ±standard deviation (ANOSIM: P<0.05) as determined by JCLASS similarity coefficient conducted at 90% sequence similarity

	Spring		**Summer**		**Autumn**	
Summer	0.769	0.001*				
Autumn	0.810	0.001*	0.546	0.001*		
Winter	0.490	0.004*	0.959	0.002*	0.848	0.002*