Stony Brook University

Title of Dissertation

A Dissertation Presented

by

Yuan Liu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Marine and Atmospheric Sciences

Stony Brook University

December 2012

Stony Brook University
The Graduate School

Yuan Liu

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Josephine Y. Aller – Dissertation co-Advisor
Professor, School of Marine and Atmospheric Sciences

Jackie L. Collier - Dissertation co-Advisor
Associate Professor, School of Marine and Atmospheric Sciences

Paul F. Kemp, Dissertation co-Advisor
Associate Director, C-MORE, School of Ocean and Earth Science
University of Hawaii

Robert M. Cerrato, Chairperson of Defense
Associate Professor, School of Marine and Atmospheric Sciences

Gordon T. Taylor
Professor, School of Marine and Atmospheric Sciences

Edward F. Delong
Department of Civil and Environmental Engineering
& Department of Biological Engineering
Massachusetts Institute of Technology

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Planktonic microbial community structure in Great South Bay, NY, where *Aureococcus anophagefferens* blooms frequently occur

by

Yuan Liu

Doctor of Philosophy

in

Marine and Atmospheric Sciences

Stony Brook University

2012

The eukaryotic and prokaryotic planktonic microbial communities in Great South Bay (GSB), NY, were characterized using Terminal Restriction Fragment Length Polymorphism (TRFLP) for both 18S rDNA and 16S rDNA in 2008 when there was an *Aureococcus anophagefferens* 'brown tide' bloom and in 2009 when there was no bloom. TRFLP data and *A. anophagefferens* cell counts using microscopy demonstrated both a spring and a fall bloom in 2008. Non-metric Multidimensional Scaling analysis of TRFLP profiles from samples collected along a ten station transect in the middle GSB from May to October (2008) and May to December (2009) revealed that while plankton community structure could change dramatically from week to week, in the same week both 18S and 16S community structure were relatively uniform across the entire transect. Nevertheless, differences in the 18S community between stations generally increased as the geographical distance between two samples increased. The 16S community showed less spatial heterogeneity. Ordination analysis of the two year data set suggested that *A. anophagefferens* was important in shaping the community structure. During 2008, the diversity of the 18S community measured with four biodiversity indices was significantly lower for samples collected during brown tides than samples collected from non-bloom conditions, while there was no significant difference for the 16S community. In general,

the 18S community was more diverse in 2009 than in 2008 while the opposite was the case for the 16S community.

According to the picoalgal niche hypothesis of Sieracki and colleagues, the phytoplankton niche is open to picoalgae in Long Island waters in early summer. The hypothesis states that *Synechococcus* usually dominates the picoalgal niche, but at times when the net growth rate of *Synechococcus* declines, similarly sized picophytoplankter *A. anophagefferens* becomes dominant. A reciprocal temporal pattern of the relative abundance of *Synechococcus* and *A. anophagefferens* was indeed observed in this study. Additionally, local similarity analysis revealed negative associations between *A. anophagefferens* and four picophytoplankton: *Synechococcus*, *Bathycoccus*, *Picochlorum* and *Picocystis*. The picoalgal niche is, therefore, proposed to be expanded to include all the above picoalgae, among which *Bathycoccus* was unique because it was the only species negatively associated with all other picophytoplankton in the non-brown tide year. In late July the picoalgal niche was taken over by *Synechococcus* and pico-chlorophytes, probably due to water temperature becoming too high for *A. anophagefferens*. Therefore, the earlier timing of *A. anophagefferens'* increase (spring) than *Synechococcus* (summer) in 2008 and the higher relative abundance of *Bathycoccus* in 2009 than 2008 especially in spring implies that *Bathycoccus*, instead of *Synechococcus*, may compete with *A. anophagefferens* for the picoalgal niche in springs of Long Island waters. Both water temperature and selective grazing of zooplankton on other picophytoplankton may have favored *A. anophagefferens* forming a fall bloom. One piece of evidence that *A. anophagefferens* might be under less grazing pressure than other phytoplankton is from local similarity analysis, where *A. anophagefferens* was negatively associated with only one group of grazers while other phytoplankton always had multiple significant associations with grazers and parasitoids such as *Telonema*, *Amoebophrya* and *Pirsonia*, which have not been reported for GSB previously. Turbulent conditions caused by wind mixing may have also helped *A. anophagefferens* outcompete other phytoplankton in 2008.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation concludes my PhD study, an extraordinary journey I could not have completed on my own. My three advisors Dr. Josephine Aller, Dr. Jackie Collier and Dr. Paul Kemp have been my resource of knowledge, insight, encouragement as well as critical comments. Jackie, in particular, spent enormous time in helping me improve the written document. I am also grateful to my committee members, Dr. Bob Cerrato, Dr. Gordon Taylor and Dr. Edward Delong, for their suggestions and advices from the beginning of this study to the completion of this dissertation. This work could not have been done without their intellectual support, moral encouragement and kind consideration.

This study was originally supported through the Ecosystem Based Management in Great South Bay, NY by NY Department of State Division of Coastal Resources. Dr. Jackie Collier, Dr. John True, Dr. Mark Fast and Undergraduate Biology (especially Dr. Marvin O'Neal, Dr. Peter Gergen and Dr. Dale Deutsch) at Stony Brook provided me with research work and teaching positions to pull through the financially difficult times. I am extremely grateful to Dr. Darcy Lonsdale and Dr. Jackie Collier for providing the financial support to conduct the field work. My upbeat fellow graduate student Marianne McNamara and the intrepid boat captain Mark DeAngelis have made the laborious field work so much fun and fruitful. My special thanks go to my boyfriend Zhenrui Cao, who went out with me on the water of Great South Bay during the coldest sampling days.

Dr. Charles Flagg, Brad Furman and Dongming Yang helped me tremendously with my understanding of wind's effect on plankton community in Great South Bay. Dr. Chris Gobler, Shelagh Palma and Dr. Yingzhong Tang generously shared their data and lab cultures with me. Dr. Shian-Ren Liu and Dr. Mike Doall in MEAD lab taught me how to use the automatic DNA sequence analyzer and I also want to thank Karolina, Garshaw and Ning for helping me in the lab.

I am fortunate enough to have worked in a loving environment like SoMAS with all the friendly faculty, staff and fellow students. I also always have my friends (whose names I do not

even want to bother to mention) around to share all the excitement, joy and disappointment with me. I am grateful to the whole process of getting a PhD because it taught me a lot on how to be a better person.

Finally, my deepest gratitude goes my dearest parents, who have given my life and brought me up the way I am. They offered me what they have, taught me how to be a content person and live a fulfilling life. For that, I am forever indebted. I also want to say thank you to my sister for growing up together with me and always understanding me. Although I have only spent three weeks with my then 10-month old niece, her smile is stuck in my mind and always makes me strong. My last notes are dedicated to my boyfriend Zhenrui, who has given me the most wonderful times in my life. His unselfish love and forgiveness constantly touches my life; his intellectual and moral support is essential to the completion of my dissertation.

Chapter 1 Introduction

Method development in isolation and characterization of pure cultures gave rise to the Golden Age of Microbiology in the early 1900's (Marsh 1999). Later investigations of microbial communities in many aquatic habitats, however, have revealed that less than 1% of the cells that can be counted by microscopy are easily cultivated in the laboratory (Ward et al. 1990, Amann et al. 1995). Although we have gained a good understanding of the life cycle and metabolism of cultivated microbes, we know very little about those that have not yet been cultivated. Additionally, many of the interactions among microbial groups in ecosystems are lost in pure cultures and so have escaped our attention. Since knowledge about the ecology of uncultivated microbes and their interactions provides critical information for us to understand the function of microbial communities in their natural environment, the goal of modern environmental microbiologists has become to understand the structures and dynamics of natural microbial communities and to resolve the environmental factors that govern, for example, their distributional patterns (Tiedje & Stahl 2002). Understanding how microbial communities are distributed in any given environment and why particular patterns exist is important for helping us predict how marine ecosystems may respond to environmental change that is underway at an unprecedented pace due to global urbanization.

This dissertation explores the spatial and temporal patterns of the planktonic microbial (both eukaryotic and prokaryotic) community in Great South Bay (GSB), a shallow coastal embayment along the south shore of Long Island, NY, USA, from late spring through fall over a two year period. These two years provided contrasting GSB plankton communities, with one year featuring a major bloom of the pelagophyte *Aureococcus anophagefferens* known as 'brown tide'. The major questions addressed by this dissertation are: How does the plankton community in GSB vary with space and time? How does the presence of the brown tide bloom affect the rest of the plankton community? What are the important planktonic microbial interactions in GSB, and which of them may be important for the development of brown tide blooms? The following part of this chapter will review recent work on microbial community characterization in aquatic environments, introduce the study site, describe the characteristics of *A. anophagefferens* and the ecological impact of brown tide blooms, and lastly, expand on the questions which will be addressed in this dissertation and their importance to the field of microbial ecology.

Background

It is generally believed in community ecology that strong spatial and temporal patterns in the distribution and abundance of taxa are governed by environmental processes, including both biotic and abiotic factors (Putman 1994, Elser & Sterner 2002), and that these processes link ecosystem function to biological diversity (Loreau et al. 2001, Hooper et al. 2005). This influence of environmental factors on microbial community structure has been recognized as the hypothesis that "everything is everywhere, and the environment selects" (de Wit & Bouvier 2006, O'Malley 2007). However, there is debate (e.g., (Fenchel & Finlay 2004)) as to whether microorganisms exhibit similar patterns of biogeography, revealed as nonrandom patterns in the distribution of biodiversity over space and time (Martiny et al. 2006), as do macroscopic organisms. Understanding the spatial and temporal patterns of marine microbial communities is especially challenging due to the fluid nature of marine environments and our limited sampling capacity in oceanographic research. In fact, in spite of an explosion of data generated by a variety of new techniques (DeLong & Karl 2005, Gilbert et al. 2011), most of the ocean is under-sampled and microbial oceanographers are still faced with the challenge to make useful generalizations from spatial/temporal snapshots. If our purpose is to describe patterns of community distribution and to determine the underlying force(s) contributing to the patterns observed, we then need to sample with fine enough spatial/temporal resolution to match the rate at which microbial communities change.

Temporal patterns in planktonic microbial communities

It is not fully understood how rapidly microbial communities respond to environmental changes (the answer is likely to be different for different ecosystems), and most studies are still conducted at sampling frequencies compatible with logistical constraints. Many studies of temporal changes in marine plankton communities have been conducted on a monthly timescale. For example, using automated ribosomal intergenic spacer analysis (ARISA) of samples collected monthly over the course of 4.5 years, Fuhrman and colleagues found that bacterioplankton community composition off the California coast was highly predictable seasonally (Fuhrman et al. 2006). Similarly, high resolution 16S rRNA tag pyrosequencing of samples collected monthly from 2003 to 2008 also revealed seasonally repeatable bacterial community structure at a station in the western English Channel (Gilbert et al. 2012), and that environmental variables explained most of the seasonal variance in bacterial community

composition. Off the New Jersey coast, a monthly study from 1995 to 1996 using Terminal Restriction Fragment Length Polymorphism (TRFLP) revealed that bacterial communities could be clearly separated by sampling time (Nelson et al. 2008). While there appear to be clear patterns of microbial community composition on monthly and longer timescales and the community structure can be highly predictable (Fuhrman et al. 2006, Gilbert et al. 2012), the rate of change in microbial communities in different oceanic systems is probably variable. In coastal environments such as lagoons, microbial community change is likely to be much faster than in the open ocean due to their smaller area, shallower depth, closeness to terrestrial input and limited water exchange with the open ocean. A good example demonstrating the necessity to conduct research with greater temporal resolution in coastal lagoons comes from a weekly flow cytometric analysis of the pico-eukaryotes in West Neck Bay, Long Island, NY, where a short-lived (less than 2 weeks) bloom caused by *Ostreococcus*-like algae was captured (O'Kelly et al. 2003) in June, 2001, for the first time in that area. Monthly or less frequent sampling is likely to miss such events.

Spatial patterns in planktonic microbial communities

Given the highly dynamic and fluid nature of marine systems, many studies of spatial variation in marine plankton community structure have been conducted along obvious environmental gradients likely to be important in structuring plankton communities. For example, Crump et al. (2004) studied the bacterioplankton community composition along the salinity gradient of the Parker River Estuary and Plum Island Sound in northeastern Massachusetts and found a unique bacterioplankton community was associated with intermediate salinities in summer and fall, when the average bacterial doubling time was much shorter than water residence time (Crump et al. 2004). Similarly, but over a much larger spatial scale, a global survey of *Synechococcus* and *Prochlorococcus* reinforced the idea that there is clear spatial partitioning of cyanobacterial linages in the four major ocean biomes: Polar, Coastal Boundary, Trade Wind and Westerly (Zwirglmaier et al. 2008). Therefore, distributional patterns of microorganisms are often found along distinct environmental gradients and on a variety of spatial scales.

Variation in microbial community structure over spatial scales without apparent environmental gradients has been much less explored and the currently available studies are

mostly on soil samples. Cho and colleagues found a negative relationship between genetic similarity of fluorescent *Pseudomonas* strains and geographic distance among 38 pristine soil samples collected from 10 sites on four continents (Cho & Tiedje 2000). A study of soil samples collected from 1 m to 100 km apart in arid Australia discovered, using distance-decay curves (community similarity versus geographical distance), that eukaryotic microbial communities were less similar with increasing distance between samples (Green et al. 2004). With the help of molecular techniques to characterize microbial communities, the above and additional studies suggest that microbial communities become less similar as geographical distance between them increases, particularly in soil samples over geographical distances as small as 0.1 km (Noguez et al. 2005, Robeson et al. 2011). In marine environments, it has been found that bacterial plankton diversity demonstrates latitudinal patterns over large spatial scales from tropical to polar regions (Fuhrman et al. 2008). However, differences in microbial community structure over small (kilometer) spatial scales and in the absence of dramatic environmental gradients have not been well studied in marine environments.

The challenges of concurrent examination of spatial and temporal patterns

Some field studies have looked at both spatial and temporal variation in hopes of getting a clearer picture of microbial community structure. For example, using TRFLP, Treusch and colleagues analyzed vertical and monthly patterns of bacterial communities at the Bermuda Atlantic Time-Series (BATS) site between 1991 and 2004 (Treusch et al. 2009). Non-metric multidimensional scaling ordination identified four distinct bacterial community compositional modes: a post-mixing spring bloom community, a stratified surface community, a deep chlorophyll maximum community, and a stratified upper mesopelagic community, which reflected changes in both light and nutrient regimes as a result of temporal and spatial variation driven by winter mixing and summer stratification. In some other cases, especially studies looking at large-scale spatial variation, data interpretation suffers from the fact that spatial variation is difficult to distinguish from temporal variation when samples are collected at different locations at different times. For example, TRFLP analysis of samples collected from 40 stations along a 15400 km transect in the Pacific Ocean from August to November 2003 indicated geographic separation of microbial communities (Baldwin et al. 2005). However, temperature was suggested as an important variable for the geographical segregation observed, so sampling from August to November while crossing the globe was a compounding factor.

While admitting this pitfall, the authors stated that the time window was already as short as logistically possible.

Therefore, given the desirability of studying both spatial and temporal patterns of microbial communities, sampling should optimally be conducted in such a way that changes in the two dimensions can be examined separately. One potential way to separately examine spatial and temporal variation is to sample over a restricted geographical distance over a long period of time. In fact, information on spatial variations of planktonic microbial communities at kilometer distance intervals without obvious environmental driving factors is scarce probably because it is often assumed that microbes are highly dispersible and within such short distances community structure would be similar. Understanding spatial variation over small distances is, however, of particular interest since first, microorganisms 1 km apart are comparable to terrestrial plants half the globe apart; and second, from the practical perspective, temporal variation is less likely to be a compounding factor for community variation seen in samples collected from geographically nearby locations within a short period of time.

Microbes, not just bacteria

Simultaneous investigation of eukaryotic and prokaryotic (both Bacteria and Archaea) components of the microbial community using molecular techniques has been much less frequent than studies focusing only on the prokaryotic community. This in part reflects the backgrounds of the people who started doing community characterization work using molecular tools. It might also be due to the greater amount of prokaryotic DNA in natural samples and the availability of databases of prokaryotic sequences such as the Ribosomal Database Project. Additionally, taxonomic identification based only on morphology is more useful for eukaryotic microorganisms due to their larger sizes and greater morphological complexity than prokaryotes. Consequently, the vast diversity of pico-eukaryotes has only relatively recently begun to be revealed using molecular methods (Diez et al. 2001, Moon-van der Staay et al. 2001). The well-accepted microbial loop concept describes trophic interactions among small eukaryotes and prokaryotes: heterotrophic Bacteria and Archaea utilize organic material released by phytoplankton (eukaryotes or prokaryotes), and are fed upon by microzooplankton such as ciliates and heterotrophic flagellates. It is perhaps a fair statement, that this framework of microbial interactions exists with big black boxes representing different groups of

microorganisms. It is the many specifics within these black boxes that are missing. As our ability to understand microbial diversity has increased with the help of molecular tools, characterization of microbial communities including both eukaryotes and prokaryotes has become feasible and makes it possible to unravel microbial interactions that are both ecologically and biogeochemically important.

Study Site

Great South Bay

Great South Bay (GSB) is a shallow coastal lagoon on the south shore of Long Island, New York, USA (Fig. 1.1.) with an average depth of only 2 m (Schubel et al. 1991). Except for a subtle salinity gradient sometimes, there are generally not well defined latitudinal environmental gradients across the ~10 km north-south axis of GSB. A bar-built estuary, GSB was formed in the past few thousand years by rising sea level and continues to change as sediments are deposited inside the embayment and the barrier islands evolve (Schubel et al. 1991). GSB extends approximately 40 km on its long axis, adjoining South Oyster Bay on its western end and Moriches Bays at the eastern end. Two narrow inlets connect GSB with the open ocean, Fire Island Inlet in the west and Moriches Inlet in the east. The fresh water input is relatively minor with a significant fraction (11%) coming from direct inflow of ground water through the seabed (Schubel et al. 1991). Water exchange between GSB and the ocean is the result of tidal forcing and wind (Schubel et al. 1991). Because of the narrowness of the inlets, although tidal currents are strong within the inlets, they dissipate quickly away from the inlets and are negligible in most of GSB. Water exchange with the Atlantic Ocean is principally driven by wind and characterized by flow through the entire Bay from one end to the other along the south shore of Long Island (Wong & Wilson 1984, Schubel et al. 1991, Wilson et al. 1991, Vieira & Chant 1993).

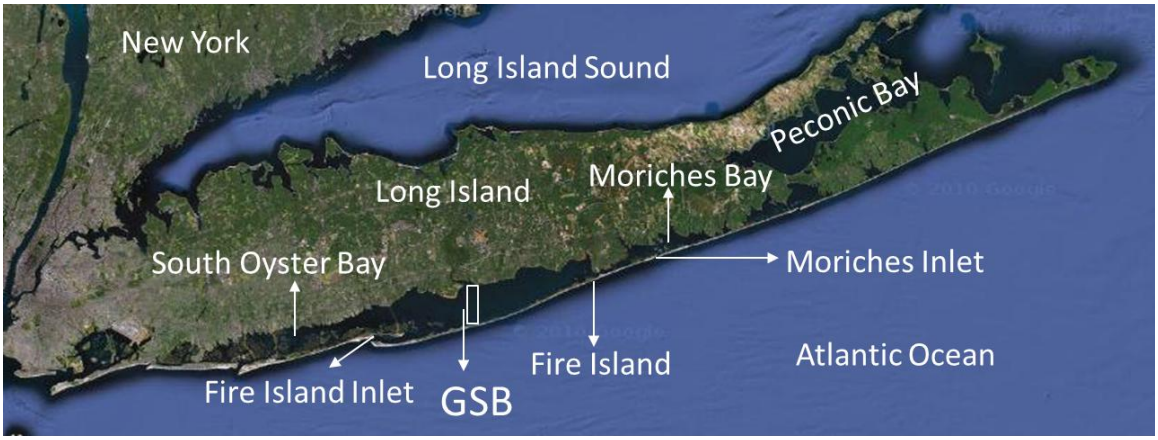


Fig. 1. 1. The location of the Great South Bay (GSB), on the south shore of Long Island, New York, USA. The rectangle roughly indicates the location of the 10-station sampling transect. A magnification of the exact study site is provided in Equation 1Fig. 3.1. in Chapter 3.

Brown Tides and their impact on GSB

From 1970 to 1978, GSB supported the largest hard clam (*Mercenaria mercenaria*) industry in the United States (LoBue & Bortman 2011), but landings of hard clams started to drop in 1979, and by 1984 were only half of the peak values. Although overharvesting is believed to be responsible for the rapid decline in hard clam biomass, harmful algal blooms known as 'brown tides' caused by the minute (~2 µm) non-motile pelagophyte *Aureococcus anophagefferens* are suggested to have contributed to the failure of hard clam recovery (Bricelj et al. 2001, Greenfield & Lonsdale 2002). *A. anophagefferens* was originally assigned to the algal class Chrysophyceae (Sieburth et al. 1988), but based on morphological, ultrastructural, and molecular data was later assigned to a new class, *Pelagophyceae*, as well as its own order, family and genus (Andersen et al. 1993). In 1985, the first *A. anophagefferens* bloom occurred in GSB. Since then, *A. anophagefferens* blooms have sporadically reoccurred in GSB, and have also appeared in bays in New Jersey, Delaware, and Maryland in the United States, Saldanha in South Africa, and recently the Bohai Sea, China (Cosper et al. 1987, Sieburth et al. 1988, Smayda 1990, Anderson et al. 1993, Probyn et al. 2001, Nuzzi & Waters 2004, Kong et al. 2012). In addition to its impact on bivalves, intensive growth of *A. anophagefferens* significantly decreases light penetration through the water column and is suggested to have caused large scale die-offs of eelgrass (*Zostera marina*) (Dennison et al. 1989, Dennison et al. 1993). Since seagrass beds are

important nursery places for some fish and shellfish, *A. anophagefferens* has also indirectly affected shellfish populations by reducing their juvenile recruitment.

The impacts of brown tide blooms on pelagic residents of GSB are less clear. We do not yet fully understand which plankton complete with and consume *A. anophagefferens*. The 'picoalgal niche hypothesis' suggests that *A. anophagefferens* and *Synechococcus* may compete with each other when a picoalgal niche develops in Long Island bays during the transition from a spring community composed mainly of diatoms and dinoflagellates to a summer community dominated by small algal cells (Sieracki et al. 1999, Sieracki 2001). According to this hypothesis, under normal conditions *Synechococcus* dominates the picoalgae size class, but if the net growth rate of *Synechococcus* declines, the picoalgal niche opens up for other algae of similar cell size, such as *A. anophagefferens*. Consistent with this hypothesis, a temporal shift in dominance between *Synechococcus* and *A. anophagefferens* was observed during 2000 in Quantuck Bay (a Long Island embayment east of GSB) (Sieracki et al. 2004). Due to the difficulty in identifying and monitoring other picoalgae, however, it is still largely unknown how picophytoplankton such as *Picochlorum atomus* (synonym *Nannochloris atomus*), identified in earlier studies as a dominant species in GSB (Ryther 1954, Lively et al. 1983), interact with *A. anophagefferens*. Microzooplankton will feed on *A. anophagefferens* (Caron et al. 1989) but there is evidence that overall grazing rates are reduced when brown tide is present (Lonsdale et al. 1996). Using microscopic image analysis of heterotrophs, Sieracki et al. (2004) suggested that the dominant pelagic grazers of *A. anophagefferens* were probably heterotrophic protists ranging in size from 7 to 20 μm, which would include larger microflagellates, heterotrophic dinoflagellates and non-loricate choreotrich ciliates. As useful as information on size classes of grazers is, interactions between specific grazer species and *A. anophagefferens* are poorly defined in the natural environment in large part due to the difficulty of accurately identifying and quantifying nano-zooplankton using microscopy.

The effect of brown tide blooms on the planktonic community is not yet well understood. It has been argued that any phytoplankton species can impact a specific marine ecosystem (e.g. cause anoxia) as well as other living marine resources (loss of biodiversity) when it persistently dominates the phytoplankton community (Zingone & Enevoldsen 2000). A meta analysis of global biodiversity surveys of phytoplankton and zooplankton species using the Shannon

diversity index, which considers both richness and evenness, suggested that phytoplankton communities were most diverse at an intermediate level of biomass and least diverse during massive blooms (Irigoien et al. 2004). While the same relationship between zooplankton biomass and zooplankton biodiversity was also discovered, a relationship between phytoplankton diversity and zooplankton diversity was not found (Irigoien et al. 2004). Using TRFLP analysis of 18S rDNA, Vigil et al. investigated microbial biodiversity from May, 2003 to August, 2004 in 6 estuarine ecosystems off the northeast U.S. coast historically prone to brown tide blooms (Vigil et al. 2009). During the 15 month sampling period, 4 to 16 samples were taken from each of the 6 estuaries, of which only 2 samples were affected by a brown tide bloom. Shannon and Simpson biodiversity indices for the 2 brown tide bloom samples were not significantly different from the remaining 64 samples. Unfortunately, this comparison between unmatched numbers of samples from bloom and non-bloom conditions makes for a weak conclusion, especially given the fact that the biodiversity baseline in each of these different estuarine systems was unknown and may initially have been very different. It remains unclear how brown tides affect plankton diversity.

Method and objectives

Our ability to distinguish different microorganisms has to come before we can start to unravel the factors which influence their distributional patterns in the environment. Current understanding of prokaryotic phylogeny is based on the highly conserved small subunit rRNA (SSU or 16S rRNA) as originally proposed by Carl Woese, who used comparative sequence analysis of 16S rRNA to study the relationships among the prokaryotes (Woese 1987). Since then, SSU rRNA genes (rDNA) have become the universal biomarker and the criteria by which Bacterial groups are identified (Giovannoni & Rappe 2000). Compared to culture-based methodologies, 16S rRNA-based phylogenetic surveys have revealed wide-spread distribution of previously unknown microbial groups including members of the Archaea (Delong 1992) and Bacteria (Giovannoni et al. 1990). As noted previously, since not all protists can be easily identified using morphological characteristics, molecular tools like 18S SSU rRNA gene (rDNA) sequence surveys have revealed novel eukaryote diversity in ecosystems as diverse as polar waters, sea ice, and tropical deciduous forest soil (Darling et al. 2000, Norris & de Vargas 2000, Moon-van der Staay et al. 2001, Gast et al. 2004). As a result, genomic technologies have gained broad application in characterizing naturally occurring microbial communities.

In this study, community fingerprinting by TRFLP was carried out for both 18S rDNA and 16S rDNA to provide a parallel understanding of temporal and spatial patterns of eukaryotic and prokaryotic members of the planktonic microbial community in GSB throughout summer and autumn during two consecutive years, 2008 and 2009. Approximately equal numbers of samples were collected during both brown tide bloom and non-bloom periods in 2008 for community comparisons. Identification of Terminal Restriction Fragments (TRFs) from the abundant planktonic microbes was also achieved by cloning and sequencing rDNA amplicons. The weekly (occasionally biweekly) sampling conducted in this study generated datasets suitable for statistical analysis that would not have been possible with more temporally sparse datasets. For example, Local Similarity Analysis (LSA) of the weekly time series produced a parameter, the Local Similarity (LS) score, whose value represents the strength of the similarity in temporal pattern between a pair of TRFs. Spatial variation of microbial communities was examined at ~1 km intervals along a north-south transect composed of 10 stations across the middle of the GSB (Fig. 1.1.). Specifically, this dissertation will examine the following:

1. What are the spatial and temporal patterns of planktonic microbial community composition and structure in GSB?
2. Are plankton communities in GSB more different as the geographic distance between them increases?
3. What are the important biotic and abiotic interactions in the planktonic community in GSB?
4. What are some of the important factors involved in brown tide bloom initiation, persistence and cessation?
5. How does the brown tide bloom affect the spatial/temporal pattern of the planktonic microbial community?
6. Does the brown tide bloom caused by *A. anophagefferens* affect planktonic community biodiversity?

Procedures used in establishing the TRFLP protocols are described in Chapter 2, followed by a description of the planktonic microbial community structure in the presence of a brown tide bloom in Chapter 3, and in a non-brown tide bloom year in Chapter 4, a comparison of bloom and non- bloom conditions in Chapter 5, and a summary chapter with future perspectives in Chapter 6.

Chapter 2 Development of a Terminal Restriction Fragment Length Polymorphism (TRFLP) analysis protocol

Introduction

Terminal Restriction Fragment Length Polymorphism (TRFLP) is a DNA fingerprinting technique with unique strength for monitoring changes in microbial community structure due to its sensitivity, throughput, and reproducibility (Moeseneder et al. 1999, Osborn et al. 2000, DeLong & Karl 2005, Schutte et al. 2008, Heidelberg et al. 2010, Gilbert et al. 2012). As its name indicates, TRFLP measures the size polymorphism of terminal restriction fragments (TRFs) carrying a detection marker (usually a fluorescent tag). Most frequently, the procedure starts from polymerase chain reaction (PCR) amplified fragments of small subunit (18S or 16S) rDNA produced from community DNA using fluorescently labeled universal primers. The resulting mixture of PCR products is then digested with one or more restriction enzymes, usually with four base-pair recognition sites, generating fluorescently labeled TRFs of different sizes. Separation, detection and quantification of these fluorescently labeled TRFs are done by running the restriction digest products on an automated DNA sequencer along with a DNA fragment size standard. Since size differences among TRFs reflect differences in the sequence of amplified rRNA genes, TRFs of different sizes theoretically represent different phylotypes, or, operational taxonomic units (OTUs).

There are, however, a variety of potential technical problems associated with using TRFLP in characterizing microbial communities. First, any problems associated with PCR reactions may confuse the interpretation of TRFLP results. In particular, partial PCR products that do not proceed to full length may be single-stranded at restriction enzyme recognition sites, rendering them unavailable to the restriction enzyme, generating ‘pseudo-TRFs’ and underestimating the abundance of the TRF expected from the full length PCR amplicons (Egert & Friedrich 2003). Post-PCR enzymatic treatment has been used to remove partial PCR products (Egert & Friedrich 2003, 2005). Mung bean endonuclease digests the single-stranded DNA parts of amplicons while Klenow fragment, the large subunit of the DNA polymerase I of *E. coli*, fills in complementary bases for the single-stranded part of PCR products. Although both have been found to be effective in eliminating pseudo-TRFs from TRFLP profiles, mung bean nuclease digestion causes underestimation of the relative abundance of amplicons affected by pseudo-TRF formation because it destroys them, while Klenow repairs partial PCR products by converting single-stranded stretches of amplicons to double-stranded molecules (Egert & Friedrich 2003, 2005). Additionally, ExoSAP (comprising the hydrolytic exoenzymes Exonuclease I and Shrimp

Alkaline Phosphatase) in theory removes leftover primers, single stranded DNA produced during PCR, and dNTPs, and is commonly used to clean up PCR products for sequencing. A comparison of the three enzymatic treatments, however, has not been reported.

The second technical issue with using TRFLP in characterizing natural microbial communities is that it is difficult to choose what restriction enzyme to use. Given that it is not possible to know all the sequences representing every organism in a natural community, it is impossible to know all TRFs that will be produced using any particular restriction enzyme for a natural environmental sample. It is therefore difficult to choose a particular restriction enzyme that can effectively distinguish all, or even most, of the different organisms in the community. In this study, our ability to answer many of the scientific questions raised in Chapter 1 relies on identification of the TRF representing the brown tide species *Aureococcus anophagefferens* and it is therefore a priority to choose an enzyme that distinguishes the TRF representing *A. anophagefferens* from those of other species.

Thirdly, there is always a difference between the length of a predicted TRF obtained using *in silico* digestion and its observed TRF, known as TRF drift (Kaplan & Kitts 2003). In this study, to get an estimate of TRF drift, TRFLP was carried out on PCR products from 8 eukaryotic microorganisms whose 18S rDNA sequences are known, and the lengths their observed TRFs were compared with their predicted TRFs. The 8 species were selected to cover as wide a taxonomic range as possible, including 7 different classes (see Materials and Methods).

Lastly, even with an optimal restriction enzyme (the choice of which depends on the scientific questions asked), different species in a natural community are likely to share the same TRF length. In fact, both this and TRF drift contribute to the difficulty of identifying observed TRFs by matching them with TRFs predicted from organisms with known sequences. In Chapters 3 and 4, to more confidently identify observed TRFs, clone libraries were constructed using the same community DNA and PCR primers as used for TRFLP analysis. Clones were randomly picked for sequencing, identified using Basic Local Alignment Search Tool (BLAST) against GenBank, and *in silico* digestion was performed to match the predicted TRFs from these sequences with the observed TRFs. Although instances of common TRFs from different organisms were indeed observed in this study, TRFs with high relative abundance always

matched predicted TRFs frequently found in clone libraries, making identification of at least the relatively abundant TRFs reliable.

Details of cloning and sequencing will be described in Chapter 3. For this chapter, three experiments were designed to address the first three technical issues raised above, to examine reproducibility of the TRFLP method and to evaluate whether the TRFLP protocol could reveal expected patterns of similarity among artificially constructed communities of the TRFLP protocol. In Experiment 1, the effectiveness of using mung bean nuclease, Klenow fragment and ExoSAP to remove partial PCR products was compared. Experiment 2 investigated TRFLP reproducibility using six restriction enzymes, where the restriction enzyme generating the least variation among replicate samples was considered to produce the most reliable data. In Experiment 3, *in silico* digestion of 8 marine eukaryote 18S rDNA sequences was conducted to pick a restriction enzyme that generates a unique TRF for *A. anophagefferens*. Subsequently, the predicted TRFs from *in silico* digestion were compared with those observed in the TRFLP profiles obtained from the 8 pure cultures, and TRF drift was calculated. Additionally, DNA extracts from the 8 cultures were mixed to create samples with different degrees of compositional overlap, and these were analyzed by TRFLP. The ordination (see Data analysis in Materials and Methods) pattern of these different DNA samples was examined to evaluate whether the TRFLP protocol developed in this chapter is able to reveal expected dissimilarities among microbial communities.

Materials and Methods

Experiment 1: Using enzymes to treat partial PCR products

Overview: Effectiveness of three enzymes (mung bean, Klenow and ExoSAP) in treating partial PCR products was compared.

1. **Sample collection and processing**

Surface water was collected using a clean bucket from Shinnecock Bay, Long Island, NY in October, 2007. 70 to 100 ml sea water was filtered through each of several replicate 47 mm diameter, 0.22 μm pore size polycarbonate membranes (Millipore, Billerica, MA). Each membrane filter was immersed in 2X lysis buffer (Countway et al. 2005) and kept at -80 ℃ until processed.

2. DNA extraction

DNA extraction was based on Countway et al. (2005) with some modifications. In brief, the membrane filters frozen in 2X lysis buffer were thawed and subjected to three rounds of 30-second vortex and 5-minute water bath incubation at 70 ℃. The lysate was separated from the membrane filters using a syringe and transferred to a clean 15 ml falcon tube. Equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol solution was added to the lysate, the mixture vortexed and then spun at 10400 rcf in an Allegra 64R Centrifuge (rotor C1010, Beckman Coulter, Brea, CA) for 3 minutes at room temperature to separate the aqueous phase from the organic phase. After a repetition of this step, two chloroform extractions were done to remove any residual phenol. DNA was then precipitated by adding 1X volume of 95% ethanol and 0.1X volume of 3M (pH 5.3) sodium acetate (NaAc) and incubating in the freezer overnight. All DNA extracts were quantified using Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen, Calsbad, CA) and diluted to concentration of 10 ng/µl for PCR.

3. Polymerase chain reaction

Partial 18S rDNA was amplified using a universal primer set (Table 2.1.) based on Countway et al. (2005). 6 reactions were run to ensure production of adequate product for subsequent procedures and to dilute "jackpot" artifacts that may have caused aberrant results in any particular PCR reaction. Briefly, each 50 µl reaction contained 0.25 µM of each primer, lX colorless buffer (Promega, Madison, WI), 2.5 mM $MgCl_2$ (Promega), 200 µM dNTPs (Promega), 300 ng/µl BSA (Sigma A7030, St. Louis, MO, (Kirchman et al. 2001)), 2.5 U of GoTaq Flexi polymerase (Promega), and 1 µl DNA template. Thermal cycling was carried out following the thermal protocol: one cycle at 95 ℃ for 2 min; 35 cycles at 95 ℃ for 30s, 55 ℃ for 30s, and 72 ℃ for 1 min; a final extension at 72 ℃ for 10min; and a hold at 4 ℃.

Table 2. 1. 18S rDNA universal primers

Primer	Sequence (5' to 3')	Target	Fluorescent Label	TRF Color Code	Reference
EukA	aacctggttgatcctgccagt	18S rDNA	FAM	B	Countway et al., 2005
Euk 570rev	gctattggagctggaattac	18S rDNA	HEX	G	Countway et al., 2005

4. Post-PCR enzymatic treatment

Replicate PCR reactions were pooled together. One quarter of the pooled PCR reactions was used for ExoSAP treatment following the manufacturer protocol: 50 µl pooled PCR reaction was mixed with 20 µl of ExoSAP-IT (USB) and the mixture was incubated at 37 ℃ for 15 minutes. To deactivate the enzymes, the mixture was incubated at 80 ℃ for 15 minutes. Three quarters of the pooled reactions was cleaned up using PCR Clean-Up System (Promega) and quantified using Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen). One third of the cleaned PCR product was set aside as the control to compare with enzyme-treated PCR products. The other two thirds of the cleaned PCR product was used for mung bean nuclease treatment and Klenow treatment. Protocols from the manufacturers were followed to carry out the enzymatic treatments. For mung bean nuclease treatment, 1 µg DNA was mixed with 5 units of enzyme and 10 µl of the provided 10X buffer in a total volume of 100 µl. The mixture was incubated at 30 ºC for an hour and the reaction was stopped by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The DNA was then purified by ethanol precipitation. For Klenow treatment, cleaned PCR product was mixed with Klenow Fragment Exonuclease Minus (Promega) at 1 µg DNA: 1 Unit Klenow ratio. The reaction volume was 100 µl and contained 40µM of each dNTP and 20 µg/ml acetylated BSA. The mixture was incubated at room temperature for an hour, after which Klenow was deactivated at 75 ℃ for 10 minutes.

5. Fragment analysis

At this point, the PCR products treated four ways (PCR cleaned up just by column, PCR cleaned up by column and treated with mung bean nuclease, PCR cleaned up by column and treated with Klenow, and PCR treated with ExoSAP) were ready to be loaded on the DNA sequencer. No restriction digestion was done for this experiment. After enzymatic treatment in the previous step, a mixture of 0.05X volume (1 µl) of glycogen and 0.1X volume (2 µl) of 3 M NaAc was added to each reaction. Next, 2.5X volume (50 µl) of 95% ethanol was added and the mixture was spun at 4 ℃ at 19300 rcf in an Allegra 64R Centrifuge (rotor F3602, Beckman Coulter, Brea, CA) for 15 minutes. The supernatant was carefully removed and 100 µl of 70% ethanol was added before a final spin at 19300 rcf at 4 ℃ for 10 minutes. After the supernatant was removed following the last spin, the DNA pellet was dried in the fume hood for 15 to 20 minutes before being re-suspended in 20 µl water. 0.5 µl GeneScanTM 600 LIZ Size Standard (Applied Biosystems, Carlsbad, California) was mixed with 10 µl HiDi and 1 µl DNA resuspension was added to the mixture, which was loaded on a 3130XL ABI DNA Analyzer (Applied Biosystems, Carlsbad, California).

6. Data analysis

Raw data were exported from Peak ScannerTM (Applied Biosystems, Carlsbad, California) into Excel files. Peaks shorter than 30 bases are likely to be from free dyes, primers, or primer dimers and were excluded from subsequent analysis. Throughout this dissertation, a TRF is represented by a number reflecting its length in bases and a letter reflecting its peak color. For example, 174B stands for the blue TRF that is 174 bases long (Table 2.1.). The percent contribution of all full length fragments (>575 bases) to the total peak area was summed and compared among 3 treatments and the control sample. The treatment with highest area contribution from full-length TRFs was considered the most effective treatment.

Experiment 2: TRFLP repeatability using 6 restriction enzymes

Overview: Identify the restriction enzyme with the highest reproducibility.

1. Sample collection and processing

Four 400-liter tanks were filled with Great South Bay water in May, 2008 (as part of a mesocosm experiment) and were not otherwise manipulated. 3 water samples were then taken from each of the four tanks (12 samples in total), filtered and stored as described in Experiment 1.

2. **DNA extraction**

Refer to Experiment 1.

3. **Polymerase chain reaction**

Triplicate PCR reactions were run using DNA extracted from each of the 12 filters following the protocol described in Experiment 1.

4. **Post-PCR enzymatic treatment**

Triplicate PCR reactions were pooled and the 12 pooled PCR products were treated with Klenow as described in Experiment 1.

5. **Restriction digestion**

PCR products were purified using PCR Clean-Up System (Promega), and quantified using Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen). 150 to 300 ng PCR products was digested with *Hae*III (Promega), *Hha*I (Promega), *Rsa*I (Promega), *Msp*I (Promega), *Sau*3AI (Promega) or *Taq*I (Promega) restriction enzyme at a ratio of 100 ng DNA: 1 Unit enzyme in 20 µl reactions according to the manufacturer's direction. For *Taq*I, there was only enough enzyme to digest 4 samples. Restriction digestion proceeded for 4 hours including a 15 min deactivation step by heating the reactions at 75 ℃. The digested PCR products were then precipitated and analyzed on the DNA sequencer following the description in step 5 of Experiment 1.

6. **Data analysis**

Raw data were exported into Excel as described in Experiment 1. Excel files containing raw data were subsequently re-arranged to include only file name, peak color, peak size, data point and peak area, the format that is required by online software T-REX (Culman et al. 2009) to do noise filtration and TRF alignment. T-REX uses an approach outlined in Abdo et al. (2006) to find true peaks and eliminate background noise. True peaks are identified as those whose area exceeds the standard deviation of all peaks (assuming zero mean). The procedure is then reiterated using only the peaks which were not identified as true peaks. The iterations continue until no new true peaks are found. TRF alignment is then performed based on the algorithm of Smith et al. (2005). Briefly, the shortest peak across all samples is identified. Peaks within 1 base are then identified and grouped into a TRF. The next smallest peak not falling into the first TRF is then identified, and subsequent iterations continue until all peaks are grouped into TRFs. Files

aligned this way by T-REX contained the area contribution of each TRF to the total TRF peak area in each sample.

TRFs contributing more than 5% to total TRF peak area averaged over the 12 samples (4 for *Taq*I) were considered to be major TRFs, and the average coefficient of variation of all major TRFs' area contribution was compared among the 6 restriction enzymes. The restriction enzyme with the smallest average coefficient of variation was considered to be the enzyme with highest reproducibility.

Experiment 3: TRF drift and TRFLP of artificial communities

Overview: TRF drift was estimated by comparing *in silico* digestion predicted TRFs and observed TRFs of 18S rDNA from 8 eukaryotic lab cultures. Mixed DNA samples with different degrees of compositional overlap were prepared using DNA extracts from 8 pure cultures, and their TRFLP profiles were compared using ordination analysis to evaluate if the pattern reflects their original composition.

1. **Sample collection and processing**

Cells from fresh cultures of *Aureococcus anophagefferens* (Pelagophyceae), *Chaetoceros gracilis* (Coscinodiscophyceae), *Dunaliella tertiolecta* (Chlorophyceae), *Emiliania huxleyi* (Haptophyceae), *Gymnodinium galatheanum* (Dinophyceae), *Isochrysis galbana* (Haptophyceae), *Pycnococcus provasolii* (Prasinophyceae), and labyrinthulid quahog parasite unknown (QPX) were collected and their cell densities were determined using a hemocytometer.

2. **DNA extraction**

For six of the cultures, duplicate samples of 10^8 cells were collected at 6000 rcf in an Allegra 64R Centrifuge (rotor F3602, Beckman Coulter, Brea, CA) for DNA extraction. For *Pycnococcus provasolii* and *Gymnodinium galatheanum*, there was only enough culture material to do one DNA extraction. Additionally, DNA extraction was done on duplicate cell culture mixtures comprising ~10^7 cells from each of the 8 species, referred to as CM (culture mix). After centrifugation, supernatant was removed and DNA was extracted from the pellet using DNeasy Blood & Tissue Kit (Qiagen, Alameda, CA). Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) was used to quantify the 16 DNA extracts. Five DNA mixes were prepared from the 8 pure culture DNA samples as follows: equal amount of DNA from all DNA extracts with QPX DNA (DM8) and without QPX DNA (DM7), equal amount of DNA from

Pycnococcus provasolii and *Gymnodinium galatheanum*, and equal volume of DNA from all DNA extracts with QPX DNA (EVDM8) and without QPX (EVDM7). The 9 DNA mixes (one DNA mix for *Pycnococcus provasolii* and *Gymnodinium galatheanum*, and duplicate DNA mixes of DM7, DM8, EVDM7 and EVDM8) were quantified using Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen, Calsbad, CA). 25 DNA samples (the aforementioned 16 DNA samples and 9 DNA mixes) were diluted to a concentration of 10ng/μl for PCR.

3. Polymerase chain reaction

Duplicate PCR reactions were run for the 25 DNA samples prepared as above following steps in Experiment 1.

4. Post-PCR enzymatic treatment

Duplicate PCR reactions were pooled and treated with Klenow as described in Experiment 1.

5. Restriction digestion

*Taq*I was used to do restriction digestion and DNA digests were run on the DNA sequencer following the steps described in Experiment 2.

6. Data analysis

Raw data were exported into Excel and noise filtration as well as TRF alignment performed as in Experiment 2. The aligned file was next transformed into a 'main matrix' for Non-metric Multidimensional Scaling (NMS) in PC-ORD (v5.32). A second matrix was also prepared to include sample identifications (and environmental parameters for later chapters). NMS is an ordination method well-suited to data that are nonnormal. It does not assume linear relationships among variables. Instead, it uses ranked distances and is compatible with any distance measure or relativization. Unlike methods such as principle component analysis and canonical correspondence analysis, which start with an assumptive model and subsequently see only configurations that fit a limited perspective (McCune & Grace 2002, Caron et al. 2009), NMS can reveal a much wider range of structures in the data. The limitation of applying NMS extensively in the past was its failure to find the solution with minimum stress and its slow computational speed for large data sets. With the increase of modern computing power these problems no longer exist and this is why McCune and Grace suggest that 'The future of

ordination is in exploiting computational power with iterative optimization methods such as nonmetric multidimensional scaling.' Autopilot mode with medium speed and thoroughness was used to do NMS ordination in PC-ORD. Sørensen (Bray-Curtis) distance was computed using PC-ORD because it performs well with ecological data where relationships between variables are nonlinear (McCune & Grace 2002).

Results

Experiment 1: Enzymatic Treatment of partial PCR products

The most dominant TRF of the control sample and the 3 enzymatic treated samples (collected from Shinnecock Bay) was 590 bases (Fig. 2.1.). Based on the TRF size distribution of the 4 samples (Fig. 2.1.), TRFs longer than 575 bases were considered full length PCR products and were tallied to compare among the 4 samples. The size range of the environmental sequences using the same primer set obtained from GSB was from 576 bases to 631 bases, which supported the '575 bases and above' grouping. The contribution to the total peak area of observed TRFs >575 bases was higher for both Klenow treatment (95.3%) and mung bean nuclease treatment (83.3%) than the control (82.7%). However, the difference between mung bean nuclease treatment and control PCR product (going through column cleanup only) was too small to be convincing. TRF 590B was present in all four treatments and was the most dominant full-length TRF (Fig. 2.1.). Mung bean nuclease treatment had TRFs that were not present in other treatments (584B and 587B) and had the fewest full-length TRFs. For the control, the remaining 17% of total peak area was contributed by TRFs along the whole size spectrum, while for mung bean nuclease treated PCR, the rest was from TRFs of about 30 ~ 40 bases, 60 bases, 80 bases, and 290~300 bases (Fig. 2.1.). ExoSAP treatment had the smallest area contribution from observed TRFs >575B mainly because three smaller TRFs (44B, 46B and 80B) contributed 31%, 34% and 18% to the total TRF peak area (Fig. 2.1.).

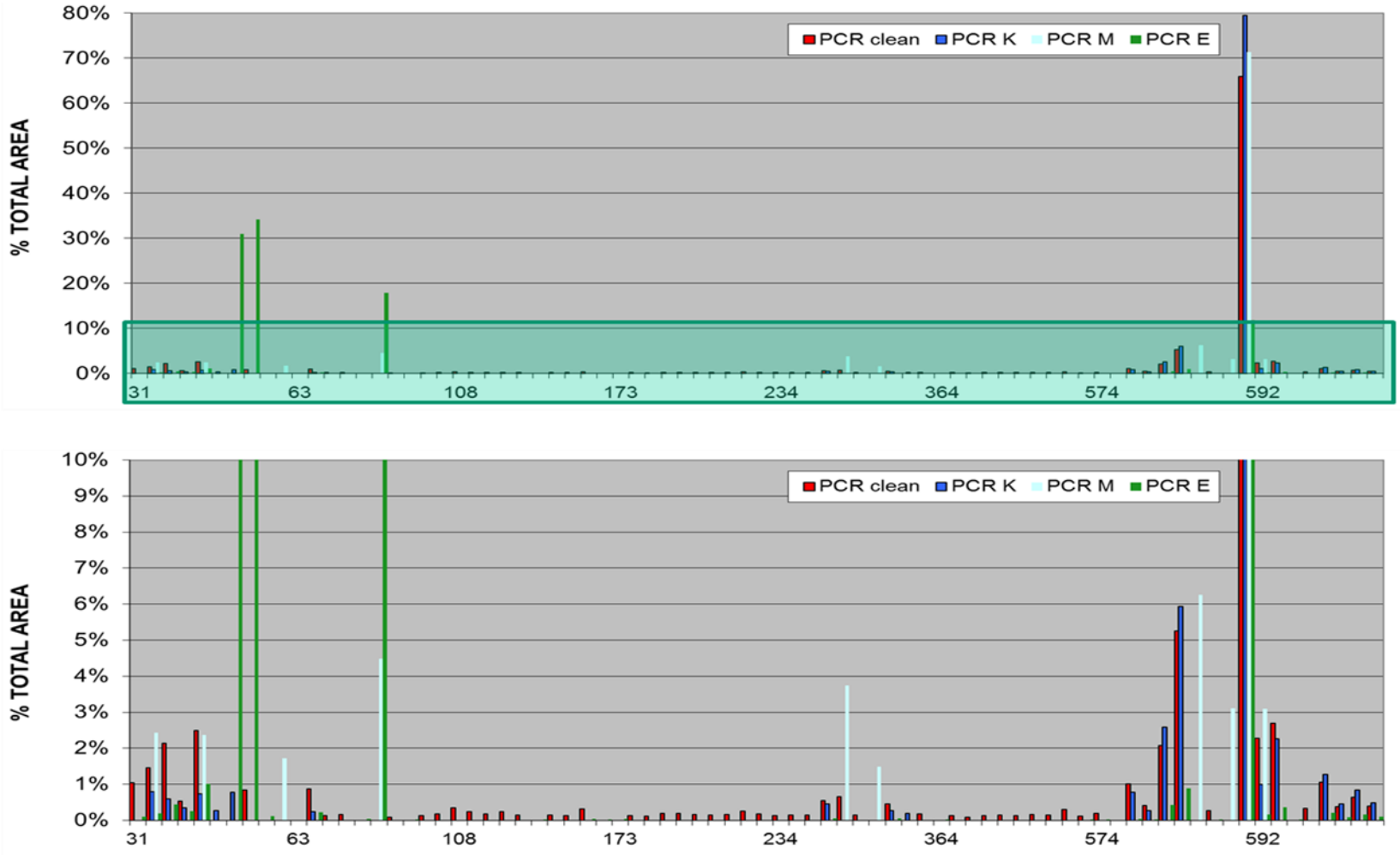


Fig. 2. 1. TRF size distribution of control PCR (PCR clean), Klenow treated PCR (PCR K), mung bean nuclease treated PCR (PCR M) and ExoSAP treated PCR (PCR E). Bottom panel is the zoomed-in view of the shaded area of the upper panel.

Experiment 2: TRFLP reproducibility

Area contributions of the major TRFs across the 12 parallel samples (collected from the mesocosm incubation tanks in GSB) were used to calculate the coefficient of variation (CV) (Table 2.2.). The restriction enzyme with the lowest CV was *Taq*I. Even when just considering the four samples digested by *Taq*I, CV for *Hae*III, *Hha*I, *Msp*I, *Rsa*I, and *Sau*3AI were 0.18, 0.20, 0.15, 0.32, 0.29, all higher than the CV for *Taq*I. *Sau*3AI had a much larger CV than other restriction enzymes and was excluded from further consideration. *Hha*I had the smallest number of peaks (suggesting least resolution capability) while *Rsa*I and *Msp*I generated a large percentage of peaks that were similar in size (Table 2.2.) so they were not considered further, either. *Hae*III had a small CV not much greater than *Taq*I and also generated many TRFs that were well separated.

Table 2. 2. Area contribution from major TRFs averaged over 12 samples (4 for *Taq*I) and their coefficient of variation for different restriction enzymes

*Hae*III (n=12)			*Hha*I (n=12)			*Rsa*I (n=12)		
Size	TRF%	CV	Size	TRF%	CV	Size	TRF%	CV
263B	5.2%	0.10	171B	8.6%	0.33	72B	33.5%	0.36
280B	27.1%	0.31	427B	32.4%	0.22	124B	7.9%	0.15
335B	6.4%	0.08	438B	33.4%	0.25	584B	26.3%	0.28
337B	11.1%	0.25				586B	7.0%	0.51
483B	17.2%	0.30						
592B	7.6%	0.18						
Average		0.20	Average		0.26	Average		0.26
*Msp*I (n=12)			*Sau*3AI (n=12)			*Taq*I (n=4)		
Size	TRF%	CV	Size	TRF%	CV	Size	TRF%	CV
238B	24.8%	0.33	259B	9.6%	0.84	73B	11.8%	0.12
369B	5.8%	0.09	285B	15.5%	0.45	174B	34.5%	0.10
376B	7.8%	0.18	296B	24.9%	0.28	203B	7.0%	0.06
379B	7.2%	0.11	592B	19.6%	0.47	376B	14.0%	0.28
381B	11.9%	0.22	594B	10.7%	0.60	530B	8.2%	0.05
479B	15.8%	0.25	603B	9.5%	0.66			
Average		0.19	Average		0.55	Average		0.12

Experiment 3: Evaluation of TRF drift and TRFLP protocol using artificial communities

1. *In silico* digestion

In silico digestion using the previously tested restriction enzymes (except for *Sau*3AI) and two additional restriction enzymes showed that *Taq*I had the greatest resolving power to distinguish the 8 species (Table 2.3.). Compared to blue fragments, the green end was more likely to produce similar-sized TRFs from multiple species. *Alu*I generated TRFs that were similar in length for *Aureococcus anophagefferens* and *Isochrysis galbana. Mse*I generated blue TRFs that were the same across all species. In fact, except for *Hae*III and *Taq*I, the other restriction enzymes generated TRFs of similar sizes (±3 bases) from several different species. However, *Hae*III did not have restrictions sites in two of the species tested. *Taq*I was the only enzyme that separated all species (except for the two prymnesiophytes *Emiliania huxleyi* and *Isochrysis galbana*) by blue TRFs differing in size by more than 3 bases.

Table 2. 3. Predicted full-length PCR product, and forward (Blue, B) and reverse (Green, G) TRF lengths using *in silico* digestion on 8 lab culture species with 7 restriction enzymes

	PCR (bp)	*Alu*I	*Hha*I	*Hae*III	*Mse*I	*Msp*I	*Rsa*I	*Taq*I
AA	605	164B/ 10G	441B/ 166G	283B/ 259G	44B/ 71G	242B/ 110G	78B/ 527G	177B/ 59G
CG	591	161B/ 10G	427B/ 164G	483B/ 108G	44B/ 71G	376B/ 110G	236B/ 355G	530B/ 59G
DT	591	161B/ 10G	429B/ 162G	216B/ 257G	44B/ 62G	221B/ 108G	128B/ 79G	317B/ 217G
EH	590	159B/ 10G	426B/ 166G	331B/ 259G	41B/ 244G	220B/ 213G	124B/ 79G	256B/ 219G
IG	593	162B/ 10G	429B/ 166G	334B/ 259G	44B/ 244G	223B/ 213G	127B/ 79G	259B/ 100G
GG	595	69B/ 10G	174B/ 161G	339B/ 256G	44B/ 70G	383B/ 210G	no cut	377B/ 58G
PP	588	161B/ 10G	424B/ 164G	no cut	44B/ 377G	274B/ 110G	126B/ 462G	180B/ 59G
QPX	592	108B/ 10G	431B/ 163G	no cut	44B/ 546G	378B/ 212G	no cut	294B/ 218G

AA: *Aureococcus anophagefferens*; CG: *Chaetoceros gracilis*; DT: *Dunaliella tertiolecta*; EH: *Emiliania huxleyi*; IG: *Isochrysis galbana;* GG: *Gymnodinium galatheanum*; PP: *Pycnococcus provasolii*; QPX: Labyrinthulid quahog parasite QPX

2. Predicted vs observed TRF for the 8 species

For *Taq*I digestion of the 8 species, the difference between the predicted and the observed TRFs, known as TRF drift (Kaplan & Kitts 2003), ranged from 0 for the blue (EukA) TRF of *Emiliania huxleyi* to 7 bases for the green (Euk570rev) TRFs of *Aureococcus anophagefferens* and *Pycnococcus provasolii* (Table 2.4.). The largest TRF drift for blue TRFs was smaller than that for green TRFs (5 bases for *Dunaliella tertiolecta* vs 7 bases for *Aureococcus anophagefferens* and *Pycnococcus provasolii*), although the fact that the green TRFs are closer to the smaller end of the size spectrum probably contributed to the larger TRF drift observed (Kaplan & Kitts 2003). As was already predicted by the *in silico* digestion, a clear advantage of the forward TRFs over the reverse TRFs was that forward TRFs were quite different among different species while several species generated reverse TRFs of overlapping observed sizes which makes comparing community structure with green TRFs difficult. It should be noted that in addition to one most abundant observed TRF, one or two other TRFs differing in size by up to 3 bases were also present. Except for the green fragments of *Gymnodinium galatheanum*, for

which the observed 52G and 55G added up to only ~half of the total area, the most abundant TRF and its close-by TRF(s) usually accounted for 80% or more of the total TRF area.

Table 2. 4. Observed and predicted forward (blue) and reverse (green) TRFs of the eight lab culture species

Strain	Observed blue TRF	TRF (%)*	Predicted blue TRF	Observed green TRF	TRF (%)*	Predicted green TRF
AA	174 B	57%	177B	52 G	79%	59 G
	176 B	31%		54G	9%	
CG	528 B	53%	530 B	53 G	78%	59 G
	527 B	19%		55 G	12%	
DT	312 B	67%	317 B	215 G	84%	217 G
	314 B	10%		217 G	6%	
	315 B	6%				
EH	256 B	73%	256 B	216 G	2%	
	258 B	10%		217 G	80%	219 G
				218 G	6%	
IG	256 B	64%	259 B	95 G	70%	100 G
	259 B	8%		97 G	14%	
GG	375 B	67%	377 B	52 G	44%	58 G
	377 B	8%		55 G	7%	
PP	177 B	71%	180 B	52 G	81%	59 G
	179 B	10%		54 G	6%	
QPX	291 B	59%	294 B	216 G	79%	218 G
	293 B	7%		217 G	5%	
	294 B	21%				

*: This value is the average of the duplicate TRFLPs. AA: *Aureococcus anophagefferens*; CG: *Chaetoceros gracilis*; DT: *Dunaliella tertiolecta*; EH: *Emiliania huxleyi*; IG: *Isochrysis galbana;* GG: *Gymnodinium galatheanum*; PP:*Pycnococcus provasolii*; QPX: Labyrinthulid quahog parasite QPX

3. Nonmetric Multidimensional Scaling (NMS)

Different species separated well from each other on the NMS biplot, with the most closely related pairs, *Emiliania huxleyi* and *Isochrysis galbana*, clustering together (due

essentially to their identical dominant TRF 256B) (Table 2.4. and Fig. 2.2.). All duplicate DNA extracts as well as duplicate DNA mixes clustered in pairs. Artificial communities constructed from different DNA mixes (CM, DM7, DM8, EVDM7, EVDM8 and PP_GG) ordinated in a way consistent with their composition. For example, PP_GG was positioned in between PP and GG; sample points representing DNA mixes without QPX (DM7 and EVDM7) were located further away from QPX than similar DNA mixes with QPX (DM8 and EVDM8).

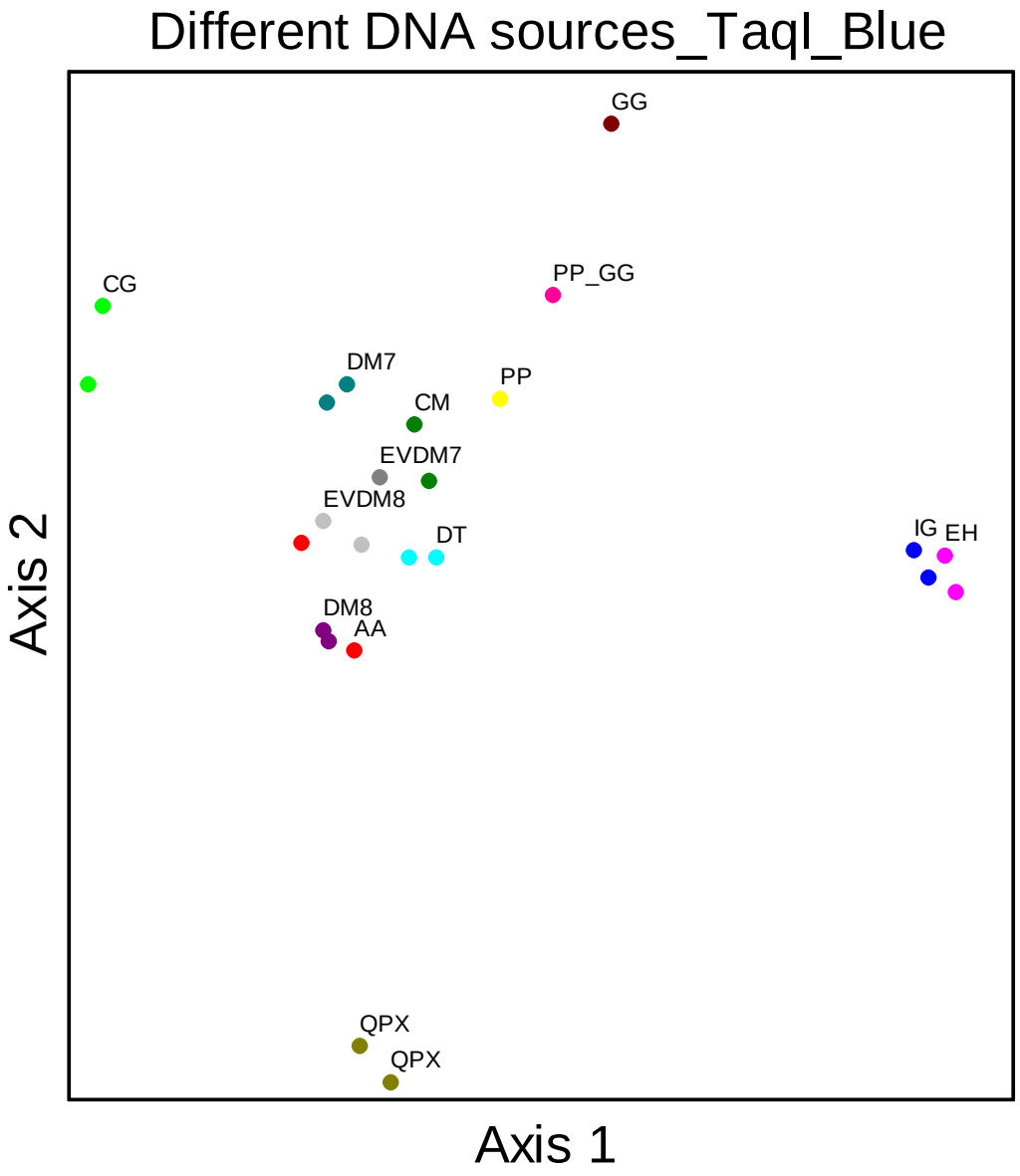


Fig. 2. 2. NMS ordination biplot of TRFLP profiles from pure culture DNA and different DNA mixes representing different artificial communities. AA: *Aureococcus anophagefferens*;

CG: *Chaetoceros gracilis*; DT: *Dunaliella tertiolecta*; EH: *Emiliania huxleyi*; GG: *Gymnodinium galatheanum* (1); IG: *Isochrysis galbana;* PP: *Pycnococcus provasolii* (1); QPX: Labyrinthulid quahog parasite QPX; CM: DNA extract of the culture mix; PP_GG: equal DNA mix of *Pycnococcus provasolii* and *Gymnodinium galatheanum* (1); DM7: equal DNA mix of all species except for QPX; DM8: equal DNA mix of all species; EVDM7: equal volume mix of all species except for QPX (1; one of the duplicates of EVDM7 was accidentally lost during processing); EVDM8: equal volume mix of all species.

Discussion

Among the 3 enzymes tested to treat partial PCR products, Klenow seemed to be the most effective. After mung bean nuclease treatment, area contribution from full-length TRFs was lower than after Klenow treatment. This agrees well with the working principle of mung bean endonuclease, which should convert partially single-stranded amplicons into shorter double-stranded DNA, resulting in underestimation of the relative abundance of the full length PCR product. Additionally, the shorter double-stranded DNA may become a source of pseudo-TRFs. The ExoSAP treatment seemed to generate numerous artifacts, reflected especially in the generation of pseudo-TRFs ~45 bases long, which were not present in the control sample and other enzymatic treatments. Since the sizes of the partial PCR products were distributed along the whole size spectrum above 30 bases (Fig. 2.1.), what Klenow treatment seemed to have achieved is filling in the missing bases for partial PCR products of various sizes and moving many of them up to the full-length, larger than 575 bases range.

The choice of the best restriction enzyme depends on the study objective. For the purpose of this study, where hundreds of microbial community compositions were to be compared in an ecosystem where information regarding *Aureococcus anophagefferens*'s dynamic in relation to the rest of the plankton community is critical, an enzyme that generates a unique TRF for *Aureococcus anophagefferens* was essential. *Taq*I appeared to produce a unique TRF for *Aureococcus anophagefferens* (considering the species included in this chapter and some other phytoplankton species tested with *in silico* digestion such as *Coscinodiscus radiatus* and *Rhodomonas salina*) and gave the lowest coefficient of variation of major TRFs among the 4 replicate samples. Using restriction enzyme *Mnl*I, another TRFLP study characterizing the planktonic eukaryotic community in estuarine ecosystems off the northeast U.S. coast achieved average standard deviation of 1.33% for all TRFs among 3 replicates, which was considered to be highly reproducible (Vigil et al. 2009). With *Taq*I, the current study established a TRFLP

protocol with average standard deviation as small as 0.90% for all TRFs among 4 replicates, suggesting that for the field samples collected in GSB, restriction digestion of a single sample should be representative without running replicates.

The TRFLP validation work using artificial communities with different degrees of compositional overlap showed clearly that the TRFLP protocol coupled with NMS analysis is capable of revealing expected dissimilarities among microbial communities. Although it seemed that there was not much separation among DNA mixes prepared in different ways (CMs, DMs, EVDMs and PP-GG, Fig. 2.1.), it does not mean that the relative peak area of TRFs representing different species is not important in determining the ordination. The distance on NMS ordination biplots does not reflect the original distance between different DNA mixes, but rather preserves their rank order. With compositionally very different single species included in the ordination analysis, the distance between sample points representing communities with mixed DNA is limited on the biplot.

The results generated in Experiment 3 show that first, the predicted TRFs can indeed be very different from the observed TRFs. Kaplan and Kitts (2003) described 5 bases difference as 'alarming' TRF drift, which was not only matched but also exceeded by the 7 bases difference for the green TRFs seen in this study (Table 2.4.). According to Kaplan and Kitts, TRF size and purine content both have important impacts on TRF drift. In this study, *Emiliania huxleyi* and *Dunaliella tertiolecta* had the largest TRF drift in their blue TRF fragment (5 bases). The difference of the observed TRF sizes of *Emiliania huxleyi* and *Dunaliella tertiolecta* was ~ 60 bases. Although Kaplan and Kitts did not specify how big of a TRF size difference would cause significant TRF drift, 60 bases was less than the TRF size difference between many other species pairs (*Aureococcus anophagefferens* and *Chaetoceros gracilis*, for example, had the largest TRF size difference of ~ 350 bases). Additionally, the purine contents of *Emiliania huxleyi* and *Dunaliella tertiolecta* were also very similar (Table 2.5.). In fact, all PCR products were labeled with FAM and HEX (Table 2.1.) and the samples were analyzed under the same conditions (run at the same time) on the DNA analyzer. This 0 vs 5 bases difference in TRF drift seemed more 'alarming' since it indicates that other factors besides TRF size, purine content, fluorescent dye and electrophoresis condition must have contributed to TRF drift. Assuming these other factors have to do with other features of the sequences of different species (which is beyond the scope of

this study), the variable TRF drifts of different species observed here make a strong argument that to identify observed TRFs for a natural community, just matching them to predicted TRFs from a general sequence database is not sufficient. This is the motivation for cloning and sequencing in this study, as briefly mentioned in introduction of this chapter. Additionally, this experiment also showed that one species may generate several TRFs with slightly different lengths (although one of them is likely to be more abundant than the others), which may be due either to slight sequence differences between 18S rRNA gene copies or *Taq* polymerase error during PCR. Either way, this and the variable TRF drift observed suggest that when analyzing complex natural communities, allowing TRFs within 2 or 3 bases to be lumped together is a realistic approach to process TRFLP data.

Table 2. 5. Purine content for the 18S partial rDNA sequences of the 8 culture species

	AA	CG	DT	EH	IG	GG	PP	QPX
(G+C)%	47.60	46.53	46.02	46.10	46.21	45.55	45.41	46.45

AA: *Aureococcus anophagefferens*; CG: *Chaetoceros gracilis*; DT: *Dunaliella tertiolecta*; EH: *Emiliania huxleyi*; IG: *Isochrysis galbana;* GG: *Gymnodinium galatheanum*; PP: *Pycnococcus provasolii*; QPX: Labyrinthulid quahog parasite QPX

Chapter 3 Planktonic microbial community structure in Great South Bay, NY in the presence of an *Aureococcus anophagefferens* (brown tide) bloom

Introduction

In 1985, the first brown tide bloom caused by *A. anophagefferens* occurred in GSB, and brown tides have reoccurred sporadically since then. A number of different mechanisms have been proposed to explain the initiation and persistence of brown tide blooms, but no single hypothesis has had complete success in predicting brown tide events and a comprehensive understanding of factors causing blooms is still lacking.

A. anophagefferens' heterotrophic capability, especially its growing advantage under low light condition, and its affinity for certain forms of N, such as NH_4^+, have been suggested to lead to its dominance in phytoplankton communities (Dzurica et al. 1989, Lomas et al. 2001, Gobler & Sanudo-Wilhelmy 2001a, MacIntyre et al. 2004, Taylor et al. 2006). The relationship between *A. anophagefferens*, nutrient availability and light, however, is not often straightforward. For example, despite of the suggested advantage DON over DIN providing with *A. anophagefferens* under low light conditions (Dzurica et al. 1989, Berg et al. 1997, Mulholland et al. 2002), *A. anophagefferens* in batch culture grew faster when given nitrate than glutamic acid under low light condition, although in semi-continuous culture *A. anophagefferens* grew faster when both nitrate and urea were present than when only nitrate was present (Pustizzi et al. 2004). Meteorological processes, particularly the combination of a prolonged drought period that may reduce estuarine flushing rates and a subsequent pulse of heavy rainfall that delivers specific micronutrients from the watershed, may have facilitated the growth and retention of *A. anophagefferens* in a few coastal embayments including GSB (Cosper et al. 1987, Cosper et al. 1989, Smayda & Fofonoff 1989). Additionally, years with less groundwater discharge seemed to favor the growth of *A. anophagefferens* in Peconic Bay, perhaps because it may preferentially utilize dissolved organic N for growth and groundwater has a higher inorganic N to organic N ratio compared to other external N sources such as river runoff and atmospheric deposition (Laroche et al. 1997).

A. anophagefferens was suggested to compete with the picophytoplankton *Synechococcus* in picoalgal niche (Sieracki et al. 1999, Sieracki 2001). Sieracki and coworkers hypothesized that in Long Island Waters where a picoalgal niche forms in summer, *A. anophagefferens* can occasionally outcompete *Synechococcus* and become the dominant species in the picoalgal niche and develop into bloom concentrations. Additionally, reduced grazing rates by planktonic protozoa have been suggested to contribute to the formation of brown tide blooms, as was observed in the contrasting conditions in GSB and Quantuck Bay in 2002, when the higher grazing rate in GSB was associated with a non-bloom condition while the lower grazing rate in Quantuck Bay was associated with a bloom condition (Gobler et al. 2004b). Caron and colleagues, however, found that instead of deterring protistan grazing activities, the higher growth rate of *A. anophagefferens* than *Synechococcus* could initiate a brown tide bloom (Caron et al. 2004). With the ability of planktonic protozoa to graze on *A. anophagefferens* well established even under bloom conditions (Gobler et al. 2002, Caron et al. 2004), understanding the interactions of specific protozoan grazers with not only *A. anophagefferens* but also other phytoplankton is needed to better evaluate whether and how preferential protozoan grazing on different phytoplankton may help initiate or maintain a brown tide bloom.

Although it is generally recognized that brown tides have negative impacts on shellfish (e.g. hard clams, oysters, and scallops) and submerged aquatic vegetation (Gastrich et al. 2003), except for *A. anophagefferens'* competing effect on similarly sized *Synechococcus* (Sieracki et al. 1999, Sieracki 2001),its influence on the rest of the plankton community is not as well studied. While cell concentration of *A. anophagefferens* was found to have a significant negative effect on ciliate population growth (Lonsdale et al. 1996), *A. anophagefferens* did not always seem to affect zooplankton abundance (Caron et al. 1989 , Boissonneault-Cellineri et al. 2001). Lack of a comprehensive understanding of how *A. anophagefferens* influences the planktonic microbial community (composition, distributional pattern and diversity) is due in part to under-sampling on both spatial and temporal scales and the focus of many studies on only particular plankton species or functional groups, often with limited taxonomic resolution.

