ABSTRACT

Title of Document: BACTERIOPLANKTON IN THE CHESAPEAKE BAY: GENETIC DIVERSITY, POPULATION DYNAMICS, AND COMMUNITY PROTEOMICS

Jinjun Kan, Doctor of Philosophy, 2006

Directed By: Dr. Feng Chen, Assistant Professor, University of Maryland Biotechnology Institute

Although the ecosystem of the Chesapeake Bay has been studied extensively, little is known about the genetic diversity, population dynamics and metabolic activity of bacterioplankton living in the Bay. In this study, clone libraries containing the rRNA operon (16S rRNA-ITS-23S rRNA) were constructed from samples collected from the Chesapeake Bay to study spatial and temporal dynamics of estuarine bacterioplankton. Major bacterial groups changed dramatically between cold and warm seasons. In the summer, *Alpha*- and *Gammaproteobacteria*, *Bacteroidetes* (*Flavobacterium-Bacteroidetes-Cytophaga*), *Cyanobacteria* and *Actinobacteria* were the dominant groups while in the winter, *Alpha*- and *Betaproteobacteria*, and *Actinobacteria* were commonly found. Clone library analysis also revealed dramatic shifts in bacterial species composition between seasons. Unique SAR11, SAR86, and *Roseobacter* clades were discovered in the Chesapeake Bay, suggesting the ecological adaptation of organisms endemic to the Bay or

perhaps, large temperate estuaries. The bacterioplankton populations were monitored from 2002 to 2004 by denaturing gradient gel electrophoresis (DGGE) of PCR-amplified 16S rRNA gene fragments. Remarkable seasonal shifts and repeatable annual patterns were identified. Temporal variation of bacterial communities was best explained by the change of chlorophyll *a* (Chl *a*) and water temperature, while other factors such as dissolved oxygen, ammonia, nitrite and nitrate, and viral abundance also contributed to the seasonal succession of bacterial populations.

In order to understand ecological functions of microbes living in the natural environment, a community-based proteomic approach was developed. Typically, a few hundred-protein spots were visualized based on two-dimensional gel electrophoresis (2-DGE) from Chesapeake Bay microbial communities (0.2 to 3.0 μm filtered fractions). Distinct seasonal patterns and noticeable spatial variations of Chesapeake Bay metaproteomes were observed and the metaproteomic patterns correlated with genetic fingerprints based on 16S rRNA-DGGE. Six protein spots were characterized by LC-MS/MS and three of them were most closely related to the genes in the Sargasso Sea metagenomic database. We proved for the first time that metaproteomics could be applied to a complex marine microbial community. Our results indicate that community proteomics has great potential to unveil novel microgeochemical functions and to link microbial functions to their population structures.

BACTERIOPLANKTON IN THE CHESAPEAKE BAY: GENETIC DIVERSITY, POPULATION DYNAMICS, AND COMMUNITY PROTEOMICS

By

Jinjun Kan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:
Assistant Professor Feng Chen, Chair
Professor Brian P. Bradley
Professor Shiladitya DasSarma
Associate Professor Russell T. Hill
Professor Frank Robb
Assistant Professor Marcelino T. Suzuki

Preface

His invisible hands are behind this work.

Dedication

This work is dedicated to my love, Yiyun.

Acknowledgments

This Ph. D dissertation, representing the culmination of my research efforts, would not have been possible without the scientific, spiritual and emotional support from many others. My warmest thanks go to my advisor Feng Chen for his strong desire for my success, creative advice, encouragement and friendship. Next, I am particularly indebted to Russell T. Hill and Marcelino T. Suzuki, who helped me with the experiments as well as provided fruitful discussions and scientific guidance along my research. I gratefully acknowledge Frank Robb, Shiladitya DasSarma and Brian P. Bradley for their advice and suggestions to my study. I also must acknowledge the generosity of Byron C. Crump and Thomas E. Hanson for allowing me to use their facilities and gel image analysis software, which were critical to the project. I would like to thank Kui Wang for collecting the Chesapeake Bay water samples and sharing insightful discussion with me. Helps from Ju Sheng, Kate O'Mara are also appreciated. Finally, I acknowledge the personnel of the Center of Marine Biotechnology who enabled many important aspects of my research.

Reaching the ultimate level of education requires great success at the preceding stages of education and support from my parents Weishu Li and Fengyao Kan. They constantly encouraged and supported their son to pursue his career goals and complete this dissertation. Also important in this regard are my two sisters, Jinhong and Jinling, and their families. Without their full support, care, and encouragement, I would not be able to stay the arduous process of finishing my Ph. D dissertation.

Table of Contents

List of Tables

List of Figures

Chapter 1: General introduction

Diversity of bacterioplankton

I. The vast majority of bacteria are not easily cultivated

Bacteria are the main form of biomass in aquatic ecosystems and are centrally involved in fluxes of energy and matter (Azam 1998). Although a broad range of cell densities of bacteria in aquatic environments has been reported (10^4-10^7 cells/ml), in most cases bacterial concentrations are about 10^5 to 10^6 cells ml^{-1} (Hobbie et al. 1977; Porter and Feig 1980; Whitman et al. 1998). Entire aquatic bacterial community is synecologically considered as an entity, "bacterioplankton". Traditionally, bacteria from aquatic environments have been studied by characterizing those that can be cultivated on enriched media plates (ZoBell et al. 1946). However, the culturable fraction of bacterioplankton was thought to only comprise a minor portion of the total community due to selective effects of the media used and the presence of inactive cells (Jannasch and Jones 1959). For seawater samples, viable counts of bacterial isolates typically differ from total counts by 2-3 orders of magnitude (Jannasch and Jones 1959; Hoppe 1976). The number of culturable bacteria was confirmed to significantly underestimate the abundance of bacterioplankton in the natural environment by direct counting methods (Daley and Hobbie 1975; Porter and Feig 1980). Such a difference between microscopic counts and viable counts has been called the "great plate count anomaly" (Staley and Konopka 1985). Since then, many efforts have been devoted to developing different cultivation conditions that mimic *in situ* soil, freshwater or marine environments (Button et al. 1993; Janssen et al. 2002;

Hahn et al. 2003; Stevenson et al. 2004). For instance, oligotrophic marine bacteria were successfully isolated by a dilution cultivation approach (Button et al. 1993). By applying a new cultivation strategy, several important marine bacterial lineages (e.g. SAR11 or *Pelagibacter ubique*, *Lentisphaera*, and the oligotrophic marine *Gammaproteobacteria* group) have now been brought into culture (Rappé et al. 2002; Cho et al. 2004a; Cho et al. 2004b). However, many common bacterial groups including SAR86, SAR202, SAR324, marine *Actinobacteria*, and SAR406 have not yet been cultivated, pointing to the need of further improvements in cultivation technology.

II. Genetic diversity of bacterioplankton

Since the first study of bacterial community structure in Sargasso Sea in 1990 (Giovannoni et al. 1990), increasing molecular evidence has demonstrated a remarkable wider diversity of aquatic bacterial world than previously thought. The small-subunit ribosomal (RNA) gene (16S rRNA gene for prokaryotes) has been widely used as a phylogenetic marker to study the microbial diversity in various environments. As of June 2006, GenBank includes 255,208 16S rRNA gene sequences and many of these sequences were associated with phyla that contain no cultured representatives (Benson et al. 2005). The Ribosomal Database Project II (Version 9.41) contained 253,813 rRNA gene sequences, among which bacterial isolates comprised only 30% of the total (Klappenbach et al. 2001). Despite this expanded diversity, in general, less than 20 major microbial clades represent most 16S rRNA genes recovered from marine environmental samples. Those include *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*,

Deltaproteobacteria, *Cyanobacteria*, *Bacteroidetes*, *Actinobacteria*, *Planctobacteria*, *Fibrobacter* and *Archaea* (Giovannoni and Rappé 2000) and many of these clades seem to be ubiquitously distributed in aquatic ecosystems.

Defining a bacterial species is not straightforward. A genospecies of bacteria has been defined based on a DNA-DNA similarity of more than 70% (Schleifer and Stackebrandt 1983; Wayne et al. 1987). Alternatively, bacteria that share less than 97% similarity of 16S rRNA gene sequence are considered as different species (Stackebrandt and Goebel 1994; Amann et al. 1995). However, these definitions of bacterial species have been questioned with rapidly increased sequence data. For instance, in a coastal assemblage, a study of the genotypic diversity of a group closely related to *Vibrio splendidus* (> 99% 16S rRNA gene identity) showed that this group consisted of more than 1000 distinct genotypes with extensive allelic diversity and size variation in genome sizes (Thompson et al. 2004). A recent study on genomic islands of co-occurring *Prochlorococcus* ecotypes with less than 1% difference on 16S rRNA gene sequences demonstrated that variations of gene islands have been acquired by lateral gene transfer or environmental stresses (Coleman et al. 2006). If extensive genomic variation is a general feature of natural aquatic bacterial populations, bacterial diversity should be much higher than was previously thought. Furthermore, identical 16S rRNA gene sequences were found among *Actinobacteria* strains isolated from different thermal niches in temperate, subtropical, and tropical freshwater habitats (Hahn and Pöckl 2005), indicating bacterial diversity assessment cannot solely rely on sequence analysis of a single gene marker and that, the diversity

of natural bacterial communities should be considered not only at a genetic level, but also at metabolic and functional levels.

Fingerprinting bacterial population dynamics

A variety of molecular tools have been applied to study the population structure and dynamics of bacterial communities in various natural environments (Giovannoni and Rappé 2000; Akkermans et al. 1999). Analysis of 16S rRNA gene clone libraries has become a common technique to investigate the composition of bacterioplankton, and provides both qualitative and semi-quantitative information on population structure. However, the analyses of clone libraries are labor intensive and time consuming and clone library analysis becomes less practical when many samples need to be analyzed. Techniques that can be used for rapid assessment of bacterial community structures over time and space are thus desired.

Several fingerprinting approaches have been developed in the 1990s as rapid tools to analyze microbial community structure. These methods include denaturing gradient gel electrophoresis (DGGE; Muyzer et al. 1993), automated ribosomal intergenic spacer analysis (ARISA; Fisher and Triplett 1999), terminal restriction fragment length polymorphism (T-RFLP; Liu et al. 1997), and length heterogeneity-PCR (LH-PCR; Suzuki et al., 1998). Based on the variable length of the internal transcribed spacer (ITS), ARISA provides higher taxonomic resolution than the 16S rRNA gene for bacterial strains or closely related species (Fisher and Triplett 1999). For instance, ecotypes of *Prochlorococcus* and *Synechococcus* can be easily identified based on the length, G+C content, and sequence of the ITS (Rocap et al. 2002). LH-PCR detects the natural length variation of 16S rRNA gene 5' end-PCR

amplicons (Suzuki et al. 1998), while T-RFLP differentiates microbial communities based on restriction patterns of PCR products digested by endonucleases. All these non-gel based techniques can be applied on a high throughput scale with good sensitivity. However, data interpretation can be complicated. For instance, it is hard to identify a specific peak due to the overlapping sizes of fragments from different bacterial taxa.

In DGGE, mixed bacterial sequences amplified by PCR can be separated based on sequence and their G+C content using a denaturing gradient polyacrylamide gel. A flanking GC-clamp attached to one primer holds different amplicons at different gradient positions. Theoretically, DGGE is sensitive enough to detect a single base difference between two sequences. A major advantage of DGGE over other fingerprinting methods is that DGGE bands can be quickly isolated from gels and sequenced. Sequence information can further be used for phylogenetic reconstruction. In addition, due to low cost, a DGGE system can be easily set up in a routine laboratory. Perhaps due to these advantages, DGGE has been extensively applied to investigate population dynamics in various aquatic ecosystems, including lakes and rivers (Casamayor et al. 2000; Øvreås et al. 1997), coastal waters (Bernard et al. 2000; Crump et al. 2004), polar regions (Murray et al. 1998; Crump et al. 2003; Bano and Hollibaugh 2002) and extreme environments (Ferris et al. 1996; Sievert et al. 1999; Nakagawa et al. 2004).

All the fingerprinting methods described above include PCR steps and therefore, inherent limitations of the PCR technique affect the measurement of microbial diversity. Minor groups may be beyond the detection limit and thus the

richness of bacterial communities is likely underestimated (Kisand and Wikner 2003). In addition, as for other PCR-based cultivation independent techniques, DGGE could be associated with possible bias introduced by PCR, such as G+C content (Dutton et al. 1993), copy number of 16S rRNA gene (Farrelly et al. 1995), template annealing and primer selection (Suzuki and Giovannoni 1996), and chimera formation (Kopczynski, et al. 1994). Thus, these fingerprinting techniques may lead to a biased view of the 'real world' when reconstructing microbial structures at population level (von Wintzingerode et al. 1997; Kisand and Wikner 2003). Since there is no single standard protocol for 16S rRNA gene analysis of environmental samples, comparisons of these fingerprinting techniques to clone library analysis is necessary to confirm the results and understand their inherent limits. Regardless of existing technological limitations, these fingerprinting techniques are still the best way to analyze the major bacterial populations within natural communities when extensive environmental samples need to be compared.

Estuarine bacterioplankton

An estuary is a semi-enclosed body of water, which freely connects with the open seawater. Estuarine ecosystems are characterized by a pronounced salinity gradient and a high load of nutrients and organic matters derived from land drainage. Due to dramatic environmental gradients and osmotic constraints, estuaries offer a special niche for microbial organisms from freshwater and marine origins. Typically, an estuarine ecosystem harbors microorganisms with both freshwater and marine origins due to the mixing effect. Bacteria from both sources undergo a strong physiological stress at changing salinities along the estuaries and result in variations

of biomass, activities and population composition (Jonas and Tuttle 1990; Shiah and Ducklow 1994; del Giorgio and Bouvier 2002; Troussellier et al. 2002; Smith and Kemp 2003).

The composition of bacterial communities have been extensively studied in a number of estuaries including the Columbia River estuary (Crump et al. 1999), the San Francisco Bay (Hollibaugh et al. 2000), the Weser estuary, Germany (Selje and Simon 2003), the Parker River estuary (Crump et al. 2004), the Rhone River estuary, France (Troussellier et al. 2002), the Ria de Aveiro estuary, Portugal (Henriques et al. 2004; Henriques et al. 2006), the Changjiang River estuary, China (Sekiguchi et al. 2002), the Moreton Bay estuary, Australia (Hewson and Fuhrman 2004), the Delaware River estuary (Cottrell et al. 2005) and the Chesapeake Bay estuary (discussed below). Most of these studies applied quick fingerprinting approaches (i.e. DGGE, TRFLP and ARISA) or FISH (fluorescence in situ hybridization) to monitor the change of bacterial community along the salinity gradient or over the time scale.

Different bacterial communities have been found along the salinity gradient in estuaries. Compared to middle estuary, the population shifts occurred quickly at low or high salinity regions, indicating the salt tolerance affects the population structure of estuarine bacterial (Henriques et al. 2006). In general, *Betaproteobacteria* shifted to *Alpha*- and *Gammaproteobacteria* from freshwater to marine sections (Bouvier and del Giorgio 2002; Sekiguchi et al. 2002; Henriques et al. 2006). In contrast, Selje and Simon (2003) reported that *Alpha*-, *Beta*-, and *Gammaproteobacteria* constituted about 10% of the community without pronounced changes among the various sections in the Weser estuary. However, the dominancy and distribution of

Bacteroidetes shifted along this estuary and the Columbia River estuary, suggesting that this commonly found particle-associated bacteria might play important roles in estuarine community development and microbial activities (Crump et al. 1999; Selje and Simon 2003).

The previous studies also provided some information on the temporal variations of estuarine bacterioplankton. Most of the previous studies were carried out in temperate estuarine ecosystems. However, the results are inconsistent. For example, the middle estuary exhibited more pronounced temporal variation than other sections in the Ria de Aveiro estuary (Henriques et al. 2006). In contrast, no significant temporal variation in bacterial community was observed in the Weser estuary (Selje and Simon 2003). These contrasting results indicated that the population dynamics of estuarine bacterioplankton was a complex biological process and was controlled by multiple local abiotic as well as biotic environmental parameters. The picture for estuarine bacterioplankton is far from complete. In order to improve our understanding the synecology of estuarine bacterioplankton, we need much more detailed insights into phylogenetic diversity and spatio-temporal dynamics of its most prominent members.

Distinct estuarine bacterial communities have been reported in the Columbia River estuary, the Parker River estuary and Plum Island Sound, and the Weser estuary (Crump et al. 1999; Selje and Simon 2003; Crump et al. 2004). Most estuarine bacterioplankton were closely related to typical freshwater or marine bacterial groups and belonged to the phyla *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria* (Crump et al. 2004). However, estuarine phylotypes occurred within a wider salt gradient than

typical freshwater or marine biota (Crump et al. 1999; Selje and Simon 2003). The development of local estuarine bacterial communities depends on both the residence time and growth rate of the bacterial populations in an estuary (Crump et al. 2004). Residence time is the average time for a parcel of water in a section of an estuary takes to leave that section (Monsen et al. 2002). It describes the time available for the development of native bacterial population, or in other words, how long the local bacterial communities have been exposed to estuarine conditions. The residence time varies significantly in different estuaries, ranging from couple days to months, depending on their hydrological features and sampling seasons (Nixon et al. 1996; Crump et al. 2004). In small estuaries, the residence time can be too short to develop local bacterial communities (Troussellier et al. 2002; Henriques et al. 2006). Relatively long residence time provide enough time for the development of the estuarine bacterial community. Therefore, in order to better understand the microbial processing of estuarine bacterioplankton, study on an estuary with a larger geographic span and longer residence time is desired.

In addition, changes in the relative growth rate and mortality affect the population distribution and influence shifts in diversity. High growth rate of certain bacterial groups from advected populations within estuaries might lead to the development of an estuarine community. For example, a unique estuarine bacterial community formed only in summer and fall when the residence time is longer than the bacterial doubling time in the Parker River estuary and Plum Island Sound (Crump et al. 2004).

The Chesapeake Bay and its bacterioplankton

The Chesapeake Bay is the largest estuary in the United States with a length of more than 300 km and a total area of tidal waters of 11,000 km^2 (www.gmu.edu/bios/bay). The average depth of the Chesapeake Bay is only about 10.6 m, and therefore it is very sensitive to environmental fluctuations such as those of temperature and winds. The Chesapeake Bay receives about 50% of its water volume from the Atlantic Ocean, and the other half of water from rivers including 50 major tributaries. Salinity in the Chesapeake Bay ranges from nearly 0 to 30 ppt from the northern Bay to the mouth of the Bay. Water temperature varies from 0 to 29ºC with a fairly predictable seasonal pattern (www.gmu.edu/bios/bay). Rivers and streams also input large amount of inorganic nutrients that affects the biological production in the Bay. Excess organic and inorganic nutrients often trigger diatom blooms in early spring and dinoflagellate blooms in late spring (Glibert et al. 1995). These algal blooms have caused dramatic changes of the ecosystem such as increased bacterial respiration (Shiah and Ducklow 1994) and oxygen depletion (Malone et al. 1986).

The Baltimore Inner Harbor is located in the northwest of the Chesapeake Bay. It is part of the mesohaline Patapsco River estuary, the fifth largest tributary of the Bay. An excess amount of nutrients (i.e. dissolved inorganic nitrogen, phosphorus etc.) are carried into the Inner Harbor with river runoff and cause rapid growth of phytoplankton (Sellner et al. 2001). Phytoplankton blooms occur frequently from late spring to summer resulting in a large and active bacterial community (Sellner et al.

2001; Kan and Chen 2004). The Pier V of Inner Harbor was chosen as a sampling site to monitor the temporal changes of bacterial community structure (Chapter 2).

Extensive ecological surveys in the Chesapeake Bay have been focused on the bulk measurement on phytoplankton biomass and production (Harding 1994; Harding et al. 2002; Malone et al. 1991; Ray et al. 1989) or bacterioplankton biomass, growth rate, respiration, and production (Jonas and Tuttle 1990; Shiah and Ducklow 1994). These early studies indicated that the biomass and metabolic activities of Chesapeake Bay bacterioplankton vary with time and space. Spatial and temporal variations demonstrate the interactive effects of substrates, temperature, and salinity in controlling the bacterial activities (Shiah and Ducklow 1994; Schultz and Ducklow 2000; Smith and Kemp 2003). Bacterial abundance, production, and growth rate are positively correlated with water temperature in non-summer seasons, but are regulated by substrate when temperature is above 20ºC (Shiah and Ducklow 1994). High bacterial biomass and production rates are commonly observed in the mid-Bay region (Ducklow and Shiah 1993; Shiah and Ducklow 1994; Ducklow et al. 1999), which are also consistent with bacterial respiration rates (Smith and Kemp 2003). Bacterioplankton activity appears to be limited by organic carbon in the upper Bay and by inorganic nutrients in the lower Bay, while likely controlled by grazing and other forms of mortality in the mid-Bay (Smith and Kemp 2003). In contrast, the bacterial biomass and production do not seem to synchronize in the tributaries of the Chesapeake Bay. For example, bacterial cell counts increased while bacterial production rate decreased along the York River, the Choptank River and the Pocomoke River (Ducklow et al. 1999; Bouvier and del Giorgio 2002).

Differences in planktonic bacterial activity in the Chesapeake Bay may also result, at least in part, from variability in the phylogenetic composition of dominant bacterial populations. Little is known about bacterial community structure and population dynamics in the Bay. Bidle and Fletcher (1995) and Noble et al. (1997) showed that the bacterial community in summer was distinct from that in winter based on the 5S rRNA fingerprints. However, it is difficult to obtain detailed information on the genotype or species composition of these bacterial assemblages based on 5S rRNA band patterns. Using FISH, Bouvier and del Giorgio (2002) found that the composition of bacterioplankton shifted along two tributaries (the Choptank and Pocomoke Rivers) of the Chesapeake Bay. *Alphaproteobacteria* dominated in the saltwater region while *Betaproteobacteria* were more abundant in the upper freshwater region. *Cytophaga* prevailed in the middle Bay at the turbidity maximum. *Gammaproteobacteria* did not show spatial trends along this transect but peaked at certain locations (Bouvier and del Giorgio 2002). Using more specific FISH probes, Heidelberg et al. (2002) showed that the distribution and abundance of *Gammaproteobacteria* and four *Vibrio* spp. strains varied seasonally along the Choptank River. These previous studies have demonstrated dynamic variations of bacterioplankton community composition in the Chesapeake Bay. However, these studies only provided the analyses of bacterial communities at either class or subclass level (too broad) or species level (too narrow). No detailed analysis of the bacterial community structure based on 16S rRNA gene has yet been reported. In general, microbial diversity in the Chesapeake Bay is still not well studied and deserves further investigation.

Factors affecting bacterioplankton biomass and population structure

Studies on spatial and temporal variations of the bacterial communities are important to understand their physiological adaptations, potential ecological functions and niche partition. A group of closely related bacteria (e.g. < 2-3% 16S rRNA sequence divergence) may adapt to unique ecological habitats or geographic locations, forming different ecotypes or geotypes (Ward 2006). For example, the vertical distribution of photosynthetic unicellular marine cyanobacteria *Prochlorococcus* and *Synechococcus* suggest the depth-variable adaptation to light and nutrients are confined to the photic zone (Rocap et al. 2003; DeLong et al. 2006). The abundance and distribution of *Prochlorococcus* ecotypes in Atlantic Ocean were affected by temperature, light, nutrients, and *Synechococcus* abundances (Johnson et al. 2006). Thus, environmental and biological factors can play roles in determining the abundance, distribution and activity of bacterial community.

Bacterioplankton communities in the open ocean differ from those in freshwater environments (e. g. González and Moran 1997; Methé et al. 1998; Glöckner et al. 1999; Cottrell and Kirchman 2000b; Zwart et al. 2002), suggesting that salinity could affect the distribution of aquatic bacterial communities as change of cellular osmotic pressure might regulate the distribution of these organisms. For instance, in a study employing FISH *Betaproteobacteria* were absent in the marine environment but constituted a dominant fraction (16%) of lake bacterioplankton (Glöckner et al. 1999). Along the salinity gradients in estuaries, bacterioplankton populations often shift from *Alpha*- and *Betaproteobacteria*, *Actinobacteria*, and *Verrucomicrobia* to *Alpha*- and *Gammaproteobacteria* (Crump et al. 1999; Bouvier

and del Giorgio 2002). Many unique phylogenetic clusters appear to be restricted to or dominate in the freshwater ecosystems (Zwart et al. 2002). Clusters representing the best known freshwater groups have been defined and belong to *Alpha*-, *Beta*-, and *Gammaproteobacteria*, *Cytophaga*, *Cyanobacteria*, *Actinobacteria* and *Verrucomicrobia* (Zwart et al 2002).

Temperature is another important factor that affects biological activities in the aquatic environment (White et al. 1991). For example, cell density and growth of marine picocyanobacteria *Synechococcus* varies dramatically with water temperature (Agawin et al. 1998; Anderson et al. 1994; McDaniel et al. 2002; Wang et al. unpubl. data). A recent survey in a Pole to Pole transect in the Atlantic Ocean showed that the abundance of different ecotypes of *Prochlorococcus* was significantly correlated with temperature (Johnson et al. 2006). Furthermore, abundance of *Gammaproteobacteria* and *Vibrio* spp. is tightly correlated with water temperature in the Chesapeake Bay (Heidelberg et al. 2002). In temperate estuaries, coupled with substrate supply, temperature can affect bacterial biomass, growth and respiration and consequently might regulate the bacterial composition (Wikner and Hagström 1991; Shiah and Ducklow 1994; Pomeroy and Wiebe 2001). Temperature may also act indirectly on microbial processes or activities by affecting other environmental (e. g. dissolved oxygen and pH) and biological parameters.

Concentration of nutrients (both organic and inorganic) also can affect the abundance, activity, and composition of microbial assemblages (Brett et al. 1999; Biddanda et al. 2001). Carbon, nitrogen, and phosphorus are important for growth and production of heterotrophic bacteria (Kirchman et al. 2003a; Caron et al. 2000).

Bacterial production increases significantly when amended with organic carbon and inorganic nutrients (Caron et al. 2000). Nutrients also trigger the growth of specific bacterial groups and thus change the community composition. For example, high concentration of ammonia in the Chesapeake Bay favors the growth of ammonia-oxidizing bacteria (AOB), resulting in higher abundance of AOB in the estuary than open oceans (Ward 1982).

Interaction between bacterioplankton and other organisms, particularly phytoplankton can affect microbial community structure. It has been reported that bacterioplankton cell abundance or production is positively correlated with phytoplankton biomass or production (Bird and Kalff 1984; Fuhrman et al. 1980; Linley et al. 1983; Marvalin et al. 1989; Robarts et al. 1994). In aquatic ecosystems, phytoplankton release a significant amount of dissolved organic matter, which can be quickly taken by bacterioplankton (Cole et al. 1988; Currie 1990; White et al. 1991; Riemann et al. 2000). Phytoplankton blooms may also cause a shift of bacterial population structure (Riemann et al. 2000). Moreover, changes in phytoplankton species composition also influences bacterial population structure. For example, bacterial groups associated with dinoflagellates were found to be different from those associated with diatoms (Pinhassi et al. 2004). It is important to note, however, that several investigations only found a weak correlation between bacterioplankton and phytoplankton abundance (Findlay et al. 1991; Le et al. 1994). In addition, no correlation between bacterial composition (detected by FISH) and Chl *a* was observed in Chesapeake Bay rivers (Bouvier and del Giorgio 2002). These findings

indicate that phytoplankton biomass is not the only force driving the succession of the bacterial communities.

Viruses (mostly bacteriophage) are probably the most abundant biological entities in seawater (Bergh et al. 1989; Fuhrman and Suttle 1993). Viruses are believed to infect dominant (or most successful) bacterial populations by 'killing the winner' (Thingstad & Lignell 1997). Through the lysis of specific host cells, viruses cause prokaryotic mortality (Proctor and Fuhrman 1990) and therefore might regulate the population structure of bacterial community (Suttle 1994; Wommack and Colwell 2000). Consequently, non-infected host populations may be stimulated by release of organic matter from lysed cells (Middelboe and Lyck 2002). A strong correlation between bacterial and viral abundance has been found in the Chesapeake Bay (Wommack and Colwell 2000). In addition, co-variation of bacterial and viral assemblage composition (Hewson et al. 2006) supports the hypothesis that viral infection may shape the composition of the host community (Thingstad and Lignell 1997; Wommack et al. 1999; Øvreås et al. 2003). Also, in a recent metagenomic survey, an unexpected large number of viral DNA sequences was obtained from replicating viruses within infected host cells (DeLong, et al., 2006), suggesting a tight interaction between bacteria and virus in the ocean. Finally, from genetic and evolutionary perspectives, viruses are able to influence genetic diversification of bacteria via horizontal gene transfer resulting from viral transduction, transformation, and conjugation (Paul 1999; Weinbauer and Rassoulzadegan 2004).

Besides salinity, temperature, phytoplankton, nutrients ("bottom-up control"), and viral lysis mentioned above, planktonic bacteria are also regulated by

bacterivorous protists ("top-down control"). Grazing is considered a major mortality factor and therefore an important selective force for aquatic bacterioplankton populations (Pace 1988; Sherr et al. 1992; Sherr and Sherr 2002). Different bacteria are not equally vulnerable to grazers and thus grazing plays an important role in structuring bacterial population dynamics (González et al. 1990; Hahn and Höfle 2001; Jürgens and Matz 2002; Suzuki 1999). Many bacterial taxa possess phenotypic plasticity and develop predation-resistant cells by altering cell size and morphology (Hahn et al. 1999 and 2000; Matz et al. 2002). For example, under the grazing pressure, bacterial communities structure can shift from a dominance of small size rod cells to filamentous cells (Corno and Jürgens 2006). All these observations suggest that grazing is an important factor affecting the biomass and population dynamics of bacterial assemblages.

Environmental and biological factors act interactively on microbes living in the aquatic environment. Multivariate statistical analysis has been applied to understand the complex interaction between microbial community and environmental factors. Nonmetric multidimensional scaling (MDS) is commonly used analytical method to describe changes in bacterial communities over time or space (Crump et al. 2003; van Hannen et al. 1999). It attempts to arrange the bacterial communities in two or three dimensions so as to identify community patterns and help to explain observed similarities or dissimilarities. However, it is not possible to link the bacterial community patterns with environmental parameters using MDS. A second multivariate analysis, canonical discriminant analysis (CDA), can be used to correlate community structure to environmental variables. Canonical discriminant analysis

(CDA) relates the variation of bacterial community to *in situ* environmental variables and determines the environmental factors that discriminate the naturally occurring community patterns via multivariate F tests (Momen et al. 1999; SAS/STAT 1992). If the canonical discriminant functions are statistically significant, bacterial communities can be distinguished and predicted based on predictor variables included in these functions.

Community genomics/proteomics

Bacteria are metabolically diverse. Aquatic bacteria play pivotal roles in natural environmental processes and thus provide a large untapped resource for the discovery of novel metabolisms, enzymes and pathways. However, as mentioned above, only a small fraction of bacterial populations can be cultivated, posing a great challenge to the understanding of the *in situ* activities and metabolism of natural bacterial assemblages. Recently, community-based genomics and proteomics have been explored as means to study microbial functions in natural environments.

Community genomics (or metagenomics) is the analysis of the collective microbial genomes contained in an environmental sample (reviewed by Riesenfeld et al. 2004). A study using whole genome shotgun sequence data from Sargasso Sea water yielded 1.0 billion base pairs of non-redundant environmental sequences and 1,184 16S rRNA gene fragments (Venter et al. 2004). Assuming 97% similarity as a cutoff, Venter et al. concluded that 1,800 genomic species of bacteria and 145 new phylotypes were present in the samples. This study alone contributed about 1.2 millions new genes and translated proteins to the public database. In the North Pacific Subtropical Gyre, DeLong et al. (2006) constructed seven genomic libraries along a

depth continuum from 10m to 4000m and in total about 64 Mbp of assembled DNA sequences were obtained.

Metagenomics provides new information into microbial diversity and new insights into potential functions of microbes living in nature. Novel metabolic activities of not-yet-cultivated microbes can be discovered through metagenomics. For example, using metagenomics it was found that an uncultured *Gammaproteobacteria*, SAR86 group contains a proteorhodopsin gene, which had previously seen in Archaea only (Béjà et al. 2000a). The discovery of the proteorhodopsin gene in SAR86 revealed a novel photosynthetic pathway in planktonic *Bacteria*. Subsequent studies showed that the proteorhodopsin gene is widely distributed among marine *bacterioplankton* and spectrally tuned at different water depths for various light sources (de la Torre, et al. 2003; Béjà et al. 2001; Venter et al. 2004), indicating the ecological adaptation of this novel type of marine phototroph. In addition, discovery of an ammonium monooxygenase gene in genomic scaffolds from *Archaea* suggests that oceanic nitrification is not solely mediated by *Bacteria* (Venter et al. 2004). Finally, bioinformatics analysis of metagenomic data obtained from different microbial ecosystems demonstrated that predicted functional gene expressions are clustered according to environments (Tringe et al. 2005), suggesting potential niche adaptation of microbes in the environment.

Although metagenomics is a powerful tool to uncover potential new functional genes, it does not provide information on whether these genes are expressed under particular environmental conditions. The proteome was defined as "the total protein complement able to be encoded by a given genome" (Wasinger et

al. 1995). Proteomics is a technique that systematically documents and analyzes the proteins expressed in biological samples. Community proteomics (metaproteomics) studies the protein expression in natural microbial assemblages. The goal of proteomics is to study the changes in protein expression, modification, and interaction on a large scale with a view to understand global, integrated processes at the protein level (Blackstock and Weir 1999). These changes are likely due to biological perturbations (Anderson and Anderson, 1998) and effects from environmental conditions (Shepard et al. 2000).

As stated in the 'central dogma' of molecular biology, biological activities can be characterized at three different levels: DNA, RNA and proteins. Proteins are the final products of gene expression. Nearly all the cellular activities are performed by enzymes that are made up of individual proteins. It is generally believed that cellular and biological functions can be better interpreted at the protein level than at the DNA and RNA levels. Furthermore, extensive studies on yeast and mammalian cells demonstrated that protein expression does not always directly correlate to mRNA expression (Pradet-Balade et al. 2001) and therefore, studying gene function at the protein level has great potential for understanding the actual biological activities.

Proteomic analysis includes the 'classic' two-dimensional gel electrophoresis (2DGE)-based approach and the non-gel-based approach. Each approach has its own strength and weakness. 2DGE was first introduced about 30 years ago (O'Farrell 1975) and is still extensively applied in current studies. In the late 1990s, gel independent proteomic techniques began to emerge, including surface enhanced laser desorption/ionization (SELDI) and liquid chromatography-mass spectrometry (LC-

MS). Benefiting from these new technologies, protein samples can be depleted or concentrated, pre-fractionated or de-complexified before downstream analysis. Gel independent methods are particularly useful when dealing with complex or limited amounts of sample. However, these approaches cannot compete with 2DGE-based proteomics with regard to protein quantification, since intensity and size of protein spots on 2D gels provide more accurate estimation on the level of gene expression than that inferred from MS spectra. In addition, 'proteome fingerprints' can be obtained by visual observations of proteome distribution patterns among different species or same species under different environmental stresses. Therefore, old-fashioned 2DGE-based proteomics still holds its merits in current proteomic studies.

In 2DGE, proteins are separated based on their isoelectric points (pI) and molecular weights (MW). In the first dimension, high voltage power enforces the individual protein species to migrate until they reach their neutral pH point (pI). Proteins with same pI are further separated based on MW in the second dimension (Fichmann and Westermeier 1999). Proteins spots of interest can be excised from the gel and characterized via Edman N-terminal sequencing or MS.

Recently, several groups of researchers have applied community proteomics to investigate functional gene expression in various microbial communities. The first application of microbial metaproteomics was used to decipher the metabolic details of enhanced biological phosphorus removal process of activated sludge wastewater (Wilmes and Bond 2004). Strong expression of proteins involved in phosphorus removal was evident as revealed by metaproteomics (Wilmes and Bond 2004). More recently, community proteomics was applied to investigate the biofilm community

associated with an acid mine drainage (AMD) (Ram et al. 2005). Metaproteomic analysis of the AMD biofilm greatly benefited from its existing metagenomic database (Tyson et al. 2004). A total of 2033 individual proteins excised from 2DGE were positively identified. Proteins linked to environmental challenges including chaperone, thioredoxins and peroxiredoxins were found to be abundant in the total proteomes. In addition, a large portion of proteins could not be assigned to particular functional categories suggesting possible novel gene products (Ram et al. 2005). These studies demonstrate the potential of the metaproteomic approach to elucidate the detailed activities of natural microbial environments.

Metaproteomics is challenged by the complexity of microbial communities. Soil and aquatic samples contain diverse microbial species. It has been estimated that there are 160 taxa of bacteria in one milliliter of seawater, and 6,400-38,000 taxa of bacteria in one gram of soil (Curtis et al. 2002). Assuming that the average genome size of environmental bacteria is ~3 Mb and 1 kb of sequence encodes one gene, one can expect to observe 4.8×10^5 expressed proteins in 1 ml of seawater (Wilmes and Bond 2006). This number could be significantly higher if considering diverse protein conformations (at least 10 times the gene number) resulting from transcriptional or translational modifications. This is beyond the resolution of current proteomic tools and consequently, metaproteomics of water or soil samples can only resolve a minute fraction of the highly expressed (abundant) proteins. Some other challenges may come from proteomic extraction, separation and identification. Since there is no general guideline for proteome extraction, protocol varies with types of environmental samples. Along with the methodology development, a few issues

should be taken into consideration: 1) Environmental proteomes should be quantitatively extracted and isolated with acceptable good reproducibility. 2) The isolation process should not affect the ability to resolve the proteins in the following steps, i.e. 2DGE. And 3) Limited environmental genomic sequences challenges characterization and identification of environmental proteomes.

A study on DOM from lake water and forest soils demonstrated that proteomic fingerprints can be used to describe presence and activity of organisms in an ecosystem (Schulze et al. 2005). More proteins (78%) originating from bacteria were found in lake DOM (dissolved organic matter) than in forest soil (50%), and the number of identified proteins and taxonomic groups significantly varied in winter and summer seasons (Schulze et al. 2005).

Microbial community genomes recovered from marine environments still contains numerous sequences without known function (Venter et al. 2004; DeLong et al. 2006). Metagenomic sequences provide a useful database for identifications of environmental proteomes. However, no proteomic studies have been conducted in the natural marine ecosystem. Future studies on metaproteomics will link the protein identification to their source and ecological roles, and thus improve our understanding on the functional pathways of environmental microbes. Comparative metaproteomics is an approach to understand how microbial processes are regulated by various environmental parameters such as light, salinity or nutrients. Finally, identification of microbial proteins *in situ* may allow uncovering of important or novel biogeochemical functions.

Objectives of the dissertation

The objectives of this dissertation are to improve our understanding of the population structure, temporal dynamics, and potential functional roles of bacterioplankton in the Chesapeake Bay. In order to reveal the monthly variation of natural bacterial community, both cultivation and cultivation independent approaches (DGGE) was applied to the water samples from the Inner Harbor during 2001-2002. In addition, the detection threshold of the DGGE method was also tested by seeding a natural bacterial community with different concentrations of known bacterial isolates (Chapter 2).

Detailed population structure of bacterioplankton in the Chesapeake Bay was determined by rRNA operon clone libraries, DGGE and LH-PCR (Chapter 3 and 4). Six libraries were constructed at three stations with samples from two different seasons. A rapid phylogenetic screening approach, ITS-LH-PCR was used to analyze the clone libraries (Chapter 4). DGGE was applied to investigate multiple-year bacterial population dynamics in the Chesapeake Bay (Chapter 5). Succession of bacterial communities was analyzed by clustering (MDS) and the variation was correlated to the environmental parameters by CDA (Chapter 5).

To study the *in situ* activities of bacterial groups occurring in the natural environment, a 2DGE-based community proteomics (metaproteomics) approach was developed and applied to Chesapeake Bay water samples. The technique was validated using a constructed community and subsequently applied to analyze Chesapeake Bay picoplankton communities (0.2-3 μm cell size). Metaproteomics of picoplankton allowed differentiation the protein expression at different stations and

seasons (Chapter 6 and Appendix B). A subset of proteins excised from the 2D gels was tentatively identified by LC-MS/MS. Despite the existing challenges (e.g. limited environmental sequences), metaproteomics has demonstrated its great potential for linking microbial processes with specific microbial populations in the natural environment.

Chapter 2: Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE)

Abstract

To understand how the composition of estuarine bacterioplankton changes on a monthly basis, microbial communities in the Baltimore Inner Harbor were investigated using denaturing gradient gel electrophoresis (DGGE) of 16S rRNA gene. As revealed by DGGE fingerprints, the composition of bacterioplankton populations in the Harbor varied from month to month, and three major seasonal patterns were identified: winter (December and January), spring (February to May) and summer-fall (June to November). Sequencing of DGGE bands showed that *Planctomycetes* and uncultured *Alphaproteobacteria* were detected in all seasons. *Roseobacter* spp. and *Rhodobacter* sp. were present only in winter and spring. Marine *Alphaproteobacteria* and *Cyanobacteria* exhibited similar seasonal patterns and appeared to be more dominant from late summer to fall. *Betaproteobacteria* were present in most months, but different phylotypes were present from spring to summer-fall. *Gammaproteobacteria* and *Bacteroidetes* were present only in winter and early spring. In addition to DGGE analysis, 48 bacterial isolates from summer and winter were cultured and characterized. Few of these bacterial isolates matched with phylotypes determined by sequencing DGGE bands, which suggested that the density of 'easy-to-culture' bacteria in the natural environment may be too low to be detected by PCR-DGGE. Bacterial seeding experiments showed that detection

thresholds for PCR-DGGE ranged from 2.5×10^3 to 1×10^4 cells ml^{-1} (0.1 to 0.4% of total cell counts), depending on the copy number of rRNA operons in the genome of individual species.

Introduction

As an important component of aquatic microbial food webs, bacterioplankton plays a significant role in the global cycle of carbon, nitrogen and other elements. Understanding how the composition of microbial community changes over time and space in a given environment could shed light on the ecological role of microbes in the natural environment. Using molecular tools, many novel microorganisms have been discovered from various environments (Akkermans et al. 1999, Giovannoni and Rappé 2000). The structure of the bacterial community changes seasonally in aquatic environments (Lee and Fuhrman 1991; Höfle et al. 1999; Yannarell et al. 2003). For example, an extensive study of lake bacterioplankton showed a distinct seasonal succession and a dramatic drop in richness and abundance in summer (Kent et al. 2004). A recent study in a Californian coastal water body showed that the composition of bacterial populations shifted between months and that temporal patchiness was seen (Fuhrman et al. 2006). In estuaries, composition of bacterial populations is due to the mixing of microbial communities from the river, estuary, and coastal ocean (Crump et al. 1999). However, the diversity and population dynamics of bacterioplankton are poorly understood in estuarine ecosystems.

The Chesapeake estuary has been shown to be very dynamic in terms of its hydrological conditions (Smith et al. 1992). The Baltimore Inner Harbor lies 14 miles (~22.5 km) from the mouth of the Patapsco River, the fifth large tributary of the

Chesapeake Bay. Baltimore Inner Harbor is part of the mesohaline Patapsco River estuary where freshwater and salt water meet. Dissolved inorganic nitrogen and phosphorus in the surface of Patapsco River estuary exceed 30 and 0.5 μM, respectively, in the summer period (Sellner et al. 2001). Phytoplankton blooms occur frequently in the Baltimore Inner Harbor from late spring throughout the summer, and the resulting biomass from dinoflagellates supports a large and active microheterotrophic community (Sellner et al. 2001). A tight association between bacteria and phytoplankton was observed during a bloom that occurred in the harbor (Kan and Chen 2004).

To date, there have been a limited number of studies on the microbial composition in the Chesapeake Bay and its sub-estuaries (Bidle and Fletcher 1995; Bouvier and del Giorgio 2002). Banding patterns of 5S rRNA demonstrated that the composition of bacterial communities in the bay varied between summer and winter (Noble et al., 1997). Using fluorescent *in situ* hybridization (FISH) analysis with taxon-specific probes, it was found that *Gammaproteobacteria* and 4 *Vibrio* spp. strains exhibit strong seasonality in the Choptank River (Heidelberg et al. 2002). However, phylogenetic composition and temporal variations in the bacterioplankton community based on 16S rRNA gene characterization have not yet been explored in the Chesapeake estuary.

Denaturing gradient gel electrophoresis (DGGE) of PCR amplified 16S rRNA genes was first introduced as a quick fingerprint method to study bacterial dynamics at the community level (Muyzer et al. 1993). DGGE has been extensively used to study microbial population composition in various environments, including lakes and

rivers (Øvreås et al. 1997; Casamayor et al. 2000), coastal waters (Bernard et al. 2000; Crump et al. 2004), polar regions (Murray et al. 1998; Bano and Hollibaugh, 2002; Crump et al. 2003) and extreme environments (Ferris et al. 1996; Sievert et al. 1999; Nakagawa et al. 2004). However, the PCR-DGGE method tends to bias towards the predominant groups within a community (von Wintzingerode et al. 1997; Casamayor et al. 2000; Kisand and Wikner 2003). The detection threshold of DGGE for a specific type of bacterium has been estimated based on mixed bacterial assemblages (Muyzer et al. 1993; Murray et al. 1996) and by double-checking microscopic counts of certain bacterial groups (Casamayor et al. 2000). Using DGGE, it is difficult to detect a bacterial population when it occupies less than 0.5 to 1% of the total bacterial community. However, more systematic studies need to be done to determine the detection threshold of the PCR-DGGE approach.

The goal of this study was to understand how the composition of an estuarine bacterioplankton community changes over time. We described the monthly variation of bacterial community structure from the Baltimore Inner Harbor using DGGE. In addition, by use of seeding a natural bacterial community with different concentrations of known bacterial isolates, we evaluated the detection threshold of the DGGE method applied in this study. The detection limit of the PCR-DGGE method is affected by the relative abundance of a population as well as by ribosomal RNA (*rrn*) operon copy numbers.

Materials and methods

Sample collection and bacterial isolation. Water (5 l) was collected monthly from March 2001 to February 2002 from Pier V, Baltimore Inner Harbor, using acid-rinsed

carboys; 250 ml of water was filtered immediately through 0.2 μm pore-size polycarbonate filters (47 mm diameter, Millipore) and the filters were stored at –20˚C for later DNA extraction. Salinity and temperature were recorded on site and salinity was measured with a 300011 refractometer (Sper Scientific).

Bacteria strains from the Inner Harbor were isolated on 1/2 YTSS (4 g Yeast Extract, 2.5 g Tryptone, 18 g Agar l^{-1} water) medium plates (Sobecky et al. 1996). The media were adjusted to varying salinities using viral particle-free water. Briefly, 50 μl and 100 μl of water were streaked on plates for each month (March 2001 to February 2002). Twenty bacterial colonies from each month were collected according to morphological characteristics including color, size, and shape of colonies. These bacteria colonies were further purified and stored at –80°C. In this study, 48 bacterial strains isolated from winter and summer were randomly picked and characterized by analysis of 16S rRNA gene sequence.