Given the presence of a brown tide bloom in GSB in 2008, this study set out to use both 18S and 16S rDNA to characterize the plankton community using the reproducible Terminal Restriction Fragment Length Polymorphism (TRFLP) technique described in Chapter 2. Weekly

sampling was conducted to investigate variations in the plankton community with season and with changing abundance of *A. anophagefferens* over a spatial scale of ~10 km with a resolution of ~1km. With the help of statistical analysis, the following questions will be addressed in this Chapter:

1. How does the planktonic microbial community change with time? What were the important environmental parameters shaping microbial community structure in GSB when there was a brown tide bloom?

2. Is there spatial heterogeneity at the scale of a few km? i.e., will planktonic microbial communities become more different as the geographic distance between them increase?

3. Does brown tide bloom have any effect on the biodiversity of plankton microbial communities?

4. What are the potential biotic and abiotic interactions in the planktonic community? What mechanisms do these interactions suggest for the progression of the brown tide bloom in GSB, 2008?

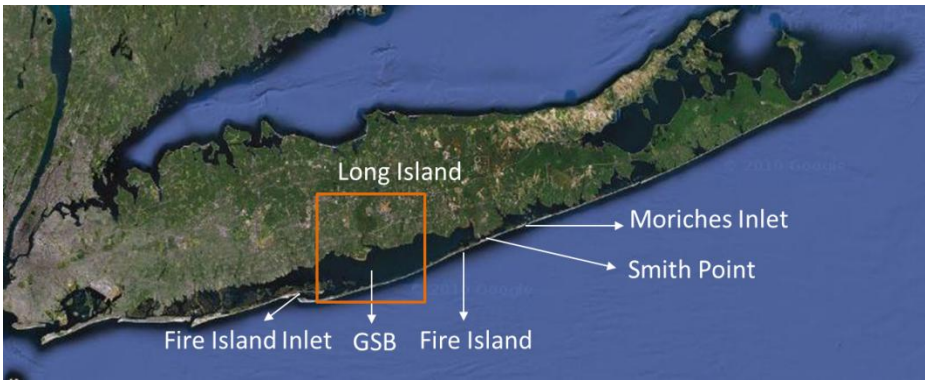
Materials and Methods

Sample collection and processing

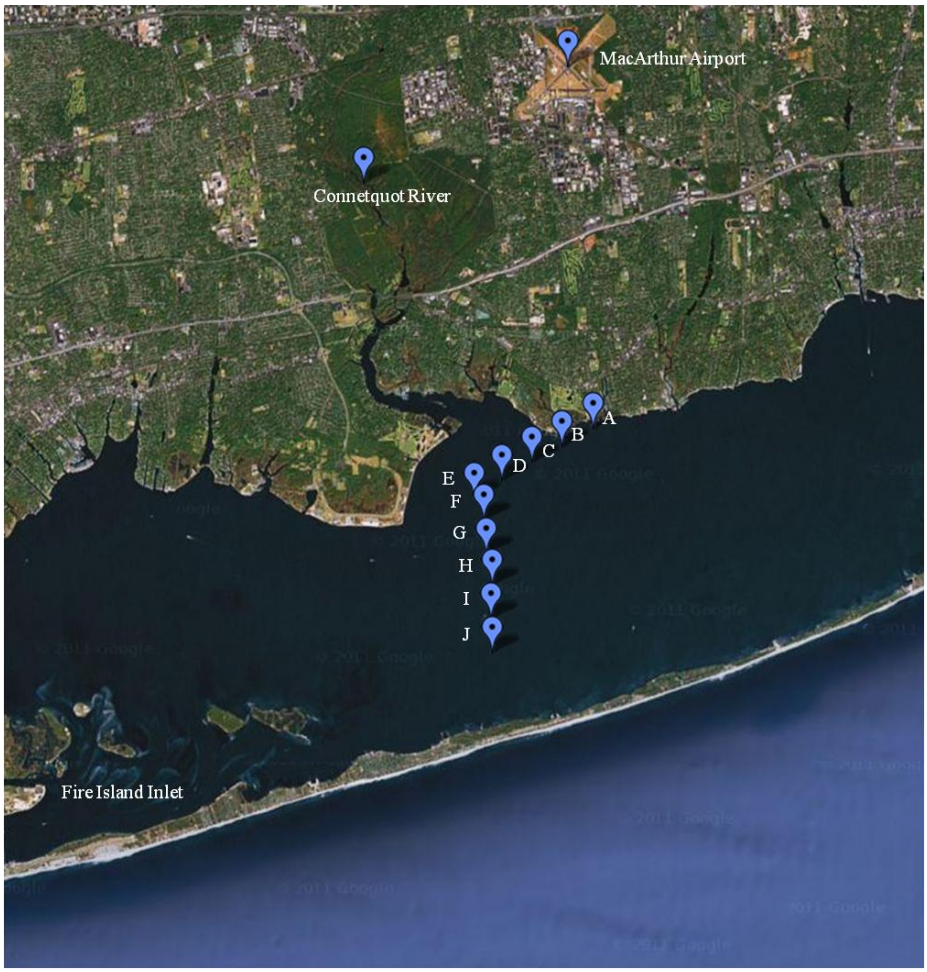
Surface water samples were collected in Great South Bay on the south shore of Long Island, New York (Fig. 3.1.A.) in 2008. There were 10 stations along the sampling transect (A to J, Fig. 3.1.B.) with ~1 km between adjacent stations; in some weeks all 10 stations were sampled but in other weeks only A to E were (Table 3.1.). The sampling season extended for 23 weeks from May 8 to October 9. Samples from 19 weeks were collected, with no samples collected in weeks 16, 18, 20 and 22 (Table 3.1.). In the rest of this paper, samples will be identified by the combination of a number representing the week of collection and a letter representing the station of collection.

Table 3. 1. Sampling date, the corresponding week number, stations sampled and samples analyzed by TRFLP in GSB, 2008

	Week number	Samples collected	Samples analyzed (18S)	Samples analyzed (16S)
8-May	1	E	E, E	E, E
15-May	2	A to E	A to E	A-C,E
22-May	3	A to J	A to F, J	A-E,G,J
28-May	4	A to E	A, E	A, E
5-Jun	5	A to J	A, E	A, J
12-Jun	6	A to E	A, E	A
19-Jun	7	A to J	A, E, J	A, E
27-Jun	8	A to E	A, E	A, E
3-Jul	9	A to J	A, E, J	A, E, J
10-Jul	10	A to E	A, E	A, E
17-Jul	11	A to J	A, E, J	A, E, J
24-Jul	12	A to E	A, E	A, E
31-Jul	13	A to J	B, E, J	E, J
7-Aug	14	A to E	A, E	A, E
14-Aug	15	A to J	A, J	A, J
28-Aug	17	A to J	A, E, J	E, J
11-Sep	19	A to J	A, D, J	D, J
25-Sep	21	A to J	A, E, J	A, E, J
9-Oct	23	A to J	A to J	A to J



A



B

Fig. 3. 1. A. Long Island, NY
B. Square area in A showing the ten sampling stations (A-J). Wind speed was recorded at MacArthur Airport and river discharge was recorded for the Connetquot River

Temperature, salinity, and dissolved oxygen were measured for surface water at each station using a YSI-85 multiparameter probe (Yellow Spring, OH). See "NMS ordination" in "Data analysis" below for other environmental parameters included in the multivariate analysis. About 25 liters of surface water was collected using a clean plastic bucket and 2 liters was immediately transferred to brown Nalgene® bottles and kept on ice until filtered. 70 to 100 ml was filtered through each of several replicate 47 mm diameter, 0.22 μm pore size polycarbonate membranes (Millipore, Billerica, MA). Each membrane filter was immersed in 2X lysis buffer (Countway et al. 2005) and kept at -80^{o}C until processed.

TRFLP, cloning and sequencing

DNA extraction, PCR, restriction digestion and data processing were done following the protocol described in Chapter Two. Briefly, phenol/chloroform was used to extract the genomic DNA. Universal primers labeled with fluorescent tags were used to amplify partial regions of 18S rDNA and 16S rDNA (Table 3.2.). PCR products were cleaned up using the Wizard® SV Gel and PCR Clean-Up System (Madison, Wisconsin, Promega) and treated with Klenow Fragment, Exonuclease minus (Promega) (Egert & Friedrich 2005). After another cleanup with the Wizard® SV Gel and PCR Clean-Up System, restriction enzyme *Taq*I (Promega) was used to digest the PCR products. The fragmented PCR product was precipitated using ethanol, resuspended in Milli-Q water and loaded on a 3130XL ABI DNA Analyzer (Applied Biosystems, Carlsbad, California). Only the terminal, fluorescently tagged fragments are detected, and are called terminal restriction fragments (TRFs). During electrophoresis, the signal of each TRF passing the fluorescence detector is captured as a peak on an electropherogram, where the X axis represents the size of the fragment (length in bases) and the Y axis represents the fluorescence intensity. Raw data from the electropherograms were exported into an Excel spreadsheet using Peak Scanner^{TM} Software v1.0 (Applied Biosystems). This Excel spreadsheet was fed into online software TREX (Culman et al. 2009) for noise filtration and TRF alignment. Since there were always more forward 18S TRFs (represented by a number indicating the fragment length and a letter B for the peak color) and forward 16S TRFs (represented by a number indicating the fragment length and a letter Y for the peak color), only the forward TRFs were analyzed. The reverse TRFs were occasionally used when assigning a taxon to a TRF (see the next paragraph) was aided by examining both forward and reverse TRFs.

Table 3. 2. 18S and 16S rDNA universal primers used in PCR

Primer	Sequence (5' to 3')	Target	Fluorescent Label	TRF Color Code	Reference
EukA	aacctggttgatcctgccagt	18S rDNA	FAM	B	(Countway et al. 2005)
Euk 570rev	gctattggagctggaattac	18S rDNA	HEX	G	(Countway et al. 2005)
341F	cctacgggaggcagcag	16S rDNA	TAMRA	Y	(Liu et al. 1997)
926R	ccgtcaattcctttragttt	16S rDNA	Rhodamine Red X	R	(Liu et al. 1997)

To identify the organisms represented by TRFs, cloning and sequencing of both 18S and 16S rDNA PCR products were performed. The same 18S and 16S primers without fluorescent tags were used in PCR, and cloning was performed using the pGEMT-Easy kit (Promega). Plasmids were purified using Wizard SV Plus Miniprep kit (Promega). Sequencing reactions were carried out using BigDye Terminator v3.1 (Applied Biosystems), cleaned up using Sephadex G50 (GE Healthcare, Buckinghamshire, UK) columns, and run on the 3130XL ABI DNA Analyzer. All sequences were identified by BLAST against GenBank (June, 2012) and aligned in Bioedit V7.0 (Hall 1999) with sequences from named organisms. In some cases, the best hits of a query sequence included very different organisms (usually 2), in which case the best hits always aligned at different locations of the query sequence. Each part of the query sequence was then run in BLAST separately and the query sequence was considered a chimera if the best hits for the different parts were the same as in the previous step. Chimeras were excluded from further analysis. *Taq*I restriction sites were located using the BioEdit restriction mapping utility and used to calculate the predicted TRF sizes for each sequence. To identify the observed TRFs, their sizes were compared to the predicted TRFs and matching was performed to allow TRF drifts up to 5 bases (Table 2.4.). Usually the more abundant a TRF's relative area contribution was, the more sequences generating this TRF were retrieved. Identification of the relatively more abundant TRFs were performed first, making the identification of the less abundant TRFs more accurate. Although the best hit of a query sequence in GenBank sometimes

indicated a species name (Table A1.), genus names were often used to report TRFs' identity throughout this study because the exact species might not yet be represented in GenBank. For query sequences that have 99% or 100% similarity (%similarity in GenBank) to those deposited in GenBank, such as *A. anophagefferens* and *Chaetoceros calcitrans* (see Results), it is most likely that the query sequences represented these species.

Data analysis

1. Temporal variation of major TRFs

The contribution of each TRF to total peak area in each sample (referred to as TRF% in this paper; for example, area contribution of 174B to total TRF peak area is referred to as 174B%) was averaged over all samples. TRFs contributing more than 5% to total area overall were considered major TRFs and plotted against time to examine the temporal patterns of the most common and abundant TRFs in the community. Based on the observations in Chapter 2 (Table 2.4.) that usually more than one TRF with up to 3 bases difference is generated from the same species, observed fragments with similar length (in most cases, ±2 bases) were binned into a new (major) TRF and different taxa with predicted TRFs of similar sizes (in most cases, ±2 bases) were assigned to this new TRF. Lumping TRFs with similar sizes into major TRFs was only done for temporal variation analysis of major TRFs.

2. Nonmetric multidimensional scaling (NMS) ordination

To visualize both temporal and spatial patterns in structure of the microbial community including all TRFs, nonmetric multidimensional scaling (NMS) ordination was performed. The data matrix generated by TREX (see previously in 'TRFLP, cloning and sequencing') contained the contribution of each TRF to total peak area for every sample and was used as the main matrix for NMS ordination in PC-ORD (v5.32, MJM, Gleneden Beach, OR). The second matrix included sample ID and environmental parameters. Dissolved oxygen was always 100% saturation and was therefore not included in the multivariate analysis. To investigate the influence of fresh water input, discharge from the Connetquot River on the sampling day and the sum of river discharge for the prior 7 days (including the sampling day) were included in the second matrix. Water discharge from the Connetquot River was obtained from USGS WaterWatch (http://waterdata.usgs.gov/ny/nwis/dv?cb_00060=on&format=gif_meas&begin_date=2008-05-

01&end_date=2008-07-30&site_no=01306460&referred_module=sw, May, 2011). Precipitation on the sampling day and the sum of precipitation for the prior 7 days (including the sampling day) at Long Island MacArthur Airport were downloaded from http://www.wunderground.com (January, 2010) for comparison to river discharge, although not included in NMS ordination (see 'NMS ordination' in Results). Lastly, average wind speed and maximum gust speed on the sampling day at Long Island MacArthur Airport were retrieved from NOAA/National Climatic Data Center (http://cdo.ncdc.noaa.gov/qclcd/QCLCD?prior=N, Jan, 2011) and included in the second matrix.

There were 61 and 55 samples (Table 3.1.) for 18S and 16S TRFLP, respectively, in the main matrices. Although arcsine square root transformation is recommended for proportional data (Sokal & Rohlf 1995), this transformation of the main matrix generated the same outliers, defined as samples more than 2 standard deviations from the mean using the Bray-Curtis distance measure, as when the original dataset was used for both 18S and 16S communities. Additionally, the ordination pattern looked the same using the original and the transformed datasets. Therefore, the original data was used as the main matrix in NMS ordination. Square root transformation was performed on the quantitative environmental parameters including temperature, salinity, river discharge and wind speed. Relativization was then carried out by adjustment to standard deviation according to b=(x-x_mean)/S for each environmental parameter, where b is the relativized square root, x is the square root, x_mean is the mean of x and S is the standard deviation of x. This relativization places environmental parameters originally with different ranges on equal numerical footing (McCune & Grace 2002).

NMS was run in autopilot mode where the stress of multiple runs at each dimensionality with the real data was compared with the stress of multiple runs with randomized data. Known as a Monte Carlo test, this analysis indicates whether NMS extracted a stronger pattern of samples along different ordination axes than expected by random chance (indicating the reliability of the solution) and also why a particular dimensionality was chosen as the final solution by PC-ORD. Each additional dimension is considered useful if it reduces the final stress by 5 or more (McCune & Grace 2002). Also, at the chosen dimensionality the final stress for the real data must be lower than that for 95% of the randomized runs (reported as $p \leq 0.05$ in the Monte Carlo test) (McCune & Grace 2002). The best solution was visualized as biplots whose X and Y axes

were extracted from the main matrix and did not necessarily represent a gradient of any parameter.

The coefficient of determination (R^2) assesses the percentage of variance in the original data matrix represented by each axis in the NMS ordination solution and is calculated with a distance-based method by comparing the distances in the ordination space with distances in the original space. In community ecology, when there are 20 or more species or operational taxonomic units (TRFs in this case), having 50% of the variation on 2 axes often indicates the existence of an explainable pattern, although as little as 30% to 50% of variation on 2 axes can be useful (McCune & Grace 2002).

The Pearson's Correlation Coefficients (r) between the environmental parameters for each sample and the position of the samples on an ordination axis were calculated to reveal any linear relationships between the ordination axis and the environmental parameters. The square value of the correlation coefficient (r^2) expresses the percentage of variation in the positions of the samples on an ordination axis that is explained by the variable in question. In PC-ORD, only environmental variables with an r^2 larger than the cutoff value 0.2 are plotted by default. This is a conservative criterion considering that the sample size in this study is large enough that even a very small correlation coefficient can be statistically significant. Additionally, r was looked up in a critical value table for the Pearson's Correlation Coefficient to determine which environmental parameters were statistically significant at p=0.001. Probability of 0.001 was chosen because I want to focus on the parameters that are most predictably related to the community structure.

3. Spatial variation

To explore differences in spatial variation between weeks, the coefficient of variation (CV) of the major TRFs' area contribution among samples was calculated for each week. Weeks with spatially homogeneous samples were identified as weeks when all major TRFs had a CV smaller than 0.5. For other weeks, the CV of at least one major TRF was greater than 0.5 and usually larger than 1. All spatial samples (5 or 10 samples depending on the length of the transect) collected in the weeks with high CV were processed by TRFLP and a spatial variation data matrix was generated as described above and imported into PC-ORD to compute the Bray-Curtis distance between every possible pair of samples collected in the same week. Geographical distance between every possible pair of stations was also calculated. To examine whether

microbial communities were more different from each other as the geographical distance between them increased, Bray-Curtis distance vs geographical distance from all weeks was plotted and the slope was analyzed. This regression was also determined separately for each week; higher slopes of the regression lines represent more spatially heterogeneous weeks.

4. Biodiversity indices

To evaluate the influence of brown tide on the diversity of the plankton community, samples were divided into bloom (18S samples with c175B% >15%; 16S samples with c298Y% > 7.4%; letter 'c' before TRFs is explained in 'Major TRFs' in Results.) and non-bloom (18S samples with c175B% <15%; 16S samples with c298Y% < 7.4%) categories and richness (N, number of TRFs), Shannon's H', Simpson's D, and Pielou's J were compared between the two categories for both 18S and 16S TRFLP. See '*Aureococcus anophagefferens* bloom' in Results for why c175B %> 15% and c298Y %> 7.4% were used for 18S and 16S bloom cutoffs. H' was calculated as H'= $-\sum_{i=1}^{N} p_i * \ln\ p_i$, Simpson's D was calculated as D=1-$\sum_{i-1}^{N}(p_i)^2$ and Pielou's J was calculated as J= $\frac{H'}{logN}$, where N is richness and equals to the total number of TRFs in each sample and p_i is the contribution of the ith TRF to total peak area (Wang et al. 2008). Larger H' and D values correspond to more diverse communities and larger J values correspond to more even communities.

Although the DNA template for PCR and subsequent PCR products were quantified and controlled at each step to maintain consistency among samples, the ABI 3130XL sequencer can still take up different amounts of DNA during sample injection. Samples with less or more DNA loaded may have artificially low or high TRF richness (N in the above equations). To minimize this unequal loading effect, normalization was performed in the following way. First, the ratio between the total TRF peak area of each sample and the smallest total TRF peak area among all samples was calculated. This normalization coefficient for each sample was used to reduce all peak heights in that sample. The resulting peaks with height under 50 arbitrary fluorescence units, the default cutoff value in Peak Scanner, were subsequently deleted. The data matrix generated this way was the diversity data matrix used to compute Richness N, Shannon's H', Simpson's D and Pielou's J.

5. Local similarity analysis (LSA)

Associations among TRFs and between TRFs and environmental factors were examined by analyzing the local similarity correlation (Ruan et al. 2006). First, the 18S and 16S main matrices used in NMS analysis were combined with the second matrix. TRFs present in fewer than 20% of the samples were removed from the matrix to minimize the effect of rare TRFs. This generated a local similarity matrix (LSM) containing 55 rows representing 55 samples (with both 18S and 16S TRFLP data) and 96 columns representing 49 18S TRFs, 41 16S TRFs and 6 environmental parameters (temperature, salinity, Connetquot River discharge, sum of river discharge for the prior 7 days, wind speed and maximum gust speed). To investigate time delayed associations between TRF pairs as well as TRF-environmental factor pairs, the averages of each TRF's area contribution, temperature and salinity from all stations were calculated for the 19 weeks, generating a data matrix containing 19 weekly average samples with the same number of TRFs and environmental factors as the LSM, which will be referred to in the rest of the paper as Local Similarity Matrix-Temporal (LSM-T). Both LSM and LSM-T were normalized according to Ruan et al. (2006). In the local similarity analysis (LSA) for the LSM matrix, for every pair of columns of data, the product of the two entries with no time delay was calculated in R according to Ruan et al. (2006). In the LSA for the LSM-T matrix, for every pair of columns of data, the product of the two entries of all time points within a time delay of 4 (mostly about one month) was calculated in R according to Ruan et al. (2006). The maximal sum of the product for each pair of columns was subsequently divided by the length of the time series and defined as the Local Similarity Score (LS). LS is similar to the Pearson's Correlation Coefficient in linear regression, which was also computed in R for all pairs of transformed data. The higher the LS is for a TRF pair or a TRF-environmental factor (TRF-EF) pair, the more similar pattern they have (or the more opposite pattern they have if LS is negative). The sign of the correlation is determined by a sign function (Ruan et al. 2006).

Negative associations in LSA are generally understood as competition or predation while positive associations not only mean that two organisms co-vary but could also indicate symbioses or parasitism (Steele et al. 2011). To decide if an LS value was statistically significant, the p value was computed as follows. For each pair of data matrix columns, data values were permuted randomly, and LS was computed for the permuted data columns as described above. Data permutation and LS computation was repeated 1000 times (Ruan et al. 2006). The

proportion of permutations when LS of permuted data equaled or exceeded the absolute values of LS of the original data was assigned to p. With 4005 TRF pairs, 540 TRF-EF pairs and 15 environmental factor pairs in LSM, multiple testing is a concern. q values, defined as the proportion of occasions when a claimed significant LS was in fact not significant, were computed in R. q values for LS of LSM-T were calculated in the same way. All LS values among 90 TRFs and 6 environmental factors were visualized as networks using Cytoscape 2.8.2 (Shannon et al. 2003). Each TRF or environmental factor is represented by a node in the network, connected by a line (LS correlation) with another TRF or environmental factor that has a similar or opposite pattern. Data filtering was performed in Cytoscape to include only the statistically significant correlations in the networks presented in Results.

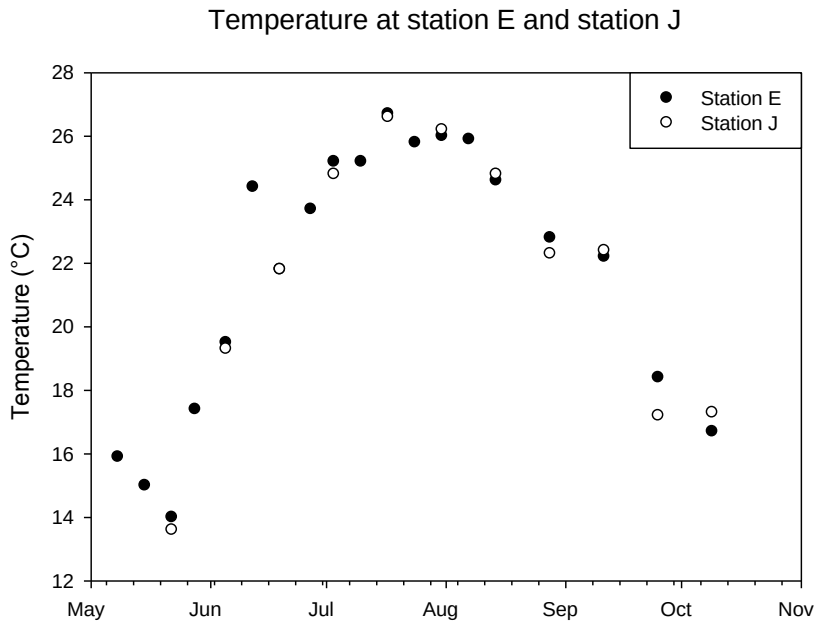
In the LSA analysis using the LSM-T matrix, although setting the predefined time delay at 4 allows recognition of significantly associated pairs with up to 1 month delay, correlations with a time delay of 4 have at most 15 points determining the relationship, and this makes it necessary to be cautious of the reliability of time-delayed associations. For example, according to LSA run using LSM, *Chaetoceros* was positively associated with temperature at p=0.001, which was supported by p value of 0.007 for Pearson Correlation (named as Ppcc later in this paper). The LSA run using LSM-T concluded, however, that *Chaetoceros* was negatively associated with temperature with a time delay of 4 weeks at p=0.005 while Pearson Correlation suggested that no significant relationship between *Chaetoceros* and temperature existed. Therefore, the original as well as the time-delayed associating pairs were plotted and Pearson Correlation coefficients were calculated to decide if the time-delayed association was reliable.

Results

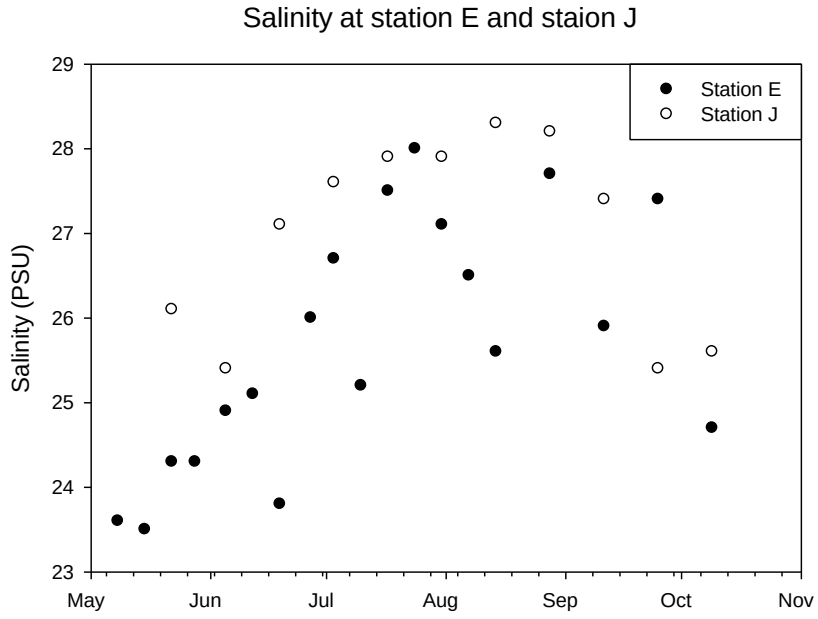
Environmental parameters

Temperature increased from 14 ℃ in May to almost 27 ℃ in July and then declined to ~17 ℃ at the end of the sampling period in October (Fig. 3.2.A). The increase in temperature from May to July was almost monotonic except for week 6 (June 12), when the temperature at station E was 25.1 ℃ and higher than both the previous week (station E=19.5 ℃ and station J=19.3 ℃) and the following week (both stations=21.8 ℃). Salinity followed a similar pattern of increase from a little over 23 in May to 28 in July followed by decline, but with greater variability between weeks and between stations (Fig. 3.2.B). For weeks when samples were collected from station A to J, salinity was always higher at J than E except for week 21 (September 25). The largest difference in salinity between station J and E was 3.3 and occurred in week 7 (June 19). Salinity followed temperature closely with a strong linear relationship between them (r=0.696) (Table 3.3.). 7-day sum of Connetquot River discharge followed the temporal pattern of river discharge on the sampling day (Fig. 3.2.D) with a strong linear relationship (r=0.881). The biggest discrepancy occurred during week 12 (July 24), when there was heavy precipitation on the sampling day (Fig. 3.2.C). The average wind speed and the maximum gust speed (Fig. 3.2.E) were linearly correlated with r being 0.815 and the highest average wind speed occurred in week 21 (September 25) at 14 mph.

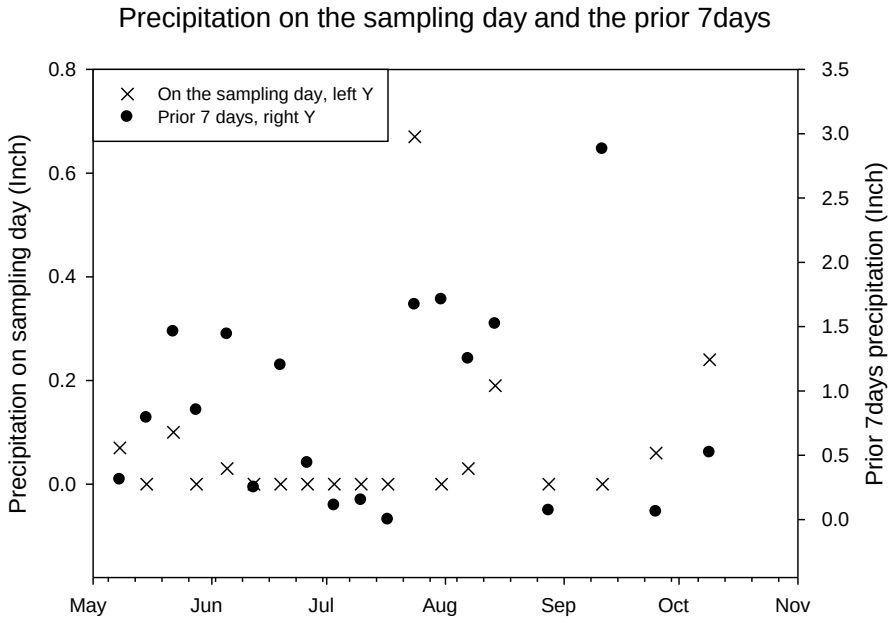
A)


Temperature at station E and station J
Station E
Station J
Temperature (°C)
28
26
24
22
20
18
16
14
12
May
Jun
Jul
Aug
Sep
Oct
Nov

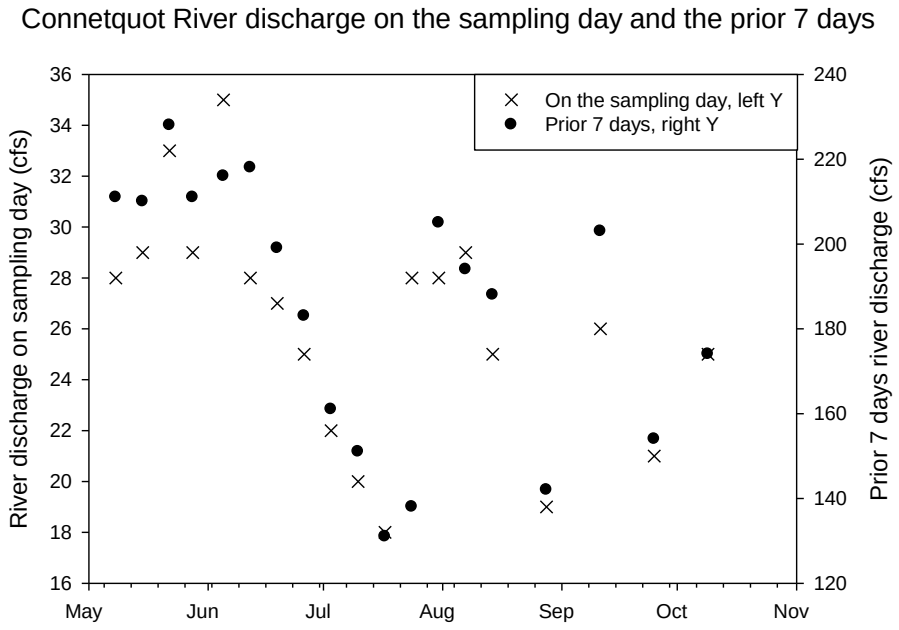
B)


Salinity at station E and staion J
Station E
Station J
Salinity (PSU)
29
28
27
26
25
24
23
May
Jun
Jul
Aug
Sep
Oct
Nov

C)


Precipitation on the sampling day and the prior 7days
On the sampling day, left Y
Prior 7 days, right Y
Precipitation on sampling day (Inch)
Prior 7days precipitation (Inch)
0.8
0.6
0.4
0.2
0.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
May
Jun
Jul
Aug
Sep
Oct
Nov

D)


Connetquot River discharge on the sampling day and the prior 7 days
On the sampling day, left Y
Prior 7 days, right Y
River discharge on sampling day (cfs)
Prior 7 days river discharge (cfs)
36
34
32
30
28
26
24
22
20
18
16
240
220
200
180
160
140
120
May
Jun
Jul
Aug
Sep
Oct
Nov

E)

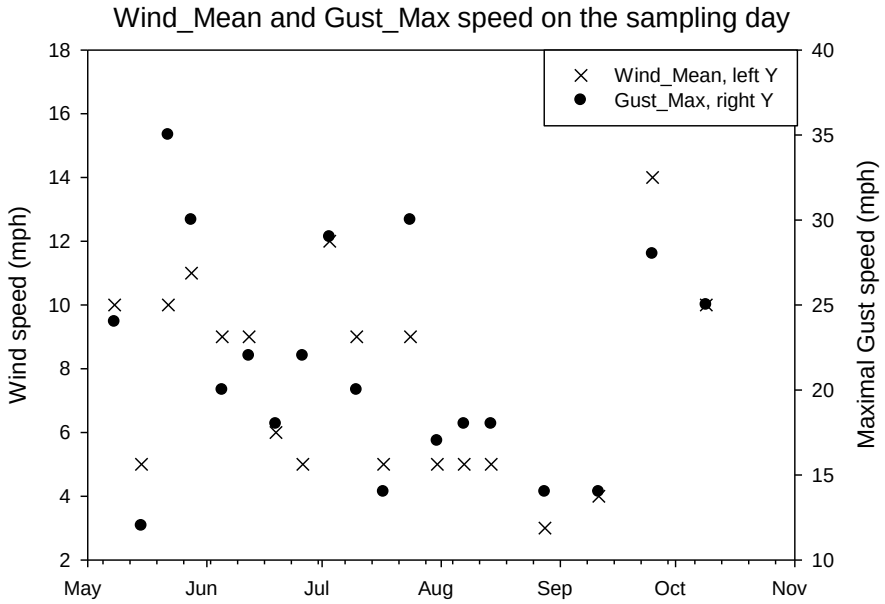


Fig. 3. 2. Environmental parameters versus time in GSB, 2008 A) Temperature at station E and J B) Salinity at Station E and J C) Precipitation on the sampling day and the sum of the prior 7 days D) Connetquot River discharge on the sampling day and the sum of the prior 7 days E) Average wind speed and maximum gust speed on the sampling day

Table 3. 3. Correlation coefficients between environmental parameters; numbers in bold represent significant correlations ($p<0.001$)

	Temp	Sal	River	7-days river	PPT	7-days PPT	Wind-Mean	Gust_ Max
Temp	1.000							
Sal	**0.696**	1.000						
River	**-0.504**	**-0.518**	1.000					
7-days river	**-0.512**	**-0.606**	**0.881**	1.000				
PPT	-0.101	0.122	0.092	-0.268	1.000			
7-days PPT	-0.023	-0.005	**0.596**	**0.543**	0.115	1.000		
Wind_Mean	-0.402	-0.205	0.103	-0.003	0.307	-0.310	1.000	
Gust_Max	-0.412	-0.121	0.346	0.185	0.407	-0.018	**0.815**	1.000

Temp: Surface water temperature; Sal: Salinity; River: Water discharge from the Connetquot River on the sampling day; 7-days river: The sum of water discharge from the Connetquot River over the prior 7 days including the sampling day; PPT: Precipitation on the sampling day; 7-days PPT: The sum of precipitation over the prior 7 days including the sampling day; Wind-Mean: Average wind speed on the sampling day; Gust_Max: Maximum gust speed on the sampling day; Temperature and salinity were the average of those at station E and J when all 10 stations were sampled and were those at station E when only 5 stations were sampled.

Temporal change of major TRFs

1 18S community

There were 5 'major' 18S TRFs (Fig. 3.3.). 256 partial 18S rDNA sequences of good quality were retrieved from clone libraries constructed from 9 samples and their predicted *Taq*I restriction fragments (Table A1.) were used to identify the major 18S TRFs.

Observed TRF 174B matched the predicted TRF from cloned and sequenced amplicons of *A. anophagefferens* nuclear rDNA. Another observed TRF, 176B, was present at average relative abundance of 3.1%, about 1/6 of 174B% (17.0%). 176B was usually present only when 174B was also present. Furthermore, 176B% demonstrated a strong linear relationship with 174B% (r=0.615) and their correlation was also captured in LSA (Fig. 3.15.A.). Therefore 174B and 176B most likely both represented TRFs of *A. anophagefferens* (Chapter 2). Area contributions of 174B and 176B were combined as c175B%, which ranged from 0 to nearly 50% of total TRF peak area (Fig. 3.3.). Dramatic changes in c175B% occurred between weeks, which resulted in its three peak values on June 5, June 26 and July 10. There was also a 'fall' peak of c175B% from late September into October.

Observed TRF 284B matched the TRF predicted from cloned *Chaetoceros calcitrans f. pumilus* 18S rDNA amplicons. A similar relationship as that between 174B and 176B was also found between 284B and 285B (linear correlation coefficient r=0.880), with average 284B% (1.4%) being 12% of average 285B% (11.8%). This was again confirmed by LSA (Fig. 3.16.A.). 284B and 285B were combined as c284B, and c284B% increased from 10% on June 5 (week 5), to over 80% on June 19 (week 7) (Fig. 3.3.). Correspondingly, c175B decreased in its area contribution from around 50% to about 10%. There was also a smaller peak of c284B% of around 35% on July 24 (week 12), when the area contribution of c175B was 0.

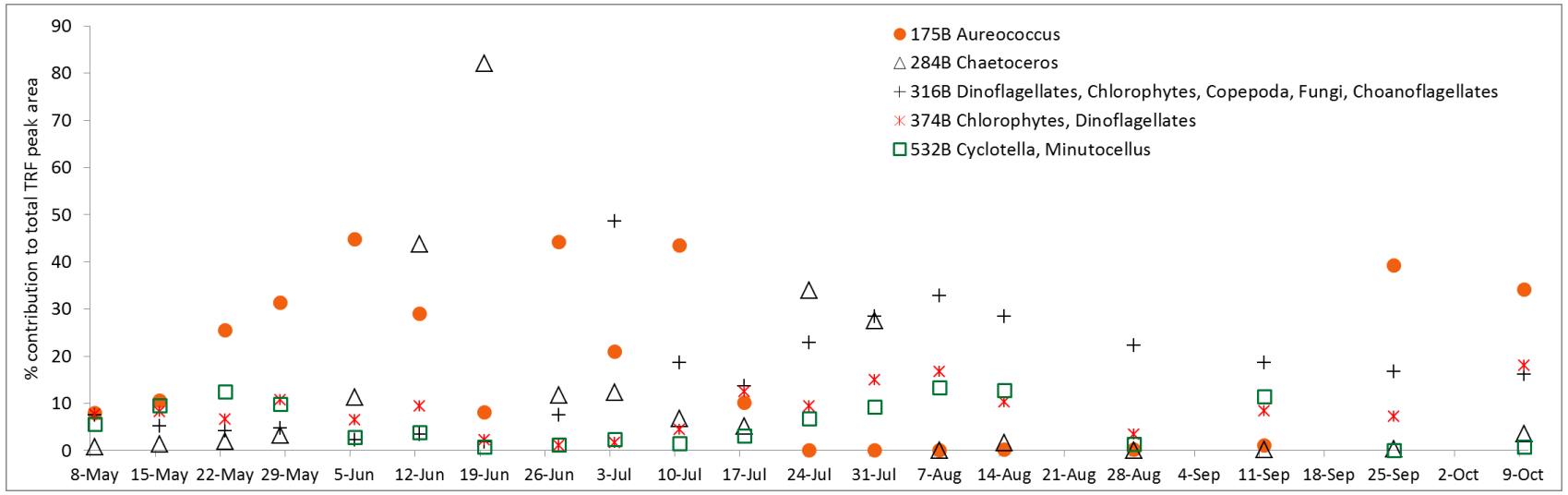
Major TRF c316B was composed of 3 observed TRFs with fragment lengths of 314, 316 and 318 bases, with no significant correlation between each pair. c316B matched the TRFs predicted for diverse taxa including dinoflagellates, chlorophytes, copepods, fungi and choanoflagellates (Table A1.). In this case, the 3 observed TRFs were lumped not because they represented a single taxon, but because it was not possible to separate the different taxa based on the observed TRF sizes. Out of the 36 sequences with predicted TRFs matching c316B, 8

belonged to heterotrophic dinoflagellates, 8 belonged to copepods and 10 belonged to chlorophytes especially *Picocystis salinarum*. c316B% started to increase on June 26 and peaked on July 3 at almost 50% (Fig. 3.3.). After that, it decreased but remained a major component, contributing 20-30% to total TRF peak area.

Major TRF c374B was also composed of 3 observed TRFs with fragment lengths of 372, 374, and 376 bases, with no significant correlation between each pair. Out of the 48 sequences that generated predicted TRFs matching c374B, 22 belonged to *Picochlorum* and other chlorophytes, while the other 26 belonged to various autotrophic dinoflagellates. c374B was present throughout the sampling season, ranging in area contribution from 1% in week 8 (June 27) to ~17% in week 14 (August 7) and week 23 (October 9) (Fig. 3.3.).

Major TRF c532B was contributed by observed TRFs 530B and 533B and matched predicted TRFs of the diatoms *Cyclotella choctawhatcheeana* (based on sequences retrieved in 2009, Table A3.) and *Minutocellus polymorphus* (Table A1.). Although *Cyclotella choctawhatcheeana* was not sequenced in 2008, its predicted TRF was the only TRF that matched the observed TRF 530B and the relative abundance of observed 530B was about 2 times higher than that of 533B. Major TRF c532B was present for most of the sampling time, with its maximal relative abundance around 15% on May 22 (week 3), August 7 (week 14), and September 11 (week 19), dropping to 0 in between (Fig. 3.3.).

The sum of the 5 major 18S TRFs' area contribution ranged from 27% on August 28 (week 17) to 95% on June 19 (week 7) (Fig. 3.3.). A common feature of the weeks when the combined area contribution of the major 18S TRFs was under 50% (weeks 1, 2, 11, 17, 19) was relatively low contribution of c175B.



Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sum of major TRFs area%	**30**	**35**	51	60	68	89	95	66	86	75	**45**	73	80	63	53	ND	**27**	ND	**40**	ND	64	ND	73

Fig. 3. 3. Contribution of major 18S TRFs to total peak area averaged among stations each week in 2008 (average CV of the contributions of all major TRFs at all sampling times is 0.47). Some TRFs with similar size (usually ±2 bases) were combined to form one major TRF, as explained in the text. Numbers in the first row below the figure represent the number of the week when samples were collected and the numbers in the second row are the summed area percentage of the 5 major 18S TRFs in that week (values less than 50% in bold); ND: no data (samples not collected these weeks)

2 16S community

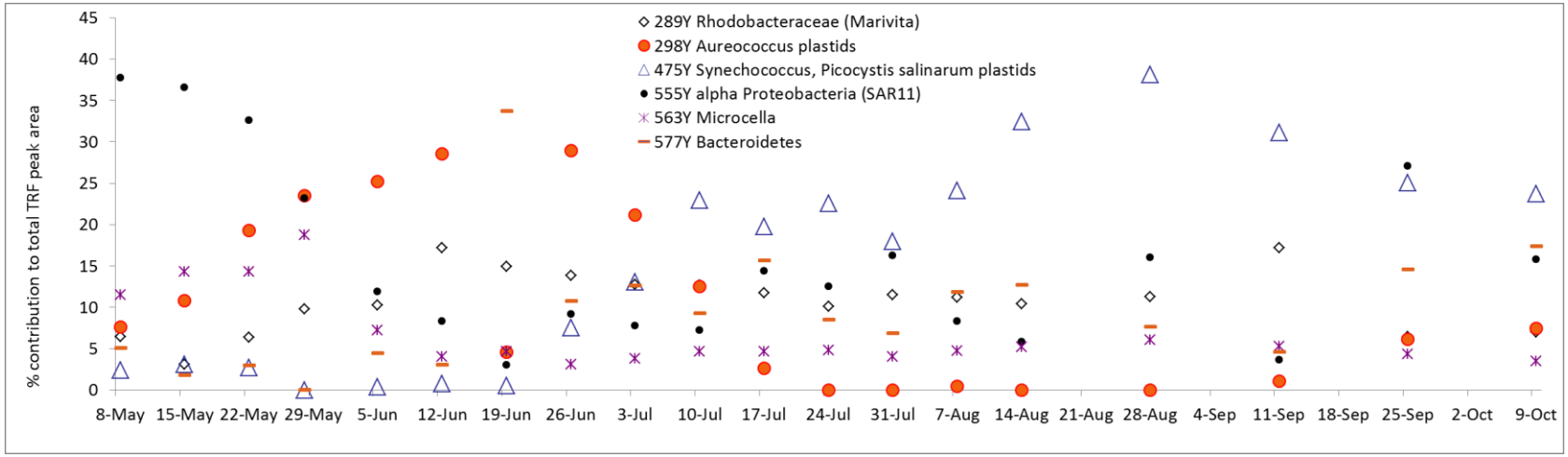
Six 16S TRFs met the criterion to be considered as major TRFs (Fig. 3.4.). 99 partial 16S rDNA sequences of good quality were retrieved from 2 clone libraries (Table A2.) and their predicted *Taq*I restriction fragments were used to identify the major 16S TRFs.

Observed TRF 297Y matched the predicted TRF from cloned and sequenced amplicons of *A. anophagefferens* plastid rDNA. Another observed TRF, 299Y, was present in all samples with 297Y at about 1/7 the contribution to total TRF area (1.5% vs 10.5%), and there was a strong linear relationship between 297Y and 299Y (r=0.711), which was also supported by LSA (Fig. 3.15.B.). Therefore, 297Y and 299Ywere combined as the major TRF c298Y to represent *A. anophagefferens* plastid rDNA. One sequence from a Bacteroidetes had the same predicted TRF as *A. anophagefferens* plastid, but out of the 13 sequences with a predicted TRF matching 297Y, only 1 Bacteroidetes sequence was retrieved. In contrast to c175B, which had its highest area contribution of ~ 50% on June 5 (week 5), June 26 (week 8) and July 10 (week 10), the area contribution of c298Y was highest on June 12 (week 6) and June 26 (week 8) at ~ 30% and did not peak in July. Although not as dramatic as c175B, c298Y also increased in area contribution from September 11 (week 19) into the fall. The temporary decrease of c298Y's area contribution on June 19 (week 7) was consistent with that of c175B. Concurrent with this drop in area contribution of c298Y, major TRF 577Y, representing *Flavobacteria* and other Bacteroidetes (Table A2.), increased dramatically in its area contribution on June 19 (week 7) (Fig. 3.4.). Five of the six sequences generating predicted TRFs matching 577Y were almost identical (identities=577/578) with a sequence retrieved from Chesapeake Bay (Kan et al. 2008) and identified as uncultured Bacteroidetes. The area contribution from 577Y was lower before week 7, at < 5%, and higher after week 7, ranging from 5% to 15%.

Major TRF c289Y was composed of the observed TRFs 288Y and 290Y(r= -0.36, also confirmed by LSA), and represented mostly *Marivita* and other Rhodobacteraceae (2 of 20 sequences with a predicted TRF matching 288Y were Actinobacteria *Yonghaparkia*). c289Y increased in its area contribution from 3% in mid-May to 17% in

mid-June, after which it stayed around 12% to 17% until September 11, then declined to 7% the following week (Fig. 3.4.). Major TRF c475Y was composed of the observed TRFs 474Y and 476Y (not significantly correlated), and represented both *Synechococcus* and *Picocystis salinarum* plastid rDNA. Inspection of *Synechococcus* and *Picocystis* plastid rDNA sequences suggested that *Rsa*I digestion would produce different TRFs, 124Y for *Synechococcus* and 295Y for *Picocystis*. TRFLP done on the sample from week 11 using *Rsa*I showed that the area contribution from *Rsa*I 124Y was 14% of the total TRF area while *Rsa*I 295Y contributed only 2%. Therefore, c475Y most likely represents more *Synechococcus* than *Picocystis*, at least in the sample collected from week 11. Major TRF 555Y matched the TRF predicted from SAR11 (*Pelagibacter*) and other alpha Proteobacteria. It decreased in its area contribution from nearly 40% in early May to ~ 5% on June 19 (Fig. 3.4.). The area contribution of 555Y fluctuated between 5% and 15% through the summer before increasing again in the fall. Major TRF c563Y was composed of 562Y and 564Y (not significantly correlated) and represented mostly the actinobacterium *Microcella* (may also include chlorophyte plastids). c563Y% was higher in May (12% to 19%) than in the rest of the sampling season (~5%) (Fig. 3.4.).

There was never a 16S TRF that was as dominant as c284B was in the 18S community (> 80% in week 7). The sum of the 6 major 16S TRFs' area contribution ranged from 57% on July 31 (week 13) to 84% on September 25 (week 21), which did not correspond with the dates of the minimal and maximal sum values for the 5 major 18S TRFs (Fig. 3.3., Fig. 3.4.).



Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sum of Major TRFs area%	71	70	78	75	59	62	61	73	79	69	69	59	57	61	67	ND	79	ND	63	ND	84	ND	75

Fig. 3. 4. Contribution of major 16S TRFs to total peak area averaged among stations each week in 2008 (average CV of the contributions of all major TRFs at all sampling times is 0.26). Some TRFs with similar size (usually ±2 bases) were combined to form one major TRF, as explained in the text. Numbers in the first row below the figure represent the number of the week when samples were collected and the numbers in the second row are the summed area percentage of the 6 major 16S TRFs in that week; ND: no data (samples not collected these weeks)

Aureococcus anophagefferens bloom

The contributions of c175B (c175B%) and c298Y (c298Y%) to total 18S and 16S TRF area (respectively) were averaged among all samples analyzed on each date and compared with immunofluorescent *A. anophagefferens* cell count data (SCDHS 2011) (Fig. 3.5.). Changes in *A. anophagefferens* cell numbers determined microscopically demonstrated a similar pattern as changes in c175B% and c298Y%, with a strong linear relationship between cell count data and TRFLP data. The linear regression coefficient between c298Y% and *A. anophagefferens* cell counts was higher than that between c175B% and *A. anophagefferens* cell counts (Fig. 3.5.) because neither cell count data from September 25 nor TRFLP data from September 18 could be included in the correlation analysis, so the better agreement between c175B% and *A. anophagefferens* cell counts towards the end of the sampling season was not captured in the linear regression. The linear model suggested that an area contribution of 15% from c175B corresponded to 200,000 *A. anophagefferens* cells per ml, which defines the bloom as the most severe category (Gastrich & Wazniak 2002). Therefore 15% area contribution of c175B is used as the threshold separating bloom and non-bloom samples for the rest of this paper. The linear model between c298Y% and *A. anophagefferens* cell counts indicated that 7.4% area contribution from c298Y corresponded to 200,000 cells per ml.

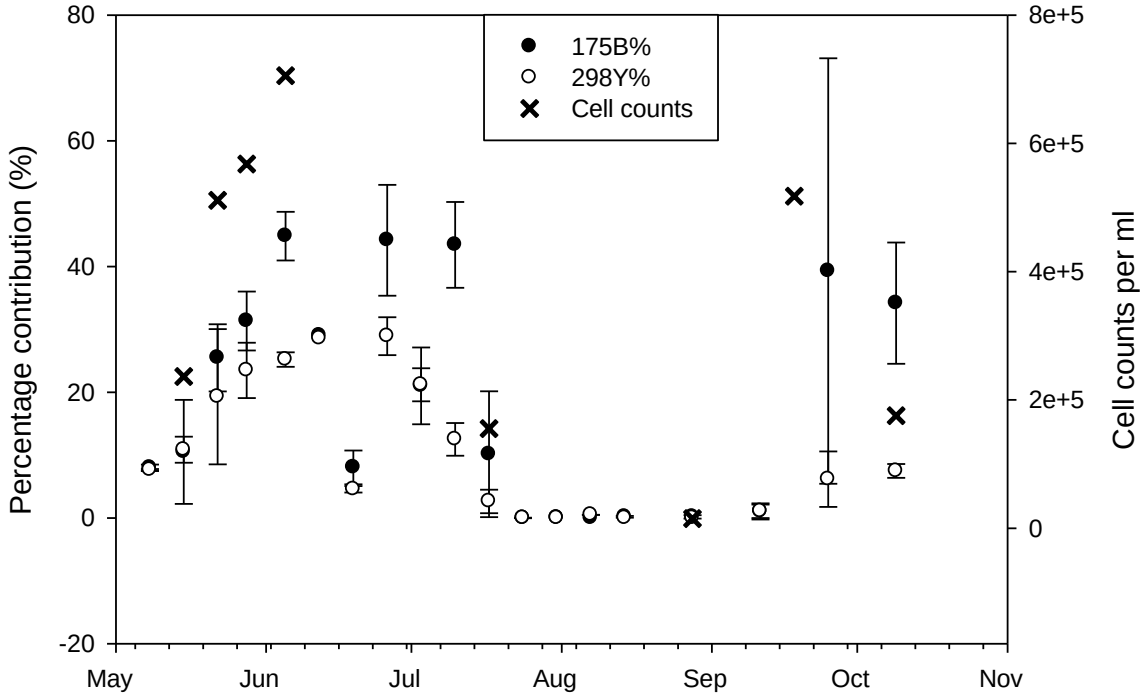


Fig. 3. 5. Average c175B% (solid circle) and c298Y% (empty circle) among all stations versus *A. anophagefferens* cell counts (cross); the difference between the minimal and maximal c175B% and c298Y% from all stations on each sampling day is shown as the height of the error bar. Cell Counts = 12890*c175B% + 49684 (r=0.801)
Cell Counts = 24948*c298Y% + 21232 (r=0.982)

Temporal and spatial variation of GSB plankton community

1 18S community

Fig. 3.6. shows the NMS ordination biplots for the GSB 2008 18S TRFLP. The distances between points on the biplots are approximately proportional to the dissimilarities measured by Bray-Curtis distance between the samples. In general, samples collected on the same day clustered together, although the spatial variation among stations was pronounced in some weeks, for example week 21. Temporal variation was generally greater than spatial variation and clear distinctions between samples from two consecutive weeks were quite common, such as from week 5 to week 6 (Fig. 3.6.). 88% of variation (R^2) among samples was represented by 3 axes, with axes 2 and 3 accounting for more variation than axis 1 (Table 3.4.). Temperature, salinity and river discharge were correlated with axis 1 (note the different sign of salinity and river discharge), wind alone was correlated with axis 2, and axis 3 was only significantly correlated with temperature (Table 3.5.). Based on axes 1 and 2, May samples were located in the area of the biplot corresponding to the highest Connetquot River discharge and the lowest temperature (Fig. 3.6.). In June, sample points moved to the region of the plot corresponding to higher temperature, salinity and wind. July sample points were spread over a large part of the biplots, while August samples moved to the upper left corner corresponding to high temperature and low wind speed. Early September samples were located in the region corresponding to lower temperature with little wind (week 19, Fig. 3.6.) while during the second half of September and early October, sample points moved to the mid-lower part of the biplot associated with windier conditions. Noticeably, samples with c175B contributing more than 15% to total TRF area (from the 'bloom' weeks 3, 4, 5, 6, 8, 9, 10, 21 and 23) were separated from the rest of the samples along axis 2 (Fig. 3.6.), indicating a strong association between wind speed on the sampling day and *A. anophagefferens* relative abundance in the community.

Table 3. 4. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 166 TRFs-dimensional space for 18S community

Axis	Increment	Cumulative
1	0.167	0.167
2	0.362	0.529
3	0.350	0.880

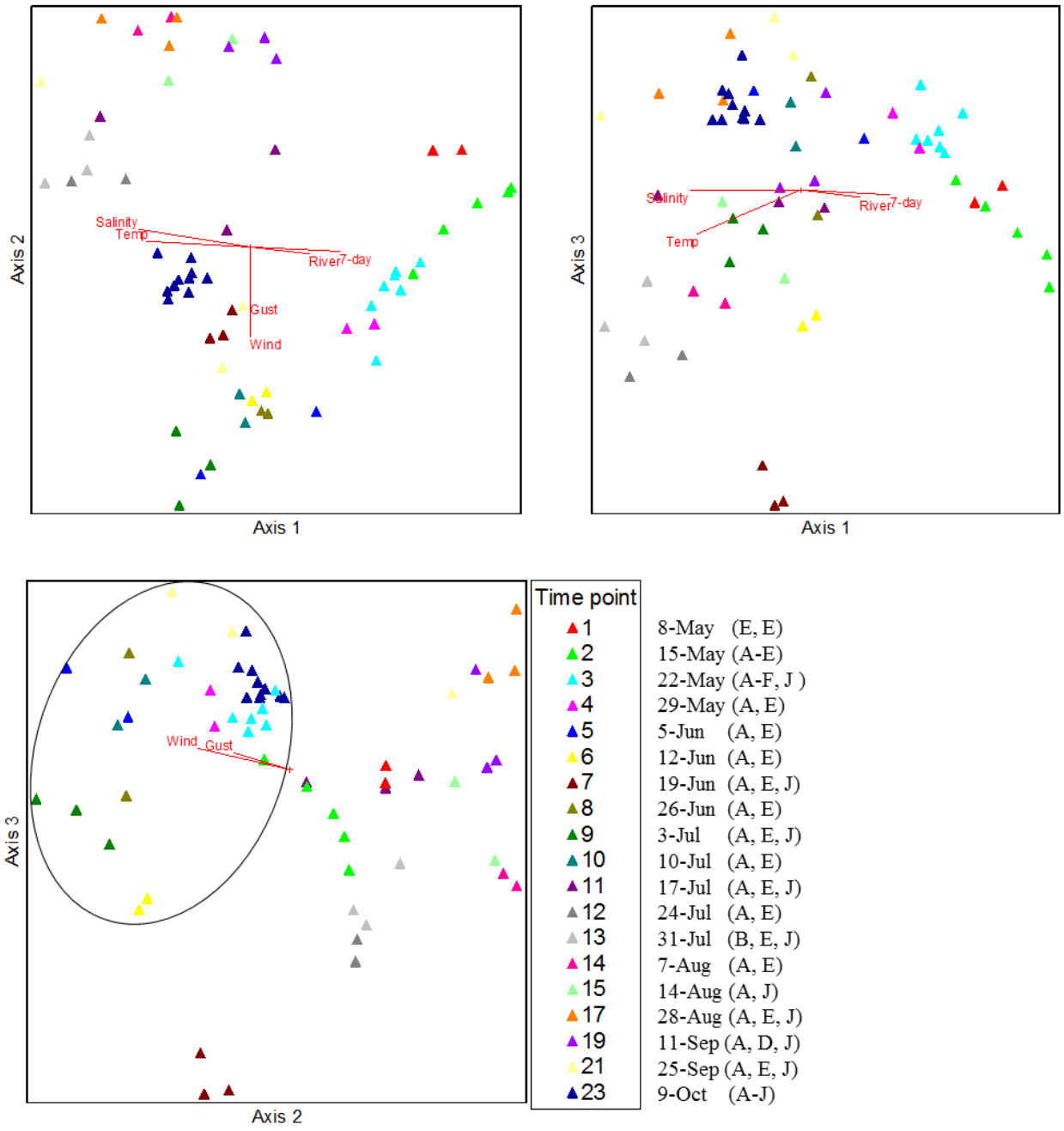


Fig. 3. 6. NMS ordination biplots of 18S TRFLP in GSB, 2008; Samples in the circle are the *A. anophagefferens* bloom samples (c175B %> 15%)

Table 3. 5. Pearson's Correlation Coefficients (r) between environmental parameters and the ordination axes for 18S community (N= 61), 16S community (N=55), and 16S community with photoautotroph TRFs artificially removed (N=55); numbers in bold represent significant r values at p=0.001 according to the statistical table of critical values

Axis / **Env**	18S			16S			16S_No autotrophs		
	1	2	3	1	2	3	1	2	3
Temp	**-0.651**	0.149	**-0.424**	**0.649**	**-0.649**	0.146	**0.667**	**-0.597**	**-0.558**
Salinity	**-0.675**	0.267	-0.018	**0.555**	-0.311	0.366	0.323	-0.441	-0.319
Wind	0.052	**-0.606**	0.290	**-0.458**	0.148	-0.189	-0.216	0.171	0.370
Gust	0.050	**-0.472**	0.256	-0.391	0.173	-0.274	-0.270	0.207	0.236
River	**0.493**	-0.166	-0.175	-0.311	**0.482**	**-0.516**	**-0.543**	0.425	-0.077
7-day	**0.605**	-0.134	-0.142	-0.363	**0.490**	**-0.526**	**-0.561**	0.427	-0.041

Env: Environmental parameters
Temp: Surface water temperature
Wind: average wind speed on the sampling day
Gust: maximum gust speed on the sampling day
River: water discharge from the Connetquot River on the sampling day
7-day: water discharge from the Connetquot River of the prior 7 days including the sampling day

2 16S community

Similar to 18S, temporal variation dominated over spatial variation and the 16S community structure of consecutive weeks was sometimes very different (Fig. 3.7.). 91% of variation among samples was represented by 3 axes with each axis accounting for a similar percentage of variation (Table 3.6.). Axis 1 was significantly associated with temperature, salinity and average wind speed on the sampling day (Table 3.5.). The other two axes were both related to river discharge, but axis 2 was also strongly associated with temperature while axis 3 was not. Based on axes 1 and 2, May samples were found in the part of the biplot corresponding with lowest temperature, highest Connetquot River discharge, and highest wind speed (Fig. 3.7.). June samples moved to the region corresponding to higher temperature and lower Connetquot River discharge. From July 3 and July 10 to the rest of July and the first half of August, samples became associated with saltier and less windy conditions. Samples collected on August 28 grouped near the early July samples. Compared to the samples collected on October 9, samples collected on two days in September were associated with saltier and less windy conditions. Spatially, the large spread of samples observed in week 21 for the 18S community is not present in the 16S community (see 'spatial variation' below for more detail). 'Bloom' samples with c298Y contributing more than 7.4% to total TRF area (weeks 1, 2, 3, 4, 5, 6, 8, 9, 10, 23) grouped to one end of axis 1 (Fig. 3.7.), consistent with the association of bloom samples with wind in the 18S community.

Table 3. 6. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 87-dimensional space for 16S community

Axis	Increment	Cumulative
1	0.350	0.350
2	0.301	0.651
3	0.263	0.913

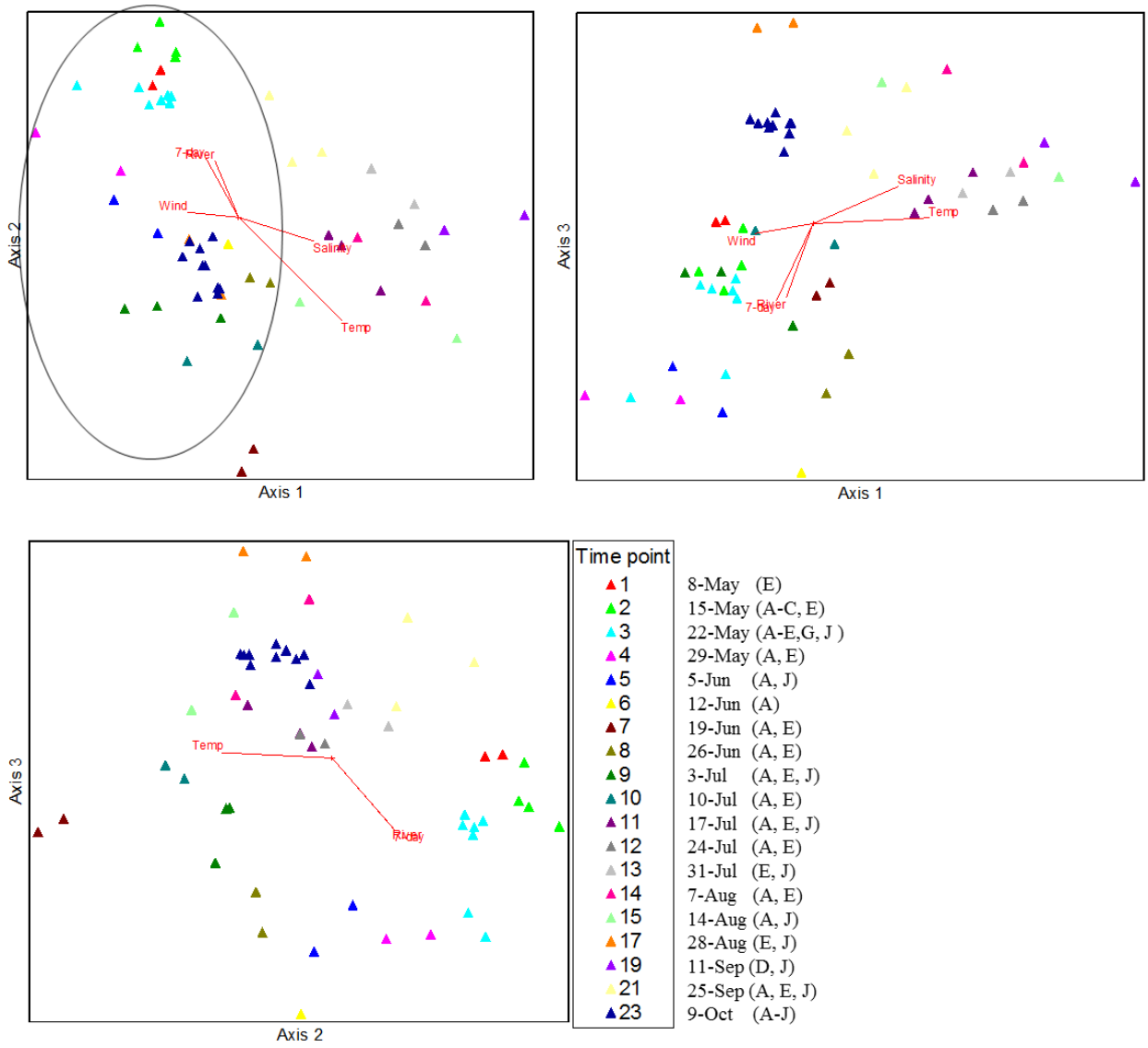


Fig. 3. 7. NMS ordination biplots of 16S TRFLP in GSB, 2008; Samples in the circle in A are those with c298Y% higher than 7.4% (except for week 17).

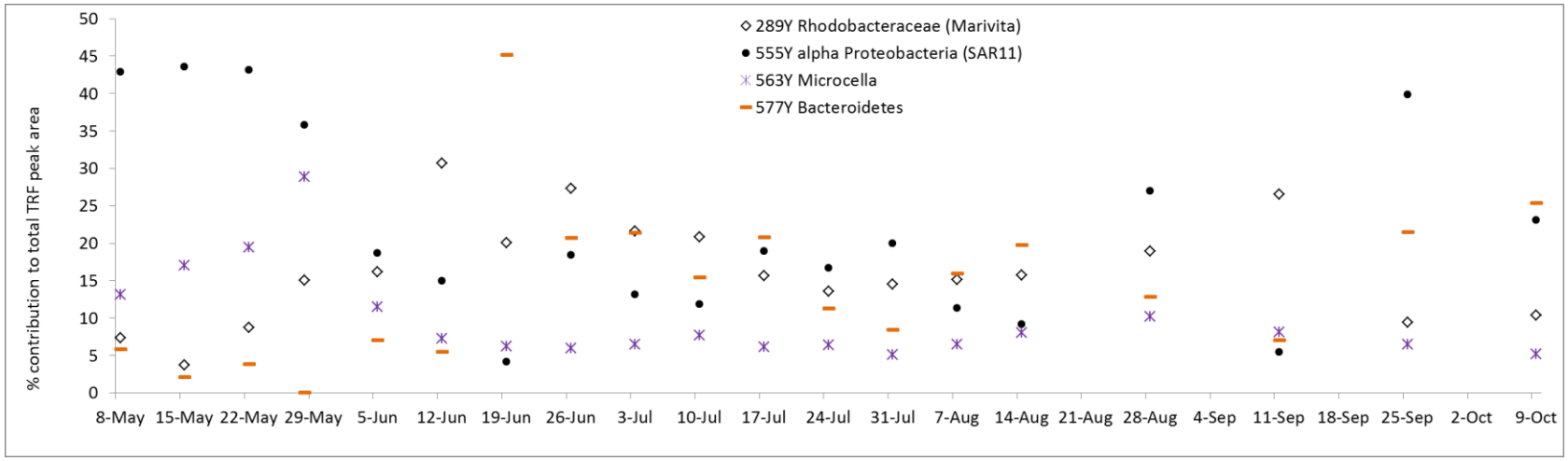
The 16S TRFLP main matrix was reconstructed by removing the six TRFs representing autotrophic groups (*Synechococcus*, plastids of *Picocystis salinarum*, plastids of *A. anophagefferens* and plastids of diatoms represented by 272B) from the main matrix to reveal spatial and temporal patterns of the heterotrophic bacteria. The area contribution of the sum of autotrophic TRFs ranged from 15% in week 13 to 87% in week 8, so the area contribution of major 16S TRFs after removing the autotrophic TRFs was recalculated (Fig. 3.8.). Major TRFs of the reconstructed community were still the same major TRFs (Fig. 3.4.) and no new TRF was 'promoted' to the major TRF category, suggesting that there was very strong dominance of the

major TRFs in the original community. Temporal patterns of the relative abundance of major TRFs for the reconstructed community were similar to those of the original community throughout the sampling season (Fig. 3.8.). The sum of area contribution of major TRFs for the original community was also similar to that of the reconstructed community, with the greatest changes occurring in week 7 and week 9 from 61% and 79% to 78% and 57%, respectively (Fig. 3.4. and Fig. 3.8.).

For the reconstructed 16S main matrix, the NMS solution was also 3-D (Fig. 3.9.). Almost 93% of the variation among samples was represented by 3 axes with axis 1 accounting for the highest percentage (Table 3.7.). Temporal variation still dominated over spatial variation, and weeks such as week 15 with greater spatial variation in the original dataset still had larger spatial variation after the autotrophic TRFs were removed (Fig. 3.7. and Fig. 3.9.). In addition, sample 9A was placed further from 9E and 9J, because differences in the contributions of other TRFs had a greater influence on the ordination after the autotrophic TRFs, which contributed ~58% to the original total TRF area in these samples, were removed. Temperature was related to all 3 axes while river discharge was only significantly correlated with axis 1 (Table 3.5.). Wind was no longer significantly associated with any of the ordination axes.

Table 3. 7. Coefficients of determination for the correlations between ordination distances and distances in the original 81-dimensional space for 16S community when autotrophic TRFs were artificially removed

Axis	Increment	Cumulative
1	0.466	0.466
2	0.246	0.713
3	0.213	0.925



Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sum of Major TRFs area%	72	81	82	82	62	60	78	75	57	59	64	51	52	51	55	ND	71	ND	48	ND	83	ND	69

Fig. 3. 8. Contribution of major 16S TRFs to total peak area in GSB 2008 averaged among stations when autotrophic TRFs were removed (average CV of the contributions of all major TRFs at all sampling times is 0.26). Some TRFs with similar size (usually ±2 bases) were combined to form one major TRF, as explained in the text. Numbers in the first row below the figure represent the number of the week and the numbers in the second row are the summed area percentage of the 4 major 16S TRFs in that week; ND: no data (samples not collected these weeks)

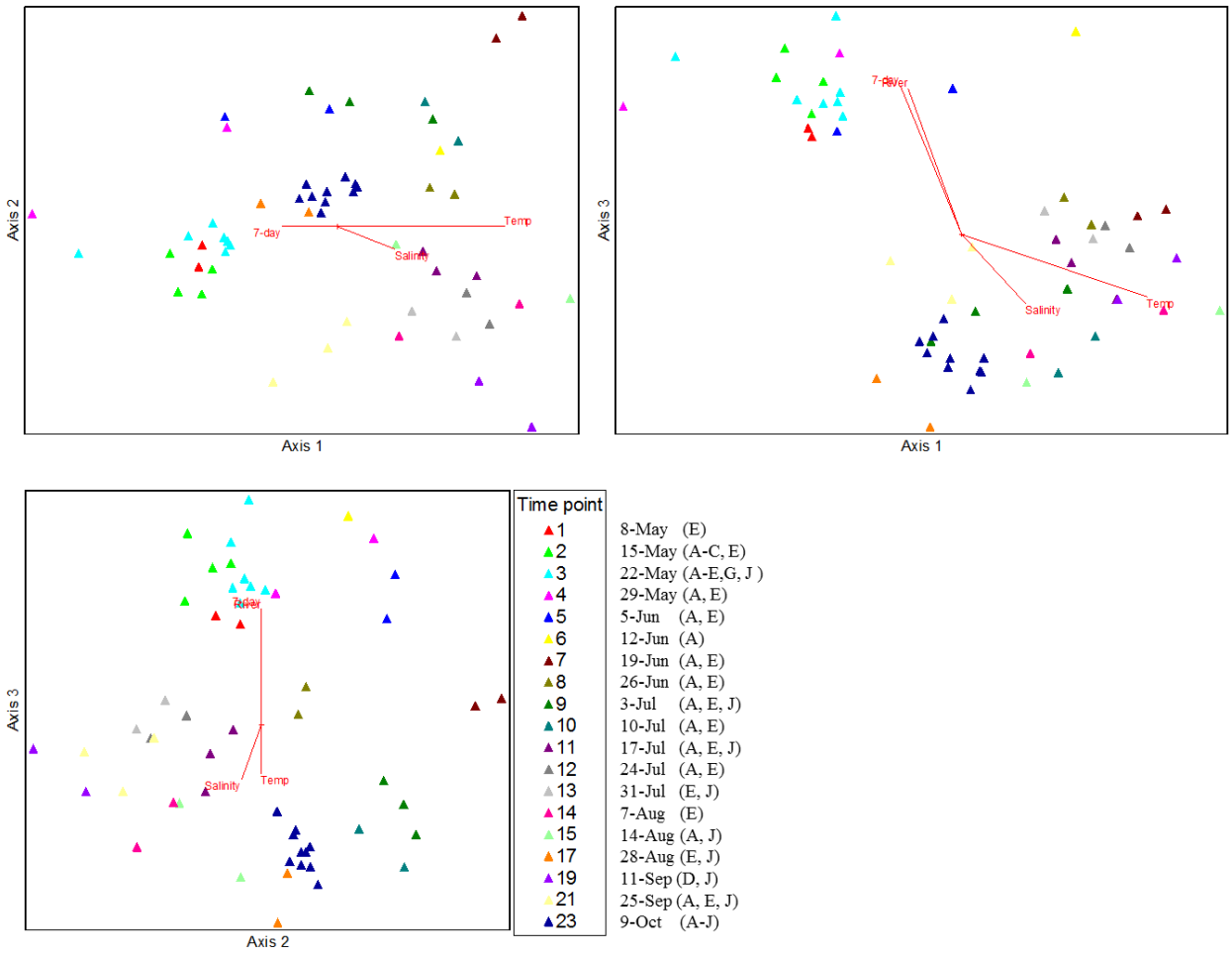


Fig. 3. 9. NMS ordination biplot of GSB 2008 16S microbial community when autotrophic TRFs were removed

Spatial variation

The NMS ordinations suggested that there might be more spatial variation in some weeks than others. Based on the CV of the major TRFs among stations (see Materials and Methods), weeks 8, 11, 12, 13, 14, 15, 19, and 21 for 18S and weeks 15 and 21 for 16S were identified as spatially heterogeneous weeks, and all samples collected in these weeks were analyzed with TRFLP (except that samples 11B, 13A, 13D, 14D, 19E, 19F and 19G were missing for the 18S community).

1 18S spatial heterogeneity

The Bray-Curtis distances between all pairs of TRFLP profiles from different stations in the same week were regressed against the geographical distances between stations (Fig. 3.10.). There was always substantial variation of the Bray-Curtis distance for each geographical distance; however, a statistically significant positive relationship between the Bray-Curtis distance and the geographical distance was indicated by a linear regression model (Fig. 3.10., $p<0.001$). The slope of the regression line indicates how much plankton communities differ as they get further from each other geographically in GSB. Piecewise regression using R (Team 2012) did not offer support for the presence of different relationships at different distances, suggesting that there was no saturation of the Bray-Curtis distance at the higher geographical distances.

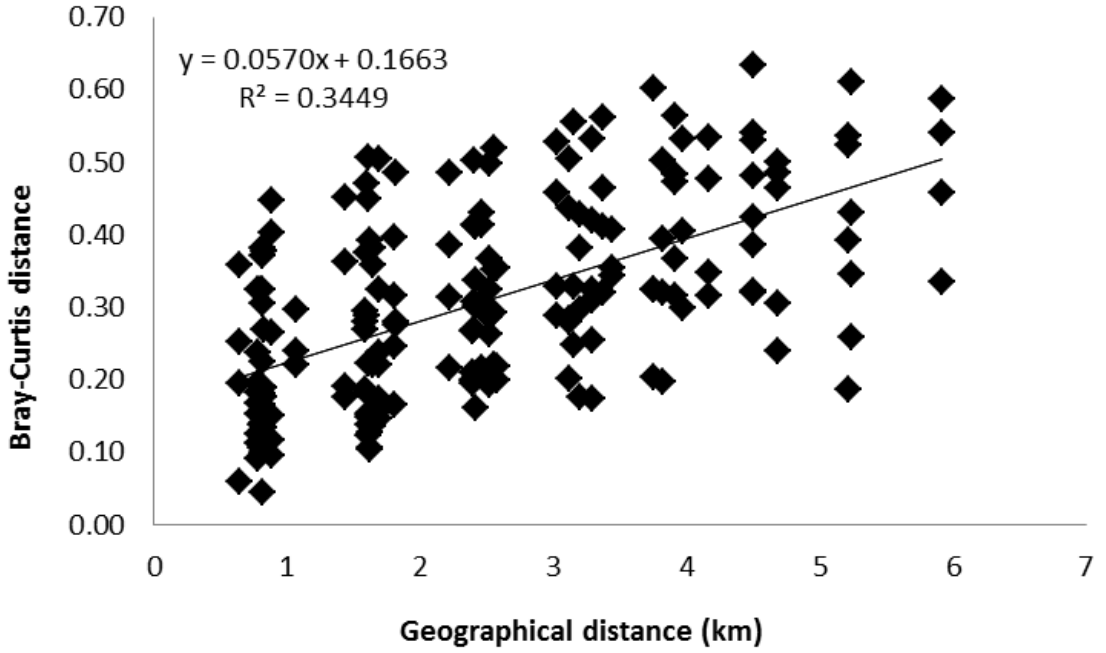


Fig. 3. 10. Linear regression (N=201) of the Bray-Curtis distance against geographical distance for 18S TRFLP in spatially heterogeneous weeks in GSB, 2008

Each week was also analyzed separately to determine whether spatial heterogeneity varied with time (Fig. 3.11.). The average of all slopes for spatially heterogeneous weeks was 0.052 with a standard deviation of 0.017. The largest slope of the regression line was from week 14 and was caused by a big difference between the community at station E and the other stations, which was reflected in the area contribution of all major (>5%) 18S TRFs in this week (Fig. 3.12.). The second largest slope was from week 21, when there was a strong along-transect trend for all major TRFs' area contribution, with c175B% decreasing and the other TRFs increasing from station A to station J (Fig. 3.12.). Week 21 was also the only week when the salinity of station E, 27.4 PSU, was higher than that of station J, 25.4 PSU (Fig. 3.2.B.), which is referred to as 'reverse salinity structure' later in this dissertation.

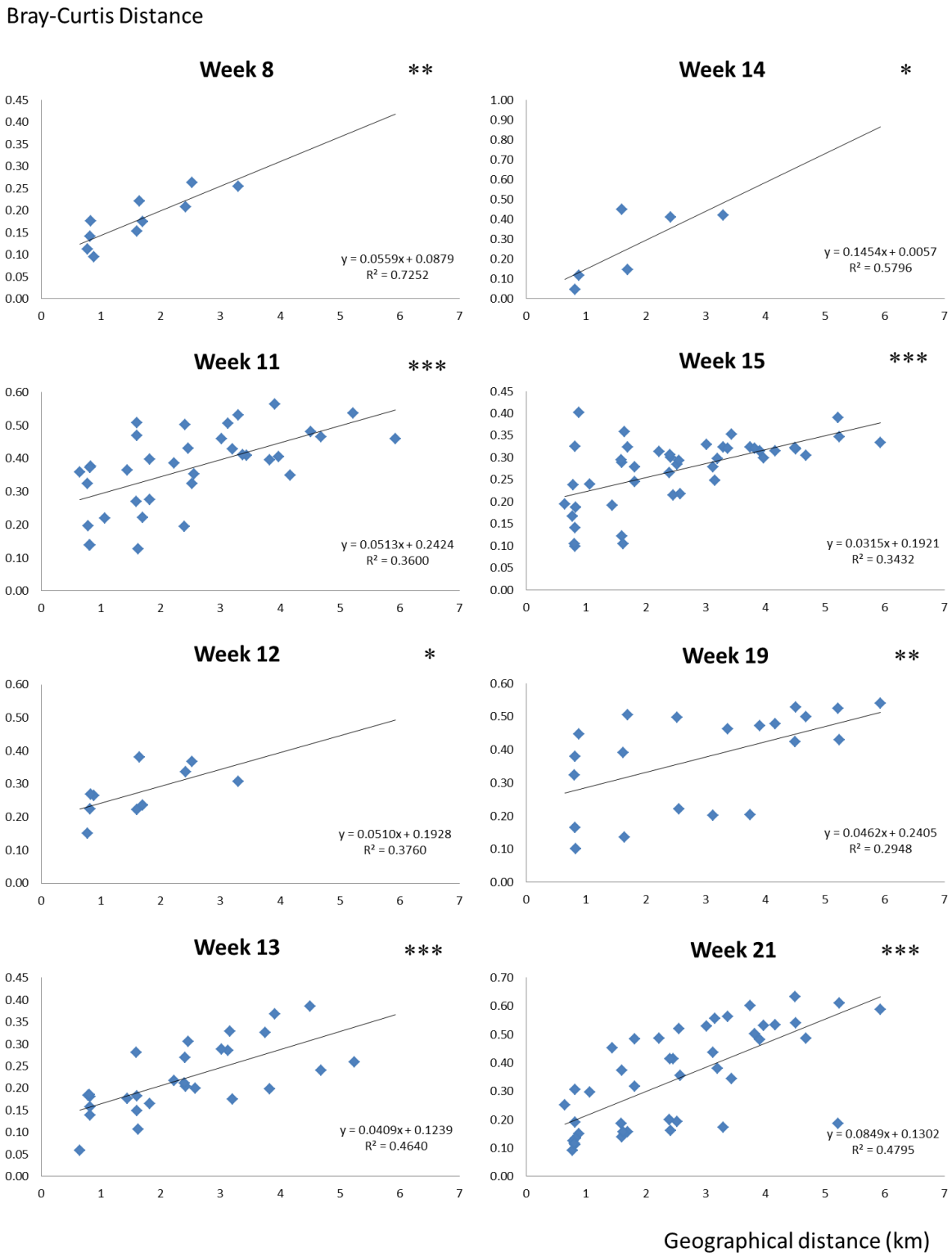


Fig. 3. 11. Linear regression of Bray-Curtis distance against geographical distance for 18S TRFLP in all spatially heterogeneous weeks during 2008
***: $p<0.001$; **: $0.001 \leq p \leq 0.01$; *: $0.01 \leq p < 0.05$

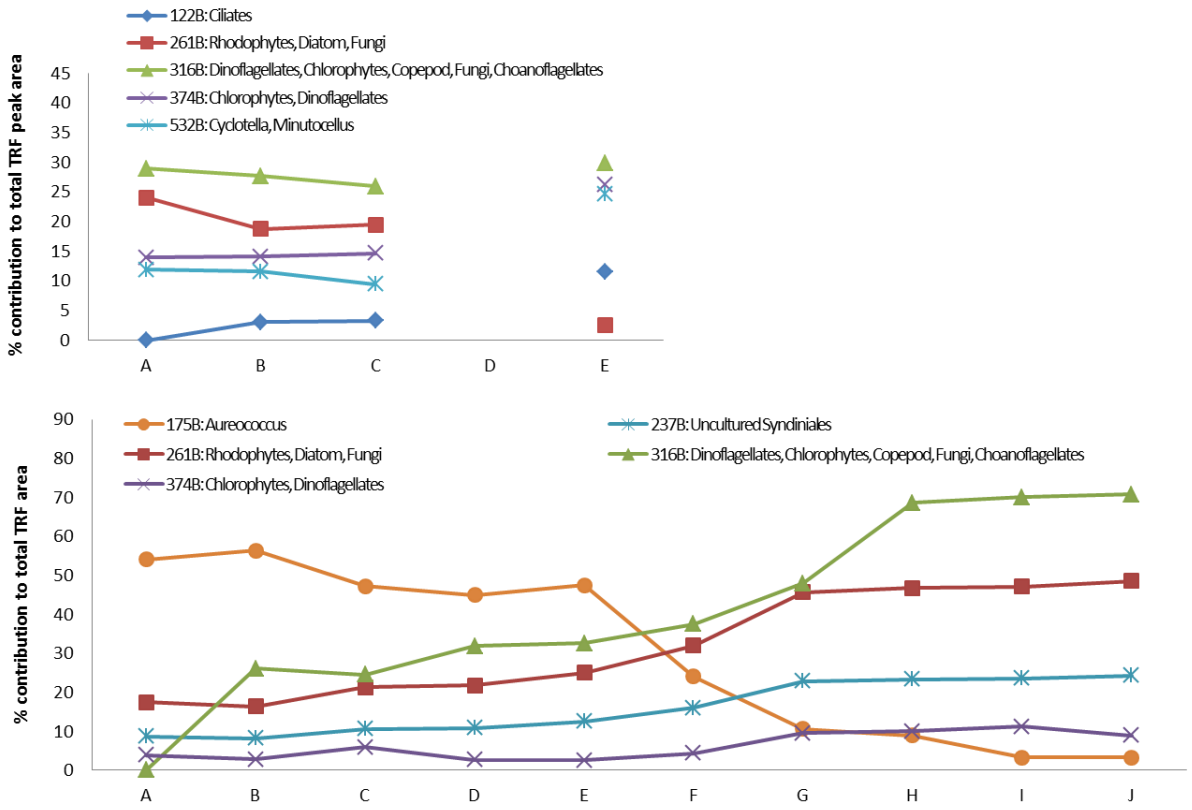


Fig. 3. 12. Area contribution of major 18S TRFs (>5% in respective weeks) along the sampling transect (station A to J) in week 14 (upper panel) and week 21 (lower panel)

2 16S community

TRFLP analysis of all samples collected in weeks 15 and 21 showed that even these two most heterogeneous weeks for the 16S TRFLP were much less spatially heterogeneous than the 18S samples collected in the same weeks: the regression of Bray-Curtis distance versus geographical distance through the 90 data points from weeks 15 and 21 for the 16S community was almost horizontal (slope was not significantly different from 0 at p=0.01) with a very small correlation coefficient (Fig. 3.13.). Regression lines for the two weeks separately had the same features (not shown). The range of the Bray-Curtis distance at larger geographical distances was less than at smaller geographical distances, though fewer data were available at the larger geographical distances. TRF c475Y was the most abundant TRF of week 15 but there was no clear along-transect trend for it or any of the major 16S TRFs (Fig. 3.14.). Similar to the pattern

seen for c175B% in week 21 (Fig. 3.12.), c298Y% (*A. anophagefferens* plastid) decreased from station F to station J.

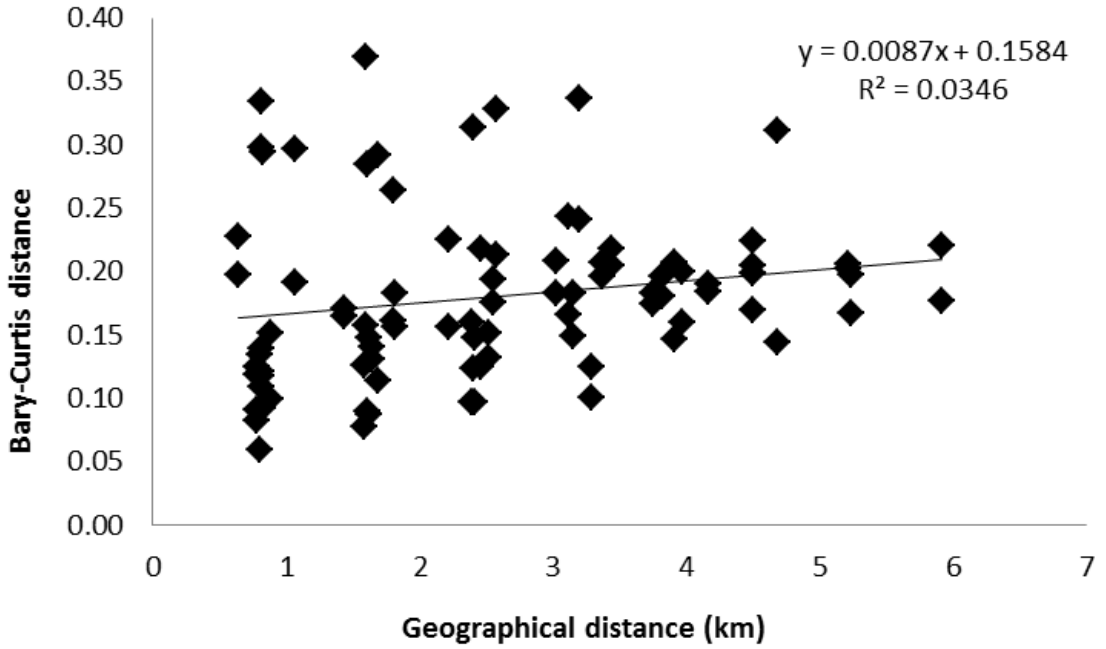


Fig. 3. 13. Linear regression (n=90) of the Bray-Curtis distance and geographical distance for 16S TRFLP in spatially heterogeneous weeks in GSB, 2008

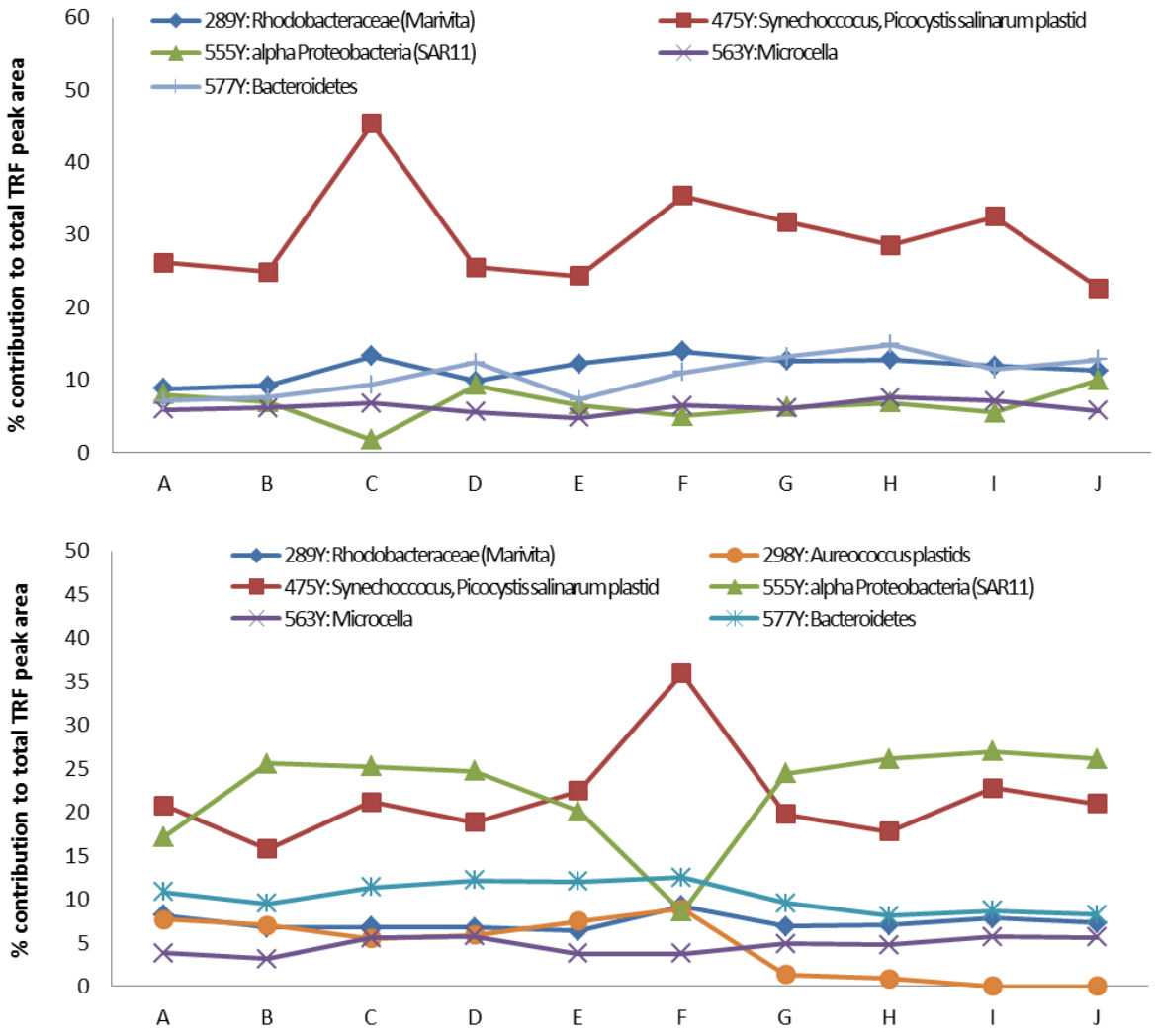


Fig. 3. 14. Area contribution of major 16S TRFs (>5% in respective weeks) along the sampling transect (station A to J) in week 15 (upper panel) and week 21 (lower panel)

Biodiversity

After normalization of total peak area (see 'Biodiversity index' in Materials and Methods) and removing 3 under-loaded samples (total peak area in these samples was less than 15% of the averaged total peak area from all samples; statistically not outliers in other analyses) as well as samples from week 7 (TRF c284B was extremely dominant in week 7 at a relative abundance of 80%, giving week 7 samples unusually low diversity), there were 55 samples in the 18S diversity data matrix. 31 of them had *A. anophagefferens*-specific TRF c175B contributing more than 15% of the total TRF peak area and were categorized as bloom samples while the other 24 were categorized as non-bloom samples. Both the richness index N and the evenness index Pielou's J were significantly higher for the non-bloom samples than for bloom samples (Table 3.8.). Diversity indices Shannon's H' and Simpson's D are influenced by both richness and evenness and were both significantly higher for non-bloom samples as well.

Table 3. 8. T tests of Richness (N), Shannon's H', Simpson's D and Pielou's J of 18S community for brown-tide-bloom and non-brown-tide-bloom samples

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	Non-bloom	Bloom	Non-bloom	Bloom	Non-bloom	Bloom	Non-bloom	Bloom
Range	8~21	5~18	1.72 ~2.91	1.26 ~2.51	1.78 ~0.94	0.66 ~0.88	0.81 ~0.98	0.71 ~0.90
Mean	14.21	11.39	2.353	1.968	0.873	0.797	0.895	0.82
Observations	24	31	24	31	24	31	24	31
P(T<=t) one-tail	=0.001		<0.001		<0.001		<0.001	

There were 51 samples in the 16S diversity data matrix after normalization and excluding 4 under-loaded samples (statistically not outliers in other analyses). 19 qualified as non-bloom samples and 32 qualified as bloom samples depending on whether c298Y% was bigger than 7.4% of the total 16S TRF peak area (see '*Aureococcus anophagefferens* bloom' in Results). Contrary to the 18S community, none of the indices was significantly different between bloom and non-bloom samples (Table 3.9.). Results were the same when categorizing bloom and non-bloom samples according to c175B% (not shown). There was no significant relationship between any of the diversity indices and environmental parameters, for either 18S or 16S communities.

Table 3. 9. T tests of Richness (N), Shannon's H', Simpson's D and Pielou's J of 16S community for brown-tide-bloom and non-brown-tide-bloom samples

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	Non-bloom	Bloom	Non-bloom	Bloom	Non-bloom	Bloom	Non-bloom	Bloom
Range	7~14	8~15	1.67 ~2.34	1.81 ~2.36	0.77 ~0.88	0.75 ~0.88	0.77 ~0.91	0.80 ~0.93
Mean	11.26	11.16	2.081	2.068	0.835	0.830	0.864	0.861
Observations	19	32	19	32	19	32	19	32
P(T<=t) one-tail	=0.412		=0.395		=0.324		=0.369	

Local Similarity Analysis

1 TRF-TRF and TRF-Environmental Factor interactions

4560 LS scores were calculated for all possible TRF pairs (TRF-TRF) as well as TRF and environmental factor (TRF-EF) pairs. 96 nodes representing all TRFs and environmental factors were linked by 730 LS scores larger than 0.3 with both p and q values less than 0.005. The highest LS, 0.73, was between Connetquot River discharge on the sampling day and the sum of the prior 7 days. LSA agrees mostly with Pearson Correlation in that only 60 out of the 730 significant LS correlations were considered non-significant according to p values for Pearson's r (Ppcc) with a 0.005 cutoff. With a less stringent Ppcc value of 0.05, only 6 out of the 730 significant LS correlations were considered non-significant. Subnetworks including major TRFs, some phytoplankton groups, and potential predators/parasites are shown in Fig. 3.15 to Fig. 3.19.; names of nodes are explained in Table 3.10.

Table 3. 10. Node abbreviations used in LSA network

Abbreviation	Observed TRF	Translation
AA1	174B	*Aureococcus anophagefferens* 1
AA2	176B	*Aureococcus anophagefferens* 2
AAP1	297Y	*Aureococcus anophagefferens* plastid 1
AAP2	299Y	*Aureococcus anophagefferens* plastid 2
Alpha	558Y	α-Proteobacteria
Amoe	292B	*Amoebophrya*
Bact1	99Y	Uncultured Bacteria 1
Bact2	259Y	Uncultured Bacteria 2
Bacter1	121Y	Bacteroidetes 1
Bacter2	122Y	Bacteroidetes 2
Bacter3	372Y	Bacteroidetes 3
Bathy	304B	*Bathycoccus*
Brach	323B	*Brachionus calyciflorus*
Cer/Po	325B	Cercozoa and Polychaetes
Cerco	267B	Cercozoa
ChaeP1	272Y	*Chaetoceros calcitrans* plastid 1
ChaeP2	273Y	*Chaetoceros calcitrans* plastid 2
Chaeto1	284B	*Chaetoceros calcitrans*1
Chaeto2	285B	*Chaetoceros calcitrans*2
Cili1	122B	Ciliate 1
Cili2	265B	Ciliate 2
Cyclo	531B	*Cyclotella choctawhatcheeana*
Cylin1	519B	*Cylindrotheca closterium* 1
Cylin2	521B	*Cylindrotheca closterium* 2
Dino1	118B	*Karlodinium* or *Amoebophrya*
Dino2	314B	Uncultured Dinoflagellate
Flavo	577Y	Flavobacteria
Gamma	304Y	γ-Proteobacteria
*KM	376B	mostly *Karlodinium micrum*
Leuco	280B	*Leucocryptos*
Mari	290Y	*Marivita, Rhodobacter*
Micro	564Y	*Microcella*
Minu	533B	*Minutocellus polymorphus*
Mixed pp1	258B	Mixed phytoplankton 1
Mixed pp2	260B	Mixed phytoplankton 2
Mixed pp3	262B	Mixed phytoplankton 3
*PA	372B	mostly *Picochlorum atomus*
PA+KM	374B	*Prochlorum atomus* and *Karlodinium micrum*

Pirso	215B	*Pirsonia*
Wolo	592B	*Woloszynskia*
Poly	379B	Polychaetes
Prym	256B	Prymnesiaphyte
PS	318B	*Picocystis salinarum*
PyP pla	562Y	*Pyramimonas parkeae* plastid
SAR11	555Y	SAR11
Skele	72B	*Skeletonema*
Strame	203B	Uncultured Stramenopile
Syn/PS1	474Y	*Synechoccocus* and *Picocystis salinarum* plastid 1
Syn/PS2	476Y	*Synechoccocus* and *Picocystis salinarum* plastid 2
Synd/Co	316B	Syndiniales and Copepod
Syndi	237B	Syndiniales
Telo	382B	*Telonema*
Yon	288Y	*Yonghaparkia*
Temp		Surface water temperature
Sal		Salinity
Wind		Average wind speed
Gust		Maximum gust speed
River		Connetquot River discharge on the sampling day
7day		Sum of Connetquot River discharge for the prior 7 days

The strong positive associations among AA1, AA2, AAP1 and AAP2 (Fig. 3.15.A.) were consistent with their identity as TRFs generated from *A. anophagefferens*. *A. anophagefferens* was positively related with wind speed and this was the only positive association captured in both AA (*A. anophagefferens* 18S TRFs 174B and 176B) and AAP (*A. anophagefferens* plastid TRFs 297Y and 299Y) subnetworks. There were more negative associations captured in both AA and AAP subnetworks than in other subnetwors, including those between *A. anophagefferens* and *Cylindrotheca*, *Minutocellus*, *Synechococcus*/*Picocystis salinarum*, and Syndiniales/Copepod. The AAP subnetwork also suggested that *A. anophagefferens* was negatively related with other pico-chlorophytes, i.e. *Picochlorum atomus* and *Bathycoccus.* There were positive associations between *A. anophagefferens* and *Chaetoceros* as well as the autotrophic dinoflagellate *Karlodinium*. However, such positive associations appeared only in the AAP subnetwork and only for one node representing *A. anophagefferens*, AAP2 (Fig. 3.15.B.).

Chaetoceros and non-*Chaetoceros* diatoms seemed to have different temporal patterns and form different associations with other TRFs. The positive associations among Chaeto1, Chaeto2, ChaeP1 and ChaeP2 were consistent with their identity as TRFs generated from *Chaetoceros* (Fig. 3.16.A. and Fig. 3.16.B.). Almost all relationships between *Chaetoceros* and other phytoplankton groups were negative, except with AAP2 (Fig. 3.16.A. and Fig. 3.16.B.). Temperature was the only environmental factor that was significantly correlated with *Chaetoceros* and the correlation was positive (Fig. 3.16.A.). The diatoms *Cyclotella* and *Minutocellus* were positively correlated with each other (Fig. 3.17.A.) and with other diatoms, *Cylindrotheca* (Cylin1 and Cylin2) and *Skeletonema*. Unlike *Chaetoceros,* which does not significantly associate with autotrophic dinoflagellates, non-*Chaetoceros* diatoms formed positive relationships with autotrophic dinoflagellates *Woloszynskia* and *Karlodinium micrum* (Fig. 3.17.A. and Fig. 3.18.A.). Additionally, while *Chaetoceros* had a positive relationship with *A. anophagefferens* and negative relationships with the blue-green picoalgae (*Synechococcus*/*Picocystis salinarum* and *Bathycoccus*), the small (~3 μm) non-*Chaetoceros* diatom *Minutocellus* was negatively associated with *A. anophagefferens* (AA1, AA2 and AAP1) and was positively associated with the blue-green picoalgae *Synechococcus*/*Picocystis salinarum*.

Three nodes representing pico-chlorophytes *Bathycoccus*, *Picochlorum atomus* and *Picocystis salinarum* were positively associated (Fig. 3.17.B.), with the first two also positively associated with *Synechococcus* (Syn/PS1, Syn/PS2). In fact, these 4 picophytoplankton (referred to as blue-green picophytoplankton later on) had similar temporal patterns and always had the same relationships with other TRFs. For example, they all had negative relationships with *A. anophagefferens* and *Chaetoceros* (AAP1, AAP2, ChaeP1, Chaeto1);, and*Picochlorum atomus* (*PA) and *Synechococcus*/*Picocystis salinarum* (Syn/PS1) had negative relationships with *Karlodinium*.

Dinoflagellates comprise ecologically diverse organisms. Autotrophic *Karlodinium* and *Woloszynskia* were positively correlated and they formed positive relationships with many common TRFs (Fig. 3.18.A.). Syndiniales (Syndi, Fig. 3.18.) is a parasitic dinoflagellate order known to attack other dinoflagellates, protozoa, ciliates, and even invertebrates (Hoek et al. 1994, Park et al. 2002, Stentiford & Shields 2005, Guillou et al. 2008). The negative association between Syndiniales (Syndi) and *Chaetoceros* was strongly supported because Syndi was

negatively associated with 3 TRFs representing *Chaetoceros* (Fig. 3.19.A.). *Amoebophrya* (Amoe, Fig. 3.18.) is a genus of Syndiniales. *Amoebophrya*'s associations were almost exclusively with 18S TRFs, with positive signs (Fig. 3.17.B. and Fig. 3.18.A.). Interestingly, *Amoebophrya* was positively associated with most diatoms except for *Chaetoceros* (Fig.3.18.). In addition to *Amoebophrya* and other Syndiniales, this study identified a few other heterotrophic eukaryotes that have not been reported in Great South Bay previously (Fig. 3.19.). *Pirsonia*, a parasitic nanoflagellate infecting planktonic diatoms (Kuhn 1998, Kuhn et al. 2004), had different associations with different groups of phytoplankton: while it was negatively associated with *A. anophagefferens* and *Chaetoceros*, it was positively associated with almost all blue-green picophytoplankton (Syn/PS2, *PA, Bathy). Similarly, while a TRF representing both Syndiniales and copepods (Synd/Co) was positively associated with all blue-green picoalgae (Fig. 3.17.B. and Fig. A2.), it was negatively associated with *A. anophagefferens* (AA1, AA2, AAP1, and AAP2) and *Chaetoceros* (Chaeto1 and Chaep1). *Telonema* is a genus of phagotrophic flagellates with cosmopolitan distribution (Lee et al. 2003, Klaveness et al. 2005). While negatively associated with most 16S TRFs including *Synechococcus* (Syn/PS1), *Telonema* was also negatively associated with *Chaetoceros* (Chaeto1). Cercozoa is a group of heterotrophic protists found to be negatively related with mostly 16S TRFs (Fig. 3.19.B.). A polychaete TRF, most likely from *Clymenura*'s planktonic larvae (Wilson 1991), was negatively related with not only some 16S TRFs but also *Picochlorum atomus* (Fig. 3.17.B.).

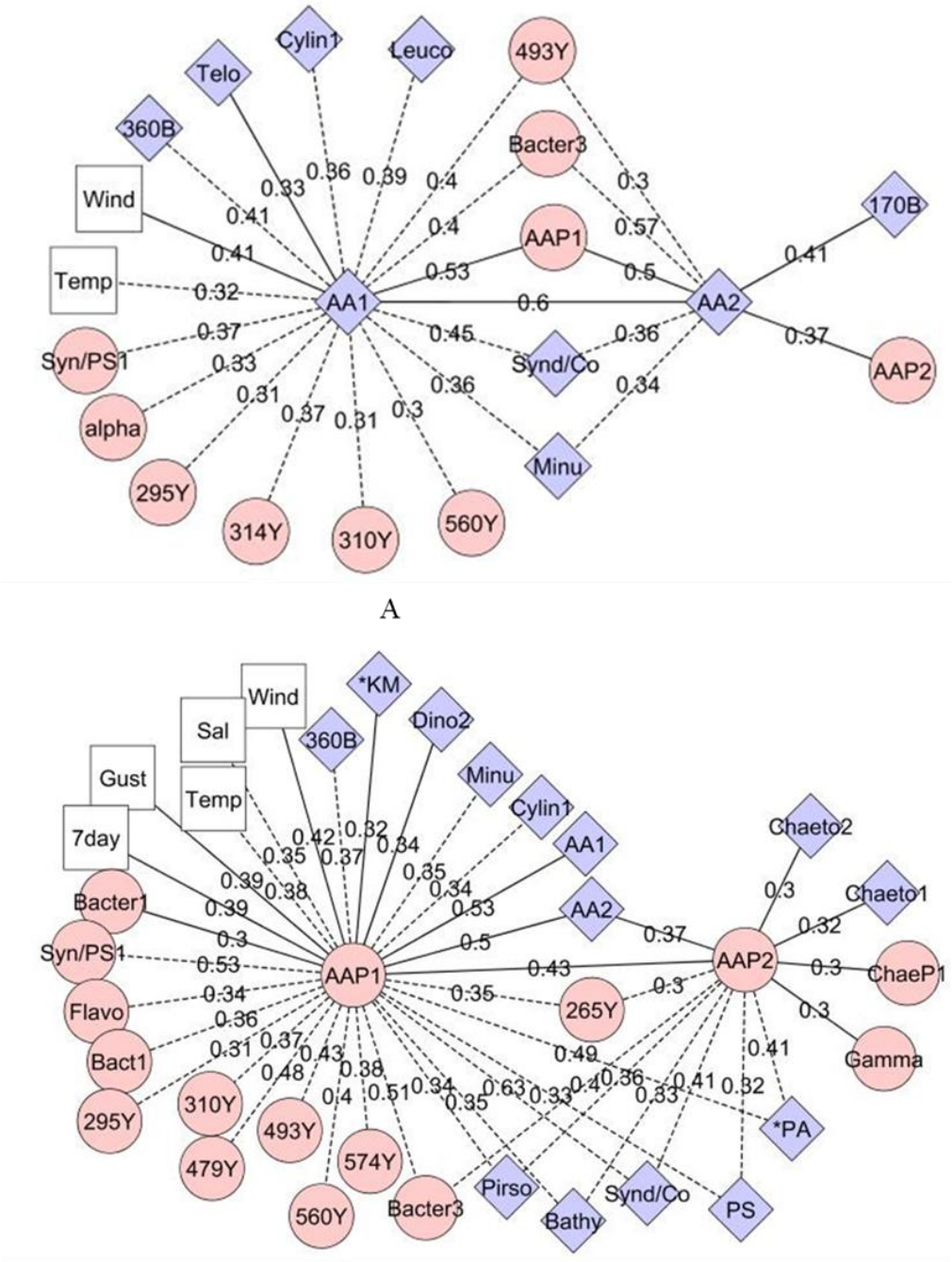
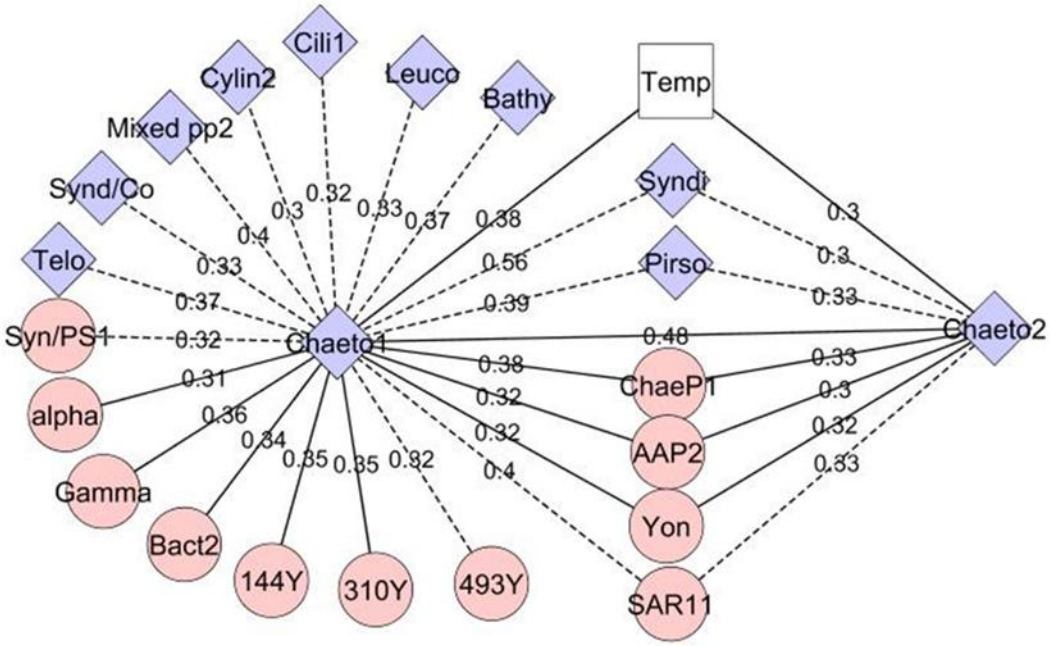
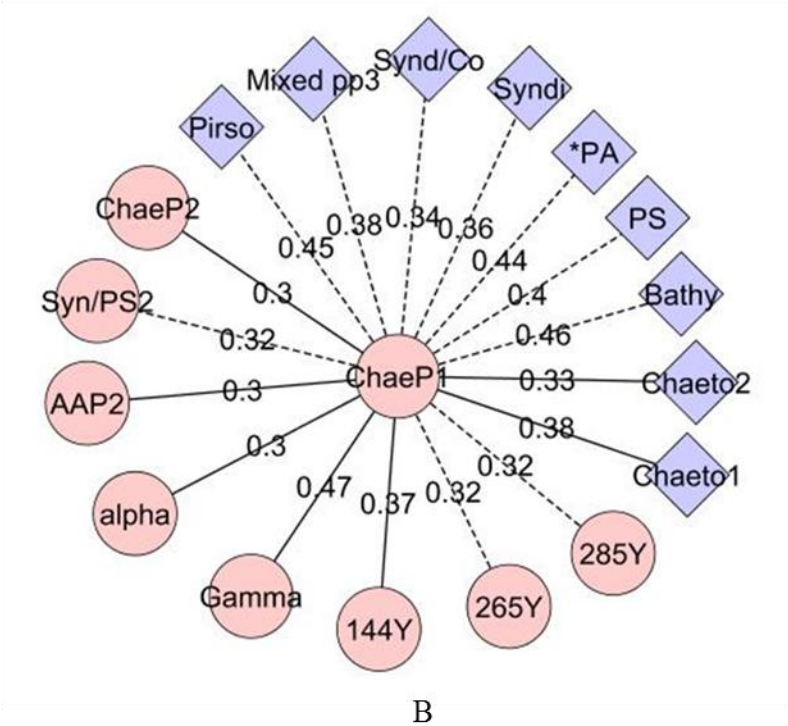


Fig. 3. 15. *A. anophagefferens* subnetworks A: With AA1 and AA2 as central nodes; B: With AAP1 and AAP2 as central nodes; Diamond: 18S nodes; Circles: 16S nodes; Square: Environmental factors; Solid lines represent positive relationships and dashed lines represent negative relationship. Numbers on the lines are LS scores. Subnetworks show only LS correlations between central nodes and their first neighbors but not among their first neighbors. The length of the connecting lines between nodes and the angle between lines are arranged just for the easiness of visualization.

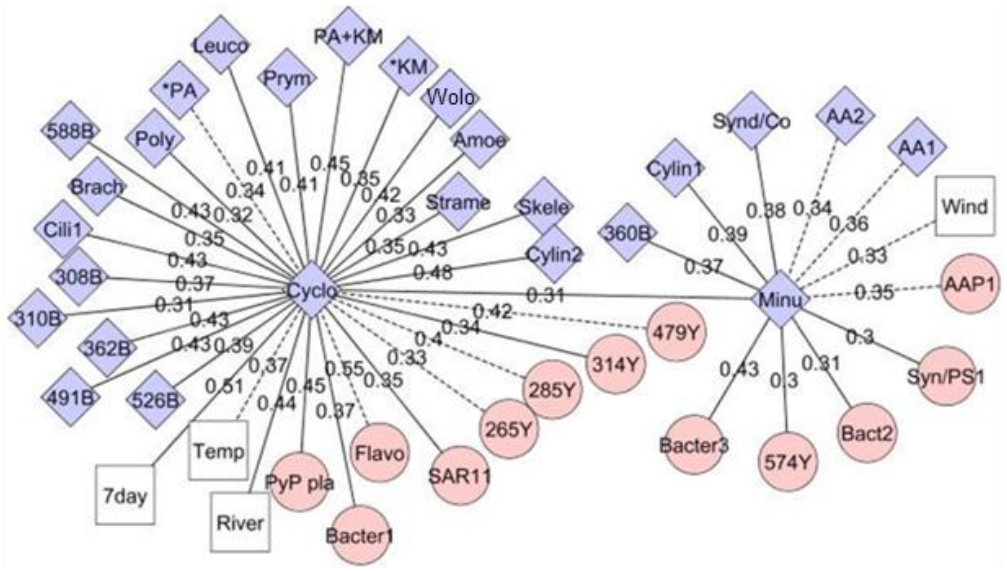


A

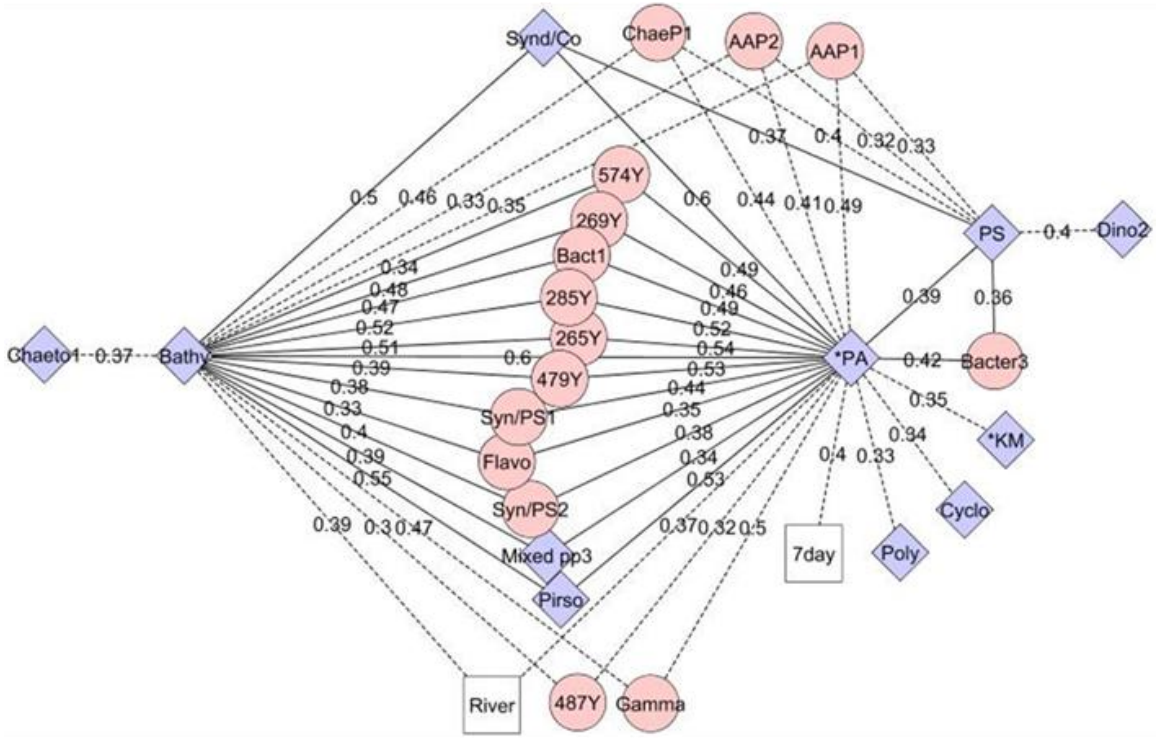


B

Fig. 3. 16. *Chaetoceros calcitrans* subnetworks A: With Chaeto1 and Chaeto2 as central nodes; B: With ChaeP1 and ChaeP2 as central nodes. For details see legend of Fig 3.15.



A



B

Fig. 3. 17. Subnetworks of two diatoms (A) and pico-chlorophytes (B).
For details see legend of Fig. 3.15.

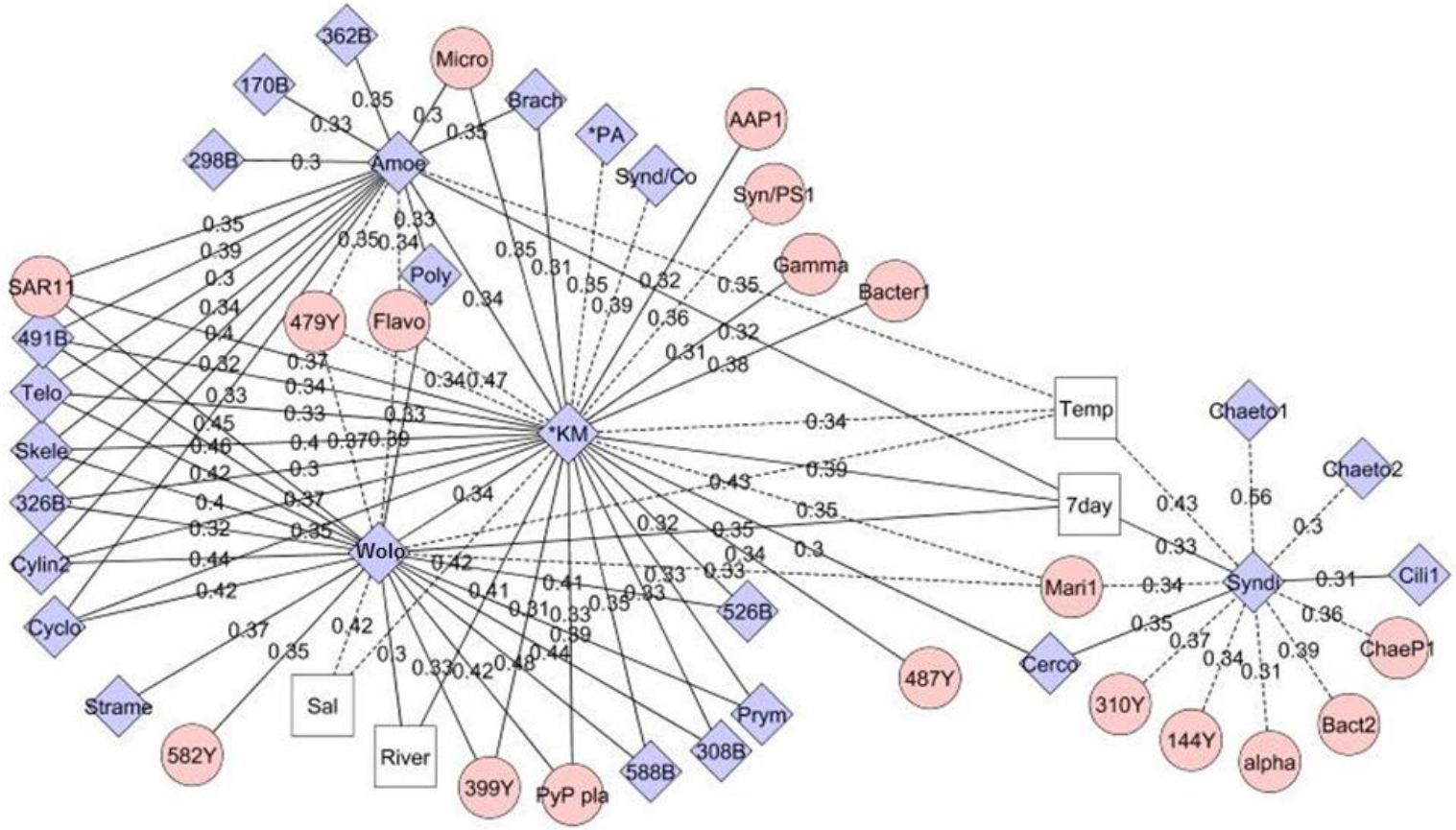
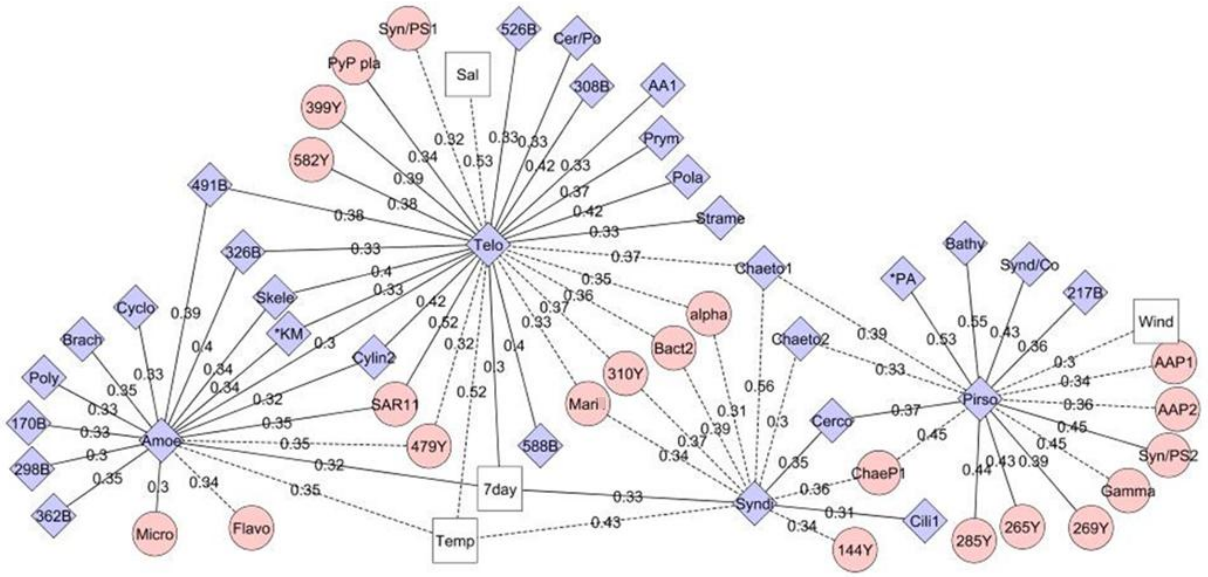
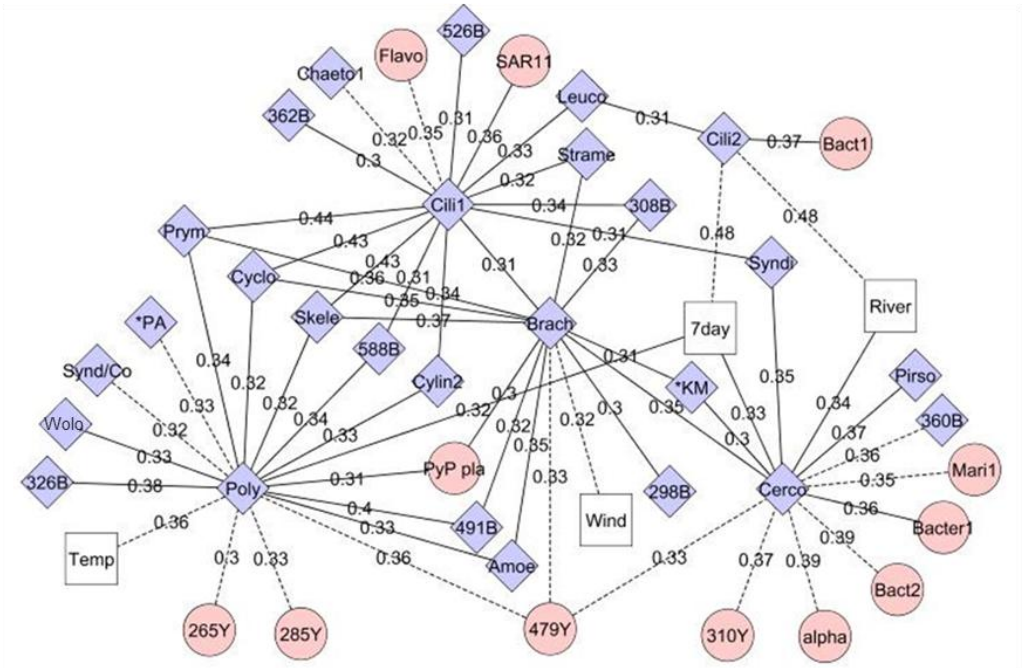


Fig. 3. 18. Subnetworks of dinoflagellates. For details see legend of Fig. 3.15.



A



B

Fig. 3. 19. Subnetworks of potential parasites and grazers A: *Telonema*, *Pirsonia*, Syndiniales and *Amoebophrya*; B: Ciliates, *Brachionus*, Cercozoa and polychaetes. For details see legend of Fig. 3.15.

2 Time-delayed associations between TRF-TRF and TRF-Environmental Factor (TRF-EF) pairs

Adding a time delay to the LS analysis sacrificed almost two- thirds of the information (from 55 samples to 19 samples; see Materials and Methods) and this resulted in both higher p values for LS correlations and fewer significant TRF-TRF and TRF-EF pairs even at p=0.05 (80 nodes and 393 significant LS correlations with $p<0.05$, $q<0.05$ and $LS>0.3$). Overall, LSA runs using LSM and LSM-T agreed with each other since the majority of the significant LS correlations observed with a time delay were already reflected in the original LSA, and time delayed LSA did not reveal strong interactions that were not already found in the original LSA. Because the time delay was set at 4 in LSM-T analysis, there were 9 possible time lags between data series (from -4 to 4), so significant associations with delay=0 should in theory be about 1/9 (11%) of the total significant associations, assuming all time delayed associations were equally likely to occur. However, among the 393 significant LS correlations identified, 141 were instantaneous correlations with delay=0, which was about 36% of the total significant associations. As the delay increased, the proportion of significant associations also decreased (±1 delay: 22%; ±2 delay: 20%; ±3 delay: 14%; ±4 delay: 8%). A few cases where significant associations were only captured in time-delayed LSA are shown below.

A. anophagefferens (AA1) and *Chaetoceros* (Chaeto1) were positively associated with a time delay of 2 weeks (Fig. 3.20., upper panel). However, *A. anophagefferens'* peak at time point 8 was unmatched by *Chaetoceros*; neither was the increasing trend of *A. anophagefferens* at the end of the sampling period reflected in the time-delayed analysis. The negative association between *Synechococcus*/*Picocystis* and Cercozoa was captured with the algae falling behind Cercozoa by 2 weeks (Fig. 3.20., middle panel). Dino2 was composed of both Peridiniales and *Amoebophrya* and its peak at time point 9 was followed closely by an increase in the relative abundance of one ciliate group (Cili1). A time delay of 3 weeks made Dino2 and Cili1 demonstrate almost opposite temporal patterns (Fig. 3.20., lower panel).

For TRF-EF pairs, significant correlations missed in the LSM run but found in the LSM-T run were mostly those with a time delay of 3 or 4 weeks. The only exception was the relationship between SAR11 and river discharge, which was not a significantly associated pair in the LSM run but was in LSM-T with a time delay of only 1 week (Fig. 3.21., middle panel).

However, the direction of the time delay indicated the relative abundance of SAR11 changed first and Connetquot River discharge followed. Neither was the negative association between AAP1 and salinity (Fig. 3.21., upper panel) convincing since the relative abundance of AAP1 was stable for the second part of the dataset while salinity was variable during the same time. The negative association between *Picocystis salinarum* (PS) and 7 days' sum of Connetquot River discharge was quite clear after moving the river discharge forward for 4 weeks (Fig. 3.21., lower panel).

The examples chosen above are among the more convincing cases (at least graphically) of LSA with a time delay because only a relatively small portion of the original data was excluded. Examination of other time-delayed associations revealed that the majority of the significant cases found only in the time delayed LSA were characterized by exclusion of original data from usually at least 5 weeks, during which the associated pairs did not show any relationship. Even in cases in Fig. 3.20. and Fig. 3.21., where the time-delayed associations did not require a significant sacrifice of the original data, these associations were never repeated by different TRFs representing the same organisms. For example, no time-delayed negative association was ever found between Cercozoa and the other Syn/PS TRF, Syn/PS1.

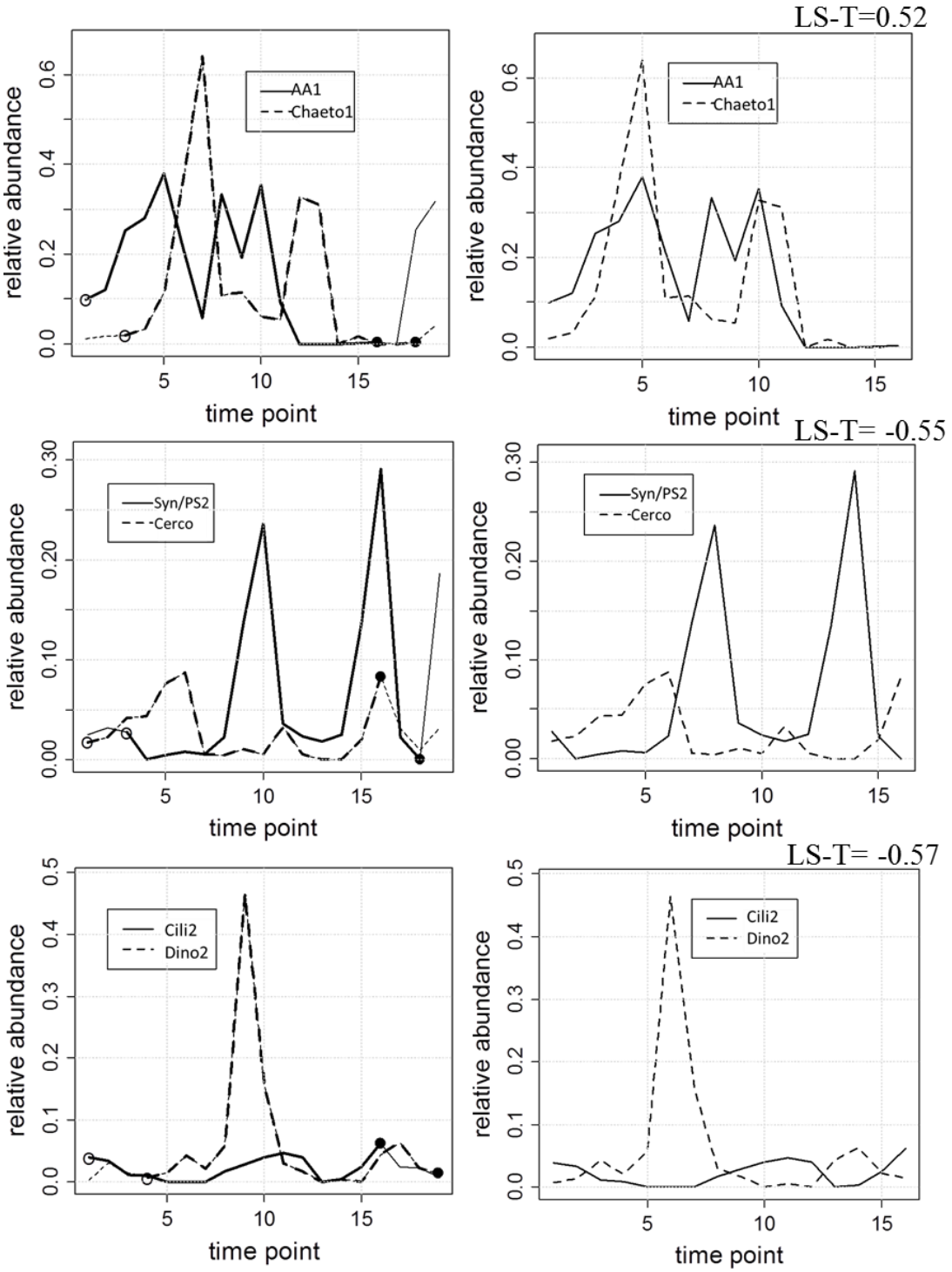


Fig. 3. 20. Significantly associated TRF-TRF pairs with time delay. Open and closed circles indicate the starting and ending points of the association. Left column: original data; right column: time-delayed alignment. Upper, middle and lower panels represent different pairs.

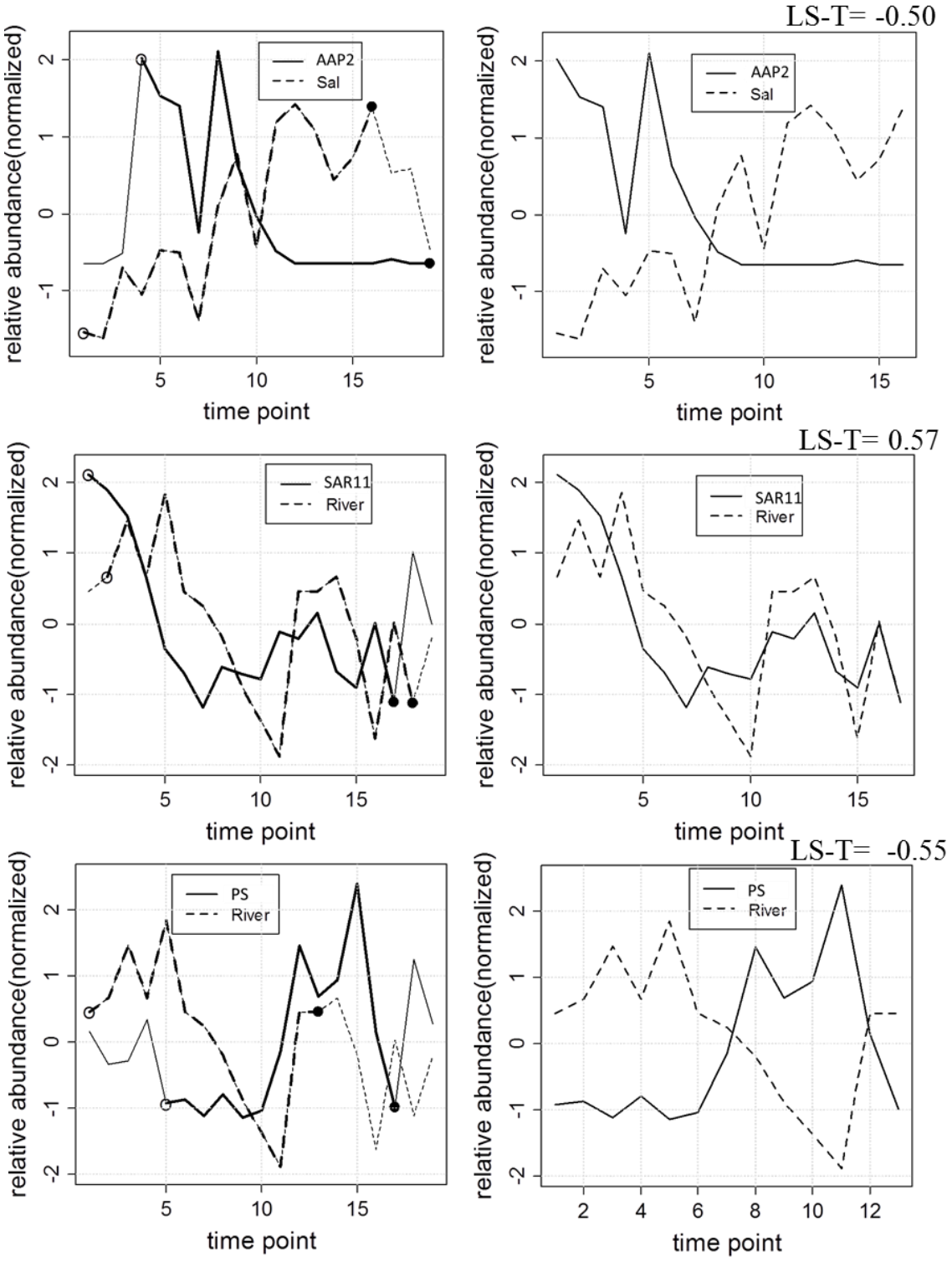


Fig. 3. 21. Significantly associated TRF-EF pairs with time delay. Open and closed circles indicate the starting and ending points of the association. Left column: original data; right

column: time-delayed alignment. Upper, middle and lower panels represent different pairs.

Discussion

Seasonal patterns in the GSB plankton community

The combination of TRFLP and cloning/sequencing enabled me to track microbial community changes through time semi-quantitatively with useful taxonomic information at least for the most abundant TRFs. I am aware of the limited resolving power of restriction enzymes, as one TRF may comprise taxonomically unrelated microbes. *Taq*I was chosen specifically because first, it generated the most reproducible results (Chapter 2). Second, *in silico* digestion of 18S sequences of a few common phytoplankton species using a variety of restriction enzymes predicted a unique TRF, 174B, representing *A. anophagefferens* only for *Taq*I. Consistent with that preliminary analysis, among all sequences retrieved in 2008, *A. anophagefferens* was the only organism with a predicted TRF matching 174B. The primer set used to amplify bacterial 16S rDNA also amplified plastid 16S rDNA of *A. anophagefferens*, *Picocystis salinarum* and a diatom (possibly *Chaetoceros calcitrans*, see below). Among the observed 16S TRFs, 297Y was specific for *A. anophagefferens*. The contributions of both c175B and c298Y to total TRF area, representing the relative abundance of *A. anophagefferens*, agreed well with microscopic cell count data representing the absolute abundance of *A. anophagefferens* (Fig. 3.5.). This enabled me to follow the progression of the brown tide bloom as well as many other important groups in the microbial community with greater temporal, spatial, and taxonomic resolution than previous studies.

Lively et al. (1983) reported that 2-4 μm chlorophytes ("small forms") and cryptophytes were consistently abundant in 9 samples collected throughout the year from September 1979 to June 1980, while diatom and dinoflagellate cell densities varied over time. From May to June, the phytoplankton community was dominated by *Rhodomonas minuta* and an unidentified 5 μm microflagellate. In this study, *A. anophagefferens*, which was absent or undetected in 1980, was a major component of the community in May and June, 2008. In Lively's study, total cell abundance of *Chaetoceros teres* and an unidentified *Chaetoceros sp.* were 79200 and 5300 cells L^{-1} on the last two sampling days, May 29 and June 12, but were absent at other sampling times. In mid June 2008, c284B representing *Chaetoceros* (most likely *Chaetoceros calcitrans* f. *pumilus* Takano) was the most dominant TRF, accounting for 45-85% of total TRF area. The almost synchronous pulse of *Chaetoceros* in mid-June nearly 30 years apart is striking.

John Ryther (Ryther 1954) reported that two "small forms" of green algae, *Nannochloris atomus* and *Stichoccoccus cylindricus* (both belonging to the class *Trebouxiophyceae*), were the most abundant species in GSB in 1952. *Nannochloris atomus*, identified in 1952 (Butcher 1952), was also recognized as a resident bloom-forming species in GSB in the 1980s (Olsen 1989), and is currently regarded as a taxonomic synonym of *Picochlorum atomus* (Henley et al. 2004). *Picochlorum atomus* rDNA was cloned and sequenced in this study and contributed to one of the major 18S TRFs, c374B. *Stichoccoccus cylindricus* was temporarily identified by Ralph Lewin with uncertainty (Lewin 1954). In addition to *Picochlorum*, 18S sequences were retrieved the pico-chlorophytes *Bathycoccus* and *Picocystis*. Pico-chlorophytes have clearly been important in GSB for at least the last 6 decades. The specific group of pico-chlorophytes dominating in GSB, however, may vary each year. For example, while *Picocystis* and *Picochlorum* were important in 2008, *Bathycoccus* was a more important pico-chlorophyte in 2009 (Chapter 4).

The presence of *Synechococcus* in GSB has been reported since the 1980s (Sieburth et al. 1988, Keller & Rice 1989, Kelly & Chistoserdov 2001, Gobler et al. 2002, Gobler et al. 2004b). TRF c475Y represents both *Synechococcus* and *Picocystis* plastid, with the former probably being the more abundant (See 'Major TRFs' in Results). The temporal pattern of c475Y% was reciprocal to that of c298Y% (the TRF specific for *A. anophagefferens* plastid) and there was a negative relationship between c475Y and c298Y according to LS analysis. Both c298Y% and c475Y% were low in the first 2 sampling weeks on May 8 and May 15 (averages of all stations over the two weeks were 9.7% and 2.8%) and the plankton community was comprised of many phytoplankton species with various cell sizes present and no single species nearly as dominant as *A. anophagefferens*, *Synechococcus* and *Chaetoceros* (in week 7) later became (Table 3.11.). In fact, the 4 biodiversity indices of all week 1 and week 2 samples (n=7) were all significantly higher than those of the rest non-bloom samples (n=17) (T test, $p<0.005$).

Table 3. 11. Phytoplankton detected in GSB on May 8 and May 15 2008 and the average area contribution from the corresponding TRFs

Diatom	**Dinoflagellate**	**Chlorophyte**	**Mixed Phytoplankton**
Skeletonema 72B: 1.2%	*Woloszynskia* and *Alexandrium* 592B~596B: 12.3%	*Bathycoccus* 304B: 1.1%	Dinoflagellate, Chlorophyte and some non-phytoplankton 314B~318B: 5.9%
Chaetoceros 284B: 1.3%			Chlorophytes and Dinoflagellates 372B~376B: 8.1%
Cylindrotheca 520B: 2.1%			Haptophyte, Chlorophyte, Rhodophyte, Diatom and Dinoflagellate 256B~265B: 9.4%
Cyclotella and *Minutocellus* 532B: 8.5%			

Brown tide bloom in Great South Bay and week 7 anomaly

After its first appearance in GSB in 1985, a monitoring effort was begun to follow *A. anophagefferens* (SCDHS 2011). Historical data (Nuzzi & Waters 2004, Gobler et al. 2005, SCDHS 2011) shows that *A. anophagefferens* has been present in GSB since at least 1985 and developed to Category 2 bloom concentrations (with potential negative effects on shellfish feeding) of more than 35000 cell/ml (Gastrich & Wazniak 2002) frequently. Both late spring-early summer blooms and fall blooms of *A. anophagefferens* have been observed in GSB with blooms always disappearing in mid-summer (Gobler et al. 2002, Gobler et al. 2004a). 2008 featured the most widespread and intense brown tide observed in GSB for more than a decade (LoBue & Doall 2010). This study (Fig. 3.5.) found that *A. anophagefferens* reached bloom abundances around mid-May and declined to below detection from late July to mid-September, followed by a fall bloom with about the same highest relative abundance as the spring-summer bloom, consistent with cell count data. Weekly samples processed using TRFLP also revealed that there were rapid changes in c175B% between consecutive weeks. For instance, the concurrent drop of c175B% in 18S TRFLP and c298Y% in 16S TRFLP (Fig. 3.3. and Fig. 3.4.) between June 12 (week 6) and June 19 (week 7), and their simultaneous recovery by the next

week, was convincing evidence that the relative abundance of *A. anophagefferens* changed dramatically within a week. No microscopic counts of *A. anophagefferens* cell abundance are available from June 19 to confirm whether this also represented an absolute decline. Rapid changes in *A. anophagefferens* cell concentration in consecutive weeks, however, are not rare. For example, cell count data collected at the west side of Smith Point shows that *A. anophagefferens* abundance increased from 5556 cells/ml on June 19 to 34641 cells/ml on June 26 (SCDHS 2011).

The distinct community structure on June 19, characterized mainly by the shift of c175B% and c284B% in 18S TRFLP as well as the shift of c298Y% and 577Y% in 16S TRFLP (Fig. 3.3. and Fig. 3.4.), led us to hypothesize that a short-lived physical force caused a temporary change in the community structure. The study site is close to the Connetquot River, a source of freshwater input. June 14 to June 16 were rainy days with thunderstorms, followed by a dry day on June 17 and another rainy day with thunderstorms on June 18 (http://www.wunderground.com). The increase in c284B% and 577Y% on June 19 could reflect a response to the cumulative effect of the weather from June 14 to June 18. With this said, June 19 was not a particular wet day and the accumulated precipitation in the previous 7 days was not the highest among all sampling days, either (Fig. 3.2.C.). It is possible that both the freshwater input (with different microbial composition than that present in GSB) accumulated through continuous rainy days and nutrient supply or speciation, especially nitrogen (see below) present in the freshwater input at that particular time, may have both facilitated the formation of the distinctive microbial communities on June 19.

Brown tide blooms have been proposed to be initiated and maintained only under particular nutrient regimes. Laroche et al. hypothesized that brown tide blooms occur in years when the nitrate-rich groundwater discharge is low (Laroche et al. 1997), but Gobler and Sanudo-Wilhelmy found that groundwater inputs can initiate a non-brown-tide phytoplankton bloom prior to brown tide events and the remineralized organic N from the non-brown-tide phytoplankton bloom can stimulate the growth of *A. anophagefferens* (Gobler & Sanudo-Wilhelmy 2001b). Taylor et al. suggested that NO_3^- tends to support a phytoplankton community that is dominated by diatoms, some chlorophytes, and *Synechococcus,* while low-to-moderate ($\leq$10 μM) concentrations of reduced N such as ammonium and urea will more likely support

dominance of *A. anophagefferens* (Taylor et al. 2006). Gobler et al. (unpublished data) measured NH_4^+ and NO_3^- concentrations at a station in GSB about 8 km to the west of our sites. Looking at NH_4^+, NO_3^-, c175B% and c284B% together (Fig. 3.22.), some temporal trends are striking. As NH_4^+ was drawn down from an initial concentration of ~5.6 μM in mid-May to barely above zero on June 5, c175B% reached its first peak, which seems to agree with Taylor et al. that moderate concentration of NH_4^+ may have favored the growth of *A. anophagefferens.* The 'low-high-low' temporal trend for both of NH_4^+ and c284B% on June 5, June 19 and July 1 suggested that NH_4^+ and *Chaetoceros* were brought into GSB together through freshwater input. From June 19 to July 1, as NH_4^+ was drawn down again, c175B% reached a secondary peak. Throughout the sampling season when both nutrient data and TRFLP data were available, NO_3^- (the nitrogen form preferred by diatoms) was always lower than NH_4^+, which may have contributed to the fact that *Chaetoceros* did not dominate over *A. anophagefferens* except on June 19. With this said, note that high NH_4^+ concentrations were not recorded prior to the increase of c175B% in October (Fig. 3.22.), which indicates that there were other factors regulating the relative abundance of the brown tide species.