Enumeration of bacteria. Subsamples of 50 ml water were fixed in 1% glutaraldehyde for total bacterial counting. Briefly, 1 ml of fixed sample was filtered onto a 0.2 μm pore-size 25 mm black polycarbonate membrane filter (Osmonics). Cells were stained with 2.5 × SYBR Gold solution for 10 min in the dark as described previously (Chen et al. 2001). Bacterial cells were enumerated under blue excitation (485 nm) on a Zeiss Axioplan epifluorescence microscope (Zeiss). At least 200 bacterial cells per sample were counted.

Extraction of nucleic acid. Bacterial genomic DNA was obtained from the filter by lysozyme, Proteinase K, and SDS (sodium dodecyl sulfate) concomitant with phenol-chloroform extraction and isopropanol precipitation (Schmidt et al. 1991). For natural

microbial assemblages, because the filter membrane was included through the extraction procedure, time allowed for enzyme reactions such as lysozyme and Proteinase K was extended overnight to avoid an incomplete reaction. After using a SpeedVac (AES1010, Savant) to dry the pellet, DNA was dissolved in double-distilled water and stored at 4°C for further analysis. DNA concentrations were estimated based on 260 nm absorbance using a SmartSpec TM 3000 (Bio-Rad).

PCR amplification of 16S rRNA genes. PCR amplification was performed in a 50 μl volume containing approximately 50 ng of template DNA, 1 × PCR buffer, 1.5 mM $MgCl_2$, 0.5 mM (each) primer, 200 mM (each) deoxynucleotide, and 2.5 U *Taq* DNA polymerase (Promega). PCR cycling was performed with a Peltier Thermal Cycler PTC-200 (MJ Research). For bacterial isolates characterization, PCR primers were 8F (AGAGTTTGATCCTGGCTCA) and 785R (CTACCAGGGTATCTAATCC) (Amann et al. 1995). The temperature-cycling conditions were as follows: 94°C for 5 min, 30 cycles at 94°C for 0.5 min, 55°C for 1 min, and 72°C for 3 min, followed by 5 min incubation at 72°C. PCR primers for DGGE were F1070 (ATGGCTGTCGTCAGCT), which is specific for most Eubacteria and R1392GC (CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCCACGGGCGGTGTGTAC). R1392GC contains a 40 bp GC-rich clamp and is based on a universally conserved region (Ferris et al. 1996). The temperature-cycling conditions were as follows: after pre-incubation at 94°C for 5 min, a total of 27 cycles were performed at 94°C for 0.5 min, T_A (annealing temperature) for 1 min, and 72°C for 3 min. In the first 20 cycles, T_A decreased stepwise by 1°C every 2nd cycle, from 65°C in the first cycle to 56°C in the 20th. In the last 7 cycles, T_A was 55°C. Cycling was followed by

5 min incubation at 72°C. Agarose gel electrophoresis was used to detect and estimate the concentration of PCR amplicons.

DGGE and banding pattern analysis. DGGE was performed using a Dcode™ Universal Mutation Detection System (Bio-Rad). Similarly sized PCR products were separated on a 1.5 mm-thick vertical gel containing polyacrylamide (acrylamide: bisacrylamide ratio of 37.5:1) and a linear gradient of the denaturants urea and formamide, which increased from 40% at the top of the gel to 65% at the bottom. Similar amounts of PCR products were loaded on the DGGE gel. Electrophoresis was performed at 60°C in a 0.5 × TAE buffer, and a voltage of 75 V was applied to the submerged gel for 16 h. Nucleic acids were visualized by staining with SYBR Gold and photographed (Øvreås et al. 1997). Banding patterns (absence and presence of bands) were analyzed by Quantity One software (Bio-Rad). A pairwise distance matrix was calculated and analyzed with weighted pair group mean average (WPGMA) cluster analysis and presented as a dendrogram.

Sequencing and phylogenetic analysis. PCR products from bacterial isolates were purified using a Qiaquick PCR purification kit (Qiagen). Purified PCR products were sequenced with primer 8F using Bigdye-terminator chemistry by an ABI PRISM310 or 3100 Genetic Analyzer (Applied Biosystems). For the bacterioplankton community, representative DNA bands were excised from the gels, re-amplified, and analyzed with DGGE again. These procedures were repeated three times. The DGGE bands were sequenced using the primer F1070.

All sequences were compared with the GenBank database using BLAST, and the closest bacterial strain matches were obtained. Phylogenetic trees were

constructed using MacVector 7.2 software package (GCG). Briefly, sequence alignment was performed with the program CLUSTAL W. Evolutionary distances were calculated using the Jukes-Cantor method (Jukes and Cantor 1969) and a distance tree was constructed using the neighbor-joining algorithm (Saitou and Nei 1987). Bootstrap values were obtained from the analysis of 1000 re-samplings of the data set.

PCR-DGGE detection limit. Different concentrations (10^2 to 10^6 cells ml^{-1}) of three bacterial strains (*Shewanella* sp., *Vibrio* sp., and *Pseudomonas* sp.) were added to a natural bacterial community collected from the Inner Harbor. Preliminary results showed that a finer range of concentrations (10^3 to 10^5 cells ml^{-1}) of bacterial isolates was required. *Vibrio* sp., *Pseudomonas* sp. and *Synechococcus* sp. CB0101 were selected because they represented groups with high (9 copies), medium (4 copies), and low (2 copies) rRNA operon copy numbers, respectively. DNA was extracted from the seeded bacterial communities following the protocol applied to natural microbial assemblages, and analyzed by PCR-DGGE following the protocols described above.

Nucleotide sequence accession numbers. Sequences of bacterial isolates and DGGE bands obtained in this study were deposited in the GenBank database under accession numbers AY789535 to AY789582 and AY654428 to AY654452.

Results

Water temperature of the Baltimore Inner Harbor varied widely across seasons. The lowest water temperature was 6°C in the winter while the highest temperature (27°C) occurred in the summer. Salinity was highest (18 ppt) in the

winter and lowest (5 ppt) in the spring. Starting in April, salinity dropped dramatically until May. From June to November salinity was relatively stable (11 to 15 ppt). The temperature and salinity recorded in this study were similar to previous and current monthly monitoring data in the Inner Harbor (http://mddnr.chesapeakebay.net/bay_cond/index.cfm). The bacterial density fluctuated from 1.2 to 3.0 × 10^6 cells ml^{-1} throughout the year, and did not exhibit an obvious seasonal pattern.

A total of 240 bacterial strains from the Inner Harbor were isolated and purified during the sampling year. Characterizing the isolated bacteria from the Inner Harbor was not the main focus of this study. Only 48 strains from summer and winter were picked and characterized based on the partial 16S rRNA gene sequence (>500 bp, Table 2-1), because our DGGE gel showed that bacterial communities were very different between these two seasons. Most isolates were similar to known cultivated bacteria in the GenBank. Seventeen out of the 27 winter isolates were most closely

Table 2-1. Identification of bacterial strains isolated from the Inner Harbor based on the partial 16S rRNA gene sequence analyses. Shaded areas showed the species that appeared both in summer and winter. N/A: not available source from the original submission.

ID	Sample collection time	Most closely related organism (based on partial 16S rRNA gene)	Identity (%)	Accesion no.	Phylogenetic group	Source (Refer to the orginal submission)	No. of isolates in warm season	No. of isolates in cold season
IH3-15	Mar. 2001	Arctic sea ice bacterium ARK10032	98	AF468385	γ	Arctic sea ice		1
IH3-10	Mar. 2001	Arctic sea ice bacterium ARK10036	96	AF468358	miscellaneous	Arctic sea ice		1
IH2-13	Feb. 2002	*Comamonas testosteroni WDL7*	97	AF538933	β	Degradation of linuron		1
IH3-6	Mar. 2001	*Flavobacterium frigidarium*	99	AF162266	Bacteroidetes	Antarctica		1
IH3-5	Mar. 2001	*Flavobacterium* sp.	99	U63938	Bacteroidetes	Northern Baltic Sea		1
IH2-10	Feb. 2002	*Idiomarina* sp. Loihi-Chm (16S)-1	94	AB049741	γ	Deep low-temperature vent Hawaii		1
IH2-7	Feb. 2002	Marine Bacterium SCRIPPS_101	97	AF359537	miscellaneous	Associated with dinoflagellates		1
IH3-16	Mar. 2001	Polar sea bacterium R7076	98	AJ295713	miscellaneous	Polar seas		1
IH3-2, IH3-8	Mar. 2001	Polar sea bacterium R7216	100	AJ295714	miscellaneous	Polar seas		2
IH2-1, IH2-6	Feb. 2002	*Pseudoalteromonas* sp. UL1	96	AF172991	γ	Marine alga *Ulva lactuca*		2
IH2-14	Feb. 2002	*Pseudomonas rhizosphaerae*	99	AY152673	γ	Spain soil		1
IH2-5, IH2-11, IH2-20	Feb. 2002	*Pseudomonas* sp. Hsa.28	96-98	AY259121	γ	Freshwater		3
IH2-2	Feb. 2002	*Pseudomonas stutzeri*	98	U65012	γ	Denitrification		1
IH2-12	Feb. 2002	*Psychrobacter maritimus*	99	AJ609272	γ	Sea ice and sediment, Sea of Japan		1
IH2-8, IH2-16, IH2-18	Feb. 2002	*Shewanella baltica* NCTC10735	97-99	AJ000214	γ	Baltic sea		3
IH3-3, IH3-11	Mar. 2001	*Shewanella frigidimarina* ACAM 588	98	U85905	γ	Antarctica		2
IH2-9, IH3-4, IH3-14	Feb. 2002, Mar. 2001	*Vibrio ordalii* NCMB2168	97-99	X74718	γ	Southern Chile		3
IH2-19, IH8-15	Feb. 2002, Aug. 2001	*Pseudomonas pseudoalcaligenes* strain LMG 1225T	96, 97	Z76666	γ	Lake Kauhako, Hawaii	1	1
IH8-2	Aug. 2001	*Aeromicrobium erythreum*	95	AF005021	Actinobacteria	Puerto Rico soil	1	
IH8-12	Aug. 2001	*Bacillus marisflavi* strain TF-11	98	AF483624	Firmicutes	Yellow Sea, Korea	1	
IH8-1	Aug. 2001	*Brevundimonas vesicularis*	96	AJ627402	α	Freshwater biofilm	1	
IH7-4	Jul. 2001	*Flavobacterium* sp. EP215	97	AF493657	Bacteroidetes	River Taff epilithon, UK	1	
IH7-14	Jul. 2001	Gamma proteobacterium GMD16F03	99	AY162108	γ	Sargasso Sea	1	
IH7-1	Jul. 2001	Marine alpha proteobacterium AS-19	100	AJ391181	α	Adriatic Sea	1	
IH7-9	Jul. 2001	*Massilia* sp. 72	95	AY177372	β	Soil	1	
IH7-6	Jul. 2001	*Microbacterium imperiale* 51-6C	100	AF526906	Actinobacteria	Spacecraft	1	
IH8-3	Aug. 2001	*Paracoccus haeundaesis* strain BC74171	98	AY189743	α	Halophilic astaxanthin-production	1	
IH8-16	Aug. 2001	*Pseudoalteromonas* sp. RE2-5b	97	AF539777	γ	Biofilms	1	
IH7-13	Jul. 2001	*Pseudoalteromonas* sp. RE2-12b	93	AF539775	γ	Biofilms	1	
IH8-19	Aug. 2001	*Pseudomonas alcaligenes* isolate LB19	97	AF390747	γ	N/A	1	
IH8-6	Aug. 2001	*Pseudomonas putida* KL33	95	AY686638	γ	N/A	1	
IH7-18	Jul. 2001	*Pseudoxanthomonas* sp. S5-25	97	AF530282	γ	Canadian paper mill	1	
IH7-3	Jul. 2001	Rape rhizosphere bacterium tsb058	99	AJ295445	miscellaneous	Rhizoplane of Oilseed Rape	1	
IH7-12, IH7-17	Jul. 2001	*Shewanella amazonensis*	96, 97	AF005248	γ	Amazonian shelf muds	2	
IH7-7	Jul. 2001	*Shewanella* sp. 184	99	AF387349	γ	Butter	1	
IH7-15	Jul. 2001	Uncultured alpha proteobacterium clone SM1E02	96	AF445680	α	Mammoth Hot Springs, Yellowstone	1	
IH7-20	Jul. 2001	Uncultured gamma proteobacterium strain GWS-BW-H33M	93	AY515442	γ	Intertidal Mudflats	1	

affiliated with strains previously isolated from cold environments including polar seas, Arctic sea ice, Antarctica, the Baltic Sea etc. These coldwater isolates were not observed in the summer months (Table 2-1).

DGGE revealed that the bacterial population structure in the Inner Harbor changed from month to month (Fig. 2-1). From winter to spring, the band patterns shifted significantly. However, from early summer to fall, the bacterial communities became relatively stable and showed similar band patterns. The monthly samples were grouped into three seasonal types: winter (December and January), spring (February to May) and summer-fall (June to November) based on the similarity dendrogram (Fig. 2-2). The number of DGGE bands in the summer-fall season (avg. = 34.0, n = 6) was higher than that in the winter and spring seasons (avg. = 25.5, n = 6). Twenty-eight bands were selected and excised. To confirm that bands from the same position in different samples represented the same organism, four additional bands (Inner Harbor [IH]-6', 11', 21', and 28') were sequenced. These 'replicated' bands were identical to IH-6, 11, 21 and 28, respectively. A total of 25 DGGE phylotypes were identified based on their 16S rRNA gene sequences. Eight sequences were most closely related with *Alphaproteobacteria*, six with chloroplasts, four with *Betaproteobacteria*, three with *Planctomycetes*, two with *Gammaproteobacteria*, one with *Bacteroidetes* and 1 with *Cyanobacteria* (*Synechococcus* sp.) phylotype (Fig. 2-3). Bands IH-1, 3, 4 and 20 could not be re-amplified and therefore no sequences were achieved.

A phylotype of uncultured *Alphaproteobacteria* (IH-9) was present all year round, but *Roseobacter* spp. (IH-14 and IH-17) and *Rhodobacter* sp. (IH-13b) only

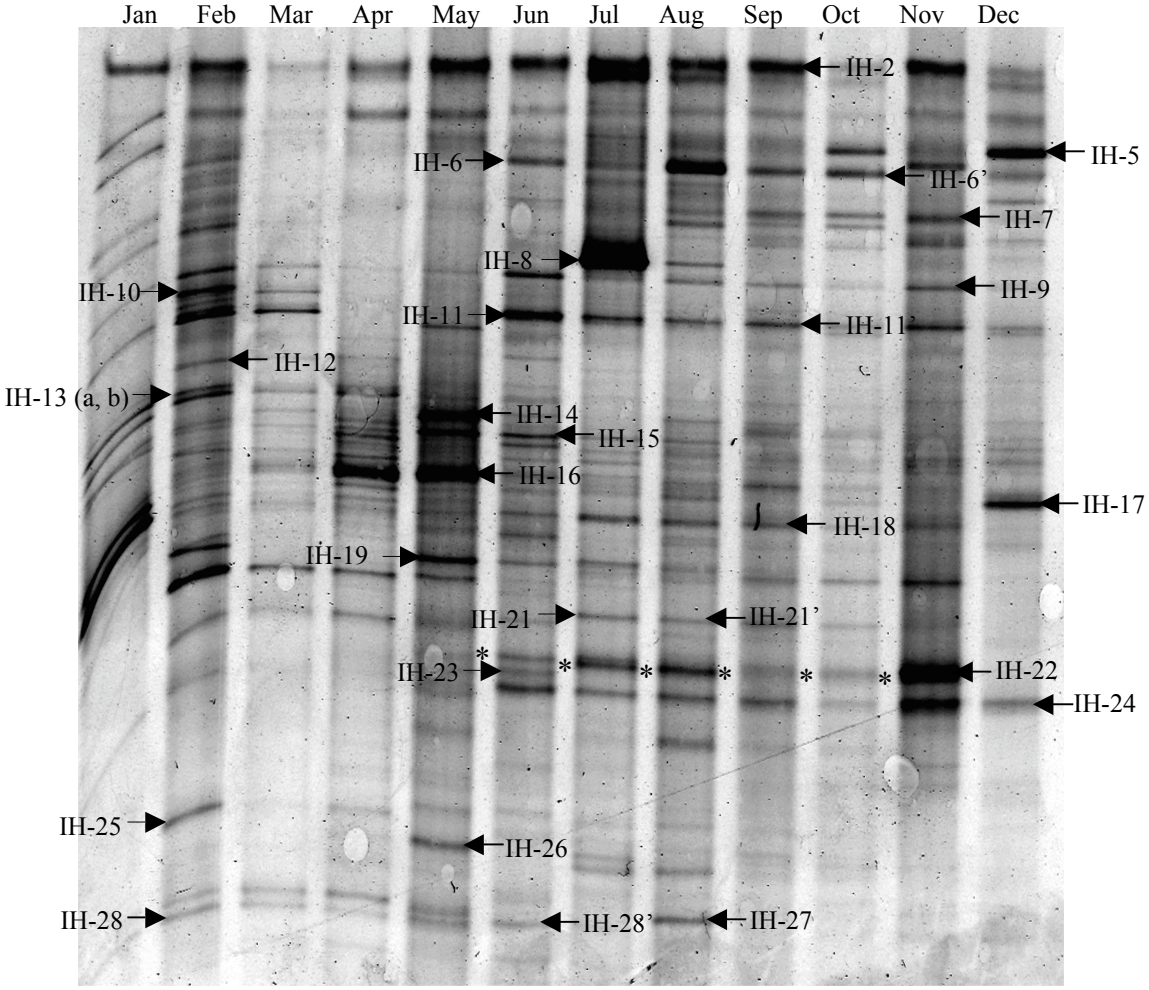


Fig. 2-1. Monthly patterns of Inner Harbor bacterial composition revealed by DGGE. Lanes 1 to 2 correspond to January and February 2002; Lanes 3 to 12 correspond to March to December 2001. Numbered bands are those excised and sequenced; *Synechococcus* sp. bands labeled with *. From July to November, *Synechococcus* sp. cell counts were 8.1×10^4, 5.2×10^4, 7.4×10^4, 2.0×10^4, and 1.1×10^4 cells ml^{-1}, respectively (Wang and Chen 2004).

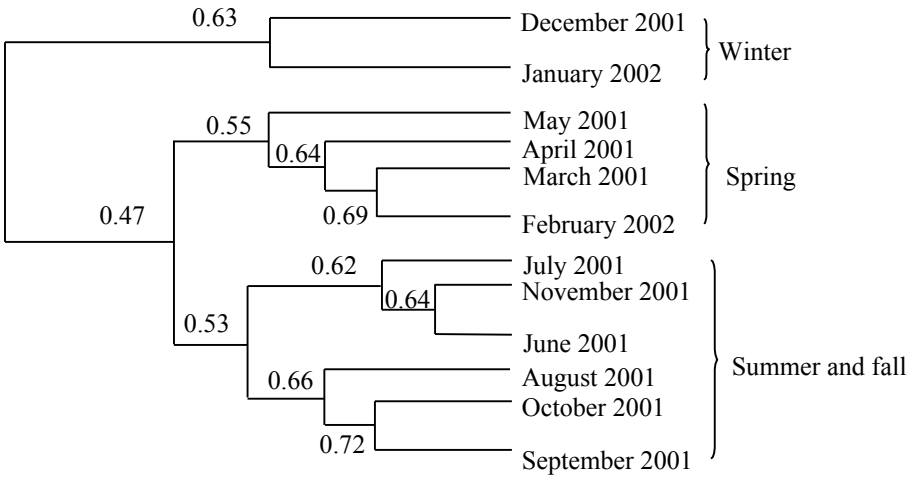


Fig. 2-2. Dendrogram based on similarity matrix of bacterioplankton DGGE fingerprint banding patterns from 12 monthly samples. Similarities (as proportions of 1) between band patterns are indicated at branch divisions.

appeared in winter and spring. Marine *Alphaproteobacteria* (IH-24) was restricted to summer-fall and winter. Phylotypes of *Betaproteobacteria* shifted between winter-spring and summer-fall, and changed from *Aquaspirllum* sp. and *Variovorax* sp. (IH-15 and IH-16; winter-spring) to *Hydrogenophaga flava* and *Alcaligenes* sp. (IH-19 and IH-21; summer-fall). *Gammaproteobacteria* (IH-12 and IH-13a) were only present in winter and early spring. *Cyanobacteria* (*Synechococcus* sp.) were commonly found in the summer-fall season but disappeared in winter and spring. Similar seasonal patterns were found in some plastids (IH-6, IH-7 and IH-8). However, in winter, different plastids (IH-5 and IH-10) became prevalent. *Bacteroidetes* (IH-28) and *Planctomycetes* (IH-27) were relatively dominant in the spring and early summer, but absent in other seasons (Figs. 2-1 & 2-3).

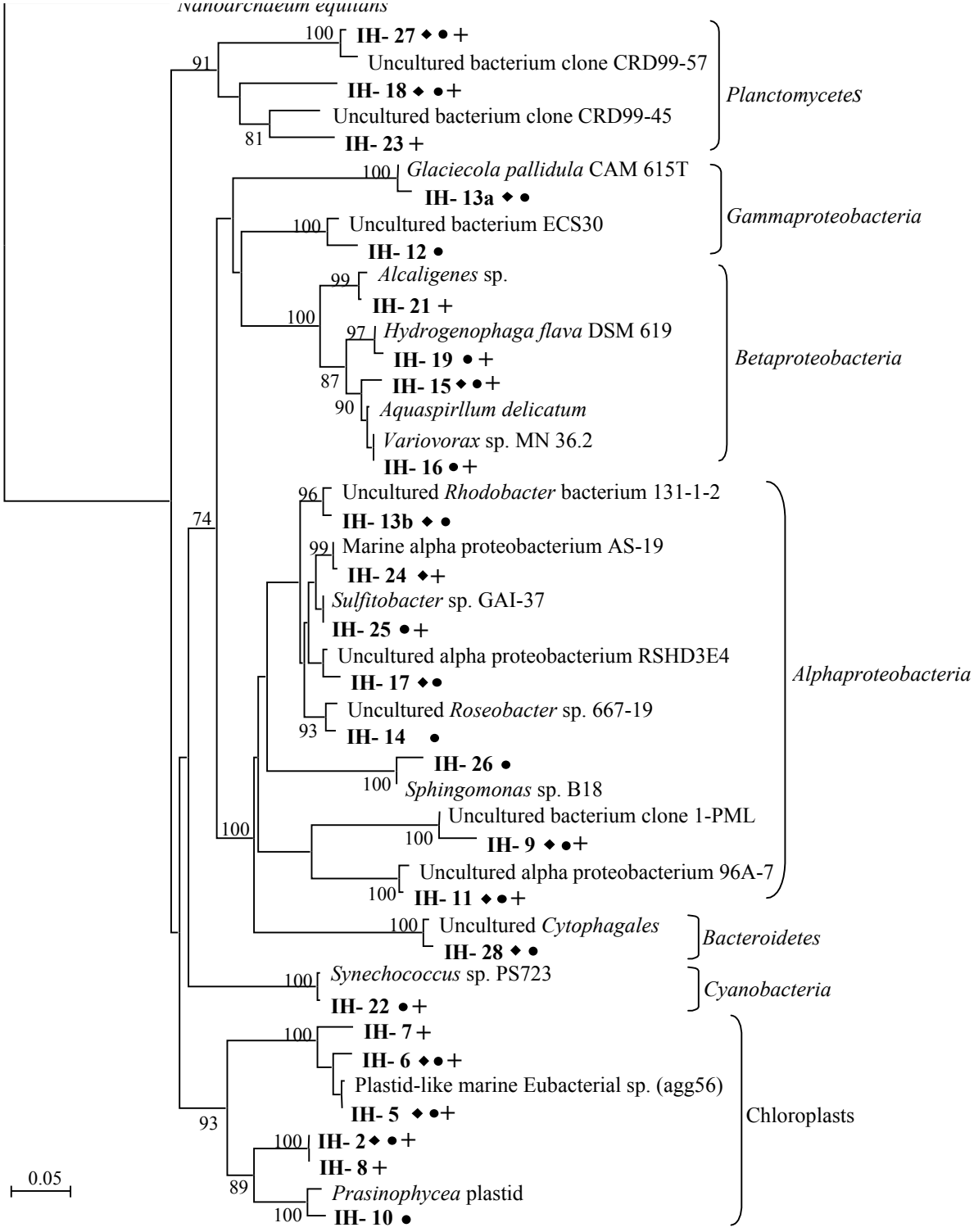


Fig. 2-3. Phylogenetic affiliations of representative 16S rRNA gene sequences from excised DGGE bands. Sequences from this study in bold. ◆: DGGE band present in winter; ●: DGGE band present in spring; +: DGGE band present in summer-fall. *Nanoarchaeum equitans* was used as an outgroup. Scale bar= 0.05 substitutions site^{-1}.

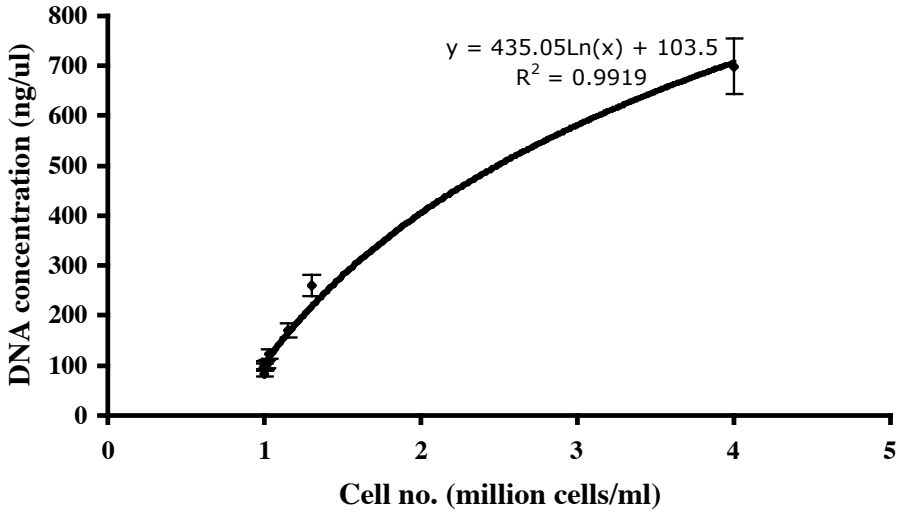


Fig. 2-4. DNA extraction efficiency of seeding experiment. Equation shows relationship between DNA concentration and total cell numbers.

Because few of the characterized heterotrophic bacteria isolates were seen in the DGGE phylotypes, additional experiments were conducted to understand the discrepancy between culture-dependent and molecular approaches. Different concentrations of cultivated bacterial cells were added to a natural microbial assemblage in order to estimate the detection threshold of DGGE. DNA extraction efficiency was tested. With the increase of cells amended, DNA concentration increased and showed a close relationship with cell densities (Fig. 2-4). The corresponding bands for *Vibrio* sp. and *Shewanella* sp. could be conveniently detected when cell densities were above 5×10^3 cells ml^{-1}, whereas *Pseudomonas* sp. presented weak bands even when the density reached 1×10^4 cells ml^{-1} (Fig. 2-5). Furthermore, when the concentration of seeded bacteria increased to 10^6 cells ml^{-1},

most DGGE phylotypes from natural water became undetectable on the gel (Fig. 2-5, Lane 8).

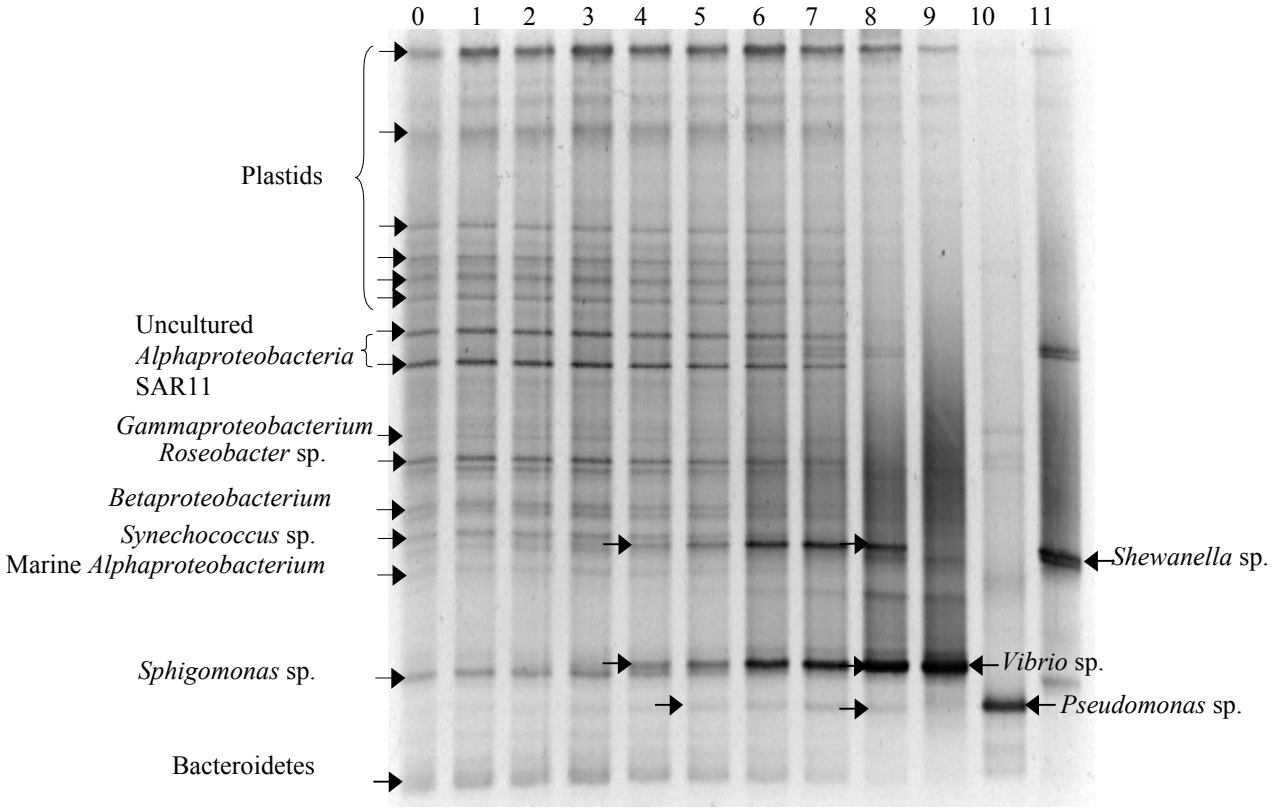


Fig. 2-5. DGGE detection threshold measured by seeding a natural microbial community with different concentrations of three bacterial strains (*Vibrio* sp., *Pseudomonas* sp., and *Shewanella* sp.). Lanes 0 to 8: negative control (0 cells), 100, 500, 1×10^3, 5×10^3, 1×10^4, 5×10^4, 1×10^5 and 1×10^6 cells ml^{-1} added, respectively. Lanes 9 to 11: DGGE fingerprints for pure cultures of *Vibrio* sp., *Pseudomonas* sp. and *Shewanella* sp., respectively. Closest phylogenetic affiliations of band sequences shown on left.

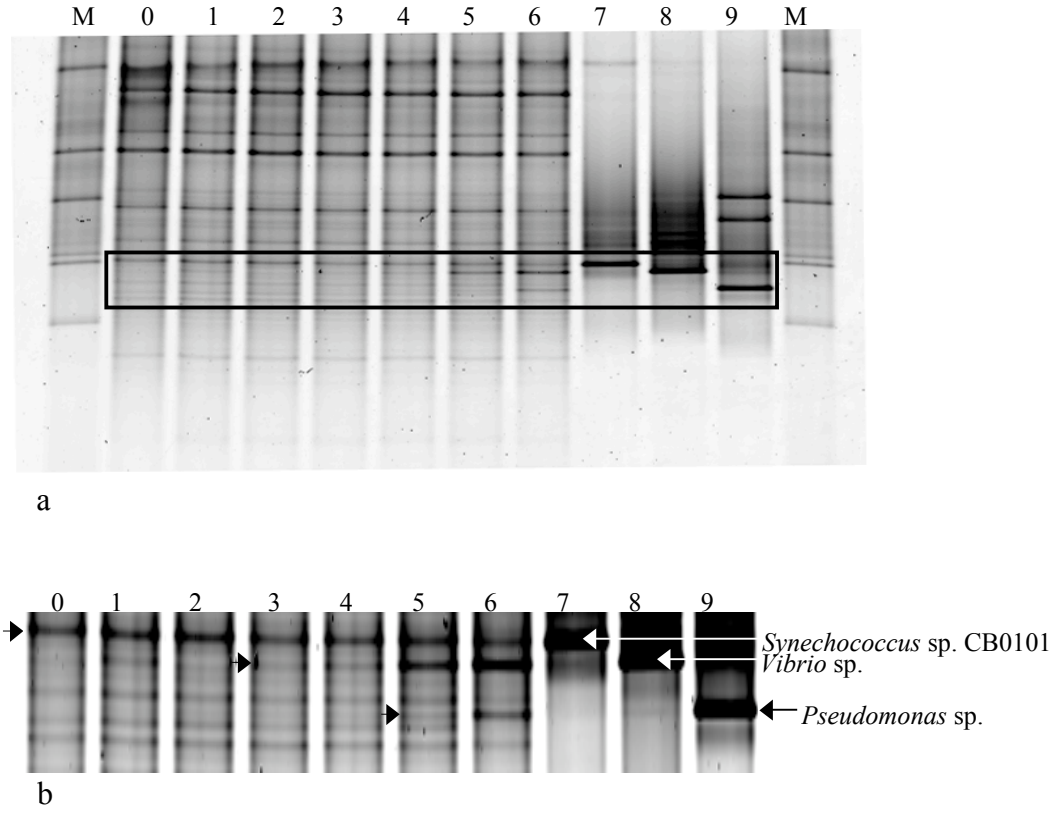


Fig. 2-6. DGGE detection threshold measured by seeding a natural microbial community with finer range concentrations of three bacterial strains (*Synechococcus* sp. CB0101, *Vibrio* sp., and *Pseudomonas* sp.). (a) Lanes 0 to 6: negative control (0 cells), 1×10^3, 2×10^3, 2.5×10^3, 5×10^3, 1×10^4, and 1×10^5 cells ml^{-1} added, respectively. Lanes 7 to 9: DGGE fingerprints for pure cultures of *Synechococcus* sp., CB0101, *Vibrio* sp., and *Pseudomonas* sp., respectively. M: marker consisting of chloroplast, uncultured α-proteobacterium, *Roseobacter* sp., β-proteobacterium, *Synechococcus* sp. and *Cytophaga* sp. phylotypes. (b) Enhanced view of subsection highlighted in (a).

The detection threshold was further analyzed by adding a more stringent range of cells (Fig. 2-6a). At a concentration of 2.5×10^3 cells ml^{-1}, the band corresponding to *Vibrio* sp. with high gene copy numbers could be detected. However, the band representing *Pseudomonas* sp. could not be seen until cell density reached 1×10^4 cells ml^{-1} (Fig. 2-6b). With an average bacterial concentration of 2.43×10^6 cells ml^{-1} in the water, the detection threshold for *Vibrio* sp. and *Pseudomonas* sp. corresponded to 0.1 and 0.4% of total bacterial cell counts, respectively. The band corresponding to the introduced *Synechococcus* sp. strain ran at the same position as a band from the natural assemblage. Sequencing of the bands showed that they were identical phylotypes. Therefore, the detection limit for *Synechococcus* sp. could not be properly evaluated.

Discussion

Based on a single rich medium, our cultured bacterial isolates provided only a 'snapshot' of 'easy-to-culture' bacteria of the whole community. It was not our intention in this study to conduct a thorough survey of Inner Harbor bacterial isolates. The region of the 16S rRNA gene used for characterizing bacterial isolates (8F to 785R) did not overlap with the region flanked by DGGE primers (1070F to 1392R). It would be ideal to compare the partial 16S rRNA sequence of bacterial isolates and DGGE bands at the same region. We tested the eight closest hits of our sequences and BLAST using a short (500 bp) and long (~1400 bp) fragment of 16S rRNA gene sequences. Six of the eight sequences (AF468358, AF538933, AJ295714, AY162673, Z76666, and AF483624) showed similar BLAST outcomes at the strain or species

level. Therefore, the comparison between bacterial isolates and DGGE phylotypes should not have been significantly affected by the non-overlapped primers.

It was intriguing to learn that few of the 48 characterized bacterial isolates matched with phylotypes from DGGE band sequences. Is the cell density of isolated bacteria in natural waters too low to be detected by PCR-DGGE? If so, what is the detection threshold of PCR-DGGE for estuarine bacterioplankton? Muyzer et al. (1993) and Murray et al. (1996) reported that PCR-DGGE is sensitive enough to detect 1 to 2% of bacterial populations in the mixed assemblage of selected bacterial strains. The introduction of serially diluted bacteria into a natural community allowed us to get a direct estimate of the detection limit based on cell number. Our experiments with seeded bacteria suggested that the detection threshold varied with different bacterial species. For example, *Vibrio* sp. could be detected at 2.5×10^3 cells ml^{-1} (Fig. 2-6), which corresponded to ~0.1% of the total bacterial population (2.1×10^6 cells ml^{-1}). However, *Pseudomonas* sp. was not detectable until cell density reached 1×10^4 cells ml^{-1} corresponding to ~0.5% of the total population (Figs. 2-5 and 2-6). Interestingly, the *Synechococcus* sp. bands detected (Fig. 2-1) corresponded to samples where *Synechococcus* counts were over 1.1×10^4 cells ml^{-1} (Wang and Chen 2004). No *Synechococcus* sp. DGGE bands were detectable when *Synechococcus* sp. cell densities were low. This result also supports a previous study on the detection limit of *Synechococcus* sp. in Lake Cisó and Lake Vilar, NE Spain (Casamayor et al. 2000).

It is known that many factors (DNA extraction, primer selection, PCR cycles, gene copy number, etc.) can influence the outcome of PCR, particularly when applied

to environmental samples (e.g. Farrelly et al. 1995; Suzuki and Giovannoni 1996; von Wintzingerode et al. 1997; Crosby and Criddle 2003). We characterized the total bacterial community in the Inner Harbor, without pre-filtration to remove phytoplankton cells. Six plastid sequences related to eukaryotic algae were identified in the DGGE fingerprints. The presence of algal DNA will affect the DGGE patterns of the bacterial community, but will also provide valuable information on the population dynamics of dominant phytoplankton. The primers used in this study yielded 320 bp amplicons, which overlapped with the V8 variable region in the 16S rRNA gene. Recently, different sets of DGGE primers were evaluated based on the bacterial community in the maize rhizosphere or rumen digesta of sheep. Universal primers based on region V3 of 16S rRNA were recommended for shorter fragments, while regions V3-V5 and V6-V8 were suggested for longer fragments (Yu and Morrison 2004). However, to our knowledge, those DGGE primers have not been compared systematically for planktonic bacterial communities. This is an important issue that warrants future study. In a recent study in the Chesapeake Bay, our DGGE band sequences matched well with clone library sequences (16S–ITS–23S region) (Chapter 3 and 4), which suggested that the major bacterioplankton populations were not distorted by the DGGE primers we used.

It is worth pointing out that the detection limit of PCR-DGGE in our study appeared to be affected by gene copy number. Ribosomal RNA gene-based molecular techniques (i.e. PCR-DGGE, LH-PCR, ARISA, TRFLP etc.) present a quantitative bias towards organisms with higher gene copy numbers (Crosby and Criddle 2003). At the time, the Ribosomal RNA Operon Copy Number Database (rrndb)

(Klappenbach et al. 2001) contained 502 entries, with 259 genome sizes and rRNA operon copy numbers available. There was no obvious relationship between rRNA operon copy number and genome size ($r^2 = 0.18$) (Fig. 2-7). We chose several bacterial strains for the detection threshold study, selected to reflect different rRNA gene multiplicity. Genomes of *Vibrio* spp. and *Shewanella* spp. typically contain 9 copies of the rRNA operon; *Pseudomonas* spp. has on average 4 copies, while *Synechococcus* spp. contains 2 copies (Fig. 2-7). Our results indicated that the detection threshold for PCR-DGGE was affected by the gene copy number of the 16S rRNA operon. Bacterial groups in natural environments with low rRNA operon copy numbers may need to reach higher cell densities to be detectable by PCR-DGGE.

The detection threshold of DGGE provides a possible explanation as to why most cultured bacteria could not be detected by DGGE. It is likely that the concentration of most bacteria grown on enriched media was low in the natural samples. On average, the total colony counts on plates were 6450 cells ml^{-1} (n = 8). Assuming that the colonies on one plate were equally derived from 10 different bacterial species, the abundance of each species accounted for 0.265% of the total cell density (avg. 2.43×10^6 cells ml^{-1}). Therefore, their concentrations were lower than the detection limit and they were absent from the DGGE gel. A limited number of bands indicated that DGGE is biased toward abundant groups in the community and underestimates actual bacterial diversity in the samples.

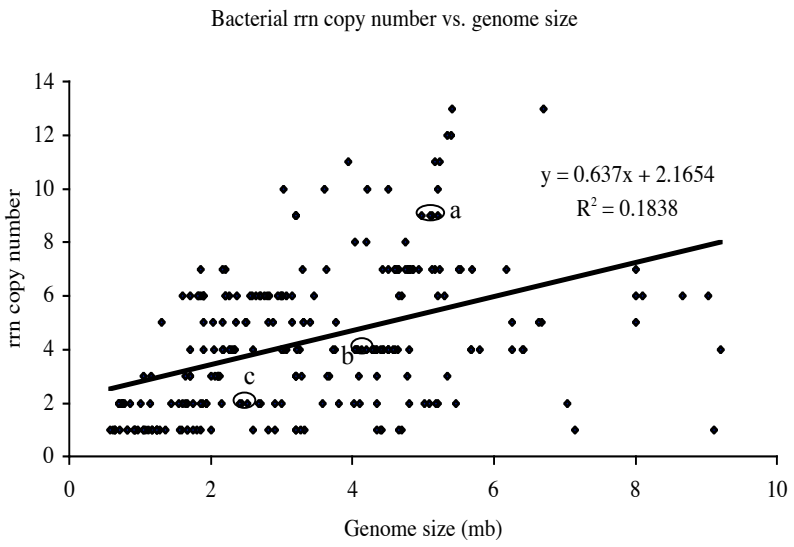


Fig. 2-7. Distribution of *rrn* operon copy numbers vs. genome size of *rrndb* bacterial entries. a, b, and c represented closest relatives of bacterial isolates added in seeding experiments; a: *Vibrio* spp. and *Shewanella* spp.; b: *Psudomonas* spp.; c: *Synechococcus* spp.. At this time, *rrndb* comprises 259 entries with both *rrn* operon copy number and genome size available.

Our results demonstrated that bacterial community structure in the estuarine ecosystem is variable between months, but that a seasonal pattern could be identified. Overall, bacterial communities in the summer-fall season were more similar to each other, whereas winter communities appeared to be distinct from spring, summer and fall. Two populations that corresponded to phylotypes of Cyanobacteria (*Synechococcus* sp.) and marine *Alphaproteobacterium* exhibited a similar seasonal pattern in the Inner Harbor. Occurrence of the *Synechococcus* sp. in the DGGE analysis supported the seasonal distribution of *Synechococcus* sp. cell densities in the harbor (Wang and Chen 2004). The abundance of *Synechococcus* sp. reached 8.14 ±

0.98×10^4 cells ml^{-1} in summer (July), and decreased to $2.57 \pm 0.53 \times 10^2$ cells ml^{-1} in winter (February). The marine *Alphaproteobacteria* phylotype (IH-24) followed a seasonal pattern similar to that of the *Synechococcus* sp.. Marine *Alphaproteobacteria* have been found in a number of marine environments (González and Moran 1997; Suzuki et al. 1997). This group of bacteria is numerically dominant (28%) in coastal waters, but not detectable in low salinity (<5 ppt) or freshwater (González and Moran 1997). In the Baltimore Inner Harbor, no marine *Alphaproteobacteria* were found in April and May when salinity was at its lowest (~ 5 ppt). However, marine α-*Proteobacteria* were also not detected in other months (January to March) when salinity was above 15 ppt, which suggested that this group may be sensitive not only to salinity, but also to water temperature.

As *Betaproteobacteria* are dominant in freshwater and have never been found in marine water (Methe et al. 1998; Glöckner et al. 1999), it is believed that this group advected into the estuary from the Patapsco River.

One *Bacteroidetes* phylotype (IH-28) with high G + C content was detected from January to June, but was not detected in other months. The *Bacteroidetes* group is abundant in marine systems (Glöckner et al. 1999) including Delaware estuarine and coastal waters (Cottrell and Kirchman 2000b; Kirchman et al. 2003b). Because these species are known to be involved in the degradation of complex macromolecules (Shewan and McMeekin 1983), they adapt well to water with high particle loads (Cottrell and Kirchman 2000b). In spring, terrestrial run-off from the Patapsco River provides the largest load of nutrients and particles to the Inner Harbor

(Boesch et al. 2001; Sellner et al. 2001), which may offer suitable environmental niches to this group.

Chesapeake estuarine bacterioplankton are composed of mixed populations from both freshwater and marine origin, and the balance of these populations may be interpreted by seasonal variability. It has been reported that bacterioplankton communities in freshwater differ from those in marine communities (e. g. González and Moran 1997; Methe et al. 1998; Glöckner et al. 1999; Cottrell and Kirchman 2000b; Zwart et al. 2002). In estuaries, dominating bacterial groups shift along the salinity gradient from α- and *Betaproteobacteria*, gram-positive bacteria, and *Verrucomicrobia* to α- and γ-*Proteobacteria* (Crump et al. 1999). Increasing precipitation that started from late March in the Inner Harbor (www.atmos.umd.edu/~climate/) resulted in a significant drop in salinity in April. Thereafter, *Betaproteobacteria* (IH-15, IH-16, and IH-19) dominated spring bacterial communities. It is likely that river run-off brings more freshwater populations into the harbor in spring. In contrast, salinities in winter and summer-fall were relatively stable but temperatures changed remarkably, which suggested that the shift in bacterial composition between summer and winter was possibly related to temperature fluctuations rather than to salinity. To understand the interaction between community shifts and environmental factors, we are currently conducting an inter-annual survey to investigate the spatial and temporal variations of Chesapeake Bay bacterioplankton.

Chapter 3: Chesapeake Bay bacterioplankton: richness and diversity revealed by 16S-23S rRNA operon clone libraries

Abstract

In comparison to freshwater and the open ocean, less is known about population structure and seasonal dynamics in estuaries, particularly those with long residence time. The Chesapeake Bay is the largest estuary in the United States but detailed analysis of microbial community composition in the Bay is still lacking. Six clone libraries based on rRNA operon (16S rRNA-ITS-23S rRNA) were constructed from samples collected at two seasons and three domains of the Chesapeake Bay to investigate the spatial and temporal dynamics of its bacterial populations. In September 2002, *Alphaproteobacteria*, *Gammaproteobacteria*, *Bacteroidetes*, *Cyanobacteria* and *Actinobacteria* were the dominant major groups. In contrast, *Alphaproteobacteria*, *Betaproteobacteria*, and *Actinobacteria* were dominant in March 2003. Dramatic seasonal shifts in bacterial species composition (changes within subgroups or clades) were observed for *Alphaproteobacteria* (*Roseobacter* clade, SAR11), *Cyanobacteria* (*Synechococcus*), and *Actinobacteria*, suggesting strong seasonal variation of taxonomic groups. Stronger seasonal than spatial variations of Chesapeake Bay bacterioplankton were also supported by denaturing gradient gel electrophoresis (DGGE) and length heterogeneity (LH)-PCR analysis. Corroboratory previous observations indicated that temperature and organic nutrients might be the main factors influencing bacterial community structure.

Introduction

Estuaries rank among the most productive and dynamic aquatic ecosystems on earth. Mixing of fresh and marine waters and significant recycling of nutrients and organic matter production provide strong environmental gradients to the microbes living in these ecosystems. A number of studies have shown that compositions of freshwater and marine pelagic microbial community are fundamentally different. Open oceans generally contain clones belonging to the SAR11, SAR116, and *Roseobacter* clades of the *Alphaproteobacteria*, *Synechococcus* group of the *Cyanobacteria* and SAR86 clade of the *Gammaproteobacteria*, and members of the *Bacteroidetes* (Britschgi and Giovannoni 1991; Crump et al. 1999; González and Moran 1997; Mullins et al. 1995; Suzuki et al. 2001). In contrast, 34 habitat-specific clusters have been identified as typical freshwater bacteria, including species associated with the *Alphaproteobacteria*, the *Betaproteobacteria*, the *Bacteroidetes*, the *Actinobacteria* and the *Verrucomicrobia* (Glöckner et al. 1999; Zwart et al. 2002). As an interface between freshwater and marine realms, estuaries contain typical taxa from both environments. Along the salinity gradient, planktonic bacteria undergo a strong physiological stress and result in variations of biomass, activities and population composition as well (Jonas and Tuttle 1990; Shiah and Ducklow 1994; del Giorgio and Bouvier 2002; Troussellier et al. 2002; Smith and Kemp 2003). In general, dominant populations shifted from *Betaproteobacteria* in freshwater to *Alpha*- and *Gammaproteobacteria* in marine sections (Bouvier and del Giorgio 2002; Sekiguchi et al. 2002; Henriques et al. 2006). In addition, estuarine bacteria

community may form specific populations indigenous to the local environments (Crump et al. 2004; Henriques et al. 2004; Hewson and Fuhrman 2004).

The Chesapeake Bay is the largest estuary in the United States where extensive ecological surveys have shown that bacterial activities are dynamic. Bacterial biomass, production and growth rate, and respiration varied over time and space (Jonas and Tuttle 1990; Shiah and Ducklow 1994; Ducklow et al. 1999). In general, biomass and bacterial activities peaks at the middle Bay region (Ducklow and Shiah 1993; Shiah and Ducklow 1994; Ducklow et al. 1999; Smith and Kemp 2003). In non-summer seasons, when temperature is below 20°C, bacterial activities changed seasonally and were positively correlated with temperature (Shiah and Ducklow 1994). Besides the environmental conditions, the dynamic of planktonic bacterial activities may result from variability of bacterial composition in the Chesapeake Bay. The population structure of Chesapeake bacterioplankton has been investigated using a variety of molecular tools including analysis of 5S rRNA patterns (Bidle and Fletcher 1995; Noble et al. 1997), fluorescent *in situ* hybridization (Heidelberg et al. 2002), and 16S rRNA-based DGGE analysis (Kan et al. 2006a; Kan et al. 2006b). However, these studies were limited to major bacterial groups (phyla and classes) or specific genera/species. Cloning and sequencing of 16S rRNA libraries have not been applied to Chesapeake Bay bacterioplankton or any other estuaries with long residence time. In order to improve our understanding of the synecology of estuarine bacterioplankton, much more detailed insights into phylogenetic diversity are needed.

In this study, we intended to investigate genetic diversity and population dynamics of Chesapeake Bay bacterioplankton by analyzing six rRNA operon clone libraries constructed from the northern, middle and southern Bay in the cold and warm season, respectively. A recently developed technique, ITS-LH-PCR (internal transcribed spacer-length heterogeneity-PCR, Suzuki et al. 2004) was used to screen bacterial clones and the genes of representative clones were sequenced. Detailed phylogenetic analysis of the sequences retrieved in the study was described separately (Chapter 4). Temporal and spatial dynamics of bacterioplankton were determined based on the clonal composition of these libraries. In addition, DGGE and LH-PCR were also applied to the same water samples for the purpose of comparison.

Materials and methods

Sample collection. Water samples were collected at three stations along the middle axis of Chesapeake Bay on September 26-30, 2002 and March 4-8, 2003 (Fig. 3-1). Three stations, 908 (39°08′ N, 76°20′ W), 818 (38°18′ N, 76°17′ W) and 707 (37°07′ N, 76°07′ W), represented the northern, middle and southern Bay, respectively. At each station, a 500 ml sub-sample was taken from a 10 liter-Niskin bottle (sampled at 2 m depth) on board the R/V *Cape Henlopen*, and filtered immediately through 0.2-μm-pore-size polycarbonate filters (47-mm diameter, Millipore, Billerica, MA). The filters were stored at -20 ˚C prior to DNA extraction. Water temperature, salinity and dissolved oxygen were recorded on board.

Chlorophyll a and nutrients analysis. Chl *a* data were kindly provided by Wayne Coats at Smithsonian Environmental Research Center. Duplicate samples (100 ml) from each station were vacuum filtered (<150 mm Hg) onto 25 mm Whatman GF/C

filters and Chl a extracted in 90% acetone for 24 h at 4˚C in the dark. Chl a concentration was determined fluorometrically using a Turner Designs 10-AU fluorometer. Nutrient data including ammonia, nitrite and nitrate, and phosphate were determined by Technicon AutoAnalyzer II at the Horn Point Analytical Services Laboratory (www.hpl.umces.edu/services/as.html). The analysis followed standard methods for chemical analysis of water and wastes proposed by USEPA (USEPA, 1983).

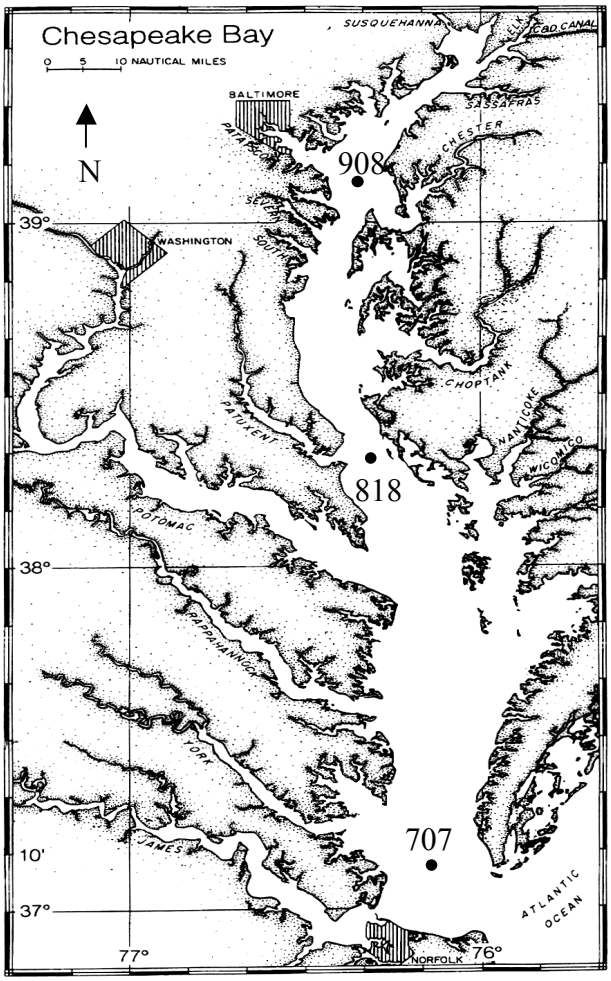


Fig. 3-1. Chesapeake Bay map showing sampling stations.

Enumeration of bacteria and viral particles. Subsamples of 50 ml of water were fixed in 1% glutaraldehyde and stored at 4 ^{o}C. For bacterial cell counts, 1 ml of fixed sample was filtered onto a 0.2-μm-pore-size black polycarbonate membrane filter (Osmonics, Minnetonka, MN). For viral particle counts, 200 ml of fixed sample was mixed with 800 ml Tris-EDTA-Sucrose buffer and filtered onto a 0.02-μm-pore-size 25 mm Anodisc membrane filter (Whatman, Maidstone, UK). Samples on filters were stained with 2.5 × SYBR Gold solution for 15 minutes in the dark as described previously (Chen et al. 2001). Both bacterial cells and viral particles were enumerated under blue excitation (485 nm) on a Zeiss Axioplan epifluorescence microscope (Zeiss, Germany). At least 200 bacterial cells or viral particles per sample were counted.

Extraction of nucleic acids. Bacterial genomic DNA was extracted as previously described (Kan et al. 2006b). DNA concentration was measured using a SmartSpec TM 3000 spectrophotometer (Bio-Rad, Hercules, CA).

Clone library analysis. Clone libraries containing a large portion of the rRNA operon (16S rRNA-ITS-23S rRNA) of bacterioplankton from the six environmental samples described above were constructed using primer set 16S-27F and 23S-1933R (Table 3-1) as previously described (Suzuki et al. 2000), except that 1) Platinum HIFI polymerase mix (Invitrogen, Carlsbad, CA) was used to provide hotstart amplification, 2) the products were A-tailed using the Qiagen A-addition kit (Qiagen, Chattsworth, CA), and 3) products were cloned using the TOPO TA (Invitrogen, Carlsbad, CA) cloning kit following the manufacturer's instructions. A total of 576 clones from 6 libraries were prescreened by a modified screening method adapted

from ITS-LH-PCR that measures the length heterogeneity of the ITS region, as well as the presence and the location of the tRNA-alanine gene within the ITS (Suzuki et al. 2004; Chapter 4). Representative clones putatively identified as different groups

Table 3-1. Primers used for clone library, DGGE and LH-PCR

Primer	Sequence (5'-- 3')	Target site	Reference
Clone library			
16S-27F	AGAGTTTGATCMTGGCTCAG	16S 7-27	Giovannoni 1991
23S-1933R	ACCCGACAAGGAATTTCGC	23S 1933-1951	Amann et al., 1995
DGGE			
16S-1070F	ATGGCTGTCGTCAGCT	16S 1055-1070	Ferris et al., 1996
16S-1392R(GC) [1]	ACGGGCGGTGTGTAC	16S 1392-1406	Ferris et al., 1996
LH-PCR			
16S-27F (FAM) [2]	AGAGTTTGATCMTGGCTCAG	16S 7-27	Giovannoni 1991; Suzuki et al., 1998
16S-355R	GCTGCCTCCCGTAGGAGT	16S 338-355	Amann et al., 1995

[1] 16S-1392R(GC) contains a 40 bp GC - rich clamp at 5' end: CGCCCGCCGC GCCCCGCGCCCGGCCCGCCGCCCCCGCCCC.
[2] 5' end labeled with the phosphoramidite dye 6-FAM

based on fragment lengths were chosen for 16S rRNA gene sequencing and phylogenetic analysis (Chapter 4).

DGGE. Partial 16S rRNA gene from each microbial community was PCR amplified using primers 16S-1070F and 16S-1392R (Muller et al. 1996; Table 3-1). PCR amplicons were subject to DGGE analysis following the methods described elsewhere (Kan et al. 2006a). Briefly, PCR products were loaded on the polyacrylamide gel with

gradient from 40-55%. Electrophoresis was run at 60°C in 1 × TAE buffer, and 70 volts for 16 h. The gel was stained with SYBR Gold (Invitrogen, Carlsbad, CA). Representative DNA bands were excised from the gel, re-amplified and re-analyzed with DGGE. PCR products were purified using the Qiagen PCR Purification Kit (Qiagen, Chattsworth, CA), according to the manufacturer's protocol and sequenced with primer 16S-1070F.

LH-PCR. Two hypervariable regions of the 16S rRNA gene (V1, *E. coli* 16S rRNA gene positions 72 to 101 and V2, *E. coli* 16S rRNA gene positions 176 to 221) were included in LH-PCR by use of 6-FAM labeled primer 16S-27F and primer 16S-355R (Table 3-1), as previously described (Suzuki et al. 1998). Sizes of peaks were analyzed by ABI Genescan software based on the GeneScan 2500 Rox size standard. Peaks less than 5 times the baseline fluorescence intensity were excluded from the analysis. The relative abundance of each peak was estimated by dividing integrated fluorescence of an individual peak by the total integrated fluorescence of all peaks.

Diversity analysis. Clone library coverage (C) was calculated by the equation $C = 1 - (n / N) \times 100$, where n is the number of unique clones and N is the total number of clones examined (Ravenschlag et al. 1999). Rarefaction curves were interpolated using the freeware program aRarefactWin (Holland 1998) with the analytical approximation algorithm (Hurlbert 1971) and 95% confidence intervals (Heck et al. 1975). Statistical methods for species richness estimation and diversity indices estimation were based on the "coverage", the fraction of the population represented by the phylotypes that have been discovered in each clone library. Coverage-based estimations for species richness, Shannon-Wiener index (H) and Simpson's index (D)

were calculated by software SPADE (Chao and Shen 2003-2005). For DGGE and LH-PCR, species richness was estimated based on the number of DGGE bands or LH-PCR peaks.

Results

The mean water temperature was 23.8°C for the samples collected in September 2002, and 2.5°C for those in March 2003 (Table 3-2). Concentrations of bacteria and viral like particles in September 2002 were higher than those in March 2003, while concentrations of Chl a, nitrate, nitrate, and dissolved oxygen in March 2003 were higher than September 2002. No significant difference between these two seasons was observed for salinity, ammonia and phosphate (Table 3-2).

Clone library analysis. A total of 576 clones from six clone libraries were analyzed. Distribution frequency of bacterial clones from each clone library was presented in Table 3-3. The mean value of distribution frequency (from the northern, middle and southern Bay) showed that composition of major bacterial groups varied between the cold and warm seasons. *Alpha*-, *Beta*-, and *Gammaproteobacteria* accounted for approximately 21, 2 and 10 % of bacterial communities in September 2002, and 49, 16, and 2% in March 2003, respectively. The FCB group accounted for 10 and 4% of bacterial communities in September 2002 and March 2003, respectively. Cyanobacteria made up 9% of bacterial communities in September 2002, but were not detectable in March 2003. *Actinobacteria* accounted for 40 and 27% of bacterial communities in September 2002 and March 2003, respectively.

Table 3-2. Environmental parameters of sampling stations on the Chesapeake Bay

	September 2002			March 2003		
	Stn. 908	Stn. 818	Stn . 707	Stn. 908	Stn. 818	Stn. 707
Water temperature (°C)	23.3	23.9	24.2	1.2	1.8	4.4
Salinity (ppt)	15.5	19.4	27.0	10.0	15.8	23.0
Dissolved oxygen (mg/L)	6.85	6.71	6.45	14.07	12.08	11.93
Chl a (ug/L)	9.0	5.0	3.0	41.6	22.5	14.9
Ammonia (uM)	1.23	0.79	0.95	1.15	0.59	-
Nitrite and Nitrate (uM)	7.96	4.27	1.37	42.0	17.60	2.83
Phosphate (uM)	1.36	0.46	0.36	0.58	0.48	0.33
Bacterial abundance (10^6 cells/ml)	6.42	4.96	4.11	1.24	0.57	0.45
Cyanobacteria abundance (10^4 cells/ml)	23.0	28.6	36.3	0.09	0.088	0.11
Viral abundance (10^7 cells/ml)	3.78	5.38	5.21	0.98	0.81	0.64

Table 3-3. Clonal composition and distribution of bacterioplankton from the Chesapeake Bay

Bacterial groups	September 2002			March 2003		
	Stn. 908 **CB01** Clone (%)	Stn. 804 **CB11** Clone (%)	Stn. 707 **CB22** Clone (%)	Stn. 908 **CB31** Clone (%)	Stn. 804 **CB41** Clone (%)	Stn. 707 **CB51** Clone (%)
Alphaproteobacteria	**13 (14.3)**	**18 (21.7)**	**23 (26.7)**	**51 (60.0)**	**33 (38.8)**	**40 (48.8)**
SAR11						
SAR11-I[1]	8 (8.8)	7 (8.4)	9 (10.5)			
SAR11-II[1]	3 (3.3)	2 (2.4)	5 (5.8)			
SAR11-III[1]		2 (2.4)	1 (1.2)			
Rhodospirillalles						
		1 (1.2)				
			3 (3.5)			
Roseobacter						
Chesapeake *Roseobacter* I [2]	2 (2.2)	6 (7.2)	4 (4.7)			
Chesapeake *Roseobacter* II [2]				1 (1.2)	1 (1.2)	5 (6.1)
Chesapeake *Roseobacter* III [2]				2 (2.4)	5 (5.9)	6 (7.3)
Chesapeake *Roseobacter* IV [2]				4 (4.7)	5 (5.9)	2 (2.4)
Chesapeake *Roseobacter* V [2]				2 (2.4)	2 (2.4)	1 (1.2)
Chesapeake *Roseobacter* VI [2]				1 (1.2)	1 (1.2)	3 (3.7)
Chesapeake *Roseobacter* VII [2]						1 (1.2)
Slope Strain D14 [2]				5 (5.9)	14 (16.5)	13 (15.9)
Arctic Sea Ice ARK9990 [2]				15 (17.6)	2 (2.4)	
Sulfitobacter mediterraneus					1 (1.2)	1 (1.2)
Rhodobacter						
				15 (17.6)	1 (1.2)	1 (1.2)
Others				5 (5.9)		
Sphigomonas				1 (1.2)		
Others						
Ahrensia kieliense					1 (1.2)	6 (7.3)
Defluvibacter lusatiae						1 (1.2)
Betaproteobacteria	**0**	**3 (3.6)**	**2 (2.3)**	**14 (16.5)**	**21 (24.7)**	**6 (7.3)**
OM 156[3]		1 (1.2)	1 (1.2)			
OM 43[3]		2 (2.4)				
Hydrogenophilus			1 (1.2)			
Beta Fuku 93				1 (1.2)		
Polaromonas				1 (1.2)		
Polynucleobacter				3 (3.5)		
GKS 98 [4]				9 (10.6)	21 (24.7)	6 (7.3)
Gammaproteobacteria	**4 (4.4)**	**12 (14.5)**	**10 (11.6)**	**2 (2.4)**	**0**	**3 (3.7)**
Gamma AGG47[5]	1 (1.1)	1 (1.2)	1 (1.2)			
SDF1-40	1 (1.1)					
Marine Gamma NOR5	1 (1.1)					
CHAB-III-7		1 (1.2)				
OM 60[5]		1 (1.2)				
Unidentified Gamma		1 (1.2)				1 (1.2)
Arctic 96B-1[5]		2 (2.4)	1 (1.2)			
Novel group - *Legionella coxi*		2 (2.4)				
Acinetobacter			1 (1.2)			
Gamma novel CB22H04			1 (1.2)			
KTC 1119			1 (1.2)			
SAR86-IV[1]	1 (1.1)	4 (4.8)	2 (2.3)			
Pseudomonas syringae				1 (1.2)		
Psychrobacter sp.				1 (1.2)		
Gamma Sva0091						1 (1.2)
Psychromonas sp.						1 (1.2)

Table 3-3 (continued)

Deltaproteobacteria	**1 (1.1)**	**1 (1.2)**	**2 (2.3)**	**0**	**0**	**1 (1.2)**
FBC	**11 (12.1)**	**12 (14.5)**	**6 (7.0)**	**2 (2.4)**	**4 (4.7)**	**3 (3.7)**
FBC ML1218M-14	1 (1.1)	1 (1.2)				
FBC ML817J6	1 (1.1)	6 (7.2)				
FBC Novel TAFB64	2 (2.2)	2 (2.4)	1 (1.2)			
Flexibacter aggregans	3 (3.3)					
Flavobacteriaceae UC1	4 (4.4)					
FBC Clone 06		1 (1.2)				
Bacteroidetes OM 273		2 (2.4)	1 (1.2)			
Sphingobacteria sp.			1 (1.2)			
Bacteroidetes AGG58			1 (1.2)			
Cytophaga novel			1 (1.2)			
Haliscomenobacter sp.			1 (1.2)			
Rhodovirga sp.				1 (1.2)		1 (1.2)
Pedobacter sp.				1 (1.2)		
Flavobacterium gelidiacus					1 (1.2)	
Cellulophaga sp.					3 (3.5)	
ATAM173_A3						1 (1.2)
Psychroserpens sp.						1 (1.2)
Cyanobacteria / *Synechococcus*	**5 (5.5)**	**12 (14.5)**	**7 (8.1)**	**0**	**0**	**0**
Plastids	**0**	**0**	**8 (9.3)**	**0**	**1 (1.2)**	**0**
Actinobacteria	**57 (62.6)**	**22 (26.5)**	**27 (31.4)**	**15 (17.6)**	**25 (29.4)**	**27 (32.9)**
Freshwater acI [6]	3 (3.3)	4 (4.8)		2 (2.4)	11 (12.9)	6 (7.3)
Freshwater acII [6]	7 (7.7)	1 (1.2)		9 (10.6)	5 (5.9)	5 (6.1)
Freshwater acIII [6]	2 (2.2)	1 (1.2)	4 (4.7)	2 (2.4)	8 (9.4)	13 (15.9)
Freshwater acIV [6]	7 (7.7)	2 (2.4)		1 (1.2)	1 (1.2)	2 (2.4)
Plankton Marine Actinobacterium	37 (40.7)	12 (14.5)	23 (26.7)			
Sediment Marine Actinobacterium		2 (2.4)				
Novel Actinobacterium CB31D05	1 (1.1)			1 (1.2)		1 (1.2)
Fibrobacteres	**0**	**1 (1.2)**	**0**	**0**	**0**	**0**
Verrucomicrobia	**0**	**2 (2.4)**	**1 (1.2)**	**1 (1.2)**	**1 (1.2)**	**2 (2.4)**
Total valid clones	**91**	**83**	**86**	**85**	**85**	**82**
Chimera or short inserts	**5**	**13**	**10**	**11**	**11**	**14**

[1] Suzuki et al. 2001
[2] Refer to Chapter 4
[3] Rappe et al. 1997
[4] Zwart et al. 2002
[5] Suzuki et al. 2004
[6] Warnecke et al. 2004

Dramatic shifts of bacterial populations between cold and warm seasons were seen at the subgroup level (Table 3-3). Clones associated with SAR11 (SAR11-I, -II, and –III, Chapter 4) were present in September 2002, but not in March 2003. Seven novel subclusters of the Roseobacter clade (I-VII, Chapter 4) were found in the Chesapeake Bay. Six subclusters (II-VII) of Chesapeake roseobacters were present only in March 2003, while subcluster I was found only in September 2002 (Table 3-3). Clones associated with *Beta-* and *Gammaproteobacteria*, and FCB group also exhibited distinct distribution patterns between cold and warm seasons. Freshwater *Actinobacteria* made up 97% of the total *Actinobacteria* group in March 2003. In contrast, planktonic marine *Actinobacteria* represented up to 68% of the total *Actinobacteria* group in September 2002 (Table 3-3).

Noticeable spatial variations were discovered along the northern, middle and southern Bay. Among the 85 clones recovered from the northern Bay in March 2003, 15 clones matched with the isolated Arctic Sea Ice ARK9990 and 20 clones matched the *Rhodobacter* group (Table 3-3). These two types of bacteria were not found in any stations in September 2002, and rarely seen in the middle and southern Bay in March 2003. Clonal composition of *Gammaproteobacteria* and FCB group varied among regions for September 2002 samples (Table 3-3). In general, clone library analysis revealed more dramatic seasonal than spatial changes of Chesapeake Bay bacterioplankton communities.

DGGE and LH-PCR. Distinct seasonal patterns of bacterioplankton populations were also seen based on DGGE analysis (Fig. 3-2). Most dominant bands appeared in September 2002 were absent in March 2003, and *vice versa*. *Alphaproteobacteria*

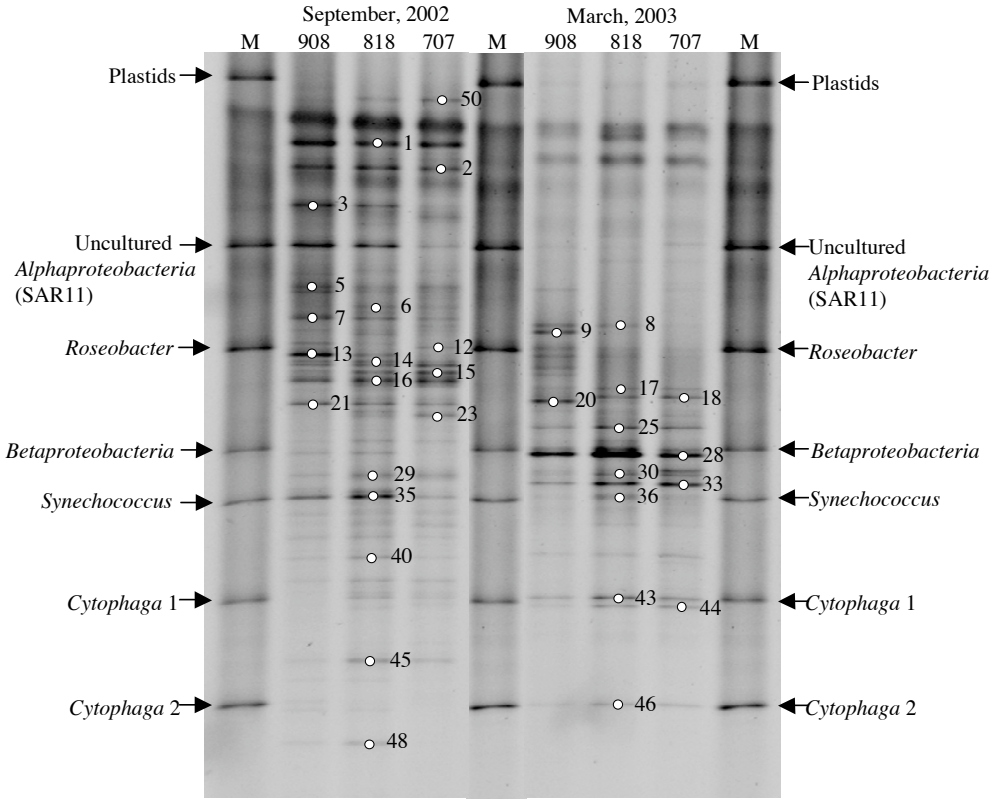


Fig. 3-2. DGGE fingerprints of bacterial communities from the Chesapeake Bay. Representative bands were excised and sequenced (Chapter 5). A total of 53 bands were excised and the missing band numbers (4, 10, 11, 19, 22, 24, 26, 27, 31, 32, 34, 37-39, 41, 42, 47, 49, and 51-53) are from other sampling months (Chapter 5). The markers covered broad range of G+C content and are made by mixture of bands excised from previous environmental samples.

(bands 1, 2, 3, 13, 29 and 40), *Gammaproteobacteria* (bands 7, 12, 14 and 21), *Cyanobacteria* (bands 15 and 35), *Actinobacteria* (bands 16 and 48) were present in September 2002. *Alphaproteobacteria* (bands 8, 9, 20, 30 and 33), *Betaproteobacteria* (bands 17 and 18), and *Actinobacteria* (bands 28, 43 and 44) were

commonly seen in March 2003. Although *Alphaproteobacteria* and *Actinobacteria* were present in both seasons, their band positions and sequences were different. Bacterial communities from the northern, middle and southern Bay exhibited similar DGGE fingerprints for each cruise.

Much like DGGE analysis, samples from the same season showed similar LH-PCR electropherogram profiles. LH-PCR showed more peaks in September 2002 than March 2003 (Fig. 3-3). Based on the calculated values of the length heterogeneity from sequenced clones and those from previous studies (Rappé et al. 1998; Suzuki et al. 1998), major groups and relative abundance were putatively assigned to each peak. Overall, the length heterogeneity of PCR products varied from 313 to 376 bp (Table 3-4). It was difficult to determine the relative abundance of each group but valuable information can be obtained based on the relative intensities of peaks. Similar to clone library and DGGE analysis, *Alphaproteobacteria* and *Actinobacteria* were the most abundant groups in both seasons while *Gammaproteobacteria* was commonly found in warm season (Table 3-4). SAR11-IA and plankton marine *Actinobacteria* were more abundant in September 2002 but Arctic Sea Ice ARK9990 & Roseobacter VI groups were exclusively found in March 2003.

Diversity estimates. Coverage of the clone libraries indicated that 60.2 to 75.3% of actual diversity of Chesapeake Bay bacterioplankton had been detected (Table 3-5). For calculations of species richness, the estimated coefficients of variation for six clone libraries were higher than 0.8. Therefore, non-parametric estimator ACE_1 (modified abundance-based coverage estimator) was applied to estimate the species richness (Chao and Shen 2003-2005). The estimated species richness for the six

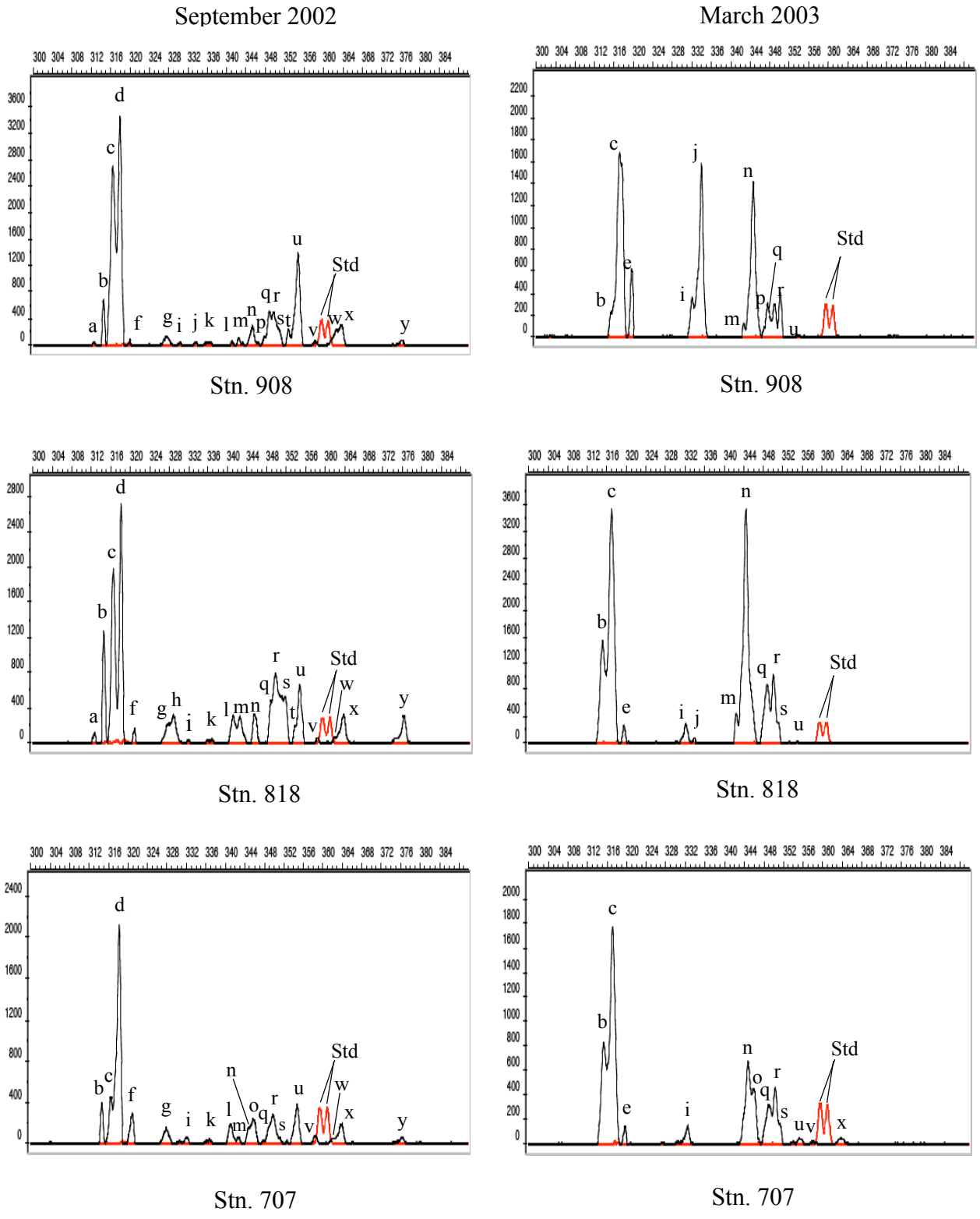


Fig. 3-3. LH-PCR analysis of bacterioplankton communities in Chesapeake Bay water samples. X-axis showed the size of the peaks in bp and y-axis was the peak intensities. Peak a-y represented different bacterial groups/subgroups. Std, peaks of size standards.

Table 3-4. Analysis of LH-PCR peaks

No.	Size (bp)	Phylogenetic groups	Taxonomic affiliation [1]	September 2002			March 2003		
				Stn.908	Stn.818	Stn.707	Stn.908	Stn.818	Stn.707
a	313	*Synechococcus*	C	0.8	1.3	-	-	-	-
b	314	Chesapeake *Roseobacter* I, III, IV & *Fibrobacteres*	α & F	5.9	10.8	6.5	3.2	12.7	15.2
c	316	Roseobacter I, II, V, VII, Rhodobacter I, Slope Strain D14, Sulfitobacter & SAR11-II, III & SAR116	α	23.2	16.8	7.5	23.7	29.2	32.5
d	317	SAR11-IA	α	29.9	23.4	35.3	-	-	-
e	318	Arctic Sea Ice Arkk 9990 & Roseobacter VI	α	-	-	-	8.8	2.1	2.6
f	319	Plastids	P	0.9	1.4	4.9	-	-	-
g	328	Rhodospirillalles 1	α	1.3	3.6	4.5	-	-	-
h	329	Rhodospirillalles 2	α	-	2.7	-	-	-	-
i	330	*Ahrensia kieliense*	α	0.2	0.4	2.2	5.1	2.5	2.7
j	331	*Pseudorhodobacter*	α	0.6	-	-	22.3	1.2	-
k	334	unknown	ND	1.1	1.0	1.0	-	-	-
l	339	Gamma	γ	0.6	2.7	3.3	-	-	-
m	340	Delta	δ	1.9	3.7	2.1	1.9	3.7	-
n	342	GKS, SAR86-II & AC I-B, C	β, γ & A	2.4	2.7	2.6	19.9	29.4	12.3
o	345	AC I-D, AC II-B & Verrucomicrobiales I	A & V	-	-	3.9	-	-	8.2
p	346	AC II-A & AC IV-A	A	1.3	-	-	4.4	-	-
q	347	Bacteroidetes & Verrucomicrobiales II	B & V	4.5	4.1	2.9	4.3	7.3	5.9
r	348	Bacteroidetes & AC III, IV-C	B & A	4.3	6.7	4.5	5.5	8.7	13.0
s	349	Bacteroidetes	B	1.9	4.5	1.8	-	2.7	3.1
t	350	Bacteroidetes & *Legionella coxi*	B & γ	1.7	1.6	-	-	-	-
u	353	Plankton marine Actinobacteria, & AC IV-B, D	A	12.0	5.6	6.3	0.7	0.5	0.9
v	358	unknown	ND	0.4	0.4	0.8	-	-	1.1
w	360	SAR86-IV	γ	2.0	1.2	1.7	-	-	-
x	361	Gamma Agg47, Arctic 96B-1 & Sva0091	γ	2.7	2.8	6.2	-	-	2.4
y	376	unknown	ND	0.4	2.6	2.0	-	-	-

[1] α, β, γ, and δ refer to the subdivisions of the *Proteobacteria*. C, *Cyanobacteria*; F, *Fibrobacteres*; A, *Actinobacteria*; B, *Flavobacterium-Bacteroidetes-Cytophaga* group; P, plastids; ND, not determined.

Table 3-5. Diversity analysis of Chesapeake Bay bacterioplankton

	Stn. 908 Sep. 2002	Stn. 804 Sep. 2002	Stn. 707 Sep. 2002	Stn. 908 Mar. 2003	Stn. 804 Mar. 2003	Stn. 707 Mar. 2003
Clone library	CB01	CB11	CB22	CB31	CB41	CB51
Phylotype richness	26	33	31	26	21	26
C (coverage, %)	71.4	60.2	64.0	69.4	75.3	68.3
Estimated richness (ACE_1)[1]	59.5 (23.9)	94.2 (38.6)	95.5 (42.8)	82.6 (40.3)	56.7 (29.1)	87.9 (45.7)
H (Shannon-Wiener' index) [1]	2.74 (0.36)	3.42 (0.19)	3.22 (0.26)	3.02 (0.17)	2.73 (0.16)	3.05 (0.16)
D (Simpson's index)[1]	0.19 (0.04)	0.07 (0.01)	0.10 (0.03)	0.09 (0.02)	0.11 (0.02)	0.08 (0.02)
DGGE						
Band richness	37	45	37	28	32	29
LH-PCR						
Peak richness	22	21	19	11	11	12

[1] Standard errors for the estimates are shown in parentheses.

clone libraries ranged from 60 to 96, respectively (Table 3-5). Northern Bay (Stn 908) in September 2002 and middle Bay (Stn 804) in March 2003 contained lower species richness than other stations in the same season. However, no significant difference was observed between the cold and warm seasons (paired t-test, $P > 0.05$). Both Shannon-Wiener's index (H) and Simpson's index (D) agreed well with this observation. DGGE band richness ranged from 28 to 45, while LH-PCR peak richness ranged from 11 to 22 (Table 3-5). In contrast to clone library results, both DGGE and LH-PCR analyses showed that bacterial species richness in September 2002 were significantly higher than March 2003 (paired t-test, $P = 0.027$ and 0.001, respectively).

Discussion

The composition of bacterioplankton in the Chesapeake Bay was distinct from previous observations of coastal ocean or freshwater systems and the variations in population structure in the Chesapeake Bay were consistent with the hypothesis that different bacterial groups are selected by temporal or spatial variation in environmental conditions.

The clone library analysis of Chesapeake Bay bacterioplankton over space and seasons added significant new information on the population structure and seasonal variations of estuarine planktonic bacteria. *Alphaproteobacteria* was one of the predominant bacterial groups in the Chesapeake Bay and was comprised of three major subgroups, SAR11, *Roseobacter*, and *Rhodobacter*. SAR11-related bacteria are known to be abundant and ubiquitous in various marine environments (Giovannoni and Rappé 2000; Morris et al. 2002; Morris et al. 2005). SAR11 related clones

appeared in the Bay only in the warm season, and this was in line with our previous observations based on the DGGE patterns in Baltimore Inner Harbor and the Chesapeake Bay (Kan et al. 2006b). This observation is also consistent with previous studies conducted in other estuarine ecosystems, with the SAR11 group in very low abundance or absent in winter and early spring (Crump et al. 1999; Crump et al. 2004; Henriques et al. 2004). In a previous study, detailed phylogenetic analysis of ITS regions separated the SAR11 clade into distinct clusters that were associated with temporal (e. g. temperature) but also geographic variations in environmental parameters (Brown and Fuhrman 2005), and thus, it appears that distribution of SAR11 group correlates with, but not necessarily a function of water temperature.

The high percentage of marine roseobacters in March 2003 clone libraries suggests that this group of bacteria thrive in colder waters. The average water temperature for March 2003 samples was 2.5 ºC. Except for Chesapeake *Roseobacter* I, most roseobacters were present in the cold season and made up more than one-third of the bacterial clones. Occurrence of marine roseobacters in winter is consistent with our multiple year investigation in the Bay based on the DGGE analysis (Chapter 5). A hallmark of cold-adaptation of microorganisms is the presence of proteins containing the cold-shock domain (Goodchild et al. 2004; Methe et al. 2005). Currently, cold-shock gene homologues are present in several *Roseobacter* genomes including *Silicibacter pomeroyi* (Moran et al. 2004), *Silicibacter* TM1040, and *Jannaschia* sp. CCS1 (www.jgi.doe.gov). For instance, *Silicibacter* TM1040 contains two cold-shock gene homologues, CSP-A1 and CSP-E (Belas, personal communication). In contrast, no cold shock gene homologues are found in "warm species" such as *Synechococcus*

spp. (www.ncbi.nlm.nih.gov/genomes/lproks.cgi). Inheriting the cold-shock genes in marine roseobacters may allow them to become more competitive in cold seasons. However, other factors may also contribute to the occurrence of roseobacters in the cold season. Roseobacters are able to turnover dimethylsulfoniopropionate released from microalgae (González et al. 2000; Zubkov 2002), and are commonly found associated with phytoplankton blooms (Alavi et al. 2001; Riemann et al. 2000). Algal blooms occur frequently in the Chesapeake Bay during winter and early spring (Glibert et al. 1995). The average concentration of Chl *a* in March 2003 was 4 times greater than that in September 2002 (Table 3-2). The concentration of Chl *a* in the northern Bay reached 41.6 ug/L in March 2003, indicating a potential phytoplankton bloom at the sampling time. Nevertheless, the possible cold-adaptation of the Roseobacter group in the Chesapeake Bay is an interesting phenomenon and more studies are needed to elucidate this distribution pattern.

Actinobacteria were another major group of bacteria in the Bay, and their clonal composition was very different between the cold and warm seasons. *Actinobacteria* are commonly found in freshwater environments (Glöckner et al. 2000; Warnecke et al. 2004). The Chesapeake Bay is greatly influenced by river runoffs and receives about half its water volume from 50 major tributaries (www.chesapeakebay.net). Freshwater *Actinobacteria* (acI-IV) occupied most of the stations in both seasons, reflecting the influence of freshwater to the Bay. Interestingly, planktonic marine *Actinobacteria* made up more than half of Actinobacterial populations in all three stations in September 2002, but did not occur in March 2003. The presence of dominant planktonic marine *Actinobacteria* in

September 2002 is likely due to marine water intrusion at the sampling time. The period between September 2001 and August of 2002 was the second driest 12 months for Chesapeake Bay watersheds in the entire 108-year record (www.atmos.umd.edu/~climate). Thus, oceanic water may have extended impact on the whole bay during the drought period. This is also reflected by the relatively high salinity (15 ppt) in the northern Bay in September 2002 (Table 3-2). Distribution of freshwater and marine *Actinobacteria* in the Bay deserves further study. Currently, no marine *Actinobacteria* have been cultivated. Understanding the occurrence of marine *Actinobacteria* and their surrounding environments may provide useful information on how to cultivate these microorganisms.

Cyanobacterial clones (mostly marine *Synechococcus*) were detected only in September libraries, suggesting that these unicellular cyanobacteria are more adapted to warm seasons. Concentrations of unicellular cyanobacteria in the Chesapeake Bay are typically low ($<10^3$ cells/ml) in winter and high (ca. 10^5 cells/ml) in summer (Wang and Chen unpublished data). Lack of cyanobacterial clone in March 2003 is consistent with low cyanobacterial counts in March 2003 (Table 3-2). Unicellular cyanobacteria are an important component of Chesapeake Bay bacterial communities during the warm season. Cyanobacterial clones accounted for 6-15% of September clone libraries, which is consistent with the fact that picocyanobacteria made up 3.6-14.1% of total bacterial counts in September 2002 samples.

Limited spatial variations on bacterial communities were seen between stations. It is noteworthy that high percentage of Arctic Sea Ice ARK9990-related clones existed in the northern Bay but not in other stations during the cold season.

Our winter cruise originally scheduled for late January 2003 was postponed to March 2003 because surface waters in the northern bay was frozen in January. Finding Arctic sea ice related bacteria in the upper Chesapeake Bay during the winter is intriguing. Our previous studies in Baltimore Inner Harbor, located in the northern Chesapeake Bay area, also identified many winter bacterial isolates that are closely related to Arctic and Antarctic clones (Kan et al. 2006b). The northern Bay also contained abundant *Rhodobacter*-related clones in the cold season. Most of the clones were affiliated with *Pseudorhodobacter* (*Agrobacterium*) *ferrugineus*, a gram-negative bacterium isolated from the sediment of Atlantic Ocean (Ruger and Höfle 1992). Where do these "cold species" live in summer? Are they present in very low abundance that typically escapes PCR detection, or sink to the bottom? The mechanism for maintaining these "cold species" and "warm species" in the Bay is still not clear and warrants future studies.

In March 2003, *Betaproteobacteria* accounted for 17 and 25% of bacterial communities in the northern and middle Bay, respectively, but only 7% of bacterial community in the southern Bay. *Betaproteobacteria* typically dominates in freshwater ecosystems (Glöckner et al. 1999; Methe et al. 1998; Zwart et al. 2002). Among the six samples, Stn 908 in March 2003 had the lowest salinity (10 ppt). The high frequency of *Betaproteobacteria* in the low salinity water suggests the influence of microbes carried by river runoffs. On the other hand, influence of marine waters can be seen in September 2002 from the spatial distribution of *Gammaproteobacteria*, that are commonly found in marine environments (Fuhrman et al. 1993; Giovannoni and Rappé 2000; Schmidt et al. 1991). The middle and southern Bay contained 15

and 12% of *Gammaproteobacteria*, respectively, while the northern bay contained only 4% of *Gammaproteobacteria*. Spatial distributions of *Beta*- and *Gammaproteobacteria*, together with salinity data suggest that the Chesapeake Bay harbors both freshwater and marine bacterial communities. In addition, certain bacterial groups might have adapted to the temperate Chesapeake estuary like some unique bacteria found in other estuaries (Crump et al. 2004; Hewson and Fuhrman 2004; Chapter 4).

Applying clone library, DGGE and LH-PCR to the same environmental samples allows us to compare these three cultivation-independent methods. Fingerprinting approaches such as DGGE and LH-PCR provide a quick "snapshot" of dominant bacterial groups and are particularly useful for the ecological survey where a large number of samples need to be analyzed. Given that three different primer sets were used for clone library analysis, DGGE and LH-PCR, presence of major bacterial groups appeared to be consistent among all three methods. We compared the 16S rRNA gene sequences retrieved from the clone libraries and the DGGE bands, approximately 70% of sequences from the DGGE bands were clustered with clone sequences from libraries. The major discrepancy between DGGE and clone library analyses were *Planctomycetes*. Four DGGE bands were identified as *Planctomycetes* (Chapter 5), but these sequences were absent in clone libraries, suggesting a possible underestimation of this group by the clone library analysis. Likely it is caused by mismatches with the 27F primer used in PCR.