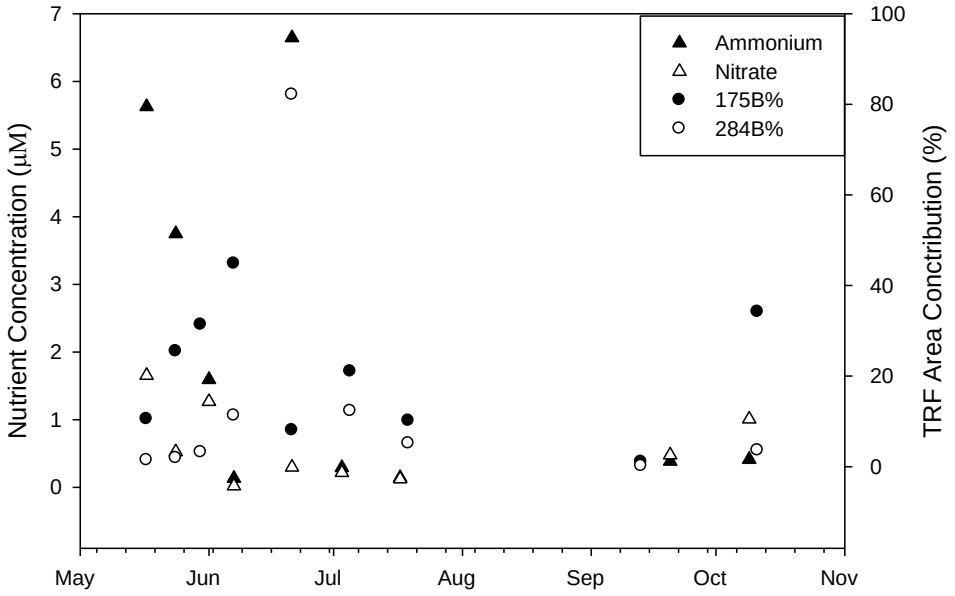


Fig. 3. 22. Average c175B% and c284B% at this study site and NH_4^+ and NO_3^- concentrations at a station 8 km to the west. Only days when both TRF data and nutrient data were available are shown here.

Spatial variation: microbial biogeography?

1 Spatial variation of the plankton community

There has been a lot of discussion surrounding the question of whether microorganisms have biogeographical distributions comparable to those of macroscopic organisms. As the focus of biogeography is the distribution of biodiversity over space and time (Martiny et al. 2006), one can see the difficulty in approaching this question right away: how do we define microbial biodiversity? This study, as many other studies (Green et al. 2004, Noguez et al. 2005, Langenheder & Ragnarsson 2007, Pommier et al. 2007, Fuhrman et al. 2008), uses an operational taxonomic unit, TRF in this case, as a measure of diversity. For the weeks examined in spatial analysis, Bray-Curtis distance increased with geographical distance for 18S but not for 16S TRFLP, suggesting that spatial structure over a 10 km scale in GSB existed for eukaryotic but not prokaryotic microbial plankton (Fig. 3.10. and Fig. 3.13.). First off, my data disagrees with Finlay and Fenchel's conclusion that small protists have no biogeographical pattern (Fenchel & Finlay 2004). On the other hand, the difference seen between 18S and 16S communities seems to be in agreement with the hypothesis that cell size and population abundance will likely determine whether organisms have cosmopolitan distributions or whether they demonstrate biogeographic patterns (Hillebrand et al. 2001). Hillebrand et al. argued that compared to multicellular organisms, unicellular organisms have higher dispersal ability and higher abundance that enable them to have ubiquitous distributions. Therefore, although there was a general trend of decreasing community similarity with increasing geographical distance for benthic diatoms and ciliates, a steeper slope between similarity and distance was found for metazoans such as corals. My study suggests that this size-based differential response of community similarity to geographical distance can be extended to the world of microorganisms, where eukaryotic cells are in general bigger and less abundant.

2 Spatial variation of *A. anophagefferens'* relative abundance

Among the 8 weeks demonstrating significant spatial structure (Fig. 3.11.), the relative abundance of *A. anophagefferens* displayed a roughly monotonic along-transect trend in week 11, week 19 and week 21 (Fig. 3.23.). Week 21 (September 25) was unique in that all the major TRFs assumed an almost monotonic along-transect trend with c175B% being the opposite of other TRFs (lower panel of Fig. 3.12.). Week 21 was also the only week when the salinity at station J (25.4 PSU) was lower than that at station E (27.4 PSU), which was an unusual condition in GSB since saltier open ocean water enters GSB through the Fire Island Inlet which is closer to station J. This unique salinity structure in week 21 was most likely the result of wind-driven advection along-shore. The average wind speed on September 25 was the highest, at 14 mph, of all sampling days (Fig. 3.2.E.). More importantly, a northeast wind had persisted for 4 days including September 25, which was quite unusual during the sampling period (September 16 to 20 had similar wind direction but samples were not taken then, Fig. A3.). Therefore the saltier water sample collected at station E (27.4 PSU) may be from water that passed Smith Point from Moriches Bay, whose salinity was higher than GSB at ~28 PSU and whose *A. anophagefferens* cell abundance doubled that in GSB (SCDHS 2011), and moved along the south shore of Long Island (Fig. 3.1.A.). Wind pushes water westward most effectively along the shoreline due to the shallower water depth (personal communication with Charles Flagg). Due to the closeness to the shoreline of stations A to D, water collected at these stations was likely most strongly influenced by the saltier water mass with relatively greater c175B%, originally from Moriches Bay. Stations H to J are located furthest away from the shoreline, and were likely less influenced by the wind-driven alongshore current originally from Moriches Bay. Therefore, water collected from these stations most likely represented GSB 'local' water with relatively lower salinity and c175B% (Fig. 3.23.). In the middle of the transect, from station E to G, the 2 water masses appeared to mix, indicated by the change in all major TRFs between stations (Fig. 3.12.). Wilson et al. pointed out that this type of along-shore water flow driven by wind accounts for ~20% of the current variance in GSB and often occurs over a time scale of ~3 days (Wilson et al. 1991).

In contrast to week 21, c175B% increased from station A to station J in week 11 and week 19 (Fig. 3.23.) and salinity at station J was higher than that at station E (Fig. 3.2.B.). Weeks 11 and 19 differed from each other in that the increase of c175B% started further north in week 11 than in week 19, the average c175B% in week 11 was higher than that in week 19, and the

average salinity in week 11 was also higher than that in week 19. No environmental parameter (temperature, salinity difference at station E and J, wind and river discharge) seemed to have driven the spatial heterogeneity of the microbial community observed in the 2 weeks.

The along-transect pattern of c175B% and salinity in weeks 11, 19 and 21 (Fig. 3.23.) seemed to suggest that there was an association between relative abundance of *A. anophagefferens* and more saline water. However, neither NMS ordination nor LSA indicated a positive correlation between any *A. anophagefferens* TRF and salinity. Contradictory observations on the effect of salinity on brown tide have been made in previous studies. Although lab cultures of *A. anophagefferens* demonstrated higher growth rates at 30 PSU than at lower salinities (22, 24, 26 and 28 PSU) (Cosper et al. 1989), samples obtained from 12 stations across the Peconic Bay (Fig. 1.1.) and GSB revealed a negative relationship between salinity and *A. anophagefferens* cell density from 26 PSU to 30 PSU (Cosper et al. 1989). A straightforward relationship between *A. anophagefferens* cell density and salinity may not exist because not only may salinity affect *A. anophagefferens*' growth directly, but probably more importantly, water in GSB with different salinities may represent waters with different nutrient loads and different planktonic competitors and predators, and all these will affect *A. anophagefferens*' growth in different ways. Imposed on this complexity are physical factors such as wind, which may have caused the unique along-transect trend for both c175B% and salinity during week 21, as discussed above.

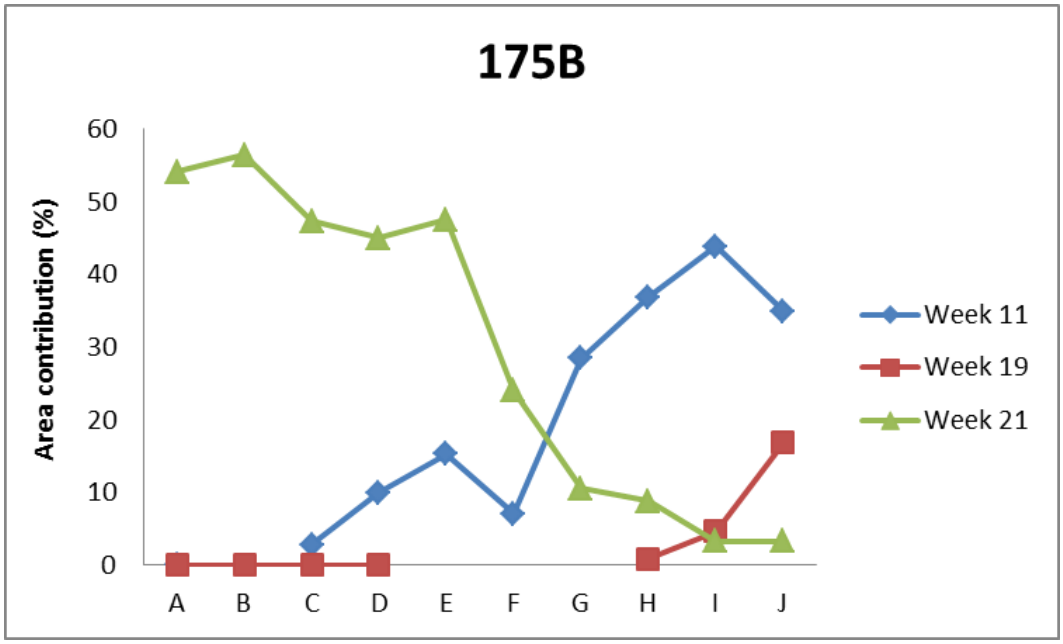


Fig. 3. 23. Area contribution of 175B to total TRF peak area along the transect in week 11, week 19 and week 21

Physical factors working on water masses, however, should not distinguish between 18S and 16S communities. The fact that a steeper slope between Bray-Curtis distance vs geographical distance was seen in 18S than in 16S TRFLP suggested an essential difference of distribution pattern between 18S and 16S communities, reflecting the existence of biogeography in the former but not so much in the latter.

Interactions within the microbial community

1 Potential microbial interactions and biodiversity

Among the phytoplankton examined in networks as central nodes, *A. anophagefferens* and *Chaetoceros* were different from the others (such as *Cyclotella*/*Minutocellus*, *Bathycoccus*/ *Picocystis*/*Picochlorum* and *Woloszynskia*/*Karlodinium*) in that their associations with other 18S TRFs were more often negative (Fig. 3.15., Fig. 3.16., Fig. 3.17. and Table 3.12.). In agreement with what was suggested by biodiversity analysis (Table 3.8. and Table 3.9.), network analysis supported the hypothesis that the bloom samples were associated with lower diversity because *A. anophagefferens* exerted more direct, competitive (or grazer-deterring) influence on the rest of the 18S community. Similarly, the biodiversity indices of samples dominated by *Chaetoceros* (weeks 6, 7, 12, 13) were significantly lower than at the rest of the samples (Table 3.13.).

Table 3. 12. Ratio of significant positive to negative interactions in LSA for *A. anophagefferens*, *Chaetoceros*, diatoms, picochlorophytes and dinoflagellates with other 18S TRFs (internal interactions such as those among AA1, AA2, AAP1 and AAP2 and that between Cyclo and Minu are not counted)

	$\frac{\#\,of+\,interactions}{\#\,of-\,interactions}$ with 18 TRFs	$\frac{\#\,of+\,interactions}{\#\,of-\,interactions}$ with 16S TRFs
A. anophagefferens (AA1, AA2)	2/8=0.25	0/10=0
A. anophagefferens (AAP1, AAP2)	5/13=0.38	3/13=0.23
Chaetoceros (Chaeto1, Chaeto2)	0/11=0	4/9=0.44
Chaetoceros (ChaeP1, ChaeP2)	0/7=0	3/4=0.75
Synechococcus/Picocystis (Syn/PS1,	6/6=1	9/7=1.29
Cyclotella and *Minutocellus*	21/3=7	5/8=0.625
Bathycoccus, *Picochlorum*, and *Picocystis*	7/5=1.4	13/20=0.65
Woloszynskia and *Karlodinium*	25/2=12.5	5/11=0.45

In contrast to 18S TRFs, the presence of a brown tide bloom did not seem to have any effect on 16S biodiversity (Table 3.9.). One possible explanation is that *A. anophagefferens* may compete more with other phytoplankton, which are mostly eukaryotic organisms. However, both AA and AAP were also negatively associated with almost all 16S TRFs they had significant associations with (Fig. 3.15. and Table 3.12.). The key to understanding the different effects of a brown tide bloom on 18S and 16S biodiversity may lie in the differences in the interactions among members of the 18S and 16S communities. There were 195 and 144 significant LSA

correlations for 18S TRF pairs and 16S TRF pairs, respectively, 17% and 18% of all possible LSA correlations. What distinguished 18S and 16S interactions was the proportion of significant negative LS correlations. For 18S, 27 out of 195 (14%) significant LS correlations were negative while 55 out of 144 (38%) were negative for 16S. The values in Table 3.12 reflect this distinction: there were proportionately more positive associations of non-*Chaetoceros* diatoms, pico-chlorophytes, and autotrophic dinoflagellates with other eukaryotes than with prokaryotes. This leads to a hypothesis that a community with proportionally more negative associations (16S community) is more stable and less likely to be disrupted by an overdominant species (*A. anophagefferens*) while a community with proportionally more positive associations (18S community), although it may support high diversity under normal conditions, is more susceptible to disruption by an overdominant species. All the 16S biodiversity indices were statistically the same between samples dominated by *Chaetoceros* and the other samples (Table 3.14.), which supported the idea that the 16S community is less likely to change with dominance of a single species.

Table 3. 13. T tests of Richness (N), Shannon's H', Simpson's D and Pielou's J of 18S community for *Chaetoceros*-dominated (CD, c284B %> 15% in weeks 6, 7, 12, 13) and non-*Chaetoceros*-dominated samples (NCD, c284B %< 15%)

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	NCD	CD	NCD	CD	NCD	CD	NCD	CD
Range	6~13	5~21	1.153 ~2.123	1.255 ~2.907	0.533 ~0.822	0.660 ~0.938	0.643 ~0.844	0.714 ~0.980
Mean	13.02	8.857	2.174	1.728	0.836	0.752	0.855	0.801
Observations	49	7	49	7	49	7	49	7
P(T<=t) one-tail	=0.001		=0.002		=0.002		=0.018	

Table 3. 14. T tests of Richness (N), Shannon's H', Simpson's D and Pielou's J of 16S community for *Chaetoceros*-dominant (CD, c284B %> 15% in weeks 6, 7, 12, 13) and non-*Chaetoceros*-dominant samples (NCD, c284B %< 15%)

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	NCD	CD	NCD	CD	NCD	CD	NCD	CD
Range	10~14	7~15	1.978 ~2.309	1.668 ~2.358	0.783 ~0.879	0.752 ~0.882	0.771 ~0.900	0.798 ~0.929
Mean	11.00	12.43	2.060	2.156	0.831	0.838	0.863	0.857
Observations	44	7	44	7	44	7	44	7
P(T<=t) one-tail	=0.015		=0.070		=0.305		=0.342	

2 A schematic model of microbial interactions and its implications for brown tide blooms

The associations among major ecological groups found in LSA network analysis, including some grazers and parasites, were organized into a schematic diagram to summarize their potential interactions (Fig. 3.24.). Associations included in this heuristic model are mostly those supported by multiple TRF pairs. For example, the positive association between Syndiniales/copepod and blue-green picophytoplankton was based on 4 positive associations between the 4 picophytoplankton and Syndiniales/copepod. Of all the significant associations, correlations between blue-green picophytoplankton and *Chaetoceros* (Fig. 3.16.), between blue-green picophytoplankton and *A. anophagefferens* (Fig. 3.15.), and between autotrophic dinoflagellates and non-*Chaetoceros* diatoms (Fig. 3.17.A. and Fig. 3.18.) were considered strongest because they were found among many pairs of TRFs comprising these groups. Neither LSA nor Pearson Correlation found a negative association between *A. anophagefferens* and *Chaetoceros* even though they dominated the ecosystem at different times. Instead, they may be positively associated, as indicated by the original LSA through the associations of AAP2 and Chaeto1&2 (Fig. 3.15.B.) and by the time delayed LSA (Fig. 3.20., upper panel). Because no positive association was found between nodes AA and nodes Chaeto/ChaeP in LSA, and only in LSA-T when the last 3 weeks were excluded (when c175B% increased but c284B% did not), the positive association (Fig. 3.24) between *A. anophagefferens* and *Chaetoceros* needs further confirmation.

The small (2-3 μm) centric diatom *Minutocellus polymorphus* has previously been isolated from *A. anophagefferens* bloom waters in Narragansett Bay, RI (Cosper et al. 1987, Sieburth et al. 1988) and also was found in the same area during a brown tide bloom (Smayda & Villareal 1989). LSA revealed that this species had a negative association with *A.*

anophagefferens (Fig. 3.15. and Fig. 3.17.A.), suggesting that competition may exist between them. The other small (3-5 μm) centric diatom *Cyclotella choctawhatcheeana* was positively associated with *M. polymorphus*, although the former did not demonstrate any relationship with *A. anophagefferens*. Based on the association between *C. choctawhatcheeana* and *M. polymorphus*, as well as the positive association between these two species and other diatoms except for *Chaetoceros calcitrans* (Fig. 3.17.A.), my data suggests that a group of diatoms excluding *Chaetoceros* shared a similar temporal pattern in GSB in 2008 and so they are illustrated as an ecological group in the schematic model (Fig. 3.24.).

According to Sieracki et al. (1999), the picoalgal niche is typically occupied by *Synechococcus*, which must be selectively removed or reduced to open the niche to *A. anophagefferens*. The 4 picophytoplankton identified in this study (*Picochlorum*, *Picocystis*, *Bathycoccus* and *Synechococcus*) were positively associated with each other (Fig. 3.17.B.) and consistently demonstrated negative relationships with both *A. anophagefferens* and *Chaetoceros* (Fig. 3.15., Fig. 3.16., and Fig. 3.17.B.). Based on the positive associations among blue-green picophytoplankton, it is suggested that *Picochlorum*, *Picocystis* and *Bathycoccus*, together with *Synechococcus*, inhabited the picoalgal niche during 2008. One interesting feature of the suggested interactions involving these 4 phytoplankton is that there were more associations of these picophytoplankton with 16S TRFs than with 18S TRFs (33 vs 12) (Table 3.12.). Organisms represented by 16S TRFs are generally smaller than those represented by 18S TRFs. The more abundant associations between picophytoplankton and organisms represented by 16S TRFs may imply that there is a higher chance for microorganisms of similar sizes to interact, probably via similar cell surface/volume ratio (therefore similar competitive abilities for nutrients) and shared grazers.

In agreement with a previous study (Taylor et al., 2006), the inverse relationship between *A. anophagefferens* and NH_4^+ concentration during May and June (Fig. 3.22.) may suggest that *A. anophagefferens* was more competitive in taking up NH_4^+ than other phytoplankton, contributing to the initiation of the first brown tide peak in mid-June. Another competitive advantage *A. anophagefferens* is thought to have over other picophytoplankton is its heterotrophic capability, which was discovered long ago and is generally accepted now (Cosper et al. 1990, Berg et al. 1997, Gobler & Sanudo-Wilhelmy 2001b, Gobler et al. 2011). In this study, *A. anophagefferens*

was found to be positively associated with wind speed, both in NMS ordination analysis and in LSA. Therefore, wind may have 'helped' *A. anophagefferens* develop and maintain its dominance by creating a nutrient and light environment favorable for it as described below.

Unlike the effect of wind discussed earlier with regard to spatial heterogeneity in week 21, when the wind-driven alongshore current needed days to develop, the wind speed on the sampling day represents instantaneous mixing of the water column, which may not only have made the water turbid but also may have brought up nutrients from the sediment that may favor *A. anophagefferens* growth. It has been observed that *A. anophagefferens* can survive extended periods under low light conditions or even in total darkness (Popels & Hutchins 2002). A model proposed by MacIntyre et al. (2004) argued that in shallow water environments, depending on the ability of microphytobenthos to regulate the light and nutrient conditions in the overlying water, benthic-water column interactions tend to be defined by two stable states: one, a clear and nutrient-poor overlying water environment facilitated by well-developed microphytobenthos that restrict sediment resuspension and dissolved nutrient efflux; two, turbid and nutrient rich water associated with microphytobenthos erosion. In the MacIntyre et al. hypothesis, the transition from the first state to the second favors bloom development and is the result of wind mixing of the water column. *Minutocellus polymorphus* was also associated with wind speed, but with a negative sign (Fig. 3.17.A.). Under the second state, *Minutocellus* may have been outcompeted by *A. anophagefferens* because the former has a lower organic matter uptake rate under low light conditions (Cosper et al. 1990).

Wind was also found to be significantly correlated with the relative abundance of the TRF representing the rotifer *Brachionus calyciflorus* with a negative sign (Fig. 3.19.B.). This agrees with the suggestion that grazing of this species is significantly reduced in turbulent environments (Miquelis et al. 1998). To test whether the instantaneous water mixing effect of wind may have affected *A. anophagefferens* abundance, a Wave Exposure Model (WEMo) (Malhotra & Fonseca 2007) which takes into consideration the shoreline of GSB, bathymetry, wind direction and wind speed, was run for time blocks from half a day to a week. The computed parameter Representative Wave Energy (RWE) is a measure of wind energy accumulated over different time blocks. None of the RWE values calculated over the above time blocks had a

better correlation with relative abundance of *A. anophagefferens* than the original wind speed data.

My data suggest that selective grazing and parasitism may be important factors regulating the dominance of different phytoplankton at different times. *Chaetoceros* was negatively associated with all potential grazers/parasites except polychaetes (Fig. 3.24.), which implies that high grazing pressure may be the reason why *Chaetoceros* never formed a persistent bloom in GSB as *A. anophagefferens* did. In contrast, only one TRF representing a mixture of parasitic Syndiniales and copepods was negatively associated with *A. anophagefferens* (Fig. 3.15. and Fig. 3.25.A.). Polychaetes (probably a *Clymenura* species' pelagic larvae) and *Telonema* may have contributed to keeping the relative abundance of blue-green picophytoplankton low at times (Fig. 3.25.B.). In June, although potential grazers of the blue-green picophytoplankton decreased, *A. anophagefferens* and *Chaetoceros* established their dominance, and *A. anophagefferens* did not leave the picoalgal niche open to the blue-green picophytoplankton. It was not until July, when *A. anophagefferen*s started to decrease, that blue-green picophytoplankton started to increase when the relative abundance of polychaetes and *Telonema* was also low. Sieburth et al. observed elevated populations of the testate amoeba *Calycomonas ovalis* A.Wulff during a brown tide bloom when the *Synechococcus* population was depressed in Narragansett Bay, RI (Sieburth et al. 1988). The authors suggested that selective grazing of *Calycomonas* on *Synechococcus* rather than *A. anophagefferens* may be associated with the brown tide bloom. My data showed that polychaetes (most likely *Clymenura* larvae) and the pico-eukaryote *Picochlorum atomus* were negatively correlated (Fig. 3.17.B.), suggesting potential grazing effect of polycheates' planktonic larvae on picophytoplankton. Although polychaetes larvae were reported to be abundant zooplankton in Long Island waters (Castro & Cowen 1991, Lonsdale et al. 1996, Deonarine et al. 2006), little is known about their grazing behavior. Whether polychaete larvae can consume *Picochlorum atomus*, as was established in other studies that polychaetes removed picoplankton from seawater (Jordana et al. 2001, Licciano et al. 2005), or alternatively, the negative association between *Calycomonas* and *Picochlorum atomus* simply reflected an opposite response of the two organisms to a common factor, awaits further clarification.

In summary, the greater ability of *A. anophagefferens* relative to other phytoplankton to utilize NH_4^+ may have contributed to its dominance in May and June. Throughout the course of

the brown tide event in 2008, *A. anophagefferens*'s ability to grow well under low light conditions that may have been caused occasionally by wind mixing of the water column (and the algae's self-shading during bloom), coupled to its heterotrophic ability and being an unattractive food source for pelagic grazers, may all have contributed to dominance of *A. anophagefferens*. Temperatures above 25 ℃ in mid-July (Fig. 3.2.A.) may have kept *A. anophagefferens*' abundance in check in the middle of the summer; previous work has indicated that the growth rate of *A. anophagefferens* is highest at 20 ℃ (~0.8 doubling/day) and declines to ~0.6 doubling/day at 25 ℃, although higher temperature was not tested (Cosper et al. 1989). Grazing of *A. anophagefferens* by zooplankton from mid-July to mid-September also cannot be ruled out as a potential loss term for the brown tide bloom during that time. The second peak of *A. anophagefferens* relative abundance in fall may be a combined effect of lower water temperature and selective grazing on the blue-green picophytoplankton by polychaete larvae and *Telonema*.

Care should be taken in explaining significant associations and one caveat is that both positive and negative associations can indicate two involved TRFs responding to a third factor, in a similar or opposite way. It should be mentioned that the negative association between *A. anophagefferens* and copepods/Syndiniales could also have been caused if the zooplankton were not able to increase until *A. anophagefferens* declined, suggesting a potential toxic effect of the picoalgae towards the zooplankton. In addition, although the parasitic Syndiniales was found to be negatively associated with *Chaetoceros* (in fact the same observation was found in 2009 as well), the parasitic Syndiniales genus *Amoebophrya* was positively associated with non-*Chaetoceros* phytoplankton. These and other observations, such as the positive associations of *Pirsonia* and all the blue-green picophytoplankton, demonstrate that GSB may host many specific trophic interactions that are not well understood. Parasitism may be particularly important in GSB (likely to be true for other coastal environments as well) considering the numerous significant associations involving Syndiniales and *Pirsonia* found in this study. These observations highlight the importance of obtaining lab cultures of these parasitic flagellates from GSB and using them to perform lab experiments to better understand their ecological roles in the natural environment.

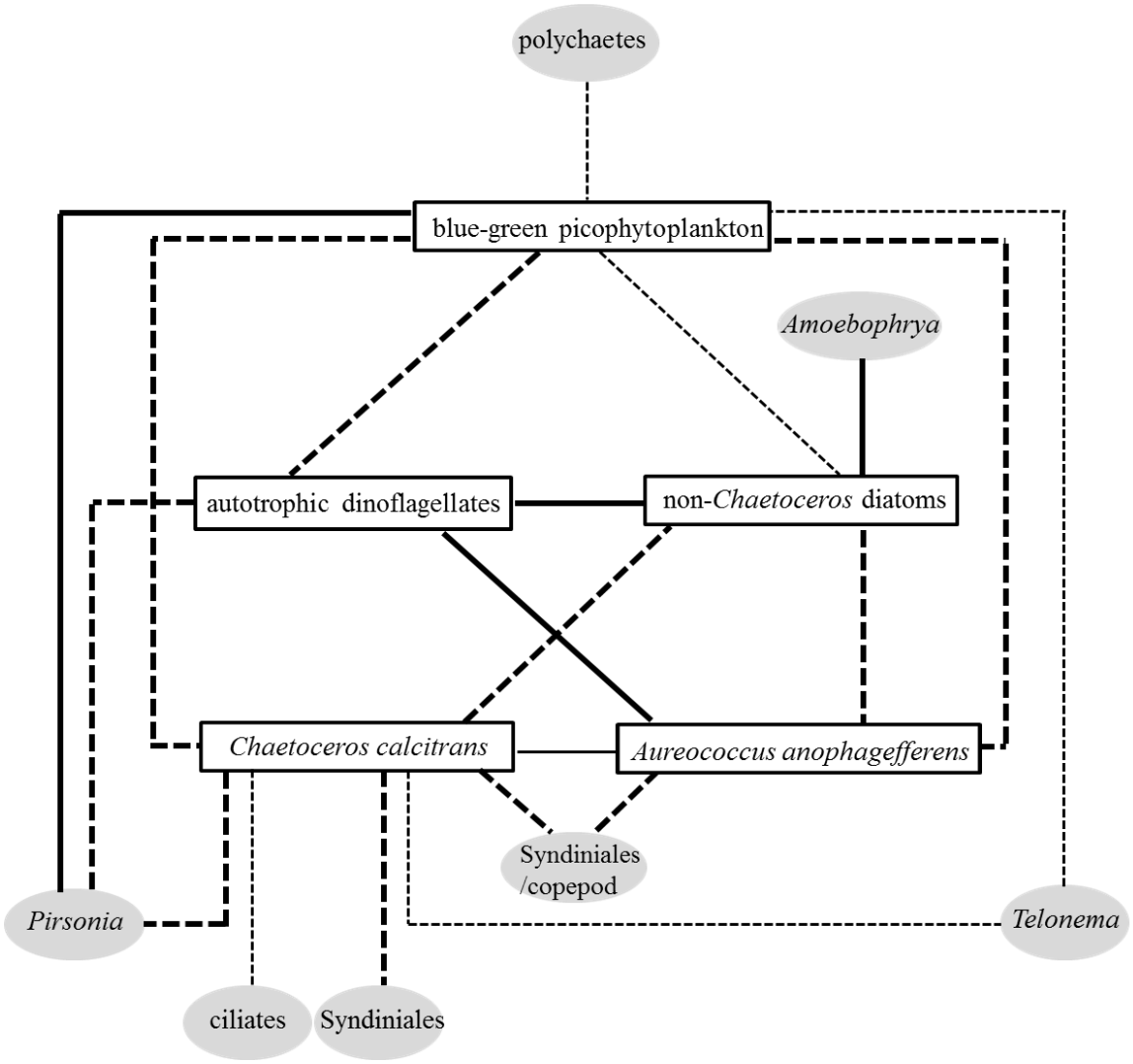
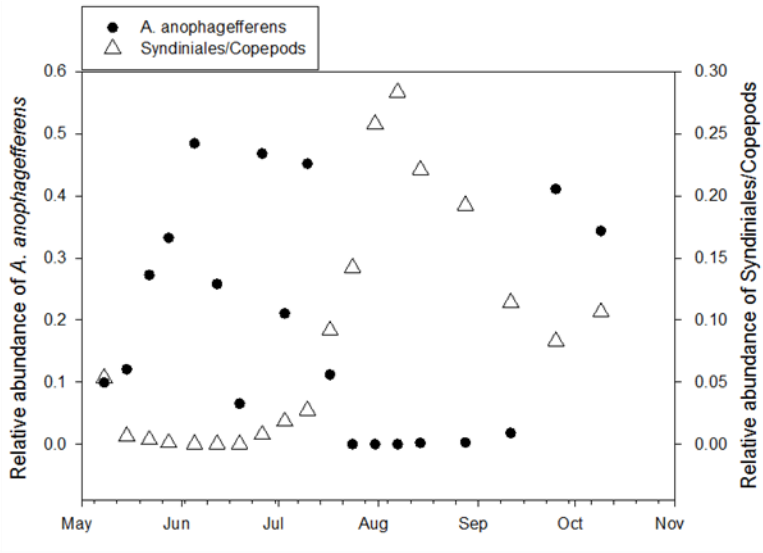
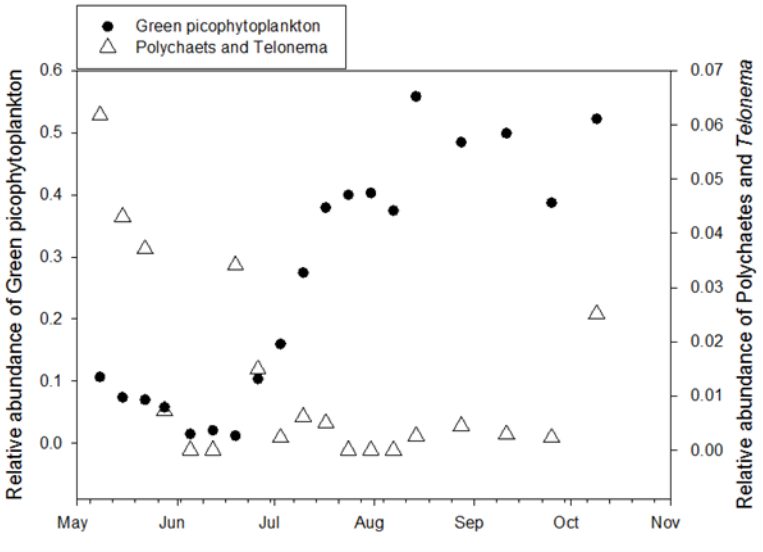


Fig. 3. 24. Associations among important ecological groups observed in this study. polychaetes: Poly; blue-green picophytoplankton: Syn/PS1, Syn/PS2, PS, *PA and Bathy; *Amoebophrya*: Amoe; autotrophic dinoflagellates: *KM and Pola; non-*Chaetoceros* diatoms: Cyclo, Cylin1, Cylin2, Minu, Skele and Strame. *Chaetoceros calcitrans*: Chaeto1, Chaeto2, ChaeP1 and ChaeP2; *Aureococcus anophagefferens*: AA1, AA2, AAP1 and AAP2; *Pirsonia*: Pirso; ciliates: Cili1; Syndiniales: Syndi; Syndiniales/copepod: Synd/Co; *Telonema*: Telo. See Table 3.10. for node abbreviations. Dashed and solid lines represent negative and positive correlations. Bold lines represent interactions observed between many TRFs comprising each group and fine lines represent interactions observed between only one pair of TRFs.



A



B

Fig. 3. 25. Temporal change of relative abundance of *A. anophagefferens*, blue-green picophytoplankton and their significantly associated grazers/parasites

Chapter 4 Planktonic microbial community structure in a non-brown tide year in Great South Bay, NY

Introduction

It is probably a fair statement that research into the dynamics of plankton communities in Great South Bay (GSB) has been influenced and stimulated by the fact that this area has been frequently perturbed by brown tide blooms in the past 3 decades. Ironically, there is little information about what the plankton community looked like before the first occurrence of a brown tide bloom in 1985. This makes it difficult to carry out, for example, ecosystem restoration work because it is largely unknown what status we want to bring the ecosystem 'back' to. Additionally, testing hypotheses about factors supporting bloom development would be best based on understanding the community under non-bloom conditions, but this 'background' information on both phytoplankton and zooplankton is not always available. Sieracki et al. (1999, 2001) hypothesized that *Synechococcus* dominates a 'picoalgal niche' in most years, but when they are selectively removed, the picoalgal niche becomes available for algae of similar size, such as the brown tide species *Aureococcus anophagefferens*. It is, however, not known if *Synechococcus* is the sole species occupying the picoalgal niche in a non-brown tide year. If there are other picoalgae, how they interact with *Synechococcus* is also unknown. For GSB, this information is of particular importance since the planktonic community is known to have been dominated by 'small forms', green picoalgae like *Nannochloris spp.* (synonym *Picochlorum*) and *Stichococcus spp.* (Ryther 1954, Lively et al. 1983), in the past. Without a better understanding of all the species inhabiting the picoalgal niche, and their interactions, the mechanisms through which *A. anophagefferens* establishes its dominance in the GSB plankton community will be difficult to determine.

The species composition of planktonic grazers in GSB is also not well understood, and therefore it is difficult to evaluate what roles grazers may play during the initiation and progression of brown tide blooms. For example, the lack of knowledge on the zooplankton species composition hinders our ability to explain conflicting observations made on the interaction between pelagic grazers and *A. anophagefferens*. While early lab experiments found that brown tide did not inhibit ciliates grazing on other microalgae (Caron et al. 1989), a lower grazing rate of total microzooplankton in the presence of *A. anophagefferens* was observed using dilution incubations (Gobler et al. 2002)of natural communities. Lab studies are constrained to testing species available in culture that are not necessarily relevant to the study sites, while field studies often only report grazing rates of the whole zooplankton community without identifying

specific groups feeding on *A. anophagefferens*. A more precise characterization of planktonic grazers in non-brown tide years is needed to compare with that in brown tide years to determine if the grazer community composition differs.

Parasites are ubiquitously distributed, strong biological control agents of plankton community structure in both marine (Park et al. 2004, Kagami et al. 2011)and fresh water environments (Ibelings et al. 2004, Kagami et al. 2007, Guillou et al. 2008, Miki et al. 2011, Rasconi et al. 2012). Parasites have the potential to directly regulate host abundance and to indirectly mediate the interspecific interactions involving host species (Miki et al. 2011). The importance of parasites is not sufficiently appreciated in many food web studies because parasite biomass is usually assumed to be small. A recent study in 3 estuaries on the Pacific coast of California and Baja California shows that parasites have substantial biomass in these ecosystems, which even exceeded that of top predators (Kuris et al. 2008). The marine parasitic dinoflagellate order Syndinalies was found to be highly diverse and abundant in various marine environments in Atlantic and Mediterranean waters (Guillou et al. 2008). All described Syndiniales are parasitoids, and obligatorily kill their host (Guillou et al. 2008). One genus of Syndiniales, *Amoebophrya*, has been found to infect and kill bloom-forming dinoflagellates in Chesapeake Bay and California coast waters (Coats & Park 2002, Mazzillo et al. 2011). Clearly, the importance of parasites/parasitoids in shaping plankton community structure is underestimated. As a matter of fact, no parasites/parasitoids have been reported for GSB to our knowledge.

There was no brown tide bloom in GSB in 2009, providing the opportunity to investigate the biodiversity and temporal and spatial patterns of the planktonic microbial community, and environmental factors influencing these patterns, in the absence of *A. anophagefferens* blooms. This information will serve as the baseline to which the brown tide year is compared. The same spatial/temporal sampling resolution and approximate synchronization of the sampling dates with those in 2008 were followed, but sampling continued into the winter. The specific the questions I will address in this chapter include:

1. What are the spatial/temporal patterns of the planktonic microbial community in a non-brown tide year? What environmental factors may be involved in shaping microbial community structure?

2. In the absence of the brown tide bloom, is there spatial heterogeneity of the microbial community across GSB? What environmental factors may be involved in generating the spatial heterogeneity?

3. Does *Synechococcus* dominate the picoalgal niche in the absence of the brown tide bloom? Are there co-occurring picoalgae? If so, are they positively or negatively correlated with each other?

Materials and Methods

Sample collection and processing

Surface water samples were collected in GSB off the south shore of Long Island, New York in 2009 at the same sampling sites as in year 2008 (Fig. 3.1.). Weekly samples were collected from May 8 to December 5 with no samples collected in weeks 19, 21, and 24 (Table 4.1.). In the rest of this paper, each sample will be identified by the combination of a number representing the week of collection and a letter representing the station of collection (e.g. the sample collected at station E in week 10 is identified as 10E).

Water sample collection and processing were done the same way as in 2008 (Chapter 3). In addition, for samples from stations A and E (when samples were taken from 5 stations) and stations A, E, F, and J (when samples were taken from 10 stations), 1980 µl seawater were fixed with 20 µl glutaraldehyde (25%, Sigma-Aldrich, St. Louis, MO) in 2 ml cryo tubes. After incubation in the dark at room temperature for 30 min, the cryo tubes were transferred to deep freezer (-80 ℃) until analyzed. Additionally, 20 to 30 ml of seawater from stations A, E, F, and J were filtered through 25 mm GF/F filters (Whatman, Piscataway, NJ) in duplicate to measure chlorophyll a concentrations (Chla). GF/F filters were extracted in 7 ml pure acetone at -20 ℃ for 12 to 16 hrs and glass vials containing the filters (covered by aluminum foil throughout) were brought to room temperature before fluorescence measurement to avoid condensation on the vials. Fluorescence readings before (FO) and after addition of 10% HCl (FA) were taken on a Turner Designs (model 10-AU) fluorometer calibrated to known Chla. Equation $\frac{(FO-FA)*7}{Filtered\ vol.}$ was used to calculate Chla in µg/L. Other environmental data was collected the same way as in 2008 (Chapter 3).

Table 4. 1. Sampling date, the corresponding week number, stations sampled (A closer to the mainland and J closer to Fire Island) and samples analyzed by TRFLP in GSB, 2009

	Week number	Samples collected	Samples analyzed (18S)	Samples analyzed (16S)
8-May	1	A to E	A, E	A, C, D, E
13-May	2	A to E	A to E	A,C
20-May	3	A to J	A to J	A to J
27-May	4	A to E	A, E	A, E
4-Jun	5	A to J	E, F, J	A,E,F, J
10-Jun	6	A to E	A, E	A, E
17-Jun	7	A to J	A, E, F, J	A, E, F, J
24-Jun	8	A to E	A, E	A, E
1-Jul	9	A to J	A, E, F, J	A, E, F, J
8-Jul	10	A to E	A, E	A
15-Jul	11	A to J	A, E, F, J	A, E, F, J
22-Jul	12	A to E	A, E	A, E
29-Jul	13	A to J	A, E, F, J	A, E, F, J
5-Aug	14	A to J	A, E, F, J	A, E, F, J
11-Aug	15	A to J	A, E, F, J	A, E, F
20-Aug	16	A to E	A, E	A, E
26-Aug	17	A to J	A, E, F , J	E, F, J
4-Sep	18	A to E	A, E	A, E
18-Sep	20	A to J	A, E, F, J	A, E, F
2-Oct	22	A to J	A, E, F, J	A, E, F, J
9-Oct	23	A to E	A, E	A, E
23-Oct	25	A to E	A, E	E
31-Oct	26	A to E	A, E	A, E
7-Nov	27	A to J	A, E, F, J	A, F, J
15-Nov	28	A to J	A, E, F, J	A, E, F, J
21-Nov	29	A to J	A, E, F, J	A, E, F, J
29-Nov	30	A to J	A, E, F, J	A, E, F, J
6-Dec	31	A to J	A to J	A to J

TRFLP, cloning and sequencing

DNA extraction, PCR, restriction digestion and data processing were done following the protocol described in Chapter 2 and kept consistent with year 2008. Briefly, phenol/chloroform was used to extract the genomic DNA. Universal primers labeled with fluorescent tags were used to amplify partial 18S rDNA and 16S rDNA (see Table 3.2. in Chapter 3 for the primer sets and fluorescent tags used). PCR products were cleaned up and treated with Klenow Fragment Exonuclease minus (Egert & Friedrich 2005). After another cleanup, restriction enzyme *Taq*I was used to digest the PCR products. The fragmented PCR product was subsequently precipitated using ethanol, resuspended in Milli-Q water and loaded on a 3130XL ABI Analyzer. Raw data from the electropherograms was exported into an Excel spreadsheet in Peak Scanner™ Software v1.0, which was subsequently fed into online software TREX (http://trex.biohpc.org/) for noise filtration and TRF alignment. Cloning and sequencing of both 18S and 16S rDNA PCR amplicons were performed to identify TRFs as in Chapter 3. Because of the constantly changing nature of GenBank, genus name was usually used to report a TRF identification even when the best hit in GenBank indicated a species.

Enumeration of Synechococcus using flow cytometry (FCM)

Frozen samples fixed with glutaraldehyde were thawed on ice, and filtered through 50 μm mesh. 1 ml filtrate was mixed with 10 μl of green fluorescent beads (~10^8/ml, Sigma-Aldrich, St. Louis, MO) and run on a BD FACSCalibur flow cytometer (BD, San Jose, CA) at low speed (12 μl/min). Data collection ended when events number reached 100,000 or collection time reached 5 min. Event rate was ~2000 to 3000 events/sec. FSC, SSC, FL1, FL2, FL3 and FL4 PMT voltages were set at 500, 375, 725, 650, 500 and 550, respectively. FSC and SSC were used as thresholds, set at 50 and 50. Enumeration of picoeukaryotes, phycoerythrin-containing *Synechococcus* (SYN-PE) and phycocyanin-containing *Synechococcus* (SYN-PC) was performed with FlowJo (V4.4.1, Tree Star, Ashland, OR). Total *Synechococcus* abundance was calculated as the division of the summed SYN-PE and SYN-PC events over the total volume analyzed. *Synechococcus* abundance was then correlated with the area contribution from the TRF representing *Synechococcus*. *Synechococcus* abundance from one sample in each of 28 sampling weeks (when c475Y% was available for correlation analysis) was enumerated on the flow cytometer.

Data analysis

1 Temporal variation

Similar to 2008, TRFs contributing more than 5% to total peak area averaged over all samples were considered major TRFs. Observed TRFs with similar fragment lengths (in most cases, ±2 bases) were binned into major TRFs as explained in Chapter 3. Sizes of the observed TRFs for the same organisms in 2009 vs 2008 differed by 1 to 3 bases due to instrumental variation. For example, major TRF c284B was composed of the observed TRFs 284B and 285B in 2008 while the observed TRFs 282B and 284B represented the same organism, *Chaetoceros*, in 2009. To avoid confusion when 2008 data and 2009 data are compared in Chapter 5, both 18S and 16S major TRFs that were the same in both years are called by the names used in 2008.

2 Nonmetric multidimensional scaling (NMS) ordination

102 and 96 samples were analyzed (Table 4.1) by 18S and 16S TRFLP, respectively. The data matrix generated by TREX contained the contribution of each TRF to total peak area for each sample and was used as the main matrix for NMS ordination in PC-ORD. The second matrix included sample ID and environmental parameters, which were transformed and relativized following the method described in Chapter 3. 18S samples from week 31 (December 6) suffered from misalignment of a few abundant TRFs and were therefore excluded from NMS analysis, leaving 92 samples.

NMS was run in autopilot mode and the solutions were evaluated by Monte Carlo test (McCune & Grace 2002). Visualization of the ordination was through biplots whose X and Y axes were extracted from the main matrix and did not necessarily represent gradients of any parameter. The coefficient of determination (R^2) was calculated to assess the percentage of variance in the original data matrix represented by each axis. The correlation coefficient (r) between the position of each sample on an ordination axis and each environmental parameter was calculated to reveal any linear relationship between the ordination axis and the environmental parameters.

3 Spatial variation

TRFLP profiles were obtained from samples collected from stations A and E in weeks when 5 stations were sampled and from samples collected from stations A, E, F, J in weeks when

10 stations were sampled. The main matrix for multivariate analysis in PC-ORD was used to compute the Bray-Curtis distance between every possible pair of samples analyzed from each week (Table 4.1). Geographical distance between these pairs of stations was also calculated. Regression analysis between Bray-Curtis distance and geographical distance was used to determine if the two were related. Because of the misalignment problem with samples collected in week 31, they were not included in spatial variation analysis.

4 Local similarity analysis (LSA)

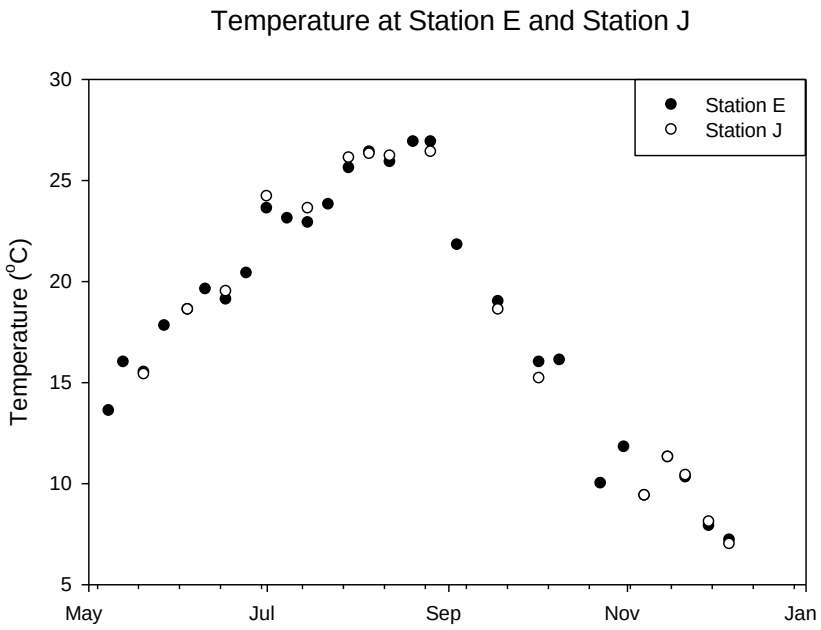
Associations among TRFs (TRF-TRF) and between TRFs and environmental factors (TRF-EF) were examined by analyzing the local similarity correlation (Ruan et al. 2006). After minimizing effects of rare TRFs as described in Chapter 3, the local similarity matrix (LSM) contained 93 samples with 57 18S TRFs, 33 16S TRFs and 7 environmental parameters (temperature, salinity, Connetquot River discharge, sum of river discharge for the prior 7 days, wind speed, maximum gust speed, and Chla). To investigate time-delayed associations between TRF-TRF as well as TRF-EF pairs, Local Similarity Matrix-Temporal (LSM-T) was generated the same way as for year 2008 (Chapter 3). LS scores for TRF-TRF and TRF-EF pairs with and without time delay were computed in R as in Chapter 3. p and q values were also computed to determine if an LS score was statistically significant. All LS scores among 90 TRFs and 7 environmental factors were visualized as networks using Cytoscape 2.8.2 (Shannon et al. 2003). Each TRF or environmental factor is represented by a node in the network, connected by a line (LS correlation) with another TRF or environmental factor that has a similar pattern (or opposite pattern if the correlation is negative). Data filtering was performed in Cytoscape to include only the statistically significant correlations, defined as those with LS score larger than 0.3 and p smaller than 0.005. With 4656 possible TRF-TRF and TRF-EF pairs, multiple testing becomes a concern and so q values were calculated. Only LS scores with $q<0.005$ were considered significant. The original as well as the time-delayed associating pairs were plotted and Pearson Correlation coefficients were calculated to decide if the time-delayed association was reliable.

Results

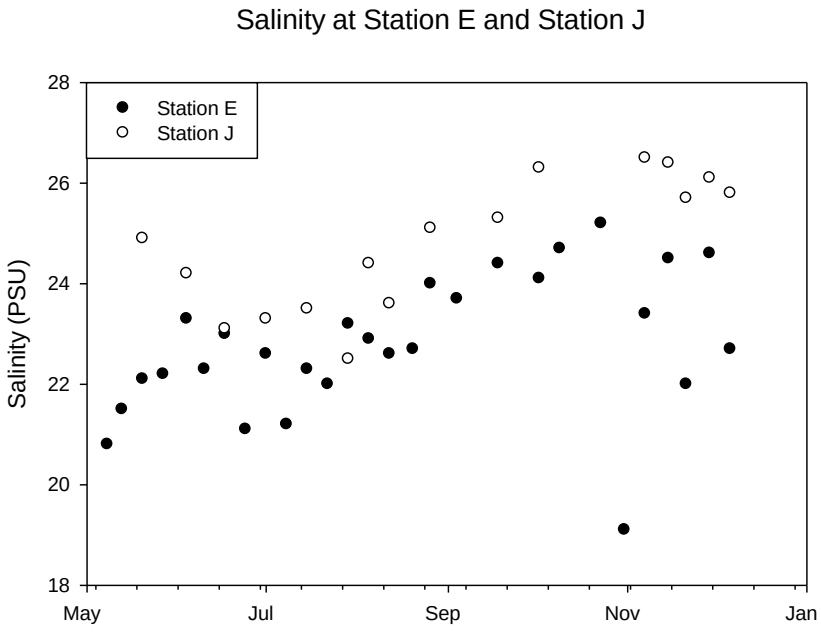
Environmental parameters

Surface temperature increased from 14 ℃ in May to 27 ℃ in August (Fig. 4.1. A). The lowest salinity, 19.1 PSU, was measured at station E on October 31 and the highest salinity, 26.5 PSU, was recorded on November 6 at station J (Fig. 4.1. B.). Unlike 2008 (Chapter 3, Table 3.3.), the correlation between salinity and temperature was negative and not statistically significant in 2009 (Table 4.2.). For weeks when samples were collected from stations A to J, salinity was always higher at J than E except for week 13 (July 29). Precipitation, particularly 7-day precipitation (Fig 4.1.C), was positively correlated with Connetquot River discharge (Table 4.2). 7-day Connetquot River discharge and river discharge on sampling day (Fig 4.1.D) were both significantly correlated with salinity (Table 4.2). The average wind speed and the maximum gust speed (Fig. 4.1. E.) were significantly correlated. The highest wind speed of ~13 mph occurred in week 26 (October 31). Among all environmental factors, Chla was only significantly correlated with temperature and its peak value of 36.18 μg/L was measured at station A in week 15 (August 11).

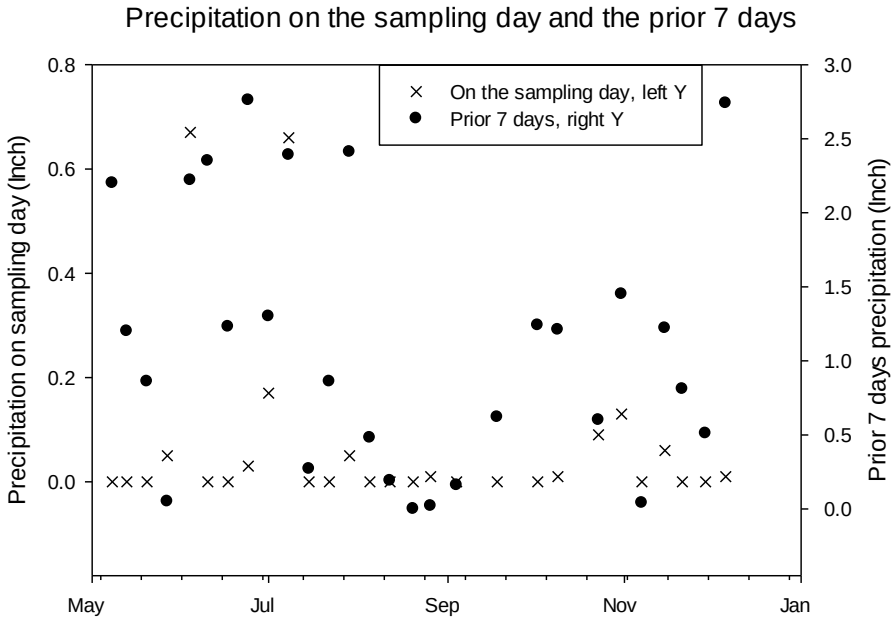
A)



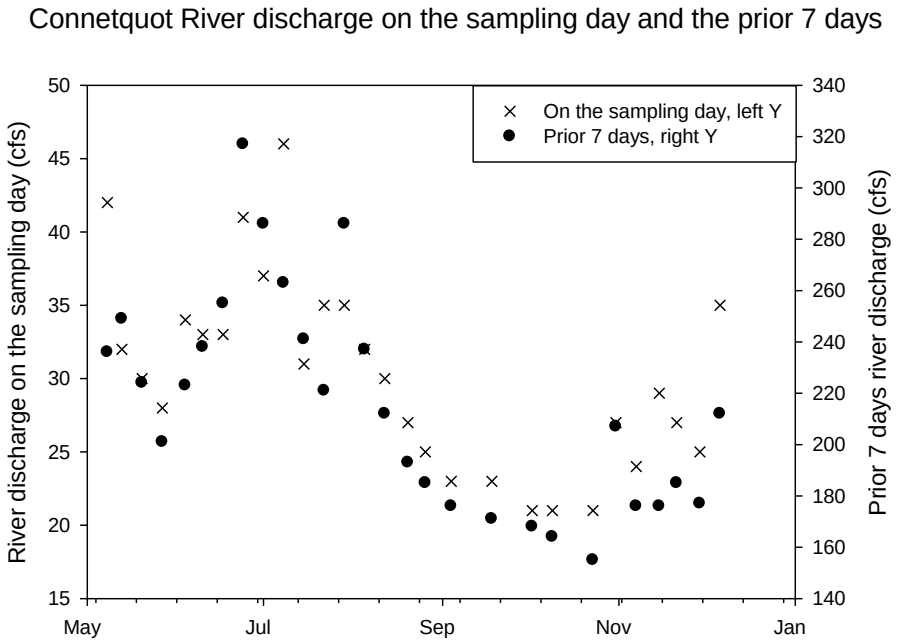
B)


Salinity at Station E and Station J
Station E
Station J
Salinity (PSU)
28
26
24
22
20
18
May
Jul
Sep
Nov
Jan

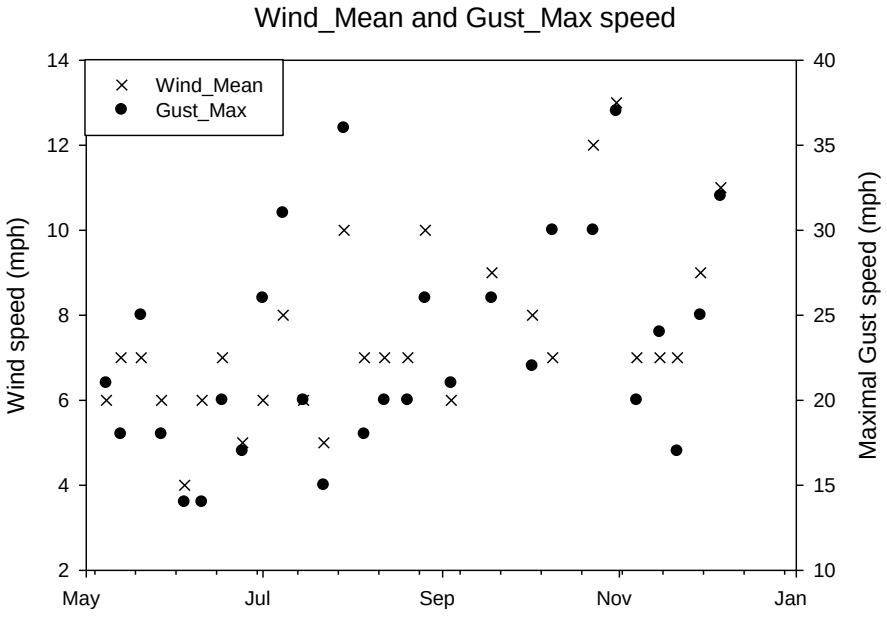
C)


Precipitation on the sampling day and the prior 7 days
On the sampling day, left Y
Prior 7 days, right Y
Precipitation on sampling day (Inch)
Prior 7 days precipitation (Inch)
0.8
0.6
0.4
0.2
0.0
3.0
2.5
2.0
1.5
1.0
0.5
0.0
May
Jul
Sep
Nov
Jan

D)


Connetquot River discharge on the sampling day and the prior 7 days
River discharge on the sampling day (cfs)
Prior 7 days river discharge (cfs)
On the sampling day, left Y
Prior 7 days, right Y
50
45
40
35
30
25
20
15
340
320
300
280
260
240
220
200
180
160
140
May
Jul
Sep
Nov
Jan

E)


Wind_Mean and Gust_Max speed
Wind speed (mph)
Maximal Gust speed (mph)
Wind_Mean
Gust_Max
14
12
10
8
6
4
2
40
35
30
25
20
15
10
May
Jul
Sep
Nov
Jan

F)

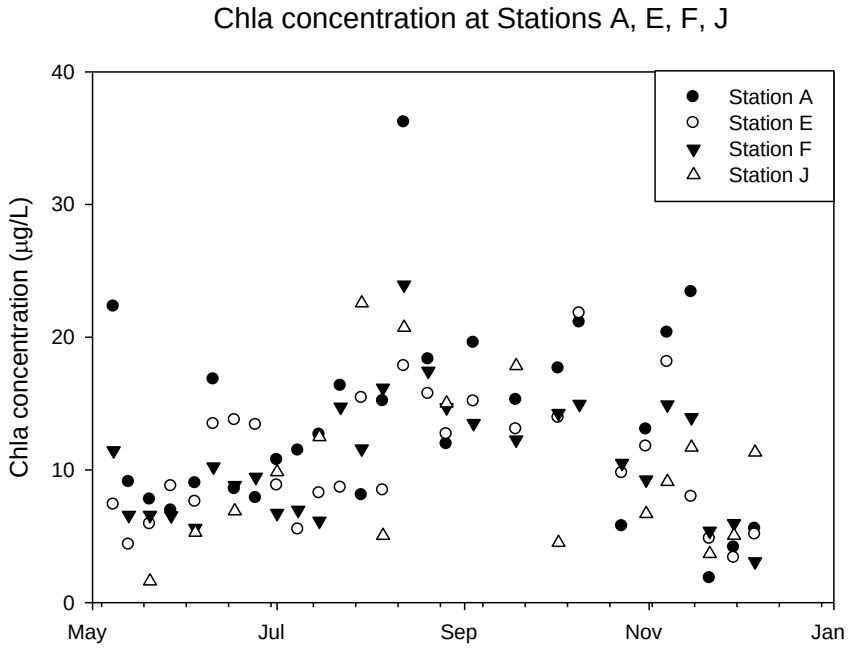


Fig. 4. 1. Environmental parameters versus time in GSB, 2009 A) Temperature at station E and J B) Salinity at Station E and J C) Precipitation on the sampling day D) The sum of water discharge from the Connetquot River over the prior 7 days including the sampling day E) Average wind speed and maximum gust speed on the sampling day F) Chla concentration at Stations A, E, F, J

Table 4. 2. Correlation coefficients between environmental parameters; numbers in bold represent significant correlations ($p<0.001$*)

	Temp	Sal	River	7-days river	PPT	7-days PPT	Wind _mean	Gust_ max	Chla
Temp	1								
Sal	-0.283	1							
River	0.163	**-0.487**	1						
7-days river	**0.435**	**-0.548**	**0.818**	1					
PPT	0.132	-0.157	**0.382**	0.209	1				
7-days PPT	-0.297	-0.190	**0.658**	**0.445**	0.322	1			
Wind_mean	**-0.375**	0.171	-0.171	-0.268	-0.201	0.248	1		
Gust_Max	-0.241	0.092	0.058	-0.011	-0.008	**0.392**	**0.818**	1	
Chla	**0.399**	-0.047	-0.196	-0.140	-0.137	-0.233	-0.085	-0.044	1

Temp: Surface water temperature; Sal: Salinity; River: Water discharge from the Connetquot River on the sampling day; 7-days river: The sum of water discharge from the Connetquot River over the prior 7 days including the sampling day; PPT: Precipitation on the sampling day; 7-days PPT: The sum of precipitation over the prior 7 days including the sampling day; Wind-Mean: Average wind speed on the sampling day; Gust_Max: Maximum gust speed on the sampling day; Chla: chlorophyll a concentration *: a stringent p value is used here to focus on the strongest associations. Temperature and salinity were the average of those at station E and J when the long transect was sampled and were those at station E when the short transect was sampled.

Temporal change of major TRFs

1 18S community

There were 6 major 18S TRFs in 2009. Four of them (c284B, c316B, c374B and c532B) were also major TRFs in 2008 (Fig. 4.2.), and c175B representing *A. anophagefferens* was the only 2008 major TRF that was not a major TRF in 2009. 351 partial 18S rDNA sequences of good quality (chimeric sequences were identified the same way as in 2008 and excluded from further analysis) were retrieved from clone libraries constructed from 10 samples and their predicted *Taq*I restriction fragments were used to identify the major 18S TRFs (Table A3.). 12 sequences with the same predicted TRF length as *A. anophagefferens* were captured in 2009, all from the copepod *Harpacticus* sp. (Table A3). Although *A. anophagefferens* may be present in GSB in 2009, its abundance must be much lower than in 2008 since the average contribution of the observed TRFs (171B and 173B) that may include *A. anophagefferens* was only 0.87% of total TRF peak area throughout the sampling season in 2009.

c258B and c264B were major TRFs only in 2009. c258B was a combination of 3 observed TRFs (256B, 258B and 260B), with no significant correlation between any pair, that matched the TRFs predicted for ciliates (e.g. *Strombidium*) and phytoplankton including chlorophytes (Mamiellales), haptophytes (*Chrysochromulina*, *Isochrysis*, *Diacronema*) and dinoflagellates (*Gymnodinium*) (Table A3.). c258B was detected throughout the 2009 sampling season with its peak value of 15.0% occurring on July 15 (Fig. 4.2.). Major TRF c264B was a combination of 4 observed TRFs (261B, 263B, 265B and 267B), with 265B and 267B significantly correlated (r=0.42), that corresponded to predicted TRFs (263 to 270 bases in length) from cloned amplicons of Cercozoa (21 out of the 30 sequences were identified as Cercozoa and 7 were identified only as uncultured eukaryotes (Table A3.)). Similar to c258B, c264B was detected in GSB throughout 2009 although its relative abundance was highest, around 15 to 20%, in September.

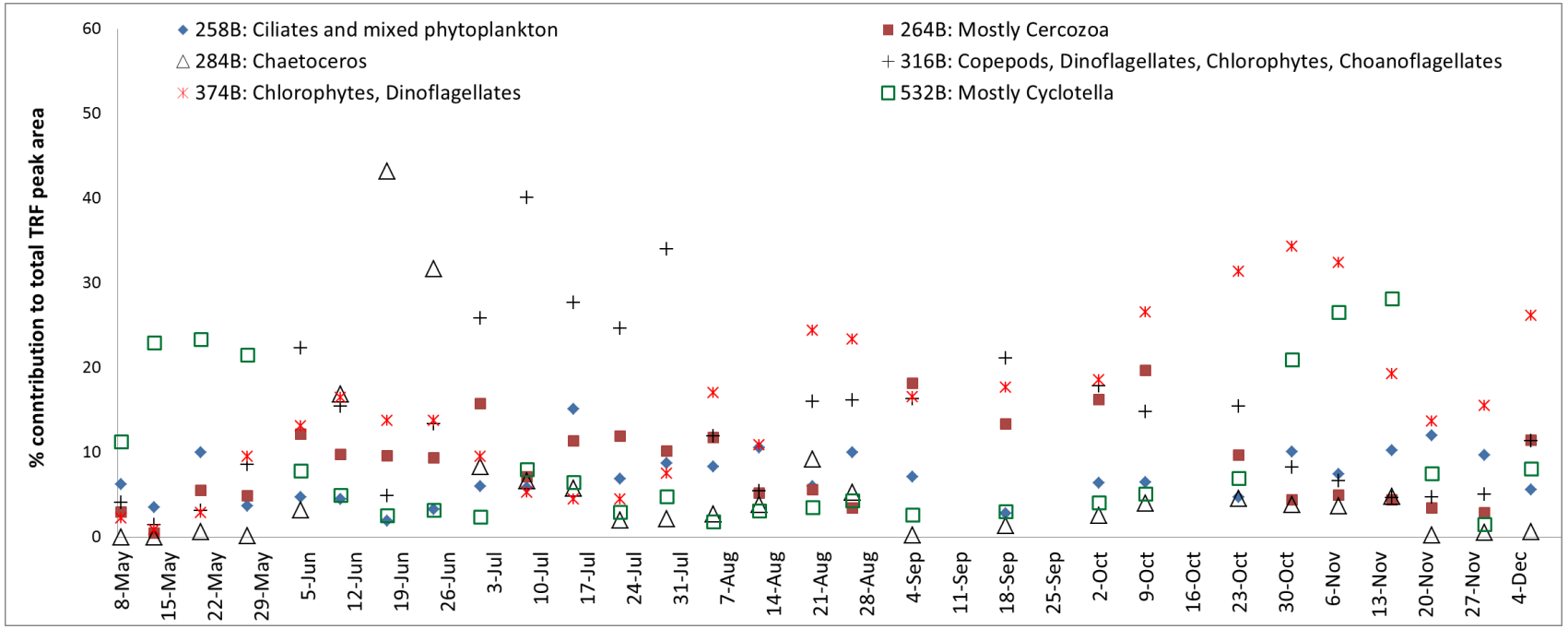
Major TRF c284B (composed of observed TRFs 282B and 284B that are significantly correlated, r=0.74) matched the TRF predicted for *Chaetoceros calcitrans*. Very similar to 2008, c284B% had its peak value in middle June (June 17, week 7), accounting for ~42% of total TRF peak area. Except for weeks 6, 7, and 8, when c284B% was above 15%, c284B contributed less than 10% to total peak area and was at times undetected.

Major TRF c316B was composed of 3 observed TRFs (313B, 315B and 317B), with no significant correlation between any pair, and matched the TRFs predicted from several taxa including copepods (*Oithona*) (55 of 88 sequences with predicted TRF length from 317 to 322 bases were from *Oithona*), dinoflagellates (*Amoebophrya, Symbiodium, Pentapharsodinium* and *Woloszynskia*), chlorophytes (*Pyramimonas*, *Picocystis*, and *Picochlorum*), and choanoflagellates. Major TRF c316B was a dominant TRF for the whole month of July, with its relative abundance ranging from 25% (July 1 and July 22) to 40% (July 8).

Major TRF c374B was composed of observed TRFs 371B, 373B and 375B, with no significant correlation between any pair, and matched the TRFs predicted from several taxa including chlorophytes (9 out of 41 sequences with predicted TRF sizes from 375B to 379B were from *Picochlorum*) and dinoflagellates (14 out of 41 sequences with predicted TRF size from 375B to 379B were from *Gymnodinium, Polykrikos, Scrippsiella,* and *Karlodinium*). c374B% reached a small peak around 16.5% on June 10, a moderate peak at the end of August (August 26) around 24.0% and its highest value on October 31 at 34.3%.

Major TRF c532B (composed of observed TRFs 529B and 532B) matched predicted TRFs of diatoms *Minutocellus polymorphus* and *Cyclotella choctawhatcheeana*, with the latter more frequently sequenced (4 vs 1 sequences). There was a clear temporal pattern for c532B, with its peak values in the colder seasons in mid-late May and early-mid November at ~25%. For the rest of the sampling season, c532B was present at 2% to 10% of total peak area.

As is shown in Fig. 4.2., the sum of the 6 major 18S TRFs' area contribution ranged from 27% on May 8 (week 1) to 82% on October 31 (week 26) and November 7 (week 27). Weeks 26 and 27 were when c374B% and c532B% together accounted for more than half of the total TRF peak area.



week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Sum of Major TRF area%	27	29	45	48	63	68	76	75	68	73	71	53	67	54	39	65	62	61	ND	59	ND	66	77	ND	73	82	82	71	41	35	63

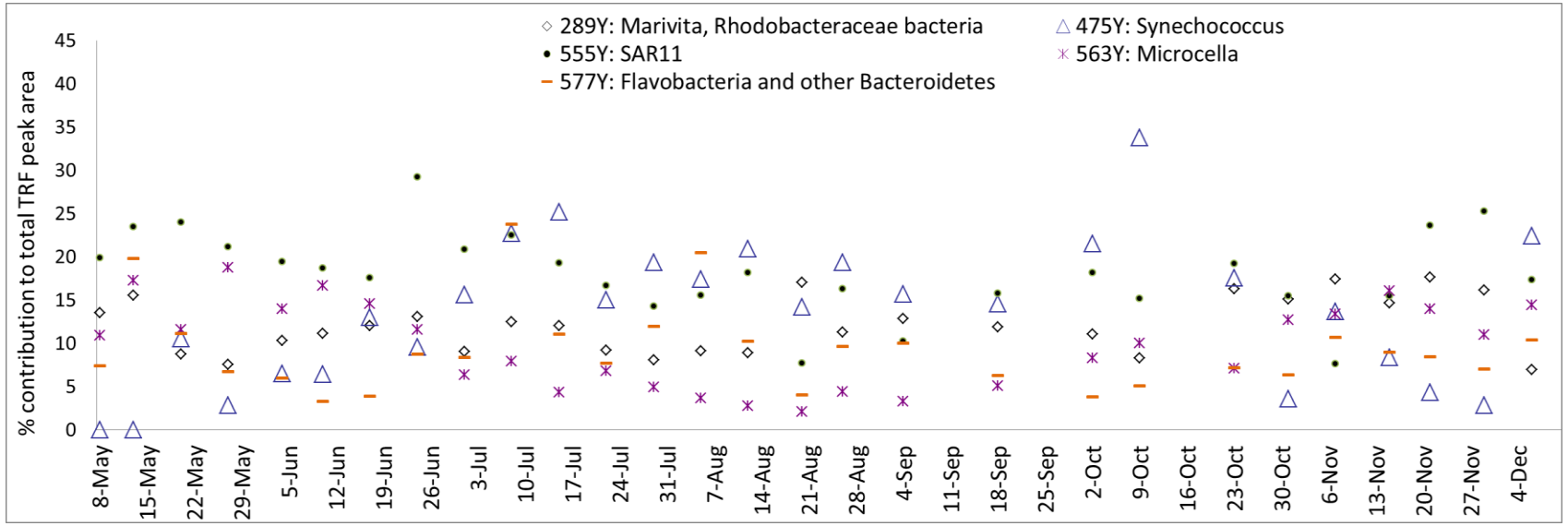
Fig. 4. 2. Contribution of major 18S TRFs to total peak area averaged among stations in 2009 (average CV of the contributions of all major TRFs at all sampling times is 0.34); Some TRFs with similar size (usually ±2 bases) were combined to form one major TRF, as explained in the text. Numbers in the first row below the figure represent the number of the week when samples were collected and the numbers in the second row are the summed area percentage of the major 18S TRFs in that week; ND: no data (samples not collected these weeks)

2 16S community

There were 5 major 16S TRFs in 2009 (Fig. 4.3.) and all of them were also major TRFs in 2008. 265 partial 16S rDNA sequences of good quality (chimeric sequences were identified the same way as in 2008 and excluded from further analysis) were retrieved from 6 clone libraries and their predicted *Taq*I restriction fragments were used to identify the major 16S TRFs (Table A4.). No plastid sequence of *A. anophagefferens* was captured, but 1 sequence representing Bacteroidetes generated a predicted TRF similar to that for *A. anophagefferens* plastid (300Y, Table A2.) that matched observed TRFs 297Y and 298Y. The average (297Y+298Y)% was only 0.22% throughout the season in 2009.

Major TRF 289Y matched the TRF predicted for *Marivita* and other Rhodobacteraceae, and 289Y% varied between 7% and 17% throughout the sampling period (Fig. 4.3.). Major TRF 475Y matched TRFs predicted from the cloned amplicons of *Synechococcus*. 475Y% was greatest from July (15.6% on July 1) to the end of October (17.6% on October 23), averaging 20.2% during this period with a peak value of 33.8% on October 9. TRF 555Y matched the TRF predicted from cloned SAR11 (*Pelagibacter*) amplicons. It varied in its area contribution from ~8% on August 20 and November 7 to ~ 30% on June 24. Major TRF 563Y matched the TRF predicted from cloned *Microcella* amplicons and there was a strong seasonal pattern of 563Y%: peak values of 563Y% occurred on May 27 and November 15 at 18.8% and 16.1%, and the smallest 563Y% was on August 20 at 2.1%. Major TRF 577Y matched the TRF predicted for *Flavobacteria* and other Bacteroidetes such as *Cytophaga* (Table A4.). There were 3 maximal values of 577Y% on May 13, July 8 and August 5 around 20%, and 577Y% ranged from 3% to 12% at the other sampling times.

The sum of the 5 major 16S TRFs' area contribution ranged from 45% on August 20 (week 16) to 89% on July 8 (week 10), which was smaller than the range for the 18S community. On July 8, when sum of the area contribution from the major 16S TRFs was the biggest, 475Y representing *Synechococcus*, 555Y representing SAR11 and 577Y representing Bacteroidetes each contributed a little over 22% to the total TRF peak area.



week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Sum of Major TRF area%	55	76	66	57	56	56	61	72	60	89	72	55	58	66	61	45	61	52	ND	53	ND	63	72	ND	67	53	63	63	68	62	72

Fig. 4. 3. Contribution of major 16S TRFs to total peak area averaged among stations in 2009 (average CV of the contributions of all major TRFs at all sampling times is 0.26); Some TRFs with similar size (usually ±2 bases) were combined to form one major TRF, as explained in the text. Numbers in the first row below the figure represent the number of the week when samples were collected and the numbers in the second row are the summed area percentage of all major 16S TRFs in that week; ND: no data (samples not collected these weeks)

Correlation of flow cytometrically determined Synechococcus abundance and c475Y%

c475Y% varied from 0 (station A on May 8 and May 13) to ~30% (station A, October 9). Changes in *Synechococcus* cell numbers determined cytometrically demonstrated a similar pattern as change in c475Y%, with a significant correlation between *Synechococcus* cell abundance and c475Y% (Fig. 4.4., r = 0.7664), suggesting agreement between absolute and relative abundance of *Synechococcus*.

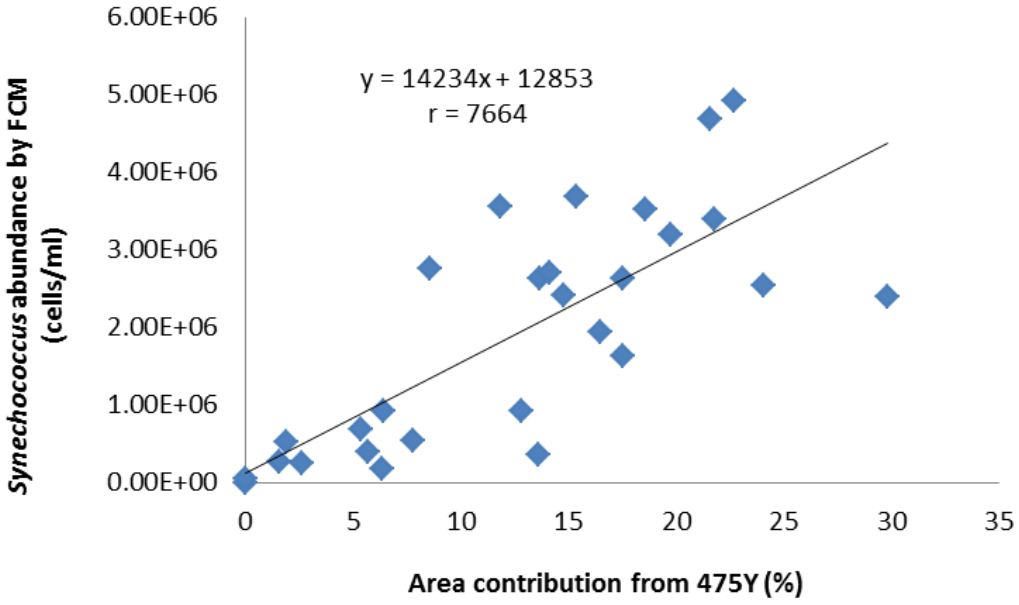


Fig. 4. 4. Correlation between *Synechococcus* cell abundance by FCM and c475Y% by TRFLP

Temporal and spatial variation of GSB plankton community

1 18S community

Three axes in NMS ordination represented 82% of the variation seen among 18S samples in 2009, with axes 1 and 2 accounting for slightly more variation than axis 3 (Table 4.3.). Temperature was associated with both axes 2 and 3, Connetquot River discharge was significantly associated with axis 3, and Chla was significantly associated with axis 2 (Table 4.4.). Although axis 1 accounted for almost 30% of the variation seen among samples, there was no environmental parameter included in the ordination analysis that was strongly related to this axis. In general, samples collected on the same day clustered together and temporal variation was greater than spatial variation (Fig. 4.5.). There was a dramatic increase in the relative abundance of major TRF c284B on June 17 (week 7, Fig. 4.2.). Sample points from week 7 were separated from the other samples most clearly on ordination axis 1, which was not only significantly associated with 282B% (r= -0.577) and 284B% (r= -0.490) contributing to c284B%, but also many others such as 371B% (r= -0.616) contributing to c374B%, 279B% representing polychaetes and 303B% representing *Bathycoccus*..

Table 4. 3. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 168 TRF dimensional space for 18S community

Axis	Increment	Cumulative
1	0.284	0.284
2	0.290	0.574
3	0.246	0.820

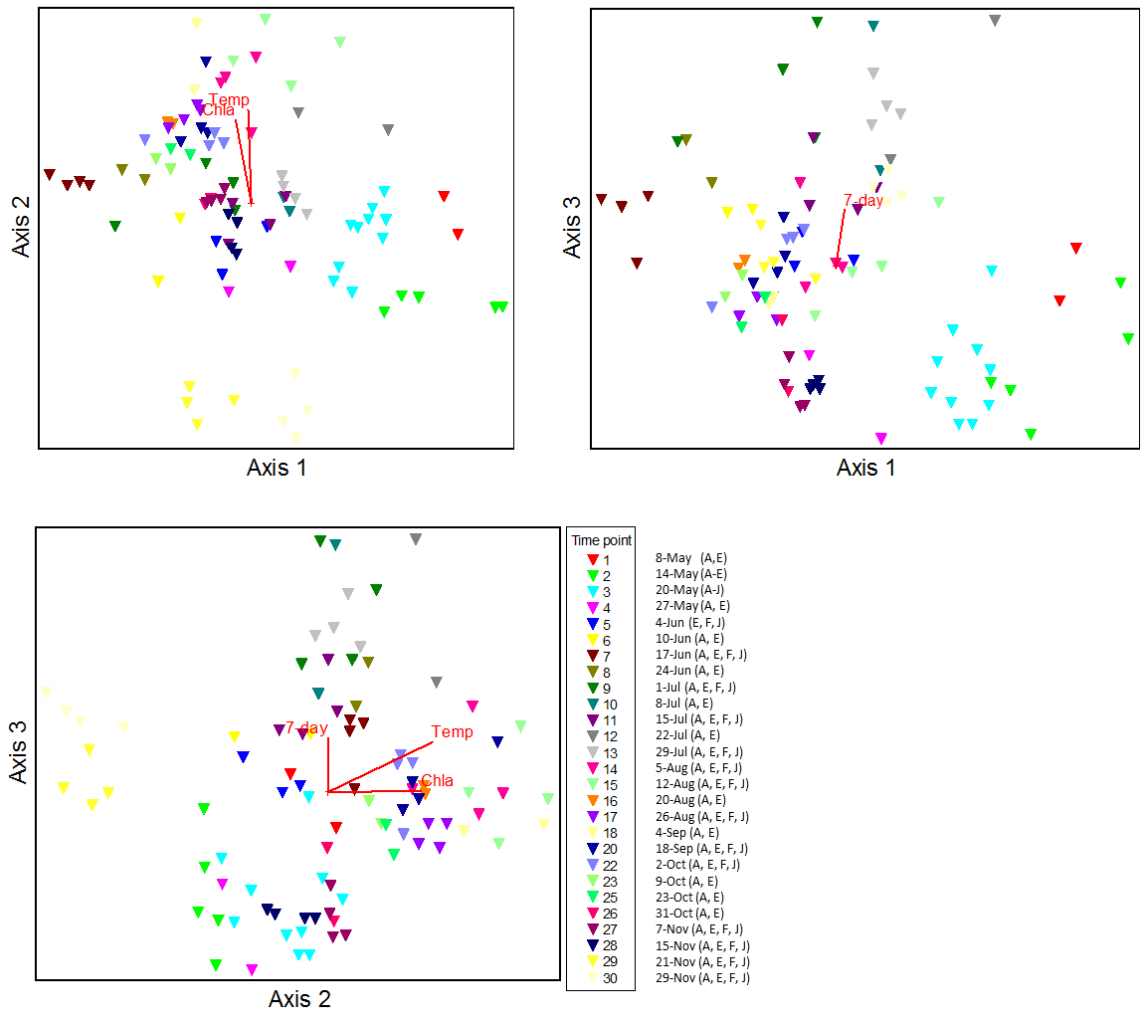


Fig. 4. 5. NMS ordination biplots of 18S microbial community in GSB, 2009

Table 4. 4. Pearson's Correlation Coefficients (r) between environmental parameters and ordination axes for 18S community (N= 92) and16S community (N=96); numbers in bold represent significant r values at p=0.001 according to the statistical table of critical values

Axis / **Env**	18S			16S		
	1	2	3	1	2	3
Temp.	-0.122	**0.644**	**0.444**	-0.083	**-0.417**	**-0.448**
Salinity	-0.225	0.051	-0.170	-0.231	-0.085	-0.060
Wind	-0.085	0.062	-0.079	**-0.310**	-0.197	-0.070
Gust	0.130	0.104	0.084	-0.294	-0.301	-0.062
River	0.270	-0.075	**0.429**	0.219	-0.103	**0.338**
7-day	0.181	-0.013	**0.461**	0.167	-0.253	0.202
Chla	-0.263	**0.608**	0.052	-0.206	-0.142	**-0.429**

Env: Environmental parameters
Temp: Surface water temperature
Wind: average wind speed on the sampling day
Gust: maximum gust speed on the sampling day
River: water discharge from the Connetquot River on the sampling day
7-day: water discharge from the Connetquot River over the prior 7 days including the sampling day
Chla: chlorophyll a concentration

2 16S community

86% of variation among samples was represented by 3 NMS ordination axes, with axes 2 and 3 accounting for slightly more variation than axis 1 (Table 4.5). Axis 1 was only significantly associated with wind speed, which was not a significant environmental factor on any of the ordination axes for 18S community (Table 4.4.). Interestingly, although river discharge on the sampling day was important for both 16S and 18S communities, salinity was not associated with any of the ordination axes for either 16S or 18S community (Table 4.4).Similar to 18S NMS, temporal variation of the 16S TRFLP was greater than spatial variation, and community structure between consecutive weeks was sometimes very different (Fig. 4.6.). Samples collected in May, June and late November (week 28) clustered together in the region of the biplot corresponding to lower temperature (Fig. 4.6. and Table 4.4.).

Table 4. 5. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 108 TRF dimensional space for 16S community

Axis	Increment	Cumulative
1	0.249	0.249
2	0.295	0.544
3	0.318	0.862

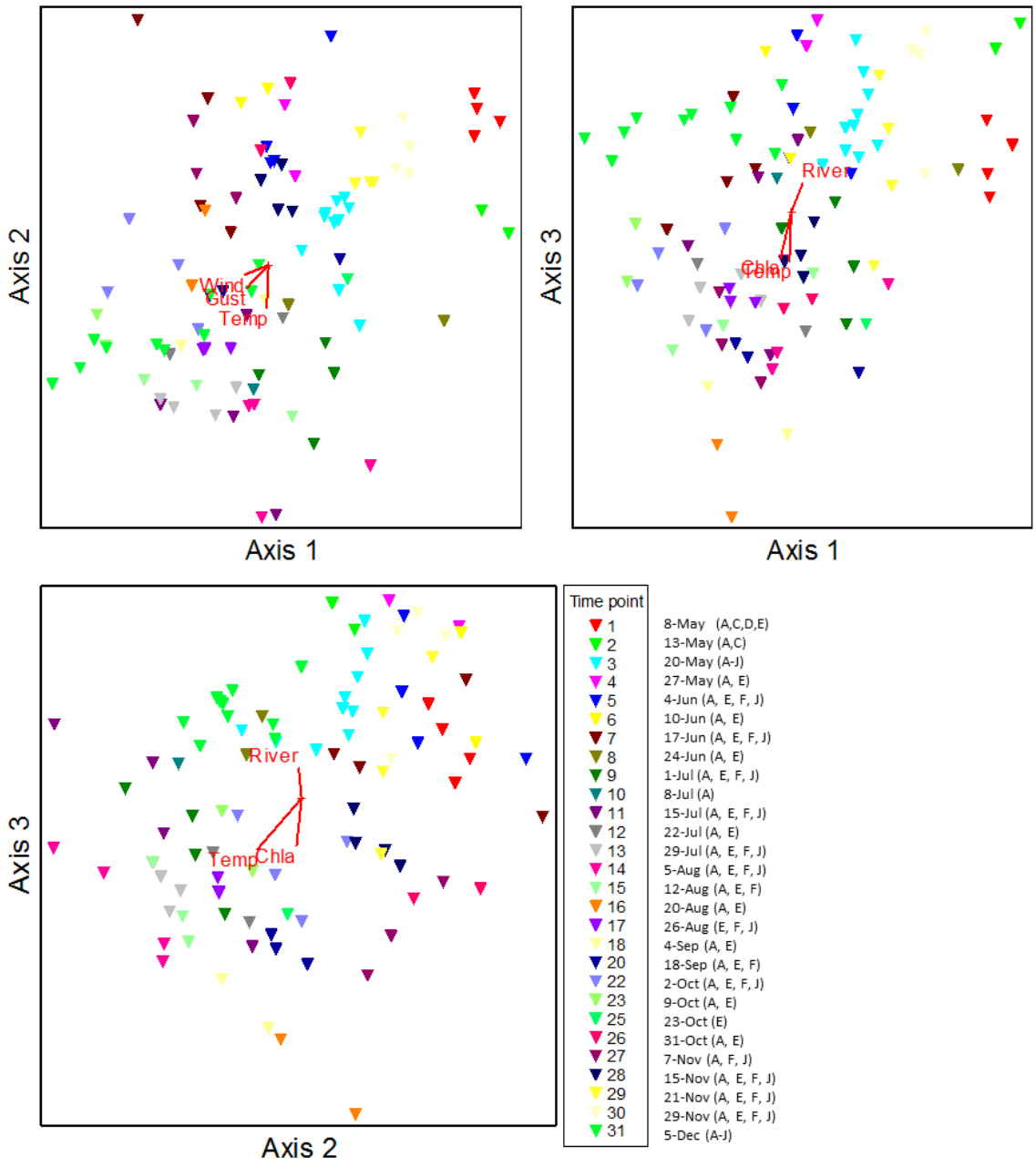


Fig. 4. 6. NMS ordination biplots of 16S microbial community in GSB, 2009

Spatial variation

1 18S TRFLP

There was a significant ($p<0.001$) positive relationship between the Bray-Curtis distance and the geographical distance for 18S samples collected from May 8 to November 29 in 2009 (Fig. 4.7.), even with the substantial variation of the Bray-Curtis distance for each geographical distance. The slope of the regression line is 0.0296, which indicates how much planktonic communities differ as they get further away from each other.

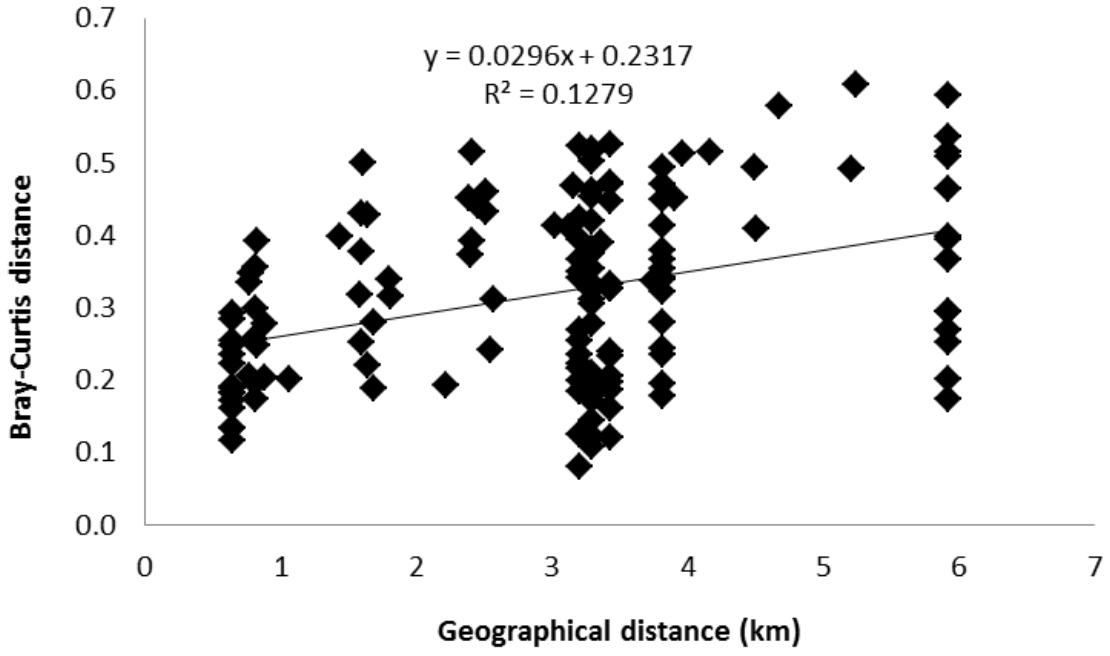


Fig. 4. 7. Linear regression (N=148) of the Bray-Curtis distance against geographical distance for 18S TRFLP in GSB, May 8 to November 29, 2009

Weeks when there were more than 2 spatial samples analyzed were also plotted separately (Fig. 4.8.). A significant correlation between Bray-Curtis distance and geographical distance was found for 8 out of 16 weeks in 18S TRFLP. The week with highest spatial variation along the transect was week 2 (May 13), when the slope of the regression line was 0.10. For the only sampling day (week 13, July 29) that was characterized by the reverse salinity structure (salinity at station E higher than that at station J), the slope of regression line was 0.0376, higher than the overall slope of 0.0296. However, there was not any along-transect pattern in the major TRFs.

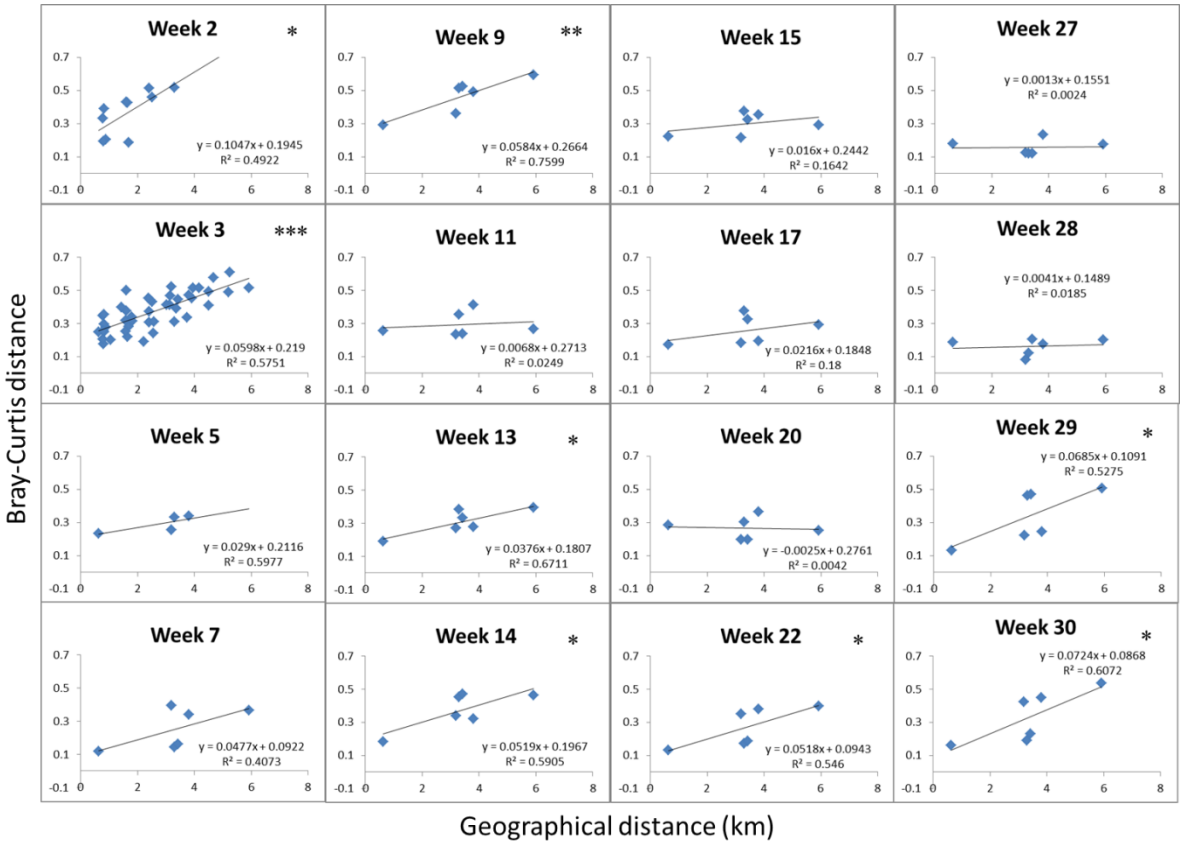


Fig. 4. 8. Linear regression of Bray-Curtis distance against geographical distance for 18S TRFLP from May 8 to November 29, 2009 ***: $p<0.001$; **: $0.001 \leq p \leq 0.01$; *: $0.01 \leq p < 0.05$

2 16S TRFLP

For 16S TRFLP, there was also a significant ($p<0.001$) increase of Bray-Curtis distance with geographical distance, but with a smaller slope than that of 18S TRFLP (Fig.4.9.). For all weeks when more than 2 spatial samples were analyzed, relatively fewer weeks demonstrated significant spatial heterogeneity in 16S TRFLP (4 of 12) than in 18S TRFLP (8 of 16) (Fig. 4.10.). Except for week 1, when no comparison was possible, all weeks with significant correlations for 16S TRFLP also had significant correlations for 18S TRFLP. Week 1 (May 8), when the slope of the regression line was 0.0537, had the greatest spatial variation along the transect. Different from 18S TRFLP, week 13 (July 29) with the reverse salinity structure did not show any significant spatial variation in 16S TRFLP. Weeks 14, 22, and 30 also showed significant relationships between Bray-Curtis distance and geographic distance for 18S but not 16S TRFLP.

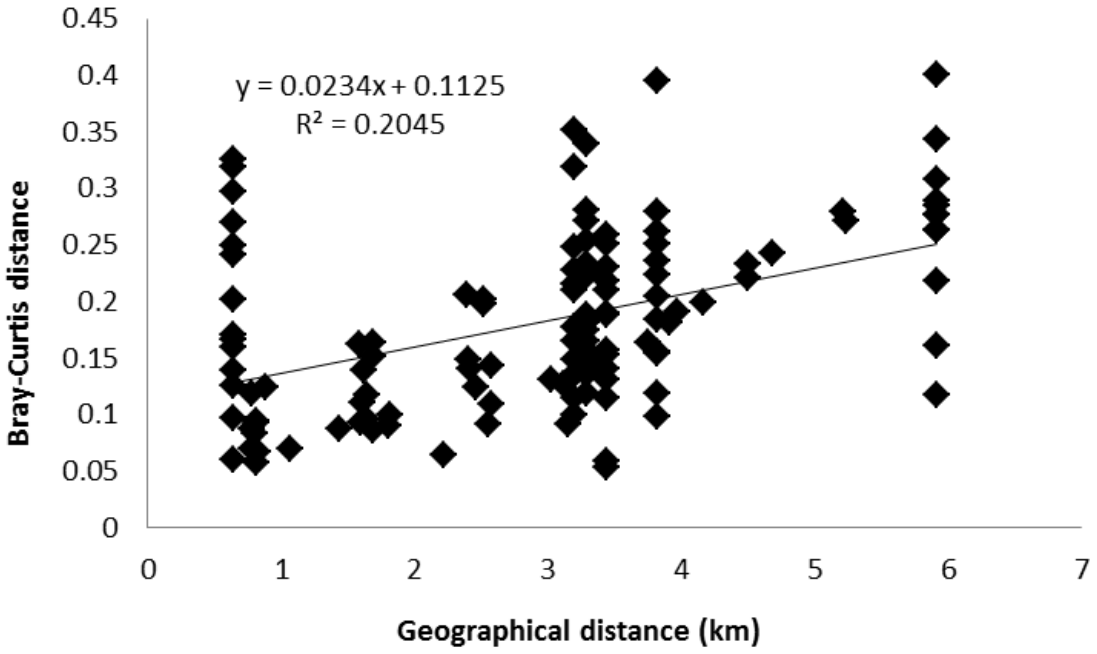


Fig. 4. 9. Linear regression (N=132) of the Bray-Curtis distance against geographical distance for 16S TRFLP in GSB, May 8 to November 29, 2009

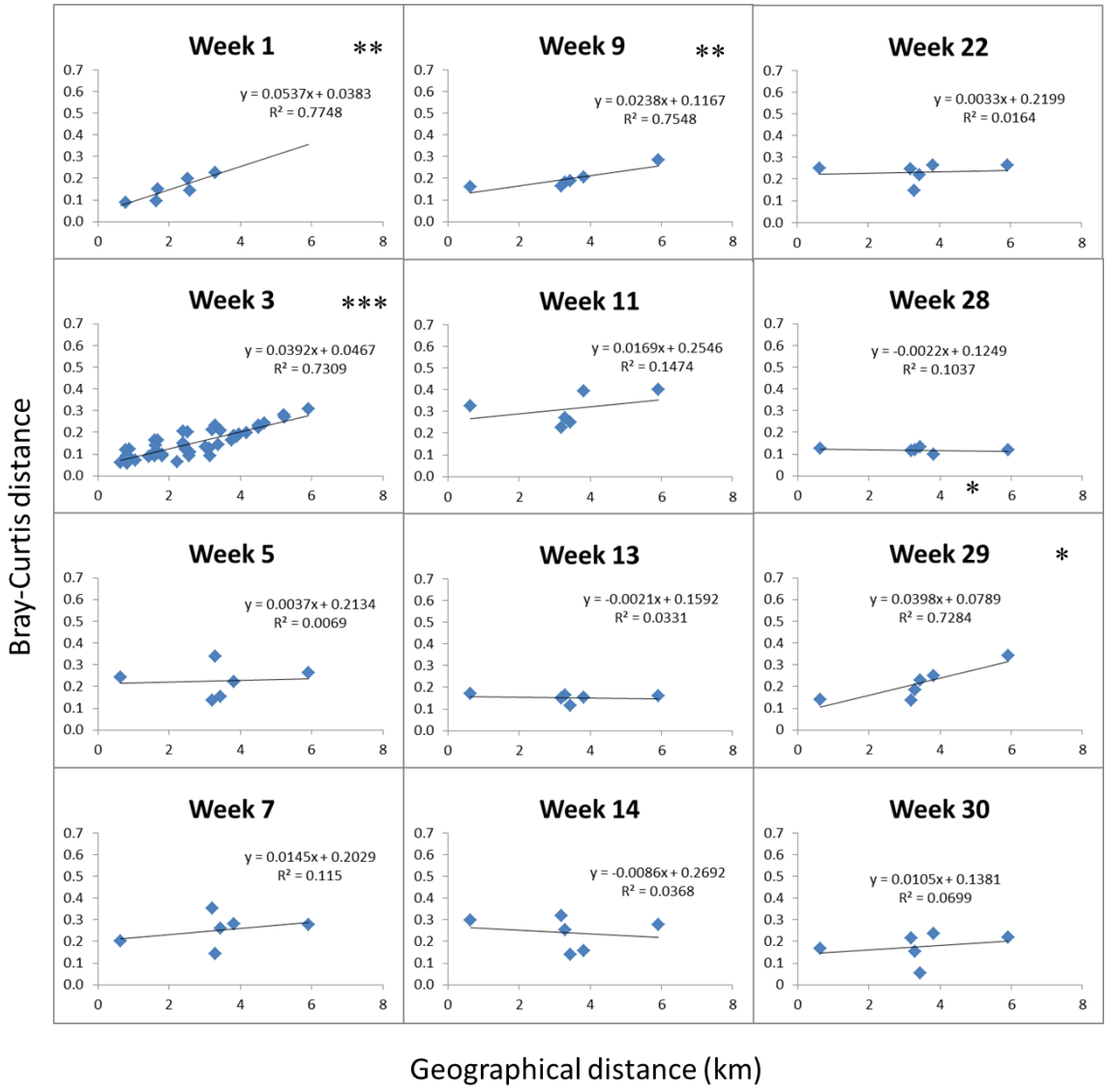


Fig. 4. 10. Linear regression of Bray-Curtis distance against geographical distance for 16S TRFLP from May 8 to November 29 ***: $p<0.001$; **: $0.001\leq p\leq 0.01$; *: $0.01\leq p<0.05$

Local Similarity Analysis

1 TRF-TRF and TRF-Environmental Factor associations

4657 LS correlations were calculated for all possible TRF pairs (TRF-TRF) as well as TRF and environmental factor (TRF-EF) pairs. 97 nodes representing all TRFs and environmental factors were linked by 433 LS correlations ≥0.3 with both p and q values ≤0.005. The highest LS correlation, 0.74, was between Connetquot River discharge on the sampling day and the sum of the prior 7 days. LSA agrees mostly with Pearson Correlation in that only 24 out of the 433 significant LS correlations were considered non-significant according to p values for Pearson Correlation Coefficient (Ppcc) of ≥0.005. With a less stringent Ppcc value of 0.05, only 4 out of the 433 significant LS correlations were considered non-significant.

Table 4. 6. Node abbreviations used in LSA network. Abbreviations with underscores are those filtered out in Cytoscape because they do not have any significant relationship with any other TRF or environmental factor.

Abbreviation	Observed TRF	Translation
Acti	264Y	Actinobacteria
Amoe	71B	*Amoebophrya*
Bact1	99Y	Uncultured Bacteria 1
Bact2	492Y	Uncultured Bacteria 2
Bacter1	297Y	Bacteroidetes 1
Bacter2	567Y	Bacteroidetes 2
Bacter3	576Y	Bacteroidetes 3
Bathy1	303B	*Bathycoccus* 1
Bathy2	305B	*Bathycoccus* 2
BathyP	100Y	*Bathycoccus* plastid
Cer/Ci1	260B	Cercozoa and Ciliate 1
Cer/Ci2	262B	Cercozoa and Ciliate 2
Cer/Ci3	264B	Cercozoa and Ciliate 3
Cerco1	180B	Cercozoa 1
Cerco2	259B	Cercozoa 2
Cerco3	266B	Cercozoa 3
ChaeP	272Y	*Chaetoceros* plastid
Chaeto1	282B	*Chaetoceros* 1
Chaeto2	284B	*Chaetoceros* 2
Cili	121B	Ciliate
Cope1	170B	Copepod 1
Cope2	297B	Copepod 2
Cope3*	313B	Mostly copepod 3
Cy/ChaeP	288Y	Cyanobacteria and *Chaetoceros* plastid
Cy/DiaP	258Y	Cyanobacteria and diatom plastid
Cyano	398Y	Cyanobacteria
Cyclo	529B	*Cyclotella*
Cylin	517B	*Cylindrotheca*
Delta	315Y	δ-Proteobacteria
Diat*	257B	Mostly unclassified Diatom
Dino1	310B	*Pentapharsodinium tyrrhenicum*
Dino2	591B	Autotrophic Dinoflagellate 2
Flavo1	143Y	Flavobacteria 1
Flavo2	313Y	Flavobacteria 2

Flavo3	372Y	Flavobacteria 3
Flavo4	486Y	Flavobacteria 4
Gam/Bet	580Y	Gamma and β-Proteobacteria
Gamma	303Y	γ-Proteobacteria
Gemi	244B	*Geminigera*
Dino3	589B	Autotrophic Dinoflagellate 3
Mari	291Y	*Marivita*
Micro	561Y	*Microcella*
Mixed pp1	315B	Mixed phytoplankton 1
Mixed pp2	317B	Mixed phytoplankton 2
Navi	519B	*Navicula*
PA*	371B	mostly *Picochlorum atomus*
Po/Cer1	322B	Polychaete and Cercozoa 1
Po/Cer2	324B	Polychaete and Cercozoa 2
Poly	279B	Polychaete
PX	488B	Stramenopile PX clade
Rhodo	374Y	*Rhodobacter*
SAR11	555Y	SAR11
Sph1	121Y	Sphingobacteria 1
Sph2	477Y	Sphingobacteria 2
Syn1	285Y	*Synechococcus* 1
Syn2	473Y	*Synechococcus* 2
Syndi1	236B	Syndiales 1
Syndi2	291B	Syndiales 2
Syndi3	289B	Syndiales 3
Telo1	177B	*Telonema* 1
Telo2	377B	*Telonema* 2
TreP	558Y	Trebouxiophyceae plastid
Temper		Water temperature
Sal		Salinity
Wind		Average wind speed
Gust		Maximum gust speed
River		Connetquot River discharge on the sampling day
7day		Sum of Connetquot River discharge for the prior 7 days

There were strong associations among Chaeto1, Chaeto2 and ChaeP, agreeing with their identity as TRFs generated from *Chaetoceros* (Table 4.6., Fig. 4.11.). *Chaetoceros* was

negatively associated with many eukaryotic phytoplankton including *Navicula, Cyclotella,* autotrophic dinoflagellates (Dino3), *Geminigera*, and *Bathycoccus*, but was positively associated with *Synechococcus* (Syn1), *Picochlorum atomus*, and PX clade (phylum Xanthophyceae and order Mischococcales) (Fig. 4.11.). The negative association between *Chaetoceros* and *Bathycoccus* as well as the positive association between *Chaetoceros* and PX clade was captured in both Cheato and ChaeP subnetworks. Among the Bacteroidetes (Sph1, Sph2, Flavo1, Flavo3, Bacter2 and Bacter3) and grazers/parasites (Cerco1, Cerco2, Cerco3, Cer/Ci3 and Syndi2) that *Chaetoceros* was associated with, the relationship was almost always positive (exceptions were ChaeP vs Bacter3 and ChaeP vs Cerco2). *Chaetoceros* was positively associated with temperature, Chla and accumulated 7 day river discharge.

There were 8 different Bacteroidetes TRFs (Bacter1 was not involved in any significant association and was filtered out), with 4 representing *Flavobacteria* and 2 representing *Sphingobacteria* (Table 4.6). For 28 possible associations among the 9 Bacteroidetes TRFs, only 4 were significant (Sph1 vs Bacter3, Flavo1 vs Bacter3, Flavo3 vs Flavo4, Sph2 vs Flavo4) and all of them were negative (Fig. 4.12). Different environmental factors were important for different Bacteroidetes: Connetquot river discharge was significantly associated with Bacter2 and Bacter3; temperature was strongly associated with Flavo3, Flavo4 and Sph2, while wind speed was significantly associated with Flavo2.

Within the subnetwork centering on PA*, Bathy1 and Bathy2, *Bathycoccus* (Bathy1, Bathy2 and BathyP) and *Picochlorum atomus* were negatively associated and they had negative and positive associations with *Synechococcus* (Syn2), respectively (Fig. 4.13.). *Picochlorum* was positively associated with Chla and negatively associated with Connetquot River discharge. Within the diatom network (Fig. 4.14.A.), for many of the TRFs *Cylindrotheca*, *Cyclotella* and *Navicula* were all associated with, *Cylindrotheca* always had the opposite association as the others. While *Cylindrotheca* and unidentified diatom (Diat*) were positively associated with Chla and temperature, *Navicula* was negatively associated with Chla and *Cyclotella* was negatively associated with temperature. This negative association between *Cyclotella* and temperature agrees with *Cyclotella*'s temporal dominance in May and late November when it was colder (Fig. 4.2.). There was no direct association detected between any two dinoflagellates

(Fig. 4.14.B.), and temperature was the only factor that was associated with more than 2 dinoflagellates (Syndi1, Dino3 and Dino2).

Planktonic grazers were grouped together to provide a higher trophic level perspective (Fig. 4.15.). Polychaetes (Poly, *Polygordius* and *Heteromastus*), copepods (Cope2, *Pseudodiaptomus*) and ciliates (Cili, *Strombidium*) were negatively correlated with *Picochlorum atomus* while Cercozoa and ciliates (Cer/Ci3) was positively correlated with *Picochlorum atomus*. Also, while one copepod TRF (Cope1) was positively correlated with Diat*, Cercozoa and ciliates (Cer/Ci1 and Cerco2) were negatively correlated with Diat*. Many of the pelagic grazers (such as Cope1, Cope2, Cer/Ci1, Cer/Ci2, Po/Cer1, Cerco1, Cerco3) were not associated with any environmental factor. Polychaetes (Poly) had the most significant associations with environmental factors, which included positive associations with river discharge and water temperature, and a negative association with wind speed. The parasitic dinoflagellate order Syndiniales, its genus *Amoebophrya*, and parasitic *Pirsonia* (1 sequence retrieved but the observed TRF matching its predicted TRF was below detection) were present in GSB in 2009. Both Syndiniales and *Pirsonia* are known to be parasitoids-obligatory killers of their host (Kuhn et al. 2004, Guillou et al. 2008). There seemed to be fewer significant associations involving parasitoids in 2009 than 2008 (Fig. 4.14.B. and Fig. 3.19.A) .

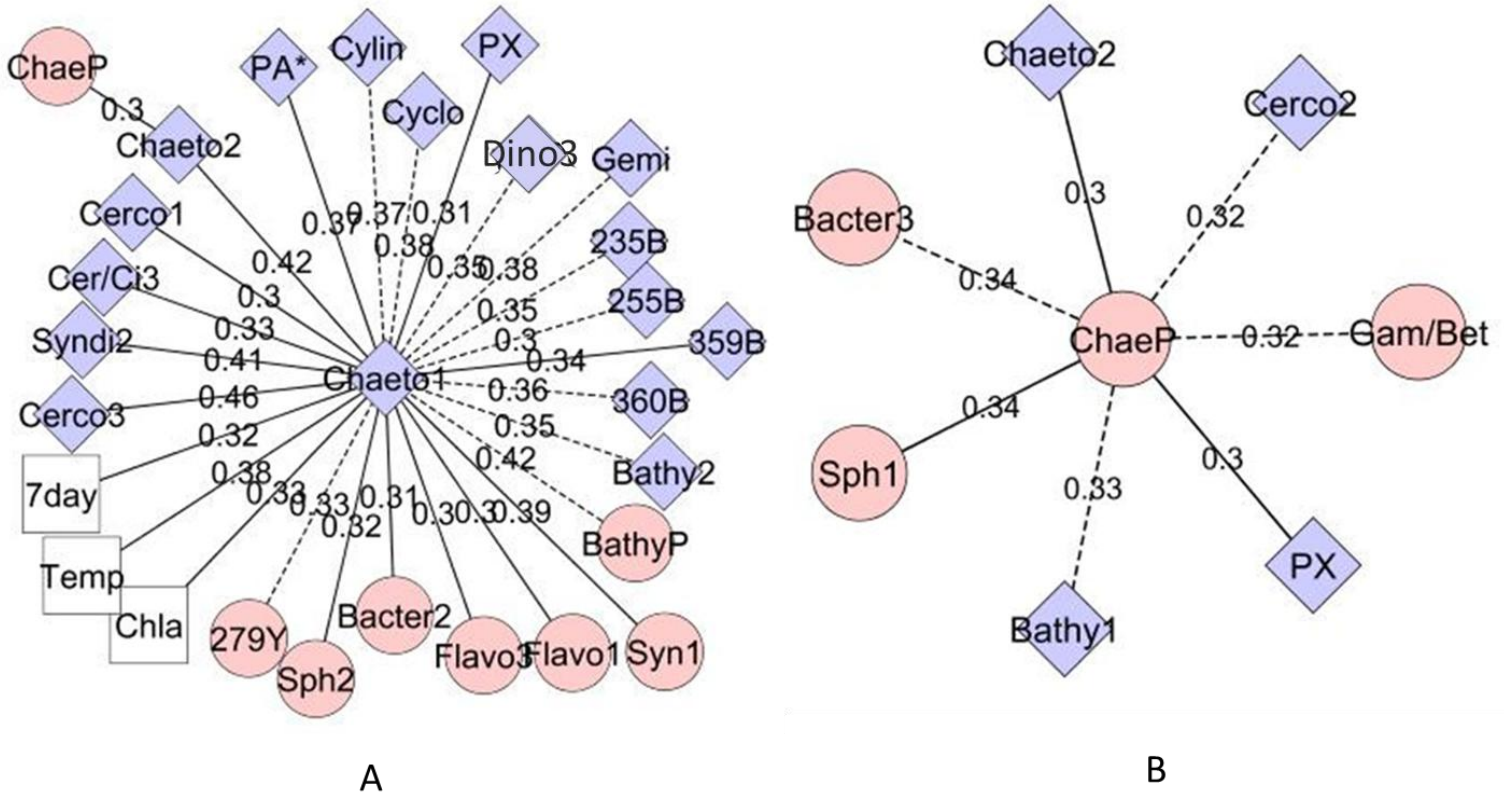


Fig. 4. 11. *Chaetoceros calcitrans* subnetworks. A: With Chaeto1 and Chaeto2 as central nodes; B: With ChaeP as central node. Diamond: 18S TRF nodes; Circles: 16S TRF nodes; Square: Environmental factors; Solid lines represent positive relationships and dashed lines represent negative relationships. Numbers on the lines are LS scores. Subnetworks show only LS correlations between central nodes and their first neighbors but not among their first neighbors. The length of the connecting lines between nodes and the angle between lines are arranged just for the easiness of visualization.

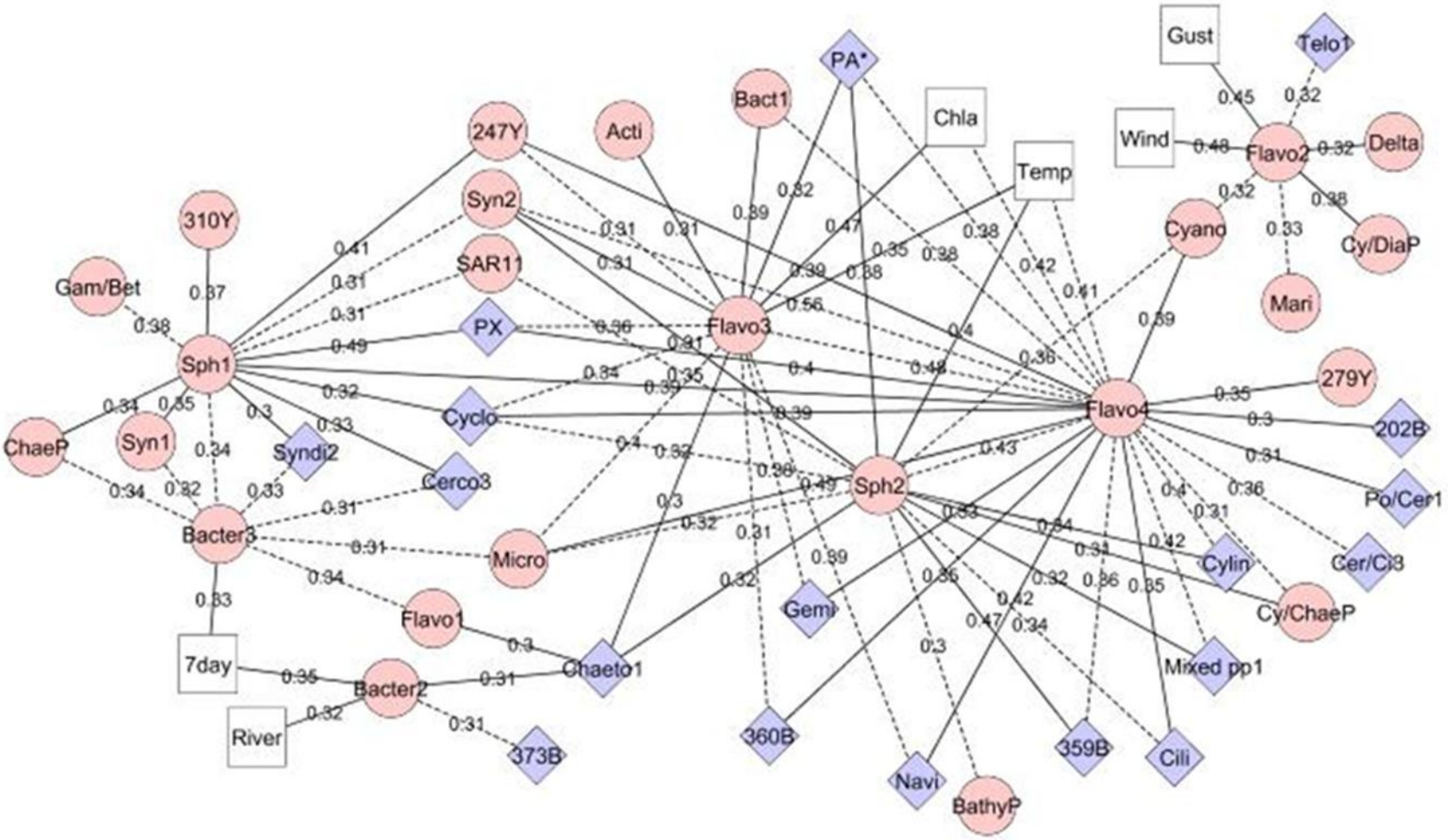


Fig. 4. 12. Bacteroidetes subnetwork; See Fig 4.10 legend for details.

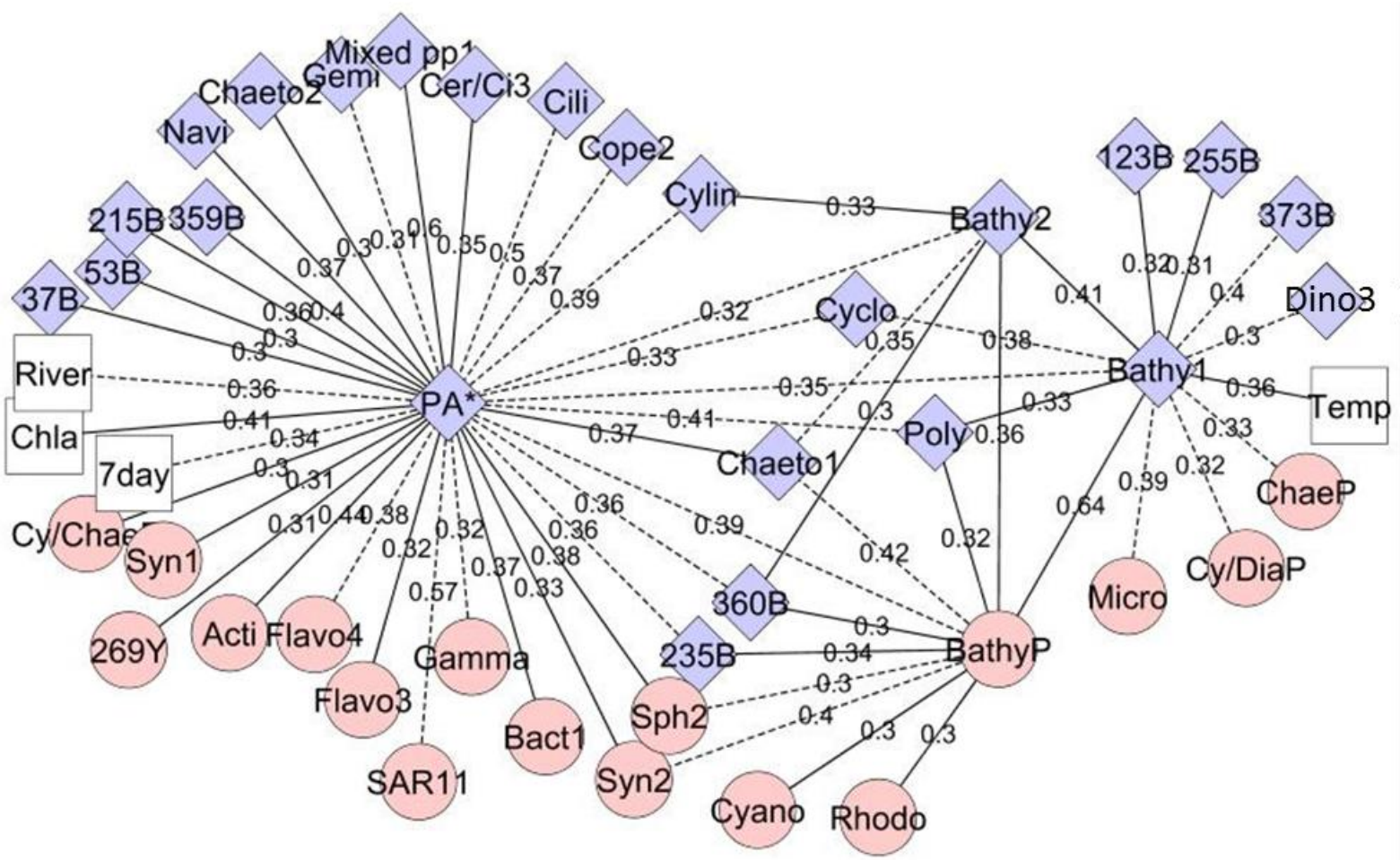


Fig. 4. 13. Subnetwork of pico-chlorophytes; See Fig 4.10 legend for details.

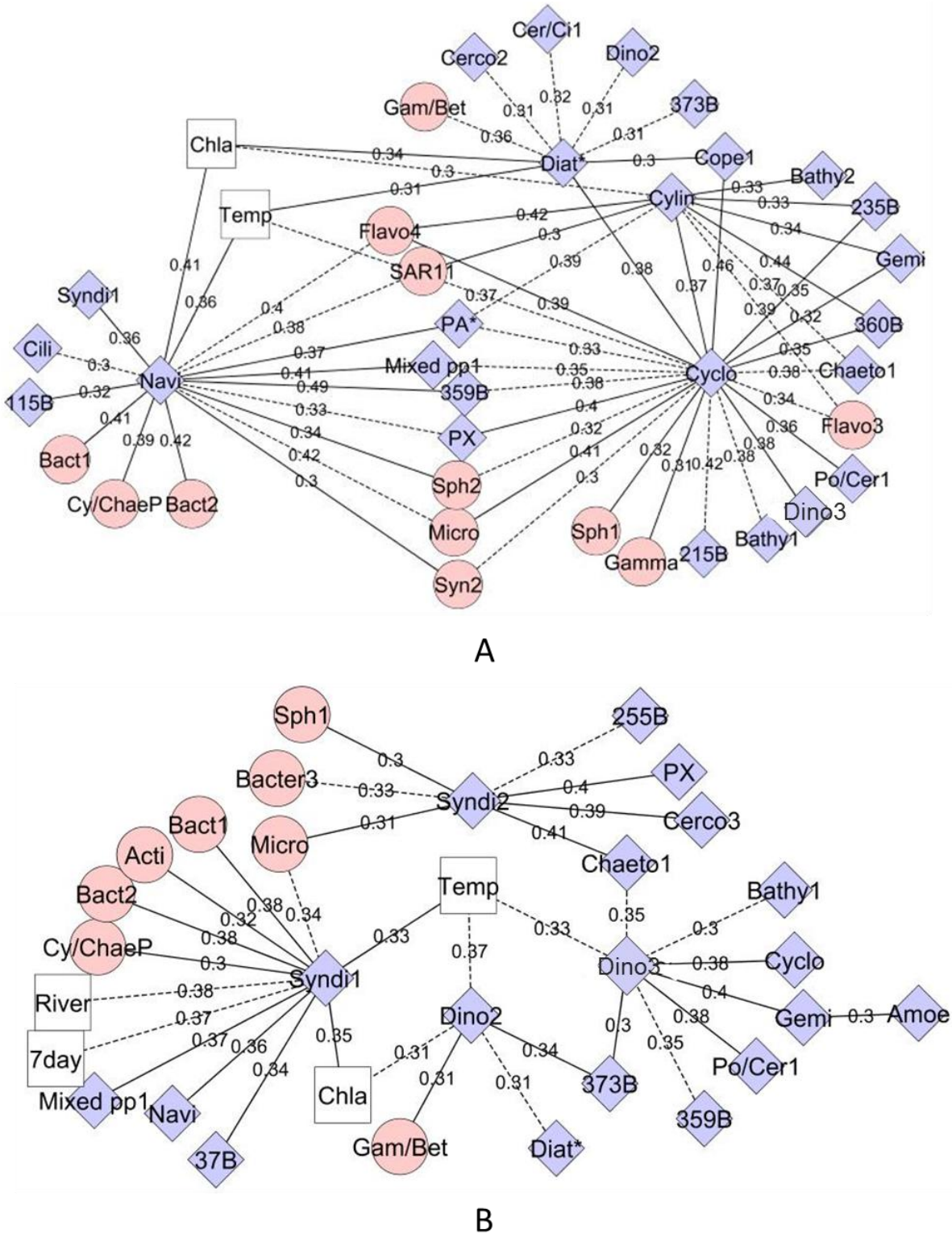


Fig. 4. 14. Subnetworks of diatoms (A) and dinoflagellates (B); See Fig 4.10 legend for details.

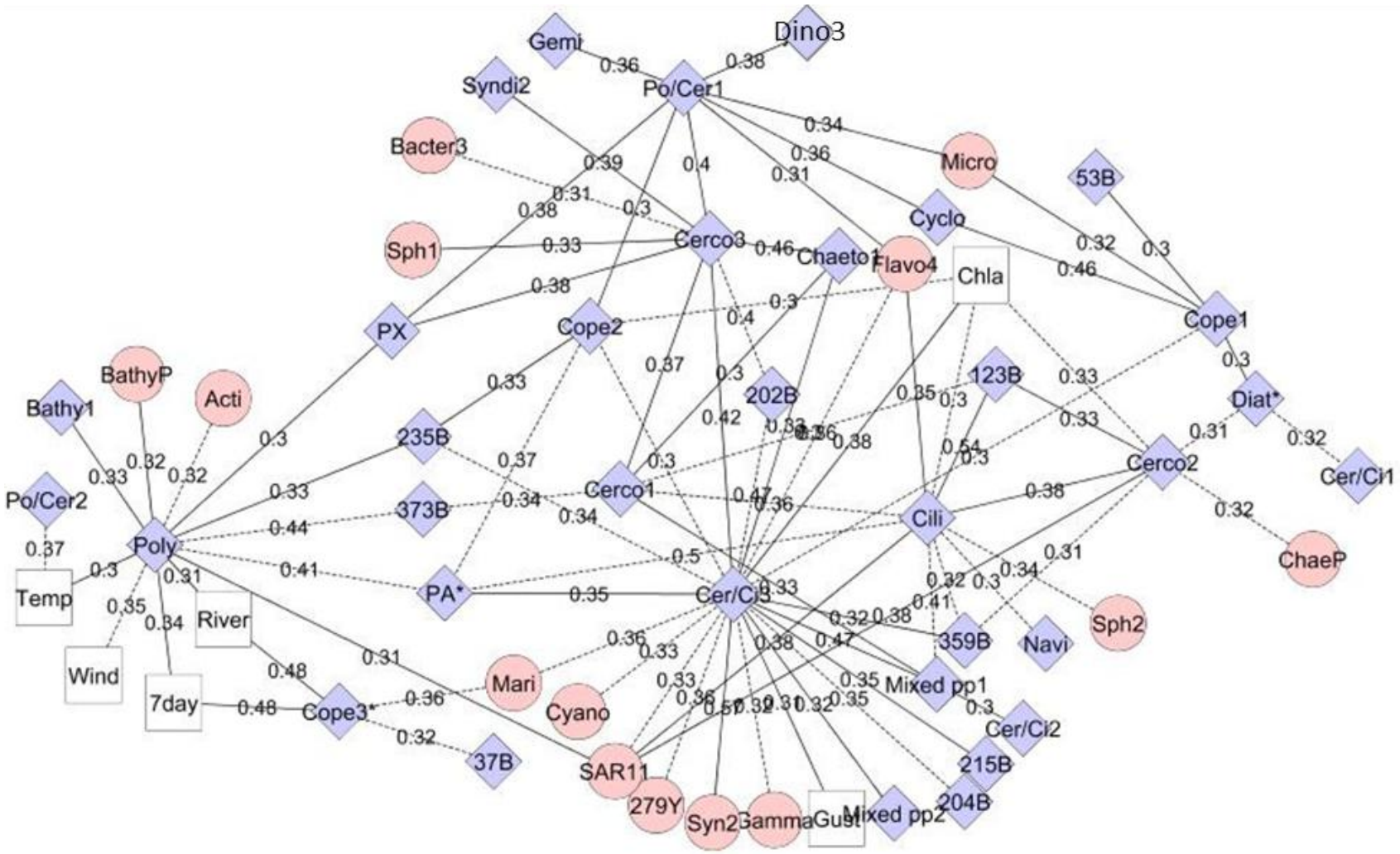


Fig. 4. 15. Subnetworks of potential grazers; See Fig. 4.10. legend for details.

2 Time-delayed associations between TRF-TRF and TRF-Environmental Factor (TRF-EF) pairs

Converting the local similarity matrixes LSM to LSM-T sacrificed more than two- thirds of the information (93 samples to 28 samples) and this resulted in both higher p values for LS correlations and fewer significant TRF-TRF and TRF-environmental factor (TRF-EF) pairs even at p=0.05 (72 nodes and 172 significant LS scores with $p<0.05$, $q<0.05$ and $LS>0.3$). Because the time delay was set at 4 in LSM-T analysis, there were 9 possible time lags between two associated organisms (from -4 to 4), so significant associations with delay=0 should in theory be about one ninth (11%) of the total significant associations, assuming all time delayed associations had an equal chance to occur. Among the 172 significant LS correlations identified, 50 were instantaneous correlations without time delay, which was about 30% of the total significant associations. As the time delay increased, the proportion of significant associations decreased (±1 delay: 21%; ±2 delay: 19%; ±3 delay: 16%; ±4 delay: 14%). A few cases when associations were significant only in the time delayed LSA are described below.

Syndiniales (Syndi2) was negatively associated with TRF 37B with a time lag of 2 weeks (Fig. 4.16., upper panel). The negative association existed, however, only after removing data collected at the beginning (4 and 6 weeks for Syndi2 and 37B, respectively) and at the end (9 and 7 weeks for Syndi2 and 37B, respectively) of the sampling season. As a matter of fact, at the end of the sampling season, Syndi2 increased while 37B was quite stable. Ciliates (Cili) and TRF 269Y were also negatively associated with a time lag of 2 weeks (Fig. 4.16., middle panel). Although they illustrated clear reciprocal temporal trends from time point 13 to time point 23 in the time-delayed alignment, the variation in the relative abundance of ciliates before week 13 corresponded to a rather stable and very low period for 269Y. Similarly, after moving forward by 3 time points, *Flavobacteria* (Flavo3) aligned better with Cercozoa/ciliates (Cer/Ci2) although there was still disagreement between the two TRFs' temporal patterns from time point 9 to time point 14 (Fig. 4.16., lower panel).

For TRF-EF pairs, similar to 2008, significant correlations missed in the LSM run but found in the LSM-T run were mostly those with a time delay of 3 or 4 weeks. The negative association between *Cylindrotheca* (Cylin) and Connetquot River discharge was quite clear after moving river discharge backward for 4 weeks and removing 5 weeks' of data (Fig. 4.17., upper

panel). Similarly, in order for the positive association between Syndiniales (Syndi2) and 7 days' sum of Connetquot River discharge to be discerned, 12 weeks of data were removed (Fig. 4.17., middle panel). A period of highest relative abundance of copepod (Cope3*, mostly *Oithona*) corresponded with the period with highest temperature after moving temperature backward for 3 weeks. However, a local minimum of Cope3* relative abundance at week 11 was not reflected in the temperature data (Fig. 4.17., lower panel).

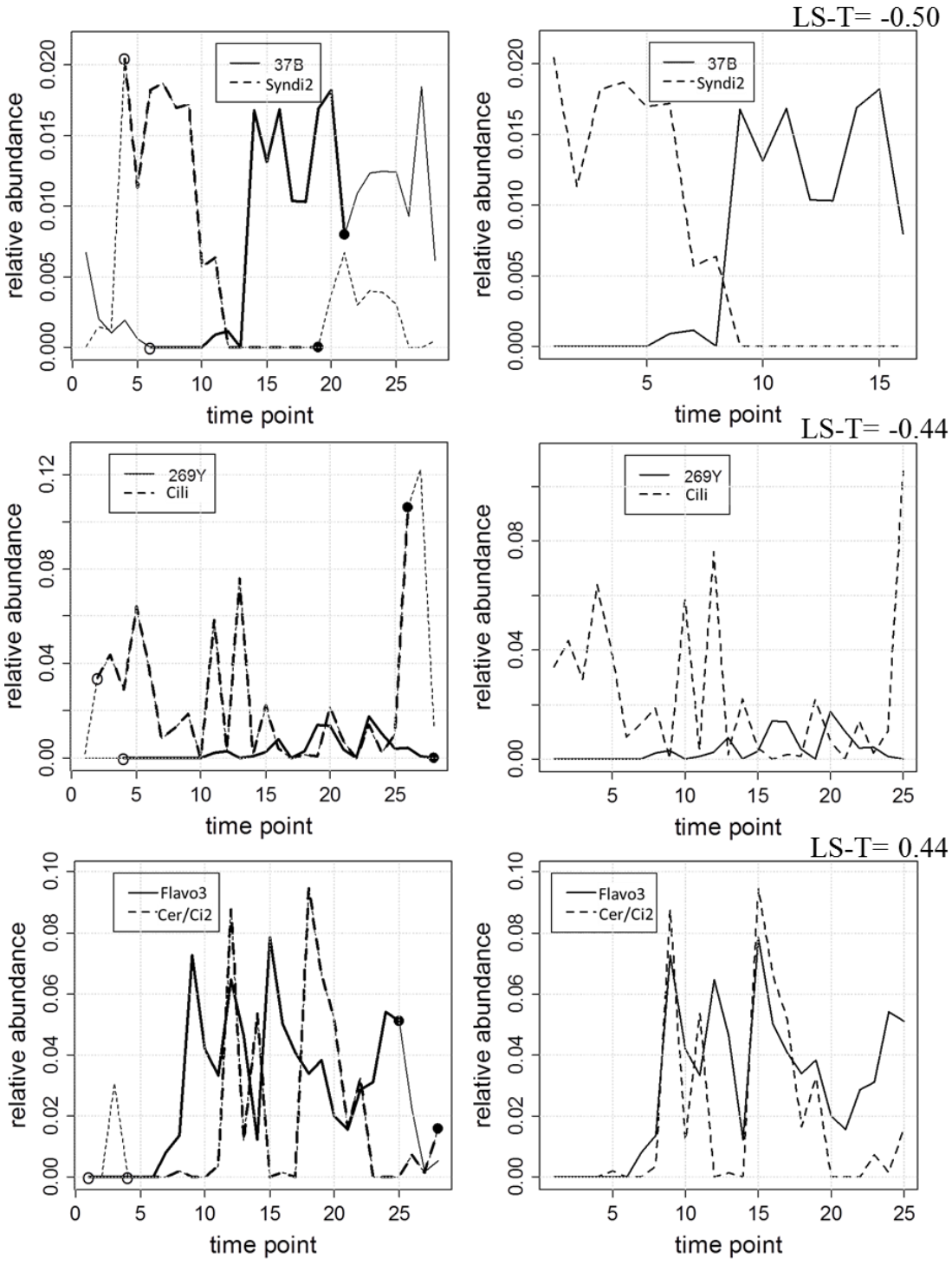


Fig. 4. 16. Significantly associated TRF-TRF pairs with time delay. Open and closed circles indicate the starting and ending points of the association. Left column: original data; right column: time-delayed alignment. Upper, middle and lower panels represent different pairs.

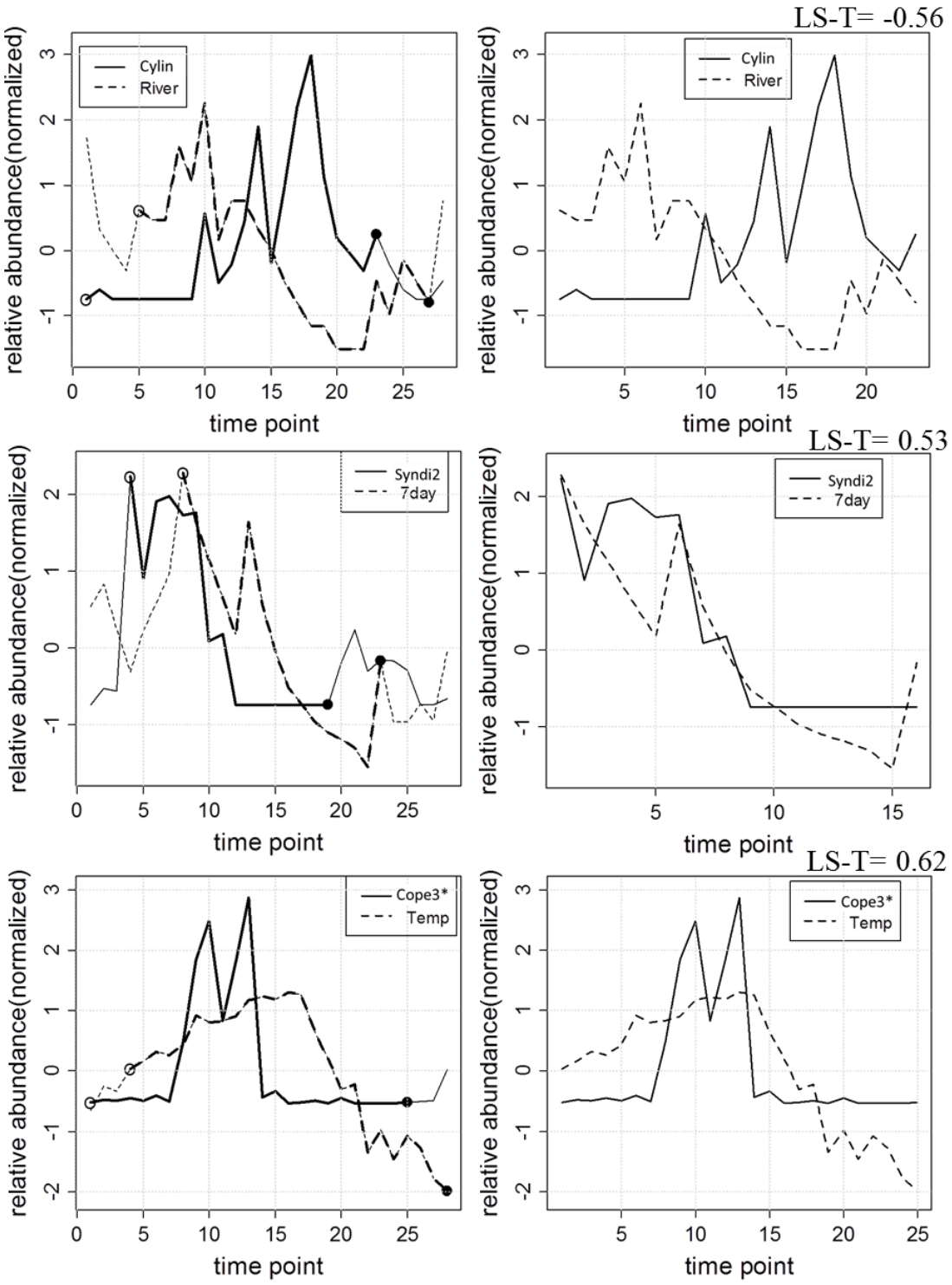


Fig. 4. 17. Significantly associated TRF-EF pairs with time delay. Open and closed circles indicate the starting and ending points of the association. Left column: original data; right column: time-delayed alignment. Upper, middle and lower panels represent different pairs.

Discussion

Plankton community composition in GSB

Major TRFs representing a variety of planktonic grazers including copepods, ciliates, and Cercozoa, were abundant in 2009. Copepod *Oithona* contributed to major TRF c316B (Fig. 4.2., Table A3.) and 18 *Oithona* partial rDNA sequences were retrieved from the sample collected at station J on July 1 (week 9), which also coincided with the time when 316B% was highest of the season. High copepod abundance was also found during late June to mid-July in GSB in 1985 and 1986 (Duguay et al. 1989), when *Acartia hudsonica* and *Acartia tonsa* were the dominant copepods, although *Oithona similis* was also reported to be present. *Acartia* was also reported to be dominant in 2001 in GSB (Cosper 2001). The current study, however, did not capture any sequence of *Acartia tonsa*, which would have been recognized because a reference sequence is available in GenBank. The dominance of *Oithona* sequences in this study may reflect some change in the GSB planktonic community over the last decade (2001 to 2009). Other copepod sequences retrieved included those of *Harpacticus*, *Sinocalanus* and Pseudodiaptomus. While copepods can potentially prey upon small phytoplankton such as *A. anophagefferens* directly, they may feed more effectively on protozoan herbivores such as ciliates and flagellates (Caron et al. 1989 , Lonsdale et al. 1996, Boissonneault-Cellineri et al. 2001). The abundance of *Oithona* in 2009 found in this study calls for more work to better characterize the copepods in GSB, which may be different from 20 years ago.

Although not a major TRF, the observed 121B representing aloricate ciliates (*Strombidium*) was present in GSB on 24 of the 28 sampling days with average relative abundance of 2.5%. In contrast, the observed TRFs (361B and 363B) representing tintinnid ciliates were present at below 0.1% throughout the sampling season. The dominance of aloricate ciliates over loricate ciliates was also found in the number of sequences retrieved for each group. 23 of 351 partial rDNA sequences had predicted TRFs the same size as aloricate ciliate *Strombidium* but there was only one loricate tintinnid sequence (*Codonellopsis nipponica*) retrieved (Table A3.). These observations agree with an earlier study in GSB in the summer of 1988, when non-loricate ciliates were the most abundant ciliates, although instead of the Oligotrichia *Strombidium* identified in this study, Choreotrichia ciliates were identified in 1988 (Caron et al. 1989). A more recent investigation of the microplankton in GSB identified

Strombidium as the most abundant aloricate ciliates (Lonsdale et al. 2006). Lonsdale et al. (2006) also found that loricate ciliates, including *Tintinnopsis* and *Favella*, were present in GSB in 1998, although contributing less to total micozooplankton biomass than aloricate ciliates.

Major TRF c264B, although composed of 4 observed TRFs, represented almost exclusively Cercozoa, a group of diverse predatory ameoboflagellates (Chantangsi et al. 2008). While there have been no grazing studies of this group in GSB, many Cercozoa are found to be both bacterivorous and eukaryovorous (Myl'nikov & Karpov 2004, Chantangsi et al. 2008). A recent study revealed that grazing of 9 genetically closely related and morphologically similar cercozoan strains had significantly different impacts on both the abundance and the composition of the bacterial communities (Gluecksman et al. 2010), pointing out Cercozoa's potential to shape bacterial community structure. In the future, lab-based experiments focusing on isolating/culturing Cercozoa from GSB water and understanding their grazing behavior on different planktonic organisms would contribute tremendously to our overall understanding of the plankton community in GSB, and probably other ecosystems as well.

The relative abundance of *Synechococcus*, represented by major TRF c475Y, started to increase from late May and became the dominant 16S TRF in July, which agrees with the trend indicated by FCM determined *Synechococcus* cell abundance in 2009 (Figure 4.4.). The timing of *Synechococcus* cell abundance increases also agreed with previous studies conducted in GSB and other Long Island embayments (Sieracki et al. 1999, Gobler et al. 2004a, Gobler et al. 2004b). Dominance of c475Y% persisted until at least the end of October, and it was present occasionally at >15% even in the colder season (November 6 and December 5). Clear understanding of *Synechococcus* dynamics, especially its interactions with other picophytoplankton, is the key to better describe the picoalgal niche, suggested by Sieracki et al. (1999) to exist in summers in Long Island waters. Although falling a little below the major TRF standard, the observed TRF 304B, representing *Bathycoccus*, was present in GSB in 2009 at 4.48% on average. With two peak values of 304B% found on May 13 and November 29 at 22% and 14%, respectively, *Bathycoccus* seemed to switch seasonally with *Synechococcus* in being the most abundant picoalgae. Potential interactions among picophytoplankton will be discussed further in the following section on microbial associations.

Cyclotella choctawhatcheeana, noted during 2003-2004 as an important diatom species in GSB by Anderson and Gobler (personal communication with Gobler) and represented by major TRF c532B in this study, was most abundant in late spring and late fall when the water temperature was ~10-15 ℃ (Fig. 4.1, Fig. 4.2., Fig. 4.14.A.). Unfortunately, Anderson and Gobler's study was conducted only during June, so the temporal pattern cannot be compared. An earlier light microscopy study by Lively et al. detected an unidentified *Cyclotella* species in GSB only in July, 1979 and May, 1980 at 5300 cells/L. At other 7 sampling times from August 1979 to June 1980, *Cyclotella sp.* was not detected (Lively et al. 1983). *Cyclotella choctawhatcheeana* was originally described from samples taken in Choctawhatcheeana Bay, FL (Prasad et al. 1990) and has been found to bloom in April, May, June and November/December in many Florida bays (Prasad et al. 1990, Livingston 2000). It is now known that *Cyclotella choctawhatcheeana* has a cosmopolitan distribution in coastal brackish water (Hakansson et al. 1993, Carvalho et al. 1995) and can tolerate a wide temperature range from 10 to 30 ℃ based on its distribution pattern in Florida (Prasad & Nienow 2006). The 'cold water species' feature of *Cyclotella choctawhatcheeana* found in this study, therefore, still waits to be confirmed.