The current study provides a basis for understanding phylogenetic diversity and seasonal variation of bacterioplankton in the Chesapeake Bay. Rarefaction

analyses of clone libraries suggest that high genetic microdiversity exists in Chesapeake bacterioplankton. It has been predicted that the number of different microbial taxa in water samples range from a few hundred to one million (Curtis et al. 2002; Acinas et al. 2004; Schloss and Handelsman 2005). Recently, by use of 454 tag sequencing strategy, bacterial communities from deep water masses were estimated that contained one to two orders of magnitude more genetic diversity than previous estimation (Sogin et al. 2006). A relatively small number of major populations dominate at certain spatial or temporal dimensions, but more low-abundance populations, defined as "rare biosphere", account for most of the diversity in the community (Sogin et al. 2006). All the PCR-based molecular approaches applied in this study focus on those major components of microbial communities, possibly due to the potential bias introduced by PCR, cloning procedures etc. Consequently, the extraordinary diverse with low-abundance members of microbial communities are likely undersampled. As a fact, we have just begun to understand the diversity and ecology of Chesapeake Bay bacterial communities.

Chapter 4: Fast screening and phylogenetic analysis of estuarine bacterial rRNA operon libraries

Abstract

A rapid phylogenetic screening approach, ITS (internal transcribed spacer)-LH (length heterogeneity)-PCR was tested and applied to six rRNA gene operon clone libraries constructed from the Chesapeake Bay. 576 clones from estuarine bacterioplankton over two seasons were classified based on the natural length variations of the ITS and the presence and location of tRNA-alanine coding genes within the ITS. Representative clones with varying ITS-LH-PCR sizes were further identified by sequencing and phylogenetic analysis, and a generally good congruence was found between ITS-LH-PCR sizes and phylogenetic relationships. Few overlaps of the ITS-LH-PCR fragment sizes occurred among members of different bacterial groups allowing putative identification of clones without sequencing. Phylogenetic reconstruction confirmed that Chesapeake Bay bacterioplankton included typical marine and freshwater organisms, although novel groups were present that had not been previously retrieved from either system. SAR11, *Roseobacter*, and SAR86 clades and clades of *Actinobacteria* represented mostly by sequences retrieved from the Chesapeake Bay, suggesting that the existence of clades adapted to the Bay and perhaps, other large temperate estuaries.

Introduction

Since the first analysis of small subunit ribosomal RNA (16S rRNA) genes from marine bacterioplankton community (Giovannoni et al. 1990), clone libraries of PCR-amplified 16S rRNA have become a common approach to study microbial population structure in a wide range of environments. Subsequent studies provide mounting evidence that the complexity of natural aquatic microbial assemblages are far beyond that of culture-based studies (DeLong et al. 1993; Fuhrman et al. 1993; Schmidt et al. 1991; Sekiguchi et al. 2002). Typically, 16S rRNA genes in clone libraries are sequenced to determine phylogenetic origins of the recovered ribotypes. One of the major bottlenecks in the analysis of clone libraries has been the screening for unique clones to avoid unnecessary sequencing by restriction fragment length polymorphism (RFLP). This type of analysis is labor intensive and time consuming and organism identification from RFLP patterns is in general difficult. Recently, a novel high-throughput PCR analysis, internal transcribed spacer length heterogeneity PCR (ITS-LH-PCR) has been developed to allow screening of large insert (BAC and fosmid) libraries (Suzuki et al. 2004). Based on the length of entire ITS region, and the location of the tRNA-alanine in the spacer, using ITS-LH-PCR it is possible to identify environmental clones down to the sub-clade level (Suzuki et al. 2004). However, this new approach has so far been limited to marine bacterioplankton ITS regions. In order to create a database of ITS lengths and tRNA positions that can be used for future screening of environmental genomic libraries, clones from different ecosystems should be analyzed.

In previous phylogenetic analysis, both habitat-specific and universally distributed marine or freshwater bacteria clades have been identified. Rivers and lakes have unique planktonic bacterial community that are distinct from adjacent soil and sediments, as well as marine bacterial flora (Warnecke et al. 2004; Zwart et al. 2002). Estuaries are regions of freshwater and marine influence and clearly, bacteria can be originated from different resources depending on turnover times (Crump et al. 2005; Henriques et al. 2006). Therefore, bacteria found in estuaries do not necessary grow and thrive in that environment. Few studies have shown that estuarine bacteria contain mixed populations of both freshwater and marine origins, as well as endemic populations to estuarine ecosystems (Crump et al. 1999; Selje and Simon 2003; Henriques et al. 2004; Hewson and Fuhrman 2004), although most of these were conducted in relatively small systems or estuaries with short residence time. Limited studies have been conducted on the composition of bacterioplankton in the Chesapeake Bay (Bidle and Fletcher 1995; Bouvier and del Giorgio 2002; Heidelberg et al. 2002). However, most of these previous studies characterized the bacterioplankton in the Chesapeake Bay either at broad level (phyla and clades) or at narrow resolution focusing on individual specific genera/species. With no previous detailed studies on rRNA gene sequencing analysis, currently little is known about Chesapeake Bay bacterioplankton diversity.

In order to better understand the bacterioplankton diversity of the Chesapeake Bay, a large estuary with long residence time at a high phylogenetic resolution, we constructed six rRNA operon clone libraries from northern, middle and southern Chesapeake Bay at two seasons. A total of 576 clones were screened by ITS-LH-PCR

and the combined fragment sizes were compared to previously sized fragments (Suzuki et al. 2004). Clones with representative and unique size combination were further sequenced and phylogenetic relationships of planktonic bacteria in the Chesapeake Bay was defined. Details on the dynamics of population structure are described elsewhere (Chapter 3).

Materials and methods

DNA sampling and library construction. Details of DNA sampling and construction of rRNA operon clone libraries were described elsewhere (Chapter 3).

Screening for clones with ITS-LH-PCR. For two libraries, CB1 and CB2, clones were grown overnight and the cells were pelleted down using U-bottom microtiter plates. Plasmids were isolated using a standard alkaline lysis protocol using a Hydra 96 microfluidic dispenser (Robbins Scientific). Plasmid DNA were precipitated using isopropanol, washed by 70% ethanol and resuspended by TE buffer. For the remaining libraries, 50 µl of cells were pelleted in 96 well PCR plates, resuspended in 20 µl of platinum Taq PCR buffer and lysed at 94ºC for 5 min.

ITS-LH-PCR was performed as previously described with some modifications (Suzuki et al. 2004). Briefly, two different primer sets were used in separate reactions for 96 clones per library: 6-FAM labeled 16S-1406F and 23S-66R to amplify the ITS region, and HEX labeled 16S-1406F and tRNA alaR for the tRNA fragment (Lane 1991; Suzuki et al. 2004) (Table 4-1). Due to mismatches of the original 23S-66R primer to *Actinobacteria* and *Bacteroidetes* groups, we modified the 66R primers to target these groups and the new primers represent a mixture of the different primers (Table 4-1). Labeled fragments were discriminated using an Applied Biosystems

3100 Genetic Analyzer. Sizes of ITS and tRNA fragments were determined by Genescan software (Applied Biosystems) using the GS2500 size standard (Applied Biosystems). *E. coli* ITS and tRNA fragments were used as positive controls. The phylogenetic identity of clones represented by different fragment pairs as well as those with no amplified fragments using either primer pair were determined by sequencing and comparison to fragment sizes previously measured (Suzuki et al. 2004; Table 4-1).

Table 4-1. Primers used in screening and sequencing of Chesapeake Bay clones

Primer	Sequence (5'-- 3')	Target site	Reference
ITS-LH-PCR			
16S-1406F-FAM	*6-FAM* -TGYACACACCGCCCGT	16S 1391-1406	Lane 1991; Suzuki et al., 2004
23S-66R	CACGTCTTTCATCGSCT	23S 50-66	Suzuki et al., 2004
23S-66R-Actino	TACGTCCTTCT/GTCGGTT	Actino 23S 50-66	this study
23S-66R-CFB	CACGTCCTTCTTCGCCA	CFB 23S 50-66	this study
16S-1406F-HEX	*HEX* -TGYACACACCGCCCGT	16S 1391-1406	Lane 1991; Suzuki et al., 2004
ITS-tRNAalaR	CTGCTTGCAAAGCAGGCGCTC	ITS-tRNA alanine	Suzuki et al., 2004
Sequencing			
16S-27F	AGAGTTTGATCMTGGCTCAG	16S 8-27	Lane 1991
16S-1074F	ATGGCTGTCGTCAGCTCGTG	16S 1055-1074	Suzuki et al., 2004
16S-1100R	AGGGTTGCGCTCGTTG	16S 1100-1115	Suzuki et al., 2004
16S-1541R	AAGGAGGTGATCCRGCCGCA	16S 1522-1541	Suzuki et al., 2000
16S-1406F	TGYACACACCGCCCGT	16S 1391-1406	Lane 1991
ITS-tRNAalaR	CTGCTTGCAAAGCAGGCGCTC	ITS-tRNA alanine	Suzuki et al., 2004
ITS-tRNAalaF	GAGCGCCTGCTTTGCAAGCAG	ITS-tRNA alanine	Suzuki et al., 2004
23S-139R	GCTGGGTTKTCTCATTCRG	23S 121-139	this study

Sequencing and phylogenetic analysis. For the vast majority of representative fragment pairs, 16S rRNA genes of the clones were fully sequenced by dideoxynucleotide termination Big Dye Chemistry v3.0 with primers 16S-27F,

1074F, 1100R, and 1541R (Suzuki et al. 2000) (Table 4-1). Plasmids were purified using the Montage (Millipore), SprintPrep (Agencourt), and Fasplasmid (Eppendorf) kits. Chimeric sequences were tested by CHIMERA DETECTION program of RDP (Maidak et al. 1997) and removed from further analysis. Non-chimerical sequences were compiled in ARB and aligned to sequences in the ARB database containing approximately 28,000 total sequences (Ludwig et al. 2004). Sequence alignments were constructed using automated aligner in the ARB_EDIT software and then manually inspected and corrected based on the conserved secondary structure of 16S rRNA genes. All sequences were added to a ca. 28,000-sequence tree in the ARB distribution (ssujun02.arb) using a filter that excluded positions where gaps outnumbered characters.

Near complete gene sequences (>1400 bp) were used to construct the phylogenetic trees. Positions with ambiguous characters or with gaps more frequent than bases were excluded from the analysis. Remaining positions were used for phylogenetic reconstruction.

The multiple sequence alignment were exported from ARB for bootstrap analysis. Bootstrapping of parsimony and distance (Jukes-Cantor Neighbor-Joining) were calculated by PAUP 4.0b10 for Macintosh program (Swofford 1998). Heuristic searches were based on 100 replicates and starting trees were obtained via stepwise addition. Sequences were added randomly with three random-addition replicates per bootstrap replicate. For bootstrapping analysis of maximum likelihood, nucleotide substitution models were determined by Modeltest Version 3.7 (Posada and Crandell 1998). The bootstrap analyses were carried out by PHYML (Guindon and Gascuel

2003). Briefly, the nonparametric bootstrap was under one category of substitution rate and fixed proportion of invariable sites (P-invar = 0.01). The input tree format was BIONJ and the tree topology was optimized. The trees were imported in ARB and short sequences obtained from clone libraries were added by ARB_PARSIMONY.

Results and discussion

ITS-LH-PCR.

There was generally good congruence between phylogenetic relationships and paired ITS-LH-PCR fragment sizes of estuarine bacterial groups (Table 4-2). High variability in the length of ITS region and tRNA content between bacterioplankton clades was observed. Size of ITS fragments varied significantly among the major bacterial groups and ranged from 347 bp for plankton marine *Actinobacteria* to 1275 bp for *Ahrensia* sp. DFL-42. Sizes of tRNA fragments ranged from 260 bp for relatives of uncultured *Bacteroidetes* group ML81771-J6 to 655 bp for *Desufotalea arctica* (Table 4-2). Few overlaps occurred among members of different clades, which allows putative enumeration and identification of clones by combined information of ITS and tRNA lengths. Based on ITS-LH-PCR, we could clearly distinguish clades and subclades of many groups (i.e. roseobacters and *Actinobacteria*) (Table 4-2). Genes coding tRNA-alanine were absent in some groups including SAR86, SAR116, and *Actinobacteria*. In some cases, tRNA-alanine fragments were observed but no corresponding ITS fragments were detected. Compared with available sequences in GenBank, we noticed that mismatches of 23S-66R primers to *Actinobacteria* and *Bacteroidetes* (data not shown). Modified version

of this primer (23S-66R-Actino and 23S-66R-CFB, Table 4-1) improved the detection of ITS fragments of clones associated with *Bacteroidetes* and *Actinobacteria* and pointed to the necessity of using these primers in future analyses.

We compared the ITS-LH-PCR fragment sizes of estuarine bacterioplankton to sequences in GenBank and sizes measured in a previous study (Suzuki et al. 2004) and in general, ITS-LH-PCR fragment sizes agreed well with previously assigned phylogenetic identities (data not shown). However, Chesapeake Bay clones contained a broader spectrum of ITS-LH-PCR fragment sizes, indicating a higher diversity of bacterial communities in estuaries. Extending the ITS-LH-PCR size database to estuarine bacterioplankton improved the putative identification of environmental clones from aquatic environments without sequencing and will facilitate future rRNA as well as genomic clone library analysis.

The ITS region displays significant heterogeneity in both length and nucleotide sequence. Both variable features provide higher taxonomic resolution than the 16S rRNA gene and have been extensively applied to distinguish strains or closely related species (Aubel et al. 1997; Chen et al. 2006; Jensen et al. 1993; Rocap et al. 2002). For instance, ecotypes of *Prochlorococcus* and *Synechococcus* were identified based on the length, G+C content, and sequence of ITS (Rocap et al. 2002). The Chesapeake Bay *Synechococcus* clones could be distinguished by ITS-LH-PCR fragments sizes (Table 4-2). Phylogeny of ITS sequences of these *Synechococcus*

Table 4-2. Measured size of the 1406F-66R and 1406F-tRNAR fragments for Chesapeake Bay rRNA operon libraries

Phylotype	1406F-66R size (bp)	1406F-tRNAala size (bp)
Phylum *Proteobacteria*		
Class *Alphaproteobacteri*		
Order *Rhizobiale*		
Defluvibacter lusatiae (93.7 - 894 bp)	1002	no
Ahrensi sp. DFL-42 (97.1 - 1301	1275	549
Order *Rhodobacterales*		
Pseudorhodobacter ferrugineum	974, 977-8, no	427, -29, -
Sulfitobacter mediterraneus	1172	581
Slope Strain D14	1142	446
	1177, -81	545-
	1163-	447-
	1200, -11	581
Arctic Pack Ice Strain Ark9990	923-	361-
Chesapeake Roseobacter I	889	406
Chesapeake Roseobacter I	1002-	353-
Chesapeake Roseobacter I	902-	445-
Chesapeake Roseobacter I	954-	441-
Chesapeake Roseobacter I	no	582-
Chesapeake Roseobacter	1085	519
	1112, -17, -	518, -46, -
Chesapeake Roseobacter III CB41G10	945, -	443-
Chesapeake Roseobacter IV	1110-3, no	367, 470-1
Chesapeake Roseobacter V CB31B07	1137, no	361, no
Chesapeake Roseobacter VI	1141, -6	363
Chesapeake Roseobacter VII	965	365
Chesapeake Rhodobacter	1003, 1026-7	433
Novel Rhodobacter DC5-50-	860	340
SAR11		
SAR11 - IA - *Pelagibacter ubique*	616, -	365, 367-9
	593-	365-6, 369-70, -
SAR11 -	600, -18	370
SAR11 - IIA	644, 551, -93	361, 366-7
SAR11 -	593-	375-
SAR11 - IIIA	663-	367, 415
SAR11 -	618,	386, -
Order *Sphingomonadales*		
marine alpha proteobacterium JP63.1	967	468
Order *Rhodospirillales*		
Sar116 - III	530	no
Sar116 -	669, no	no
Novel Rhodospirillalles	605, no	372, 484, no
Class *Betaproteobacteri*		
Beta OM43	743	363
GKS	1060, -76	447, -
	1081	491
	1036, -62, -	452-
	1101, -12, -	501-
	1105, -18	468, -
	no	447, -67, -69,
GKS group Novel	936, -	374-
Hydrogenophilu sp.	760	370
OM15	695,	338, -
Polaromonas	917	456
Polynucleobacter necessarius	720, -	353, 357-8
Class *Gammaproteobacteri*		
Acinetobacte	916	394
Arctic 96B-	807, 947	289, 315
	857, -	327
CHAB-III-7	669	345
Gamma AGG47	666, -73, -76,	347-
Gamma Novel CB22H04	1006	379
Gamma	962	359
KTC111	954	477
marine Gamma	no	no
novel NOR5-like	960	486
Legionella Nove	548, -56, -	280-1, no
OM6	844	280
Pseudomonas syringae	723	397
Psychrobacter sp.	814	353
Psychromonas sp.	959	428
Sar86 - II	413	no
Sar86 - IV	375-	no
SDF1-	954	no
Unidentified Gamma	432	no

Phylotype	1406F-66R size	1406F-tRNAala size
Class *Deltaproteobacteria*		
Desulfotalea sp.	872	655
Marine Delta Prot	622, -4, -6,	376,
Syntrophobacteraceae CB11F12	721	no
Phylum *Bacteroidetes*		
Class *Flavobacteria*		
ATAM173 A3	774	425
Flavobacteriaceae UC1	671,	368,
Flavobacterium gelidiacus	632	426
Psychroserpens sp.	753	450
Class *Sphingobacteri*		
FBC ML1218M-	973,	470,
FBC	708, -	260, -
Haliscomenobacte sp.	555	no
Pedobacte sp.	786	456
Rhodovirga sp.	1117	no
Sphingobacteri sp.	no	543
Unclassified		
Bacteroidetes	595	386
Bacteroidetes OM	600	382
	612	389
	762	449
FBC Clone	no	460
FBC Novel TAFB64	no	481
Flexibacter aggregans	no	466
Cytophaga novel	698	400
Phylum *Cyanobacteria*		
Synechococcus	822	448
	1056	485-6
	1123	487
	943, -62, -	458-9
	954, -72, -74, -76	464-5
	964, 1006-7	457, -60
	993, 995-6	460-1
Diatom	no	no, 320-36
Prasinophyte Plastid	467	324
Phylum *Actinobacteria*		
Order *Actinomycetales*		
Family *Microbacteriaceae*		
Freshwater AC II-	559, -61, 590-1, no	no
Freshwater AC II-B	605-	no
Freshwater AC	616-	no
Freshwater AC III-	622-	no
Unclassified		
Freshwater AC I-B	506	no
Freshwater AC I-C	409-10, 413	no
Freshwater AC I-D (Novel)	554, -57, -65, no	no
Unclassified *Actinobacteria*		
Freshwater AC IV Novel CB11A12	486	no
Freshwater AC IV-	511,	no
Freshwater AC IV-B	540	no
Freshwater AC IV-C	497,	no
Freshwater AC IV-D (Novel)	489-	no
Actinobacterium Novel CB31D05	544	no
Plankton Marine Actinobacterium	347, 350-1, no	no
Plankton Marine Actinobacterium	360, -	no
Sediment Marine Actinobacterium	461	no
Phylum *Fibrobacteres*		
Fibrobactere	no	no
Phylum *Verrucomicrobi*		
Verrucomicrobiales Group I	796, 969, -80, 1138	397, 435, 566, -75
	969	566
	980	575
	1138	435

clones confirmed the divergence between marine cluster A and B *Synechococcus* clades (Chen et al. 2006). In addition, combined sizes of ITS and tRNA fragments allowed easy identification of different subclusters of freshwater *Actinobacteria*, roseobacters, and different phylotypes of *Gammaproteobacteria* and *Bacteroidetes*. Finally, variable paired sizes of ITS-LH-PCR for closely related phylotypes based on 16S rRNA gene sequence were commonly observed (i.e. *Roseobacter* clade, Slope strain DI4, *Betaproteobacteria* GKS group). Multiple copies of rRNA operons present in a bacterial genome are likely the cause. Nevertheless, ITS-LH-PCR is a fast approach to prescreen multiple clone libraries with high resolution and reliable identifications.

Phylogeny of estuarine bacterioplankton.

Alphaproteobacteria. *Alphaproteobacteria* was a major component in Chesapeake Bay bacterioplankton and members of the SAR11, *Roseobacter* and *Rhodobacter* clades were the most abundant groups. Clones belonging to the SAR11 clade only appeared in the three clone libraries sampled from September 2002 (Chapter 3), indicating the seasonal effects on the microbial population structures.

Four main groups of the SAR11 clade, SAR11-I, II, III and IV were obtained in the reconstructed phylogeny (Fig. 4-1). The phylogeny of SAR11-I and II were in good agreement with previous classification schemes, supported by 16S rRNA gene, ITS, and 23S rRNA gene phylogeny (Suzuki et al. 2001). Group I can be divided into three subgroups. Twelve clones, mostly retrieved from middle and southern Bay, were clustered with *Pelagibacter ubique* in SAR11-IA (Suzuki et al. 2001). SAR193 and one clone from Monterey Bay (MB12A07) formed a subgroup SAR11-IB, while

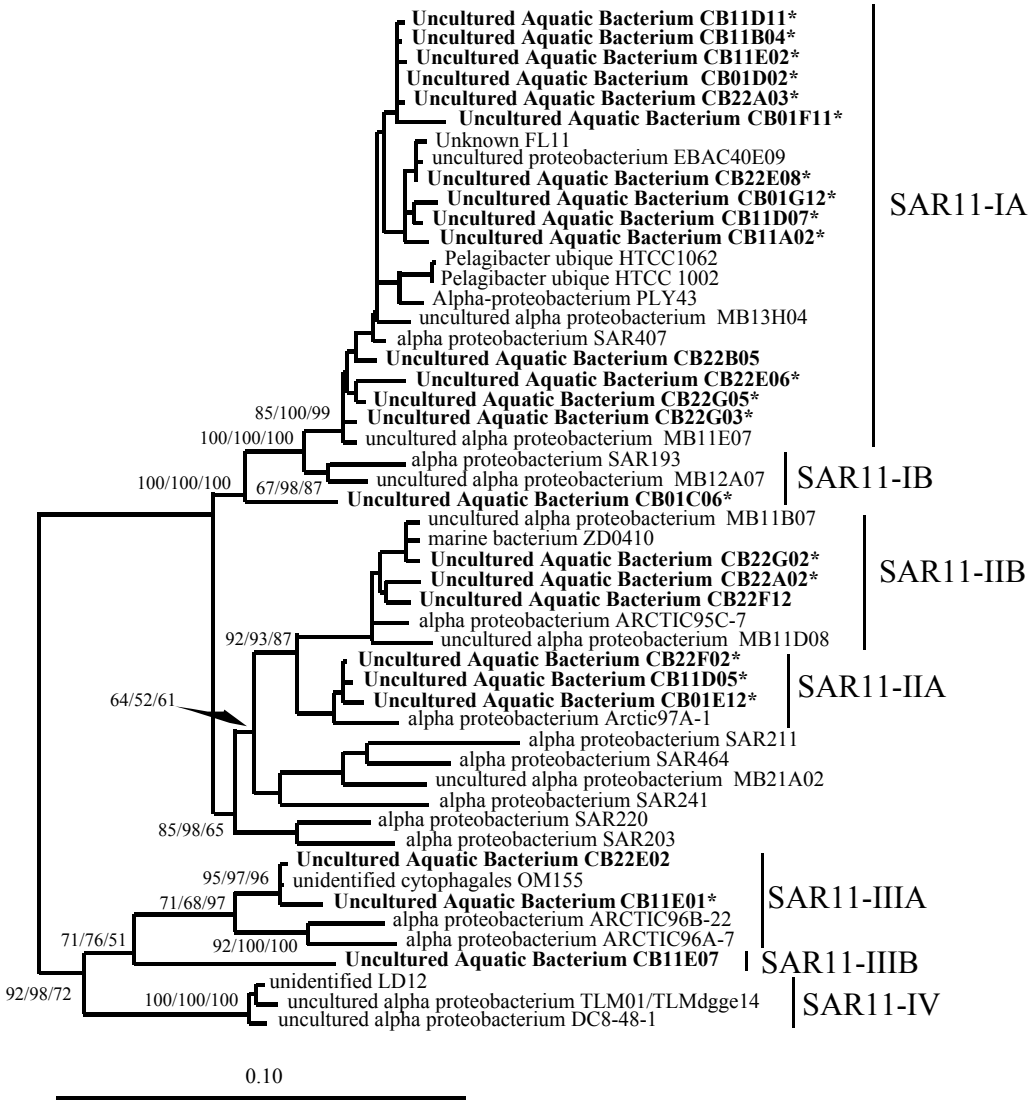


Fig. 4-1. Phylogenetic reconstruction (DNA distance - NJ) of the SAR11 clade. Clone sequences obtained in this study are shown in bold. Bootstrap values at nodes were calculated using maximum likelihood (before the first slash), distance (Jukes-Cantor Neighbor-Joining) (before the second slash), and parsimony (after the second slash). Δ: bootstrap value lower than 50 or the branch collapsed. Bootstrap values not relevant to the interpretation of suggested subgroups were omitted. *: Short sequences added to the original tree by ARB_PARSIMONY. Scale bar indicates 10% estimated sequence divergence.

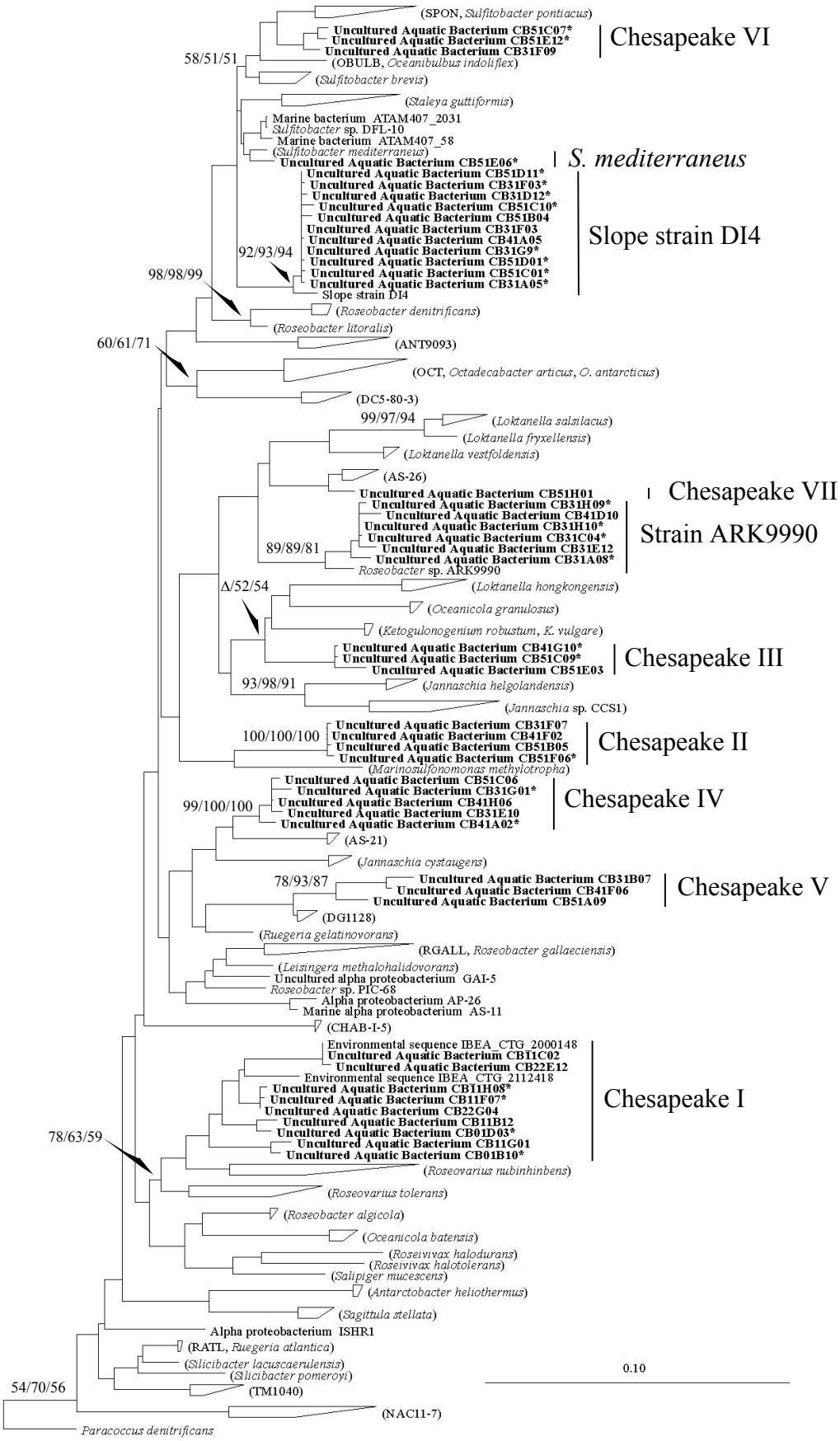

(SPON, Sulfitobacter pontiacus)
Uncultured Aquatic Bacterium CB51C07*
Uncultured Aquatic Bacterium CB51E12*
Uncultured Aquatic Bacterium CB31F09
Chesapeake VI
58/51/51
(OBULB, Oceanibulbus indoliflex)
(Sulfitobacter brevis)
(Staleya guttiformis)
Marine bacterium ATAM407_2031
Sulfitobacter sp. DFL-10
Marine bacterium ATAM407_58
(Sulfitobacter mediterraneus)
Uncultured Aquatic Bacterium CB51E06*
S. mediterraneus
Uncultured Aquatic Bacterium CB51D11*
Uncultured Aquatic Bacterium CB31F03*
Uncultured Aquatic Bacterium CB31D12*
Uncultured Aquatic Bacterium CB51C10*
Uncultured Aquatic Bacterium CB51B04
Uncultured Aquatic Bacterium CB31F03
Uncultured Aquatic Bacterium CB41A05
Uncultured Aquatic Bacterium CB31G9*
Uncultured Aquatic Bacterium CB51D01*
Uncultured Aquatic Bacterium CB51C01*
Uncultured Aquatic Bacterium CB31A05*
Slope strain DI4
92/93/94
98/98/99
(Roseobacter denitrificans)
(Roseobacter litoralis)
60/61/71
(ANT9093)
(OCT, Octadecabacter arcticus, O. antarcticus)
(DC5-80-3)
99/97/94
(Loktanella salsilacus)
(Loktanella fryxellensis)
(Loktanella vestfoldensis)
(AS-26)
Uncultured Aquatic Bacterium CB51H01
Chesapeake VII
Uncultured Aquatic Bacterium CB31H09*
Uncultured Aquatic Bacterium CB41D10
Uncultured Aquatic Bacterium CB31H10*
Uncultured Aquatic Bacterium CB31C04*
Uncultured Aquatic Bacterium CB31E12
Uncultured Aquatic Bacterium CB31A08*
Strain ARK9990
89/89/81
Roseobacter sp. ARK9990
Δ/52/54
(Loktanella hongkongensis)
(Oceanicola granulosus)
(Ketogulonogenium robustum, K. vulgare)
Uncultured Aquatic Bacterium CB41G10*
Uncultured Aquatic Bacterium CB51C09*
Uncultured Aquatic Bacterium CB51E03
Chesapeake III
93/98/91
(Jannaschia helgolandensis)
(Jannaschia sp. CCS1)
Uncultured Aquatic Bacterium CB31F07
Uncultured Aquatic Bacterium CB41F02
Uncultured Aquatic Bacterium CB51B05
Uncultured Aquatic Bacterium CB51F06*
100/100/100
Chesapeake II
(Marinosulfonomonas methylotropha)
Uncultured Aquatic Bacterium CB51C06
Uncultured Aquatic Bacterium CB31G01*
Uncultured Aquatic Bacterium CB41H06
Uncultured Aquatic Bacterium CB31E10
Uncultured Aquatic Bacterium CB41A02*
Chesapeake IV
99/100/100
(AS-21)
(Jannaschia cystaugens)
Uncultured Aquatic Bacterium CB31B07
Uncultured Aquatic Bacterium CB41F06
Uncultured Aquatic Bacterium CB51A09
Chesapeake V
78/93/87
(DG1128)
(Ruegeria gelatinovorans)
(RGALL, Roseobacter gallaeciensis)
(Leisingera methalohalidovorans)
Uncultured alpha proteobacterium GAI-5
Roseobacter sp. PIC-68
Alpha proteobacterium AP-26
Marine alpha proteobacterium AS-11
(CHAB-I-5)
Environmental sequence IBEA_CTG_2000148
Uncultured Aquatic Bacterium CB11C02
Uncultured Aquatic Bacterium CB22E12
Environmental sequence IBEA_CTG_2112418
Uncultured Aquatic Bacterium CB11H08*
Uncultured Aquatic Bacterium CB11F07*
Uncultured Aquatic Bacterium CB22G04
Uncultured Aquatic Bacterium CB11B12
Uncultured Aquatic Bacterium CB01D03*
Uncultured Aquatic Bacterium CB11G01
Uncultured Aquatic Bacterium CB01B10*
Chesapeake I
78/63/59
(Roseovarius nubinhinbens)
(Roseovarius tolerans)
(Roseobacter algicola)
(Oceanicola batensis)
(Roseivivax halodurans)
(Roseivivax halotolerans)
(Salipiger mucescens)
(Antarctobacter heliothermus)
(Sagittula stellata)
Alpha proteobacterium ISHR1
(RATL, Ruegeria atlantica)
(Silicibacter lacuscaerulensis)
(Silicibacter pomeroyi)
(TM1040)
54/70/56
(NAC11-7)
Paracoccus denitrificans
0.10

Fig. 4-2. Phylogenetic reconstruction (DNA distance - NJ) of the *Roseobacter* clade. Full or partial of 16S rRNA gene sequences were aligned to 41 known major lineages (names in parentheses) reviewed by Buchan et al. (2005).

clone CB01C06 may represent an additional subgroup. SAR11-IIA and SAR11-IIB were two distinct subgroups, which included six clones obtained from this study, two from Monterey Bay (Suzuki et al. 2001), two from the Arctic Ocean (Bano and Hollibaugh 2002), and one from the North Sea after an algal bloom (Zubkov et al. 2002). Two newly proposed groups, SAR11-III and SAR11-IV, were separated from SAR11-I and SAR11-II. SAR11-III appeared to have costal distribution and was composed of two Arctic clones (Bano and Hollibaugh 2002), three clones from the Chesapeake Bay and one clone from the continental shelf off Cape Hatteras, North Carolina (Rappé et al. 1997). Unique Chesapeake Bay clone CB11E07 was the only member in a proposed novel subgroup SAR11-IIIB. SAR11-IV was a monophyletic group supported by high bootstrapped values, containing freshwater SAR11 clones. With more clones added into SAR11 clade, novel groups or subgroups emerged. All the SAR11 sub-groups contain no cultivated representatives except for the subgroup SAR11-IA that contains *Pelagibacter ubique* (Rappé et al. 2002). These environmental clones were retrieved from diverse habitats and no clear separation exists between marine and estuarine SAR11.

The Chesapeake Bay contains many novel and unique Roseobacter lineages (Fig. 4-2), named Chesapeake *Roseobacter* I-VII. *Roseobacter* I was the cluster containing the sequences from the September and they were closely related to two

sequences from Sargasso Sea metagenomic database (IBEA_CTG_2000148 and 2112418) (Venter et al. 2004). Chesapeake *Roseobacter* II, III, IV, V, and VI were uniquely represented by Chesapeake Bay clones in the March samples, suggesting they may represent endemic groups to the Chesapeake Bay. Chesapeake *Roseobacter* VII only contained one clone from South Bay and it was relatively closely related to AS-26 clade. In addition, many clones were associated with strains isolated from Arctic (Strain ARK9990) (Brinkmeyer et al. 2003) or North Atlantic continental slope (Slope strain DI4) (Teske et al. 2000), indicating a somewhat ubiquitous distribution of these species.

Roseobacter represents one of the nine major clades of marine bacterioplankton (Giovannoni and Rappé 2000; González and Moran 1997). Typically roseobacters comprise up to 15-20% of ocean and estuarine bacterial communities (Giovannoni and Rappé 2000; González and Moran 1997; Selje et al. 2004) and hold an overwhelming diversity within the group (Buchan et al. 2005). In cold season, the Chesapeake Bay contained 35.3% to 39.1% of clones associated with roseobacters, which showed a high-diversity of phylotypes (Chapter 3). To some extent, the phylogenetic placement of roseobacters is difficult to resolve. Although our reconstruction was consistent with 41 previously defined lineages (Buchan et al. 2005), instability of tree branching patterns between these lineages were observed by different phylogenetic reconstructions. Thus, the phylogenetic relationships among these lineages were not clear. The seven unique *Roseobacter* lineages remained stable in different types of tree reconstructions, indicating that they are clearly distinct from other *Roseobacter* lineages. Interestingly, no sequences from other estuaries (Crump

et al. 1999; Sekiguchi et al. 2002; Selje and Simon 2003) were clustered together with Chesapeake *Roseobacter* lineages, suggesting they may represent habitat-specific populations adapted to the Bay or other large estuaries with long residence time.

The phylogenetic placement of clones in the SAR116 clade was consistent with previous description (Suzuki et al. 2001). One clone (CB22G09) from this study and one clone from Sargasso Sea metagenomic database (IBEA_CTG_1958364) (Venter et al., 2004) were clustered with group III. In addition, two clones (CB22C04 and CB22D08), combined with 27 symbiotic clones from marine sponge *Halichondria okadai*, formed a new group SAR116-IV (data not shown). Most of *Rhodobacter* clones (68%) were associated with *Pseudorhodobacter ferrugineum*, an organism isolated from northeastern Atlantic Ocean bottom sediments (Ruger and Höfle 1992).

Gammaproteobacteria and SAR86. High genetic diversity of *Gammaproteobacteria* was observed in the Chesapeake Bay, and SAR86 was the most abundant group present during the sampling times particularly in September 2002 (Chapter 3). Four distinct phylogenetic groups (SAR86-I, II, III and IV) were observed within the SAR86 clade (Fig. 4-3). Within SAR86-IV, one fully sequenced clone was tightly clustered with another five partially sequenced clones exclusively found in the Bay. They formed a unique and monophyletic group distinct from three previously defined groups of SAR86 (SAR86-I, II and III) (Suzuki et al. 2001). Because only one fully sequenced clone (CB11A08) was included in the preliminary tree construction as well as bootstrap analysis, no bootstrapping support for the SAR86-IV clade was obtained. However, consistent phylogenetic placement among

different phylogenetic construction methods suggest the existence of SAR86-IV as a unique novel group.

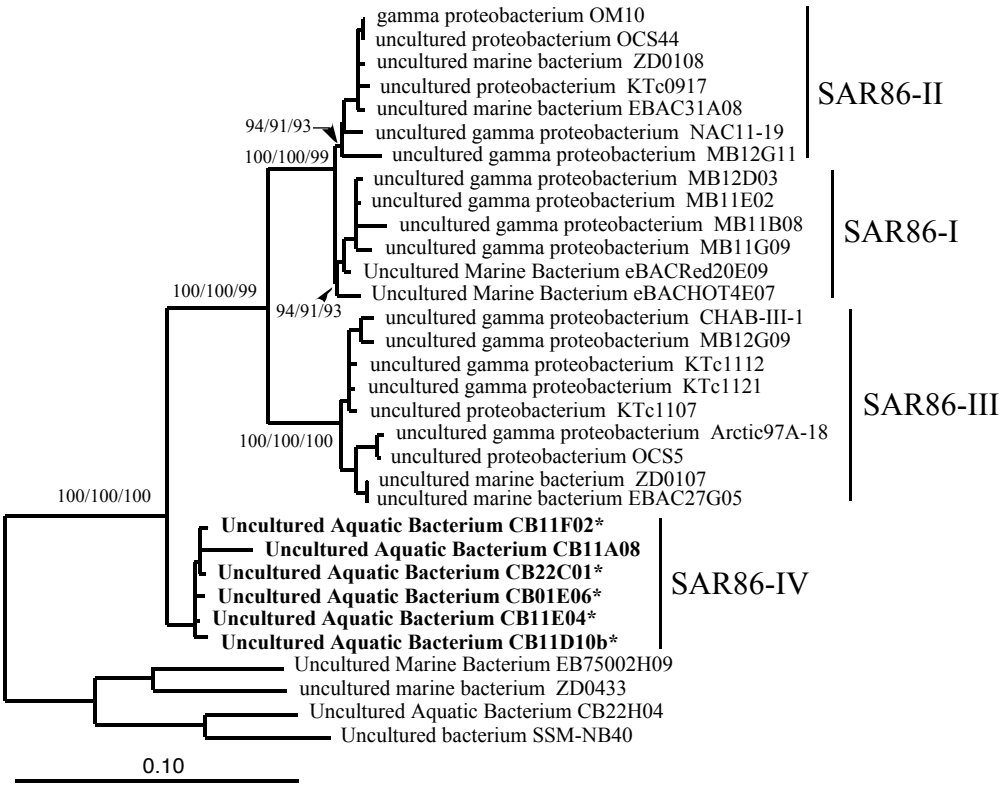


Fig. 4-3. Phylogenetic reconstruction (DNA distance - NJ) of the SAR86 clade.

Actinobacteria. Clones affiliated with freshwater *Actinobacteria* observed in all six clone libraries were more abundant in March 2003 than September 2002 (Chapter 3, Table 4-3). The preliminary placement of *Actinobacteria* clones by ARB-PARSIMONY showed that many clones belonged to four previously defined distinct phylogenetic clusters of freshwater *Actinobacteria* (acI, II, III and IV, Fig. 4-6; Warnecke et al. 2004).

Sequences affiliated with freshwater cluster acI were positioned within subcluster acI-B and C (Warnecke et al. 2004), while eight clones were clustered with unidentified bacterium rJ7, rJ14, and Actinomycetales bacterium GP-5 forming a novel cluster (acI-D, Fig. 4-4). Bootstraps strongly support that proposed acI-D is distinct from other subclusters. In agreement with Warnecke et al. 2004, it appears that acI are autochthonous components of aquatic microbial communities since these organisms are almost exclusively found in freshwater and estuaries.

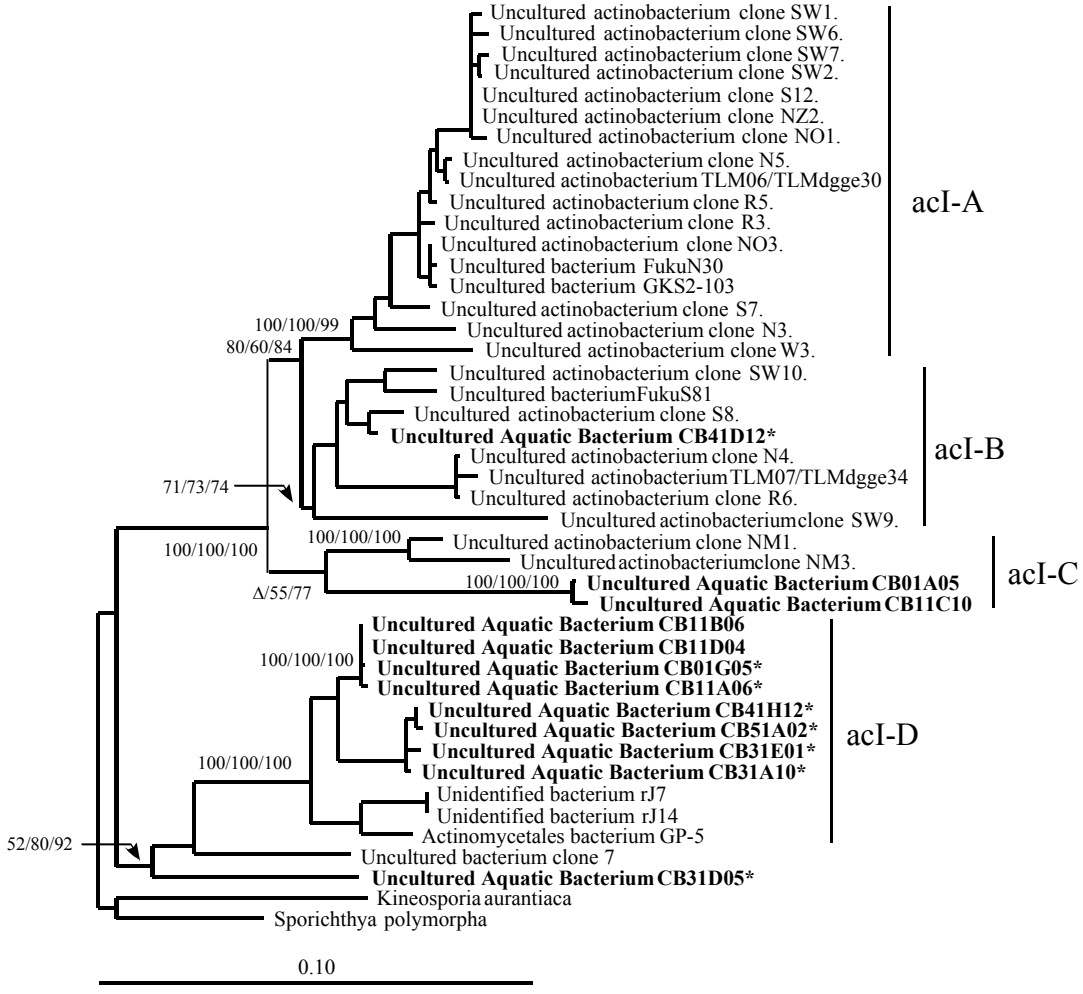


Fig. 4-4 Phylogenetic reconstruction (DNA distance - NJ) of *Actinobacteria* ACI clade.

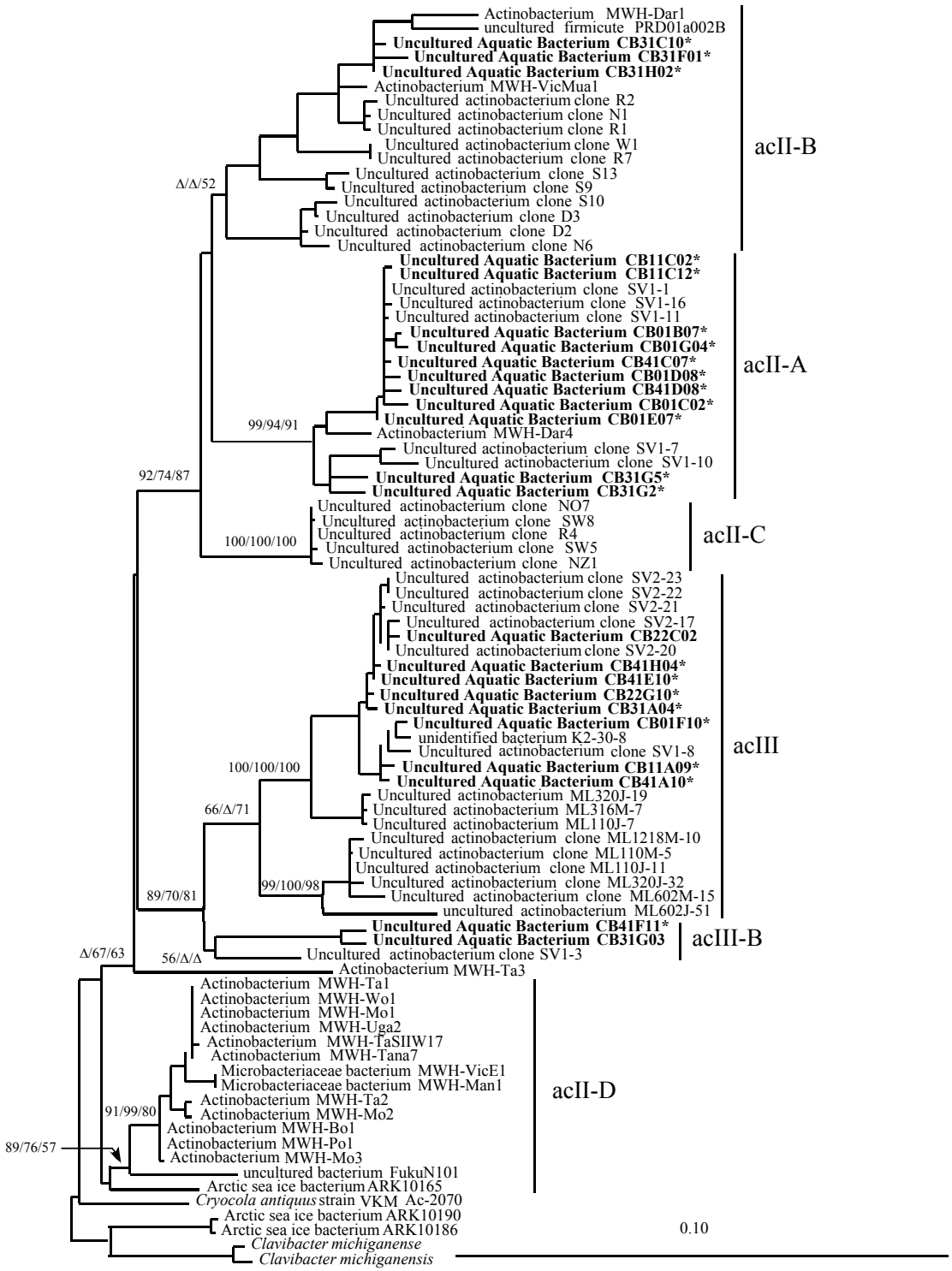


Fig. 4-5 Phylogenetic reconstruction (DNA distance - NJ) of *Actinobacteria* ACII and ACIII clades.

Clones associated with clusters acII and acIII were closely related to those obtained from Changjiang river (Sekiguchi et al. 2002) and seven lakes from Europe (Warnecke et al. 2004). Clones obtained from this study fell into subcluster acII-A, acII-B and cluster acIII (Fig. 4-5). Same as acI, Freshwater acII represented another group absent in marine environments. So the fact that these groups were found at salinity ranging from 10 to 27 ppt in the Chesapeake Bay was quite remarkable. Actinobacterial cluster acIII previously only contained sequences derived from two meromictic lakes (Lake Saelenvannet and Mono Lake) (Humayoun et al. 2003; Warnecke et al. 2004). In our clone libraries, two clones (CB41F11 and CB31G03) were clustered with Lake Saelenvannet clone SV1-3, which might represent a novel subcluster of acIII.

Cluster acIV is predominantly constituted of sequences from freshwater and estuarine environments (Zwart et al. 2002). In reconstructed phylogeny of acIV, subcluster acIV-A and acIV-B (Fig. 4-6) were consistent with the phylogeny proposed by Warnecke et al. (Warnecke et al. 2004). In addition, four Chesapeake Bay clones including two fully sequenced 16S rRNA genes were clustered with clones from Mono Lake, Hawaiian Archipelago and Arctic pack ice. They formed two novel subclusters (acIV-C and acIV-D) within acIV (Fig. 4-6).

Other bacterial groups. *Betaproteobacteria* was more commonly found in the Chesapeake Bay during March 2003 (Chapter 3). Subgroup GKS98 (Zwart et al. 2002) was the most abundant subgroup. High microdiversity was observed in *Bacteroidetes* and clone sequences were positioned within diverse groups (Chapter 3). Unique and diverse marine cluster B *Synechococcus* were found in the

Chesapeake Bay and detailed phylogeny based on 16S rRNA gene, ITS and *rbc*L gene were discussed elsewhere (Chen et al. 2006).

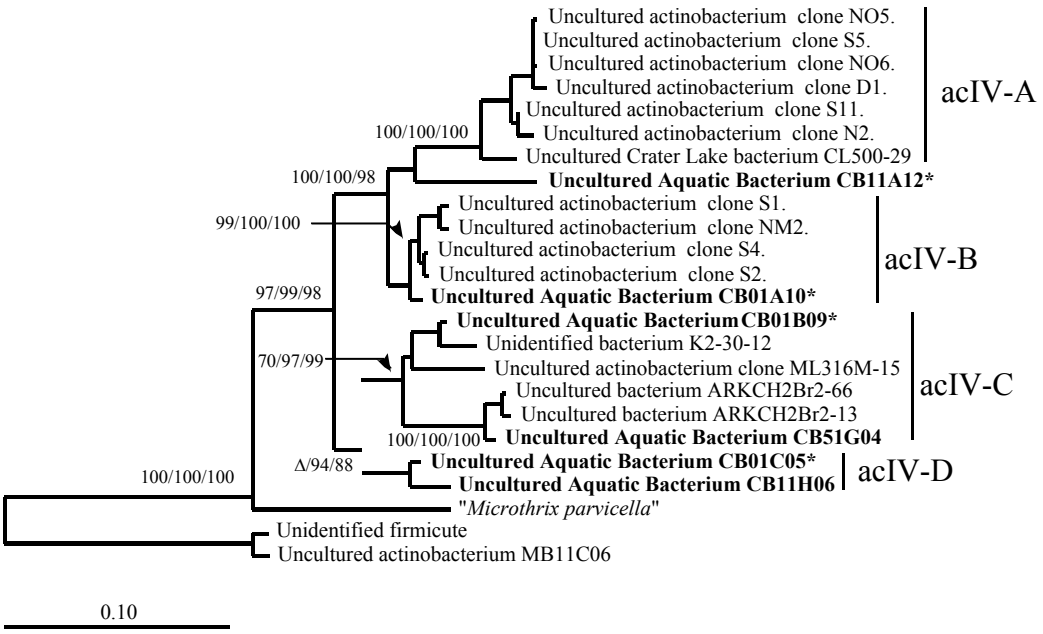


Fig. 4-6. Phylogenetic reconstruction (DNA distance - NJ) of *Actinobacteria* ACIV clade.

Considering the phylotypes obtained in this study, moderate coverage (60.2-75.3%) of bacterial species were detected in the clone libraries (Chapter 3). However, a high level of heterogeneity within *Gammaproteobacteria* and *Bacteroidetes* were observed in the Chesapeake Bay, suggesting that estimates of species richness within Chesapeake Bay bacterial clusters could be underestimated. Although ribosomal RNA gene only provides a conservative estimate of the actual diversity, detailed and comprehensive analysis of an costal bacterioplankton community identified highly

micro-diverse phylogenetic clusters within bacterial groups, based on 99% similarity cutoff of 16S rRNA gene (Acinas et al. 2004). Currently, by applying metagenomic approach, high microdiversity within natural communities has been observed in surface water of Sargasso sea (Venter et al. 2004) and along the depth continuum in the North Pacific Subtropical Gyre (DeLong et al. 2006). Estuarine ecosystems are more dynamic than open oceans in terms of physical, chemical and biological gradients and thus more diverse and discrete populations coexist in the estuary. Such diversification, most likely genetic variation of rRNA genes instead of novel ecotypes, are believed to be regulated by the local environments. Dynamic environments in estuaries enforce niche separations and thus contribute to emerging of unique bacterial groups.

Conclusion

The results reported here provide the first picture of genetic diversity and population dynamics of bacterioplankton in the Chesapeake Bay. The ITS-LH-PCR identification and subsequent phylogenetic construction supported the Chesapeake Bay contains bacteria originating from freshwater and marine origins, although there appear to be novel groups that was not previously retrieved from either system. ITS-LH-PCR offers a fast, high-throughput, and informative approach to prescreen environmental clones. SAR11, roseobacters, SAR86, and *Actinobacteria* contained sequences recovered exclusively in this study and the clusters were maintained by different phylogeny reconstructions, suggesting that indigenous organisms occur in the Bay, and perhaps other large temperate estuaries. However, to date, no cultivated strains from these groups have been isolated and there is no information regarding

their physiological and ecological properties. We believe that future studies combined with cultivation efforts will offer the basis for better understanding the metabolic traits and biogeochemical relevance of bacterioplankton in estuarine ecosystems.

Chapter 5: Bacterioplankton community in the Chesapeake Bay: predictable or random assemblages

Abstract

We monitored bacterioplankton communities from the Chesapeake Bay over two years (2002-2004) using denaturing gradient gel electrophoresis (DGGE) of PCR-amplified 16S rRNA gene. Chesapeake Bay bacterioplankton exhibited a repeatable annual pattern and strong seasonal shifts. In winter the bacterial communities were dominated by *Alphaproteobacteria* and *Actinobacteria* while in summer the predominant bacteria were members of *Alphaproteobacteria*, *Gammaproteobacteria*, *Cyanobacteria*, *Actinobacteria*, *Planctomycetes*, and *Bacteroidetes*. Phylotypes of *Alphaproteobacteria* and *Actinobacteria* present in warm seasons were different from those in cold seasons. Relatively stable communities were present in summer-fall across the sampling years while winter communities were highly variable interannually. Temporal variations in bacterial communities were best explained by changes of chlorophyll *a* (Chl *a*) and water temperature but dissolved oxygen, ammonia, nitrite and nitrate, and viral abundance also contributed significantly to the bacterial seasonal variations.

Introduction

Over the last two decades, our view of aquatic bacterial communities has changed considerably with the application of molecular techniques. With the advantages of cultivation-independence, molecular techniques determine the structure

of bacterial communities by characterizing indicative macromolecules, generally rRNA genes, directly isolated from the environments (Giovannoni et al. 1990; Ward et al. 1990). Community fingerprinting approaches such as denaturing gradient gel electrophoresis (DGGE) provide powerful tools for comparing bacterial communities (Muyzer et al. 1993). DGGE is a quick fingerprint technique and it can separate different PCR fragments even with single base-pair difference on a denaturant gradient gel (Muyzer et al. 1993). Diversity profiles from different microbial communities can be compared according to their gel patterns and the sequences of representative bands. Simultaneous comparisons of DGGE fingerprint patterns allow rapid assessment of changes in bacterial community structures over time and space.

Temporal variation in bacterial communities is an important and complex ecological process. Dramatic seasonal variations of bacterial community structures have been observed in marine, estuarine and freshwater ecosystems (Höfle et al. 1999; Pinhassi and Hagström 2000; Selje and Simon 2003), regardless of their environmental characteristics. Physiological predisposition and nutritional tolerance of dominant bacteria tend to maintain stable communities during certain seasons (Pinhassi and Hagström 2000). Meanwhile, it appears likely that bacteria are also influenced by abiotic characteristics and microbial food web structures of aquatic ecosystems (Yannarell and Triplett 2005). Previous studies have reported that population structures of bacterioplankton are correlated with salinity (Bouvier and del Giorgio 2002; Crump et al. 1999), nutrients (Biddanda et al. 2001), pH and water clarity (Yannarell and Triplett 2005), substrates resource (Crump et al. 2003), phytoplankton and Chl *a* (Murray et al. 1998; Pinhassi et al. 2004), grazing (Höfle et

al. 1999), and viral lysis (Fuhrman and Suttle 1993; Suttle 1994). However, given the indigenous characteristics among diverse aquatic ecosystems, environmental variables affecting the bacterial communities may also vary by site, time, and experiment.

The Chesapeake Bay contains strong physical, chemical, and biological gradients and provides a representative ecosystem to study the dynamics of estuarine bacterioplankton communities. Banding patterns of 5S rRNA showed that the compositions of bacterial communities from the Chesapeake Bay varied between summer and winter (Bidle and Fletcher 1995; Noble et al. 1997). By use of fluorescence in situ hybridization (FISH), Heidelberg et al. (2002) showed that *Gammaproteobacteria* exhibited strong seasonality in a Chesapeake Bay tributary (Choptank River). An annual DGGE fingerprint of bacterial community at Baltimore Harbor has shown that bacterial structure was more stable in summer-fall than winter and spring (Kan et al. 2006b). None of these studies in the Chesapeake Bay examined bacterioplankon dynamics inter-annually, however, making it unclear whether the bacterial community patterns vary from year to year. Moreover, little effort has been made to understand what environmental factors contribute to annual changes in bacterial communities. The fact that bacterial communities are affected by temperature (Heidelberg et al. 2002) and salinity (Bouvier and del Giorgio 2002) suggests that bacterioplankton in eutrophic habitats are regulated by hydrological factors in addition to nutrient availability. Bulk measurements of bacterial abundance and secondary production in the Chesapeake Bay were also found strongly dependent on water temperature (Shiah and Ducklow 1994). Thus, seasonal patterns of bacterial

communities should reflect the effects of aquatic environments. If the suite of environmental factors responsible for structuring the Chesapeake bacterial communities are known, then samples with similar values for these variables would be expected to contain rather similar bacterioplankton communities. Therefore, environmental variables, either stable or fluctuating on seasonal cycles, may be used to predict and interpret the occurrence of seasonality of bacterioplankton communities.

Multivariate analysis of variance is an appropriate statistical tool for defining variations of communities and relating the variations to changes of environmental variables. Nonmetric multidimensional scaling (MDS) attempts to arrange the bacterial communities in a space with certain dimensions (usually two or three dimensions) so as to identify community patterns and help to explain observed similarities or dissimilarities. MDS has been extensively applied to describe changes in bacterial communities over time or space (Crump et al. 2003; van Hannen et al. 1999). However, it is not possible to link the bacterial community variations with environmental changes using MDS. Canonical discriminant analysis (CDA), another multivariate analysis, can be used to determine what environmental variables discriminate the naturally occurring patterns. CDA classifies the variables and determines the optimal combination of variables via multivariate F tests. If the canonical discriminant functions are statistically significant, bacterial communities can be distinguished and predicted based on predictor variables included in these functions.

In this study, the population structures of bacterioplankton were investigated by DGGE at three stations along the main stem of the Chesapeake Bay. Sampling included three summer-fall seasons in two consecutive years (2002-2004). We reported seasonal variations of major phylotypes of bacterioplankton in the Chesapeake Bay and described the annual patterns that occurred in the middle and southern Bay from 2002 to 2004. We took band richness (alpha diversity) as a diversity index, and HARMONIC analysis of the diversity indicated a repeatable seasonal pattern in the Chesapeake Bay. By use of distance matrix constructed from DGGE band profiles (absence and presence of the bands), MDS defined the population structures in a multiple dimension space and samples with similar communities plot close to one another in 2D plots. Finally, the environmental variables that may explain or predict the bacterial seasonal patterns were determined using CLUSTER analysis and CDA.

Materials and methods

Sample collection. Water samples were collected at three stations along the middle axis of the Chesapeake Bay from Sep 2002 to Oct 2004 (Fig. 5-1). Station N (39°08′ N, 76°20′ W), M (38°18′ N, 76°17′ W), and S (37°07′ N, 76°07′ W) represented the northern, middle, and southern Bay, respectively. At each station, 500 ml surface water samples (below 2 m) were collected from 10 liter-Niskin bottles mounted on a CTD rosette on board the R/V *Cape Henlopen* and filtered immediately through 0.2-μm-pore-size polycarbonate filters (47-mm diameter, Millipore, Billerica, MA). The filters were stored at -20˚C. Water temperature, salinity and dissolved oxygen were

recorded on board. A subsample of 50 ml water was frozen at -20˚C for nutrient analysis.

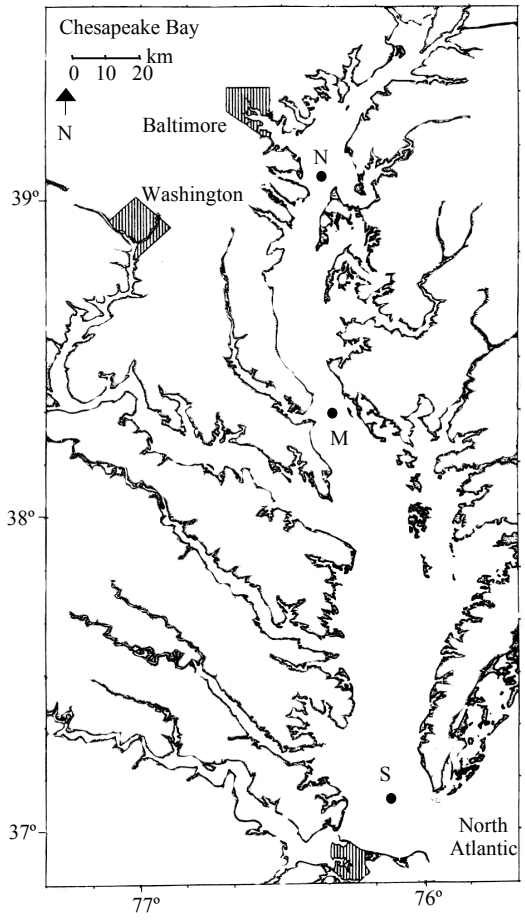


Fig. 5-1. Chesapeake Bay map showing sampling stations. N, M, and S represent the northern, middle, and southern Bay, respectively.

Chl *a* and nutrients analysis. Refer to the protocol described in Chapter 3.

Enumeration of bacteria and viral particles. As previously described in Chapter 3.

Extraction of nucleic acid and PCR amplification of 16S rRNA gene. As previously described in Chapter 3.

DGGE and banding patterns analysis. DGGE was performed as previously described (Crump et al. 2003; Muyzer et al. 1993) with modifications. Briefly, same amount of PCR products were separated on a 1.0-mm-thick vertical gel containing polyacrylamide (acrylamide-bisacrylamide, 37.5:1) and a linear gradient of the denaturants (urea and formamide), increasing from 40% at the top of the gel to 55% at the bottom. Electrophoresis was run in a DGGE-2001 system (C.B.S Scientific) at 65°C in a 0.5 × TAE buffer, and at 75 V for 22 h. Nucleic acids were visualized by staining with SYBR Gold (Øvreås et al. 1997) and photographed with a ChemiDoc imaging system (Bio-Rad). Defined as at least 5% of the most intense band in the sample, bands were scored as present or absent using the GelcomparII software package (Applied Maths). The numbers and positions of the bands on the gel were determined based on the vertical position of the bands in ladders. Banding patterns were compared with matching bands (absence and presence) and binary data was exported to Microsoft Excel for further statistical analysis.

Statistical analysis. All the statistical analysis described below was performed with HARMONIC regression, CLUSTER, MDS, and CANDISC procedures of the SAS System (SAS/STAT, 1992).

Harmonic regression analysis. To analyze the annual pattern of DGGE band richness, we conducted harmonic regression analysis (also known as trigonometric regression or cosinor regression). In this linear regression model, the predictor variables are trigonometric functions of a single variable, usually a time-related variable. We used least-square techniques to obtain parameter estimates of the equation:

$$Y_{jt} = \beta_0 + \mu_j + \sum(\beta_{1k}\sin(k\omega t) + \beta_{2k}\cos(k\omega t)) + \varepsilon, \text{ where } k=1, 2, \ldots n.$$

Y_{jt} is the band richness from DGGE gel; μ_j represents the jth year effect; β_{1k} and β_{2k} are estimated parameters for a given k value; ω is the frequency expressed in terms of radians per unit time, that is: $2\pi/12$, where π is the constant *pi*=3.1415… and 12 is the frequency of an annual cycle (12 months); the variable t is a continuous numeric value converted from time variable (e.g., starting time point 1 is Sep 2002 and 9 months later, Jun 2003 would be converted to 10). Significant first-order terms, i.e., k=1, indicate a dome-shaped annual pattern, and second-order terms (k=2) indicate a bimodal annual pattern, and so on.

Cluster analysis. To examine the relationship between bacterial communities, cluster analysis (Ward's minimum-variance method) was performed. The distance matrix was calculated and constructed by Jaccard coefficient based on the binary data from DGGE band patterns.

Nonmetric multidimensional scaling (MDS). MDS was performed based on the distance matrix. The differences between bacterial community DGGE patterns were illustrated in two-dimension MDS plots. The band patterns with the higher similarity are plotted closer and the band patterns with the lower similarity are located further apart. In order to judge the degree to which this ordination matches the distance matrix, the stress value of MDS was examined. Stress value less than 0.1 indicated a good ordination with little risk of misinterpretation of banding patterns (Clarke 1993).

Canonical discriminant analysis (CDA). Physical, chemistry, and biological variables of Chesapeake Bay water were analyzed by CDA to identify their relative contribution in discriminating among the DGGE band patterns of bacterial communities. Nine variables included temperature, salinity, *C*hl *a*, dissolved oxygen,

ammonia, nitrite and nitrate, phosphate, bacterial abundance and viral particle abundance. Because of significant pairwise correlations for some of the independent variables ($p < 0.05$), total canonical structure (TOC) was used to explain canonical discriminant functions (CDFs) (Momen et al. 1999).

Sequencing and BLAST. Representative DNA bands were excised from the gels and sequenced as previously described (Kan et al. 2006b). All sequences were compared with the GenBank database using BLAST, and the phylogenetic trees were constructed as previously described (Kan et al. 2006b).

Nucleotide sequence accession numbers. Sequences of DGGE bands obtained in this study were deposited in the GenBank database under accession no. DQ206714 – DQ206762.

Results

Hydrological conditions varied markedly in the Chesapeake Bay (Table 1). Salinity fluctuated in the sampling years and no annual trend was discovered. In the middle Bay, the salinity varied from 10.1 to 15.6, except Sep 2002 when salinity reached 19.4. In the southern Bay, salinity exhibited stronger fluctuation than middle Bay and varied from 15.4 to 26.8. Water temperature exhibited a repeatable fluctuation, reaching the highest above 25°C in summer and lowest 1°C in winter (Table 5-1). Winter was defined by the low water temperature, which sometimes extended to early spring (e.g., Mar 2003 and Feb 2004). Bacterial and viral abundance followed a similar trend as temperature: high in summer and low in winter. In contrast, dissolve oxygen, ammonia, nitrite and nitrate peaked in the cold season and reached their lowest levels in warm season. Concentrations of nitrite and

nitrate were higher in the middle Bay than in the southern Bay. Chl *a* concentrations increased from early spring and peaked in summer (Table 5-1).

Seasonal dynamics of Bacterioplankton community in 2002-2003.

DGGE banding patterns showed that Chesapeake Bay bacterioplankton communities exhibited a great deal of seasonal variability (Fig. 5-2). Bacterial communities in early spring (Mar and Apr 2003) contain many unique populations that were not found in summer-fall (Sep 2002, Jul, Aug, and Oct 2003). Bacterioplankton community succession over 18 months was observed in the MD plot (Fig. 5-3). Bacterioplankton populations in the northern, middle, and southern Bay exhibited similar seasonal shifts. Cold season communities (Mar and Apr 2003) shifted to a transitional community in early summer (Jun 2003) and after July, the community formed relatively stable summer-fall communities (Jul 2003, Aug 2003, Sep 2002, and Oct 2003).

Forty-nine phylotypes were obtained from the representative DGGE bands and the closest phylogenetic affiliations were shown in Fig. 5-4 and Fig. 5-5. In warm seasons, phylotypes associated with *Alphaproteobacteria* (e. g., band 1, 2, 3, 4, 13, 40), *Gammaproteobacteria* (e. g., band 7, 12, 14, 21), *Cyanobacteria* (e. g., band 15, 27, 35, 38), *Actinobacteria* (e. g., band 16, 34, 48), *Planctomycetes* (e. g., band 39, 47), and *Bacteroidetes* (e. g., band 37, 45) were commonly seen. However, in cold seasons *Alphaproteobacteria* (e. g., band 8, 9, 19, 20, 30 31, 32, 33), *Betaproteobacteria* (e. g., band 11, 17, 18), and *Actinobacteria* (e. g., band 28, 43, 44) affiliated phylotypes were found. Although *Alphaproteobacteria* and *Actinobacteria* were present in both warm and cold seasons, the composition of

Table 5-1. Measurements of water quality variables and bacterial and viral abundances for middle Bay and southern Bay stations

Month	Water temperature (°C)	Salinity	Dissolved oxygen (mg L^{-1})	Chl *a* (μg L^{-1})	Ammonia (μmol L^{-1})	Nitrite and nitrate (μmol L^{-1})	Phosphate (μmol L^{-1})	Bacterial abundance (10^6 cells mL^{-1})	Viral abundance (10^7 cells mL^{-1})
Sep 2002 (902)[a]	24.4[b]/24.2[c]	19.4/26.8	6.71/6.45	5.0/3.0	0.79/0.95	4.27/1.37	0.46/0.36	4.96/4.11	5.38/5.21
Mar 2003 (303)	1.7/4.4	15.6/22.3	12.08/11.93	22.5/14.9	0.59/3.0	17.6/2.83	0.48/0.33	0.57/0.45	0.81/0.64
Apr 2003 (403)	8.4/9.4	10.5/24.0	10.45/10.19	20.9/32.1	3.61/2.6	41.3/0.81	0.32/0.39	1.52/1.44	1.8/1.24
Jun 2003 (603)	16.9/18.2	12.7/17.2	7.79/9.85	38.5/29.7	3.41/0.65	11.5/1.2	0.42/0.31	3.8/4.64	2.74/1.73
Aug 2003 (803)	27.6/26.3	11.2/17.1	3.06/3.48	9.6/7.3	1.46/0.44	4.27/0.1	0.21/0.2	2.26/3.73	1.68/2.0
Oct 2003 (1003)	19.6/20.6	13.7/19.1	7.51/7.75	7.0/12.3	0.66/0.77	6.7/1.58	0.18/0.23	0.79/0.51	0.71/0.15
Feb 2004 (204)	1.0/3.8	10.5/15.4	11.77/7.94	5.0/4.8	0.61/3.55	15.6/12.3	0.18/0.29	0.65/0.86	0.26/0.1
Mar 2004 (304)	6.2/7.5	13.7/21.8	7.7/9.24	6.5/8.7	1.16/2.6	18/11.3	0.14/0.29	0.3/1.15	0.08/0.31
May 2004 (504)	16.5/16.5	10.5/20.7	4.75/7.15	17.1/7.5	0.46/0.69	19.4/9.81	0.17/0.19	2.78/1.83	0.99/0.71
Jun 2004 (604)	22.7/20.9	10.1/19.8	3.88/3.5	15.8/10.5	0.59/1.11	11.7/1.85	0.36/0.44	4.76/5.62	0.95/0.92
Aug 2004 (804)	26.1/25.6	13.7/24.1	6.22/6.49	6.3/8.0	0.59/1.25	4.72/0.72	0.34/0.67	3.17/3.74	3.9/3.08
Oct 2004 (1004)	16.6/16.5	12.2/16.3	4.66/5.9	18.6/20.4	0.6/0.7	5.0/1.4	0.23/0.3	3.39/3.82	3.59/3.73

[a] Numbers in parentheses stand for corresponding month and year (same in figures below)

[b] Middle Bay

[c] Southern Bay

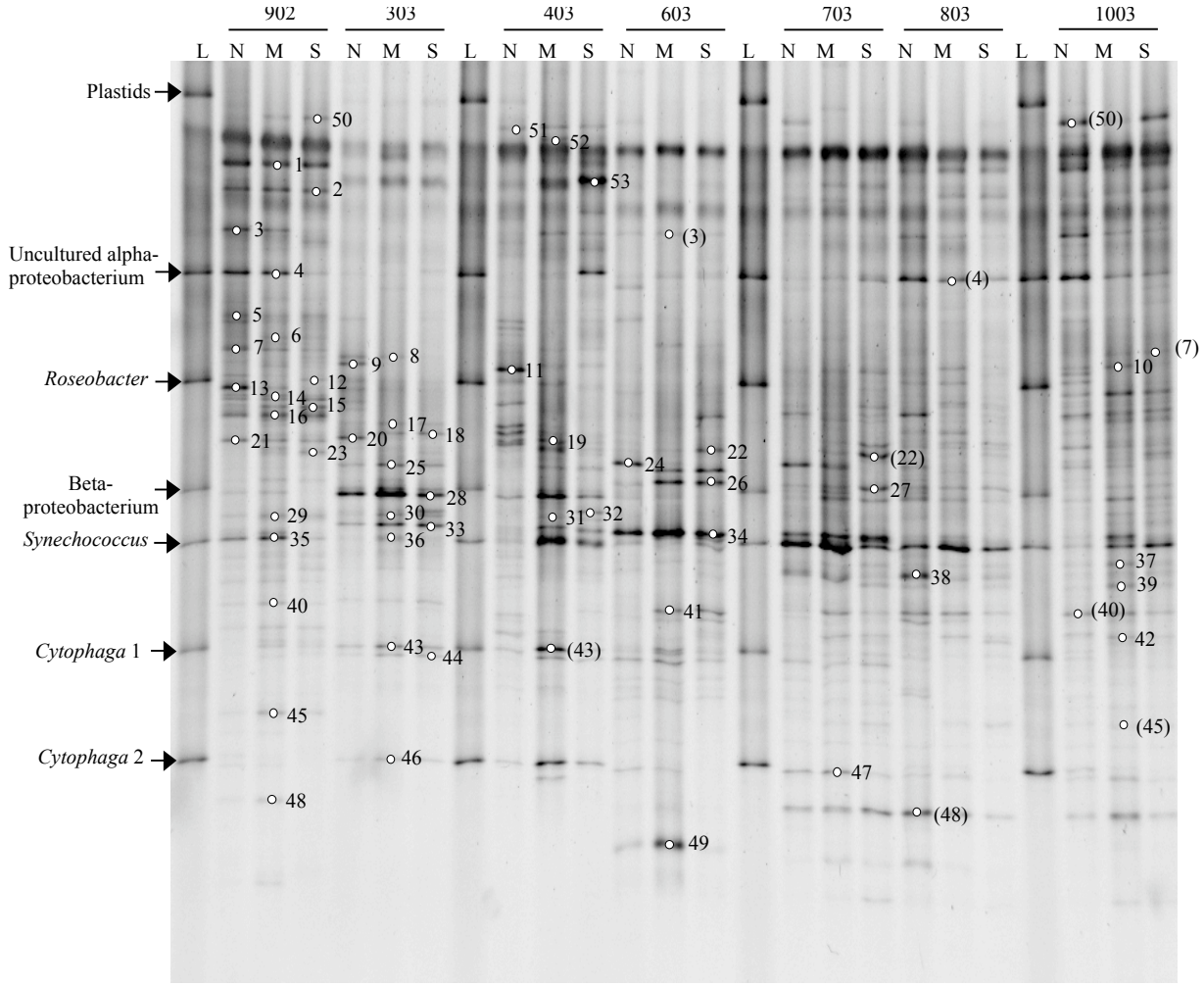


Fig. 5-2. Annual DGGE patterns (Sep 2002 to Oct 2003) of Chesapeake Bay bacterioplankton communities. No. 1 to 53 are representative bands excised and sequenced. Band (3), (4), (7), (22), (40), (43), (45), (48), and (50) are additional bands that are sequenced to confirm that the bands at the same vertical position contain the same sequence. N, M, and S represent the northern, middle, and southern Bay (Fig. 5-1). L is DGGE band marker consisting of six different bacteria and one plastid as indicated.

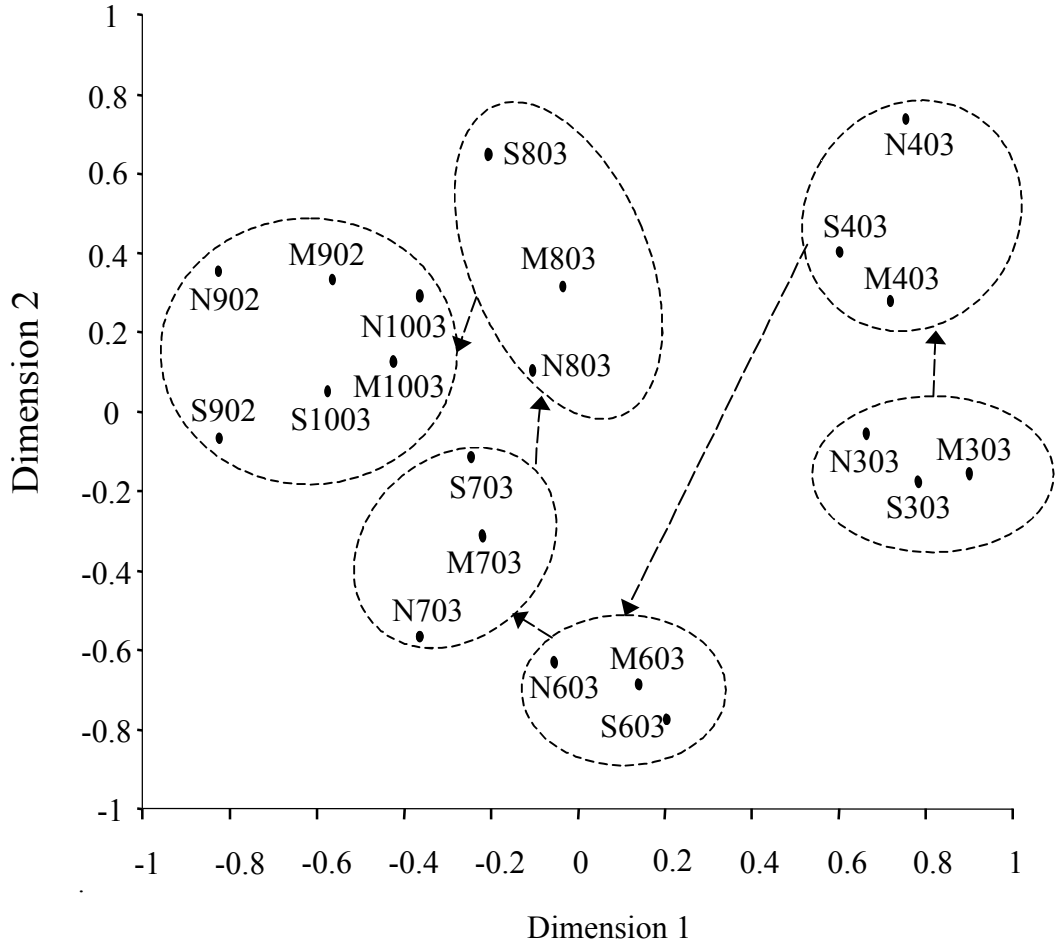


Fig. 5-3. MDS plots for DGGE banding patterns shown in Fig. 5-2. Sampling months are indicated next to each point. N, M, and S represent the northern, middle, and southern Bay (Fig. 1). Stress = 0.039.

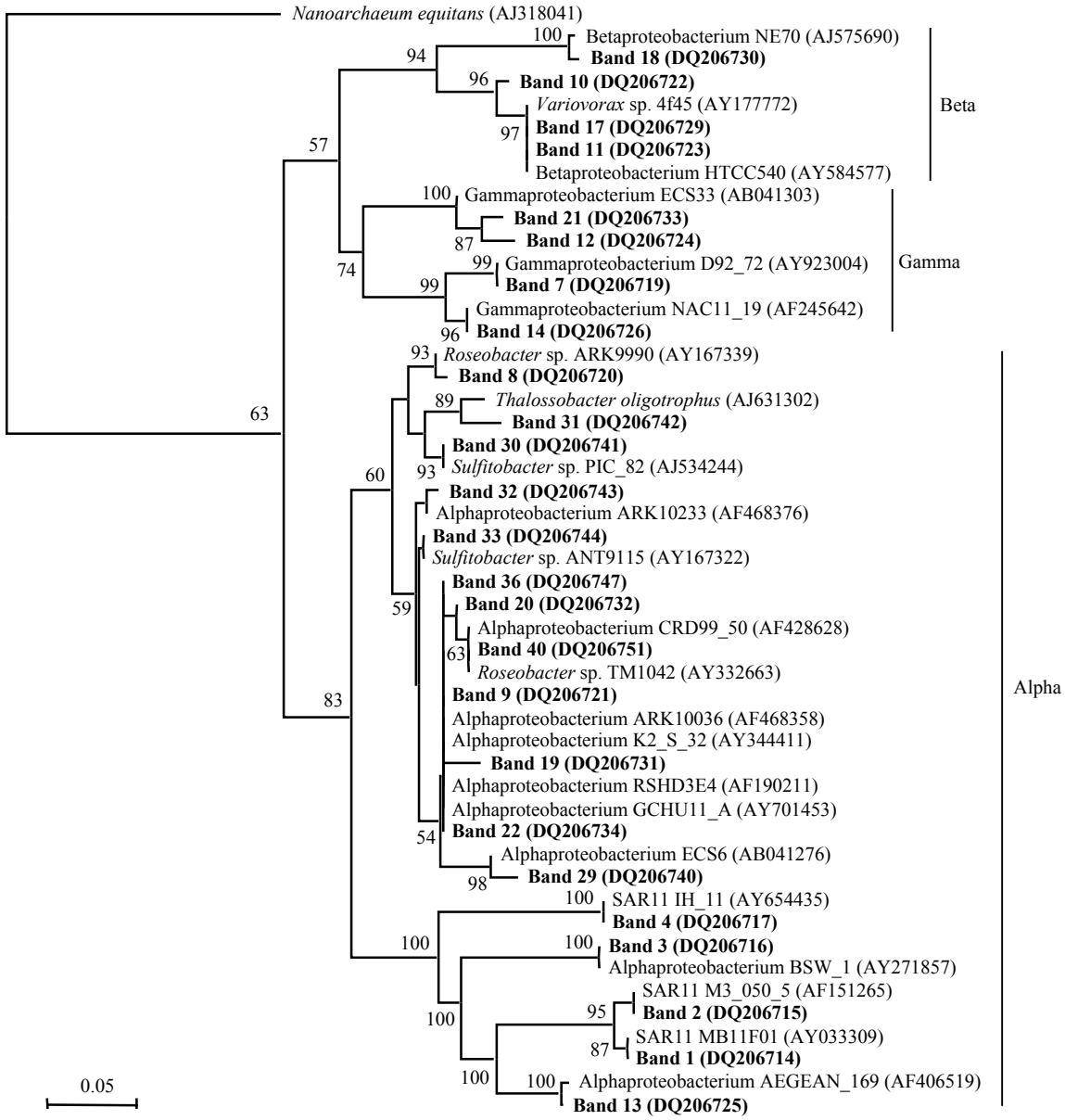


Fig. 5-4. Phylogenetic affiliations of DGGE band sequences related to *Proteobacteria*. Bands were excised from DGGE gel shown in Fig. 5-2. Sequences from this study are in boldface type. Bootstrap values were based on 1000 replicated trees. *Nanoarchaeum equitans* is used as an outgroup. Scale bar represents 0.05 substitutions per site.

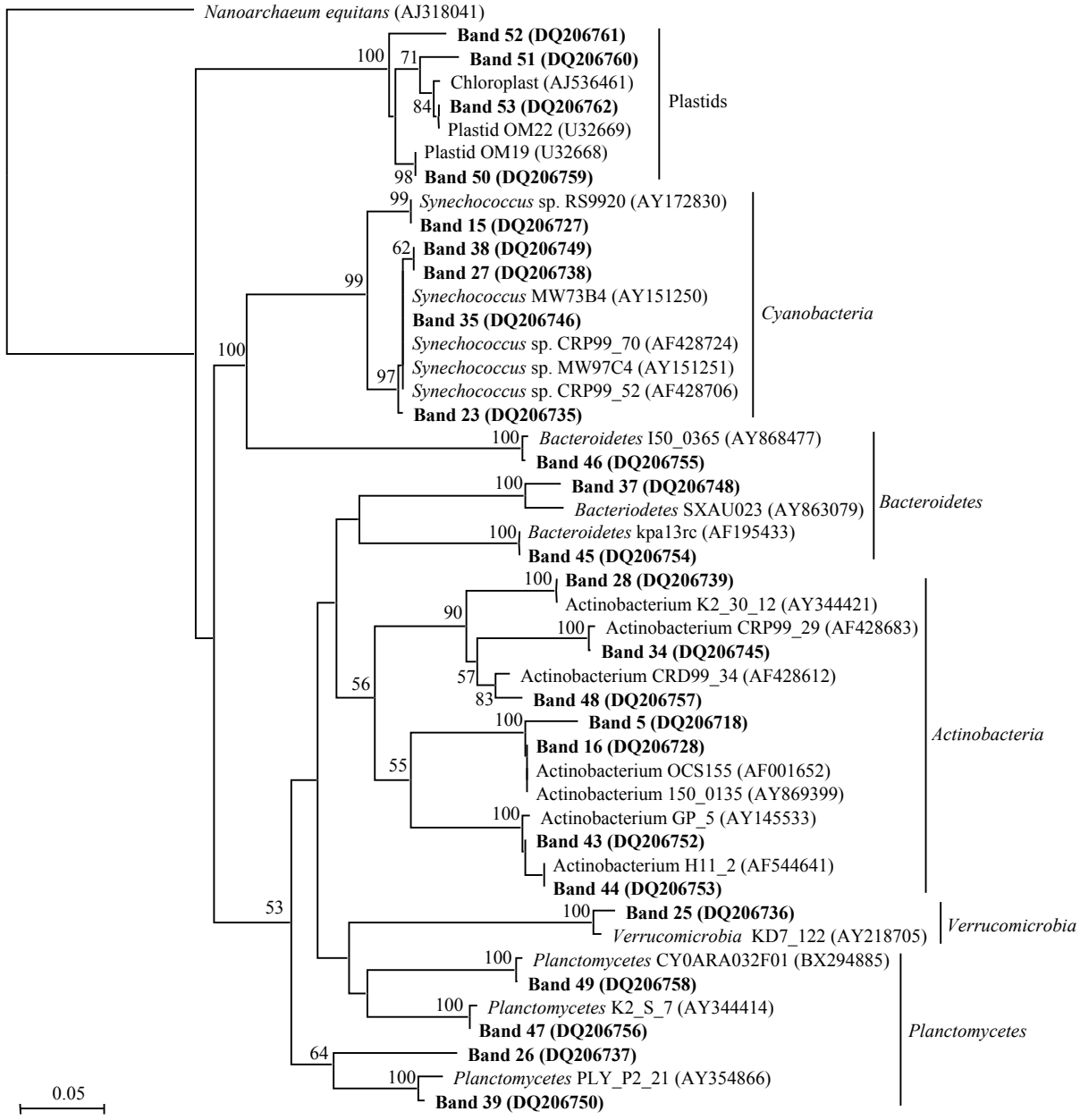


Fig. 5-5. Phylogenetic affiliations of DGGE band sequences related to Plastids, *Cyanobacteria*, *Bacteroidetes*, *Actinobacteria*, *Verrucomicrobia* and *Planctomycetes*. Bands were excised from DGGE gel shown in Fig. 5-2. Sequences from this study are in boldface type. Bootstrap values were based on 1000 replicated trees. *Nanoarchaeum equitans* is used as an outgroup. Scale bar represents 0.05 substitutions per site.

phylotypes shifted. Transient populations, including phylotypes related to *Alphaproteobacteria* (e. g., band 22), *Planctomycetes* (e. g., bands 26, 49), and *Actinobacteria* (e. g., band 34) were present in Jun 2003 (603). Sequences of band (3), (4), (7), (22), (40), (43), (45), (48), and (50) are identical to the bands at the same vertical positions (i.e. bands 3, 4, 7, 22, 40, 43, 45, 48, and 50). Bands 6, 24, 41, and 42 failed to be re-amplified and therefore no sequences were obtained.

Inter-annual patterns for community structure and bacterial richness in 2002-2004.

Pattern-forming bands were identified and highlighted in 24 samples (middle and southern Bay) from Sept 2002 (902) to Oct 2004 (1004, Fig. 5-6). Bacterial communities in summer and fall were relatively stable compared to those in winter. Eighteen common bands were shared among samples from Sep 2002 (902), Aug 2003 (803), Oct 2003 (1003), Aug 2004 (804), and Oct 2004 (1004). In contrast, in cold seasons 10 common bands were present, among which five bands appeared only in Mar 2003 (303), Apr 2003 (403), Feb 2004 (204), and Mar 2004 (304). In addition, six unique pattern-forming bands were found in cold seasons. Three of them were observed in Mar 2003 (303), and Apr 2003 (403) and the other three were present only in Feb 2004 (204), and Mar 2004 (304).

Cluster analysis grouped the 24 bacterial communities into 4 classes, winter 2003, winter 2004, early summer 2003, and summer-fall 2002-2004 (data not shown). MDS analysis on these samples highlights the annual succession of the bacterial

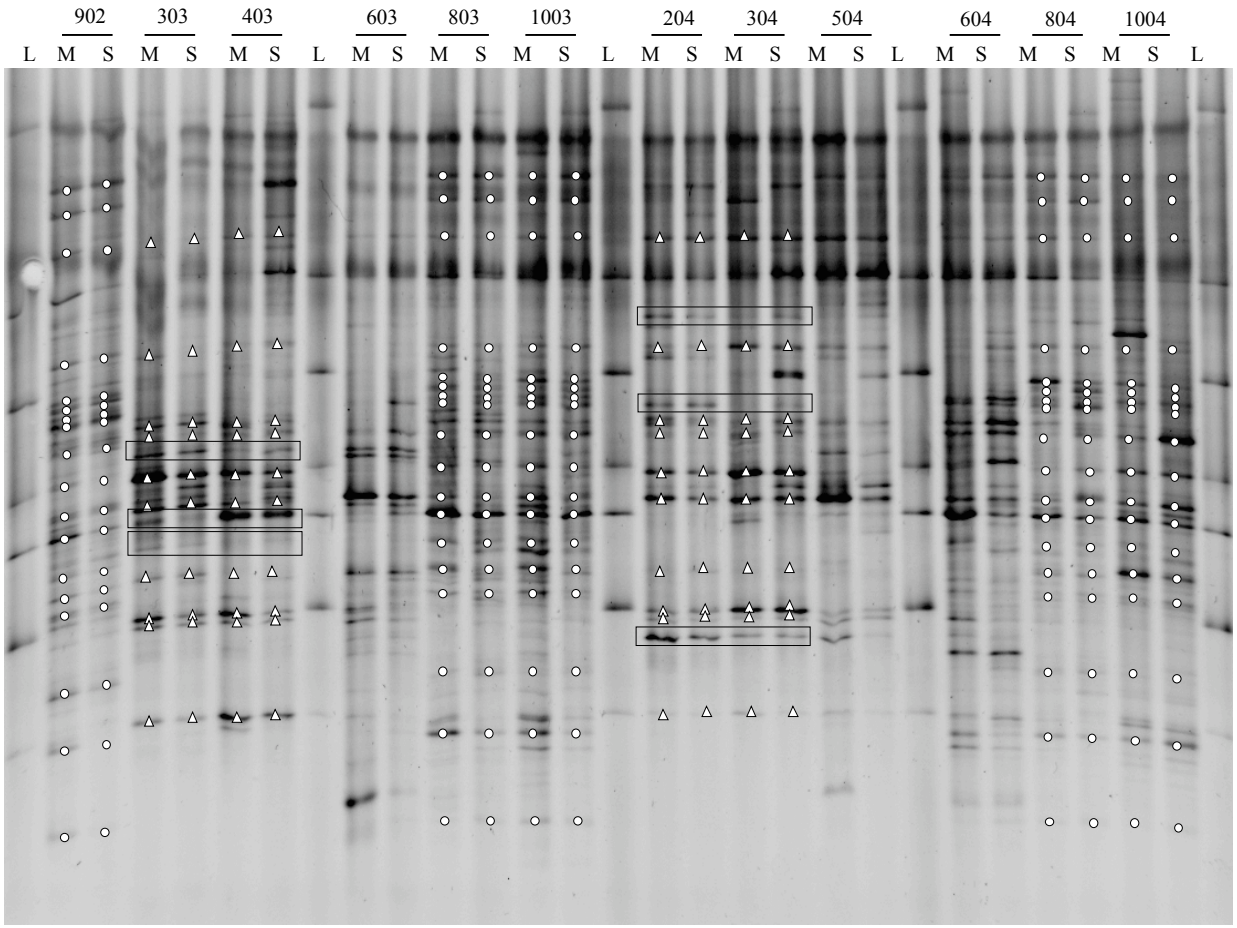


Fig. 5-6. Inter-annual variations (Sep 2002 to Oct 2004) of Chesapeake Bay bacterioplankton communities revealed by DGGE. Only samples from the middle (M) and south Bay (S) were analyzed. Symbols categorize the bands as important pattern-forming bands in summer-fall (open circle) and winter (open triangle). Unique bands appeared in winter 2003 or 2004 are shown in the rectangular box. L, DGGE band marker (same as Fig. 5-2).

communities in the Chesapeake Bay (Fig. 5-7). Although the samples from different years showed variability, generally the bacterial communities shift between winter and summer-fall communities. Samples from Jun 2003 (d and d', Fig. 5-7) were different from either winter or summer-fall communities and could represent transient populations.