Chaetoceros calcitrans, represented by TRF c284B, reached its greatest relative abundance in week 7 (June 15) (Fig. 4.2.). c284B% that day was also the highest area contribution ever achieved by any major TRFs on any sampling day. Continuous (about a week) rainy/stormy days persisted before water samples were collected on June 15 in GSB. Rain/storm events can potentially have different effects on the plankton community in a shallow embayment like GSB. First, they can bring in a species (*C. calcitrans* in this case) native to the terrestrial or freshwater environment. Second, they can bring in specific nutrient forms (through overland runoff or groundwater discharge) that favor the growth of *C. calcitrans*. Community composition in week 7 was characterized at stations A, E, F, and J with c284B% being similar at the three stations closer to the mainland (A, E, F), around 47%, but much lower, 24%, at station J. Therefore, the first possibility of terrestrial/freshwater input seemed to be supported by the spatial pattern of c284B%. However, salinities at station E and J were almost the same (23.0 PSU vs 23.1 PSU), which does not seem to suggest that there was any 'fresh water signal' gradient along the transect. Additionally, since there was no spatial pattern for any 16S major TRFs (data not shown), the first hypothesis could only work assuming that the terrestrial input source carried a 16S community that was similar to that present in GSB. It is not possible to test

whether the second scenario had anything to do with the unique 18S TRFLP signal or the c284B% along-transect gradient on June 15 due to the lack of nutrient data at the sampling sites and in the groundwater discharge before, during and after June 15. In the future, a better characterization of nutrient composition, especially N species, at the sampling site in GSB, from Connetquot River input and groundwater discharge into GSB will help understand the driving force of the outburst of *C. calcitrans* ~ mid-June, in both years and maybe in other years as well.

Spatial heterogeneity in 18S and 16S communities

Spatial heterogeneity was present in both 18S and 16S communities in 2009, although it was stronger in the 18S community. Comparison of the individual weeks showed that spatially heterogeneous weeks for the 16S community were always also spatially heterogeneous weeks for the 18S community but not vice versa. This indicates that as one moves away from a location, the chance of getting a different 18S community is greater than that of getting a different 16S community. One intuitive explanation is that prokaryotes have a more cosmopolitan distribution due to their larger absolute population sizes, while for the generally larger eukaryotes with smaller populations, endemism is more common and the chance of getting something different as one moves away from one location is therefore larger. This 'cosmopolitan distribution' vs 'endemism' distinction has been suggested to separate cells smaller from those larger than 1 millimeter at the global scale (Fenchel & Finlay 2004), but data generated in this study implies that although maybe less often seen, even the prokaryotic community can exhibit spatial variability over a spatial scale of 10 km. One caveat of this is the greater chance of sampling a different eukaryotic community than a different prokaryotic community by processing a limited volume of water sample. The lower eukaryotic cell abundance within 100 ml of sea water (the volume of sea water filtered in this study) certainly increases the chance of collecting different eukaryotic cells at a different location, although there should not be any spatial pattern of this random effect.

One may still wonder then what factor(s) make some weeks spatially heterogeneous while other weeks are not. Is this determined by any of the environmental factors included in this study? Since the presence of spatial heterogeneity did not have any clear seasonal pattern, temperature can probably be ruled out. The salinity difference between station E and J was from 0.1 to 3.7, and spatially heterogeneous weeks were not always associated with a larger salinity

gradient across the bay. It is worth mentioning that southern wind almost across the bay was prevalent on July 29 (week 13, the week with 'reverse salinity structure'), which was different from the along-shore feature of the wind causing 'reverse salinity structure' in 2008. How about freshwater input through river discharge, which may have a separate effect from salinity on microbial community composition since it may serve as a foreign source of microorganisms to GSB? Spatially heterogeneous weeks were associated with wide ranges of both river discharge and accumulated river discharge. Assuming that terrestrial/freshwater input was the cause of the unique signal of TRF c284B in week 7 (Fig. 4.2.), the presence of spatial heterogeneity in neither 18S nor 16S communities in that week indicates that river discharge alone was not capable of producing spatial heterogeneity for the whole community, perhaps reflecting the rapid dilution of fresh water into the bay. It seems that no single environmental factor can easily explain the higher spatial heterogeneity in certain weeks. Various environmental factors may work together or play important roles in defining spatial structure of the planktonic microbial community at different times.

Microbial interactions in GSB in 2009

Synechococcus represented by major TRF c475Y was highly abundant, at ~20% of total peak area from July to October (Fig. 4.3.). Defining its interactions with other picophytoplankton is essential to better understand the hypothesized picoalgal niche when brown tide blooms are absent. The picoalgal niche hypothesis states that a succession from larger to smaller phytoplankton cells occurs in Long Island bays from April to May, and under non-brown tide bloom conditions *Synechococcus* dominates the picoalgae class (Sieracki 2001). This study found a significant positive relationship between *Synechococcus* and *Picochlorum atomus* (Fig. 4.13.), and they were probably both dominant picoalgae from July to November (week 7 to week 25) (Fig. 4.18.). *Synechococcus* and *Picochlorum atomus* shared a similar temporal trend, which was different than that of *Bathycoccus. Bathycoccus* belongs to the algal class Prasinophyceae, which are common in temperate and cold waters (Gescher et al. 2008). Low relative abundance of *Bathycoccus* from June until November may reflect their preference for lower temperature. *Synechococcus* and *Picochlorum atomus* are of similar enough cell size (~2μm) and may have similar nutrient uptake kinetics. Their positive association thus implies that they may avoid competing with each other by utilizing different nutrient forms.

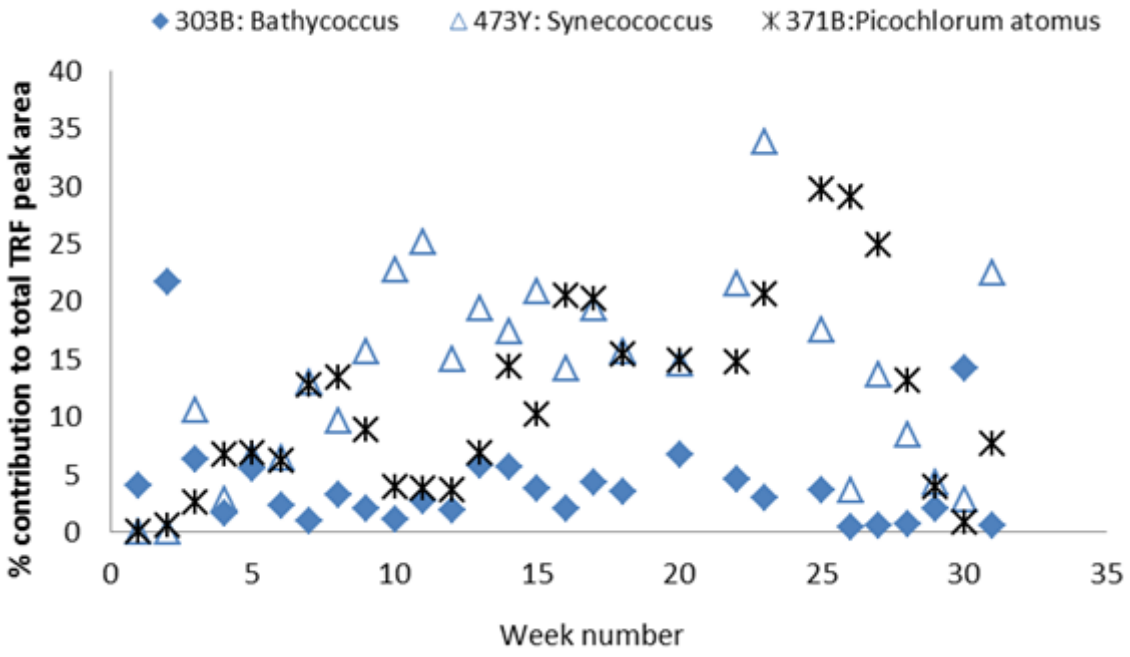


Fig. 4. 18. Temporal pattern of relative abundance of *Bathycoccus*, *Synechococcus* and *Picochlorum* represented by area contribution of their respective observed TRFs

Chaetoceros was different from other phytoplankton in that there were proportionally more positive associations between *Chaetoceros* and TRFs representing potential zooplankton grazers while there were proportionally more negative associations between *Chaetoceros* and other 18S TRFs representing phytoplankton and unidentified organisms (including other diatoms such as *Navicula* and *Cyclotella*) and unknown groups (Fig. 4.11. and Table 4.7.). The negative associations could simply mean that two organisms had different temporal pattern. Alternatively, if the negative associations between *Cheatoceros* and non-grazer organisms indeed suggest competition, the proportionally more negative associations with non-grazers for *Chaetoceros* means that *Chaetoceros* experienced more competition than other co-occurring phytoplankton species.

Table 4. 7. Positive and negative interactions of *Chaetoceros*, diatoms, chlorophytes and dinoflagelates with other 18S TRFs (internal interactions such as that between Chaeto1 and Chaeto2 are not counted)

	Planktonic grazer		Phytoplankton and unidentified	
	# of +	# of -	# of +	# of -
Chaetoceros	4	1	4	10
Cylindrotheca, Cyclotella, unidentified diatom, and *Navicula*	4	3	14	11
Chlorophytes (*Picochlorum and Bathycoccus*)	3	3	11	9
Autotrophic Dinoflagellates (Dino2 and Dino3)	1	0	5	4

LSA revealed several examples of closely related taxa that have distinctly different relationships with the rest of the plankton community. For example, phagotrophic grazers *Telonema antarcticum* and *Telonema subtilis* represented by observed TRFs 177B (Telo1) and 377B (Telo2) were negatively associated with different plankton (Fig. 4.14.B.). While *Telonema antarcticum* was negatively associated with *Flavobacteria* (Flavo2), *Telonema subtilis* was negatively associated with mixed phytoplankton (Mixed pp2). These together imply that different members of the genus *Telonema* may graze on different plankton with different cell sizes. Flavo1, Flavo2, Flavo3 and Flavo4 were also associated with different TRFs in the Bacteroidetes subnetwork (Fig. 4.12.). Flavo1 and Flavo2 did not have any association with each other or with Flavo3 and Flavo4. The only direct association among them was between Flavo3 and Flavo4, but with a negative sign. Furthermore, the same taxon represented by different TRFs can be associated with different environmental factors (Flavo2 and Flavo3, Flavo4) or the same environmental factor but in different ways (Flavo3 and Flavo4). Even organisms from the same genus may not share similar temporal patterns, suggesting that they may have different ecological functions.

Lastly, time delayed associations between organisms should be considered with care. For example, the negative association between TRF 37B and Syndi2 was captured only in the time delayed LSA (Fig. 4.16., upper panel), when a large part of the original data was removed. While being a useful tool in pointing out potentially important relationships between organisms during part of the sampling period, too much manipulation of the original data runs the risk of making this type of analysis a figure matching procedure, over-emphasizing chance patterns. Instead of simply embracing the conclusions indicated by time-delayed LSA, which could be meaningful

since relationship between TRF-TRF and TRF-EF pairs could be different during different seasons, it is also important that these results be carefully used to design laboratory experiments for testing explicit hypotheses that are field-based.

Chapter 5 Comparison of planktonic microbial community in Great South Bay under brown tide bloom and non-brown tide bloom conditions

Introduction

The goal of this chapter is to compare and synthesize observations made concerning the 18S and 16S planktonic communities in the brown tide year (2008) and the non-brown tide year

(2009) in Great South Bay. Major differences in the planktonic community composition between the two years will be examined, and spatial and temporal patterns of planktonic communities incorporating both years will also by analyzed with and without *Aureococcus anophagefferens*. Through these comparisons, this chapter aims to reveal if and how the brown tide species *A. anophagefferens* imposes any effect on the planktonic microbial community. In particular, the questions to be addressed in this chapter include:

1. How do the temporal patterns of some of the planktonic microbes identified in this work differ in years with and without brown tide blooms in GSB?

2. Does the spatiotemporal structure of microbial communities in a brown tide year differ from that in a non-brown tide year? If so, does the brown tide bloom contribute to the difference?

3. Is the biodiversity of 18S and 16S communities different in bloom and non-bloom years? In addition, if the relationship of the biodiversity between the two years is different for 18S and 16S communities, what feature of 18S and 16S communities might explain this difference?

4. What are the potential biotic and abiotic interactions found in GSB in both 2008 and 2009? What interactions are unique to each year that may reflect the effects of brown tide?

Materials and Methods

In order to compare and contrast the GSB planktonic communities during 2008 and 2009, 18S and 16S TRFLP profiles from May 8 to October 9 in 2008 and 2009 were compared (Table

5.1.). Due to the closeness of the sampling dates in the two years (differing by 2 days at the most), sampling dates in 2008 were used in text and figures except on dates when samples were collected only in 2009, in which case sampling dates in 2009 were used.

Table 5. 1. Sampling date, the corresponding week number, and samples compared between 2008 and 2009 in GSB

	Week number	Samples analyzed (18S)-2008	Samples analyzed (18S)-2009	Samples analyzed (16S)-2008	Samples analyzed (16S)-2009
8-May	1	A, E	E	A,C,D,E	E
13-May	2	A to E	A to E	A,C	A-C,E
20-May	3	A to J	A to F, J	A to J	A-E,G,J
27-May	4	A, E	A, E	A, E	A, E
4-Jun	5	E, F, J	A, E	A,E,F,J	A, J
10-Jun	6	A, E	A, E	A, E	A
17-Jun	7	A, E, F, J	A, E, J	A, E, F, J	A,E
24-Jun	8	A, E	A, E	A, E	A, E
1-Jul	9	A, E, F, J	A, E, J	A, E, F, J	A, E, J
8-Jul	10	A, E	A, E	A	A, E
15-Jul	11	A, E, F, J	A, E, J	A, E, F, J	A, E, J
22-Jul	12	A, E	A, E	A, E	A, E
29-Jul	13	A, E, F, J	B, E, J	A, E, F, J	E, J
5-Aug	14	A, E, F, J	A, E	A, E, F, J	E
12-Aug	15	A, E, F, J	A, J	A, E, F	A, J
20-Aug	16	A, E	NA	A, E	NA
26-Aug	17	A, E, F , J	A, E, J	E, F , J	E, J
4-Sep	18	A, E	NA	A, E	NA
11-Sep	19	NA	A, D, J	NA	D, J
18-Sep	20	A, E, F, J	NA	A, E, F	NA
25-Sep	21	NA	A, E, J	NA	A, E, J
2-Oct	22	A, E, F, J	NA	A, E, F, J	NA
9-Oct	23	A, E	A to J	A, E	A to J

Environmental parameters

Due to the close associations between river discharge and 7-day river discharge as well as between maximum gust speed and average wind speed in both years (Table 3.3. and Table 4.3.), only 7-day river discharge and average wind speed were graphed for both years. T tests were

performed to determine if there were differences between years for the 8 environmental parameters (for temperature and salinity, data for station E and J were compared separately). For temperature and salinity, to test if the environmental gradient along the sampling transect was larger in either year, the difference of these two parameters between the two stations was also compared using T tests.

Nonmetric multidimensional scaling (NMS) ordination

1. **Putting two years together**

The main matrixes used for NMS ordination in Chapter 3 and Chapter 4 were combined and manually aligned for both 18S and 16S communities. Since there was both instrumental and analytical (fragment binning is dependent on TRF composition of a data matrix) variation, TRFs generated from the same organisms could be assigned a different size at different times. To align data collected for the two years, major TRFs in both years were used as internal size markers and the minor TRFs were aligned accordingly. TRFs that were present in fewer than two samples (1.5% of total samples included) were removed (Treusch et al. 2009).

The second matrix was generated by pooling the original environmental data, including surface water temperature, salinity, Connetquot River discharge, 7-day river discharge, average wind speed and maximal gust speed. Square root transformation and relativization was performed on the quantitative parameters in the second matrix (temperature, salinity, river discharge, and wind speed) following the method described in Chapter 3.

NMS was run in autopilot mode and the solutions were evaluated by Monte Carlo tests. Visualization of the ordination was through 3D plots and biplots whose axes were extracted from the main matrix and did not necessarily represent gradients of any parameter. The coefficient of determination (R^2) was calculated to assess the percentage of variance in the original data matrix represented by each axis. The correlation coefficient (r) between the environmental parameters and the position of each sample on each ordination axis was calculated to reveal any linear relationship between the ordination axis and the environmental parameters.

2. Artificial removal of *A. anophagefferens*

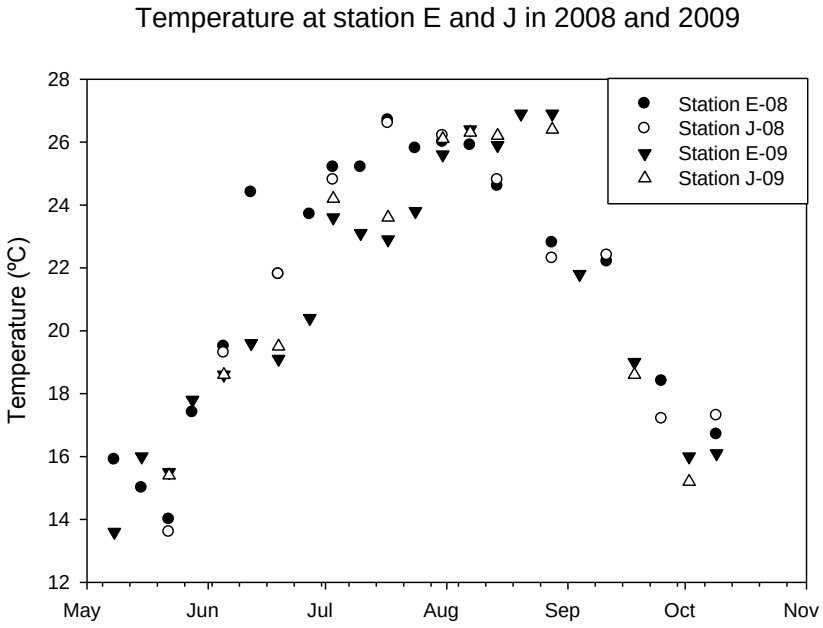
Because it was impossible to have sampled the microbial community without *A. anophagefferens in situ* in 2008, exclusion of *A. anophagefferens* in a post hoc way was performed by removing major TRFs c175B and c298Y to mimic a non-bloom situation, leaving the rest of the community unchanged. To do this, observed TRFs 174B and 176B as well as 297Y and 299Y were removed from the main matrix containing both years' data and the area contribution of the remaining TRFs to total peak area was recalculated as described in Chapter 3. The resulting main matrix was then analyzed using NMS ordination to examine whether the natural (2009) and the artificial (2008) non-brown tide communities look the same. This ordination was compared with the original ordination to see if the extent of separation between 2008 and 2009 samples was changed by the removal of *A. anophagefferens* TRFs.

Results

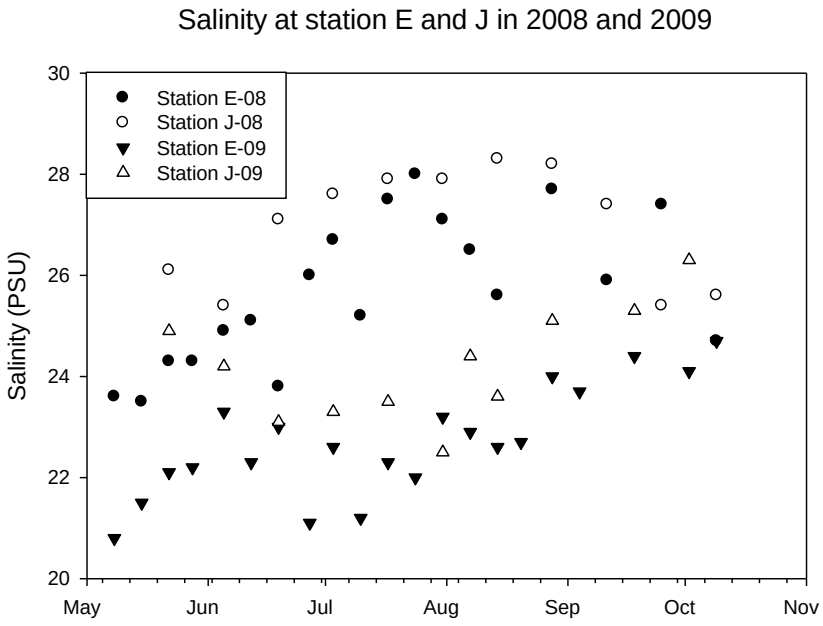
Environmental parameters

The average temperatures at stations E and J were compared between 2008 and 2009 (May to October) using T tests and no significant difference was found (results not shown). However, differences in water temperature at certain times of the year were apparent: for most of the sampling days between June 8 and July 24, water temperature was higher in 2008 than 2009 (Fig. 5.1.A.); water temperature was similar in late July, then higher in August 2009 than 2008. The highest temperature was comparable, but the peak value in 2008 occurred earlier (July 17) than in 2009 (August 28). Similar to temperature, although T tests found no significant difference in either precipitation (both on the sampling day and the 7-day accumulated) or the wind speed between the two years, 7-day precipitation in 2009 was generally higher than in 2008 before August and lower than in 2008 after August (Fig. 5.1.D.), and it was windier in 2008 from May 8 to July 24, after which it was mostly windier in 2009 (Fig. 5.1.F.). GSB was significantly saltier in 2008 than 2009, agreeing with the T test result that there was significantly more fresh water input from the Connetquot River in 2009 (Fig. 5.1.B.&E., Table 5.2.). However, when the salinity difference between station E and station J was compared between years, no significant difference was found (data not shown), suggesting that the salinity gradient across the bay was comparable in 2008 and 2009.

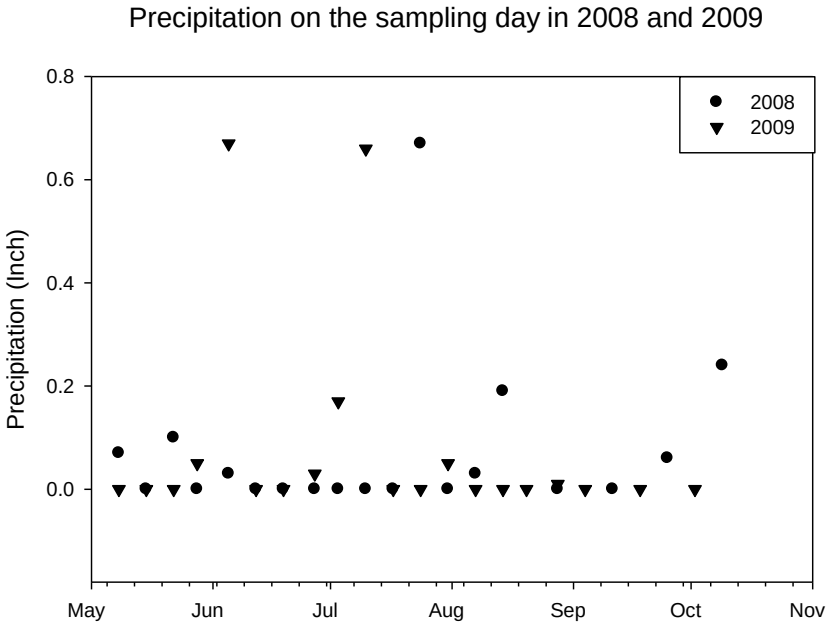
A)


Temperature at station E and J in 2008 and 2009
Temperature (ºC)
28
26
24
22
20
18
16
14
12
May
Jun
Jul
Aug
Sep
Oct
Nov
Station E-08
Station J-08
Station E-09
Station J-09

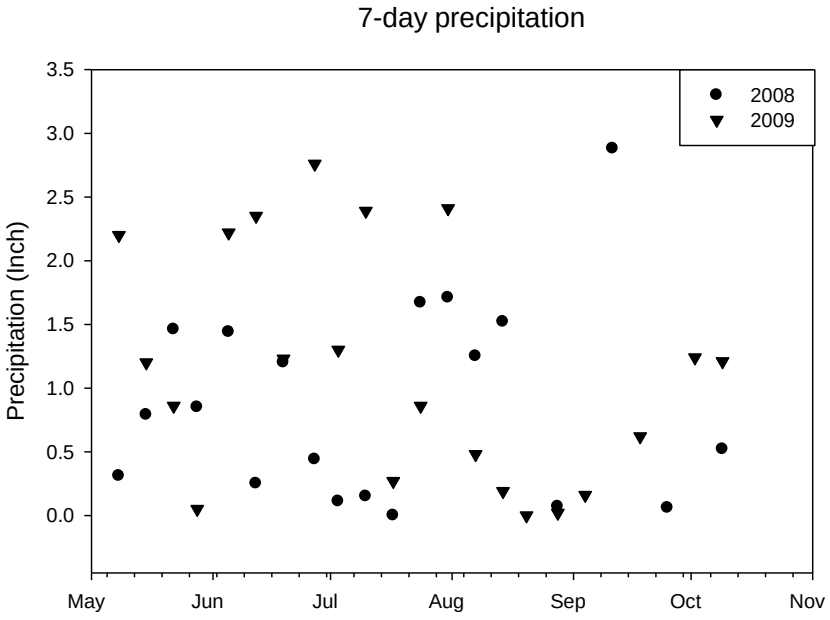
B)


Salinity at station E and J in 2008 and 2009
Salinity (PSU)
30
28
26
24
22
20
May
Jun
Jul
Aug
Sep
Oct
Nov
Station E-08
Station J-08
Station E-09
Station J-09

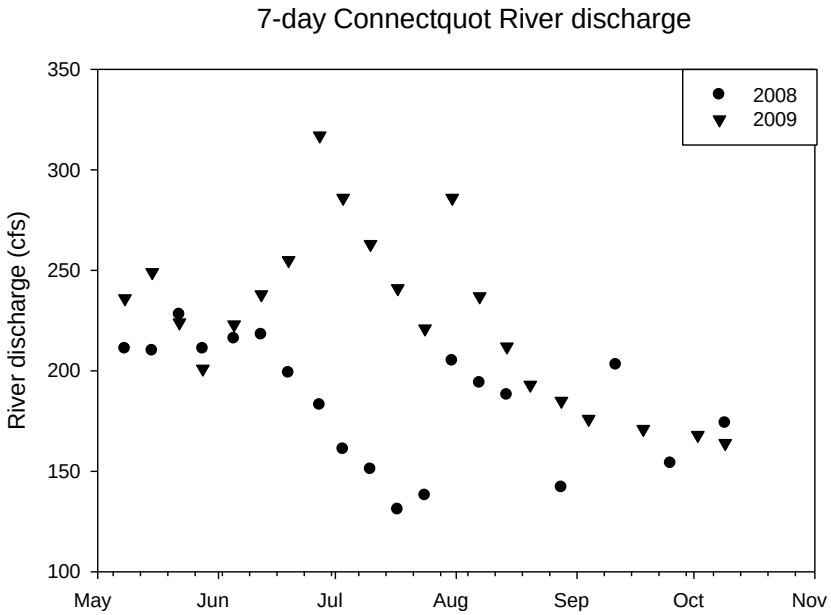
C)


Precipitation on the sampling day in 2008 and 2009
Precipitation (Inch)
0.8
0.6
0.4
0.2
0.0
2008
2009
May
Jun
Jul
Aug
Sep
Oct
Nov

D)


7-day precipitation
Precipitation (Inch)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2008
2009
May
Jun
Jul
Aug
Sep
Oct
Nov

E)



F)

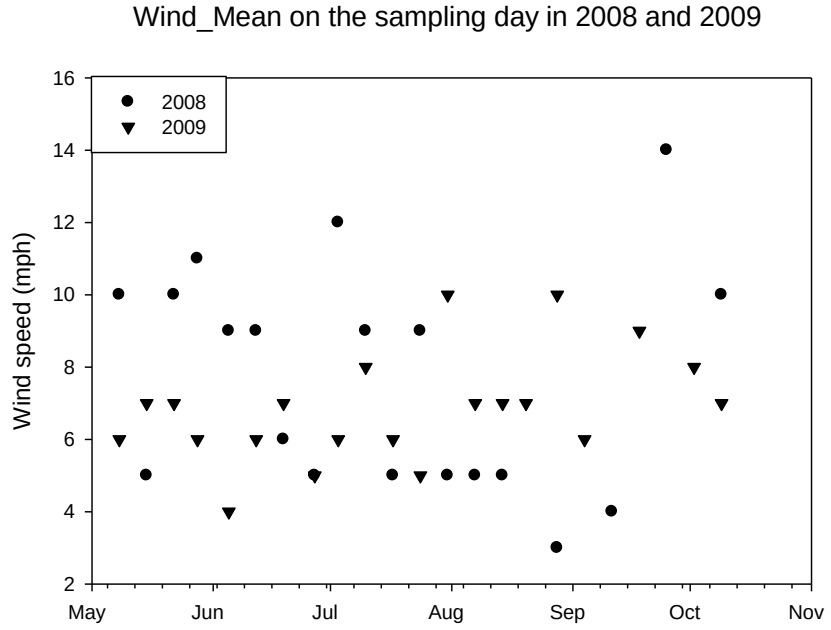


Fig. 5. 1. Environmental parameters versus time in GSB in 2008 and 2009 A) Temperature at station E and J B) Salinity at Station E and J C) Precipitation on the sampling day D) Sum of precipitation over the prior 7 days including the sampling day E) Sum of the Connetquot River discharge over the prior 7 days including the sampling day F) Average wind speed on the sampling day

Table 5. 2. T tests of salinity at station E, salinity at station J and 7-day Connetquot River discharge for 2008 and 2009. All other environmental parameters were not significantly different between 2008 and 2009.

	Salinity at E		Salinity at J		7-day river discharge	
	2008	2009	2008	2009	2008	2009
Mean	25.67	22.70	26.99	24.20	185.11	226.0
Observations	19	21	11	11	11	11
P(T<=t) one-tail	<0.001		<0.001		<0.001	

Major TRFs

1. 18S major TRFs

Unlike 2008, observed TRFs 174B and 175B were present in 2009 only as minor TRFs (average 0.87%) and the only cloned amplicons from 2009 producing a similar predicted TRF represent the copepod *Harpacticus*. With c175B missing as a major TRF in 2009, c258B and c264B became major TRFs (Fig. 5.2.). Major TRF c258B representing ciliates and mixed phytoplankton in 2009 corresponded with observed TRFs that were present as minor TRFs (average 3.67%) in 2008. The relative abundances of organisms contributing to c258B may have been different in the two years. For example, only 1 out of 20 sequences with predicted TRF size from 258B to 264B was a ciliate sequence in 2008 while 6 out of 21 sequences with predicted TRF size from 258B to 264B were ciliate sequences in 2009 (Table A1. and A3), indicating that more aloricate ciliates may have been present in 2009. Major TRF c264B representing Cercozoa in 2009 corresponded with observed TRFs (also representing Cercozoa) that were present as minor TRFs (average 2.73%) in 2008.

Major TRF c284B, which represented *Chaetoceros* in both years, had a sharp peak in contribution to total TRF area (c284B%) in mid-June (week 7) in both years (Fig. 5.3.). The secondary peak of c284B% in late July (~week 12) in 2008 at ~49%, however, was not repeated in 2009, although there was a small peak in late August (~week 16) at ~10%. Major TRF c316B was composed of several taxa that were similar in the two years and the temporal pattern of c316B% was similar both years (Fig. 5.3.) with its lowest contribution to the community in

spring and early summer (except during weeks 5 and 6 in early June, 2009). Major TRF c374B, although composed of similar taxa for the two years, demonstrated somewhat different temporal patterns, especially in May (weeks 1 and 2), late June (weeks 7 and 8) and late August (week 17) (Fig. 5.3.). The chlorophyte *Picochlorum* was an important contributor to major TRF c374B in both years while various dinoflagellates also contributed to c374B. For example, sequences from *Karlodinium micrum* were cloned in both years while sequences from *Polykrikos kofoidii* were cloned only in 2009 (Table A1). Major TRF c532B, comprising observed TRFs 530B and 533B, was contributed by both *Cyclotella* and *Minutocellus* in 2008 while almost completely by *Cyclotella* in 2009, according to both the ratio of 530B% over 533B% and the number of sequences of the two species (Table A1 and Table A3). No *Cyclotella* sequence was cloned in 2008, suggesting there might be some bias against its amplicon during cloning. c532B% peaked in May (~week 3) in both years though the peak value was about twice big in 2009 as in 2008 (25.5% vs 12.6%). Two smaller peaks in early August (~week 14) and mid-September (week 19) in 2008 were absent in 2009.

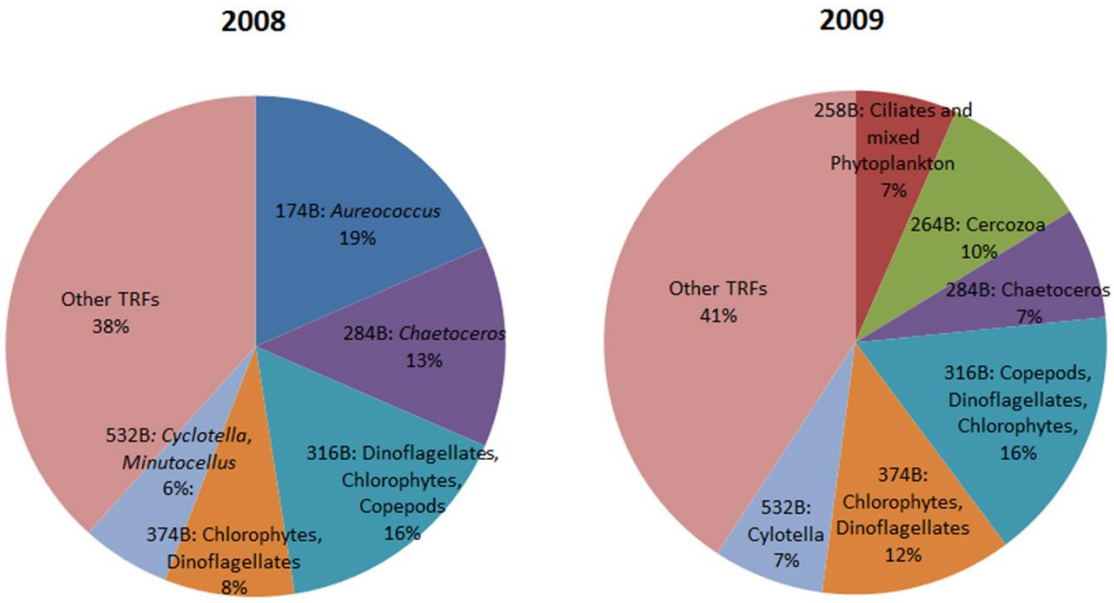


Fig. 5. 2. Comparison of major and other 18S TRFs' area contribution to total TRF peak area averaged over stations and sampling times in 2008 and 2009

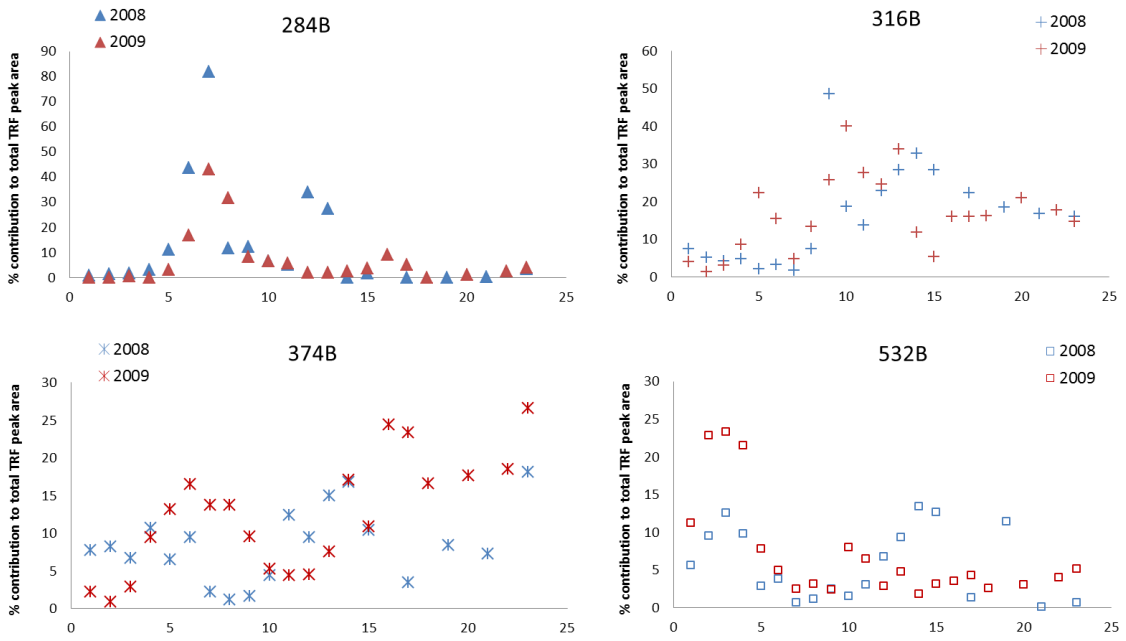


Fig. 5. 3. Contribution of the 4 common 18S major TRFs to total peak area averaged over stations each week in 2008 and 2009; X axis represents week number as in Table 5.1.

2. 16S major TRFs

Except for the absence of major TRF c298Y in 2009, the relative abundance of other major TRFs was almost identical in the two years (Fig. 5.4.). Observed TRF c298Y was present as a minor TRF in 2009 and matched the TRF predicted from a cloned uncultured Bacteroidetes 16S rDNA amplicon; c298Y% in 2009 was on average 0.22%, consistent with the low abundance of the corresponding Bacteroidetes in the clone libraries (1 out of 265 16S sequences).

Major TRF c289Y represented the alpha-proteobacterium *Marivita* in both years and did not show a strong seasonal pattern (Fig. 5.5.). Major TRF c475Y represented *Synechococcus* in both years, and c475Y% generally increased throughout the sampling season in both years, though the increase began later in 2008 than in 2009. *Picocystis salinarum* plastids also contributed to c475Y, although probably less than *Synechococcus*. This is because first, no *Picocystis* plastid sequence was cloned in 2009 (Table A4.) and second, *Rsa*I digestion of a 2008 sample from which both *Synechococcus* and *Picocystis salinarum* sequences were cloned suggested that *Picocystis* was ~ 1/7 the relative abundance of *Synechococcus* (Chapter 3). Major

TRF 555Y represented the alpha-proteobacterium SAR11 in both years. There was larger temporal variation of 555Y% in 2008, especially from May to late July (week 1 to week 13), inversely related to the temporal change of c298Y% in 2008 (Fig. 3.4.). Major TRF c563Y was contributed by the Actinobacterium *Microcella* in both years and had its greatest relative abundance in May, declining earlier in 2008 than 2009 (Fig. 5.5). Major TRF 577Y represented Bacteroidetes in both years, with a greater variety of taxa cloned and sequenced in 2009 than 2008 (Table A2. and Table A4.), although this may simply reflect that more sequences were retrieved in 2009 than in 2008 (265 vs 99). 577Y% was greatest in the summer, though it could vary more than 2-fold on consecutive weeks, sometimes coincident with changes in other major TRFs (such as the increase of 577Y% coincident with the decline of c298Y% in mid-June 2008).

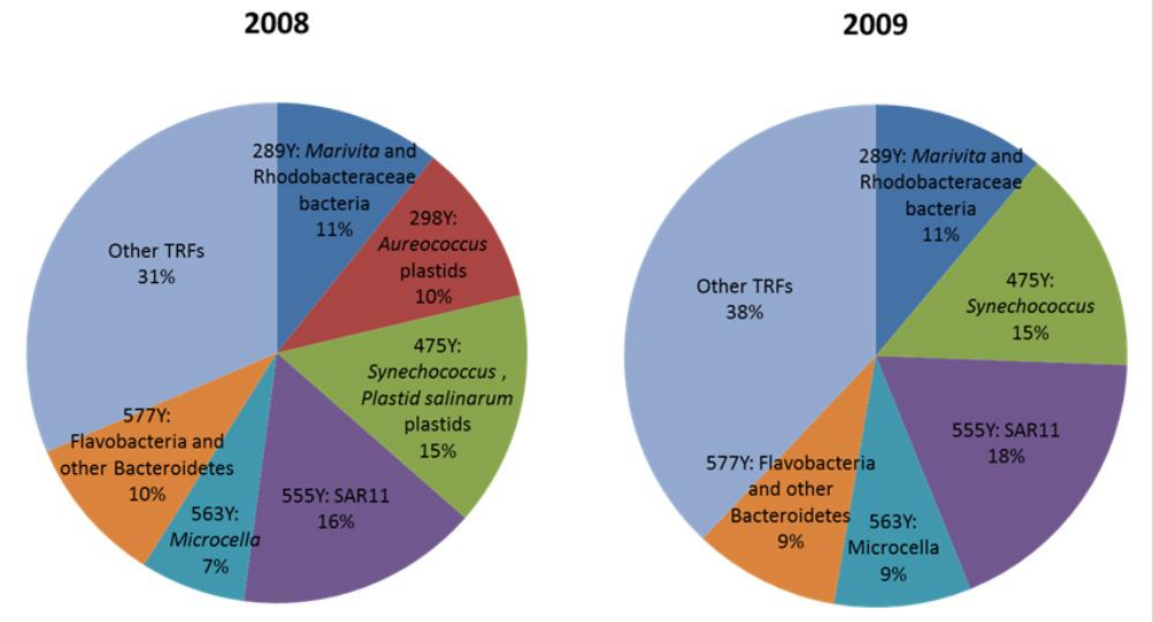


Fig. 5. 4. Comparison of major and other 16S TRFs' area contribution to total TRF peak area averaged over stations and sampling times in 2008 and 2009

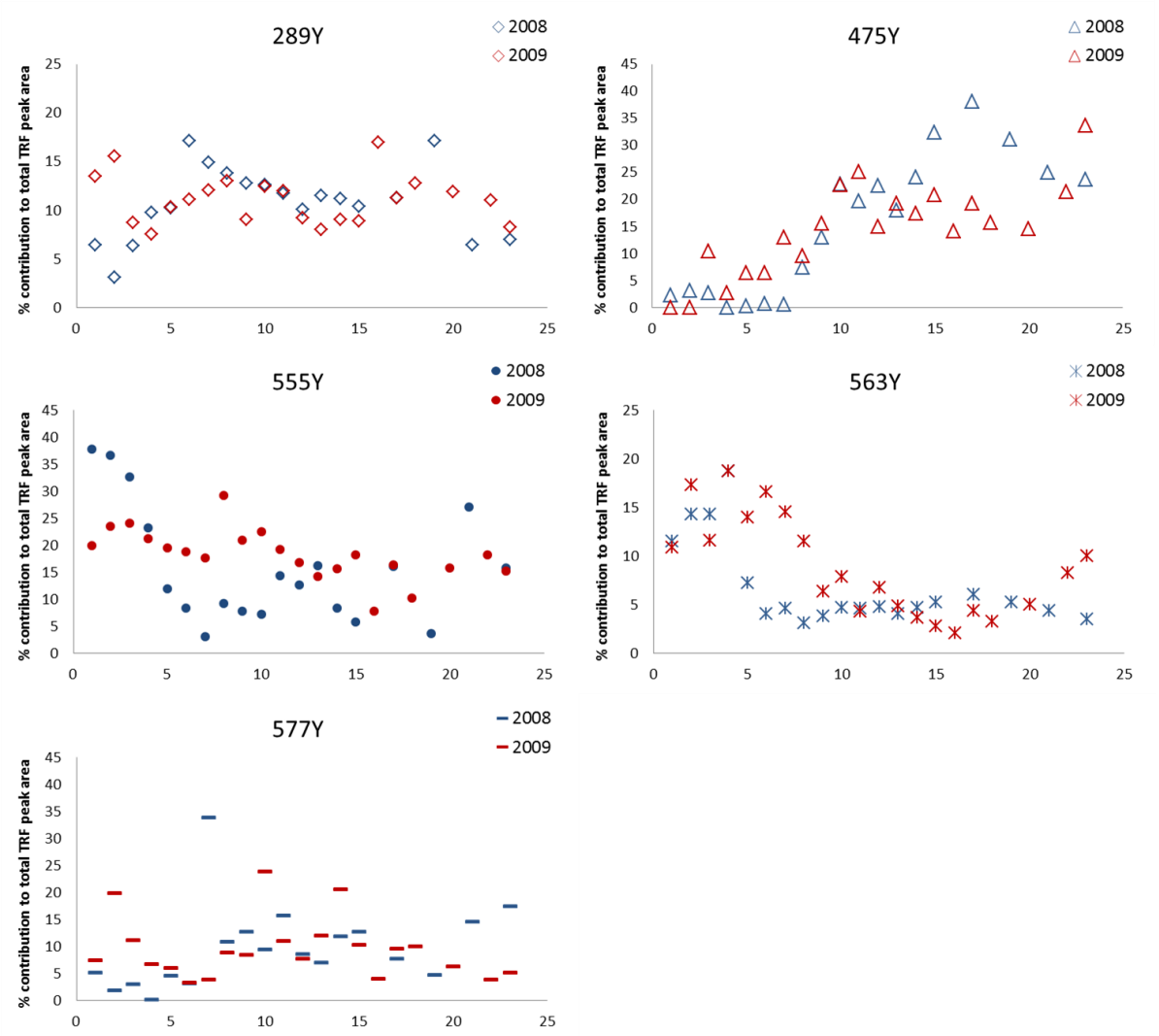


Fig. 5. 5**.** Contribution of the 5 common 16S major TRFs to total peak area averaged over stations each week in 2008 and 2009

Number of sequences retrieved

The emergence of major TRFs c258B (including ciliates) and c264B (Cercozoa) representing planktonic grazers in 2009 suggested there might be relatively more pelagic grazers in GSB when the brown tide bloom was not under way. According to the proportion of sequences retrieved (Table 5.3.), many grazers were indeed relatively more abundant in 2009 than 2008. The percentage increase was greatest for copepods and ciliates, whose sequences accounted for ~6-7 times more cloned amplicons in 2009 than 2008. In contrast, parasitic dinoflagellates Syndiniales and parasitic *Pirsonia* were more abundant in 2008, and were both negatively associated with *Chaetoceros* according to LSA in 2008 (Fig. 3.24.). Sequences of the rotifer *Brachionus calyciflorus* were only retrieved in 2008, and had only 1 significant negative association with TRF 479Y in LSA (Fig. 3.19.).

In 2008, proportionally more sequences were retrieved for the diatom *Minutocellus* and the pico-chlorophytes *Picocystis* and *Picochlorum* (Table 5.3.). In fact, for both *Minutocellus* and *Picocystis* (Table 5.3.), only 1 sequence was retrieved in 2009 out of 351 sequences in total. In contrast, proportionally more sequences of *Synechococcus* and *Cyclotella* were retrieved in 2009. Proportions of *Chaetoceros* and *Bathycoccus* sequences of the two years were comparable. Interestingly, although only the TRF representing *Chaetoceros* was a major TRF in both years, the proportions of *Chaetoceros* sequences were lower than those of *Bathycoccus* in both years. This inconsistency between TRFLP and sequencing data may reflect a bias against cloning *Chaetoceros* amplicons. For comparison between years, however, this type of bias should not matter.

Table 5. 3. Contribution of grazers, parasitoids, and phytoplankton sequences to total number of sequences retrieved each year. 256 and 351 18S partial rDNA sequences were retrieved in 2008 and 2009; 99 and 265 16S partial rDNA sequences were retrieved in 2008 and 2009.

Grazers and parasitoids (%)								
	Cercozoa	Ciliates	Copepod	*Pirsonia*	Polychaete	Syndiniales	*Telonema*	*Brachionus*
2008	6.6 (17)	2.3 (6)	3.5 (9)	2.3 (6)	3.1 (8)	12.9 (33)	0.4 (1)	6.0 (15)
2009	9.4 (33)	12.3 (43)	21.7 (76)	0.3 (1)	5.7 (20)	7.1 (25)	1.7 (60)	0.0 (0)
Phytoplankton (%)								
	Aureococcus	*Synechococcus**	*Minutocellus*	*Chaetoceros*	*Picochlorum*	*Picocystis*	*Bathycoccus***	*Cyclotella*
2008	5.1 (13)	5.1 (13)	5.1 (13)	1.6 (4)	7.4 (19)	3.9 (10)	2.7 (7)	0.4 (1)
2009	0 (0)	15.5 (54)	0.3 (1)	1.4 (5)	3.1 (10)	0.3 (1)	2.3 (8)	1.1 (4)

*: the only sequence tallied based on 16S rDNA libraries

**: This includes 1 *Ostreococcus* sequence from both years, which has the same predicted TRF length as *Bathycoccus*

Microbial community in the presence and absence of <u>A. anophagefferens</u>

1. A two-year view

81% of the variation among all 18S samples was represented by 3 axes, with axes 2 and 3 each accounting for about a third of the variation and axis 1 accounting for only ~ 13% (Table 5.4.). As when the two years were ordinated separately, sample points did not cluster by station. Samples also did not cluster by the week number, suggesting the community structure at the same time of the two years was different (figures not shown). In fact, there was discernible separation between 2008 and 2009 samples on axis 1 (Fig. 5.6.), which was significantly associated with salinity, river discharge (Table 5.8.), and TRFs representing *A. anophagefferens* (174B, r=0.475 and 176B, r=0.372). 174B and 176B were even more strongly associated with axis 2 (r=0.-729 and r=-0.520), which was the only axis significantly associated wind speed (Table 5.8.).

Table 5. 4. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 130-dimensional space (TRFs) for 18S community

Axis	Increment	Cumulative
1	0.126	0.126
2	0.366	0.492
3	0.320	0.813

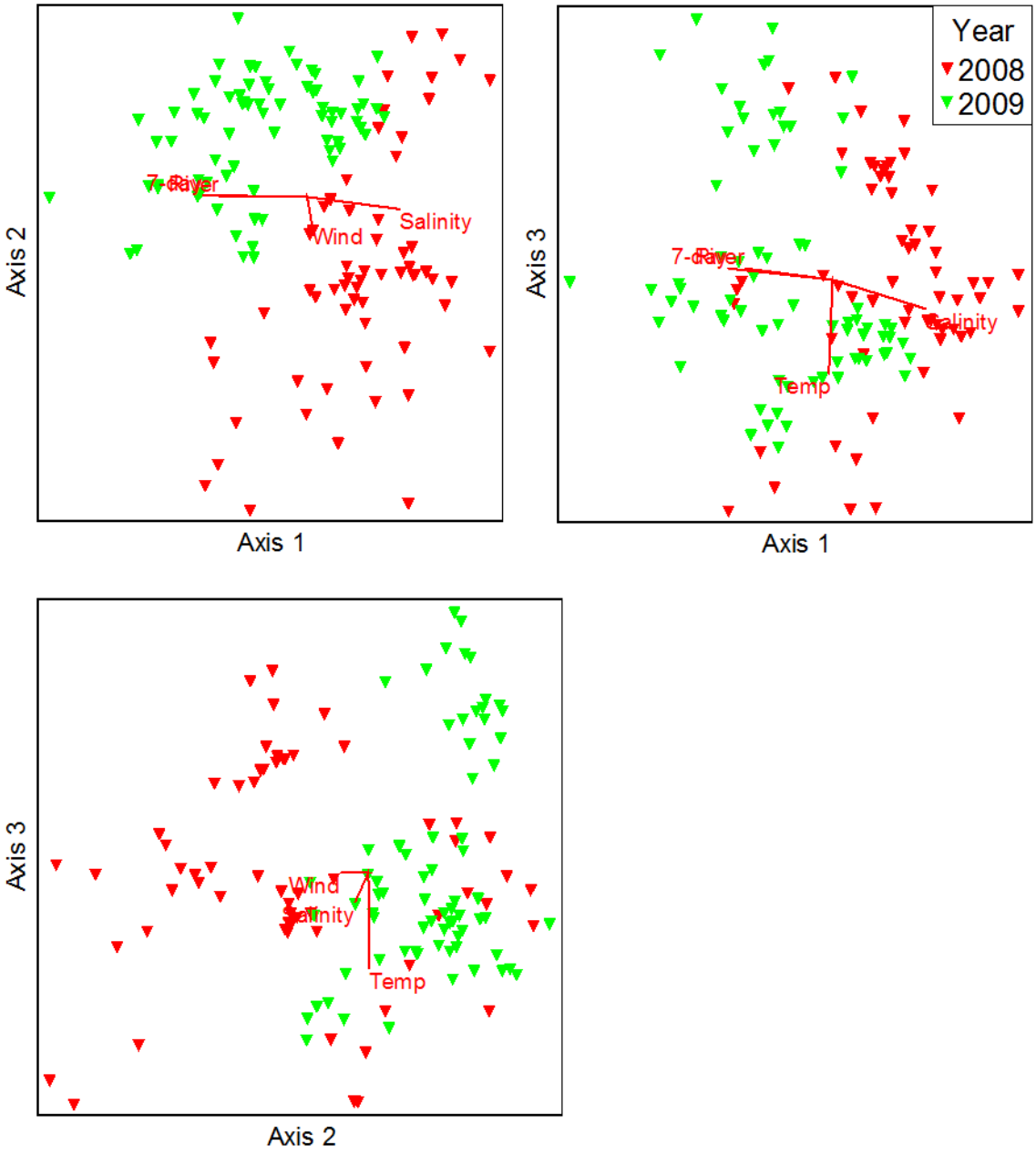


Fig. 5. 6. NMS ordination of 18S TRFLP from the two years

Only 2 axes were needed to represent 85% of the variation among samples in the combined 16S TRFLP, with axis 2 accounting for more of the variation (Table 5.5.). Separation between 2008 and 2009 samples was most clear along axis 2 (Fig. 5.7.), which was significantly correlated with salinity, river discharge and wind speed (Table 5.8.) and the relative abundance of 297Y (r=-0.627). The other major TRFs most strongly associated with axis 2 were 289Y (representing Rhodobacteraceae, r=0.551) and 563Y (representing *Microcella*, r=0.722).

Table 5. 5. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 78-dimensional space (TRFs) for 16S community

Axis	Increment	Cumulative
1	0.377	0.377
2	0.469	0.847

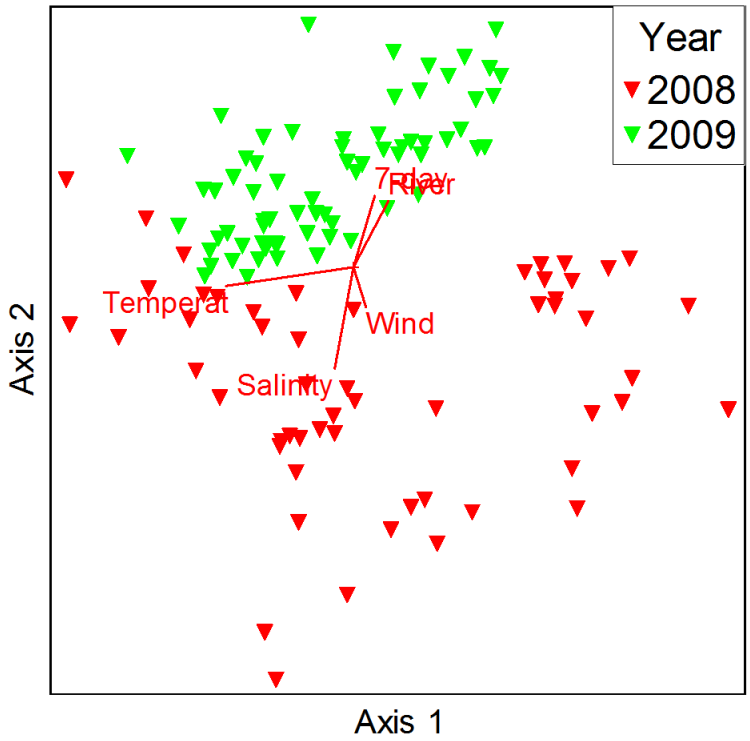


Fig. 5. 7. NMS ordination of 16S TRFLP from the two years.

2. A two year view with *A. anophagefferens* TRFs artificially removed

After the removal of *A. anophagefferens* TRFs, 3 axes represented 83% of the variation in the 18S community (Table 5.6.). Separation of samples from different years was not as clear

as before *A. anophagefferens* TRFs were removed (Fig. 5.6. and Fig. 5.8.). There was still some separation between years on axis 2 as demonstrated most clearly on the biplot axis 3 vs axis 2. TRFs most strongly associated with axis 2 was 284B and 285B, both representing *Chaetoceros calcitrans* (r=0.805 and r=0.642, respectively), which was probably the main reason why week 5 to 7 samples in 2008 (circled samples in Fig. 5.8.) were separated from rest of the samples. In these weeks in 2008, c285B% ranged mostly between 50% to ~100% after *A. anophagefferens* TRFs were removed, higher than c285B% of the *Chaetoceros calcitrans* dominant weeks (6, 7 and 8) in 2009 at 20% to 50%. Other TRFs that were also important in separating the two years included 304B and 305B, both representing *Bathycoccus* (r= -0.496 and r= -0.406, respectively), and 530B representing *Cyclotella*, which was reflected in the separation of May samples between years (samples in squares) when means of (304B%+305B%) and 530B% were both significantly higher in 2009 than in 2008 (T test, p=0.001). A striking difference after *A. anophagefferens* TRFs were removed was that wind speed was no longer associated with any axis (Table 5.8.).

Table 5. 6. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 128-dimensional space (TRFs) for 18S community in both years with 174B and 176B (*A. anophagefferens*) removed

Axis	Increment	Cumulative
1	0.192	0.192
2	0.312	0.504
3	0.326	0.830

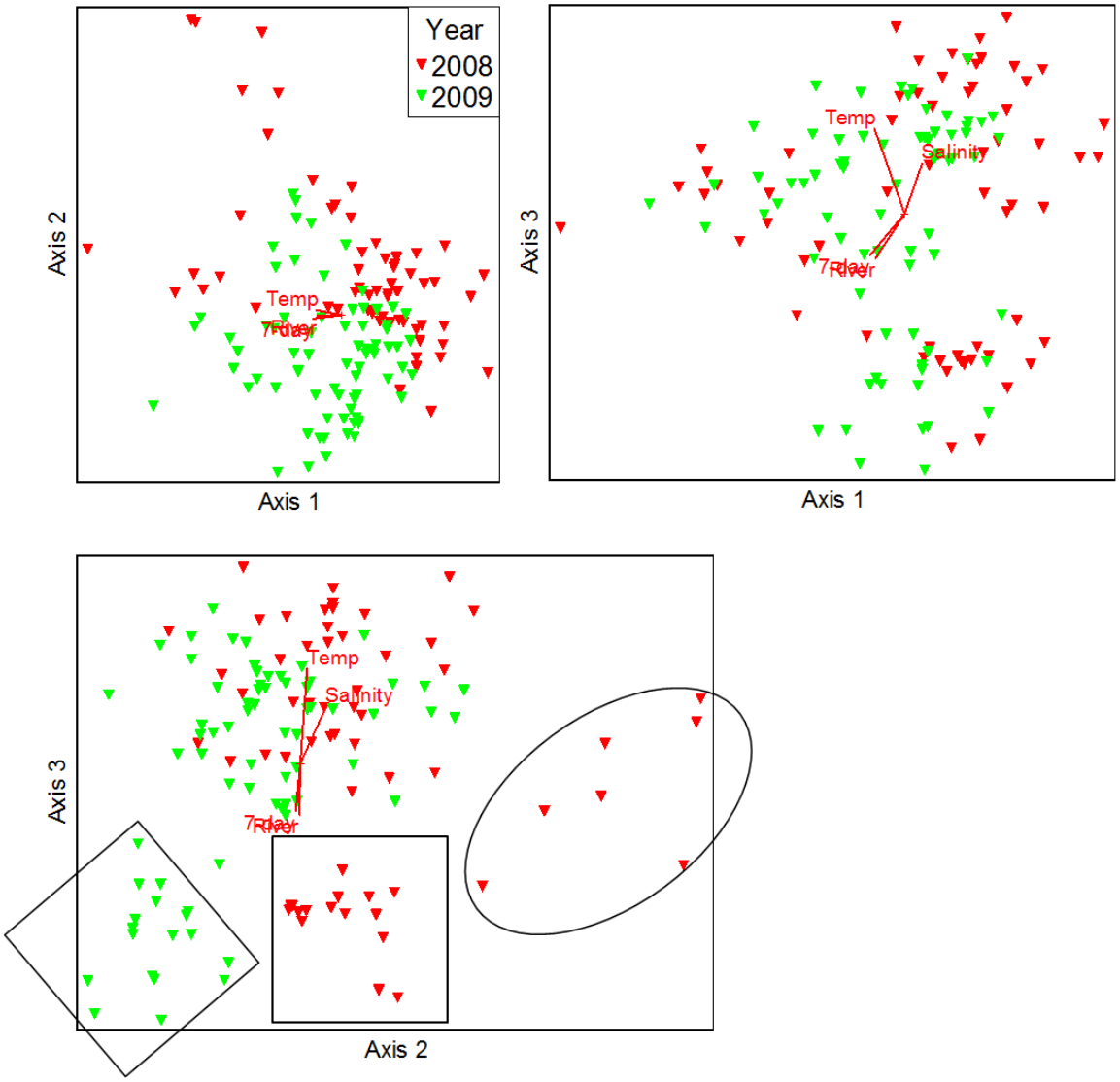


Fig. 5. 8. NMS ordination of 18S TRFLP without *A. anophagefferens* TRFs from the two years; each square indicates samples collected in weeks 1, 2 and 3, and circle indicates samples collected in weeks 5, 6 and 7.

After the *A. anophagefferens* TRF was removed from the combined 16S TRFLP, 3 axes were needed to represent 89% of the variation 16S community (Table 5.7.). Axis 3 was

associated with all environmental factors except for maximum gust speed (Table 5.8.) and this was also the axis along which 2008 samples were separated from 2009 (Fig. 5.9.), although the separation was probably less distinct without the *A. anophagefferens* TRF (Fig. 5.7.). The major TRFs significantly associated with axis 3 include 563Y (*Microcella*, r=-0.749) and 577Y (Bacteroidetes, r=0.636). Unlike the 18S community, the removal of *A. anophagefferens* plastid TRFs did not result in the disappearance of wind as a significant environmental factor.

Table 5. 7. Coefficients of determination (R^2) for the correlations between ordination distances and distances in the original 77-dimensional space (TRFs) for 16S community in both years with 298Y* (*A. anophagefferens* plastids) removed

Axis	Increment	Cumulative
1	0.384	0.384
2	0.206	0.590
3	0.300	0.890

*: Compared with the 16S TRFLP of 2008 and 2009 in part 1, only 1 fewer TRF (298Y) is included in this ordination step because observed 297Y and 299Y were already combined when aligning 16S TRFLP including two years in part 1.

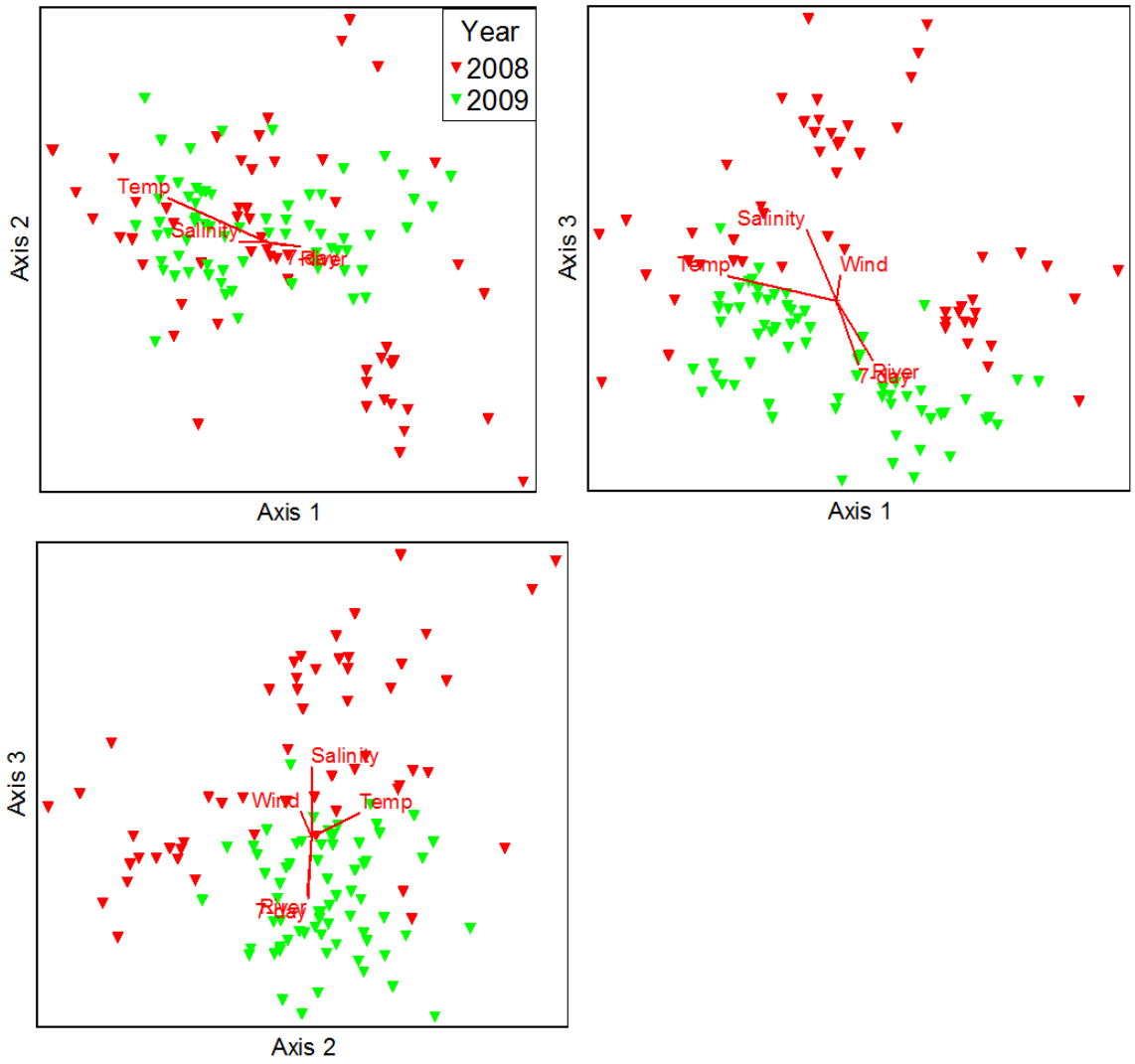


Fig. 5. 9 NMS ordination of 16S TRFLP without the *A. anophagefferens* TRF from the two years

Table 5. 8. Pearson's Correlation Coefficients (r) between the environmental parameters and the ordination axes for 18S TRFLP of both years, 18S TRFLP of both years without *A. anophagefferens* TRFs, 16S TRFLP of both years and 16S TRFLP of both years without the *A. anophagefferens* TRFs; Numbers in bold represent significant r values ($p<0.001$)

Axis	18S-2 yrs (N=133)			18S-2 yrs-Artificial* (N=133)			16S-2 yrs (N=123)		16S-2 yrs-Artificial* (N=123)		
Env	1	2	3	1	2	3	1	2	1	2	3
Temp.	-0.120	0.054	**-0.618**	**-0.334**	0.151	**0.557**	**-0.621**	-0.241	**-0.646**	**0.432**	**0.308**
Salinity	**0.615**	-0.226	**-0.342**	0.251	**0.290**	**0.430**	-0.239	**-0.552**	**-0.340**	0.003	**0.525**
Wind	0.140	**-0.331**	-0.015	-0.076	0.061	0.095	0.193	**-0.347**	0.117	-0.203	**0.310**
Gust	-0.040	-0.150	0.012	-0.143	-0.075	0.016	0.139	-0.164	0.066	-0.210	0.113
River	**-0.594**	0.021	0.216	**-0.328**	-0.046	**-0.406**	**0.326**	**0.449**	**0.375**	-0.130	**-0.478**
7-day	**-0.652**	0.088	0.214	**-0.355**	-0.118	**-0.392**	0.253	**0.465**	**0.289**	-0.121	**-0.494**

Env: Environmental parameters
Temp: Surface water temperature
Wind: average wind speed on the sampling day
Gust: maximal gust speed on the sampling day
River: water discharge from the Connetquot River on the sampling day
7-day: water discharge from the Connetquot River of the prior 7 days including the sampling day
*: *A. anophagefferens* TRFs were removed from the main data matrix

Biodiversity

Richness (N), Shannon's H', Simpson's D, and Pielou's J for samples collected from May 8 to October 9 in 2008 and 2009 were compared using T tests. 18S TRF richness (N) in 2009 was significantly higher than that in 2008 (Table 5.9.). The evenness index Pielou's J, however, was not significantly different between the two years. Diversity indexes Shannon's H' and Simpson's D, which are influenced by both richness and evenness, were both significantly higher in 2009 than in 2008 at $p<0.05$ (Table 5.9.). Contrary to the 18S TRFLP, 16S TRF richness, Shannon's H' and Simpson's D were significantly lower in 2009 than in 2008, while Pielou's J for 16S TRFLP was significantly higher in 2009 than in 2008 (Table 5.10.).

Table 5. 9. T tests of Richness (N), Shannon's H', Simpson's D and Pielou's J of 18S TRFLP for 2008 and 2009

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	2008	2009	2008	2009	2008	2009	2008	2009
Range	5~21	10~22	1.153 ~2.907	1.353 ~2.770	0.533 ~0.938	0.510 ~0.921	0.643 ~0.954	0.527 ~0.953
Mean	12.50	14.93	2.12	2.3	0.825	0.852	0.849	0.856
Observations	56	72	56	72	56	72	56	72
P(T<=t) one-tail	<0.001		=0.002		=0.018		0.285	

Table 5. 10. T tests of Richness's (N), Shannon's H', Simpson's D and Pielou's J of 16S TRFLP for 2008 and 2009

	Richness N		Shannon's H'		Simpson's D		Pielou's J	
	2008	2009	2008	2009	2008	2009	2008	2009
Range	7~15	3~12	1.668 ~2.235	1.030 ~2.252	0.752 ~0.882	0.619 ~0.880	0.771 ~0.929	0.838 ~0.990
Mean	11.2	7.671	2.07	1.83	0.832	0.811	0.862	0.92
Observations	51	70	51	70	51	70	51	70
P(T<=t) one-tail	<0.001		<0.001		=0.007		<0.001	

Potential interactions revealed by LSA

With a longer time series dataset used for 2009, relatively fewer significantly associated TRF pairs were observed for both 18S and 16S TRFs than in 2008 (Table 5.11.). Furthermore, this difference was much larger in 18S TRFLP (17% vs 7%) than in 16 TRFLP (18% vs 13%). Of the proportionally fewer significant associations between 18S TRF pairs identified in 2009, 37% were negative associations, which was greater than in 2008 (14%). On the other hand, only a slightly greater proportion of significant associations between 16S TRFs was negative in 2009 than 2008.

Table 5. 11. Percentage of significant associations between all possible TRF-TRF pairs and percentage of negative associations of all significant associations in 2008 and 2009

	18S-2008	18S-2009	16S-2008	16S-2009
significant associations	17%	7%	18%	13%
significant negative associations	14%	37%	38%	43%

In 2008, all the associations between *Synechococcus*, *Picochlorum*, *Bathycoccus* and *Picocystis* (blue-green picophytoplankton) were positive and *A. anophagefferens* was negatively associated with all of them (Fig. 3.15. and Fig. 3.17.B.). In contrast, in 2009 *Bathycoccus* was negatively associated with both *Synechococcus* and *Picochlorum* (Fig. 4.12.). Blue-green picophytoplankton were negatively associated with *Cyclotella* in both years (Fig. 3.17. and Fig. 4.12.) and positively associated with *Cylindrotheca* in both years (Fig. A1. and Fig. 4.12.). *Chaetoceros* was unique among the diatoms because it always had negative associations with other diatoms in both years (Fig. 3.16. and Fig. 4.10.), while associations among other diatoms were always positive (Fig. 3.17.A and Fig. 4.13.A). *A. anophagefferens* was negatively associated with *Cyclotella*, blue-green picophytoplankton and *Minutocellus* (Fig. 3.15.). The group 'autotrophic dinoflagellates' was positively associated with *Cyclotella* in both years and negatively associated with blue-green picophytoplankton in both years (Fig. 3.18. and Fig. 4.13.B.).

Associations between grazers/parasitoids and phytoplankton were different between the two years; pairs of TRFs associated in 2008 generally did not show any relationship in 2009. This was partly caused by a compositional change of zooplankton. For example, the TRF representing the potentially important diatom parasite *Pirsonia* in 2008 was below detection in

2009 (although one sequence was retrieved). The negative association between polychaetes and *Picochlorum atomus* was captured in both years, although interestingly, instead of *Clymenura* in 2008 (Fig. 3.17.B.), the polychaete species *Polygordius jouinae* was associated with *Picochlorum atomus* in 2009 (Fig. 4.12.).

Both *Chaetoceros* and *Synechococcus* were positively associated with water temperature in both years. *A. anophagefferens* and *Cyclotella*, on the other hand, were negatively associated with temperature. Among major phytoplankton identified here, blue-green picophytoplankton had the most associations with river discharge. *Picochlorum* was negatively associated with Connetquot River discharge in both years while for *Bathycoccus*, a negative association was found only in 2008 (Fig. 3.17. and Fig. 4.12.). With 4 weeks delay, *Picocystis* was negatively associated with Connetquot River discharge in 2008 (Fig. 3.21.). Wind speed, however, did not have common associations between years, which was partly due to the absence of TRFs representing *A. anophagefferens* and *Minutocellus polymorphus* in 2009, the two organisms significantly associated with wind speed in 2008.

Discussion

Planktonic grazers and parasitoids in GSB

Ciliates, copepods, Cercozoa, polychaete larvae (see below), Syndiniales, *Telonema*, and *Pirsonia* were common planktonic grazers/parasites in both years in GSB. Compared to 2008, many planktonic grazers were relatively more abundant in 2009, reflected not only in the difference of area contribution of TRFs representing grazers, but also in the higher proportion of grazer sequences retrieved in 2009 (Table 5.3.).

Both tintinnids (e.g. *Codonellopsis nipponica*) and aloricate ciliates (e.g. *Strombidium*) were found in GSB and sequence data suggested that aloricate species were more dominant (26 aloricate vs 9 tintinnids sequences). This agrees with Lonsdale et al. (1996) and Caron et al. (1989), who found previously that aloricate ciliates were the most abundant ciliate group in GSB. Amplicons representing copepod *Oithona* were cloned and sequenced in 2008 and 2009 (8 vs 55; TableA3.) while no sequence was retrieved from the reported dominant copepod species in GSB, *Acartia hudsonica* and *Acartia tonsa* (Duguay et al. 1989). With *Acartia tonsa*'s 18S rDNA sequence available in GenBank, matching of the sequences captured in this study with *Oithona* sequences in GenBank may indeed mean that the current dominant copepod species is *Oithona* in GSB. Alternatively, there might be some bias involved in the sampling process that prevented *Acartia* species from being caught or maybe there was PCR or cloning bias against *Acartia* as well. One minor TRF, 122B, represented exclusively ciliates (most likely *Strombidium*). During the spring-summer brown tide in 2008 and the following 3 weeks, the relative abundance of this ciliate was almost never as high as during the same period in the subsequent non-brown-tide year (Fig. 5.10.), suggesting that the brown tide bloom may have had a negative impact on this ciliate. Note that this potential negative association was not captured in LSA in 2008, which emphasizes the point that although some negative associations are reflected as opposite temporal trends of two organisms, some other negative associations may be indicated by year to year differences.

Cercozoa were an important group in both years and were even responsible for a major TRF in 2009. However, like elsewhere, cercozoan grazing activity, and even their presence, in GSB waters is very little studied. This is mainly because the diversity and importance of Cercozoa is only beginning to be appreciated through molecular evidence (Chantangsi et al. 2008) and much of the focus is still on establishing monocultures of these amoeboflagellates

(Glücksman 2011). Planktonic polychaete larvae were reported to be present in GSB in 1988 (Castro & Cowen 1991), and Lonsdale et al. (1996) also identified polychaete larvae as one of the most common microzooplankton taxa in GSB during 1991. These observations are consistent with my finding that meroplanktonic polychaete larvae are common in GSB. This study revealed many different polychaete species in GSB (Table A1. & A3.) with many present in both years. In Narragansett Bay, the abundance of meroplanktonic polychaete larvae was negatively correlated with brown tide concentration (Smayda & Fofonoff 1989). My data showed that although there was no association between polychaetes and *A. anophagefferens* (Fig. 3.15.), negative association between polychaetes and picoalgae *Picochlorum atomus* existed in both years, suggesting potential grazing effect of polychaete larvae on picophytoplankton. Polychaetes are reported to be able to effectively remove picoplankton/bacteria from seawater (Jordana et al. 2001, Licciano et al. 2005). Grazing studies of planktonic polychaetes (larvae) on *A. anophagefferens* and *Picochloru*, to my knowledge, has not yet been done for GSB. Such information should be made available because it will shed light on an potentially important trophic interaction related to brown tide bloom development/maintenance.

Sequences of parasitic dinoflagellates belonging to the order Syndiniales (Hoek et al. 1994, Stentiford & Shields 2005, Guillou et al. 2008) and parasitic *Pirsonia* (Kuhn 1997, 1998, Kuhn et al. 2004) were retrieved in both years (Table 5.3.). The near synchronization of 238B% (representing exclusively Syndiniales), with two peaks in August and September in both years (Fig. 5.10.), suggested that Syndiniales was probably not affected by the presence of brown tide bloom. This is consistent with LSA in 2008 where the central node representing Syndiniales was not associated with *A. anophagefferens* (Fig. 3.24.).

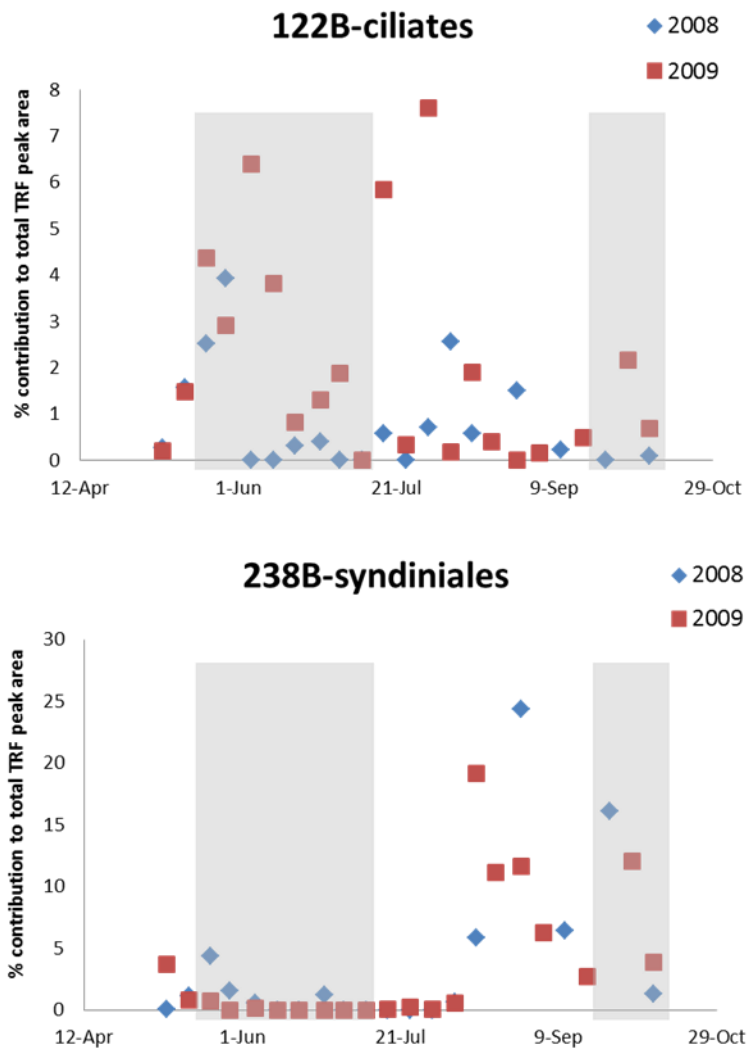


Fig. 5. 10. 122B% and 238B% averaged over all stations in each week in 2008 and 2009. Shaded areas represent bloom period in 2008 according to c175B%.

Picophytoplankton in GSB and their relationship with A. anophagefferens

A reciprocal temporal pattern of *A. anophagefferens* and *Synechococcus* existed in 2008, with c475Y% starting to increase in late June and being dominant for the rest of the sampling season (Fig. 5.11.). Dominance of *Synechococcus* in summer has been reported previously for Long Island bays (Sieracki et al. 1999, Sieracki 2001, Sieracki et al. 2004, Gobler et al. 2004a, Gobler et al. 2004b) and according to this study, the temporal pattern of *Synechococcus* was generally similar between years, although c475% started to increase about a month earlier in 2009 than 2008 and the peak of c475Y% around 40% in late August, 2008 was absent in 2009 (Fig. 5.11.). In addition to *Synechococcus*, *Picocystis* plastids also contributed to c475Y in 2008 (probably less so in 2009 since only one *Picocystis* sequence was retrieved vs 10 in 2008), and it is possible that *Picocystis* may have helped build up the c475Y% peak in August 2008. The picoalgal niche hypothesis suggests it is only when *Synechococcus* is selectively removed that *A. anophagefferens* has the window of opportunity to take over the picoalgal niche and bloom (Sieracki et al. 1999, Sieracki 2001). Earlier timing of increase in c175B% than that in c475Y% in 2008 seemed to suggest the opposite: in the brown-tide year, *A. anophagefferens* was the first to establish its dominance in picoalgal niche. It was not until July (probably because the water temperature was too high for *A. anophagefferens*) that *Synechococcus* established its dominance in the picoalgal niche in 2008. While c475Y was dominant from mid-summer until fall, the relative abundance of *A. anophagefferens* was low. There was a secondary *A. anophagefferens* bloom in October, 2008 (SCDHS 2011), which could have been supported in part by the lower water temperature during October that was optimal (Cosper et al. 1989) for *A. anophagefferens*. Selective removal of *Synechococcus* could also have made the picoalgal niche open for *A. anophagefferens*, which was suggested by the potential grazing effect of *Telonema* on *Synechococcus*, indicated by the negative association between *Telonema* on *Synechococcus* in 2008. Such negative association between *Telonema* and *Synechococcus* was, however, not observed in 2009.

Observed TRF 304B matched the TRF predicted for *Bathycoccus* (Table A1.) and had average contributions to total TRF area of 1.8% and 4.4% in 2008 and 2009, respectively. With no brown tide bloom in 2009, the similarly sized *Bathycoccus* became more relatively abundant compared to 2008, especially in May and June when 304B% was generally higher and achieved a peak of 21.6% (Fig. 5.11.). The presence of *Bathycoccus* (possibly *Bathycoccus prasinos*) has

not been reported in GSB previously. In June 2001, a picoplankton bloom lasting for less than 2 weeks was detected by flow cytometry in West Neck Bay, Peconic Bay, Long Island, and was identified as an *Ostreococcus*-like alga by epifluorescence microscopy and transmission electron microscopy (O'Kelly et al. 2003). This was the first time that pico-eukaryotes of the family Mamiellales were reported in Long Island waters. My data demonstrated that *Bathycoccus prasinos* (with occasionally sequenced *Ostreococcus*) was present in GSB in both years and was probably more abundant in the non-brown tide year, suggesting not only are species of Mamiellales widely distributed in Long Island bays, but also they may be negatively affected by the brown tide bloom, which is also supported by LSA (Fig. 3.17.). In this sense, *Bathycoccus* may be the picoalgae that compete with *A. anophagefferens* in the picoalgal niche in spring.

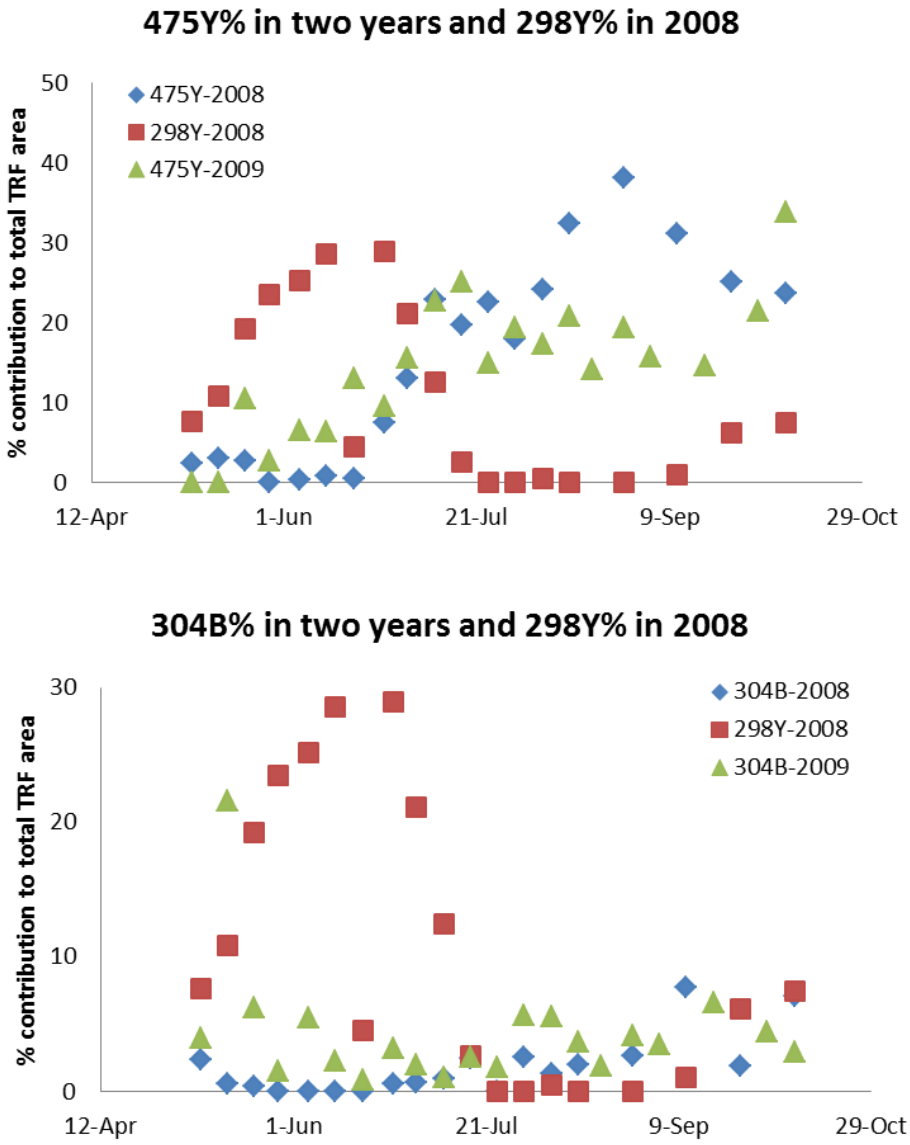


Fig. 5. 11. c475Y% and 304B% averaged over all stations in each week in 2008 and 2009. c298Y% in 2008 is also graphed for comparison.

Effect of brown tide bloom on the rest of the plankton community

1. Biodiversity

In 2008, for the 18S community, the non-brown tide bloom samples had significantly higher biodiversity indices compared with brown tide bloom samples in the same year, while no effect of brown tide bloom on 16S community biodiversity indices was observed (Table 3.8. and Table 3.9.). This, and the significantly higher 18S TRF richness, Shannon's H and Simpson's D in 2009 than 2008 agreed with a global biodiversity meta-analysis that phytoplankton and zooplankton are least diverse during massive blooms (Irigoien et al. 2004). In contrast, 16S community diversity was higher in the presence of brown tide blooms in 2008. Although differences other than the presence of brown tide in 2008 may have contributed to the higher biodiversity of the 16S community than in 2009, the opposite relationships of the biodiversity indices between the years for 16S and 18S communities are still striking.

2. Separation of microbial communitiesbetween years

NMS ordination including data from both years demonstrated that there were discernible differences in microbial community structure between the two years and that the effect of *A. anophagefferens* itself on 18S TRFLP was more apparent than that on 16S TRFLP (Fig. 5.6., Fig. 5.7., Fig. 5.8. and Fig. 5.9.).

For 18S TRFLP, separation between 2008 samples and 2009 samples became less clear with *A. anophagefferens* TRFs removed (Fig. 5.6. and Fig. 5.8.), suggesting that the main difference between years was the presence of *A. anophagefferens* in 2008. This agrees with the observation that *A. anophagefferens* TRFs were the only major TRF strongly associated with the axis separating the two years. Concurrent with the removal of *A. anophagefferens* TRFs, wind speed was no longer associated with any of the ordination axes, reemphasizing the correlation observed between *A. anophagefferens* and wind speed (Fig. 3.6. and Fig. 3.15). There was still some separation of May samples between years, which was most likely to have been caused by the different relative abundance of *Bathycoccus* and *Cyclotella* in May in the two years, suggesting relative abundance of *Bathycoccus* and *Cyclotella* also contributed to the separation of the two years, besides *A. anophagefferens*. On the other hand, separation of *Chaetoceros* dominant samples (weeks 5, 6, 7) in 2008 after removing *A. anophagefferens* TRFs, from the

other samples (Fig. 5.8.) may indicate the possibility of both *Chaetoceros*' contribution to separation and a mathematical effect since c285B% in these samples were ~50-100%, making them distinguishable from others.

For 16S TRFLP, separation between 2008 and 2009 was observed, with or without the removal of *A. anophagefferens* TRFs. In fact, the same set of environmental factors (temperature, salinity and wind speed) aligned with the separation between years with or without the removal of *A. anophagefferens* TRFs (Fig. 5.7.and Fig. 5.9.). Unlike 18S community, wind speed was still important after *A. anophagefferens* TRFs were removed in 16S community. In fact, wind speed was associated with other 16S TRFs besides *A. anophagefferens.* Flavobacteria (Flavo2), for example, was significantly associated with speed, according to LSA (Fig. 4.11), although Flavo2 was not associated with the axis separating the two years. Before *A. anophagefferens* TRFs were removed, major TRFs representing *A. anophagefferens,* Rhodobacteraceae and *Microcella* were all strongly associated with the axis separating the two years. After *A. anophagefferens* TRFs were removed, observed TRFs representing *Microcella* (563Y) and Bacteroidetes (577Y) were significantly associated with the axis separating the two years. This at least suggests that *Microcella* was also different between the two years, with or without *A. anophagefferens.* The difference of *Microcella*'s temporal patterns between years, however, did not seem to be related to the brown tide, according to the major TRF analysis and LSA (Fig. 3.4., Fig 4.4. and Fig 3.15.).

To summarize, *A. anophagefferens* was the main factor separating the two years for the 18S community while *Bathycoccus* and *Cyclotella* were also responsible for the separation of the two years at least during May. *A. anophagefferens* also contributed to the separation of the two years for the 16S community, but because other organisms also contributed to the separation, there was still separation between the two years after removing *A. anophagefferens* TRFs.

18S community versus 16S community

Simultaneous profiling of 18S and 16S communities revealed some differences between 18S TRFLP and 16S TRFLP, which all suggest that the16S community is more compositionally stable and less likely to be influenced by the brown tide bloom (although compared to 18S community, the better separation between years after *A. anophagefferens* TRFs were moved does seem to suggest 16S community's composition was more different between years). It is already

discussed previously that biodiversity of 16S community was less affected by the brown tide bloom. Two other perspectives demonstrating greater stability of 16S TRFLP will be discussed in this part.

1. **Sum of major TRFs' area contribution**

Sum of major TRFs' peak area (SA) varied more for the 18S TRFLP than for 16S TRFLP in each year (Fig. 5.12.). F tests confirmed that the variance of SA for 18S major TRFs was significantly larger than 16S major TRFs in both years ($p<0.001$ for 2008 and $p=0.05$ for 2009). For 18S TRFLP, it was quite common for non-major TRFs to be present at >5% occasionally while for 16S TRFLP, although the relative importance of the 6 major TRFs did change, there were rarely occasions when other TRFs contributed more than 5%. That 16S major TRFs represent higher taxonomic levels than 18S major TRFs should not be the reason (at least not the sole reason) for their less variable sum of major TRFs, since both major 18S and 16S TRFs included taxa from species to phylum. It may be that there are more dynamic interactions in the 18S community, which could have contributed to the difference observed. It is also possible that the 16S community is simply more compositionally stable, either with time (as demonstrated in this analysis) or with geographical distance (next part).

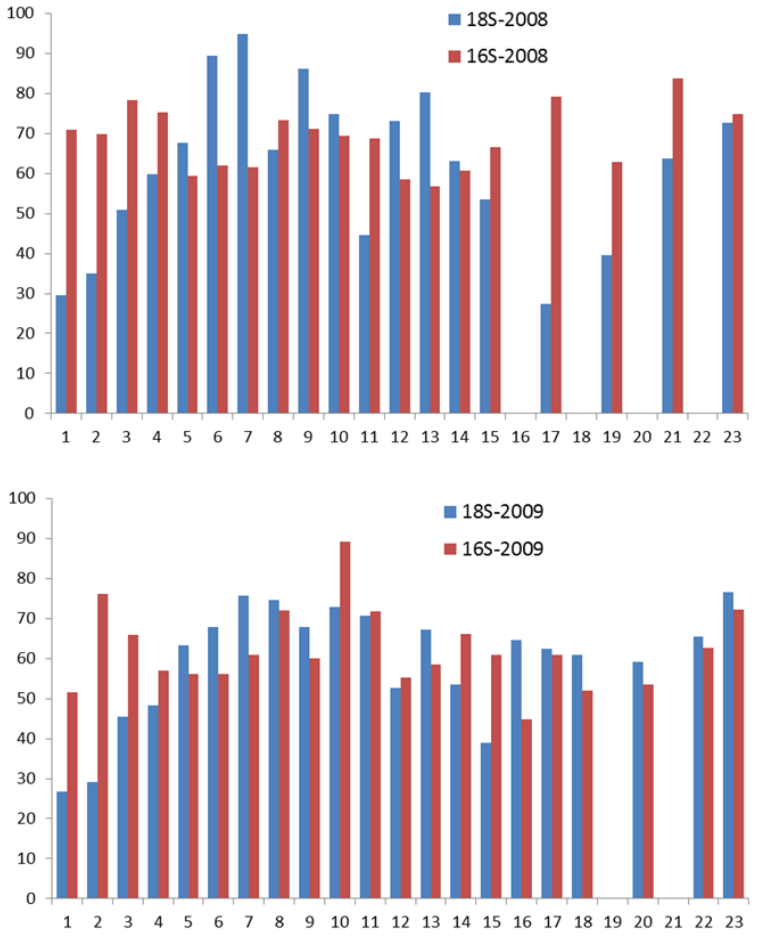


Fig. 5. 12. Sum of the major TRFs' area contribution in 18S and 16S communities in the two years

2. Spatial heterogeneity

18S TRFLP was spatially heterogeneous along the sampling transect on September 25, 2008, as discussed in Chapter 3 (Fig. 3.12.). 16S TRFLP on the same sampling day, however, demonstrated no spatial heterogeneity (Fig. 3.14.). In 2009, when samples collected from all weeks were considered together, both 18S TRFLP and 16S TRFLP demonstrated spatial heterogeneity with 18S TRFLP having a larger slope of change with geographical distance (Fig. 4.6. and Fig. 4.8.). Examination of individual weeks also identified proportionally more spatially heterogeneous weeks for 18S than 16S communities (Fig. 4.7. and Fig. 4.9.). These lines of evidence together support the idea that the 18S community is more spatially heterogeneous than the 16S community over the spatial scale covered in this study. One source of spatial variability

in 18S TRFLP could be from random sampling effects, since within the 70 to 100 ml water filtered, some protists may be present at only a few per ml. This random effect, however, should not result in a particular pattern in spatial heterogeneity (i.e. higher Bray-Curtis distance with larger geographical distance). The along-transect community pattern on the aforementioned spatially heterogeneous day, September 25, 2008, supports the contention that 18S TRFLP is more spatially variable than 16S TRFLP. Wind data and the reverse salinity structure that day suggest that wind-driven current along-shore may have induced mixing of different water masses and generated the monotonic along-transect patterns for all 18S TRFs contributing to more than 5% of total TRF area (Fig.3.12.). In fact, this was such a unique day that this monotonic along-transect trend for all major 18S TRFs was not seen on any other sampling day; neither was the mixing of water masses from station E to station H so clearly reflected by the slopes of different TRFs. With such a strong physical mixing event in place, the random sampling effect mentioned above, even if it existed, was overwhelmed (Fig. 3.12.). The fact that no spatial heterogeneity was observed for 16S TRFLP even in this extreme case is evidence that the 16S community hosted a lower spatial variability. The other sampling day with reverse salinity structure in week 13 (July 29), 2009 was not associated with wind features as in 2008, implying that there might be other complex water dynamics in GSB that are not identified yet.

Biotic and abiotic interactions in GSB

LSA in both years did not detect significant and meaningful associations with a time lag, showing that microbial interactions in GSB are effectively instantaneous on the weekly scale of sampling used here. Compared with 16S microorganisms, the larger decrease in percentage of significant associations from 2008 to 2009 in the 18S community (Table 5.11.) suggests that 18S microorganisms might have become much less interactive with each other transitioning from the brown tide year to the following non-brown tide year. The brown tide bloom seemed to have less impact on the 16S community, as discussed previously in community ordination pattern, biodiversity, variance of sum of major TRFs and spatial heterogeneity. Here again, the 16S community seemed more stable and less likely to be impacted by the brown tide bloom. It was hypothesized in Chapter 3 that proportionally more negative associations means a more stable community structure. The non-brown tide year, 2009, indeed had proportionally more negative TRF-TRF associations than in 2008, more so for 18S TRFLP than 16S TRFLP (Table 5.11.).

The current study is in general agreement with the picoalgal niche hypothesis (Sieracki et al. 1999, Sieracki 2001), although the details of how *Synechococcus* was selectively removed (which according to the hypothesis, is necessary for *A. anophagefferens* to take over) need to be further studied. LSA for both years revealed many strong associations between blue-green picophytoplankton pairs. In the brown tide bloom year, all blue-green picophytoplankton had positive associations with each other and together, they had negative associations with *A. anophagefferens* (Fig. 5.13.); with the absence of brown tide bloom in 2009, *Bathycoccus* was negatively associated with other picophytoplankton (Fig. 5.13.). In fact, as temperature was negatively associated with *A. anophagefferens* and positively associated with *Synechococcus* (Fig. 3.15. and Fig. A1.), it is likely to be an important factor influencing the succession of different species within the picoalgal niche. The only trophic interaction involving the blue-green picophytoplankton detected in both years (also the only case when the same potential grazer-prey relationship was found in both years) was the negative association between *Picochlorum atomus* and planktonic polychaete larvae (Fig. 3.17.B. and Fig. 4.12.). Selective grazing of testate amoeba on *Synechococcus* has been suggested to occur during brown tide blooms (Sieburth et al. 1988); our two-year data indicate that polychaete larvae may have preferentially removed *Picochlorum*. Investigation of what factors influence this potential grazer-prey relationship would be a step forward to understand if and how grazing by planktonic polychaete larvae can affect brown tide bloom development.

Minutocellus appeared to be relatively more abundant in 2008 (Table 5.3.). Co-occurrence of *Minutocellus* and brown tide has been reported in previous studies (Cosper et al. 1987, Sieburth et al. 1988). In fact, a study on the first brown tide bloom event in Narragansett Bay, RI, claimed that the co-occurrence of *Minutocellus* with the brown tide bloom was 'novel' to that area (Smayda & Villareal 1989), which may suggest that this diatom species was at least not present in large enough abundance to be identified before a brown tide bloom hit that area. The data from 2008 in this study showed that not only did the two species co-occur in 2008, but also there was significant negative association between *Minutocellus* and *A. anophagefferens* (Fig. 3.15. and Fig. 3.17.A., Fig. 5.13.), which could have been established through competition for organic nutrient uptake under low light condition (Dzurica et al. 1989, Cosper et al. 1990), among other factors. Based on its similar cell size to *A. anophagefferens* and the negative LS

score with *A. anophagefferens*, *Minutocellus* is suggested to also inhabit the picoalgal niche, together with blue-green picophytoplankton and *A. anophagefferens*.

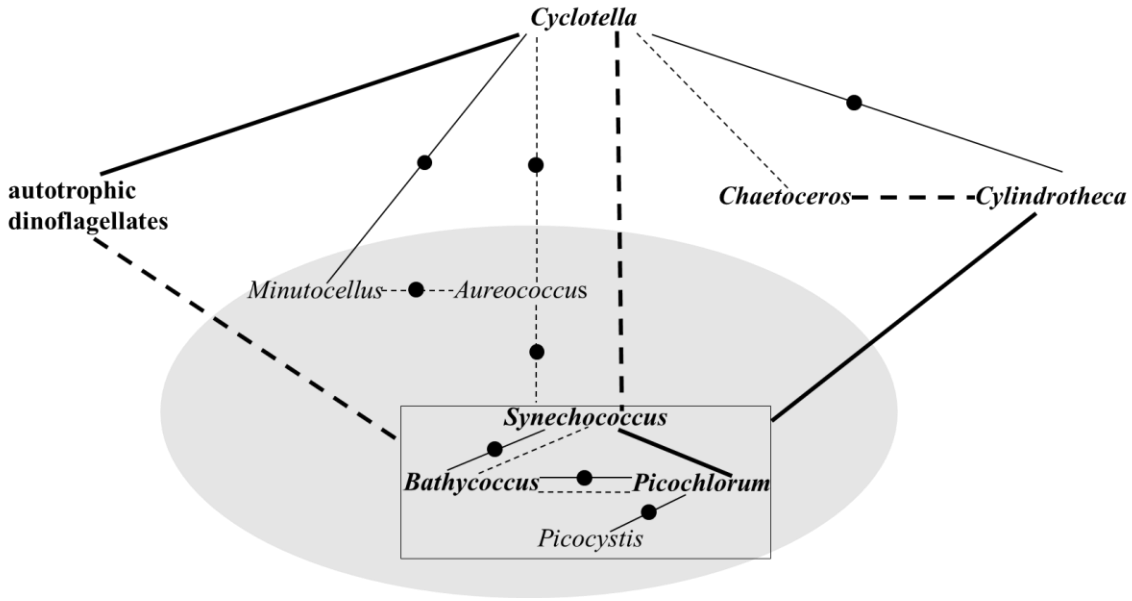


Fig. 5. 13. A schematic model for phytoplankton interactions; dashed and solid lines represent negative and positive correlation, respectively; fine lines with dots represent interactions identified only in 2008; fine lines without dots represent interactions identified only in 2009; bold lines represent interactions identified in both years; Phytoplankton names in bold represent those present in both years; square encloses blue-green picophytoplankton and shaded area encloses the proposed picoalgal niche

Given that different phytoplankters have different optimal growing temperatures, it was not surprising to find that temperature was closely associated with many major groups of phytoplankton (Fig. 5.14.). In this study, *Chaetoceros* and *Synechococcus* were found to be positively related to water temperature in both years. The positive association between *Chaetoceros* and temperature was reflected in that the diatom was most abundant only in June and July when water temperature was higher than 20 ℃, while the association between *Synechococcus* and temperature was established mostly because the change in *Synechococcus* abundance was more or less simultaneous with the change in water temperature from May to July and after September. *A. anophagefferens*, with an opposite temporal pattern to that of *Synechococcus*, was negatively associated with temperature. In 2008, relative abundance of *A. anophagefferens* started to increase in May when the water temperature was ~15 ℃ and started

to decrease in July when the water temperature was ~25 ℃, which agreed with the estimate of optimal temperature for *A. anophagefferens* growth from 20 to 25 ℃ based on lab experiments (Cosper et al. 1989). *Cyclotella* had a clear seasonal pattern with its maximum abundance in May and November (Fig. 4.2) and was negatively associated with temperature in both years. Additionally, repeated negative associations between autotrophic dinoflagellates and temperature (also Chla in 2009) were observed in both years (Fig. 3.18. and Fig. 4.14. B.). Considering all these associations between phytoplankton and temperature, GSB water seems to be dominated by nano- diatoms and dinoflagellates in spring, followed by dominance of picoalgae in summer. In brown tide year, while the initiation of the brown tide may be influenced by many environmental factors, the species succession within the picoalgal niche may be controlled mostly by temperature.

Wind speed was associated with both *A. anophagefferens* and *Minutocellus* but with an opposite sign (Fig. 5.14.). In shallow embayments like GSB, wind mixing can increase water column turbidity and bring up nutrients from the benthic environment into the water column and create a nutrient/light environment favorable for *A. anophagefferens* (Chapter 3). In fact, low light conditions can favor organic N (urea, glutamic acid) uptake by *A. anophagefferens* while inhibiting that by *Minutocellus* (Dzurica et al. 1989). The negative association between *Picochlorum* and river discharge in both years suggests that fresh water input has a negative impact on this pico-chlorophyte but the mechanism for such association is not clear. Since river discharge both affects salinity and also can be a medium to bring in nutrients that promote the growth of phytoplankton other than *Picochlorum*, laboratory studies designed to separate the effects of salinity and nutrients on the pico-algae are needed.

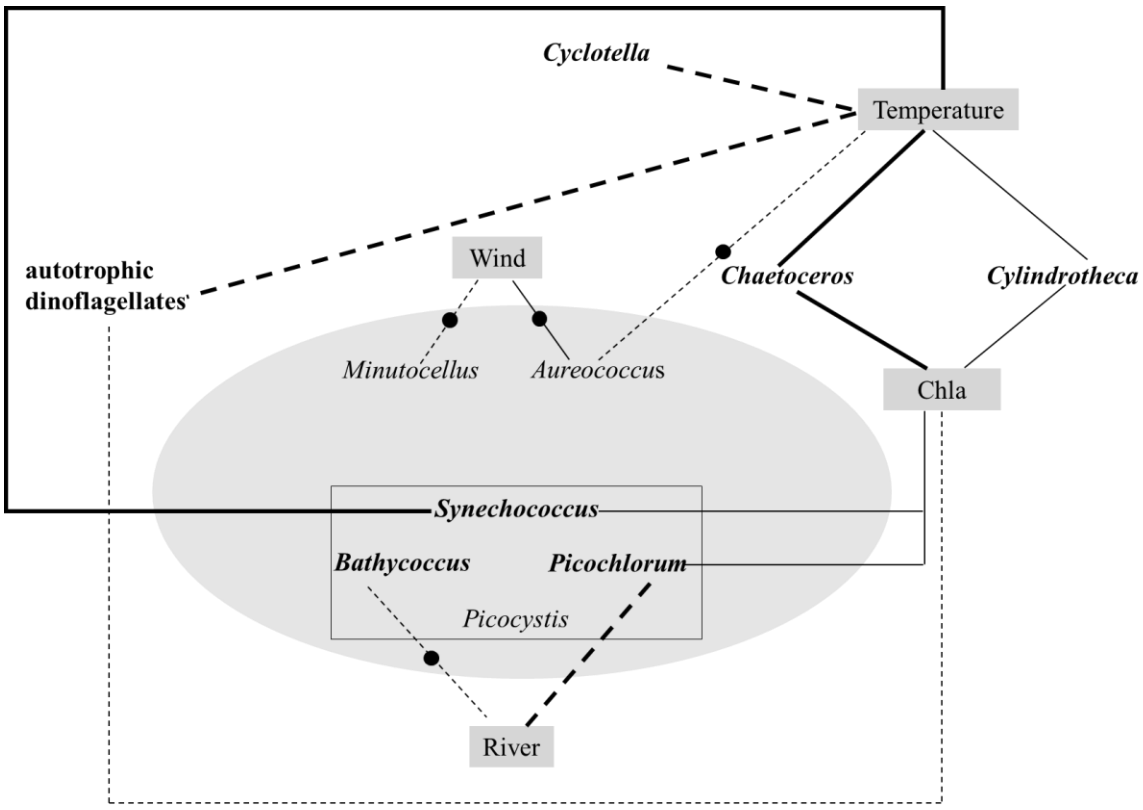


Fig. 5. 14. A schematic model for phytoplankton and environmental factor interactions; See legend in Fig. 5.13.

Lastly, it should be pointed out that with the massive amount of information generated in the network analysis, I have focused on only those associations that were most relevant to the brown tide species, such as its potential competitors, grazers, and environmental factors that were significantly associated. It should not be ignored, however, that 16S organisms include not only *Synechococcus* and other phytoplankton plastids, but also numerous heterotrophs. Information produced in this study should undergo more data mining (especially as our understanding of certain microorganisms gets deeper) to formulate explicit hypotheses guiding laboratory-based studies on model microorganisms, which in turn will help conduct field observations and sampling at appropriate spatial and temporal scales.

Chapter 6 Summary and future perspectives

Overview of this work

In the context of increasing interests and needs to better understand the factors promoting brown tide blooms, especially now that the geographic occurrence of these blooms has spread from the northeastern US coast to South Africa and the east coast of China (Cosper et al. 1987, Probyn et al. 2001, Gobler et al. 2005, Vigil et al. 2009, Probyn et al. 2010, Kong et al. 2012), this work took a spatially and temporally intensive approach and has produced a large amount of information on the GSB plankton community that was unavailable before. While trying to understand changes in the planktonic community associated with the brown tide bloom caused by *Aureococcus anophagefferens*, fundamental questions in microbial oceanography were also addressed in this study and some of the answers may be transferable to other coastal ecosystems.

Method validation

Combining the DNA-fingerprinting Terminal Restriction Fragment Length Polymorphism (TRFLP) method with cloning/sequencing, this study mapped the plankton community composition in GSB during two years. *Taq*I restriction enzyme was chosen due to its ability to produce a TRF specific for *A. anophagefferens*. The TRFLP method established in this study is of high reproducibility, and in combined with the NMS ordination analysis, it is able to reveal the spatial and temporal patterns in the microbial community. Agreement between c175B% and microscopy cell counts of *A. anophagefferens* in 2008 and between c475Y% and flow cytometer counts of *Synechococcus* in 2009 demonstrated that quantification of at least the major TRFs was reliable using this TRFLP method, which gives us the confidence to use this data set to track the plankton dynamics in GSB with a variety of statistical analyses.

Major conclusions

Spatiotemporal structure of the planktonic community in Great South Bay

In GSB, temporal variation was greater than spatial variation, reflected as close clustering of spatial samples from the same week and often clear separation of samples from consecutive weeks in the ordination analysis. With this said, during about half of the 2009 sampling weeks the 18S community showed spatial heterogeneity, i.e. communities became more and more different as they got further apart. For the 16S community, only about one third of the weeks in 2009 were spatially heterogeneous. In 2008, a difference between the 18S and 16S communities also existed; all the weeks selected statistically as being potentially spatially heterogeneous were

indeed spatially heterogeneous for the 18S while none were for the 16S community. Although the random chance of sampling (from ~100 ml seawater) a different 18S community from different locations is higher than that for 16S community due to lower abundances of 18S microorganisms in sea water, such an effect should not have a spatial pattern, and it is most likely that the weaker spatial heterogeneity of the 16S community was real. This contention is supported by the unique spatial structure observed on September 25, 2008.

On September 25, 2008, a largely alongshore wind direction (from the northeast) had lasted for the previous 3 days and wind speed was also the highest on that day of the entire 2008 sampling season. Both these features of wind may have facilitated water movement along the shore of Long Island (Wilson et al. 1991) from the saltier Moriches Bay to GSB, and mixing of GSB water and Moriches Bay water may have caused all the major 18S TRFs to demonstrate an along-transect pattern of two end-member communities near each transect end with mixing in between. Even with such a strong physical force in place, the 16S community demonstrated no spatial structure and, which indicates that the 16S community was similar even in different water masses with different 18S TRF composition. Wind-driven current alongshore was the only factor that was found to explain the spatial heterogeneity over the sampling transect, and only in one sampling week. The spatial heterogeneity observed in the other weeks did not seem to be related to any environmental factors considered-temperature, salinity, salinity gradient over the transact or river discharge.

Brown tide effect on the planktonic community

When both years' samples were ordinated together using NMS, separation between years was seen in both 18S and 16S TRFLP. After removing *A. anophagefferens* TRFs from the ordination data matrix, the separation between the two years became less distinct for both 18S and 16S communities, but was still discernible, suggesting that the presence of *A. anophagefferens* was not the only difference between the two years. When 2008 data were analyzed alone, the brown tide bloom weeks (when c175B>15% and 298Y>7.8%) became indistinguishable from the other weeks for 18S TRFLP and less distinct for 16S TRFLP, suggesting that bloom weeks were unique mostly because of the large abundance of *A. anophagefferens* TRFs in 18S and 16S TRFLP. With the *A. anophagefferens* TRFs removed,

wind speed was no longer associated with any of the 18S ordination axes, consistent with the significant relationship between *A. anophagefferens* abundance and wind speed found in LSA.

Potential Biotic and abiotic interactions in GSB plankton communities and their implications for brown tide bloom initiation and progression

A few groups of phytoplankton were found to have significant relationships with each other in both years: blue-green picophytoplankton including *Synechococcus, Bathycoccus, Picochlorum* and *Picocystis*, *Chaetoceros* (always negatively associated with other diatoms), autotrophic dinoflagellates, and a group of non-*Chaetoceros* diatoms. In 2008, TRFs representing *A. anophagefferens* and the minute diatom *Minutocellus polymorphus* were also relatively abundant, and they had a negative relationship with each other. In 2008, the four blue-green picophytoplankton were positively associated with each other, and were all negatively associated with *A. anophagefferens*, suggesting that the blue-green picophytoplankton competed together with *A. anophagefferens* for the picoalgal niche (Sieracki et al. 1999, Sieracki 2001). Phagotrophic flagellates of the genus *Telonema* and planktonic larvae of the polychaete *Clymenura clypeata* were negatively associated with blue-green picophytoplankton, suggesting that predator-prey relationships may exist between the picoalgae and these grazers. *Chaetoceros* was negatively associated with most of the grazers/parasites while *A. anophagefferens* was negatively associated with only one grazer group, a TRF representing both Syndiniales and copepods. Based on these associations and the observation during May-June, 2008, that NH_4^+ concentration measured at a station close to our study sites (~8 km) demonstrated an opposite temporal trend to that of 174B% for four weeks, the following scenario was suggested to explain *A. anophagefferens* bloom progression.

Between May and June in 2008 when the temperature was below 22 ℃ and within the optimal temperature range for *A. anophagefferens* growth (Cosper et al. 1989), faster growth rate of *A. anophagefferens* than other phytoplankton when provided with moderate concentration of reduced N species such as NH_4^+ (~5-6 μM) (Taylor et al. 2006) may have helped it establish its dominance. As the water temperature increased to ~ 25 ℃ in July, the blue-green picophytoplankton may have outcompeted *A. anophagefferens* in the picoalgal niche because the water temperature was too high for the latter (Cosper et al. 1989), leaving *A. anophagefferens*' relative abundance below detection limit during late July and the whole August. The reason for

the secondary brown tide bloom in fall was less clear, but selective removal of the blue-green picophytoplankton by *Telonema* and polychaete larvae may have contributed to it. In contrast, from August to late September, the suggested grazer(s) of *A. anophagefferens* (Syndiniales/copepod) had been decreasing in relative abundance, perhaps providing *A. anophagefferens* with a window of opportunity to again outcompete the blue-green picophytoplankton at that time. Throughout the whole sampling season, wind mixing stirs up the water column, creating dark conditions under which *A. anophagefferens* is suggested to take up organic nutrients more effectively than similarly sized cells such as *Minutocellus polymorphus* (Dzurica et al. 1989, Cosper et al. 1990, which is in line with wind's positive and negative relationship with A. anophagefferens and Minutocellus polymorphus (respectively) in LSA.)

In the following non-brown tide year, although the associations between *Synechococcus* and *Picochlorum* were still positive as in the brown tide year (relative abundance of *Picocystis* was below detection limit), *Bathycoccus* had negative relationship with *Synechococcus* and *Picochlorum*, suggesting that *Bathycoccus* may have replaced *A. anophagefferens* in the non-brown tide year. Additionally, the relative abundance of *Bathycoccus* was ~ 4.4% in the non-brown tide year, a value borderline that for a major TRF, emphasizing again its potential ecological importance in GSB, where the presence of this species has not been reported before.

Future perspectives

Long term study of planktonic community and environmental factors in GSB is needed

This study highlighted the necessity of having long-term monitoring programs in place to better understand coastal ecosystems. Weekly sampling revealed some short-lived yet striking features of the microbial community in GSB which would not have been captured with less frequent sampling. The almost simultaneous 'week 7 anomaly' occurring in both years is one example of a feature revealed by weekly sampling that will be difficult to understand without longer term study. The 'week 7 anomaly' might have something to do with fresh water input. Since groundwater discharge accounts for greater than 20% of the freshwater input into GSB (Bokuniewicz 1980, Bokuniewicz & Pavlik 1990), one possible explanation is that groundwater discharge relatively enriched with NO_3^- (Capone & Bautista 1985, Capone & Slater 1990, Laroche et al. 1997) was high during that time in both years due to prior rainy days, and may have fostered the growth of the diatom *Chaetoceros*. Since groundwater inputs of inorganic N,

point sources of N, and various exchanges with the coastal ocean, among other processes, are variable in time and space (Laroche et al. 1997), long term studies of planktonic community structure as well as N nutrient characterization from both river input and other sources are needed to unravel the secret of the 'week 7 anomaly'.

Another example of how weekly sampling benefited our understanding of the dynamic nature of GSB (and probably most temperate coastal embayments) is the effect of wind on the planktonic community structure over short time scales. The negative association between wind speed and *Minutocellus polymorphus* as well as the positive association between wind speed and *A. anophagefferens* was already discussed. Similar to *Minutocellus polymorphus*, the rotifer *Brachionus calyciflorus* was negatively associated with wind, consistent with the suggestion that grazing rate of this rotifer species is significantly reduced in turbulent environments (Miquelis et al. 1998). All these associations between organisms and wind speed involve the instantaneous effect of wind in mixing the water column, supported by the result that none of the wave energy index RWE computed by WEMo for time blocks from half a day to a week had better correlations with the relative abundance of *A. anophagefferens* than the original wind speed data. In addition to the instantaneous effect, wind-driven current along shore may have shaped the spatial structure of microbial community as discussed previously in this chapter. The temporal scale of wind-driven alongshore transport, according to both this study and a previous one (Wilson et al. 1991), is around 3-4 days. Therefore, unlike other open water types of coastal environments where there is a larger buffering zone (less enclosed and larger water depth) to make them more resistant to wind effect, (Fuhrman et al. 2006, Gilbert et al. 2012), shallow and usually more enclosed coastal lagoons are more sensitive to the constantly occurring environmental events such as wind (especially when a coastal system harbors different water sources such as GSB and Moriches Bay), making prediction of the microbial communities in coastal lagoons more challenging. This study emphasizes that wind is a very important factor to be considered in understanding the spatiotemporal patterns of planktonic communities within shallow, semi-enclosed coastal environments and encourages further study to develop more robust wind models based on long-term observations that are site-specific.

Implications for conducting similar coastal environments

The spatiotemporally intensive approach of this study has provided data unavailable for GSB previously. Based on the results generated, the following suggestions are made for conducting similar studies in the future.

First, how often do we need to sample? In this study, the majority of the significant and reasonable associations were found in the LSA without time delay, suggesting that the interactions between microorganisms in GSB were effectively instantaneous at this time scale. Therefore, monthly or even biweekly sampling is likely to miss major changes in the microbial community. The 'week 7 anomaly' in both years, for example, was less likely to be discovered had the sampling frequency were 2 weeks or longer. It is safe to say that our current understanding about oceanic processes is limited by our sampling ability and different environments (even all being coastal) may host different rates of changes in microbial community. To get some insight into the temporal change dynamics, temporally intensive sampling (daily) can be carried out initiatively for a short term (1~2 weeks) and preliminary results collected to decide if temporally intensive observation is needed. Alternatively, we could resort to autonomous sampling devices deployed in the field (e.g. the Environmental Sample Processor by Scholin team at MBARI), which is probably the future of conducting observational studies in the ocean.

Second, how intensely should we sample spatially? The spatial sampling was done the way described in this study because one of the scientific questions I want to address was whether there is spatial structure along a short transect in GSB. While this intensity of sampling was essential to understanding the physical processes occurring during the most spatially heterogeneous week (September 25, 2008), if the purpose of a monitoring program is simply to capture the temporal dynamics in the plankton microbial community, samples taken from one station should be reasonably representative of the study area. The supporting evidence for this statement is in the NMS ordination patterns, where spatial samples tended to clustered together.

The third question then becomes: what location should we sample in the study area? Except for September 25, 2008, there was no evidence that any particular spatial trend of the major TRFs existed at other times. Depending on the specific scientific question, decisions can be made accordingly. For example, if one wants to see how the plankton community structure

changes with time under the influence of river runoff, sampling right off the river mouth would be most effective. If a transect into the river plume is to be designed, sampling orientation should be adjusted based on salinity rather than having stations set in space.

Lastly, what environmental data should be collected? The ordination analyses suggest that temperature, salinity, river discharge and wind speed (average value of the day should suffice) are all important factors to include. Interestingly, although salinity and river discharge were always significantly correlated, there were times when only river discharge was significantly associated with an ordination axis, suggesting that river water may have its own effect on community structure that is not related to salinity. Chlorophyll a concentration was collected only in 2009, when it was significantly associated with temperature. In fact, since it was always associated with ordination axes that were also associated with temperature, chlorophyll a does not seem to provide extra information in understanding microbial community change if temperature is already included. A caveat is that a significant association between chlorophyll a and temperature needs to be established first for a study area before chlorophyll a is excluded from analysis. Additionally, for GSB and probably also similar shallow coastal lagoons, wind can be a major force in generating turbulence and turbidity in the water column, changing the light environment organisms experience. In order to back up the potential of wind to affect turbidity, Secchi disk readings are desirable data to have. Lastly, it would be optimal to have the concentrations of major nutrient species. Depending on the particular type of environment under investigation, different nutrient species may be of particular importance. Using this study for an example, since *A. anophagefferens* and other phytoplankton have been shown to grow at different rates when NH_4^+, NO_3^- or urea is the nitrogen source (Taylor et al. 2006), measuring these three nutrient species may have provided useful information regarding the dynamics of brown tide blooms.

Better understanding of the potential interactions revealed in LSA

The strength of LSA is that it considers all the 18S and 16S TRFs and searches for all significant correlations between different members, which is impossible to do in lab experiments. The current study, however, has not explored extensively the associations involving heterotrophic bacteria revealed by LSA. With the enormous amount of information generated in LSA, one effective way of using these network analyses would be to have particular questions in

mind and look for relevant information. Data mining on the dataset should continue, especially as more and more interactions involving heterotrophic bacteria and phytoplankton are revealed to guide us where to look (Amin et al. 2012).

A better understanding of the nature of the positive and negative associations detected by LSA is needed. Although negative associations usually indicate competition or predation (Steele et al. 2011), explanation of positive associations is more complex. As suggested by Steel et al. (2011), positive association in LSA may indicate endosymbiosis or parasitism. However, some Syndiniales are obligatory killers of their host (parasitoids) and are able to prevent reproduction of host cells in a few days. *Amoebophrya ceratii,* for example, was found to cause rapid decline of several bloom-forming dinoflagellates within a month (Chambouvet et al. 2008). Under such circumstances, a negative association between parasites and hosts is likely to be found. In this study, negative relationships between parasitic Syndiniales and *Chaetoceros* were found in both years while *Amoebophrya* was positively associated with non-*Chaetoceros* phytoplankton in 2008. The host-parasite/parasitoids interactions in aquatic systems are in general not well understood. It is possible that while some parasites are instantaneously lethal, some others can be lethal but not as instantaneously and the others are not lethal. These different types of parasites could establish different types of interactions with their host cells and whether they have positive or negative association with their hosts in LSA might also depend on sampling timescale relative to host/parasite interaction. Study of the parasites' life cycle and host/parasite specificity is probably the key to understanding the parasite interactions identified in LSA, and conducting direct lab feeding/infection experiments based on suggestions made by LSA is an efficient way to start understanding these complex interactions.

While a lot of similarities between the two years were identified in the schematic models of interactions, many differences also existed between years and some of them are due to different microbial compositions in the two years. For example, *Pirsonia* was relatively more abundant in 2008 than 2009 and was positively associated with all blue-green picophytoplankton in 2008, but no such associations existed in 2009. If the positive association indicates a parasite-host relationship, then infection by *Pirsonia* in 2008 may be a mechanism through which *Synechococcus* (Sieracki et al. 1999), as well as other picophytoplankton, was selectively removed, opening the picoalgal niche to *A. anophagefferens*. A direct way to test the hypothesis

that *Pirsonia* infection of blue-green picophytoplankton opens the picoalgal niche to *A. anophagefferens* would be to conduct experiments in the lab where *Pirsonia* is exposed to a mixed potential diet of the four blue-green picophytoplankton and other similarly sized cells such as *A. anophagefferens* and *Minutocellus polymorphus*.

To sum up, the fine spatial and temporal sampling resolution of this study offered great insight into planktonic community structure in a brown tide year and a non-brown tide year in GSB. Spatial and temporal variations of 18S and 16S communities were compared. The temporal trend of relatively abundant organisms was examined, and the spatial heterogeneity along a ~10 km transect in the middle of GSB was also investigated. Potential mechanisms through which *A. anophagefferens* established and maintained its dominance were identified and the effect of brown tide blooms on community structure and diversity was also reported. LSA revealed many potentially important biotic and abiotic associations and they are very helpful in formulating explicit hypotheses to design lab experiments, which will in turn guide field work to be conducted at the appropriate spatial and temporal scales.

Reference

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. Microbiological Reviews 59:143-169

Amin SA, Parker MS, Armbrust EV (2012) Interactions between Diatoms and Bacteria. Microbiology and Molecular Biology reviews 76:667-684

Andersen RA, Saunders GW, Paskind MP, Sexton JP (1993) Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. Journal of Phycology 29:15

Anderson DM, Keafer BA, Kulis DM, Waters RM, Nuzzi R (1993) An immunofluorescent survey of the brown tide chrysophyte *Aureococcus anophagefferens* along the northeast coast of the United States. Journal of Plankton Research 15:563-580

Baldwin AJ, Moss JA, Pakulski JD, Catala P, Joux F, Jeffrey WH (2005) Microbial diversity in a Pacific Ocean transect from the Arctic to Antarctic circles. Aquatic Microbial Ecology 41:91-102

Berg GM, Glibert PM, Lomas MW, Burford MA (1997) Organic nitrogen uptake and growth by the chrysophyte *Aureococcus anophagefferens* during a brown tide event. Marine Biology 129:377-387

Boissonneault-Cellineri KR, Mehta M, Lonsdale DJ, Caron DA (2001) Microbial food web interactions in two Long Island embayments. Aquatic Microbial Ecology 26:139-155

Bokuniewicz H (1980) Groundwater seepage into Great South Bay, New York. Estuarine and Coastal Marine Science 10:437-444

Bokuniewicz H, Pavlik B (1990) Groundwater seepage along a barrier island. Biogeochemistry 10:257-276

Bricelj VM, MacQuarrie SP, Schaffner RA (2001) Differential effects of *Aureococcus anophagefferens* isolates ("brown tide") in unialgal and mixed suspensions on bivalve feeding. Marine Biology 139:605-615

Butcher RW (1952) Contributions to our knowledge of the smaller marine algae. Journal of the Marine Biological Association of the United Kingdom 31:175-191

Capone DG, Bautista MF (1985) A groundwater source of nitrate in nearshore marine sediments. Nature 313:214-216

Capone DG, Slater JM (1990) Interannual patterns of water-table height and groundwater derived nitrate in nearshore sediments. Biogeochemistry 10:277-288

Caron DA, Gobler CJ, Lonsdale DJ, Cerrato RM, Schaffner RA, Rose JM, Buck NJ, Taylor G, Boissonneault KR, Mehran R (2004) Microbial herbivory on the brown tide alga, *Aureococcus anophagefferens*: results from natural ecosystems, mesocosms and laboratory experiments. Harmful Algae 3:439-457

Caron DA, Lim I, Kunze H, Cosper EM, Anderson DM (1989) Trophic interactions between nano- and microzooplankton and the "brown tide". In: Cosper EM, Bricelj, V. M. and Carpenter, E. J. (ed) Novel Phytoplankton Blooms: Causes and Impacts of Recurrent Brown Tides and Other Unusual Blooms, Book 35. Springer-Verlag, Berlin

Caron DA, Worden AZ, Countway PD, Demir E, Heidelberg KB (2009) Protists are microbes too: a perspective. Isme Journal 3:4-12

Carvalho LR, Cox EJ, Fritz SC, Juggins S, Sims PA, Grasse F, Batterbee RW (1995) Standardizing the taxonomy of saline lake *Cyclotella spp*. Diatom Research 10:229-240

Castro LR, Cowen RK (1991) Environmental factors affecting the early life history of bay anchovy Anchoa mitchilli in Great South Bay, New York. Marine Ecology-Progress Series 76:235-247

Chambouvet A, Morin P, Marie D, Guillou L (2008) Control of Toxic Marine Dinoflagellate Blooms by Serial Parasitic Killers. Science 322:1254-1257

Chantangsi C, Esson HJ, Leander BS (2008) Morphology and molecular phylogeny of a marine interstitial tetraflagellate with putative endosymbionts: Auranticordis quadriverberis n. gen. et sp (Cercozoa). Bmc Microbiology 8

Cho JC, Tiedje JM (2000) Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. Applied and Environmental Microbiology 66:5448-5456

Coats DW, Park MG (2002) Parasitism of photosynthetic dinoflagellates by three strains of Amoebophrya (Dinophyta): Parasite survival, infectivity, generation time, and host specificity. Journal of Phycology 38:520-528

Cosper EM (2001) Water Column Productivity. In: Tanski J, Bokuniewicz H, Schlenk C (eds) Impacts of Barrier Island Breaches on Selected Biological Resources of Great South Bay, New York, Stony Brook, NY

Cosper EM, Dennison W, Milligan A, Carpenter EJ, Lee C, Holzapfel J, Milanese L (1989) An examination of the environmental factors important to initiating and sustaining "brown tide" blooms. In: Cosper EM, Bricelj VM, Carpenter EJ (eds) Novel phytoplankton blooms. Springer-Verlag

Cosper EM, Dennison WC, Carpenter EJ, Bricelj VM, Mitchell JG, Kuenstner SH, Colflesh D, Dewey M (1987) Recurrent and persistent brown tide blooms perturb coastal marine ecosystem. Estuaries 10:284-290

Cosper EM, Lee C, Carpenter EJ (1990) Novel "brown tide" blooms in Long Island embayments: a search for the causes, Vol. Elsevier, New York

Countway PD, Gast RJ, Savai P, Caron DA (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. Journal of Eukaryotic Microbiology 52:95-106

Crump BC, Hopkinson CS, Sogin ML, Hobbie JE (2004) Microbial biogeography along an estuarine salinity gradient: Combined influences of bacterial growth and residence time. Applied and Environmental Microbiology 70:1494-1505

Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H, Buckley DH (2009) T-REX: software for the processing and analysis of T-RFLP data. Bmc Bioinformatics 10

Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R, Brown AJL (2000) Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. Nature 405:43-47

de Wit R, Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environmental Microbiology 8:755-758

Delong EF (1992) Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the United States of America 89:5685-5689

DeLong EF, Karl DM (2005) Genomic perspectives in microbial oceanography. Nature 437:336-342

Dennison WC, Marshall GJ, Wigand C (1989) Effect of "Brown Tide" shading on eelgrass (*Zostera marina* L.) distributions. In: Cosper EM, Bricelj VM, Carpenter EJ (eds) Novel phytoplankton blooms: causes and impacts of recurrent brown tides and other unusual blooms. Springer-Verlag, Berlin

Dennison WC, Orth RJ, Moore KA, Stevenson JC, Carter V, Kollar S, Bergstrom PW, Batiuk RA (1993) Assessing water quality with submersed aquatic vegetation. Bioscience 43:86-94

Deonarine SN, Gobler CJ, Lonsdale DJ, Caron DA (2006) Role of zooplankton in the onset and demise of harmful brown tide blooms (Aureococcus anophagefferens) in US mid-Atlantic estuaries. Aquatic Microbial Ecology 44:181-195

Diez B, Pedros-Alio C, Massana R (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Applied and Environmental Microbiology 67:2932-2941

Duguay LE, Monteleone DM, Quaglietta CE (1989) Abundance and distribution of zooplankton and ichthyoplankton in Great South Bay, New York during the brown tide outbreaks of

1985 and 1986. In: Cosper EM, Bricelj VM, Carpenter EJ (eds) Novel Phytoplankton Blooms: Causes and Impacts of Recurrent Brown Tides and Other Unusual Blooms

Dzurica S, Lee C, Cosper EM, Carpenter EJ (1989) Role of environmental variables, specifically organic compounds and micronutrients, in the growth of the chrysophyte *Aureococcus anophagefferens*. In: Cosper EM, Bricelj VM, Carpenter EJ (eds) Novel Phytoplankton Blooms: Causes and Impacts of Recurrent Brown Tides and Other Unusual Blooms. Springer-Verlag, Berlin, Germany

Egert M, Friedrich MW (2003) Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. Applied and Environmental Microbiology 69:2555-2562

Egert M, Friedrich MW (2005) Post-amplification Klenow fragment treatment alleviates PCR bias caused by partially single-stranded amplicons. Journal of Microbiological Methods 61:69-75

Elser JJ, Sterner R (2002) Ecological Stoichiometry: the biology of elements from molecules to the biosphere Vol c2002. Princeton University Press, Princeton

Fenchel T, Finlay BJ (2004) The ubiquity of small species: Patterns of local and global diversity. Bioscience 54:777-784

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences of the United States of America 103:13104-13109

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:7774-7778

Gast RJ, Dennett MR, Caron DA (2004) Characterization of protistan assemblages in the Ross Sea, Antarctica, by denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 70:2028-2037

Gastrich MD, Lathrop R, S. H, Weinstein MP, Danko M, Caron DA, Shaffner R (2003) Brown Tide Bloom Assessment Project in NJ Coastal Waters: 2000-2002. In: Environmental Assessment and Risk Analysis Element Research Project Summary, New Jersey, US

Gastrich MD, Wazniak CE (2002) A Brown Tide Bloom Index based on the potential harmful effects of the brown tide alga, Aureococcus anophagefferens Aquatic Ecosystem Health & Management 5:435 - 441

Gescher C, Metfies K, Frickenhaus S, Knefelkamp B, Wiltshire KH, Medlin LK (2008) Feasibility of assessing the community composition of prasinophytes at the Helgoland roads sampling site with a DNA microarray. Applied and Environmental Microbiology 74:5305-5316

Gilbert JA, Meyer F, Bailey MJ (2011) The Future of microbial metagenomics (or is ignorance bliss?). Isme Journal 5:777-779

Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, Somerfield P, Fuhrman JA, Field D (2012) Defining seasonal marine microbial community dynamics. Isme Journal 6:298-308

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

Giovannoni SJ, Rappe MS (2000) Evolution, diversity and molecular ecology of marine prokaryotes. In: Kirchman DL (ed) Microbial ecology of the oceans. Wiley-Liss, New York

Glücksman E (2011) Taxonomy, biodiversity, and ecology of Apusozoa (Protozoa). Ph.D., University of Oxford,

Gluecksman E, Bell T, Griffiths RI, Bass D (2010) Closely related protist strains have different grazing impacts on natural bacterial communities. Environmental Microbiology 12:3105-3113

Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV, Zhang Y, Collier JL, Wurch LL, Kustka AB, Dill BD, Shah M, VerBerkmoes NC, Kuo A, Terry A, Pangilinan J, Lindquist EA, Lucas S, Paulsen IT, Hattenrath-Lehmann TK, Talmage SC, Walker EA, Koch F, Burson AM, Marcoval MA, Tang Y-Z, LeCleir GR, Coyne KJ, Berg GM, Bertrand EM, Saito MA, Gladyshev VN, Grigoriev IV (2011) Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. Proceedings of the National Academy of Sciences of the United States of America 108:4352-4357

Gobler CJ, Boneillo GE, Debenham CJ, Caron DA (2004a) Nutrient limitation, organic matter cycling, and plankton dynamics during an *Aureococcus anophagefferens* bloom. Aquatic Microbial Ecology 35:31-43

Gobler CJ, Deonarine S, Leigh-Bell J, Gastrich MD, Anderson OR, Wilhelm SW (2004b) Ecology of phytoplankton communities dominated by *Aureococcus anophagefferens*: the role of viruses, nutrients, and microzooplankton grazing. Harmful Algae 3:471-483

Gobler CJ, Lonsdale DJ, Boyer GL (2005) A review of the causes, effects, and potential management of harmful brown tide blooms caused by *Aureococcus anophagefferens* (Hargraves et Sieburth). Estuaries 28:726-749

Gobler CJ, Renaghan MJ, Buck NJ (2002) Impacts of nutrients and grazing mortality on the abundance of *Aureococcus anophagefferens* during a New York brown tide bloom. Limnology and Oceanography 47:129-141

Gobler CJ, Sanudo-Wilhelmy SA (2001a) Effects of organic carbon, organic nitrogen, inorganic nutrients, and iron additions on the growth of phytoplankton and bacteria during a brown tide bloom. Marine Ecology-Progress Series 209:19-34

Gobler CJ, Sanudo-Wilhelmy SA (2001b) Temporal variability of groundwater seepage and brown tide blooms in a Long Island embayment. Marine Ecology-Progress Series 217:299-309

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M, Beattie AJ (2004) Spatial scaling of microbial eukaryote diversity. Nature 432:747-750

Greenfield DI, Lonsdale DJ (2002) Mortality and growth of juvenile hard clams *Mercenaria mercenaria* during brown tide. Marine Biology 141:1045-1050

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environmental Microbiology 10:3349-3365

Hakansson H, Hajdu S, Snoeus P, Longinova L (1993) *Cyclotella hakanssoniae* Wendker and its relationship to *C. caspia* Grunow and other similar brackish water *Cyclolella* species. Diatom Research 8:333-347

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Heidelberg KB, Gilbert JA, Joint I (2010) Marine genomics: at the interface of marine microbial ecology and biodiscovery. Microbial Biotechnology 3:531-543

Henley WJ, Hironaka JL, Guillou L, Buchheim MA, Buchheim JA, Fawley MW, Fawley KP (2004) Phylogenetic analysis of the "Nannochloris-like" algae and diagnoses of Picochlorum oklahomensis gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). Phycologia 43:641-652

Hillebrand H, Watermann F, Karez R, Berninger UG (2001) Differences in species richness patterns between unicellular and multicellular organisms. Oecologia 126:114-124

Hoek Cvd, Mann DG, Jahns HM (1994) Algae : an introduction to phycology Vol. Cambridge University Press, Cambridge

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setala H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75:3-35

Ibelings BW, De Bruin A, Kagami M, Rijkeboer M, Brehm M, van Donk E (2004) Host parasite interactions between freshwater phytoplankton and chytrid fungi (Chytridiomycota). Journal of Phycology 40:437-453

Irigoien X, Huisman J, Harris RP (2004) Global biodiversity patterns of marine phytoplankton and zooplankton. Nature 429:863-867

Jordana E, Gremare A, Lantoine F, Courties C, Charles F, Amouroux JM, Vetion G (2001) Seasonal changes in the grazing of coastal picoplankton by the suspension-feeding polychaete Ditrupa arietina (OF Muller). Journal of Sea Research 46:245-259

Kagami M, de Bruin A, Ibelings BW, Van Donk E (2007) Parasitic chytrids: their effects on phytoplankton communities and food-web dynamics. Hydrobiologia 578:113-129

Kagami M, Helmsing NR, van Donk E (2011) Parasitic chytrids could promote copepod survival by mediating material transfer from inedible diatoms. Hydrobiologia 659:49-54

Kan J, Evans SE, Chen F, Suzuki MT (2008) Novel estuarine bacterioplankton in rRNA operon libraries from the Chesapeake Bay. Aquatic Microbial Ecology 51:55-66

Kaplan CW, Kitts CL (2003) Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. Journal of Microbiological Methods 54:121-125

Keller MD, Rice RL (1989) Effect of nutrient enrichment on natural populations of the brown tide phytoplankton *Aureococcus anophagefferens* (Chrysopheceae). Journal of Phycology 25:636–646

Kelly KM, Chistoserdov AY (2001) Phylogenetic analysis of the succession of bacterial communities in the Great South Bay (Long Island). Fems Microbiology Ecology 35:85-95

Kirchman DL, Yu LY, Fuchs BM, Amann R (2001) Structure of bacterial communities in aquatic systems as revealed by filter PCR. Aquatic Microbial Ecology 26:13-22

Klaveness D, Shalchian-Tabrizi K, Thomsen HA, Eikrem W, Jakobsen KS (2005) *Telonema antarcticum* sp nov., a common marine phagotrophic flagellate. International Journal of Systematic and Evolutionary Microbiology 55:2595-2604

Kong FZ, Yu RC, Zhang QC, Yan T, Zhou MJ (2012) Pigment characterization for the 2011 bloom in Qinhuangdao implicated "brown tide" events in China. Chinese Journal of Oceanology and Limnology 30:361-370

Kuhn S, Medlin L, Eller G (2004) Phylogenetic position of the parasitoid nanoflagellate pirsonia inferred from nuclear-encoded small subunit ribosomal DNA and a description of Pseudopirsonia n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov. Protist 155:143-156

Kuhn SF (1997) Infection of *Coscinodiscus spp.* by the parasitoid nanoflagellate *Pirsonia diadema* .1. Behavioural studies on the infection process. Journal of Plankton Research 19:791-804

Kuhn SF (1998) Infection of *Coscinodiscus spp.* by the parasitoid nanoflagellate *Pirsonia diadema*: II. Selective infection behaviour for host species and individual host cells. Journal of Plankton Research 20:443-454

Kuris AM, Hechinger RF, Shaw JC, Whitney KL, Aguirre-Macedo L, Boch CA, Dobson AP, Dunham EJ, Fredensborg BL, Huspeni TC, Lorda J, Mababa L, Mancini FT, Mora AB, Pickering M, Talhouk NL, Torchin ME, Lafferty KD (2008) Ecosystem energetic implications of parasite and free-living biomass in three estuaries. Nature 454:515-518

Langenheder S, Ragnarsson H (2007) The role of environmental and spatial factors for the composition of aquatic bacterial communities. Ecology 88:2154-2161

Laroche J, Nuzzi R, Waters R, Wyman K, Falkowski PG, Wallace DWR (1997) Brown Tide blooms in Long Island's coastal waters linked to interannual variability in groundwater flow. Global Change Biology 3:397-410

Lee WJ, Brandt SM, Vors N, Patterson DJ (2003) Darwin's heterotrophic flagellates. Ophelia 57:63-98

Lewin RA (1954) A Marine *Stichococcus sp.* which requires Vitamin BI2 (Cobalamin). Journal of General Microbiology 10:93-96

Licciano M, Stabili L, Giangrande A (2005) Clearance rates of Sabella spallanzanii and Branchiomma luctuosum (Annelida : Polychaeta) on a pure culture of Vibrio alginolyticus. Water Research 39:4375-4384

Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Applied and Environmental Microbiology 63:4516-4522

Lively JS, Kaufman Z, Carpenter EJ (1983) Phytoplankton ecology of a barrier-island estuary - Great South Bay, New York. Estuarine Coastal and Shelf Science 16:51-68

Livingston RJ (2000) Origin and succession of plankton blooms and effects on secondary production in Gulf Coast estuaries. In: Eutrophication processes in coastal systems. CRC Press, Boca Raton, Florida

LoBue C, Bortman M (2011) Hard Clams, Hard Lessons: The Shellfish Renaissance. All Solutions

LoBue C, Doall M (2010) Restoring Hard Clams to Great South Bay. In, The Nature Conservancy on Long Island, 250 Lawrence Hill Rd., Cold Spring Harbor

Lomas MW, Glibert PM, Clougherty DA, Huber DR, Jones J, Alexander J, Haramoto E (2001) Elevated organic nutrient ratios associated with brown tide algal blooms of *Aureococcus anophagefferens* (Pelagophyceae). Journal of Plankton Research 23:1339-1344

Lonsdale DJ, Cosper EM, Kim WS, Doall M, Divadeenam A, Jonasdottir SH (1996) Food web interactions in the plankton of Long Island bays, with preliminary observations on brown tide effects. Marine Ecology-Progress Series 134:247-263

Lonsdale DJ, Greenfield DI, Hillebrand EM, Nuzzi R, Taylor GT (2006) Contrasting microplanktonic composition and food web structure in two coastal embayments (Long Island, NY, USA). Journal of Plankton Research 28:891-905

Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, Hooper DU, Huston MA, Raffaelli D, Schmid B, Tilman D, Wardle DA (2001) Ecology - Biodiversity and ecosystem functioning: Current knowledge and future challenges. Science 294:804-808

MacIntyre HL, Lomas MW, Cornwell J, Suggett DJ, Gobler CJ, Koch EW, Kana TM (2004) Mediation of benthic-pelagic coupling by microphytobenthos: an energy- and material-based model for initiation of blooms of *Aureococcus anophagefferens*. Harmful Algae 3:403-437

Malhotra A, Fonseca MS (2007) WEMo (Wave Exposure Model): Formulation, Procedures and Validation. NOAA Technical Memorandum NOS

Marsh TL (1999) Teminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. Current opinion in microbiology

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach AL, Smith VH, Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4:102-112

Mazzillo FFM, Ryan JP, Silver MW (2011) Parasitism as a biological control agent of dinoflagellate blooms in the California Current System. Harmful Algae 10:763-773

McCune B, Grace JB (2002) Analysis of Ecological Communities, Vol, Gleneden Beach, Oregon

Miki T, Takimoto G, Kagami M (2011) Roles of parasitic fungi in aquatic food webs: a theoretical approach. Freshwater Biology 56:1173-1183

Miquelis A, Rougier C, Pourriot R (1998) Impact of turbulence and turbidity on the grazing rate of the rotifer *Brachionus calyciflorus* (Pallas). Hydrobiologia 386:203-211

Moeseneder MM, Arrieta JM, Muyzer G, Winter C, Herndl GJ (1999) Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 65:3518-3525

Moon-van der Staay SY, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409:607-610

Mulholland MR, Gobler CJ, Lee C (2002) Peptide hydrolysis, amino acid oxidation, and nitrogen uptake in communities seasonally dominated by *Aureococcus anophagefferens*. Limnology and Oceanography 47:1094-1108

Myl'nikov AP, Karpov SA (2004) Review of diversity and taxonomy of cercomonads. Protistology 3 201-217

Nelson JD, Boehme SE, Reimers CE, Sherrell RM, Kerkhof LJ (2008) Temporal patternsof microbial community structure in the Mid-Atlantic Bight. Fems Microbiology Ecology 65

Noguez AM, Arita HT, Escalante AE, Forney LJ, Garcia-Oliva F, Souza V (2005) Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. Global Ecology and Biogeography 14:241-248

Norris RD, de Vargas C (2000) Evolution all at sea. Nature 405:23-24

Nuzzi R, Waters RA (2004) Long-term perspective on the dynamics of brown tide blooms in Long Island coastal bays. Harmful Algae 3:279-293

O'Kelly CJ, Sieracki ME, Thier EC, Hobson IC (2003) A transient bloom of *Ostreococcus* (Chlorophyta, Prasinophyceae) in West Neck Bay, Long Island, New York. Journal of Phycology 39:850-854

O'Malley MA (2007) The nineteenth century roots of 'everything is everywhere'. Nature Reviews Microbiology 5:647-651

Olsen PS (1989) Development and distribution of a brown-water algal bloom in Barnegat Bay, New Jersey with perspective on resources and other red tides in the region In: Cosper EM, Bricelj VM, Carpenter EJ (eds) Novel Phytoplankton Blooms: Causes and Impacts of Recurrent Brown Tides and Other Unusual Blooms, Book 35. Springer, New York

Osborn AM, Moore ERB, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environmental Microbiology 2:39-50

Park MG, Cooney SK, Yih W, Coats DW (2002) Effects of two strains of the parasitic dinoflagellate *Amoebophrya* on growth, photosynthesis, light absorption, and quantum yield of bloom-forming dinoflagellates. Marine Ecology-Progress Series 227:281-292

Park MG, Yih W, Coats DW (2004) Parasites and phytoplankton, with special emphasis on dinoflagellate infections. Journal of Eukaryotic Microbiology 51:145-155

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P, Tunlid A, Hagstrom A (2007) Global patterns of diversity and community structure in marine bacterioplankton. Molecular Ecology 16:867-880

Popels LC, Hutchins DA (2002) Factors affecting dark survival of the brown tide alga *Aureococcus anophagefferens* (Pelagophyceae). Journal of Phycology 38:738-744

Prasad A, Nienow JA (2006) The centric diatom genus *Cyclotella*, (Stephanodiscaceae : Bacillariophyta) from Florida Bay, USA, with special reference to *Cyclotella choctawhatcheeana* and *Cyclotella desikacharyi*, a new marine species related to the *Cyclotella striata* complex. Phycologia 45:127-140

Prasad A, Nienow JA, Livingston RJ (1990) The genus *Cyctotetla* from Choctawhatchee Bay, Florida with special reference to *C. striata* and *C. choctawhatcheeana sp.* nov. Phycologia 29:419-436

Probyn T, Pitcher G, Pienaar R, Nuzzi R (2001) Brown tides and mariculture in Saldanha Bay, South Africa. Marine Pollution Bulletin 42:405-408

Probyn TA, Bernard S, Pitcher GC, Pienaar RN (2010) Ecophysiological studies on *Aureococcus anophagefferens* blooms in Saldanha Bay, South Africa. Harmful Algae 9:123-133

Pustizzi F, MacIntyre H, Warner ME, Hutchins DA (2004) Interaction of nitrogen source and light intensity on the growth and photosynthesis of the brown tide alga *Aureococcus anophagefferens*. Harmful Algae 3:343-360

Putman RJ (1994) Community ecology, Vol. Chapman & Hall, London; New York

Rasconi S, Niquil N, Sime-Ngando T (2012) Phytoplankton chytridiomycosis: community structure and infectivity of fungal parasites in aquatic ecosystems. Environmental Microbiology 14:2151-2170

Robeson MS, King AJ, Freeman KR, Birky CW, Martin AP, Schmidt SK (2011) Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. Proceedings of the National Academy of Sciences of the United States of America 108:4406-4410

Ruan QS, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun FZ (2006) Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532-2538

Ryther JH (1954) The ecology of phytoplankton blooms in Moriches Bay and Great South Bay, Long Island, New York. Biologycal Bulletin 106:198-209

SCDHS (2011) Surface water quality monitoring data provided by the Suffolk County Department of Health Services Office of Ecology. In, Yaphank, N.Y.