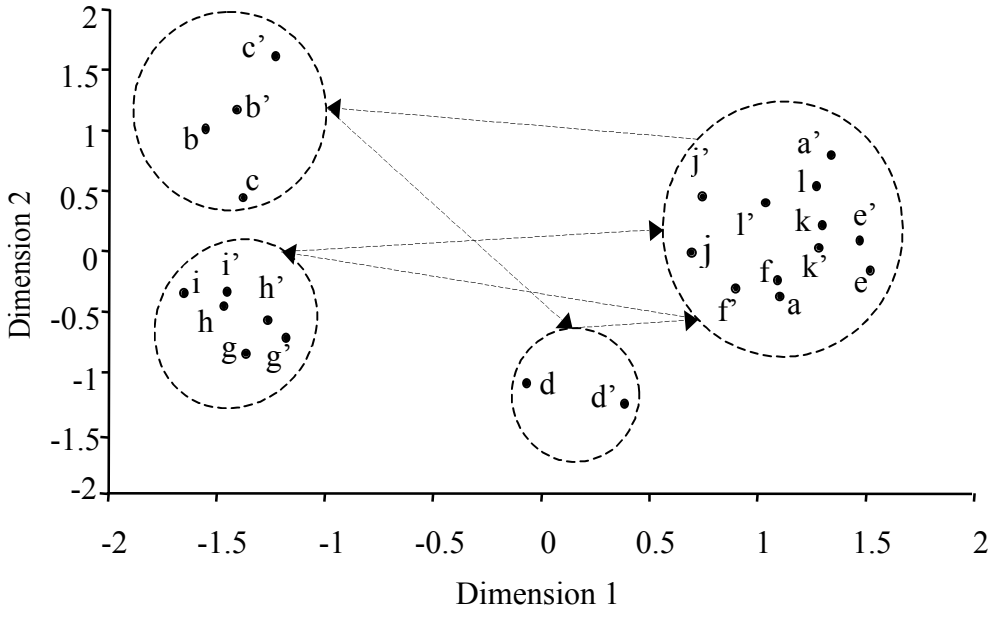


Fig. 5-7. MDS plots for DGGE banding patterns collected from 2002-2004. Each sample point is labeled with a letter. Letters a, a' to l, and l' correspond to sampling months and stations as shown in Fig. 5-6. a, M902; a', S902; b, M303; b', S303; c, M403; c', S403; d, M603; d', S603; e, M803; e', S803; f, M1003; f', S1003; g, M204; g', S204; h, M304; h', S304; i, M504; i', S504; j, M604; j', S604; k, M804; k', S804; l, M1004; l', S1004. Stress =0.028.

Although the DGGE band richness of bacterioplankton in the Chesapeake Bay varied seasonally, no significant difference was observed between the middle and southern Bay during the sampling period (paired *t*-test, $p = 0.18$, df = 11). Band richness of middle and southern Bay were used for the harmonic regression analysis. Only the 1st order cosine parameter was significant indicating a simple dome-shaped repeatable annual pattern (Fig. 5-8 and Table 5-2). The richness is well correlated with month, the time variable used in this study. In winter, low DGGE band richness was observed while summer communities contained more diverse populations. The lowest band richness (26) was observed in Feb 2004 and the highest (47) appeared in Aug and Oct 2003. Increased band richness occurred in spring to early summer. Band richness remained low following winter in Jun 2003, but was high in Jun 2004 (Fig. 5-8).

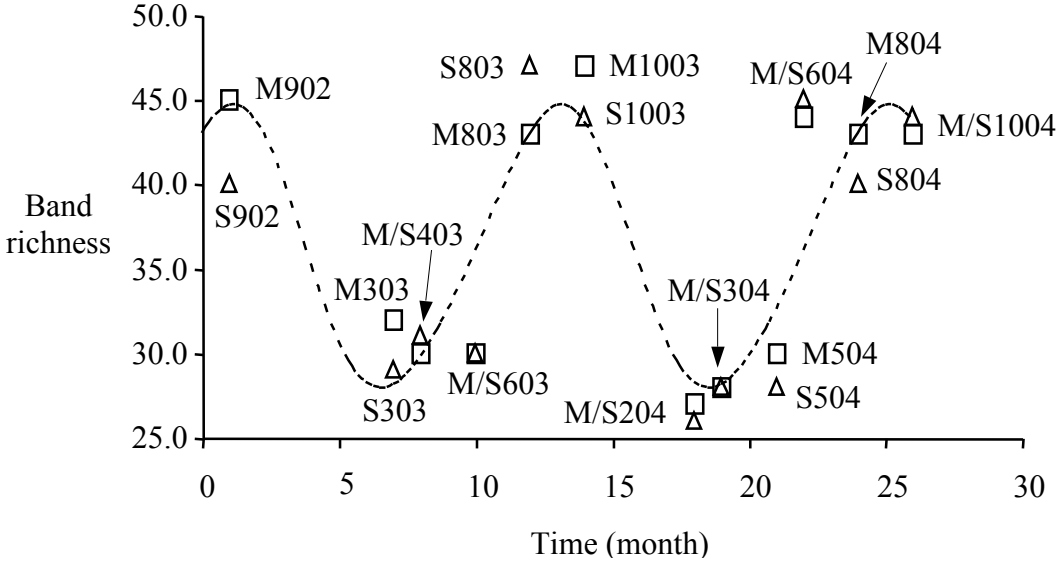


Fig. 5-8. Time series of DGGE band richness of Chesapeake Bay bacterioplankton from Sep 2002 to Oct 2004. Sampling months refer to Fig. 6. Open square: middle Bay (M); open circle: southern Bay (S).

Table 5-2. Harmonic regression parameter estimates for the annual pattern of bacterial species richness.

Bacterial species richness		$R^2 = 0.76$, df = 23, $p < 0.0001$		
Parameter	Estimates	S	*t* value	*p* value
Intercept, β_0	36	1.36	26.51	<0.0001
1 st order sine, β_{11}	3.77	2.02	1.87	0.08
1 st order cosine, β_{21}	7.38	1.60	4.62	0.0002
2 nd order sine, β_{12}	0.66	1.90	0.35	0.73
2 nd order cosine, β_{22}	- 0.36	1.28	- 0.28	0.78

Canonical discriminant analysis (CDA) of annual patterns of bacterial communities.

We included four bacterial community classes and nine independent variables in our CDA and hence three canonical discriminant functions (CDFs) were computed. Only the first CDF (CDF 1) and the second CDF (CDF 2) were significant and accounted for 99% of the variance (Table 5-3). Thus, the bacterial community-environment relationships were well characterized by the first two CDFs. In good accordance to MDS, bacterial communities from winter always plotted separately from summer communities, and winter communities from two different years were also easily distinguished (Fig. 5-9). The samples collected in Jun 2003 stood out in relation to other communities.

The correlation between the original variables and the loadings of variables for a given CDF were evaluated by total canonical structure (TOC). Among loadings on CDF 1, Chl *a* was the most significant loading variable ($p < 0.0001$). Dissolved

Table 5-3. Canonical discriminant functions (CDFs) and their correlations

CDF	Canonical correlation	Eigen value (proportion, cumulative)	Approximate *F* (numerator df, denominator df)	*p* [a]
1	0.96	10.53 (0.65, 0.65)	4.70 (27, 36)	< 0.0001
2	0.92	5.5 (0.34, 0.99)	2.76 (16, 26)	0.01
3	0.33	0.12 (0.01, 1.00)	0.24 (7, 14)	0.97

[a] The significance of individual CDFs can be inferred from eigen value or p value < 0.05

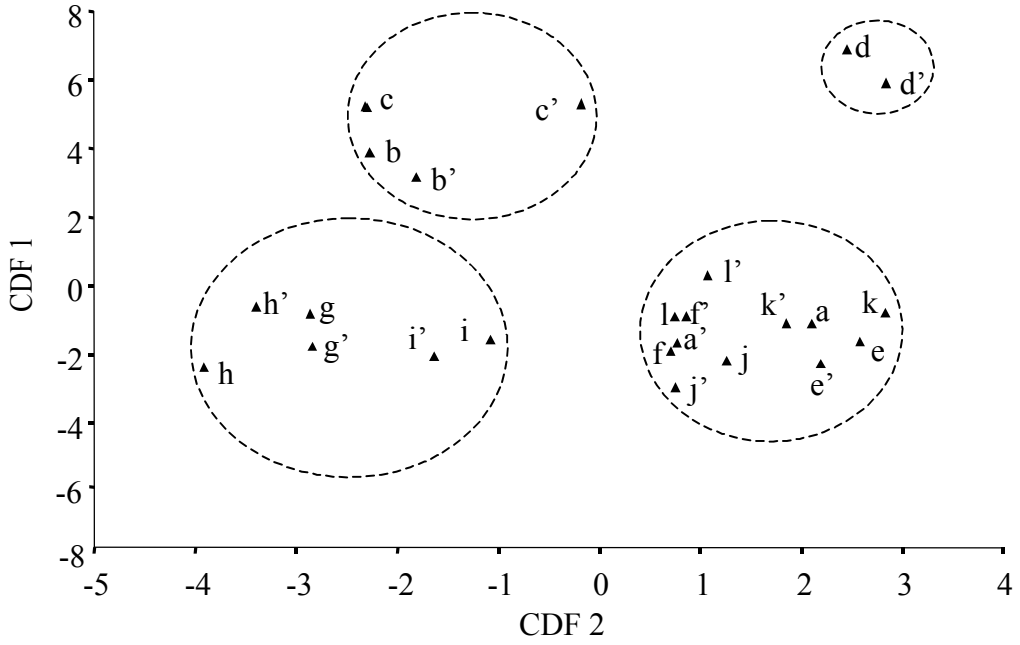


Fig. 5-9. Separation of the bacterioplankton communities collected in 2002-2004, based on the first and second canonical discriminant functions (CDF 1 and CDF 2). Labels a to l' are same as Fig. 5-7.

Table 5-4. Total canonical structure (TOC) and its significance (p) for 3 canonical discriminant functions

Variable	TOC (p) [a]	
	CDF 1	CDF 2
Chl *a*	0.85 (< 0.0001)	0.22 (0.3)
Temperature	- 0.42 (0.05)	0.84 (< 0.0001)
Salinity	0.02 (0.94)	0.02 (0.94)
Dissolved oxygen	0.66 (0.0004)	- 0.52 (0.0093)
Ammonia	0.53 (0.0084)	- 0.32 (0.13)
Nitrite and nitrate	0.25 (0.24)	- 0.59 (0.0024)
Phosphate	0.06 (0.79)	- 0.08 (0.7)
Bacterial abundance	- 0.16 (0.47)	0.71 (0.0001)
Viral abundance	-0.12 (0.58)	0.62 (0.0012)

[a] CDF 3 is not included because it has no significant role in discriminating bacterial communities

Table 5-5. Pairwise correlation coefficients between independent variables

	Chl *a*	Temperature	Salinity	Dissolved oxygen	Ammonia	Nitrite and nitrate	Phosphate	Bacterial abundance	Viral abundance
Chl *a*									
Temperature	-0.18								
Salinity	-0.14	0.1							
Dissolved oxygen	0.27	**-0.78** [a]	0.16						
Ammonia	0.31	**-0.45** [a]	0.1	**0.4** [a]					
Nitrite and nitrate	0.11	**-0.51** [a]	**-0.55** [a]	0.33	0.35				
Phosphate	-0.06	-0.18	0.19	0.093	**0.47** [a]	-0.03			
Bacterial abundance	0.04	**0.7** [a]	0.08	**-0.65** [a]	-0.34	-0.4	-0.03		
Viral abundance	-0.01	**0.57** [a]	0.21	-0.31	-0.19	-0.31	0.02	**0.56** [a]	

[a] Bolds are significant at $p < 0.05$

oxygen, and ammonia ($p = 0.0084$) also contributed significantly to CDF1 (Table 5-4). Decreases in Chl *a*, dissolved oxygen and ammonia corresponded with a transition of bacterial community from winter 2003 to winter 2004. The bacterial community of Jun 2003 was different from the majority of summer-fall communities and could be discriminated by CDF 1 as well.

Multiple significant variables were observed in CDF2. These variables included water temperature ($p < 0.0001$), bacterial abundance ($p = 0.0001$), viral abundance ($p = 0.0012$), nitrite and nitrate ($p = 0.0024$), and dissolved oxygen ($p = 0.0093$). All these variables were associated with the transition of bacterial communities from winter to summer-fall (Table 5-4, Fig. 5-9). However, bacterial abundance, viral abundance, nitrite and nitrate, and dissolved oxygen covaried with water temperature to some extent. Water temperature correlated positively with bacterial and viral particle abundances and negatively with nitrite and nitrate and dissolved oxygen (Table 5-5). So the variations between bacterial communities along CDF2 could be possibly triggered by temperature. Thus, temperature, Chl *a*, dissolved oxygen, ammonia, nitrite and nitrate, bacterial and viral abundance could generally discriminated the 24 bacterial communities into four distinct groups (Fig. 5-9).

Discussion

Chesapeake Bay bacterioplankton communities experienced strong seasonal succession from 2002 to 2004. The temporal differences in community structure were greater than the spatial differences during any sampling month. This result was consistent with previous studies in the Chesapeake Bay and other estuaries (Kan et al.

2006a; Noble et al. 1997; Selje and Simon 2003). DGGE fingerprints of bacterial communities and MDS plots indicated that the composition of bacterioplankton differed from winter and summer, and supported our results of LH (length heterogeneity) -PCR and clone library (Chapter 3&4). Changes in community composition between winter/early spring and summer were rapid rather than gradual, suggesting that few phylogenetic groups were able to overcome the environmental stresses over seasons. Although it is not clear how community replacement occurs, seasonality of bacterial succession may link to the environmental variables and intrinsic activity of the major phylotypes in the communities. Bacterial counts and bacterial growth followed the same trend (Wikner and Hagström 1991) indicating that this pattern was also reflected in the population size and activity.

Our DGGE fingerprints demonstrated reoccurring annual patterns in Chesapeake bacterioplankton. During annual succession, summer-fall communities appeared to be more stable than winter communities. Significant pattern-forming bands in summer-fall communities recurred in three years, suggesting that they represent an indigenous estuarine community. This stability is likely due to high bacterial growth rates and a relatively long residence time allowing estuarine bacterioplankton to overwhelm allochthonous populations of marine and freshwater populations (Crump et al. 2004). However, considerable inter-annual variations were observed in winters. Recurrent stable summer-fall bacterioplankton communities and variable winter communities appear to be regular features of this annual pattern. Seasonally variable but annually reassembling bacterioplankton communities have been reported in a high mountain lake (Pernthaler et al. 1998), California coastal

water (Fuhrman et al. 2004), and two temperate rivers (Crump et al. unpubl. data). However, one study conducted over three consecutive years on a humic lake in the Northen High-lands State Forest in Wisconsin indicated little similarity of bacterial community composition from year to year (Kent et al. 2004), suggesting that population dynamics may vary by site because of indigenous characteristics of the aquatic system.

Our results provide plausible explanations for seasonal variations of bacterial communities in the Chesapeake Bay. The annual shift in bacterial compositions appeared to be associated to the environmental variables. Successful classification of bacterioplankton by use of environmental variables (Fig. 5-9) suggested that the Chesapeake Bay undergoes predictable seasonal changes from year to year. Four classes of bacterioplankton resulted from cluster analysis were reconstructed along linear functions (CDF 1 and CDF 2) that were computed by CDA. Among nine hydrological and biological factors used for CDA, Chl *a*, temperature, dissolved oxygen, ammonia, nitrite and nitrate, bacterial abundance, and viral abundance corresponded significantly to changes in the bacterial communities.

Chl *a* was the most important variable in CDF1. Chl *a* and phytoplankton are important forces in structuring bacterial communities and archaeal communities (Kan and Chen 2004; Murray et al. 1998). During phytoplankton bloom senescence, bacterial abundance, cell activity in hydrolytic enzyme and growth rates increases substantially, which are potentially associated with significant shifts in bacterioplankton species composition (Riemann et al. 2000). Recent studies indicated that not only the phytoplankton biomass but the differences in phytoplankton species

composition also lead to pronounced shifts in bacterioplankton composition (Pinhassi et al. 2004). In the Chesapeake Bay, surface Chl *a* concentration increases in early spring and remains high during summer with moderate fluctuations from Jul to Sep 1991 (Malone et al. 1991). Significant difference of Chl *a* concentration was observed between winter/early spring 2003 and 2004. The appearance and disappearance of unique phylotypes of bacterial communities and changes in the relative abundance (i.e. band intensity) demonstrated that the population structure in winter 2003 was different from winter 2004, and thus the variation is likely associated with phytoplankton (diatom) blooms. Furthermore, high concentration of Chl *a* associated with samples in Jun 2003 explained why that bacterial community stood out from other communities. However, one study showed that there was no relationship between bacterial metabolism or composition and the distribution of Chl *a* along two transects of Chesapeake Bay rivers (Bouvier and del Giorgio 2002). These results suggest that changes in Chl *a* alone are not enough to drive the bacterial community successions.

Bacterioplankton are also affected by nutrients (Biddanda et al. 2001). Ammonia and nitrate are important nitrogen sources for heterotrophic bacteria (Kirchman et al. 2003a). For example, elevated ammonia concentrations favor the growth of ammonia-oxidizing bacteria (AOB), which were found to be more abundant in the Chesapeake Bay than other marine environments (Ward 1982). Subsequent observations of the depth distribution of ammonia oxidation rates indicated that most nitrification occurs in the surface waters (Ward and O'Mullan 2002). Our results showed that shifts in surface water bacterial communities were

significantly related to changes of ammonia ($p = 0.0084$) in CDF 1, and to changes in nitrite and nitrate ($p < 0.05$) in CDF 2. Another significant factor for both CDF 1 and CDF 2 was dissolved oxygen. Although hypoxia is generally restricted to the bottom waters of the Chesapeake Bay, surface water dissolved oxygen fluctuated remarkably over the seasons (Table 1). The annual spring inflow of fresh water initiates hypoxic and anoxic conditions in the Bay by delivering nutrients, increasing stratification, lowering salinity and affecting the residence time of the water (Boicourt 1992). Therefore, dissolved oxygen could be an important environmental factor affecting the temporal succession of bacterial communities in the Chesapeake Bay.

Another important source of variation in CDF2 was viral particle counts. Viruses cause prokaryotic mortality through host-specific cell lysis, and can influence bacterial community composition in various ways (Wommack and Colwell 2000). The seasonal correspondence of abundance and community patterns of both host and virus indicates that viruses hold the potential to structure the host community compositions (Wommack, unpubl. data). Apart from killing infected cells, viral lysis causes release of new materials including cytoplasmic and structuring material from host cells which can be important substrates stimulating the growth of non-infected bacterial populations (Middelboe and Lyck 2002). Furthermore, gene swapping through transduction, transformation, and conjugation probably influences the host speciation and diversification (Paul 1999). Therefore, viruses can affect the host community composition by 'killing the winner' (Thingstad and Lignell 1997), stimulating non-infected bacteria (Middelboe and Lyck 2002) and generating genetic

variability of bacteria through virus-mediated gene transfer (Weinbauer and Rassoulzadegan 2004).

The most significant variable in CDF 2 was water temperature. In temperate estuaries, temperature is considered to be an interactive limiting factor coupled with substrate supply to control bacterial biomass, growth, and respiration (Pomeroy and Wiebe 2001; Shiah and Ducklow 1994; Wikner and Hagström 1991). Correlation of water temperature and seasonality of *Gammaproteobacteria* implies that water temperature is also important in regulating bacterial community structure (Heidelberg et al. 2002). Within a moderate range, temperature could affect that how bacteria respond to changes in DOM supply (Kirchman and Rich 1997) and consequently affect the bacterial composition. Our DGGE band patterns showed that seasonal changes in water temperature were paralleled by shifts in bacterioplankton compositions. In the CDA, temperature successfully discriminated winter and summer-fall communities (Fig. 5-9). Meanwhile, bacterial abundance, viral abundance, nitrite and nitrate and dissolved oxygen correlated with water temperature to some extent (Table 5-5). This leads to the conclusion that water temperature may be an important environmental force triggering the seasonal variation of bacterioplankton communities in the Chesapeake Bay.

Surprisingly, no strong relationship between bacterial community and salinity was observed. The salinity range of the transect was between 10 and 20 and varied with season. Previous studies suggested a relationship between estuarine salinity gradients and the composition of estuarine bacterial communities (Crump et al. 1999) and, in particular, *Alpha*- and *Betaproteobacteria* (Bouvier and del Giorgio 2002).

Compared to significant seasonal variations, our MDS analysis on bacterial communities 2002-2003 showed moderate spatial variations along the Bay. However, salinity did not play a significant role in discriminating the community structures over the seasonal variations. Because of the long residence time of Chesapeake Bay water, indigenous bacterioplankton communities may remain relatively stable along the salinity gradient. The dominant bacterial groups in the Chesapeake Bay are probably able to resist changes in osmotic pressure with the adaptations of physiological features. Another minor variable is phosphate concentration. Since phosphate concentration remains relatively high and stable in the Bay, and it is not considered to be a limiting factor for microbial communities.

As a quick fingerprint technique, DGGE biases towards the abundant populations within a community (Kan et al. 2006b; Muyzer et al. 1993). The composition of the entire assemblage is not completely described based on the representative bands selected to sequencing. The minor groups are undetectable or form smearing bands on the gel and thus escape further characterized. Many factors including bias by PCR and other steps of molecular analysis can influence the outcome of PCR and therefore DGGE underestimate the diversity and complexity of natural microbial communities. Our statistical analyses are mainly based on DGGE band patterns and therefore it only provides a "snapshot" of the bacterioplankton dynamics in the Chesapeake Bay. Bacterial activity in aquatic ecosystem is very complicated and linking bacterial distribution to the environmental parameters is not straightforward. Limited by sampling size and cruise frequency, statistical analyses only provide partial view of the "real world" or even "false-positive" information. For

instance, CDA is able to identify the parameters regulating the population patterns observed, however, the direct correlation are still missing. All these limitations point to the necessity of further studies focusing on specific groups with more frequent samples.

Conclusion

We have shown that Chesapeake Bay bacterioplankton communities exhibited pronounced seasonal changes and repeatable annual patterns. Replacement of major phylotypes of bacteria from winter to summer-fall indicated that the dominant groups could not survive seasonal changes in environmental conditions. Covariations of the structure of bacterioplankton with environmental variables measured in this study were well constructed in MDS and CDA. We interpret the seasonal succession of bacterial community structure primarily as an interactive consequence of variations in several environmental factors. Temperature, Chl *a*, dissolved oxygen, nutrients and viruses all appear to play significant roles in structuring the bacterial communities in the Chesapeake Bay. However, considering the substantial phylogenetic, physiological, and metabolic diversity contained within these communities, it can be expected that they contain organisms with the ability to adapt to a wide range of environmental stresses. Thus, further studies of significant factors that contribute to the success of defined groups of bacteria or the total community will increase our understanding of estuarine microbial processes.

Chapter 6: Community proteomics of Chesapeake Bay microbial assemblages

Abstract

Microbes in the natural marine environment contain diverse species and complex processes. As the vast majority of microbes in the sea are still difficult to be cultivated, understanding the relationships between microbial diversity, microbial metabolism and their biogeochemical roles is one of great challenges facing microbial ecologists today. Metaproteomics is an approach to identity proteins present in microbial communities. In this study, this approach was validated using Chesapeake Bay microbial communities (0.2 to 3.0 micron). To obtain sufficient proteins, microbes in 20 L seawater were concentrated to ca. 150 ml using tangential flow ultrafiltration. The protein profiles based on the two-dimensional gel analysis showed that Chesapeake Bay microbial communities contained proteins with pI 4-8 and molecular masses between 10-80 kDa. Replicated middle Bay metaproteomes shared ~92% of all detected spots, but only shared 30% and 70% of common protein spots with upper and lower Bay metaproteomes. The metaproteomic patterns in the three stations were reflected by the variation of population structure based on the denaturing gradient gel electrophoresis. MALDI-TOF analysis of 34 highly expressed proteins produced no significant matches to known proteins in the database, suggesting that peptide fingerprints were not sufficient for identifying metaproteomes. Seven Chesapeake Bay proteins were analyzed and identified by LC-MS/MS sequencing and three of them matched hypothetical proteins annotated in the

Sargasso Sea metagenome. These three proteins include a predicted aminopeptidase, a subunit of the NADH:ubiquinone oxidoreductase or complex I, and a hypothetical protein with unknown function, respectively. They are of marine microbial origin and correlate with abundant Chesapeake Bay microbial lineages, *Bacteroidetes* and *Alphaproteobacteria*. Our results represent the first metaproteomic study of dynamic and highly complex marine microbial communities. As a culture-independent approach, metaproteomics has a great potential for unveiling microbial functional proteins and linking them with microbial population structure and microgeochemical cycling.

Introduction

Bacterioplankton contribute significantly to both primary production and biomass in the ocean and coastal water (Campbell et al. 1994; Li 1994). With an average concentration of approximately 10^6 cells ml^{-1}, bacterioplankton is an important catalyst of biogeochemical processes including oceanic carbon and nitrogen cycles (Hobbie et al. 1977; Azam 1998). Studying bacterioplankton is challenging because most groups either have never been cultivated (Giovannoni et al. 1990; Amann et al. 1995) or grow to very low density in the laboratory (Rappé et al. 2002). Culture-independent molecular approaches have indicated that environmental bacterial communities are more complex and diverse than previously thought (Giovannoni et al. 1990; Amann et al. 1995; Ward et al. 1990). Metagenomics is the direct cloning, sequencing, assembly and annotation of DNA from microbial communities and has been applied to waters, soils and extreme environments (Béjà et al. 2000b; Rondon et al. 2000; Tyson et al. 2004; Venter et al. 2004). A recent

metagenomic study of the Sargasso Sea revealed that substantial complex microbial diversity exists in the ocean: 148 novel bacterial phylotypes and more than a million of previously unknown genes were discovered and annotated (Venter et al. 2004). As genomic data accumulates from pure cultures and environmental communities, it becomes critical to understand gene expression and protein function. While metagenome sequences provide valuable information on potential functions, accurately predicting ecological function from sequence is nearly impossible without information on what proteins are synthesized under specific conditions (Lopez 1999; Petersohn et al. 2001; Eymann et al. 2002). To address this question, post-genomic molecular approaches such as microarrays to monitor mRNA abundance (Conway and Schoolnik 2003) have been developed. In addition, as proteins/proteomes are the ultimate functional products of genes/genomes, proteomic studies of microbial communities (metaproteomics) are an obvious approach to advance our understanding of microbial community function.

Metaproteomics can provide a direct measurement of functional gene expression in terms of the presence, relative abundance and modification state of proteins (Blackstock and Weir 1999; Wilmes and Bond 2004). Proteomics and metaproteomics rely on two-dimensional gel electrophoresis (2D-PAGE) coupled with mass spectrometry (MS) based protein identification relying on mass based (MALDI-TOF MS) or sequence based (LC-ESI-MS/MS) methods. These techniques have only been applied in limited scope to environmental microbial communities. One-dimensional gel electrophoresis (1D-PAGE) coupled with radioactive labeling or enzymatic activity assay has been used to study proteins induced in response to

environmental stresses (Ogunseitan 1997; Ogunseitan 1998). However, little concrete information on the sequences or identities of induced proteins emerged from these studies. A metaproteomic approach was applied to a laboratory-scale activated sludge bioreactor resulting in the identification of three highly expressed proteins presumably originating from an uncultured *Rhodocyclus*-type polyphosphate–accumulating organism (Wilmes and Bond 2004). More recently, using genomic and mass spectrometry-based proteomic methods, metaproteomes from an acid mine drainage (AMD) microbial biofilm community have been identified and linked their *in situ* functions to the challenging environments (Ram et al. 2005). However, all these studies are dealing with low-complexity microbial communities. So far, no studies have yet applied proteomic approaches to natural aquatic microbial communities.

Estuaries represent one of the most complex and productive ecosystems. The Chesapeake Bay is the largest estuary in United States (Fig. 6-1). It has received a great deal of attention because of its large geographic span and economic significance. With strong environmental gradients, it provides an ideal model system for integrated investigations on composition and function of microbial communities. In this study, we developed a metaproteomic approach to document microbial community protein profiles along a transect of the Chesapeake Bay. Significant differences were noted between proteomes collected at different sites and metaproteome patterns accurately predicted the relationship of sites as determined by 16S rRNA gene PCR-DGGE (denaturing gradient gel electrophoresis). Furthermore, proteins identified from Chesapeake Bay samples appeared to originate from marine

bacterioplankton. This study demonstrates that metaproteomic approaches can be successfully applied to naturally occurring and complex microbial communities in their native habitats.

Materials and methods

Eight bacterial strains isolated from upper Chesapeake Bay (Baltimore Inner Harbor) were used in this study. Based on 16S rRNA gene sequences, these bacteria have been identified as *Vibrio vulnificus*, Marine *Bacillus* sp., *Marinomonas* sp., *Psychrobacter pacificens*, *Pseudomonas* sp., *Pseudoalteromonas* sp., *Shewanella* sp., and *Hahella* sp. respectively (Kan unpublished). These bacteria were grown in 1/2 YTSS broth (4 g Yeast Extract, 2.5 g Tryptone per liter dissolved in *in situ* water) and harvested at the exponential growth stage using centrifugation (10,000 × *g*, 5 min, 4ºC).

To determine if microbial community analysis by 2D SDS-PAGE is feasible and representative, a simple artificial mixed microbial community was constructed using three bacterial strains of differing size: *Chlorobium tepidum* strain WT2321 (~0.5-0.8 μm cell length), *Escherichia coli* strain JM109 (~1.2-1.6 μm cell length), and an uncharacterized strain of *Pseudomonas fluorescens* (~8-10 μm cell length) (kindly provided by G. A. O'Toole, Dartmouth University). Protein content per cell for each strain was determined by measuring protein via a modified Bradford assay (Bio-Rad) and direct cell counting on replicate samples for each organism. Communities containing the same amount of protein for each strain were constructed by mixing appropriate volumes of pure cultures. The mock community was then diluted into 5 l of 10 mM potassium phosphate buffer (pH = 7.2) to specific cell

densities and the cells recovered. Total protein extracts of the mock community and each member strain were made by pelleting cell samples in a microcentrifuge and extracting proteins by resuspending in 5 M urea + 2 M thiourea + 2 % (w/v) CHAPS + 2 % (w/v) SB 3-10 + 40 mM Tris + 0.2 % (w/v) BioLyte 3-10 (sequential extraction reagent 3, Bio-Rad) at room temperature and vortexing for 2 minutes.

Picoplankton communities were collected at three stations along the middle axis of the Chesapeake Bay on 7 June 2003 aboard the R/V *Cape Henlopen* (Fig. 6-1). The stations 858, 804 and 707 represent the upper, middle and lower Bay, respectively. At each station, 0.2 g of chloramphenicol (Fisher Scientific, NJ) and 2 ml Protease inhibitor cocktail II (CalBiochem, CA) were added to 20 l of surface water (1 m below) to stop protein synthesis and inhibit activities of proteases. Samples were pre-filtered through 3-µm-pore-size polycarbonate filters (142-mm diameter; Millipore, Bedford, MA) to remove large particles and eukaryotes. The filter was replaced every 5 liters. Microbial cells in the filtrate were concentrated to a final volume of 150 ml using a tangential-flow ultrafiltration (30,000MW cutoff) as described elsewhere (Chen et al. 1996). Duplicate water samples were collected at station 804. Microbial cells in the retentate were pelleted using GS-15R centrifuge (Beckman, Fullerton, CA) at 13,000 × g, 4ºC for 10 minutes. The collected cells were rinsed with TS washing buffer (Tris-HCl 10 mM, Sucrose 250 mM, pH 7.6) and resuspended with 0.5 ml of extraction buffer. The extraction buffer consisted of 0.01 M Tris-HCl, pH 7.4, 1 mM EDTA, 7 M urea and 2 M thiourea, 10% (v/v) glycerol, 2 % CHAPS, 0.2 % amphylotes, 0.002 M Tributyl phosphine (TBP), DNase (0.1 mg/ml), RNase (0.025 mg/ml) and proteinase inhibitor cocktail (CalBiochem, CA).

TBP, DNase, RNase and proteinase inhibitor cocktail were freshly added to the buffer prior to applying to samples. Cells were stored frozen until further processing.

To estimate the recovery efficiency of ultrafiltration, bacterial cells were counted before and after ultrafiltration. Bacterial cells were stained with SYBR Gold (Molecular Probes Inc., Eugene, Oreg.) following the protocol described previously

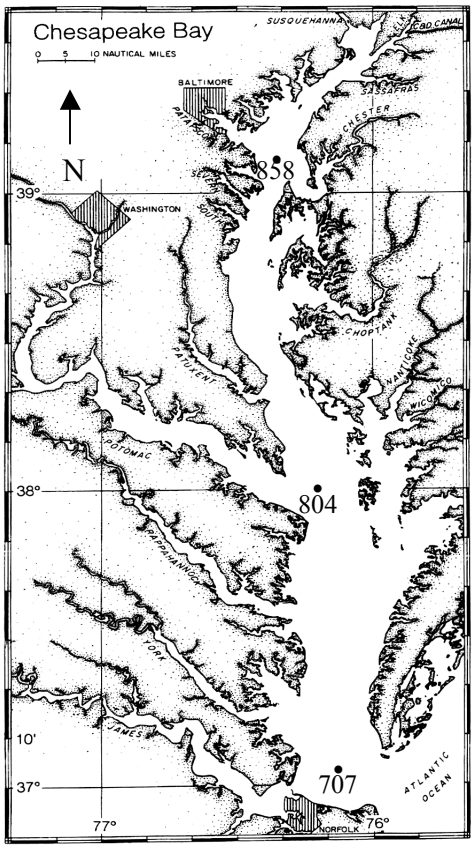


Fig. 6-1. Metaproteome sampling stations on the Chesapeake Bay.

(Chen et al. 2001). Bacterial cells were enumerated under blue excitation (485 nm) on a Zeiss Axioplan epifluorescence microscope (Zeiss) using 63× Antiflex Neoflua oil objective lens. At least 200 bacterial cells per sample were counted.

For 1D-PAGE, proteins from natural microbial communities and cultured bacteria were extracted using lysis buffer (50 mM Tris-HCl, 2% SDS, 10% v/v Glycerol, 0.1 M DTT, 0.01% Bromophenol Blue, pH 6.8). Cells suspended in buffer were heated in a boiling water bath for 2 minutes followed by centrifugation (10,000 x *g*, 4 ºC for 3 min). The supernatant was collected and 20 µg protein for each was loaded onto polyacrylamide gels. Silver staining was applied to 1D-PAGE gels.

For 2D-PAGE samples, cell suspensions were passed through a French Pressure cell (SLM Aminco) at 20,000 lb/in^2 twice and then incubated on ice for 20 minutes. During the ice incubation, samples were vortexed for 15 sec every 5 minutes. Large cellular debris was removed by centrifugation (10,000 × *g*, 4 ºC for 5 min). Proteins in the supernatant were precipitated with trichloracetic acid and resuspended in extraction buffer. Protein concentration of the sample was determined using the RC DC protein assay kit (Bio-Rad, Hercules, CA). Extracted proteins were stored at -80 ºC.

The first dimension separation of proteins was carried out in the immobilized pH gradient (IPG) strips (11cm, pH 3-10 or 4-7) on a Bio-Rad Protean IEF Cell system (Bio-Rad, Hercules, CA). Each 2D-PAGE was conducted using 100 µg of total protein. The IEF program was: 250V for 20 min followed with a linear ramp to 8000V for 2.5 hr, and 8000V for a total 40,000 V-hr with a rapid ramp. After the first dimension, the IEF strips were equilibrated in freshly made Buffer 1 (6 M Urea, 2% SDS, 0.05 M Tris/HCl pH 8.8, 50% Glycerol) and Buffer 2 (6 M Urea, 2% SDS, 0.375 M Tris/HCl pH 8.8, 20% Glycerol and 0.5 g Iodoacetamide) (Bio-Rad, Hercules, Calif), respectively.

The second dimension of 2D-PAGE were performed using 8-16% gradient precast polyacrylamide gels (Bio-Rad, Hercules, CA) following the manufacturer's instructions. The gels were stained with SYPRO Ruby (Bio-Rad, Hercules, CA) after electrophoresis and scanned using a Typhoon 9410 fluorescent Imager (Amersham, NJ) with 488nm excitation and emission filter 610 BP30.

Images were analyzed and quantitatively compared using the Z3 proteomics software package (Compugen, Israel). Gel images were compared in multiple gel mode using the total density in gel method for spot quantification. All gels were subjected to the same spot detection parameters followed by automated matching. Pairwise comparisons of gels were inspected and matches edited manually to eliminate poor quality or low intensity matches. When automatic matching failed, the number of matched and unmatched spots was estimated by manual examination of overlaid 2D SDS-PAGE images.

Protein spots were manually excised from gels using Pasteur pipettes and digested as described by Mann et al. (1996). Tryptic peptides were analyzed both via MALDI-TOF and LC-MS/MS. MALDI spectra were acquired on a Bruker (Billerica, MA) Biflex III MALDI mass spectrometer operating in reflectron mode with delayed extraction. External calibration was performed using Calibration Mixture 2 from the Sequazyme Peptide Mass Standards Kit (Applied Biosystems, Foster City, CA). LC-MS/MS was performed on a Micromass (Beverly, MA) Q-TOF Ultima API-US coupled to a Micromass capLC. Tryptic digests were separated using both a C18 trapping column for washing and concentrating (LC Packings (Sunnyvale, CA) 300 μm x 5 mm C18) and a C18 analytical column for enhanced separation (LC Packings

180 µm x 15 cm C18). The solvent system consisted of 95% 0.1% formic Acid, 5% acetonitrile for the aqueous phase and 95% acetonitrile, 5% 0.1% formic Acid for the organic phase. A 60/60 gradient (to 60% organic in 60 min) running at 1 µl/min was employed with most peptides eluting by ~30% organic. The LC eluent was electrosprayed directly into the Q-TOF using the nanosprayer source. Data dependent scanning was used with both MS and MS/MS spectra being acquired during an LC run. Spectra were processed and deconvoluted using programs found with the Micromass operating system, MassLynx v. 3.5.

MALDI-TOF peak lists were searched against protein sequence databases using the Matrix Science Mascot web interface (http://www.matrixscience.com/search_form_select.html). Deconvoluted MS/MS spectra were analyzed using a demonstration version of PeaksStudio 3.0 software (Bioinformatics Solutions Inc., Canada) for *de novo* sequence prediction. All sequences for each protein spot were used as queries in MS-BLAST searches as described by Shevchenko et al. (Shevchenko et al. 2001) via the MS-BLAST web interface (http://dove.embl-heidelberg.de/Blast2/msblast.html).

Results

Epifluorescence microscopic counts showed that concentrated microbial communities mainly contained free-living bacteria (~ 95%). The recovery efficiency of bacterial cells using the tangential flow ultrafiltration system was 75±5% (data not shown). With the average concentration of 2.5×10^{6} cells ml^{-1} in the starting water samples, the density of microbial cells in the ultrafiltration retentate was about 2.5×10^{8} cells ml^{-1}. Thus, about 3.75×10^{10} cells were analyzed in each sample.

Extracts typically contained between 140 and 192 μg of protein giving a value range of 3.7×10^{-15} to 5.1×10^{-15} g protein cell^{-1}. This value is significantly lower than that

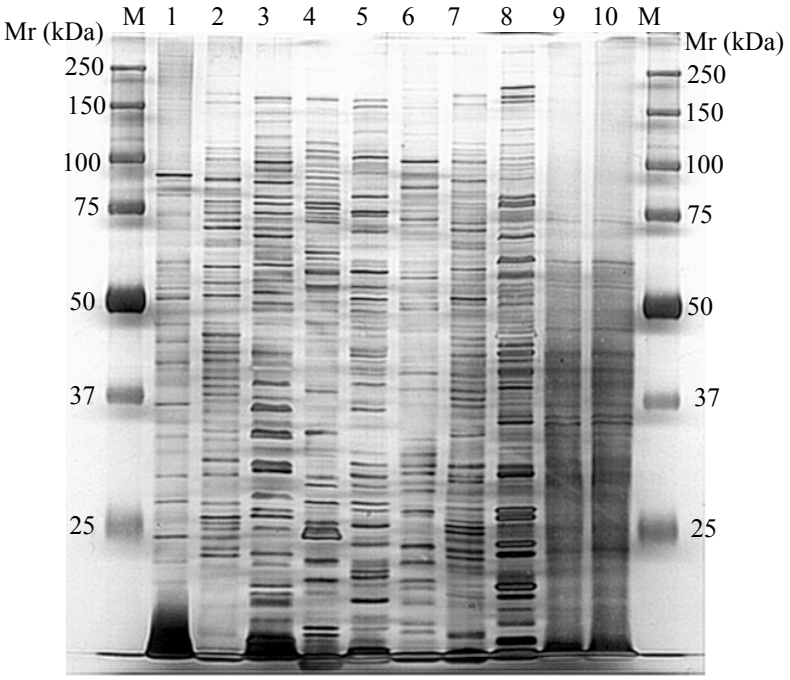


Fig. 6-2. 1D-PAGE patterns of total proteins obtained from eight different bacterial isolates and environmental water samples from Baltimore Inner Harbor. M, Marker; Mr, molecular weight; Lanes 1 – 8 correspond to *Vibrio vulnificus*, Marine *Bacillus* sp., *Marinomonas* sp., *Psychrobacter pacificens*, *Pseudomonas* sp., *Pseudoalteromonas* sp., *Shewanella* sp., and *Hahella* sp.. Lanes 9 and 10 are duplicated environmental microbial communities. For each lane, 20 μg of protein is loaded and the gel is stained by silver staining.

determined for cultured strains in this study and in general for marine bacteria (60-330×10^{-15} g protein cell^{-1}, Zubkov et al. 1999). It remains to be determined whether

this discrepancy indicates that the extraction protocol needs further optimization or is a fundamental property of microbial cells in environmental samples.

Individual proteins from cultivated marine bacteria were well resolved by 1D-PAGE and produced distinct patterns when 8 Chesapeake Bay bacterial isolates were compared (Fig. 6-2). The observed molecular masses ranged from ~10 to 250 kDa (Fig. 6-2, lanes 1-8) whereas proteins from microbial community samples were < 80 kDa (Fig. 6-2, lanes 9 and 10). Overall resolution was much poorer in community samples as evidenced by less sharply defined bands in these samples. This blurring effect was also noted in a very simple mixed microbial community described below and was not dependent on sampling manipulations (data not shown).

Artificial community consisting of *Chlorobium tepidum* strain WT2321, *Escherichia coli* strain JM109 and an uncharacterized strain of *Pseudomonas fluorescens* was analyzed by 2D-PAGE. Preliminary experiments indicated that a 300 ml sample containing 1 x 10^7 cells per ml of the community could be successfully analyzed by 2D-PAGE. Analysis by 1D-PAGE afforded greater sensitivity, ~1 x 10^4 cells per ml, but resolution of individual bands was poor as noted above. Protein assays on samples of the community before dilution and recovery and after indicated that the metaproteomic sample preparation recovered ~ 30% of the total microbial protein present in the original community sample.

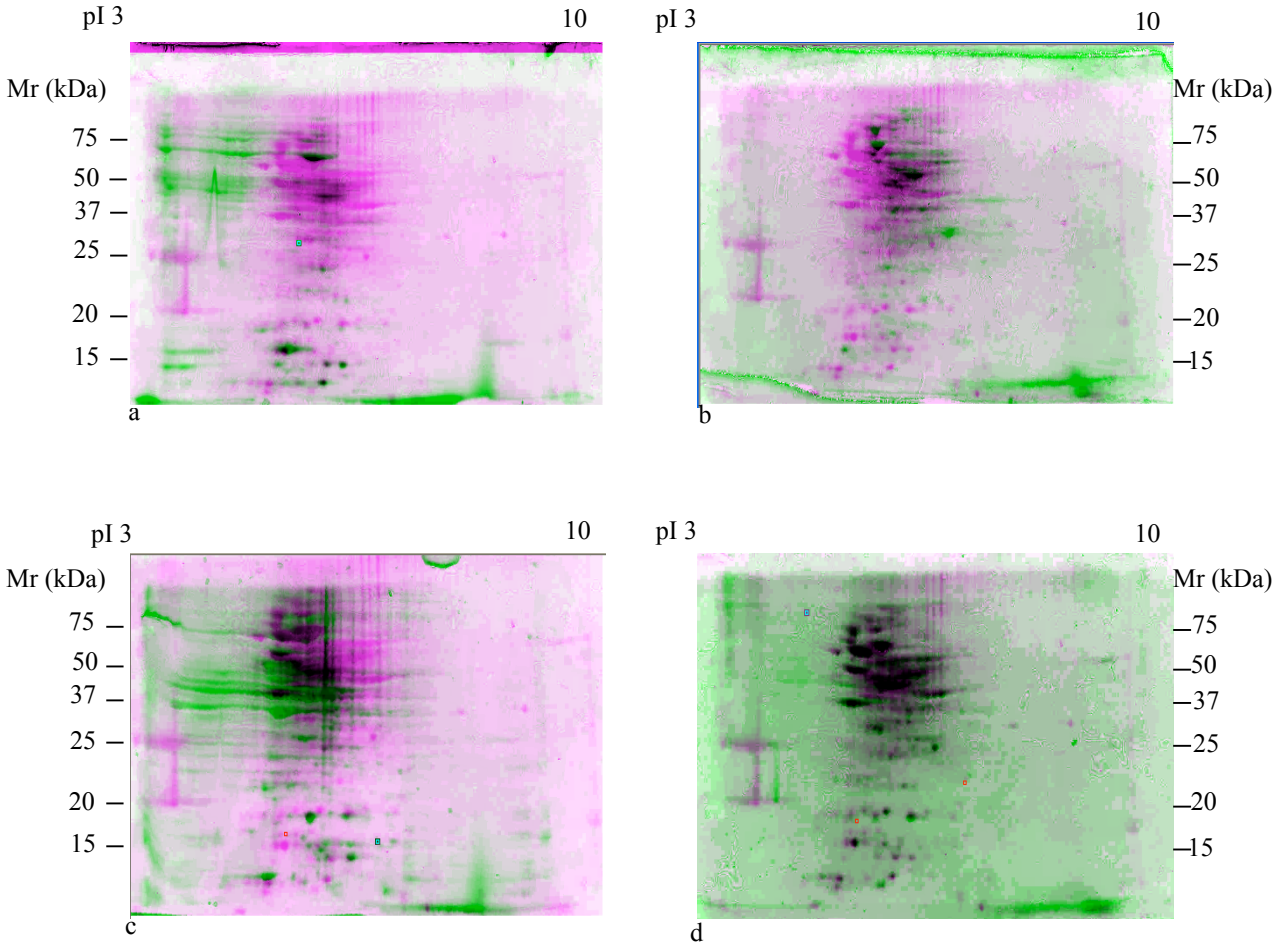


Fig. 6-3. The harvesting protocol for microbial communities does not bias against different types of bacteria. Proteomes of *Chlorobium tepidum* (a), *Escherichia coli* (b) and *Pseudomonas fluorescens* (c) and the metaproteomes of an artificially constructed community containing all three organisms (d) were overlain and compared to the metaproteomes of the artificial community after dilution and recovery using Compugen Z3 software. Green or pink colored protein spots are unmatched. Gray or black spots are matched. Total 100 μg proteins are loaded on each polyacrylamide gel and the gels are stained by SYPRO Ruby. pI, isoelectric point; Mr, molecular weight.

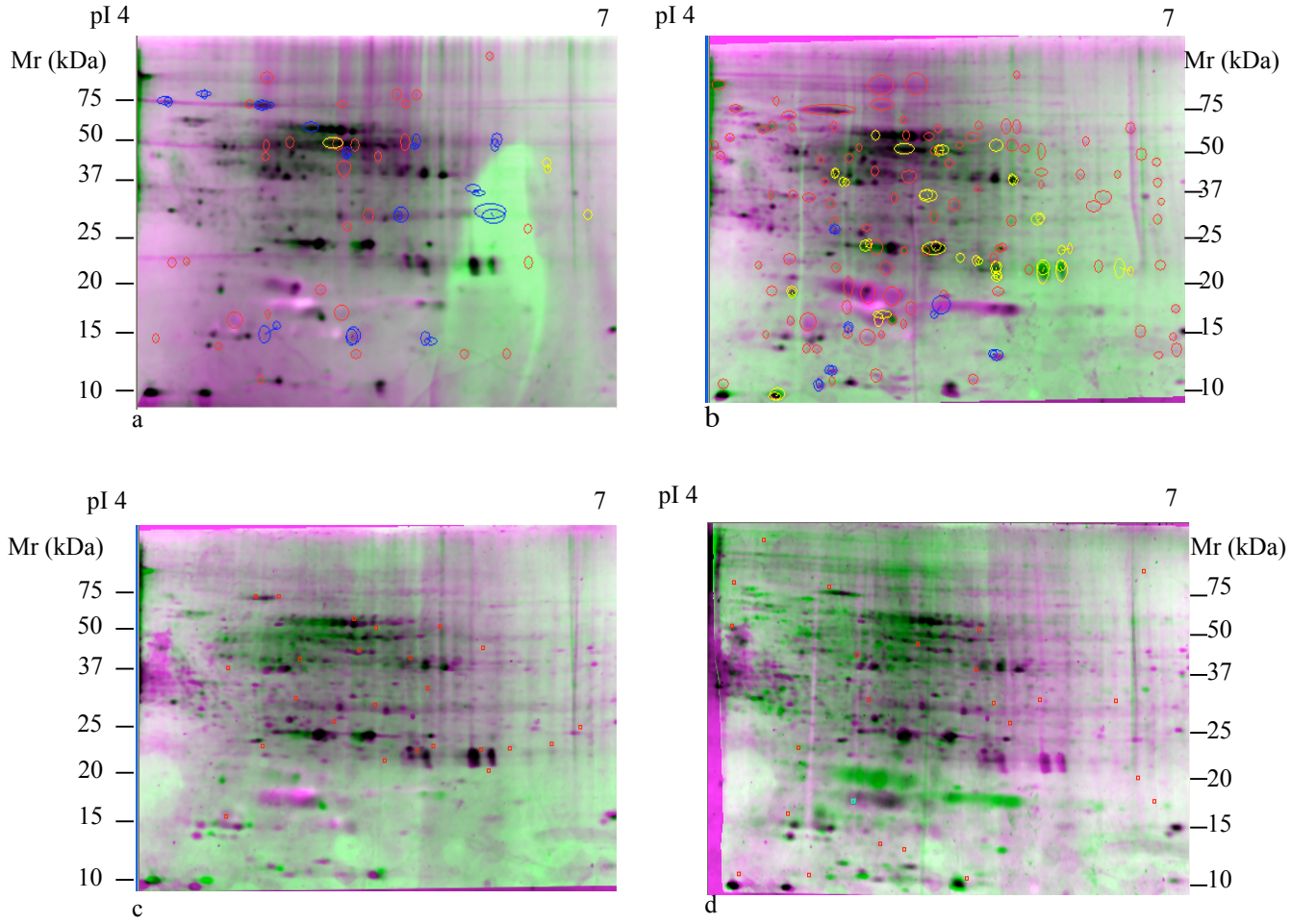


Fig. 6-4. Comparisons of Chesapeake Bay metaproteomes. (a) Independent samples from Station 804, 804a and 804b; (b) Station 804a vs. Station 707; (c) Station 804a vs. Station 858; (d) Station 707 vs. 858. Image overlays were constructed with Compugen Z3 software. Spots circled in red are unmatched, those in yellow and blue are differentially expressed at a level of $\geq$ 3-fold between images. No unmatched or differential spots are shown in c and d because software based matching of these images failed. Red marks in panels c and d are alignment points used to produce the pictured overlay. Quantitative results of matching are reported in Table 1. A total of 100 μg protein is loaded on each polyacrylamide gel and the gels are stained by SYPRO Ruby. pI, isoelectric point; Mr, molecular weight.

Typical results from a 2D-PAGE experiment are shown in Fig. 6-3. The overlays indicate that 2D-PAGE patterns from single strains of community members only match a fraction of protein spots present in the mock metaproteome sample (Fig. 6-3, a-d). This is qualitatively observed as a large number of green or pink protein spots in the overlay views showing unmatched protein spots. Each individual strain is expected to contribute only one third of the protein content of the community. In contrast, when a sample of the community prior to dilution and recovery is compared to a mock metaproteome that had been subjected to sample handling protocols, almost perfect matching of the samples is seen as evidenced by the large proportion of dark grey to black spots (Fig. 6-3, d) when these images are overlain. Thus, no individual member of the community, which covers the range of cell sizes in the environmental samples, is selectively excluded by the sampling protocol.

In this study, in order to optimize the protein extraction of aquatic microbial communities, different protocols that varied all steps in protein extraction and purification were tested including (i) sample collection (filtration on membrane filter, tangential flow concentration with centrifugation); (ii) washing buffer to remove ambient salts and polysaccharides; (iii) extraction buffer (standard lysis buffer, SDS-PAGE buffer, urea-thiourea-CHAPS buffer); (iv) reducing agent (dithiothreitol (DTT) vs. tributyl phosphine (TBP));(v) cell lysis method (freeze-thaw, French pressure cell); (vi) protein precipitation (acetone vs. TCA); (vii) IPG strip range (pH 3-10 vs. pH 4-7); and (viii) staining method (Commassie blue, silver, SYPRO Ruby). From these trials, the following protocol emerged: (i) tangential flow concentration with centrifugation; (ii) TS washing buffer (Tris 10mM, Sucrose 250mM); (iii) urea-

thiourea-CHAPS lysis buffer with TBP; (iv) lysis via French pressure cell; (v) TCA precipitation; (vi) First dimension pH 4-7 IPG strip; (vii) SYPRO Ruby staining. However, given the indigenous characteristics among diverse microbial communities, extraction of metaproteomes may vary by site, time and experiment as well.

Metaproteome images from different Chesapeake Bay stations in the upper (station 858), middle (station 804, replicates a and b) and lower Bay (station 707) were compared (Fig. 6-4, a-d). A number of protein spots were shared by all samples. Some of these are proteins present in RNase, DNase and protease inhibitor cocktail in the extraction buffer (data not shown), but a number of proteins appear to be common in all samples examined. These are black to dark grey spots in the image overlays (Fig. 6-4, a-d). A first level of quantitative comparison determined the specific numbers of protein spots shared between samples (Table 6-1). The total number of spots compared for each sample is relatively low as the analysis was restricted to spots with sufficient quality and intensity to permit subsequent attempts at protein identification. As expected, replicate metaproteome images from the middle Bay are more similar to one another than the metaproteomes of other stations, sharing ~92 % of all detected spots. Furthermore, the lower and middle Bay metaproteomes are significantly more similar to one another than either is to the upper Bay metaproteomes with ~70 % of all detected spots in common. The upper Bay metaproteomes only shared about ~30 % of detected spots with either the middle or lower Bay metaproteomes.

Relative spot intensity was extracted from comparisons of middle Bay to middle Bay and middle Bay with lower Bay metaproteome images. This was not

possible with the upper Bay sample as manual matching was employed due to the low level of similarity between samples. Again, as expected, the number of differentially expressed proteins (≥ 3-fold change in matched spot intensity) was nearly twice as large when comparing middle Bay to a lower Bay metaproteomes as when comparing the replicated middle Bay samples (Table 6-1). These results indicate that both qualitative and highly quantitative comparisons between sites and between time series samples at the same site will be possible using the approaches developed in this study.

Table 6-1. Quantitative comparison of Chesapeake Bay metaproteomes

Samples compared	spots[a]	unmatched[a]	differential[a,b]
804a	207	7	3
vs. 804b	189	26	13
	396	33 (8.3 %)	16 (4.0 %)
804a	207	37	23
vs. 707	198	86	6
	405	123 (30.3 %)	29 (7.1 %)
804a	207	156[c]	--[d]
vs. 858	155	104[c]	--
	362	160 (71.8 %)	--
707	198	142[b]	--
vs. 858	155	99[b]	--
	353	241 (68.3 %)	--

[a] Spots from first gel, second gel and the sum are listed. Numbers in parentheses show the percentage of the total.
[b] Matched spots that are ≥ 3-fold more intense than the comparative image
[c] Estimated by manual comparison of detected spots. Software was unable to match images
[d] No differential comparison possible as software based matching failed

A total of 41 protein spots were excised from a number of 2-D gels reflecting various molecular weights, charges and relative abundance. Following MALDI-TOF

MS, seven spots failed to yield interpretable MS profiles, while the remaining 34 proteins exhibited clear and distinct MS peaks. Database searches using the MASCOT search engine with varying parameter settings (peptide mass tolerance from 0.5 to 3 Da, missed cleavages from 1 up to 5) produced no significant matches for these 34 proteins. Subsequent publications from other laboratories and our own

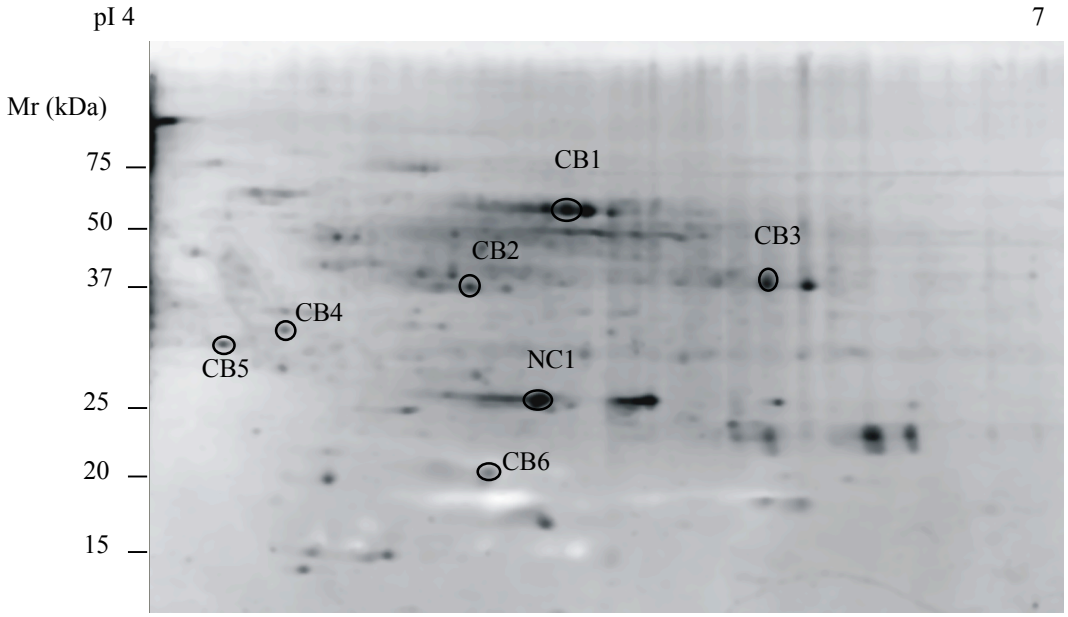


Fig. 6-5. Proteins selected for identification from middle Chesapeake Bay (station 804). Total 100 μg protein are loaded on polyacrylamide gel and the gel is stained by SYPRO Ruby.CB1-CB6 samples are common to Chesapeake Bay stations while NC1 is found on negative control gels containing DNase, RNase and protease inhibitors. Results of protein identification are reported in Table 2 and 3. pI, isoelectric point; Mr, molecular weight.

simulations using known protein sequences (Liska and Shevchenko 2003; Habermann et al. 2004; Hanson, unpublished data) suggest that greater than 97 % amino acid sequence identity is required to provide a positive match when searching with MALDI-TOF MS data.

Seven individual proteins (Fig. 6-5) isolated from middle Chesapeake Bay (station 804) metaproteome samples were further analyzed by both MALDI-TOF MS and LC-MS/MS sequencing coupled to MS-BLAST searching (Table 6-2). MALDI-TOF MS failed to provide identification for any of these samples, similar to the samples described above. LC-MS/MS based searches provided tentative identities for three Chesapeake Bay metaproteome samples. These were identified as homologues of hypothetical proteins annotated in the recently reported Sargasso Sea metagenome (Venter et al. 2004). Information on potential functions of these proteins was obtained by downloading the full-length proteins from the Sargasso Sea database and searching them against known databases by BLASTP (Table 6-3). The Sargasso Sea metagenome hypothetical protein corresponding to sample CB1 is not significantly similar to any known proteins in sequence databases. Sample CB3 may correspond to subunit 7 of the NADH:ubiquinone oxidoreductase (complex I) while sample CB6 is similar to a family of predicted aminopeptidase with unspecified functional significance. The tandem mass spectra of samples CB2, CB3 and CB5 had no match with any known proteins or hit keratin and bovine serum albumin that possibly came from background.

Table 6-2. Identification of proteins from Chesapeake Bay station 804 metaproteomes (Fig. 6-5)

Sample	pI	MW	MALDI ID?[a]	MS/MS ID?[b]	Peptides Matched	Score[c]	Accession
NC1	5.1	29 kDa	No	No	-	-	-
CB1	5.3	60 kDa	No	Sargasso sea metagenome	2	110	EAH98995.1
CB2	4.9	40 kDa	No	Bovine serum albumin	2	138	P02769
CB3	5.7	42 kDa	No	Sargasso sea metagenome	3	116	EAH45127.1
CB4	4.4	35 kDa	No	Keratin	2	117	Q9DCV7
CB5	4.2	33 kDa	No	No	-	-	-
CB6	5.0	20 kDa	No	Sargasso sea metagenome	2	88	EAC65279.1

[a] MASCOT search as described in Materials and Methods
[b] MS-BLAST search as described in Materials and Methods
[c] For a description of scoring, see reference Shevchenko et al. 2001

Table 6-3. BLASTP analysis of Sargasso Sea metagenome hits

Sample	Accession	Best hit	E-value	Organism	Accession
CB1	EAH98995.1	Hypothetical protein	0.47	*Plasmodium berghei*	CAI00437
CB3	EAH45127.1	NADH:UQ oxidoreductase (49 kDa, subunit 7)	1×10^{-63}	*Cytophaga hutchinsonii*	ZP_00309190
CB6	EAC65279.1	Predicted aminopeptidase	2×10^{-16}	*Novosphingobium aromaticivorans*	ZP_00305215

Discussion

In this study, we deliberately focused on exploring the proteome profiles from bacterioplankton communities between 0.2 and 3.0 microns in size by the choice of prefiltration and ultrafiltration cut-off sizes. Although the epifluorescence microscopy observation confirmed that the major components are bacterioplankton (~95%), small numbers of eukaryotic microbes were possibly included. These likely did not affect the overall protein profiles observed as analyses were restricted to abundant proteins, which would give the best chance for positive identification.

Metaproteomic approaches have thus far only been applied to laboratory scale bioreactors with a specialized community selected for phosphate removal (Wilmes and Bond 2004) and a low-complexity natural microbial biofilm (Ram et al. 2005). Extending this approach to complex environmental samples was not trivial. Initial studies comparing isolated strains, artificial communities and natural community samples by 1D-PAGE indicated that more resolving power was needed to deal with even simplified communities (data not shown). Thus, a metaproteomic approach utilizing 2D-PAGE and MS based protein identification was adopted. The experimental protocol outlined in this study was designed to avoid metaproteome changes arising from bias in the sample collection or handling. This was tested using artificial constructed bacterial assemblage containing 3 different species with varied cell sizes and we found no significant biases.

The protocol was also field tested by comparing replicated samples from the middle Chesapeake Bay to each other and comparing a range of samples from upper, middle and lower Chesapeake Bay stations. The replicated samples shared more than

~92 % of proteins indicating that the metaproteomic approach applied in this study was robust. Furthermore, significant differences were noted when the middle Bay metaproteomes was compared with lower Bay and upper Bay metaproteomes with only 70 % and 30 % of protein spots in common. This pattern can be likely and partially explained by the difference among the population structures of these samples. Genetic fingerprints indicated that upper Bay bacterioplankton community was different from the middle and lower Bay (Fig. 6-6). Clustering analysis based on presence/absence of DGGE bands showed that the similarity between middle Bay to lower Bay was 64% while the upper Bay only shared 46% similarity to both of middle Bay and lower Bay. Finally, relative spot abundance was also much more tightly correlated when the replicated middle Bay samples were compared to each other than when they were compared to the lower Bay sample. These results demonstrate the approach outlined here is sufficiently sensitive to detect both coarse (shared spots) and fine (relative spot abundance) quantitative differences between samples, even when relatively low numbers of spots are included in the analysis. This is critical for any comparative approach.

This study, in addition to others, indicates that protein identification is the major challenge for metaproteomics (Wilmes and bond 2004; Ram et al. 2005; Liska and Shevchenko 2003; Habermann et al. 2004). Although distinct mass spectra from 34 protein spots were obtained by MALDI-TOF MS, no significant matches were found in sequence databases. MALDI-TOF generally requires at least 97 % amino acid sequence identity between query and target to find a significant match (Shevchenko et al. 2003; Hanson unpublished). It seems unlikely that many proteins

in environmental samples will share this level of identity with proteins in sequence databases derived from cultured organisms. Post-translational modifications of proteins also account for the difficulty in the identifications. Thus, MALDI-TOF MS is unlikely to be useful for metaproteomic approaches.

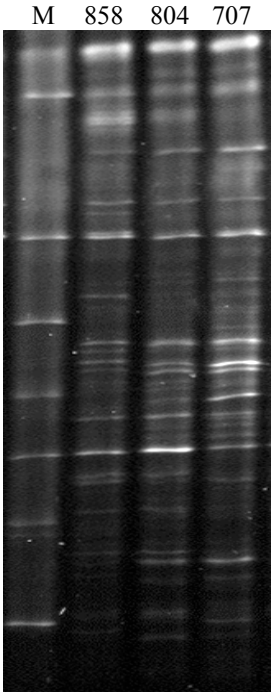


Fig. 6-6. DGGE fingerprints of bacterioplankton communities in the Chesapeake Bay. 858, 804 and 707 are sampling stations. M: marker.

In contrast, LC-MS/MS or N-terminal sequencing coupled to MS-BLAST searching is able to provide tentative identification for metaproteomes. However, the abundance of most proteins is too low to be identified through the venue of N-terminal sequencing. In the community proteomic analysis of a natural acid mine drainage microbial biofilm, the proteins could be identified by MS and assigned to

five most abundant microbes because of the availability of metagenomic data. But the relative high likelihood of false-positive protein identification requires matching of two or more peptides per protein for confident detection (Ram et al. 2005). Therefore, caution is required for interpretation of the data. In this study, three Chesapeake Bay metaproteome samples matched different hypothetical proteins annotated in the Sargasso Sea metagenome (Venter et al. 2004). This result strongly supports a marine origin for these sequences as would be expected for a large number of proteins in the Chesapeake Bay, particularly in lower and middle Bay samples where there is significant salinity. Even with tentative identities, extending that identity to function must be done with some care. The Sargasso Sea metagenome hypothetical protein corresponding to sample CB1 is not significantly similar to any known proteins in sequence databases giving no clues to its function. Sample CB3 may correspond to subunit 7 of the NADH:ubiquinone oxidoreductase or complex I (Table 6-3). Complex I is a key component of most membrane bound electron transport chains that is responsible for the transfer of electrons from cytoplasmic NADH pools to the membrane bound quinone pool coupled to proton motive force generation. Subunit 7 is a peripheral membrane protein of the quinone reduction core of complex I (Zickermann et al. 2003). The organism containing the closest match is *Cytophaga hutchinsonii*, a member of the *Bacteroidetes* assemblage of organisms, which is a substantial fraction of many marine communities (Cottrell and Kirchman 2000b). A current study on population structure of Chesapeake Bay bacterioplankton showed that *Bacteroidetes* group accounts for ~10% of total community in summer time (Chapter 3).

Sample CB6 is similar to a family of predicted aminopeptidases with unspecified functional significance. The closest matching protein is from *Novosphingobium aromaticivorans*. While *N. aromaticivorans* is normally considered terrestrial, other *Novosphingobium* and related *Sphingobium* and *Sphingopyxis* strains are widely distributed. As an important component of the *Alphaproteobacteria*, these groups can be detected in and isolated from marine and estuarine environments (Ostrowski et al. 2004; Sohn et al. 2004; Chapter 3 & 4). This identification along with that of CB3 support an aquatic bacterial origin for these proteins that is consistent with their presence in the Chesapeake Bay.

Unanswered questions remain regarding the applicability of metaproteomics to natural communities. These include the following: Does a focused protein spot on a 2D SDS-PAGE gel from an environmental sample contain one protein or multiple proteins? What type of information is required to infer identity of spots between different samples? What is the sensitivity of metaproteomics to changes in community composition and the physiological status of community members? How can functional inferences provided by metaproteomics be further tested? Will the approach outlined here be applicable to other systems such as soils, sediments, and extreme environments? Clearly, much more work and complementary approaches need to be applied to these problems.

Conclusion

To our knowledge, this study represents the first application of a metaproteomic approach to a high-complexity aquatic microbial community. The main goals of this study were to develop a method capable of collecting planktonic

microbial proteins in quantities suitable for analysis by 2D-PAGE. This was accomplished and attempts were made to identify a subset of these proteins. These attempts reinforced the notion that sequence based methods (LC-MS/MS) will be required to make any headway in protein identification in natural systems. Future studies will identify a much larger number of proteins from Chesapeake Bay microbial communities to address the questions raised above and provide insights into microbial community dynamics and function.

Chapter 7: Summary and future perspectives

This dissertation was devoted to study genetic diversity, population dynamics, and *in situ* functions of Chesapeake Bay bacterial community. Four major results can be summarized from the research described in this dissertation. First, in detailed studies of population structure of Chesapeake Bay bacterial communities unique Chesapeake/estuarine phylotypes were discovered (Chapter 3 and 4). Second, dramatic seasonal variations and repeatable annual patterns occurred in the Bay and in the Inner Harbor (Chapter 2 and 5). Third, seasonal succession of Chesapeake bacterioplankton correlated most significantly with water temperature and phytoplankton biomass. Finally, we proved the concept that community-based proteomics can be applied to explore biological and ecological functions of marine bacterioplankton (Chapter 6). Several interesting observations resulting from this dissertation work deserve further investigation, and addressing these questions could lead to a better understanding of the microbial ecology of estuarine ecosystem.

As an interface between freshwater and Atlantic ocean, the Chesapeake Bay estuary contains strong environmental gradients that provide diverse niches for bacterioplankton living in the Bay. Our studies showed that Chesapeake ecosystem harbors microbes from both freshwater and oceanic origins. Meanwhile, many unique or novel bacterial groups like those in SAR11, *Roseobacter*, and SAR86 clades were found in the Chesapeake Bay, suggesting that certain endemic populations may adapt to these specific niches. Although the composition of estuarine bacterial community has not yet been well defined, unique estuarine populations have also been reported in

other estuaries (Crump et al. 1999 & 2004; Selje and Simon 2003) indicating that estuarine bacterioplankton are not only composed of advected populations from adjacent sources including rivers, soil and oceans. The residence time of water mass in the Chesapeake Bay, in the order of months is certainly much longer than the doubling time of bacteria in the Bay (Nixon et al. 1996) and permits the development of stable local bacterial populations in the Chesapeake Bay. It will be interesting to isolate some of these Bay specific bacterial strains. Further characterization of these bacteria could provide new insight into physiological adaptation, niche partitioning, and community organization of Chesapeake Bay bacterioplankton.

Strong salinity gradients in the estuarine ecosystems might influence the composition of bacterioplankton communities (Giovannoni and Rappé 2000; Crump et al. 2004). Previous studies have reported the significant spatial variations of bacterial community structure along the salinity gradients of estuaries (Crump et al. 1999 and 2004; Bouvier and del Giorgio 2002; Sekiguchi et al. 2002; Hewson and Fuhrman 2004; Henriques et al. 2004) and these changes are likely related to the osmotic stress which negatively affects the cell survival (Barcina et al. 1997). Distinct bacterial assemblages from low (freshwater), intermediate (estuary), and high (ocean) salinity regions were observed in estuaries (Crump et al. 1999; Bouvier and del Giorgio 2002; Sekiguchi et al. 2002; Hewson and Fuhrman 2004; Henriques et al. 2004). In general, *Betaproteobacteria* are dominant in freshwater, while *Alpha-* and *Gammaproteobacteria* dominate in marine waters. Furthermore, bacterial communities in freshwater and high salinity regions exhibited more pronounced spatial variations than middle estuary (Henriques et al. 2006). Clustering analysis of

bacterial community indicated that 5 ppt appears to be a practical salinity cutoff for separating freshwater and estuarine bacterial communities (Henriques et al. 2006). However, the salinity boundary for separating estuarine and marine pelagic bacterial communities has not been identified. In our studies, water samples were collected from stations in the Chesapeake Bay where salinities varied from 5 to 27 ppt and perhaps most of the estuarine adapted bacteria can tolerate such a salinity range. Unfortunately, our sampling did not cover freshwater and oceanic waters. Therefore, no significant spatial variations of bacterial community were observed in this thesis. It will be interesting to extend sample collection to the major rivers and the Atlantic offshore water in future studies. A systematic comparison of bacterial community structure from freshwater, estuarine, coastal and oceanic waters could provide a better understanding on the impact of salinity on Chesapeake Bay bacterial populations. Salt tolerance and salinity effects could also be studied by using representative bacterial strains isolated from the Bay.

Results from clone library analysis, DGGE, and LH-PCR showed seasonal succession and annually repeatable patterns of the Chesapeake Bay bacterioplankton. These findings have many significant ecological implications. First, distinct distribution patterns of "warm-species" and "cold-species" may shed light on understanding the ecology of universally distributed species. Our culture collections of heterotrophic bacteria from Baltimore Inner Harbor indicated that many "cold-adapted species" were closely related to the Arctic or Antarctic bacterial species (Chapter 2). It is not likely to that these "cold-species" that annually repopulate the Inner Harbor have Polar origins . At the mean time, we know very little about their

survival strategy of bacterial groups that become undetectable by PCR. It will be important to understand how the switch between cold and warm species is regulated in a temperate estuary like the Chesapeake Bay.

Several Chesapeake *Roseobacter* groups were retrieved in cold season samples. Cold-adaptation in this group could be related to the wide distribution of cold-shock gene homologues in several marine *Roseobacter* genomes (Moran, et al. 2004; www.jgi.doe.gov; Belas, personal communication).. At present, mechanisms regulating the seasonal distribution of Chesapeake Bay roseobacters are still not clear, and warrant further studies. In contrast, marine *Synechococcus* and SAR11 related species were detectable only in the warm season, which may possibly relate with high water temperature or intensity of sunlight. Because temperature is an important factor that affects bacterial activity and community structure (Chapter 5; Heidelberg et al. 2002; Johnson et al. 2006), long-term monitoring bacterial populations should be integrated as part of ecological assessment of global warming. Currently, many Chesapeake *Synechococcus* spp. have been isolated and cultivated in the laboratory. However, no Chesapeake SAR11 group has yet been isolated. Cultivation of these bacteria will allow us to study their physiology, which could help us better understand their dominance in particular times.

Environmental factors (e.g. temperature, light and nutrients) that controlling the distribution of bacterial populations are not well understood. In Chapter 5, temporal variations of bacterial community during 2002-2004 were best explained by changes of Chl *a*, water temperature, nutrients availability and abundance of viral-like particle. Close association between bacterial dynamics and environmental parameters

implies that complex patterns of physical, chemical and biological variations interactively drive the distribution and diversification of bacteria in the water column. In a recent investigation along a meridional transect in the Atlantic Ocean (AMT), distributions of six different ecotypes of *Prochlorococcus* showed significant correlation with water temperature, light, *Synechococcus* (a potential source competitor) and complex relationships with nutrients (Johnson et al. 2006). The statistics used in that study and in this dissertation, showed varying degree of success explaining patterns of community structure and demonstrated that yet additional factors may also affect population dynamics in the environment. For instance, both studies lack protist-grazing rates, which may play an important role on determining the bacterial seasonal patterns (González et al. 1990; Hahn and Höfle 2001; Jürgens and Matz 2002). To better understand what environmental factors drive microbial community dynamics, further studies could be carried out using cultivated bacteria or mesocosm/ microcosm experiments.

Monitoring of microbial community over multiple years offer a good opportunity to study the ecology and interactions among diverse microbial groups in the Chesapeake Bay. Besides the data from two and half year samples included in this thesis, currently more than four-year samples of phytoplankton, bacterioplankton and virioplankton have been collected for the Microbial Observatory project in the Chesapeake Bay. Multiple year data analyses will help to corroborate or refute the findings in this thesis including, for example, the repeatable patterns of bacterioplankton community composition in the Chesapeake Bay. Summer-fall communities were more stable and are more likely predictable from environmental

factors. In contrast, winter communities varied in year 2003 and 2004. In a recent study, Fuhrman et al. (2006) have demonstrated annual recurrence of bacterial communities at the California coast that was predictable from environmental conditions. Compared to this coastal study site, the Chesapeake Bay experiences more dynamic environmental gradients and local events such as algal blooms may result in the less predictable patterns in winter. Nevertheless, repeatable bacterial community patterns have important ecological implications especially in dynamic estuarine ecosystems. Furthermore, co-monitoring of bacterial and viral communities over a long time scale will help us better understand whether viruses might control host community structure in natural environments. If variations of bacterioplankton community structure are synchronous with changes of viral community, it may provide a link between these two populations.

Since the study of bacterial community structure and population dynamics involved PCR-based molecular techniques, potential limitations associated with PCR and other steps of molecular processes may have affected the results obtained in my studies. PCR approaches may underestimate the complexity of microbial assemblages especially when dealing with natural assemblages containing high microbial diversity. In addition PCR-DGGE shows bias towards dominant groups while skipping the minor components within the natural communities (Chapter 2). In order to investigate these minor groups, group-specific primers or fluorescence-labeled specific probes can be designed and used to more efficiently detect bacterial groups. For example, FISH could target *Cytophaga*-like bacteria and *Betaproteobacteria* that are frequently underestimated by clone library or DGGE analyses (Cottrell and Kirchman 2000b;

Castle and Kirchman 2004). In addition, a high percentage (15-41%) of planktonic marine *Actinobacteria* was present in the Chesapeake Bay. Whether this high frequency of marine *Actinobacteria* in our clone libraries was caused by PCR bias is not known, but it could be confirmed by FISH using specific probes.

Application of proteomics to environmental samples allows us to study the major expressed proteins from a microbial community. Proteomic patterns obtained from the Chesapeake Bay microbial communities have shown distinct seasonal "protein fingerprints" in winter and summer. Interestingly, MDS analyses suggest that seasonal succession of these proteomic fingerprints is very similar as population structure variations revealed by DGGE. This result implies that changes on bacterial community could have a significant impact on microbially mediated processes in the Bay. Thus, proteomics is not only useful to study expressed proteins, but could also be an alternative tool to explore dynamics of metabolic functions of microbial communities. Another useful application of proteomic fingerprinting is that it can differentiate very closely related bacterial strains. For example, distinct proteome patterns were obtained from several *Roseobacter* species that share nearly identical 16S rRNA gene sequences (manuscript in preparation). Finally, characterization of highly expressed proteins holds the potential to explore important microbial processes or activities in the natural environment. Several microbial proteins from Chesapeake Bay microbial community have the closest matches with the functional genes in the Sargasso Sea metagenome database, suggesting that it is critical to have an extensive metagenome database prior to the application of community proteomics. Genome sequences from marine bacteria (e.g. *Silicibacter pomeroyi*, Moran et al. 2004;

Pelagibacter ubique, Giovannoni et al. 2005) and marine microbial communities are rapidly accumulating, and will continue to improve the accuracy of protein searching for community proteomics. Bioinformatic analysis of gene and proteins sequences from the marine ecosystem is only at its infancy and the application of advanced molecular tools will continue to uncover the secrets of ocean's life.

Appendix A

Co-monitoring bacterial and dinoflagellate communities by denaturing gradient gel electrophoresis (DGGE) and SSU rRNA gene sequencing during a dinoflagellate bloom

Abstract

Dinoflagellates are unicellular eukaryotic protists that dominate in all coastal waters, and are also present in oceanic waters. Despite the central importance of dinoflagellates in global primary production, the relationship between dinoflagellates and bacteria are still poorly understood. In order to understand the ecological interaction between bacterial and dinoflagellate communities, Denaturing Gradient Gel Electrophoresis (DGGE) and SSU rRNA gene sequencing were applied to monitor the population dynamics of bacteria and dinoflagellates from the onset to disappearance of a dinoflagellates bloom occurred in Baltimore Inner Harbor from April 15 to 24, 2002. Although *Prorocentrum minimum* is the major bloom forming species under the light microscopy, DGGE method with dinoflagellate specific primers demonstrated that *Prorocentrum micans, Karlodinium micrum* and *Gyrodinium uncatenum* were also present during the bloom. Population shifts among the minor dinoflagellate groups were observed. DGGE of PCR-amplified 16S rRNA gene fragments indicated that *cyanobacteria*, *Alpha-*, *Beta-*, and *Gammaproteobacteria*, *Flavobacterium-Bacteroidetes-Cytophaga* (FBC), and *Planctomycetes* were the major components of bacterial assemblages during the

bloom. DGGE analysis showed that Cytophagales and *Alphaproteobacteria* played important roles at different stages of dinoflagellates bloom. We demonstrated here that DGGE can be used as a rapid tool to simultaneously monitor population dynamics of both bacterial and dinoflagellates communities in aquatic environments.

Introduction

Dinoflagellates are important primary producers in both coastal and oceanic waters, and could play remarkable ecological roles on pelagic energy flow and nutrient cycling (Cole et al. 1982; Doucette et al. 1998). Many species of dinoflagellates are also capable of forming massive algal blooms. It is expected that availability of organic matters changed dramatically at different stages of an algal bloom. Bacterial biomass and production are known to be correlated with amount of organic matters released from bloom-forming species (Palumbo et al. 1984; Smith et al. 1995; Riemann et al. 2000). However, there is only limited information on the phylogenetic affiliations of bacteria associated with marine algal blooms (González et al. 2000). Moreover, little is known about the effect of bacterial succession on population structure of bloom-forming species.

The development of molecular approaches greatly enhanced our ability to study the population diversity of microorganisms in marine environments (Giovannoni et al. 1990; Ward et al. 1990; Lopez-Garcia et al. 2001; Moon-van der Staay et al. 2001). Recently, the DGGE technique has been widely used as a rapid method to examine the complexity of microbial communities including prokaryotes and eukaryotes (Muyzer et al. 1993; van Hannen et al. 1999; Bano and Hollibaugh 2002). Theoretically, DGGE can separate different PCR fragments even with single GC pair difference (Muyzer et al. 1993; Ferris et al. 1996). Therefore, diversity profile from different microbial communities can be compared according to their gel patterns and the sequences of representative bands.

In this study, we monitored bacterial and dinoflagellate population succession during a dinoflagellate bloom using DGGE method. PCR primers used in the study are specific for eubacteria and dinoflagellates, respectively. Bacterial and dinoflagellate communities at different stages of bloom were compared based on their DGGE fingerprints. The major DGGE bands were sequenced and identified based on the phylogenetic relationship with known species from the GenBank database.

Materials and methods

Sample collection. An algal bloom with dark brown color was observed at Inner Harbor, Baltimore on April 15, 2002. Water samples were collected daily from the pier 6 of Inner Harbor during the bloom period (April 15-24, 2002) using a bucket. Water temperature and salinity were recorded, respectively, when the samples were taken. Water samples (250 ml) were filtered through 0.2-μm-pore-size polycarbonate filters (47-mm diameter; Millipore, Bedford, Mass.) immediately after collection. Microbes retained on the filters were stored at -20˚C for further analysis. Meanwhile, additional 50 ml water samples were fixed by 1% glutaraldehyde for total bacterial and dinoflagellate counting. Microbial cells were stained by SYBR Gold (Molecular Probes, Inc., Eugene, Oreg.) as described by Chen et al. (2001) and enumerated using an epifluorescence microscope, Zeiss Axiplan (Zeiss, Germany). At least 200 cells were counted for bacteria and dinoflagellates, respectively.

Nucleic acid extraction. Total DNA was extracted according to a protocol developed by Schmidt et al. (1991) with minor modifications. DNA from bacteria and dinoflagellates were extracted by treating with lysozyme and proteinase K followed

by phenol extraction and isopropanol precipitation. DNA were dissolved in ddH_2O and stored at 4°C for further analysis.

PCR amplification of SSU rRNA gene. PCR amplification was performed in a 50-µl volume containing approximately 100 ng of template DNA, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.5 mM (each) primer, 200 mM (each) deoxynucleotide, and 2.5 U of Taq DNA polymerase (Promega, WI, USA).

The primers used to amplify eubacterial 16S ribosomal RNA (rRNA) gene were 1070F (eubacteria) and 1392R(GC) (universal), which contained a 40 bp GC-rich clamp (Ferris et al. 1996). The oligonucleotide primers for PCR amplification of dinoflagellate 18S rRNA gene were EUK4618R and DinoF (GC) (Oldach et al. 2000). The sequences, target sites and specificity of the primers are shown in Table A-1.

PCR amplification was carried out using a PTC-200 thermal cycler (MJ Research, Waltham, Mass.). For 16S rRNA gene, PCR program included an initial activation at 94°C for 5 min followed by 27 cycles using a touchdown PCR program developed by Muyzer (1993) to minimize nonspecific amplification. The 27 cycles were performed at 94°C for 0.5 min, T_A for 1 min, and 72°C for 3 min. In the first 20 cycles, T_A decreased by 1°C, stepwise, each two cycles, from 65°C in the first cycle to 56°C in the 20th. In the last five cycles, T_A was 55°C. Cycling was followed by 5 min of incubation at 72°C. The PCR cycle for dinoflagellates was performed as described by Oldach et al. (2000). One activation step at 95°C for 15 min was followed by 40 cycles of 95°C for 30s, 55°C for 30 s and 72°C for 40 s and then a

final extension step at 72°C for 5 min. Agarose gel (1%) electrophoresis was used to detect the PCR products.

Table A-1. Oligonucleotide sequences used for DGGE analyses

Primer	Sequence (5' to 3')	Target site	Specificity	Reference
1070F	ATGGCTGTCGTCAGCT	16S (1055-1070)[b]	Bacteria	Amann, et al. 1995; Ferris et al. 1996
1392R (GC)[a]	ACGGGCGGTGTGTAC	16S (1392-1406)[b]	universal	Amann, et al. 1995; Ferris et al. 1996
DinoF (GC)[a]	CGATTGAGTGAGTGATCCGGTGAATAA	Dino18S	Dinoflagellates	Oldach et al. 2000
EUK4618R	TGATCCTTCTGCAGGTTCACCTAC	18S	universal	Oldach et al. 2000

[a]The GC clamp sequence is CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCC
[b]*Escherichia coli* numbering of the ribosomal RNA operon

DGGE analysis and sequencing. DGGE was performed using the Dcode^{TM} Universal Mutation Detection System (Bio-Rad, Hercules, Calif.). PCR products were separated on a 1.5-mm-thick vertical gel containing polyacrylamide (acrylamide-bisacrylamide, 37.5:1) and a linear gradient of the denaturants urea and formamide, increasing from 40% at the top of the gel to 65% at the bottom. Equal amount of PCR products were loaded on the DGGE gel. Electrophoresis was performed at 60°C in a 0.5×TAE buffer, and 70 V of electricity was applied to the submerged gel for at least 16 h. DNA bands were visualized by staining with SYBR Gold and photographed (Øvreås et al. 1997).

Prominent DNA bands were excised from the gels, re-amplified and electrophoresed again in DGGE gels (at lease twice). PCR products were purified by Qiaquick PCR purification kit (Qiagen Inc., Valencia, CA). Prism Ready Reaction Dye Deoxy Termination Kit (Applied Biosystems, Foster City, Calif.) was used for sequencing in conjunction with *Taq* polymerase in a 373A DNA sequencer (Applied Biosystems, Foster City, Calif.). The PCR primers without GC clamp were used for sequencing. The sequences were submitted to GenBank and blasted against the NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/).

Phylogenetic analysis. Sequence alignment and phylogenetic tree reconstruction was performed using MacVector 7.1 program (Accelrys, San Diego, Calif.). Evolution distance was calculated by the Jukes-Cantor method (Jukes and Cantor 1969) and a distance tree was constructed with the neighbor-joining algorithm (Saitou and Nei 1987).

Results

Microbial abundance, water temperature and salinity. During the bloom period (April 15 to 24, 2002), microscopic examination indicated that *Prorocentrum minimum* was the dominant bloom-forming species (data not shown). *P. minimum* reached maximum cell density (2×10^5 cells/ml) on April 17 and decreased to 1700 cells/ml on April 24. Total bacterial counts were more than 5×10^6 cells/ml on April 19, and declined to 2×10^6 cells/ml on the day when the *P. minimum* cell density started to declined (Fig. A-1). During the bloom, bacterial abundance showed a positive correlation with dinoflagellate density ($r^2=0.985$, $p<0.05$). However, bacterial abundance had a postponed shift in relation to dinoflagellates.

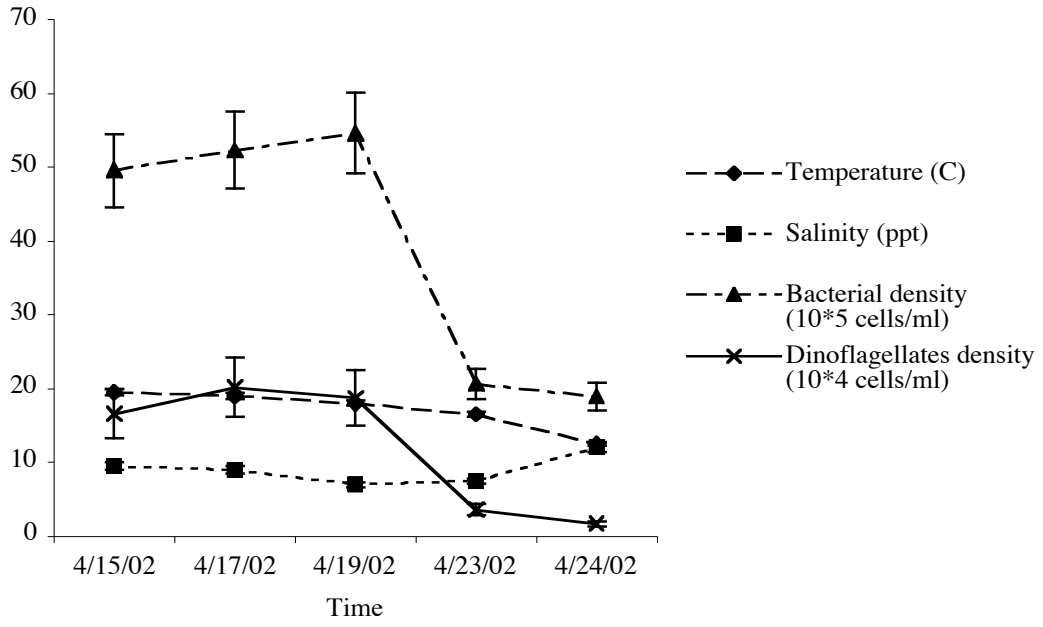


Fig. A-1. Bacterial and dinoflagellate counts, water temperature and salinity during the bloom.

Population structure of dinoflagellates. Five different dinoflagellate populations were detected by DGGE throughout the bloom period, and four of them were sequenced and identified. During the bloom period, *P. minimum* (band D3) was always detected and other minor dinoflagellate populations shifted dramatically according to their DGGE profiles (Fig. A-2, lanes 6-10). For examples, species D1 and D2 were relatively more abundant at the beginning of bloom, and species D4 emerged towards the end of bloom. Species D3 were detected at all stages of bloom. Phylogenetic analysis based on the 18S rRNA gene sequences indicated that species D1, D2, D3, and D4 are closely related to *P. micans*, *K. micrum*, *P. minimum*, and *G. uncatenum*, respectively (Fig. A-3).