Schubel JR, Bell TM, Carter HH (1991) The Great South Bay, Vol. State University Plaza, Albany, NY

Schutte UME, Abdo Z, Bent SJ, Shyu C, Williams CJ, Pierson JD, Forney LJ (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Applied Microbiology and Biotechnology 80:365-380

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498-2504

Sieburth JM, Johnson PW, Hargraves PW (1988) Ultrastructure and ecology of *Aureococcus anophagefferens* gen. et sp. nov. (Chrysopheceae): the dominant picoplankter during a bloom in Narragansett Bay, Rhode Island, summer 1985. Journal of Phycology 24:416–425

Sieracki ME (2001) The effects of microbial food web dynamics on the initiation of brown tide blooms. Brown Tide Research Initiative—report #6 3-4

Sieracki ME, Gobler CJ, Cucci TL, Thier EC, Gilg IC, Keller MD (2004) Pico- and nanoplankton dynamics during bloom initiation of *Aureococcus* in a Long Island, NY bay. Harmful Algae 3:459-470

Sieracki ME, Keller MD, Cucci TL, Their E (1999) Plankton community ecology during the bloom initiation period of the Brown Tide organism *Aureococcus anophagefferens* in coastal embayments of Long Island, N.Y. EOS 80:285

Smayda TJ (1990) Novel and nuisance phytoplankton blooms in the sea: evidence for a global epidemic. In: Graneli E, Sundstrom B, Edler L, Anderson DM (eds) Toxic Marine Phytoplankton. Elsevier, New York

Smayda TJ, Fofonoff P (1989) An extroadinary noxious "brown tide" in Narragansett Bay. II. Inimical effects. In: Okaichi T, Andersen DM, Nemoto T (eds) Red Tides: Biology, Environmental Science and Toxicology. Elsevier, New York, USA

Smayda TJ, Villareal TA (1989) The 1985 ‘brown tide’ and the open phytoplankton niche in Narragansett Bay during summer., Vol 35. Springer–Verlag, Berlin

Sokal RR, Rohlf FJ (1995) Biometry, Vol. W. H. Feeman & Co. , New York

Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow C-ET, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, Fuhrman JA (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. Isme Journal 5:1414-1425

Stentiford GD, Shields JD (2005) A review of the parasitic dinoflagellates Hematodinium species and *Hematodinium*-like infections in marine crustaceans. Diseases of Aquatic Organisms 66:47-70

Taylor GT, Gobler CJ, Sanudo-Wilhelmy SA (2006) Speciation and concentrations of dissolved nitrogen as determinants of brown tide *Aureococcus anophagefferens* bloom initiation. Marine Ecology Progress Series 312:67-83

Team RC (2012) R Program. In: R: A Language and Environment for Statistical Computing

Tiedje JM, Stahl DA (2002) Microbial ecology and genomics: a crossroads of opportunity. In: Tiedje JM, Stahl DA (eds) American Academy of Microbiology Critical Issues Colloquia Report. American Academy of Microbiology, Washington, DC, USA

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

Vieira MEC, Chant R (1993) On the contribution of subtidal volume fluxes to algal blooms in Long Island estuaries. Estuarine Coastal and Shelf Science 36:15-29

Vigil P, Countway PD, Rose J, Lonsdale DJ, Gobler CJ, Caron DA (2009) Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. Aquatic Microbial Ecology 54:83-100

Wang M, Karlsson C, Olsson C, Adlerberth I, Wold AE, Strachan DP, Martricardi PM, Aberg N, Perkin MR, Tripodi S, Coates AR, Hesselmar B, Saalman R, Molin G, Ahrne S (2008) Reduced diversity in the early fecal microbiota of infants with atopic eczema. Journal of Allergy and Clinical Immunology 121:129-134

Ward DM, Weller R, Bateson MM (1990) 16S ribosomal-RNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345:63-65

Wilson RE, Wong KC, Carter HH (1991) Aspects of circulation and exchange in Great South Bay. In: Schubel JR, Bell TM, Carter HH (eds) The Great South Bay. State University of New York Press, Albany

Wilson WH (1991) Sexual reproductive modes in polychaetes - classification and diversity. Bulletin of Marine Science 48:500-516

Woese CR (1987) Bacterial evolution. Microbiological Reviews 51:221-271

Wong KC, Wilson RE (1984) Observations of low frequency variability in Great South Bay and relations to atmospheric forcing. Journal of Physical Oceanography 14:1893-1900

Zingone A, Enevoldsen HO (2000) The diversity of harmful algal blooms: a challenge for science and management. Ocean & Coastal Management 43:725-748

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, Scanlan DJ (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environmental Microbiology 10:147-161

Appendix

Table A1. Putative identification of TRFs predicted from *in silico* digestion of 256 partial 18S rDNA sequences in 2008. Sequences were identified using BLAST in GenBank. 5 chimeric sequences were identified and are not included. *: interpretation is the same as sample ID **: brown tide bloom condition (c175B>15%)

Predicted TRF size (bases)	Putative ID (GenBank accession no. of BLAST hit)	Phylogenetic classification	% Similarity	Library ID *	# of clones	Cloned from bloom** sample?
78	*Skeletonema marinoi* (AJ632216)	Diatom	99	2B	4	YES
120	*Amoebophrya sp.* (AY208894)	Dinoflagellate	93	23F	1	YES
120	*Karlodinium micrum* (EF492506)	Dinoflagellate	99	2B	1	YES
120	Naked dinoflagellate (AM503930)	Dinoflagellate	99	13J	1	NO
125	Uncultured Ciliate (EU143864)	Ciliate	98	2B	1	YES
175	*Aureococcus anophagefferens* (AF118443)	Pelagophyte	100	21J	1	YES
177	*Aureococcus anophagefferens* (AF118443)	Pelagophyte	100	11J	5	YES
177	*Aureococcus anophagefferens* (AF118443)	Pelagophyte	100	23F	2	YES
177	*Aureococcus anophagefferens* (AF118443)	Pelagophyte	99	2B	4	YES
177	*Aureococcus anophagefferens* (AF118443)	Pelagophyte	100	2D	1	YES
183	*Pavlova salina* (L34669)	Haptophyte	98	12A	1	NO
184	*Cryothecomonas longipes* (AF290540)	Cercozoa	96	11J	1	YES
188	*Prorocentrum micans* (AY803739)	Dinoflagellate	98	19D	1	NO
205	Uncultured eukaryote (AY665065)	Uncultured Eukaryote	98	17A	1	NO
205	Uncultured eukaryote clone (AY665065)	Uncultured Eukaryote	98	19D	1	NO
205	Uncultured stramenopile (EU143919)	Uncultured stramenopile	99	2B	1	YES

205	Uncultured stramenopile (EU143919)	Uncultured stramenopile	95	2B	2	YES
212	Uncultured ciliate (GQ402482)	Ciliate	97	12A	1	NO
218	*Pirsonia verrucosa* (AJ561113)	Pirsonia	97	17A	6	NO
226	Uncultured Eimeriidae (EF024716)	Apicomplexa	91	21J	1	YES
226	Uncultured freshwater eukaryote (AB622340)	Uncultured Eukaryote	97	13J	1	NO
239	Uncultured syndiniales (EU793369.1)	Dinoflagellate	99	17A	1	NO
239	Uncultured syndiniales (EU793917.1)	Dinoflagellate	99	17A	1	NO
240	Uncultured syndiniales (EU793917.1)	Dinoflagellate	91	21J	9	YES
240	Uncultured syndiniales (EU793340.1)	Dinoflagellate	95	2B	1	YES
240	Uncultured syndiniales (EU793917.1)	Dinoflagellate	99	17A	1	NO
240	Uncultured syndiniales (EU793917.1)	Dinoflagellate	99	19D	5	NO
240	Uncultured syndiniales (EU793917.1)	Dinoflagellate	99	2B	1	YES
247	Uncultured cryptophyte (EU143940)	Cryptophyte	99	17A	1	NO
254	*Chrysochromulina parkeae* (AM490994)	Haptophyte	97	19D	1	NO
258	*Crustomastix stigmatica* (AJ629844)	Chlorophyte	97	13J	1	NO
258	Uncultured Emiliania (FJ537301)	Haptophyte	94	13J	1	NO
258	Uncultured Emiliania (FJ537301)	Haptophyte	97	2D	1	YES
259	*Chrysochromulina cymbium* (AM491018)	Haptophyte	99	13J	1	NO
259	*Chrysochromulina cymbium* (AM491018)	Haptophyte	99	19D	1	NO
259	*Chrysochromulina simplex* (AM491021)	Haptophyte	93	21J	1	YES
259	*Isochrysis sp.* (DQ079859)	Haptophyte	99	2D	1	YES
259	*Rhodotorula mucilaginosa* (X84326)	Yeast	99	11J	1	YES
259	Uncultured Ciliate (GU067973)	Ciliate	94	2B	1	YES
259	Uncultured marine eukaryote (DQ103802)	Uncultured Eukaryote	95	2B	1	YES
260	*Diacronema vlkianum* (AF106056)	Haptophyte	97	11J	1	YES
260	*Gymnodinium beii* (U41087)	Dinoflagellate	98	17A	1	NO

260	*Gymnodinium beii* (U41087)	Dinoflagellate	99	2B	1	YES
260	*Skeletonema dohrnii* (AJ632211)	Diatom	94	2B	1	YES
261	Uncultured fungus (EU144011)	Fungus	96	17A	1	NO
262	*Planktoniella sol* (AJ535173)	Diatom	98	19D	1	NO
262	Uncultured rhodophyte (AY343928)	Rhodophyte	98	17A	2	NO
263	Uncultured fungus (EU143997)	Fungus	98	13J	1	NO
264	*Chrysochromulina cymbium* (AM491018)	Haptophyte	94	23F	1	YES
264	*Heterocapsa triquetra* (AF022198)	Dinoflagellate	93	19D	1	NO
265	*Prorocentrum micans* (AY803739)	Dinoflagellate	91	2B	1	YES
266	Uncultured eukaryote (AY885023)	Uncultured Eukaryote	95	13J	1	NO
267	Uncultured ciliate (GQ402488)	Ciliate	96	12A	2	NO
267	Uncultured ciliate (GQ402488)	Ciliate	96	17A	1	NO
267	Uncultured eukaryote (DQ103797)	Uncultured Eukaryote	100	21J	1	YES
268	*Protaspis sp.* (FJ824125)	Cercozoa	97	2B	1	YES
268	Thaumatomonadida environmental sample (EF024026)	Cercozoa	91	17A	1	NO
268	Uncultured cercozoa (AY620328)	Cercozoa	93	17A	3	NO
269	*Cryothecomonas longipes* (AF290540)	Cercozoa	95	2D	1	YES
269	*Protaspis sp* (FJ824125)	Cercozoa	98	13J	1	NO
269	*Protaspis sp.* (FJ824125)	Cercozoa	98	11J	1	YES
270	*Allas sp.* (AY268040)	Cercozoa	96	23F	1	YES
270	*Cryothecomonas aestivalis* (AF290539)	Cercozoa	97	23F	1	YES
270	Uncultured cercozoan (AY620309)	Cercozoa	93	11J	1	YES
270	Uncultured marine picoeukaryote (FR874327)	Uncultured Eukaryote	98	19D	1	NO
271	Cercomonadida environmental sample (EF024287)	Cercozoa	89	17A	1	NO
271	*Paulinella chromatophora* (X81811)	Cercozoa	96	21J	1	YES

281	*Ichthyosporea sp.* (GU810144)	Fungi/metazoa	97	2B	1	YES
282	*Leucocryptos marina* (DQ980481)	Katablepharidophyte	96	12A	1	NO
282	Uncultured Katablepharidaceae (HM769614)	Katablepharidophyte	98	23F	1	YES
286	*Chaetoceros calcitrans*	Diatom	99	12A	1	NO
286	*Chaetoceros calcitrans*	Diatom	99	13J	3	NO
293	Uncultured marine Syndiniales (FN598302)	Dinoflagellate	89	17A	1	NO
294	Uncultured syndiniales (EU793266)	Dinoflagellate	92	13J	1	NO
295	*Amoebophrya sp.* (AF472553)	Dinoflagellate	99	19D	1	NO
295	*Amoebophrya sp.* (AF472553)	Dinoflagellate	96	2D	1	YES
295	Uncultured Syndiniales (FN598275)	Dinoflagellate	96	2B	1	YES
305	*Ostreococcus lucimarinus* (CP000592)	Chlorophyte	99	23F	1	YES
306	*Bathycoccus prasinos* (AY425314)	Chlorophyte	98	13J	1	NO
306	*Bathycoccus prasinos* (FN562453)	Chlorophyte	100	19D	4	NO
309	*Bathycoccus prasinos* (FN562453)	Chlorophyte	95	21J	1	YES
310	Uncultured marine eukaryote (AF363187)	Uncultured Eukaryote	97	17A	1	NO
317	*Amoebophrya sp.* (AY208893)	Dinoflagellate	91	13J	1	NO
317	Choanoflagellida sp. (EF432538)	Choanoflagellate	95	2D	1	YES
317	Uncultured eukaryote (EU333080)	Uncultured Eukaryote	88	12A	1	NO
317	Uncultured syndiniales (EU785285)	Dinoflagellate	95	17A	1	NO
318	*Oithona sp.* (JF781539.1)	Copepod	97	12A	1	NO
318	*Oithona sp.* (JF781539.1)	Copepod	97	19D	6	NO
318	*Oithona sp.* (JF781539.1)	Copepod	97	21J	1	YES
318	*Picocystis salinarum* (AF125167)	Chlorophyte	97	13J	1	NO
318	Uncultured fungus (AM114806)	Fungus	93	23F	1	YES
318	Uncultured marine Syndiniales (FN598457)	Dinoflagellate	97	17A	3	NO
318	Uncultured marine Syndiniales (FN598457)	Dinoflagellate	95	17A	2	NO

319	*Amoebophrya sp.* (AY208892)	Dinoflagellate	95	23F	1	YES
319	*Pfiesteria-like sp.* (AY590482)	Dinoflagellate	99	21J	1	YES
319	*Pyramimonas gelidicola* (EU141942)	Chlorophyte	95	2D	1	YES
320	*Picocystis salinarum* (AF153313)	Chlorophyte	99	11J	2	YES
320	*Picocystis salinarum* (AF153313)	Chlorophyte	99	12A	1	NO
320	*Picocystis salinarum* (AF153313)	Chlorophyte	98	13J	2	NO
320	*Picocystis salinarum* (AF153313)	Chlorophyte	95	19D	1	NO
320	*Picocystis salinarum* (AF153313)	Chlorophyte	99	23F	1	YES
321	Dinophyceae sp (AM408889)	Dinoflagellate	97	17A	1	NO
321	*Picocystis salinarum* (AF153313)	Chlorophyte	98	13J	2	NO
321	Uncultured fungus (AM114806)	Fungus	94	21J	1	YES
322	*Dactylopusia sp.* (EU380295)	Copepod	96	2B	1	YES
322	*Thraustochytrium multirudimentale* (AB022111)	Labyrinthulida	94	21J	1	YES
323	Uncultured eukaryote (AY885011)	Uncultured Eukaryote	94	2D	1	YES
325	*Brachionus calyciflorus* (GQ503607)	Rotifer	98	17A	1	NO
325	*Brachionus calyciflorus* (GQ503607)	Rotifer	98	19D	5	NO
325	*Heteromastus filiformis* (AF508118)	Polychaete	98	2D	1	YES
326	*Protaspis sp.* (FJ824125)	Cercozoa	94	2D	1	YES
327	*Pectinaria gouldii* (DQ790091)	Polychaete	99	19D	1	NO
327	*Pectinaria gouldii* (DQ790091)	Polychaete	99	2B	1	YES
327	Uncultured cercozoan (FN263033)	Cercozoa	96	11J	2	YES
337	*Sabellaria alveolata* (AY340442)	Polychaete	98	21J	1	YES
339	*Boccardiella ligerica* (AY527061)	Polychaete	97	21J	1	YES
339	*Hydroides pseudouncinatus* (DQ140403)	Polychaete	97	12A	2	NO
346	uncultured cryptophyte (FN690456)	Cryptophyte	95	2B	1	YES
351	*Anoplodactylus erectus* (DQ389934)	Sea spider	99	2D	1	YES

357	*Karlodinium micrum* (EF492506)	Dinoflagellate	96	2B	1	YES
374	*Picochlorum atomus* (AB080303)	Chlorophyte	97	13J	1	NO
374	*Protoperidinium minutum* (GQ227501)	Dinoflagellate	98	2B	1	YES
375	*Chlorella sp.* (Y14950)	Chlorophyte	97	13J	1	NO
375	*Dinophyceae sp.* (AY434686)	Dinoflagellate	98	19D	1	NO
375	*Dinophyceae sp.* (AY434686)	Dinoflagellate	97	19D	1	NO
375	*Gyrodinium cf. gutrula* (FN669511)	Dinoflagellate	100	21J	1	YES
375	*Picochlorum atomus* (AB080303)	Chlorophyte	99	11J	2	YES
375	*Picochlorum atomus* (AB080303)	Chlorophyte	100	12A	2	NO
375	*Picochlorum atomus* (AB080303)	Chlorophyte	100	13J	4	NO
375	*Picochlorum atomus* (AB080303)	Chlorophyte	99	17A	1	NO
375	*Picochlorum atomus* (AB080303)	Chlorophyte	99	19D	1	NO
375	*Picochlorum atomus* (AB080303)	Chlorophyte	99	21J	3	YES
375	*Picochlorum maculata* (AB080302)	Chlorophyte	99	12A	1	NO
375	*Picochlorum maculata* (AB080302)	Chlorophyte	97	13J	1	NO
375	*Picochlorum maculata* (AB080302)	Chlorophyte	99	17A	2	NO
375	*Picochlorum maculata* (AB080302)	Chlorophyte	97	19D	1	NO
375	*Nanochlorum eucaryotum* (X06425)	Chlorophyte	99	17A	1	NO
375	Uncultured *Woloszynskia* (GU067825)	Dinoflagellate	96	17A	1	NO
376	*Picochlorum maculata* (DQ779991)	Dinoflagellate	96	2B	1	YES
376	*Heterocapsa niei* (EF492499)	Dinoflagellate	99	23F	1	YES
376	*Prorocentrum triestinum* (DQ004734)	Dinoflagellate	97	12A	1	NO
376	*Prorocentrum triestinum* (DQ004734)	Dinoflagellate	99	13J	1	NO
376	*Prorocentrum triestinum* (DQ004734)	Dinoflagellate	99	17A	1	NO
376	*Prorocentrum triestinum* (DQ004734)	Dinoflagellate	95	23F	1	YES
376	Uncultured prasinophyte (AY425300)	Chlorophyte	98	13J	1	NO

376	*Woloszynskia cincta* (FR690459)	Dinoflagellate	99	23F	1	YES
376	*Woloszynskia cincta* (FR690459)	Dinoflagellate	99	2B	1	YES
377	Dinophyceae sp. (AY434686)	Dinoflagellate	98	21J	2	YES
377	*Karlodinium micrum* (EF492506)	Dinoflagellate	97	2B	6	YES
377	*Karlodinium micrum* (EF492506)	Dinoflagellate	100	2D	2	YES
377	*Scrippsiella sp.* (AM494499)	Dinoflagellate	99	19D	1	NO
377	*Thecate dinoflagellate* (AM503929)	Dinoflagellate	96	19D	1	NO
377	Uncultured marine dinoflagellate (FN598242)	Dinoflagellate	99	19D	1	NO
380	*Clymenura clypeata* (AF448152)	Polychaete	93	21J	1	YES
383	*Telonema subtilis* (AJ564772)	Telonema	97	17A	1	NO
522	*Cylindrotheca closterium* (DQ082742.1)	Diatom	97	19D	1	NO
522	*Cylindrotheca closterium* (DQ082742.1)	Diatom	99	19D	2	NO
523	*Cylindrotheca closterium* (DQ082742.1)	Diatom	99	19D	3	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	99	12A	2	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	99	13J	1	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	99	13J	1	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	100	13J	1	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	100	17A	1	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	100	17A	2	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	100	19D	1	NO
536	*Minutocellus polymorphus* (HQ912568)	Diatom	99	21J	2	YES
541	*Woloszynskia cincta* (FR690459)	Dinoflagellate	98	2B	1	YES
593	*Cyclotella choctawhatcheeana* (AM712618)	Diatom	98	2D	1	YES
594	*Minutocellus polymorphus* (HQ912568)	Diatom	97	17A	1	NO
595	*Woloszynskia cincta* (FR690459)	Dinoflagellate	99	2B	1	YES
595	*Woloszynskia cincta* (FR690459)	Dinoflagellate	99	2D	3	YES

595	*Woloszynskia cincta* (FR690459)	Dinoflagellate	99	2D	3	YES
596	*Minutocellus polymorphus* (HQ912568)	Diatom	97	17A	1	NO
596	*Woloszynskia cincta* (FR690459)	Dinoflagellate	95	2B	1	YES
597	*Woloszynskia cincta* (FR690459)	Dinoflagellate	95	2D	1	YES
598	*Alexandrium hiranoi* (AY641564)	Dinoflagellate	99	2B	1	YES

Table A2. Putative identification of TRFs predicted from *in silico* digestion of 99 partial 16S rDNA sequences in 2008. Sequences were identified using BLAST in GenBank 1 chimeric sequence was identified and was not included. *: interpretation is the same as sample ID **: brown tide bloom condition (c175B>15%)

Predicted TRF size (bases)	Putative ID (Genebank accession no. of BLAST hit)	Phylogenetic classification	% Similarity	Library ID *	# of clones	Cloned from bloom sample?
20	Uncultured Verrucomicrobia (DQ302114.1)	Verrucomicrobia	98	11J	1	No
101	Uncultured bacterium (HM128701.1)	Uncultured Bacteria	99	11J	10	No
124	Crenotrichaceae bacterium (FJ745018.1)	Bacteroidetes	99	4E	3	Yes
261	Uncultured cyanobacterium (EU641616.1)	Cyanobacteria	94	11J	1	No
272	Uncultured bacterium (FJ960270.1)	Uncultured Bacteria	100	11J	1	No
275	*Asterionellopsis glacialis* plastid (FJ002233.1)	Diatom plastid	97	4E	1	Yes
275	*Haslea crucigera* plastid (AF514849.1)	Diatom plastid	98	4E	1	Yes
277	Uncultured actinobacterium (EF471698.1)	Actinobacteria	99	4E	1	Yes
287	Uncultured alpha proteobacterium (GQ348825.1)	α-Proteobacteria	99	11J	1	No
292	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	4E	1	Yes
292	*Yonghaparkia sp.* (AM945590.1)	Actinobacteria	99	11J	2	No
293	*Marivita sp.* (GU137308.1)	α-Proteobacteria	99	11J	3	No
293	*Marivita sp.* (GU137308.1)	α-Proteobacteria	99	4E	1	Yes
293	Uncultured Rhodobacteraceae (JN625678.1)	α-Proteobacteria	99	11J	1	No
293	Uncultured Rhodobacteraceae (AY515451.1)	α-Proteobacteria	99	4E	1	Yes
293	*Marivita litorea* (NR_044513)	α-Proteobacteria	99	11J	3	No
293	Uncultured Rhodobacteraceae (HQ242427.1)	α-Proteobacteria	99	4E	1	Yes

293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	11J	3	No
293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	4E	3	Yes
294	Uncultured Rhodobacteraceae (HQ242427.1)	α-Proteobacteria	99	4E	1	Yes
300	*Aureococcus anophagefferens* plastid (GQ231541.1)	Pelagophyte plastid	100	11J	2	No
300	*Aureococcus anophagefferens* plastid (GQ231541.1)	Pelagophyte plastid	100	4E	10	Yes
300	Uncultured Bacteroidetes bacterium (FJ484372.1)	Bacteroidetes	91	11J	1	No
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	99	11J	1	No
374	Uncultured Bacteroidetes bacterium (FJ666172.1)	Bacteroidetes	94	4E	1	Yes
418	Uncultured bacterium (HM128701.1)	Uncultured Bacteria	99	11J	1	No
465	Uncultured bacterium (GU940997.1)	Uncultured Bacteria	99	4E	1	Yes
478	*Picocystis salinarum* plastid (FN563078.1)	Chlorophyte plastid	98	11J	3	No
479	*Synechococcus sp.* (FJ497740.1)	Cyanobacteria	99	11J	5	No
558	Uncultured SAR11 (JN591851.1)	α-Proteobacteria	99	4E	2	Yes
558	Uncultured SAR11 (JN591851.1)	α-Proteobacteria	96	11J	1	No
558	Uncultured alpha proteobacterium (HQ242001.1)	α-Proteobacteria	99	11J	1	No
559	Uncultured bacteria (EU237456.1)	Uncultured Bacteria	99	11J	1	No
559	Uncultured alpha proteobacterium (HQ242041.1)	α-Proteobacteria	99	11J	1	No
559	Uncultured alpha proteobacterium (HM057640.1)	α-Proteobacteria	99	4E	1	Yes
559	Uncultured alpha proteobacterium (FM242318.1)	α-Proteobacteria	99	11J	1	No

559	Unidentified alpha proteobacterium (U70679.2)	α-Proteobacteria	99	4E	1	Yes
560	*Pavlova lutheri* plastid (AF545628.1)	Haptophyte plastid	98	4E	1	Yes
560	Uncultured alpha proteobacterium (FJ429391.1)	α-Proteobacteria	99	11J	1	No
560	Uncultured alpha proteobacterium (FJ429391.1)	α-Proteobacteria	99	4E	1	Yes
562	Uncultured gamma proteobacterium (FJ745203.1)	γ-Proteobacteria	99	11J	1	No
564	*Pyramimonas parkeae* plastid (FN563104.1)	Chlorophyte plastid	95	4E	2	Yes
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	4E	1	Yes
568	uncultured *Microcella sp.* (EU703273.1)	Actinobacteria	99	11J	2	No
568	uncultured *Microcella sp.* (EU703273.1)	Actinobacteria	99	4E	5	Yes
573	Uncultured Bacteroidetes bacterium (AY580715.1)	Bacteroidetes	99	4E	1	Yes
575	Uncultured SAR406 cluster (EF471569.1)	Fibrobacteres/Acido-bacteria	96	11J	1	No
578	uncultured Bacteroidetes bacterium (EF471714.1)	Bacteroidetes	99	11J	5	No
578	Uncultured *Flavobacterium sp.* (GU230419.1)	Bacteroidetes	96	4E	1	Yes
585	Uncultured gamma proteobacterium (GU230314.1)	γ-Proteobacteria	99	4E	2	Yes
587	Uncultured delta proteobacterium (AY580412.1)	δ-Proteobacteria	87	11J	1	No

Table A3. Putative identification of TRFs predicted from *in silico* digestion of 351 partial 18S rDNA sequences in 2009. Sequences were identified using BLAST in GenBank 3 chimeric sequences were identified and were not included. *: interpretation is the same as sample ID

Predicted TRF size (bases)	Putative ID (GenBank accession no. of BLAST hit)	Phylogenetic classification	% Similarity	Library ID*	# of clones
69	Uncultured cercozoan (AY620338.1)	Cercozoa	99	13E	1
73	*Amoebophrya sp.* (AF472554.1)	Dinoflagellate	89	9E	1
73	*Amoebophrya sp.* (AY208892.1)	Dinoflagellate	92	9E	1
73	*Amoebophrya sp.* (AY208893.1)	Dinoflagellate	94	13E	1
82	Uncultured cercozoan (AY620330)	Cercozoa	99	9E	1
120	Uncultured alveolate (EU143871.1)	Alveolate	97	13E	1
126	*Strombidium* cf. basimorphum (FJ480419.1)	Ciliate	98	13E	5
126	*Strombidium* cf. basimorphum (FJ480419.1)	Ciliate	99	31F	2
126	*Strombidium* cf. basimorphum (JF791016.1)	Ciliate	99	5E	1
126	Uncultured ciliate (EU143873.1)	Ciliate	97	15J	1
126	Uncultured ciliate (EU143873.1)	Ciliate	96	3A	7
126	Uncultured ciliate (EU143873.1)	Ciliate	96	5E	2
126	Uncultured ciliate (EU143873.1)	Ciliate	96	7J	1
126	Uncultured ciliate (EU143877.1)	Ciliate	93	15J	1
126	Uncultured ciliate (EU143877.1)	Ciliate	96	3A	2
126	Uncultured ciliate (FJ939037.1)	Ciliate	99	5E	1
175	*Chlamydaster sterni* (AF534709.1)	Centroheliozoa	94	9J	1
176	*Harpacticus sp.* (EU380285.1)	Copepod	99	31F	1
176	*Harpacticus sp.* (EU380285.1)	Copepod	99	5E	1
176	*Harpacticus sp.* (EU380285.1)	Copepod	97	7J	10
181	*Telonema antarcticum* (AJ564773.1)	Telonema	99	1A	4
181	Uncultured marine eukaryote (EF526806.1)	Uncultured Eukaryote	95	5E	1
184	*Cryothecomonas longipes* (AF290540.1)	Cercozoa	96	9E	1
184	*Mataza hastifera* (AB558956.1)	Cercozoa	99	9J	1

191	Uncultured kathablepharid (EU143994.1)	Katablepharidophyte	94	9E	1
205	Uncultured eukaryote (AY665065.1)	Uncultured Eukaryote	99	15J	1
205	Uncultured marine eukaryote (FJ221511.1)	Uncultured Eukaryote	99	31F	1
213	Uncultured ciliate (GQ402482.1)	Ciliate	98	9E	1
239	Uncultured syndiniales (EU793340.1)	Dinoflagellate	97	15J	1
240	Uncultured syndiniales (EU793699.1)	Dinoflagellate	1	1A	1
240	Uncultured syndiniales (EU793917.1)	Dinoflagellate	99	15J	8
247	Uncultured cryptophyte (EU143940.1)	Cryptophyte	99	31F	1
257	Ciliate sp. (AM412525.1)	Ciliate	94	5E	1
258	*Chrysochromulina scutellum* (AJ246274.1)	Haptophyte	99	9E	1
258	*Novistrombidium orientale* (FJ422988.1)	Ciliate	92	1A	1
258	*Strombidium conicum* (FJ422992.1)	Ciliate	90	1A	1
258	Uncultured chlorophyte (FJ410565.1)	Chlorophyte	98	15J	1
258	Uncultured chlorophyte (FJ410565.1)	Chlorophyte	98	9E	1
258	Uncultured Oligotrichia (FN598239.1)	Ciliate	93	13E	1
258	*Varistrombidium kielum* (DQ811090.1)	Ciliate	92	1A	1
259	*Chrysochromulina simplex* (AM491021)	Haptophyte	100	2D	2
259	*Chrysochromulina strobilus* (FN599060.1)	Haptophyte	99	5E	1
259	*Isochrysis galbana* (HM149541.1)	Haptophyte	99	9E	1
259	Uncultured ciliate (GU067973.1)	Ciliate	93	13E	3
260	*Diacronema vlkianum* (AF106056.1)	Haptophyte	98	15J	1
260	*Gymnodinium beii* (U41087.1)	Dinoflagellate	99	9J	2
261	*Amastigomonas mutabilis* (AY050182.1)	Apusozoa	95	9E	1
262	*Planktoniella sol*	Diatom	99	5E	1
262	*Planktoniella sol*	Diatom	99	9E	2
262	Uncultured ciliate (GU067973.1)	Ciliate	98	13E	1
262	Uncultured Coscinodiscophyceae (FN598276.1)	Diatom	98	9E	1
262	Uncultured rhodophyte (AY343928.1)	Rhodophyte	98	9E	1
263	Uncultured cercozoan (JF698746.1)	Cercozoa	98	9E	1
264	*Vampyrella pendula* (HE609035.1)	Cercozoa	92	9E	1

266	*Ochromonas sp.* (U42381.1)	Chrysophyte	97	9E	1
267	Thaumatomastigidae environmental sample (EF023480.1)	Cercozoa	94	13E	1
267	Thaumatomastigidae environmental sample (EF023480.1)	Cercozoa	94	31F	1
267	Uncultured cercozoa (AY620350.1)	Cercozoa	93	7J	1
267	Uncultured cercozoan (AY620338.1)	Cercozoa	99	31F	1
267	Uncultured marine cercozoan (HM997295.1)	Cercozoa	98	5E	1
267	Uncultured marine eukaryote (EF526806.1)	Uncultured Eukaryote	99	13E	2
267	Uncultured marine eukaryote (EF526806.1)	Uncultured Eukaryote	99	5E	1
267	Uncultured marine eukaryote (EF526806.1)	Uncultured Eukaryote	100	7J	3
268	Choanoflagellida sp. (EF432538.1)	Choanoflagellate	99	31F	1
268	*Duboscquella sp.* (AB295041.1)	Dinoflagellate	99	9E	1
268	Uncultured cercozoan (AY620328.1)	Cercozoa	93	9E	3
268	Uncultured cercozoan (AY620355.1)	Cercozoa	96	9E	1
269	Cercozoa sp. (FJ824126.1)	Cercozoa	96	31F	1
269	Cercozoa sp. (FJ824126.1)	Cercozoa	96	9E	1
269	*Cryothecomonas longipes* (AF290540.1)	Cercozoa	93	7J	1
269	*Cryothecomonas longipes* (AF290540.1)	Cercozoa	96	9E	1
269	*Paulinella chromatophora* (X81811.1)	Cercozoa	97	7J	1
269	*Protaspis sp.* (FJ824125.1)	Cercozoa	98	9E	1
269	Uncultured cercozoa (AY620350.1)	Cercozoa	98	9E	1
269	Uncultured cercozoa (AY620336.1)	Cercozoa	96	31F	1
269	Uncultured cercozoa (FN263033.1)	Cercozoa	94	31F	1
270	*Allas sp.* (AY268040.1)	Cercozoa	97	9E	1
282	*Heteromastus filiformis* (AF508118.1)	Polychaete	99	1A	3
282	Uncultured Katablepharidaceae (HM769614.1)	Katablepharidophyte	99	5E	1
282	Uncultured Katablepharidaceae (HM769614.1)	Katablepharidophyte	99	9E	2
283	*Polygordius jouinae* (DQ153064)	Polychaete	99	2D	5
283	*Polygordius jouinae* (DQ153064)	Polychaete	96	3A	1
284	*Polygordius jouinae* (DQ153064)	Polychaete	99	1A	1
286	*Chaetoceros calcitrans* (EU240880.1)	Diatom	99	7J	3

286	*Chaetoceros calcitrans* (EU240880.1)	Diatom	99	9E	2
291	Uncultured labyrinthulid (JF791020.1)	Labyrinthulida	94	5E	1
293	Uncultured marine Syndiniales (FN598302.1)	Dinoflagellate	89	31F	1
294	*Amoebophrya sp.* (AY208893.1)	Dinoflagellate	91	5E	1
294	Uncultured syndiniale (EU793266.1)	Dinoflagellate	93	5E	1
294	Uncultured syndiniale (EU793266.1)	Dinoflagellate	93	7J	1
294	Uncultured syndiniale (EU793266.1)	Dinoflagellate	93	9E	1
294	Uncultured syndiniale (EU793266.1)	Dinoflagellate	93	9J	1
295	*Amoebophrya sp.* (AY208893.1)	Dinoflagellate	91	7J	1
295	*Amoebophrya sp.* (AY208893.1)	Dinoflagellate	91	9E	1
295	*Amoebophrya sp.* (AY208893.1)	Dinoflagellate	91	9J	1
296	*Sinocalanus sinensis* (GU969142.1)	Copepod	97	2D	2
299	*Pseudodiaptomus annandalei* (AY629258.1)	Copepod	95	3A	5
299	*Pseudodiaptomus annandalei* (AY629258.1)	Copepod	95	7J	1
301	Uncultured stramenopile (JQ782066.1)	Uncultured Stramenopile	99	9E	1
303	*Capitella capitata* (U67323.1)	Polychaete	96	1A	1
304	*Ostreococcus tauri* (Y15814.1)	Chlorophyte	94	5E	1
306	*Bathycoccus prasinos* (FN562453.1)	Chlorophyte	96	2D	4
306	*Bathycoccus prasinos* (FN562453.1)	Chlorophyte	96	13E	1
306	*Bathycoccus prasinos* (FN562453.1)	Chlorophyte	96	9E	2
308	*Bathycoccus prasinos* (FN562453.1)	Chlorophyte	96	2D	1
309	*Mantoniella squamata* (X73999)	Chlorophyte	98	2D	1
312	Ciliate sp. (AM412525.1)	Ciliate	96	3A	1
314	*Pentapharsodinium tyrrhenicum* (JF790993.1)	Dinoflagellate	92	15J	1
317	*Amoebophrya sp.* (AY775285.1)	Dinoflagellate	92	5E	1
317	*Amoebophrya sp.* (HM483395.1)	Dinoflagellate	92	13E	2
317	*Oithona sp.* (JF781539.1)	Copepod	97	9E	24
317	*Pentapharsodinium tyrrhenicum* (JF790993.1)	Dinoflagellate	92	9E	1
317	*Symbiodinium sp.* (U10893.1)	Dinoflagellate	88	9J	1

317	Uncultured dinoflagellate (GU067963.1)	Dinoflagellate	87	31F	1
318	Choanoflagellida sp. (EF432538.1)	Choanoflagellate	98	31F	6
318	*Oithona sp.* (JF781539.1)	Copepod	97	13E	7
318	*Oithona sp.* (JF781539.1)	Copepod	97	31F	2
318	*Oithona sp.* (JF781539.1)	Copepod	96	5E	3
318	*Oithona sp.* (JF781539.1)	Copepod	97	9J	18
318	*Oithona sp.* JF781539.1)	Copepod	97	9E	1
318	*Pseudotontonia simplicidens* (JF791039.1)	Ciliate	96	9E	6
318	*Pseudotontonia simplicidens* (JF791039.1)	Ciliate	93/99	9J	2
318	*Pyramimonas gelidicola* (EU141942.1)	Chlorophyte	99	31F	1
318	*Pyramimonas gelidicola* (EU141942.1)	Chlorophyte	98	9E	3
320	*Picochlorum maculata* (AB080302.1)	Chlorophyte	94	5E	1
320	*Picocystis salinarum* (AF153313.1)	Chlorophyte	99	15J	1
320	Uncultured fungus (GU067817.1)	Fungus	98	5E	1
321	Dinophyceae sp (AM408889.1)	Dinoflagellate	98	13E	2
321	Dinophyceae sp (AM408889.1)	Dinoflagellate	99	5E	1
321	Dinophyceae sp (AM408889.1)	Dinoflagellate	99	9E	1
321	*Woloszynskia cincta* (FR690459.1)	Dinoflagellate	96	31F	1
322	*Pseudodiaptomus annandalei* (AY629258.1)	Copepod	95	3A	1
322	Uncultured fungus (EU144007.1)	Fungus	99	9E	1
323	Uncultured cercozoan (EU143884.1)	Cercozoa	91	9E	1
323	Uncultured fungus (EU144007.1)	Fungus	99	13E	1
324	*Pseudobodo tremulans*	Bicosoecida	94	15J	1
324	Uncultured eukaryote (JF273982.1)	Uncultured Eukaryote	99	9E	1
325	*Heteromastus filiformis* (AF508118.1)	Polychaete	99	3A	2
325	Uncultured stramenopile (JQ782065.1)	Uncultured Stramenopile	99	2D	1
326	*Cryothecomonas aestivalis* (AF290539.1)	Cercozoa	96	9E	1
326	*Protaspis sp.* (FJ824125.1)	Cercozoa	93	9E	1
327	*Cryothecomonas aestivalis* (AF290539.1)	Cercozoa	99	15J	1

327	*Pectinaria gouldii* (DQ790091.1)	Polychate	99	1A	4
327	Uncultured cercozoan (FN263033.1)	Cercozoa	97	31F	2
327	Uncultured cercozoan (FN263033.1)	Cercozoa	98	9E	2
329	*Hormathiidae gen.* (AF052890.1)	Cnidarian	99	15J	3
329	*Hormathiidae gen.* (AF052890.1)	Cnidarian	98	31F	1
329	*Pirsonia verrucosa* (AJ561113.1)	Pirsonia	95	3A	1
330	*Eteone longa* (AF448155.1)	Polychaete	98	15J	1
339	*Boccardiella ligerica* (AY527061.1)	Polychaete	98	3A	1
340	*Hydroides pseudouncinatus* (DQ140403.1)	Polychaete	98	7J	1
365	*Pseudopedinella elastica* (U14387.1)	Dictyochophyte	96	3A	1
367	*Codonellopsis nipponica* (FJ196072.1)	Ciliate	94	9E	1
372	*Bolidomonas pacifica* (HQ912557.1)	Heterokontophyta	91	9E	2
372	Uncultured chlorophyte (EU143974.1)	Chlorophyte	93	9E	1
375	*Picochlorum atomus* (AB080303.1)	Chlorophyte	98	15J	1
375	*Picochlorum atomus* (AB080303.1)	Chlorophyte	98	5E	1
375	*Picochlorum atomus* (AB080303.1)	Chlorophyte	98	9E	4
375	*Picochlorum maculata* (AB080302.1)	Chlorophyte	99	9E	3
376	*Bolidomonas pacifica* (AF167154)	Heterokontophyta	98	2D	1
376	*Gymnodinium aureolum* (DQ779991.1)	Dinoflagellate	97	1A	1
376	*Gymnodinium aureolum* (DQ779991.1)	Dinoflagellate	97	3A	1
376	*Gymnodinium aureolum* (DQ779991.1)	Dinoflagellate	95	5E	3
376	*Nannochloropsis salina* (AF045050.1)	Eustigmatophyte	98	13E	1
376	Uncultured fungus (GU067958.1)	Fungus	95	13E	1
376	Uncultured marine dinoflagellate (FN598290.1)	Dinoflagellate	99	1A	3
376	*Woloszynskia cincta* (FR690459.1)	Dinoflagellate	99	3A	1
377	Dinophyceae sp. (AY434686.1)	Dinoflagellate	99	5E	1
377	Dinophyceae sp. (AY434687.1)	Dinoflagellate	98	1A	3
377	Dinophyceae sp. (AY590479.1)	Dinoflagellate	99	9J	1
377	*Gymnodinium catenatum* (DQ779990.2)	Dinoflagellate	99	7J	1
377	*Gymnodinium sp.* (AF022196.1)	Dinoflagellate	99	31F	1

377	*Karlodinium micrum* (EF492506.1)	Dinoflagellate	99	15J	1
377	*Karlodinium micrum* (EF492506.1)	Dinoflagellate	100	9E	1
377	*Polykrikos kofoidii* (AB466293.1)	Dinoflagellate	100	15J	2
377	*Scrippsiella trochoidea* (EF492513.1)	Dinoflagellate	97	5E	2
377	*Takayama cf. pulchellum* (AY800130.1)	Dinoflagellate	93	31F	1
377	Uncultured marine dinoflagellate (FN598242.1)	Dinoflagellate	99	1A	1
377	Uncultured marine dinoflagellate (FN598242.1)	Dinoflagellate	99	31F	1
378	*Gymnodinium galatheanum* (AF172712.1)	Dinoflagellate	99	9E	1
379	*Stylochus zebra* (AF315604.1)	Flat worm	97	1A	3
381	*Telonema subtilis* (AJ564772.1)	Telonema	99	31F	2
496	Uncultured eukaryote (AY885011.1)	Uncultured Eukaryote	100	3A	1
496	Uncultured eukaryote (AY885011.1)	Uncultured Eukaryote	100	5E	6
522	*Peridinium balticum* (Y10566.2)	Dinoflagellate	99	3A	1
523	*Navicula sp.* (EU090030.1)	Diatom	99	5E	1
532	*Cyclotella choctawhatcheeana* (AM712618.1)	Diatom	99	31F	1
532	*Cyclotella choctawhatcheeana* (AM712618.1)	Diatom	99	3A	2
532	*Cyclotella choctawhatcheeana* (AM712618.1)	Diatom	99	9E	1
534	Uncultured eukaryote (AY885011.1)	Uncultured Eukaryote	97	3A	1
536	*Minutocellus polymorphus* (HQ912568.1)	Diatom	99	9E	1
595	*Lepidodinium viride* (JF791033.1)	Dinoflagellate	98	1A	1
595	*Woloszynskia cincta* (FR690459.1)	Dinoflagellate	99	3A	1
596	*Picochlorum atomus* (AB080303.1)	Chlorophyte	98	9E	1
596	*Takayama* cf. pulchellum (AY800130.1)	Dinoflagellate	95	31F	1

Table A4. Putative identification of TRFs predicted from *in silico* digestion of 265 partial 16S rDNA sequences in 2009. Sequences were identified using BLAST in GenBank; 1 chimeric sequence was identified and was not included. *: interpretation is the same as sample ID

Predicted TRF size (bases)	Putative ID (GenBank accession no. of BLAST hit)	Phylogenetic classification	% Similarity	Library ID*	# of clones
18	Uncultured Lentisphaerae bacterium (GQ350847.1)	Chlamydiae/Verrucomicrobia	96	129P	1
18	Uncultured Verrucomicrobium sp. (FM242338.1)	Chlamydiae/Verrucomicrobia	98	129P	3
18	Uncultured Verrucomicrobia bacterium (GQ249493.1)	Chlamydiae/Verrucomicrobia	97	441P	1
20	Uncultured Verrucomicrobia bacterium (FJ902261.1)	Chlamydiae/Verrucomicrobia	99	171P	1
68	*Pseudorhodobacter ferrugineus* (AY701453.1)	α-Proteobacteria	99	81P	1
99	Uncultured delta proteobacterium (EF471640.1)	δ-Proteobacteria	99	149P	1
99	Bacteriovorax sp. (DQ631734.1)	δ-Proteobacteria	94	339P	1
99	Uncultured beta proteobacterium (JN233149.1)	β-Proteobacteria	99	339P	1
101	Uncultured bacterium (HM128701.1)	Uncultured bacteria	99	171P	2
101	Uncultured bacterium (HM128701.1)	Uncultured bacteria	99	339P	2
102	Uncultured cyanobacterium (HQ242079.1)	Cyanobacteria	98	81P	1
122	Uncultured beta proteobacterium (EF471531.1)	β-Proteobacteria	100	129P	1
122	Uncultured beta proteobacterium (EF471531.1)	β-Proteobacteria	95	149P	1
124	Uncultured Sphingobacteriales bacterium (HQ692020.1)	Bacteroidete	97	171P	1
124	Uncultured Sphingobacteriales cum 'Crenotrichaceae' bacterium (FJ745018)	Bacteroidete or γ-Proteobacteria	100	171P	1
145	Uncultured Sphingobacteriales bacterium (HQ692025.1)	Bacteroidete	94	149P	1
147	Uncultured Flavobacterium sp. (FJ745055.1)	Bacteroidete	93	81P	1
237	Uncultured cyanobacterium (HQ821701.1)	Cyanobacteria	99	1P	1

259	Uncultured delta proteobacterium (JQ579792.1)	δ-Proteobacteria	94	81P	1
262	*Arcocellulus mammifer* plastid (FJ002193.1)	Diatom plastid	99	441P	1
263	Uncultured bacterium (GU940749.1)	Uncultured bacteria	99	129P	1
263	Uncultured bacterium (GU940749.1)	Uncultured bacteria	99	149P	2
268	Uncultured actinobacterium (JN591931.1)	Actinobacteria	99	171P	1
272	Uncultured alpha proteobacterium (GQ472826.1)	α-Proteobacteria	99	129P	1
272	Uncultured gamma proteobacterium (HM057720.1)	γ-Proteobacteria	99	339P	1
272	Uncultured gamma proteobacterium (EU350886.1)	γ-Proteobacteria	100	441P	1
274	Uncultured delta proteobacterium (HM057716.1)	δ-Proteobacteria	100	339P	1
275	*Coscinodiscus radiatus* plastid (AJ536462.1)	Diatom plastid	99	441P	1
275	*Chaetoceros calcitrans* plastid (FJ002215.1)	Diatom plastid	99	81P	1
277	Uncultured actinobacterium (JX011174.1)	Actinobacteria	100	149P	1
277	Uncultured actinobacterium (JX011174.1)	Actinobacteria	100	171P	1
277	Uncultured Actinomycetales (FJ744987.1)	Actinobacteria	99	171P	1
289	Uncultured bacterium (HM127595.1)	Uncultured bacteria	98	171P	1
289	Uncultured *Synechococcus sp.* (EF598931.1)	Cyanobacteria	92	171P	1
289	*Synechococcus sp.* (FJ497737.1)	Cyanobacteria	99	81P	4
292	*Chaetoceros calcitrans* plastid (FJ002215.1)	Diatom plastid	92	81P	1
293	*Marivita sp.* (HQ871858.1)	α-Proteobacteria	100	129P	2
293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	129P	1
293	Uncultured Rhodobacteraceae bacterium (AY515451.1)	α-Proteobacteria	99	129P	1
293	Uncultured Rhodobacteraceae bacterium (JN591912.1)	α-Proteobacteria	99	129P	3
293	Uncultured Rhodobacteraceae bacterium (JN625678.1)	α-Proteobacteria	99	149P	1
293	Uncultured alpha proteobacterium (AB491826.1)	α-Proteobacteria	94	171P	1

293	Uncultured Rhodobacteraceae bacterium (AY515451.1)	α-Proteobacteria	99	171P	1
293	Uncultured Rhodobacteraceae bacterium (DQ234245.2)	α-Proteobacteria	99	171P	1
293	Uncultured Rhodobacteraceae bacterium (JN591912.1)	α-Proteobacteria	99	171P	1
293	*Yonghaparkia sp.* (AM945590.1)	Actinobacteria	99	171P	1
293	*Marivita sp.* (HQ871858.1)	α-Proteobacteria	99	1P	2
293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	1P	3
293	*Loktanella sp.* (JN699161.1)	α-Proteobacteria	100	339P	1
293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	339P	6
293	Uncultured Roseobacter sp. (EU600639.1)	α-Proteobacteria	99	339P	1
293	*Marivita litorea* (NR_044513.1)	α-Proteobacteria	100	441P	1
293	*Marivita sp.* (HQ871858.1)	α-Proteobacteria	99	441P	1
293	Uncultured alpha proteobacterium (EF471665.1)	α-Proteobacteria	99	441P	1
293	Uncultured Rhodobacteraceae bacterium (FN582321.1)	α-Proteobacteria	99	441P	1
293	Uncultured Rhodobacteraceae bacterium (HQ242302.1)	α-Proteobacteria	99	441P	1
293	Uncultured Rhodobacteraceae bacterium (JN591912.1)	α-Proteobacteria	99	441P	1
293	Uncultured *Sulfitobacter sp.* (AY664060.1)	α-Proteobacteria	93	441P	1
293	*Marivita sp.* (GU137308.1)	α-Proteobacteria	99	81P	3
293	Uncultured alpha proteobacterium (EF471665.1)	α-Proteobacteria	99	81P	3
293	Uncultured gamma proteobacterium (EU328105.1)	γ-Proteobacteria	91	81P	1
293	Uncultured Rhodobacteraceae (JN591912.1)	α-Proteobacteria	99	81P	2
293	Uncultured Rhodobacteraceae bacterium (HQ242427.1)	α-Proteobacteria	99	81P	1
293	Uncultured Rhodobacteraceae bacterium (JN591912.1)	α-Proteobacteria	99	81P	2

300	Uncultured Sphingobacteriales bacterium (HQ692025.1)	Bacteroidete	91	171P	1
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	99	129P	2
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	99	149P	1
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	98	171P	1
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	99	1P	1
307	Uncultured gamma proteobacterium (EU703314.1)	γ-Proteobacteria	99	81P	1
313	Uncultured Sphingobacteria bacterium (AM279196.1)	Bacteroidete	99	129P	1
313	Uncultured organism (JN453936.1)	Uncultured bacteria	91	1P	1
317	*Fluviicola sp.* (AB517714.1)	Bacteroidete	96	441P	4
317	Uncultured Bacteroidetes bacterium (AB266004.1)	Bacteroidete	97	441P	1
317	Uncultured Bacteroidetes bacterium (AM930460.1)	Bacteroidete	95	441P	1
318	Uncultured proteobacterium (GU323610.1)	Uncultured proteobacterium	98	149P	1
318	Uncultured proteobacterium (GU323610.1)	Uncultured proteobacterium	98	171P	1
374	Uncultured Flavobacteria bacterium (FN433428.1)	Bacteroidete	99	81P	1
379	Uncultured Rhodobacteraceae bacterium (DQ421650.2)	α-Proteobacteria	95	441P	1
400	Uncultured cyanobacterium (GU552226.1)	Cyanobacteria	94	171P	1
401	Uncultured cyanobacterium (EU276548.1)	Cyanobacteria	94	171P	1
404	Uncultured gamma proteobacterium (EF471467.1)	γ-Proteobacteria	99	129P	1
404	Uncultured gamma proteobacterium (FJ745205.1)	γ-Proteobacteria	97	129P	1
462	Uncultured SAR11 (JN547466.1)	α-Proteobacteria	94	1P	1
477	Uncultured crenarchaeote (AY454657.1)	Archae	94	171P	1
479	*Synechococcus sp.* (FJ497740.1)	Cyanobacteria	100	129P	2
479	*Synechococcus sp.* (JF306683.1)	Cyanobacteria	99	129P	1
479	*Synechococcus sp.* (JQ927353.1)	Cyanobacteria	99	129P	1

479	Uncultured *Synechococcus sp.* (EF471564.1)	Cyanobacteria	99	129P	2
479	*Synechococcus sp.* (FJ497740.1)	Cyanobacteria	100	149P	1
479	Uncultured *Synechococcus sp.* (EF471564.1)	Cyanobacteria	99	149P	2
479	*Synechococcus sp.* (AF448061.1)	Cyanobacteria	99	171P	1
479	Uncultured *Synechococcus sp.* (EF471564.1)	Cyanobacteria	99	171P	4
479	Cyanobacterium UCYN-A (CP001842.1)	Cyanobacteria	99	339P	1
479	*Synechococcus sp.* (FJ497740.1)	Cyanobacteria	100	339P	1
479	*Synechococcus sp.* (JF306683.1)	Cyanobacteria	99	339P	2
479	Uncultured cyanobacterium (GQ349093.1)	Cyanobacteria	92	339P	1
479	Uncultured *Synechococcus sp.* (EF471564.1)	Cyanobacteria	99	339P	1
479	*Synechococcus sp.* (JF306660.1)	Cyanobacteria	100	441P	1
479	*Synechococcus sp.* (JF306683.1)	Cyanobacteria	99	441P	6
479	*Synechococcus sp.* (JQ927353.1)	Cyanobacteria	99	441P	3
479	*Synechococcus sp.* (JQ927353.1)	Cyanobacteria	99	81P	1
479	Uncultured *Synechococcus sp.* (EF471564.1)	Cyanobacteria	99	81P	5
480	Uncultured Sphingobacteria (JN233042.1)	Bacteroidete	99	129P	1
482	Uncultured Sphingobacteria (JN233042.1)	Bacteroidete	99	149P	2
482	Uncultured Bacteroidetes bacterium (JQ580307.1)	Bacteroidete	98	149P	1
482	Uncultured Bacteroidetes (EF471480.1)	Bacteroidete	99	171P	1
482	Uncultured Bacteroidetes (EU703423.1)	Bacteroidete	92	171P	2
482	Uncultured Sphingobacteria (JN233042.1)	Bacteroidete	99	441P	1
482	Uncultured Sphingobacteria (JN233042.1)	Bacteroidete	99	81P	1
483	Uncultured *Lewinella sp.* (EU878147.1)	Bacteroidete	95	149P	1
483	Uncultured *Lewinella sp.* (EU878147.1)	Bacteroidete	95	81P	4
488	Uncultured Flavobacterium sp. (FJ744962.1)	Bacteroidete	99	441P	1

495	Uncultured delta proteobacterium (JQ579792.1)	δ-Proteobacteria	90	149P	3
557	SAR11 (JN591851.1)	α-Proteobacteria	99	129P	1
558	SAR11 (HQ242440.1)	α-Proteobacteria	99	129P	2
558	SAR11 (JN591858.1)	α-Proteobacteria	99	129P	1
558	SAR11 (JN591851.1)	α-Proteobacteria	99	149P	2
558	SAR11 (JN591851.1)	α-Proteobacteria	99	171P	3
558	SAR11 (HQ242440.1)	α-Proteobacteria	99	171P	1
558	SAR11 (JN591851.1)	α-Proteobacteria	99	1P	5
558	SAR11 (JN233111.1)	α-Proteobacteria	99	339P	1
558	SAR11 (JN591858.1)	α-Proteobacteria	99	441P	1
558	SAR11 (JN591851.1)	α-Proteobacteria	99	81P	2
559	SAR11 (JN591862.1)	α-Proteobacteria	99	441P	1
559	*Azospirillum sp.* (AB545654.1)	α-Proteobacteria	91	171P	1
559	Uncultured alpha proteobacterium (HQ163325.1)	α-Proteobacteria	100	1P	1
559	Uncultured alpha proteobacterium (FR647510.1)	α-Proteobacteria	99	81P	2
559	Uncultured alpha proteobacterium (HQ163325.1)	α-Proteobacteria	100	81P	1
560	Alpha proteobacterium (DQ486492.1)	α-Proteobacteria	99	129P	1
560	*Azospirillum sp.* (AB545654.1)	α-Proteobacteria	91	129P	1
560	*Roseibacterium elongatum* (AB601471.1)	α-Proteobacteria	93	149P	1
560	Uncultured alpha proteobacterium (EF215757.1)	α-Proteobacteria	99	149P	1
560	Alpha proteobacterium (DQ486492.1)	α-Proteobacteria	99	1P	1
561	Uncultured Sphingobacteria bacterium (JN233042.1)	Bacteroidete	99	149P	1
561	*Nannochloris sp.* plastid (AY702135.1)	Chlorophyte plastid	95	81P	1
564	Uncultured cyanobacterium (HQ242646.1)	Cyanobacteria	100	81P	1
564	Uncultured Oscillatoriales cyanobacterium	Cyanobacteria	95	81P	1

	(FJ790639.1)				
568	Uncultured Actinomycetales (FJ744815.1)	Actinobacteria	100	129P	1
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	129P	1
568	Uncultured Actinomycetales (FJ744815.1)	Actinobacteria	99	149P	1
568	Microbacteriaceae bacterium (FJ214966.1)	Actinobacteria	99	171P	1
568	Uncultured Actinomycetales (FJ744815.1)	Actinobacteria	99	171P	1
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	171P	1
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	1P	1
568	Microbacteriaceae bacterium (FJ214966.1)	Actinobacteria	99	339P	2
568	Uncultured *Microcella sp.*(EU703273.1)	Actinobacteria	99	339P	2
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	441P	2
568	Uncultured *Microcella sp.* (EU703273.1)	Actinobacteria	99	441P	6
568	Uncultured *Microcella sp.* (EU703410.1)	Actinobacteria	99	81P	1
568	Uncultured *Microcella sp.* (EU703273.1)	Actinobacteria	99	81P	5
570	Uncultured Bacteroidetes bacterium (JQ580307.1)	Bacteroidete	90	129P	1
570	Uncultured Bacteroidetes bacterium (JQ580307.1)	Bacteroidete	90	339P	1
574	Uncultured Flavobacteria bacterium (DQ289539.1)	Bacteroidete	93	441P	1
577	Uncultured Bacteroidetes/Chlorobi (DQ431894.1)	Bacteroidete	95	171P	1
578	Uncultured Flavobacteriales bacterium (EU328017.1)	Bacteroidete	97	129P	1
578	Uncultured Bacteroidetes bacterium (EF471714.1)	Bacteroidete	99	149P	4
578	*Cyclobacterium linum* (EU419941.1)	Cyanobacteria	100	171P	1
578	Uncultured Bacteroidetes bacterium (EF471714)	Bacteroidete	99	171P	1
578	Uncultured Flavobacteriaceae bacterium (FJ745043.1)	Bacteroidete	95	171P	1
578	*Algoriphagus sp.* (FJ196000.1)	Bacteroidete	98	1P	1
578	Uncultured *Cytophaga sp.* (AB015587.1)	Bacteroidete	97	339P	1

578	*Winogradskyella sp.* (HM368527.1)	Bacteroidete	99	339P	1
578	*Algoriphagus sp.* (FJ196000.1)	Bacteroidete	98	441P	2
578	Uncultured *Cytophaga sp.* (EF657862.1)	Bacteroidete	96	441P	1
579	Uncultured Flavobacteria bacterium (EU600658.1)	Bacteroidete	99	129P	1
579	Uncultured Flavobacteria bacterium (EU600658.1)	Bacteroidete	99	81P	2
581	*Flavobacterium sp.* (AF321038.2)	Bacteroidete	98	81P	1
581	Uncultured beta proteobacterium (DQ234170.2)	β-Proteobacteria	92	81P	1
584	*Arenicella xantha* (AB500096.1)	γ-Proteobacteria	97	1P	1
585	*Arenicella xantha* (AB500096.1)	γ-Proteobacteria	97	441P	3
585	*Arenicella xantha* (AB500096.1)	γ-Proteobacteria	97	81P	1
586	Uncultured gamma proteobacterium (FJ745253.1)	γ-Proteobacteria	99	129P	1
586	Uncultured gamma proteobacterium (AY580743.1)	γ-Proteobacteria	99	1P	1
586	*Spongiibacter sp.* (AB526337.1)	γ-Proteobacteria	99	81P	2
587	Uncultured beta proteobacterium (HM057738.1)	β-Proteobacteria	100	171P	2
587	Uncultured beta proteobacterium (HM057738.1)	β-Proteobacteria	100	441P	1
587	Uncultured gamma proteobacterium (GQ850579.1)	γ-Proteobacteria	93	441P	1
588	Uncultured delta proteobacterium (JF319278.1)	δ-Proteobacteria	90	339P	1
588	Uncultured delta proteobacterium (JF319278.1)	δ-Proteobacteria	90	441P	1

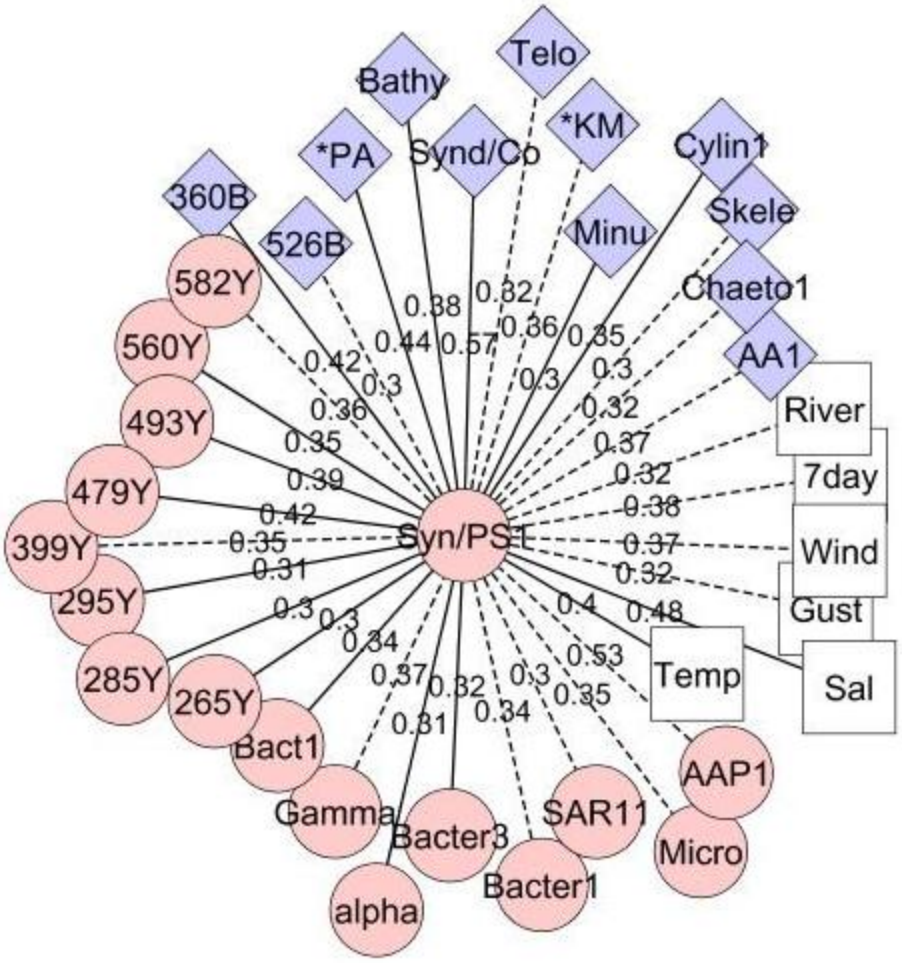


Fig. A1. *Synechococcus*/*Picocystis salinarum* sub network in 2008 Diamond: 18S nodes; Circles: 16S nodes; Square: Environmental factors; Solid lines represent positive relationship and dashed lines represent negative relationship. Numbers on the lines are LS scores. Sub networks show only LS correlations between central nodes and their first neighbors but not between their first neighbors. The length of the connecting lines between nodes and the angle between lines are arranged just for the easiness of visualization.

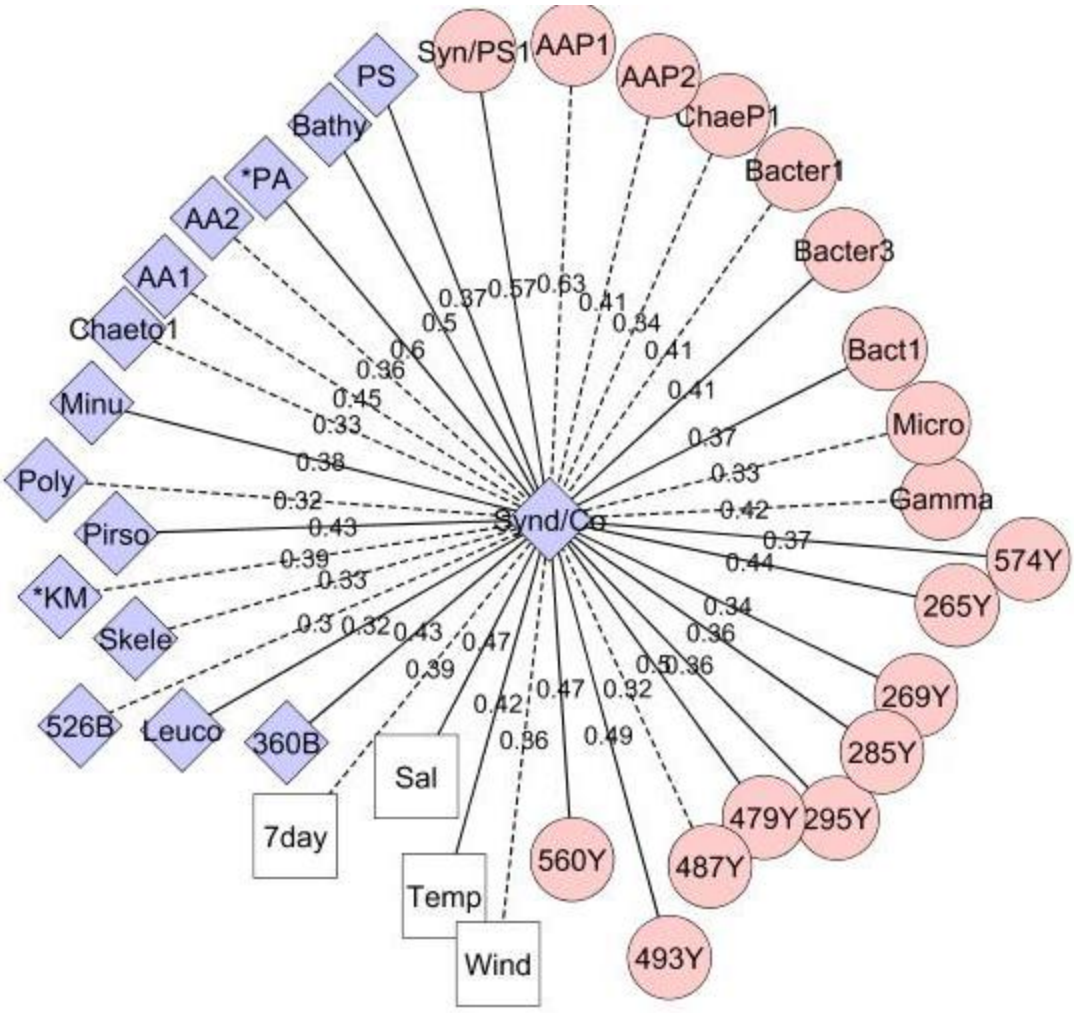


Fig. A2. Syndiniales/copepod sub network See Fig. A1 for legend details.

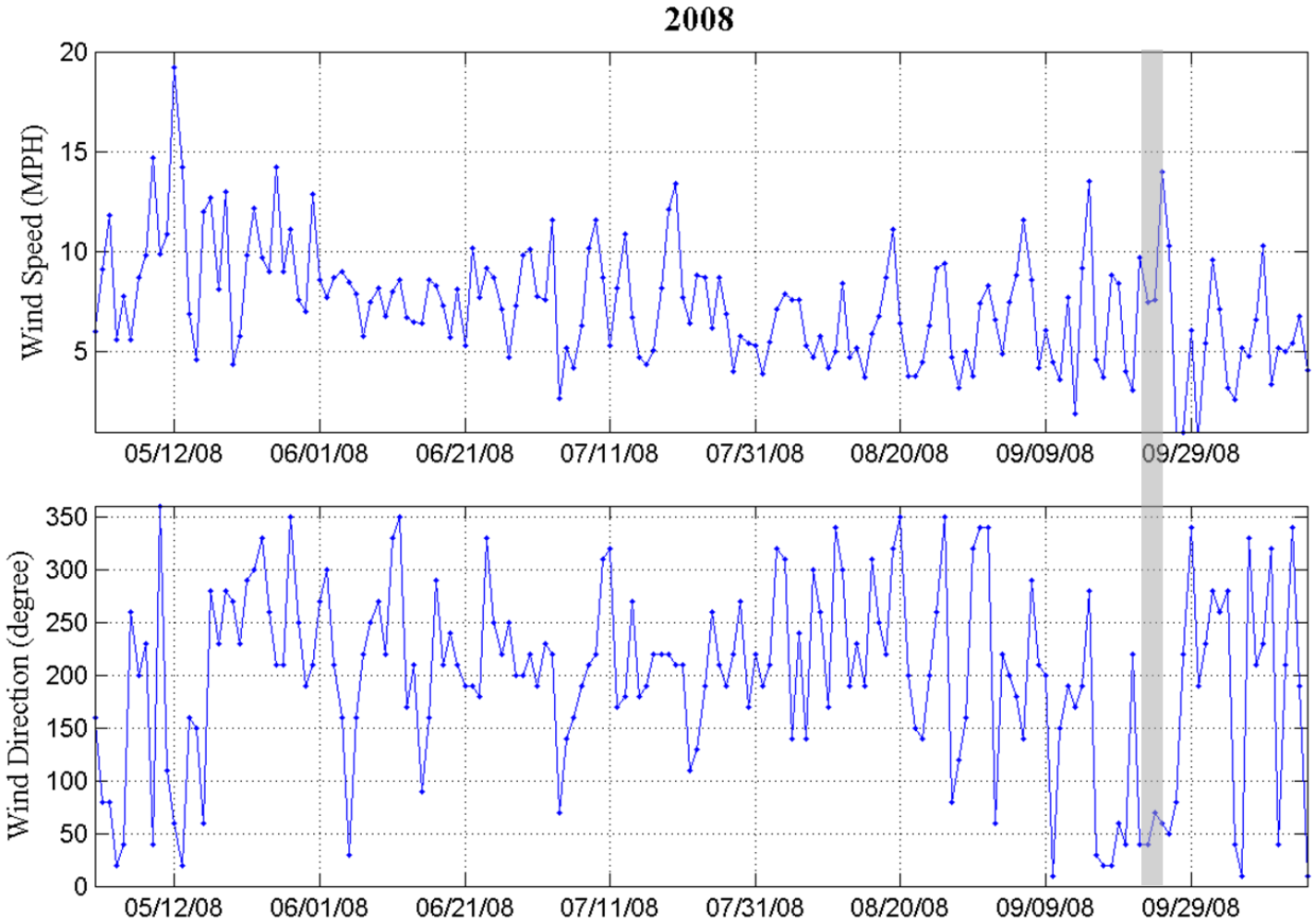


Fig. A3. Daily average of wind speed and direction throughout the sampling period in 2008 at Islip Airport, Long Island, NY Shaded area highlights September 23, 24 and 25.