Phylogenetic diversity of bacterial communities. Bacterial communities were much more complex than dinoflagellate communities. Typically, about 20 major bands were visible on the DGGE gel. In general, bacterial communities appeared to be stable during the bloom. A total of 19 bands were excised and sequenced (Fig. A-2, lane 1-5). Bands B14 and B17 were dual bands that could be separated better after re-amplification. There were multiple bands in B18 but only one band was re-amplified and sequenced in the following analysis. Bands B2 and B16 could not be re-amplified and thereof their sequences were not available. Bands from different samples with identical vertical positions in the DGGE gel were assumed to have identical sequences. The 19 bacterial sequences were closely related with *Cyanobacteria* and plastids (29%), *Alphaproteobacteria* (19%), *Betaproteobacteria* (14%), and

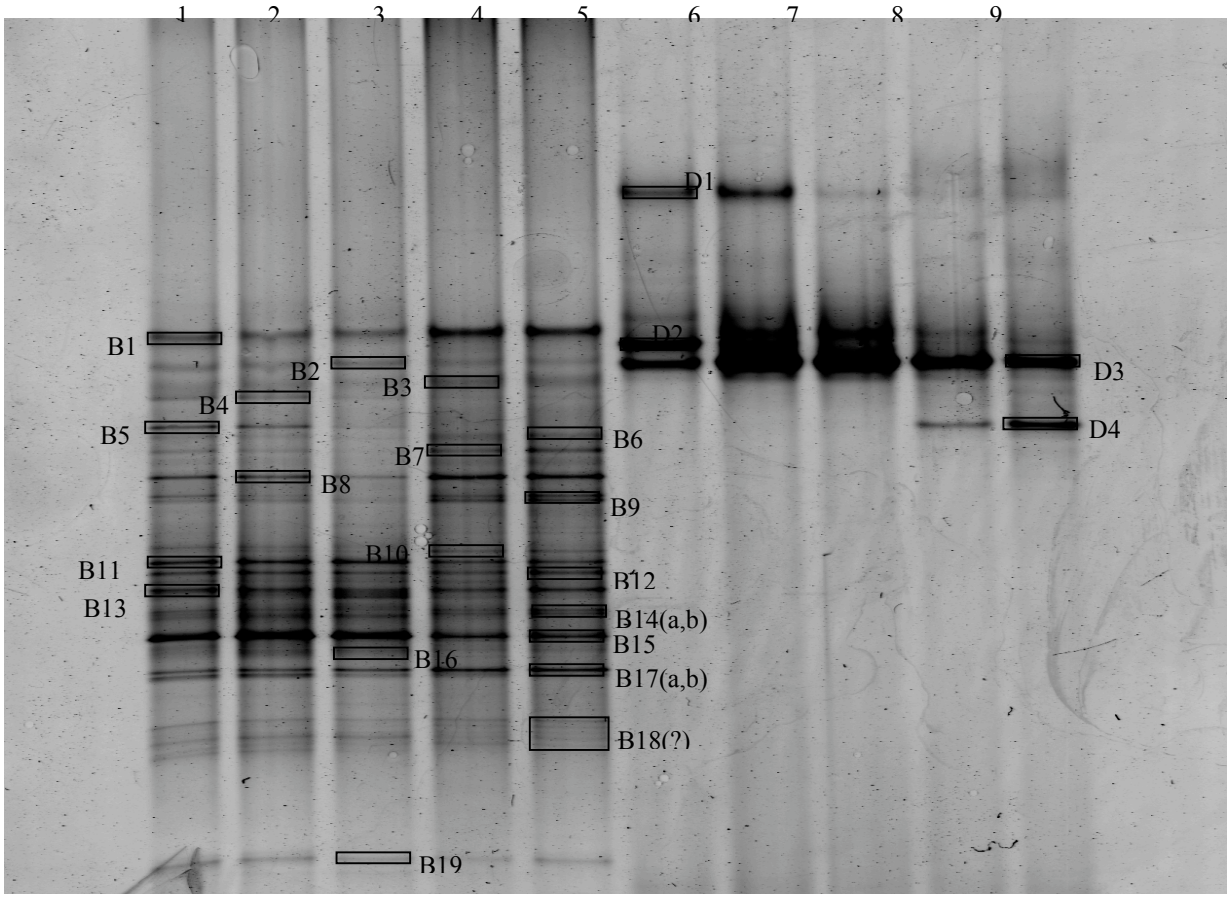


Fig. A-2. DGGE fingerprints of bacterial (lanes 1-5) and dinoflagellate populations (lanes 6-10). Sequenced bands from both communities were indicated. Lanes 1-5 represented the dinoflagellate community profiles at day 1, 3, 5, 9 and 10 of the bloom period. Bands excised for sequencing are indicated by boxes.

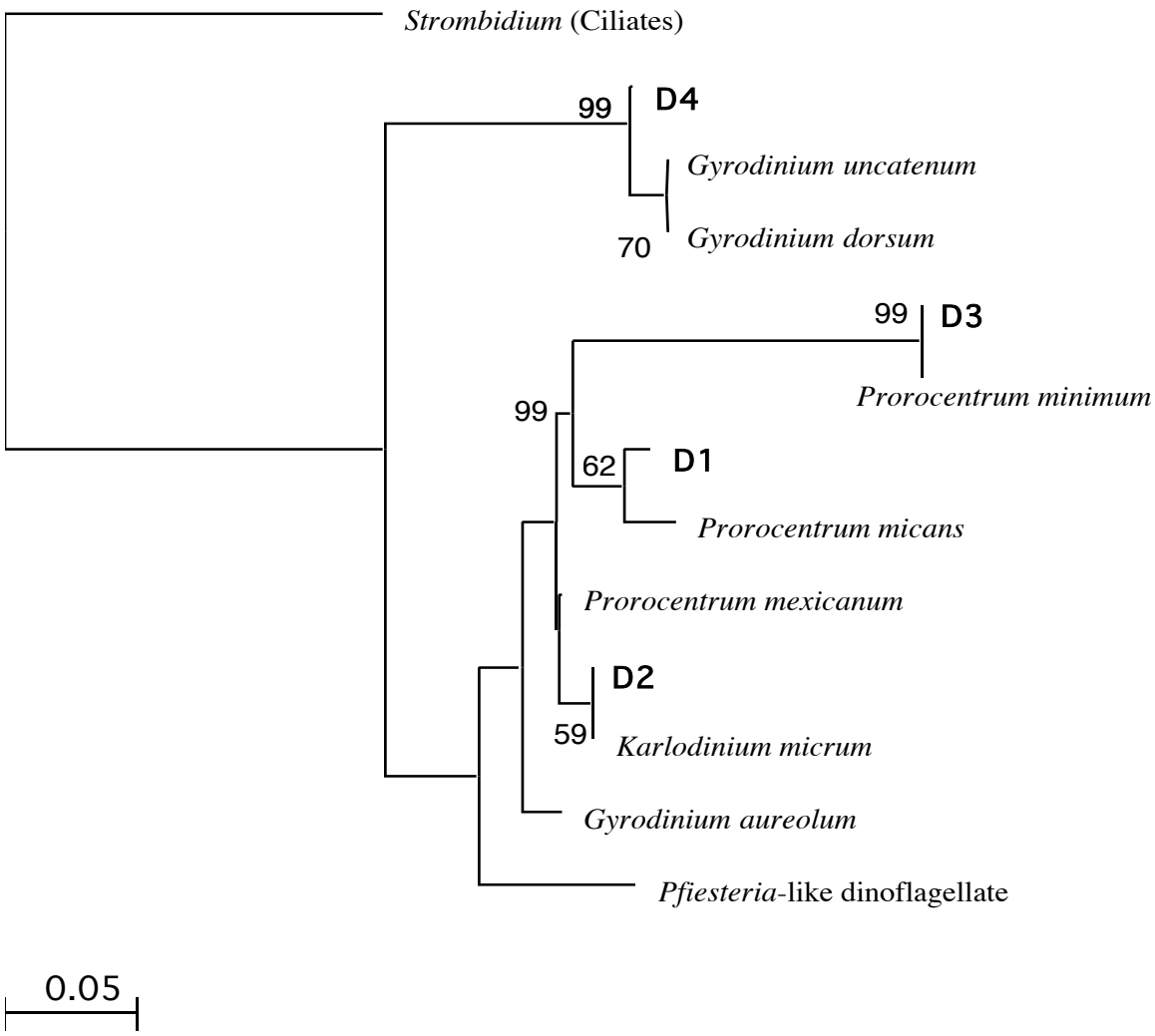


Fig. A-3. Phylogenetic tree based on 18S rRNA sequences of dinoflagellate species. D1, D2, D3, and D4 represented the bands excised from the dinoflagellate DGGE gel. A ciliate (*Strombidium* sp.) was used as an outgroup. The scale bar indicates substitutions per nucleotide position. Bootstrap values lower than 50% are not shown.

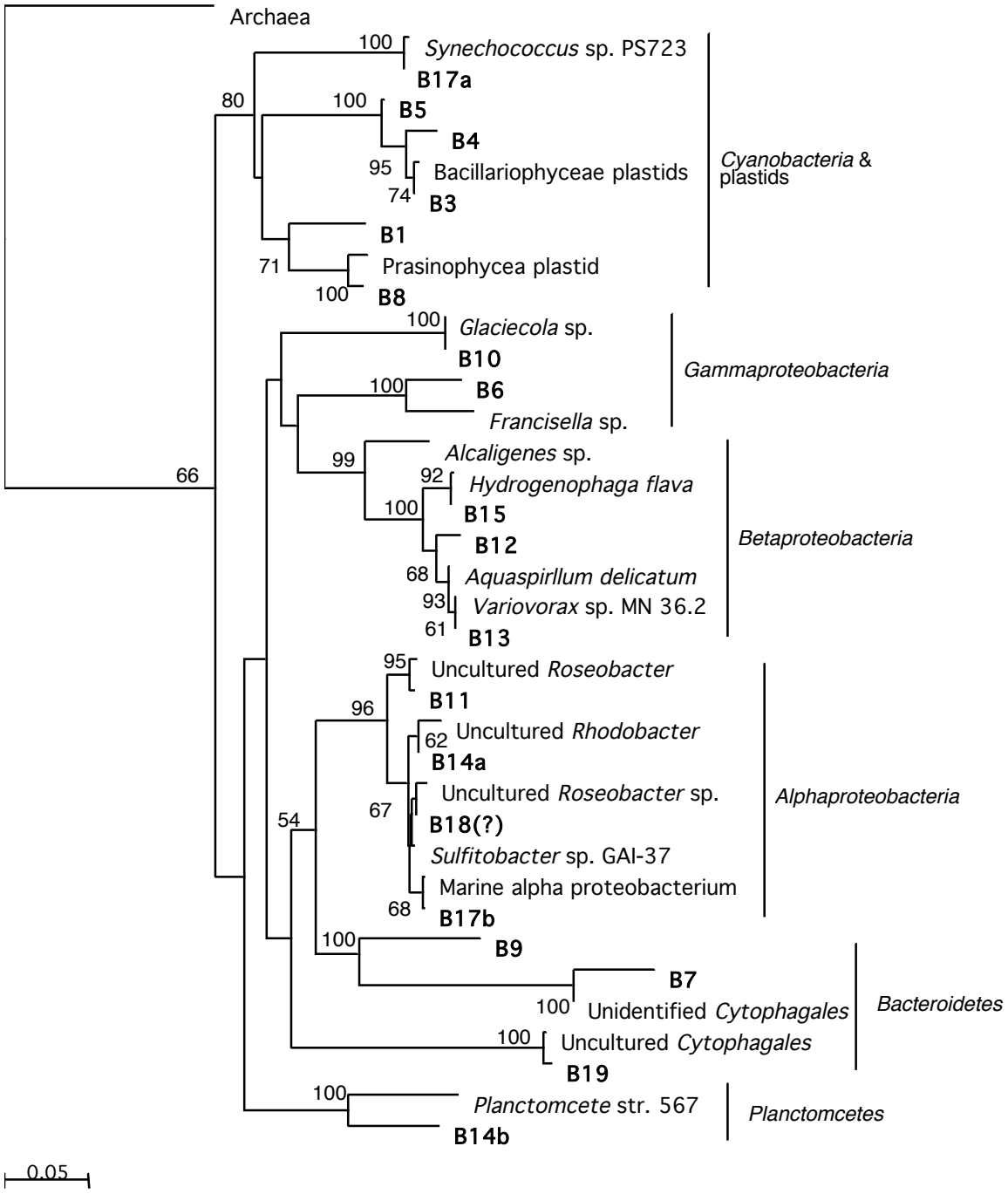


Fig. A-4. Phylogenetic affiliations of representative 16S rRNA gene sequences taken from the bloom samples. B1 to B19 represented 19 bands excised from bacterial DGGE gel. An *Archaea* (*Nanoarchaeum equitans*) was used as an outgroup. The scale bar indicates substitutions per nucleotide position. Bootstrap values lower than 50% are not shown. B14 and B17 had two bands sequenced and shown as a and b. B18 (?) had multiple bands but only one was re-amplified and sequenced.

Gammaproteobacteria (10%), *Flavobacterium-Bacteroidetes-Cytophaga* (FBC) (14%), and *Planctomycetes* (5%) respectively.

As shown in Fig. A-4, six bands (B1, B3, B4, B5, B8, and B17a) were related to photosynthetic organisms. B17a shared high similarity with *Synechococcus* sp. PS723. All other five bands were closely related with phytoplankton plastids. Four bands (B11, B14a, B17b and B18) were associated with *Alphaproteobacteria*. Bands B12, B13, and B15 were related to *Betaproteobacteria*. Three bands (B7, B9 and B19) were affiliated with *Flavobacterium-Bacteroidetes-Cytophaga* (FBC) group, which composed 14% of the whole community.

Bacteria-dinoflagellate interaction. An interesting interaction between bacteria and dinoflagellates were observed with epifluorescence microscopy (Fig. A-5). Fig. A-5A is a typical view *of P. minimum* cells during the peak of bloom. When the bloom began to fade, the nuclei of dinoflagellates became brighter and bacterial cells began to increase around the dinoflagellate nuclei (Fig. A-5B and 5C). It is common to see that bacterial cells aggregated around the nucleus of *P. minimum* when the cells were dying, as the bloom crashed (Fig. A-5D).

Discussion

Dinoflagellate blooms occur frequently from late spring to summer at the Baltimore Inner Harbor. Located in the northern part of the Chesapeake Bay, spring runoff from the Patapsco River provides Baltimore Inner Harbor with large inputs of organic matter and nutrients that in turn trigger dynamic responses of both dinoflagellate and bacterial communities. The close relationship between bacterial

and phytoplankton biomass has been studied (Pinhassi et al. 1999; Riemann et al. 2000; Fandino et al. 2001). However, very few studies have been conducted to study

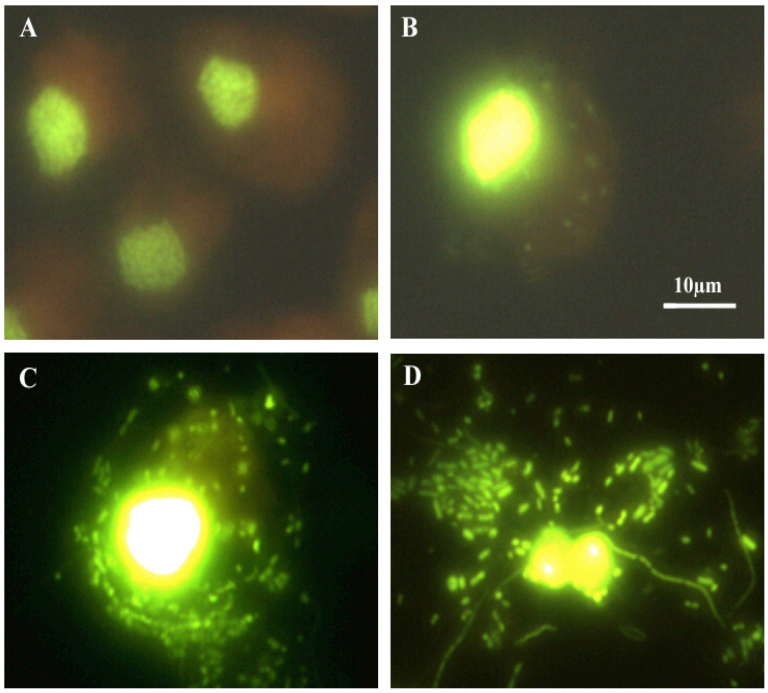


Fig. A-5. Epifluorescence microscopic images showing development of bacteria associated with *P. minimum* cells at the different stages of a bloom. A, normal *P. minimum* cells during peak of bloom; B and C, bacteria associated *P. minimum* cells at the bloom declining stage; D, nucleus from a lysed *P. minimum* cells at the late stage of bloom.

population composition of bacteria and phytoplankton (González et al. 2000; Fandino et al. 2001). Differentiating similar phytoplankton species by microscopy can be

problematic due to their similar morphological characteristics. In this study, we demonstrated that multiple dinoflagellate species could be detected using the PCR-DGGE method during the *P. minimum* bloom. For example, *P. micans* and *P. minimum* were well separated on the DGGE gel. Our results also indicated that the DGGE method was not biased for the dominant species like *P. minimum*. The minor species like *K. micrum* and *G. uncatenum* that were difficult to find by light microscopy were detectable with dinoflagellate specific PCR primers. With such a detection sensitivity, the DGGE method can be a useful tool to monitor population shift of phytoplankton.

During the bloom, phylotypes related to *Alphaproteobacteria* (bands B11, B14a, B17b and B18) were present in all the analyzed samples. Intensity of bands B11 (*Roseobacter*) and B17b (marine *Alphaproteobacteria*) were much higher in bloom samples indicating that these groups of bacteria could be numerically abundant in the bloom samples. The *Roseobacter* lineage made up over 20% of the bacterial rRNA gene associated with a *Emiliania hyxleyi* bloom (González et al. 2000). A high proportion of *Alphaproteobacteria* was also found in a mesocosm diatom bloom (Reimann, et al. 2000) and a *Lingulodiniuim polyedrum* bloom off the southern California coast (Fandino et al. 2001). It is thought that *Roseobacter* may play a role in cycling of organic sulfur compound produced during the bloom (González et al. 2000). *Alphaproteobacteria* could accelerate uptake of amino acid in marine waters (Cottrell and Kirchman, 2000a).

Three *Cytophagale* phylotypes were identified during the bloom. The *Cytophagale* lineage (band B7) appeared to increase in band intensity during the late

stage of bloom. *Cytophaga*-related species are abundant in marine environments (Glöckner et al. 1999) and known to be involved in the degradation of complex macromolecules (Shewan and McMeekin 1983). *Cytophaga* phylotypes were more abundant during the late stage of a diatom bloom (Riemann et al. 2000). Organic particles retrieved from decay of phytoplankton provide *Cytophaga* with an ecological niche for colonization and hydrolysis of organic matters.

Gammaproteobacteria (B6 and B10) also emerged at the late stage of the bloom. Consistent with our results, Fandino et al. (2001) found that *Gammaproteobacteria* showed phylotype richness and predominance of abundance in a dinoflagellate bloom. SAR86, the ubiquitous cluster of *Gammaproteobacteria* was also found to be prominent in heterotrophic bacterial communities during a North Atlantic algal bloom (González et al. 2000).

Bands B13 and B15 sharing high similarity with *Betaproteobacteria* also showed high intensities in bloom samples. Recently, transmission electron microscopy (TEM) and *in situ* hybridization showed that *Betaproteobacteria* were the major endonuclear and endocytoplasmic bacteria in dinoflagellates (Alverca et al. 2002). Intracellular symbiotic bacteria associated with different dinoflagellate species have been studied for a long time (Silva 1978). For example, FBC groups were found in cytoplasm of the dinoflagellates cells, but were absent from the nucleus (Alverca et al. 2002, Biegala et al. 2002). Although a tight interaction between bacterial and dinoflagellate cells were observed in our study, it is not clear whether the bacteria are associated with nucleus or cytoplasm of dinoflagellates (Fig. 5). *In situ* molecular

tools (e.g. FISH and *in situ* PCR) coupled with laser confocal microscopy can be exploited to further understand the interaction between bacteria and dinoflagellates.

Some potential biases associated with PCR and DGGE methods have been discussed elsewhere (Muyzer 1999; Diez et al. 2001). For eukaryotes, some bias may be due to rRNA gene copy numbers because eukaryotes usually have very high copy numbers for some genes (Long and Dawid 1980). It is worth pointing out that DGGE is not a strictly quantitative approach, but its high reproducibility demonstrates that it reflects the major variations of PCR-amplifiable phylotypes in natural communities (Riemann et al. 1999). Changes in band patterns and intensity of same bands likely reflect the changes in the relative abundance of compositions of microbial communities.

Appendix B

Applications of community proteomics to marine microbial communities

I. Comparative proteomics of Chesapeake Bay microbial community

The first successful community proteomics on complex microbial assemblages encouraged further applications. In order to compare the metaproteomes from different locations and seasons, Chesapeake Bay microbial community were collected from northern (858), middle (804) and southern station (707) in Jun, Aug 2004, and Feb 2005, respectively. Comparative proteomics analysis on Chesapeake Bay samples over space and time indicated that proteomics provide enough resolution to differentiate spatial and temporal variations of microbial communities (Fig. B-1).

The variations may come from different sources. First, the structural difference may contribute the variation of functional gene expression. Chesapeake Bay bacterioplankton experienced dramatic seasonal variations in population structure. Multiple year investigation on population dynamics indicated that distinct communities exist in winter and summer. This provides a good explanation on the significant difference observed in proteomic patterns from Summer 2004 and Winter 2005 (Fig. B-2). Secondly, it is likely differential gene expression or protein modifications are common when similar microbial assemblages are under environmental gradients. MDS analysis showed that Chesapeake Bay DGGE fingerprints and proteomics patterns were clustered by sampling time instead of

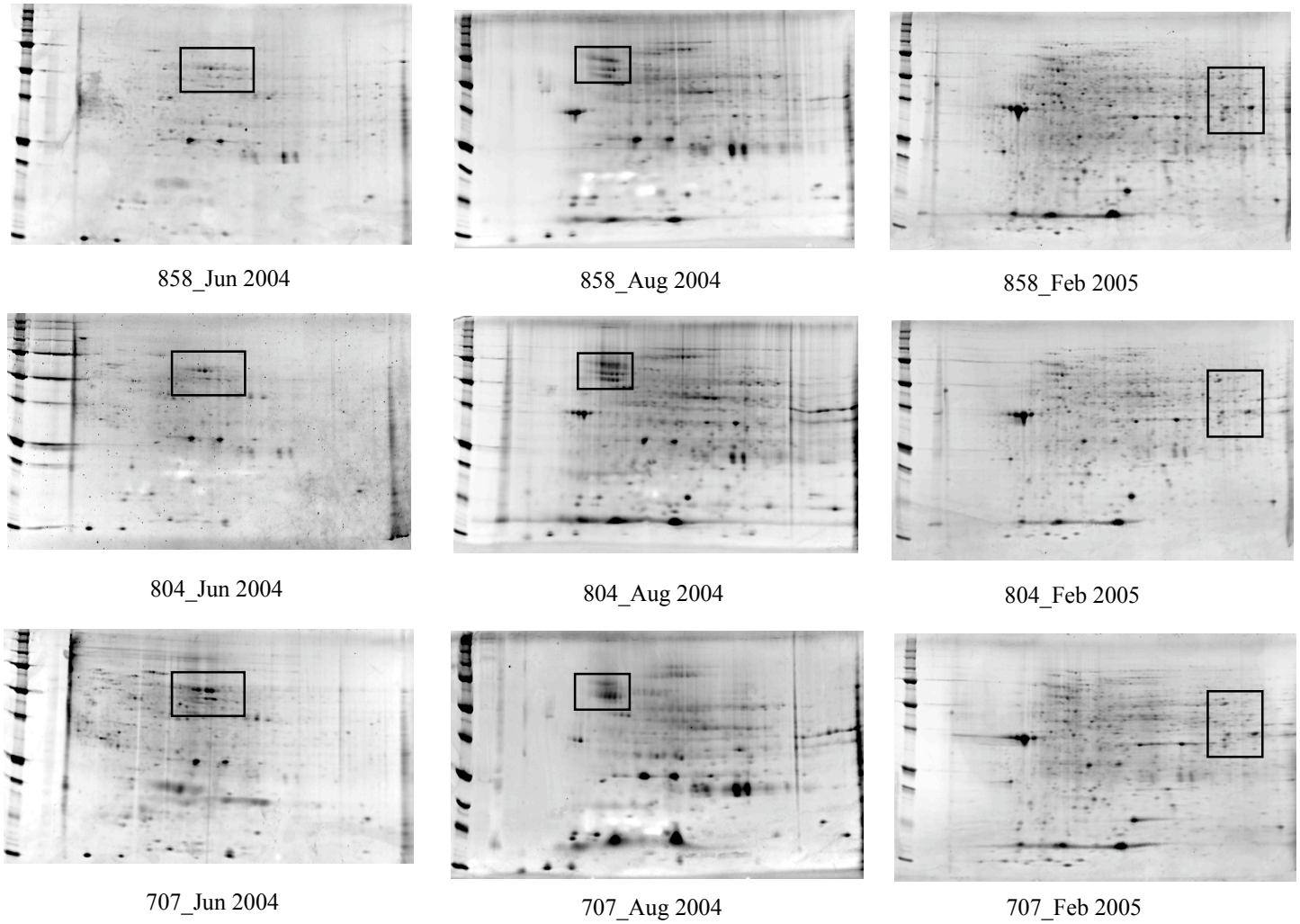


Fig. B-1. 2DGE images of metaproteomes obtained from the Chesapeake Bay. Unique protein spots patterns from different seasons were highlighted.

spatial distribution (Fig. B-2), indicating stronger temporal variations than spatial ones. Further characterization of the differentially expressed proteins is on the way and it will provide insights on their biogeochemical significance and ecological adaptation as well.

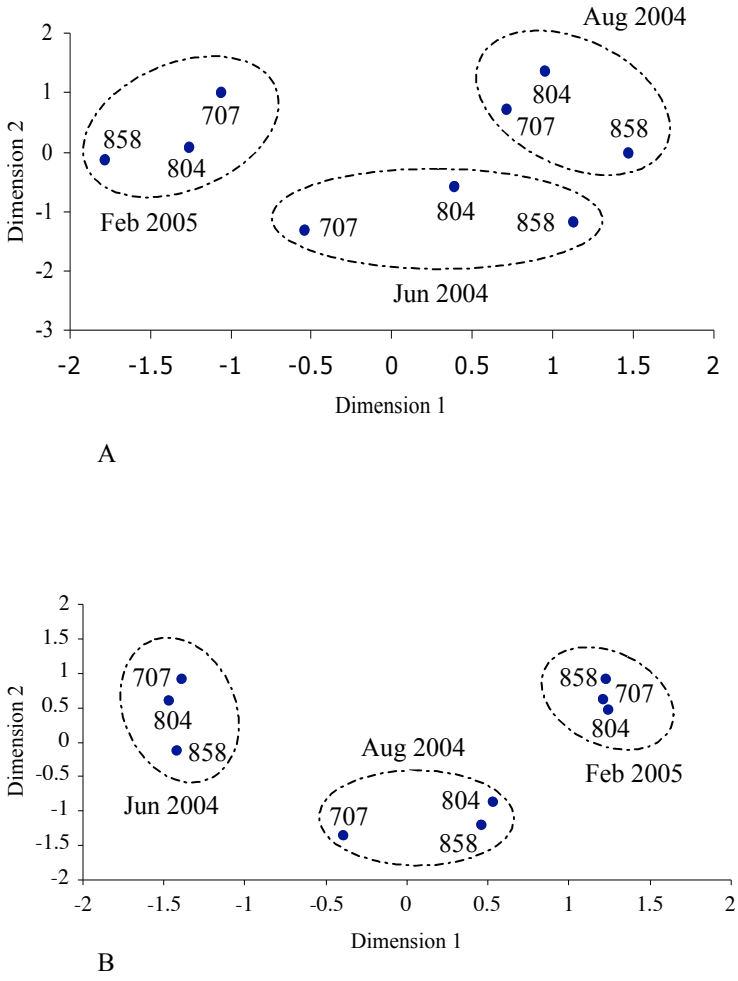


Fig. B-2. MDS plots for DGGE banding patterns (A) and community proteomic patterns (B). Stress values for A is 0.1 and for B is 0.06.

II. Community proteomics on marine sponge-microbe complexes

Sponges form symbiotic relationships with high diversity of heterotrophic and autotrophic bacteria, protozoa, and macroalgae. In some cases, the complex microbial communities can compromise more of the biomass comes from symbionts than from the sponge itself (Wilkinson 1978). It is apparent that the symbionts can fulfill diverse roles that are important for both host and bacteria (see review by Haygood et al. 1999). For instance, it has been known that sponges contain many bioactive compounds of biomedical interest and these compounds may be produced by the sponge-associated microbes (Lee et al. 2001; Faulkner 2002; Hill 2004). Thus, studying the bacteria associated with marine sponges will improve our understanding of the symbiotic relationship within sponge-microbe complexes and provide a potential for natural products discovery. Isolating or harvesting the symbiotic bacteria from sponges, however, can be difficult because not all symbionts are readily cuturable. To study *in situ* activities of symbiotic bacteria, we extracted proteins from sponge tissues, which contain proteins from both the sponge and symbiotic bacteria. Different species of sponges presented distinct proteome patterns (Fig. B-3, A-C), suggesting that 2DGE-based proteomics can be successfully applied to explore the functional gene expression of the sponge and the associated bacteria. Further characterization of the abundant proteins will help us identify the source of proteins, and understand the symbiotic physiology and synthetic pathways of bioactive compounds.

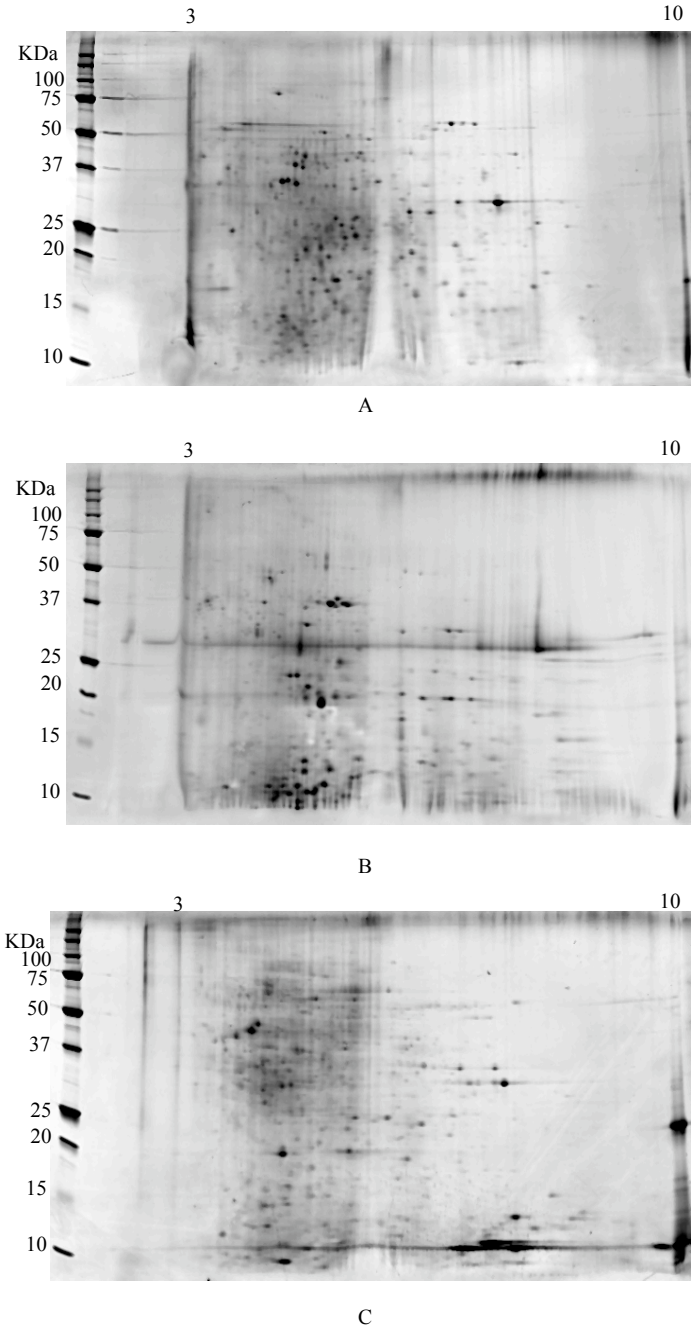
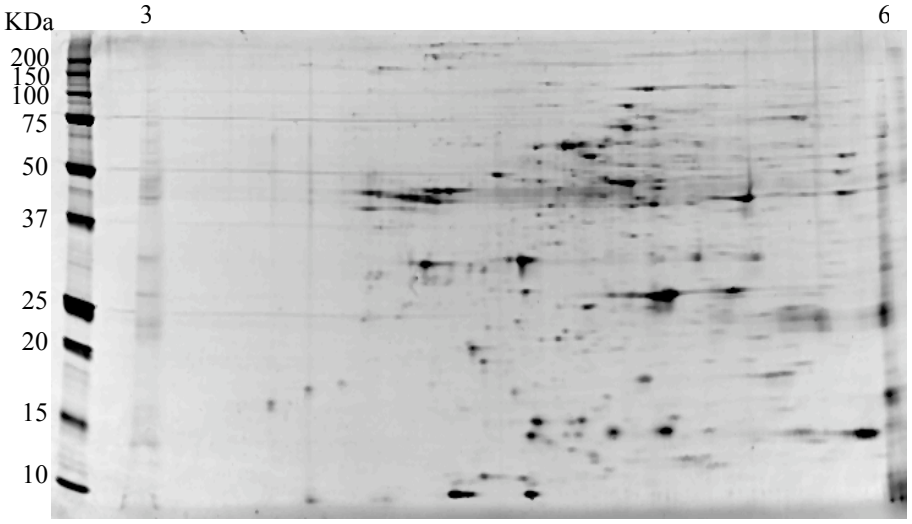


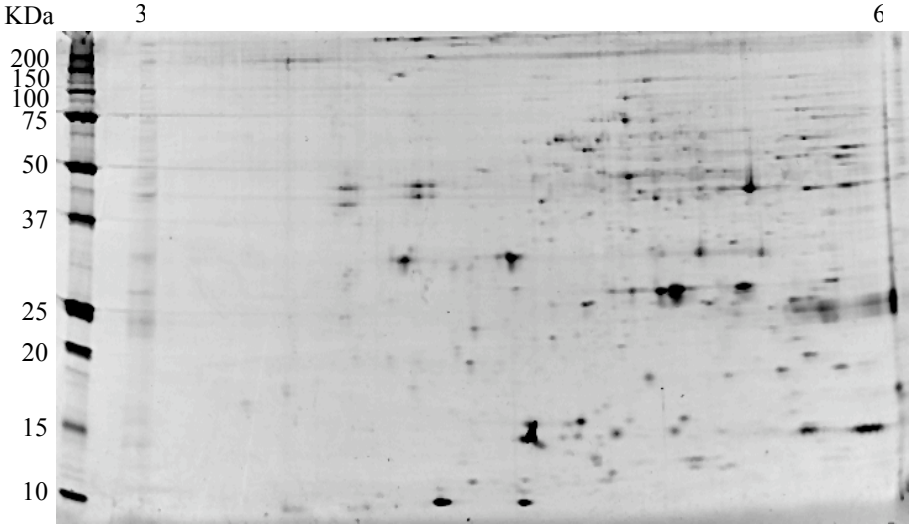
Fig. B-3. 2DGE analysis of sponge-microbe complexes. A, *Microciona prolifera* from the Chesapeake Bay; B, *Mycale laxissima* from Key Largo, Florida; C, *Xestospongia testudinaria* from Indonesia.

III. Community proteomics on episymbiotic bacterial communities from hydrothermal vent polychaete worm *Alvinella pompejana*

Deep-sea hydrothermal vent environment is characterized by high temperature, high concentrations of hydrogen sulfide and heavy metals (Edmond and von Damm 1985). Episymbiotic bacterial community associated with *Alvinella pompejana* is dominated by two filamentous *Epsilonproteobacteria* phylotypes and a *Spirochete* phylotype (Haddad et al. 1995; Cary et al. 1997; Campbell and Cary 2001). While these bacteria have not been cultivated, the role of the epibionts in the symbiotic association with *A*. *pompejana* is unclear. Two samples collected at the same location showed similar proteomic patterns (Fig. B-4). Acidic proteins of hydrothermal vent episymbiotic bacteria demonstrated the adaptation of the bacteria to the low pH ambient environment.



3838-250



3838-263

Fig. B-4. 2DGE analysis of episymbiotic bacterial communities from hydrothermal vent polychaete worm *A. pompejana*. Samples 250 and 263 were collected from site 3838.

Appendix C

Media, Buffers and Protocols

Media

1/2 YTSS (Yeast Extract-Tryptone-Sea Salt) Broth

4 g Yeast Extract

2.5 g Tryptone

200 ml 2.5 × Sea Salt Solution

800 ml DI (deionized) H_2O

Autoclave sterilize

2.5 × Sea Salt Solution

100 g Sea Salt

1 Liter DI H_2O

Autoclave sterilize

1/2 YTSS (Yeast Extract-Tryptone-Sea Salt) Agar

4 g Yeast Extract

2.5 g Tryptone

15 g Bacto-agar

200 ml 2.5 × Sea Salt Solution

800 ml DI H_2O

Autoclave sterilize

Modified 1/2 YTSS Broth for estuarine bacteria

4 g Yeast Extract

2.5 g Tryptone

1 Liter *in situ* viral-particle-free water*

Autoclave sterilize

* Obtained from tangential flow ultrafiltration system

Modified 1/2 YTSS Agar for estuarine bacteria

4 g Yeast Extract

2.5 g Tryptone

15 g Bacto-agar

1 Liter *in situ* viral-particle-free water*

Autoclave sterilize

* Obtained from tangential flow ultrafiltration system

Luria Bertani (LB) Broth

10 g Tryptone

5 g NaCl

5 g Yeast Extract

1 Liter DI H_2O

Autoclave sterilize

Luria Bertani (LB) Agar

10 g Tryptone

5 g NaCl

5 g Yeast Extract

15 g Bacto-agar

1 Liter DI H_2O

Autoclave sterilize

1/2 Marine Broth 2216

17.5 g 2216 marine broth powder

1 Liter DI H_2O

Autoclave sterilize

1/2 Marine Agar 2216

17.5 g 2216 marine broth powder

15 g Bacto-agar

1 Liter DI H_2O

Autoclave sterilize

Buffers

TE buffer

PH 7.4, 7.6 or 8.0)

10 mM Tris-Cl (pH 7.4, 7.6 or 8.0)

1 mM EDTA (pH 8.0)

10 × TBE (Tris-Borate-EDTA)

108 g Tris base

55 g Boric Acid

40 ml 0.5 M NaEDTA (pH 8.0)

DI H_2O to 1 Liter

Autoclave sterilize

50 × TAE (Tris-Acetate-EDTA)

242 g Tris base

57.1 ml Glacial Acetic Acid

100 ml 0.5 M NaEDTA (pH 8.0)

DI H_2O to 1 Liter

Autoclave sterilize

6 × gel loading buffer

25 mg Bromophenol Blue

25 mg Xylene Cyanol

4 g Sucrose

DI H_2O to 10 ml

Sucrose Lysis Buffer

16 ml 0.5 M NaEDTA (pH 8.0)

10 ml 1 M Tris-Cl (pH 8.3)

51.3 g Sucrose

DI H_2O to 200 ml

Autoclave sterilize

Pre-wash buffer for protein extraction

1 ml 1M Tris-Cl (pH 7.6)

8.55 g Sucrose

DI H_2O to 100 ml

Autoclave sterilize

Protein Extraction buffer (pH 6.8) for SDS-PAGE

50 µl 1 M Tris-Cl

100 µl 50 mM EDTA

2.4 g powder Urea

0.76 g powder Thiourea

500 µl Glycerol

500 µl 10% SDS

DI H_2O to 5 ml and aliquot to tubes (5 × 1 ml), store at –20°C

Protein Extraction buffer for 2DGE

2.1 g Urea

760 mg Thiourea

200 mg CHAPS (zwitterionic detergent)

25 µl 40% Carrier ampholytes

10 µl of 0.1% Bromophenol Blue

DI H_2O to 5 ml and aliquot to tubes (5 × 1 ml), store at –20°C

Add 10 µl 200mM TBP (Tributyl Phosphine) to each tube prior to use

Equilibration buffer I

36 g Urea

20 ml 10% SDS

3.3 ml 1.5 M Tris-Cl (pH 8.8)

40 ml 50% Glycerol

2 g DTT (Dithiothreitol)

DI H_2O to 100 ml

Equilibration buffer II

36 g Urea

20 ml 10% SDS

3.3 ml 1.5 M Tris-Cl (pH 8.8)

40 ml 50% Glycerol

2.5 g Iodoacetamide

DI H_2O to 100 ml

Protocols

I. Extraction of DNA from aquatic microbial communities

1. Collect 500 ml samples and filtrate it through 0.22 μm pore-size polycarbonate membrane filter (47 mm in diameter, Millipore). Note: change the filter when the flow rate is too slow.
2. Place the filter(s) into a Whirl-Pak bag and store the bag at –20 °C (only if you are not going to extract the DNA right after this).
3. Thaw the filters in ice and add 2 ml pre-lysis buffer (Tris-HCl, 0.1 M; EDTA, 0.1 M; sucrose, 0.8 M), 10 μl lysozyme (200 μg/μl), and incubate the sample at 37°C for 30 min.
4. Add 10 μl proteinase K (20 mg/ml) and SDS (final concentration: 1%), respectively, and incubate the sample at 37°C overnight.
5. Divide the sample into two microcentrifuge tubes, 1ml for each tube.
6. Set the incubator temperature at 65°C. Add 100 μl CTAB (10%) +NaCl (1.4 M) to each tube, and incubate at 65°C for 30 min.
7. Add equal volume of phenol-chloroform-isomyl alcohol (25:24:1) to each tube, mix and centrifuge both tubes at 13,000 rpm for 15 min.
8. Take supernatant, and add equal volume of chloroform-isomyl alcohol (24:1) to the supernatant, centrifuge at 13,000 rpm for 5 min.
9. Repeat 5-6 if necessary.

10. Take supernatant, add equal volume of isopropanol or two volume of cold ethanol (-20 °C), place the tubes at –80 °C for 20 min or –20 °C for 2hr.
11. Centrifuge both tubes at 13,000 rpm for 15 min, wash DNA pellet with 70% ethanol twice.
12. Dry DNA pellet with Speedvac (~10 min).
13. Add 50 µl double distilled water to one tube and keep the other tube dry as a backup.

II. Denaturing gradient gel electrophoresis

Equipment:

1. Dcode Universal Mutation Detection System (Bio-Rad cat. #: 1709080)
2. Power supply
3. Gradient maker (Hoefer SG50)
4. Stirrer (VWR 200 mini stirrer)
5. Peristaltic pump and tubing (Millipore cat. #: XX80 ELO 85)
6. 22 filling needle (Becton Dickinson 22G1 cat. #: 5155)

Materials:

1. PCR products for DGGE analysis
2. 2. 50 × TAE (see previous buffer formula)
3. Deionized Formamide (Fisher cat. #: BP227-500)
4. Urea (Gibco cat. #: 15505050)
5. 40% Acrylamide:bisacrylamide (37.5:1) (Bio-Rad cat. #: 161-0148)

6. Denaturant Stock Solution A (DSSA) (0% denaturant)

   20 ml 40% Acrylamide:bisacrylamide (37.5:1)

   2 ml 50 × TAE

   DI H_2O to 100 ml

7. Denaturant Stock Solution B (DSSB)

   20 ml 40% Acrylamide:bisacrylamide (37.5:1)

   2 ml 50 × TAE

   42 g Urea

   40 ml Formamide

   DI H_2O to 100 ml

8. 10% Ammonium Persulfate (APS) (Bio-Rad cat. #: 161-0700)

   0.1 g Ammonium Persulfate

   DI H_2O to 1 ml

9. TEMED (Bio-Rad cat. #: 161-0801) stored at 4 °C

10. 6 × gel loading buffer (see previous buffer formula)

Casting and Electrophoresis:

1. Clean glass plates with water and ethanol, rinse and dry them well.
2. Assemble the glass plate according to Dcode manual.
3. Make 12 ml of both low and high denaturant solutions according to the following table.

Denaturant (%)	DSSA (ml)	DSSB (ml)
0	12	0
5	11.4	0.6
10	10.8	1.2
15	10.2	1.8
20	9.6	2.4
25	9	3
30	8.4	3.6
35	7.8	4.2
40	7.2	4.8
45	6.6	5.4
50	6	6
55	5.4	6.6
60	4.8	7.2
65	4.2	7.8
70	3.6	8.4
75	3	9
80	2.4	9.6
85	1.8	10.2
90	1.2	10.8
95	0.6	11.4
100	0	12

4. Add 120 µl of APS and 6 µl of TEMED to each denaturant solution just prior to pouring to the gradient maker. Make sure all valves are closed on the gradient maker. Pour the low denaturant solution to entrance chamber. Slowly open the entrance chamber valve and then quickly close it (eliminate air bubbles between the chambers).
5. Pour the high denaturant solution into the exit chamber.
6. Turn on the stirrer and mix the denaturant solution in the exit chamber.
7. Turn on the peristaltic pump with lowest setting and open the exit valve. Turn the vacuum pressure up to the setting of 45-50.

8. Use the filling needle to pour the gel, keeping it close to the side of the gel plates.
9. When all the liquid has been poured into the sandwich, remove the needle and rinse the chamber and tubing with distilled water.
10. Place the comb in the sandwich.
11. Make the stacking gel in the exit chamber (5 ml stacking gel: 5 ml DSSA, 50 µl APS and 2 µl TEMED) and slowly pour on the top of the running gel (pump setting below 30).
12. Allow the gel to solidify for at least two hours.
13. Follow the instruction and set up the Dcode apparatus. Fill the electrophoresis tank with 7 liter of 1 × TAE buffer.
14. Set the temperature controller to 60 °C and wait until the buffer reached the desired temperature.
15. Add gel loading buffer (6 ×) to PCR products and load appropriate amount of samples.
16. Set the constant voltage to 70 V and run the gel overnight or for at least 16 hours.
17. After the run, turn off the power, pump and heater and remove the top.
18. Disassemble the gel and leave the gel on the top of the larger glass plate. Place the gel and the plate into staining buffer. (Note: Be very careful at this stage not to rip the gels, separate slowly).
19. Stain gels in 1:10,000 solution of SYBR Gold (Invitrogen/Molecular Probes cat. #: S11494), slowly rocking for 15 minutes.

20. Visualize the gel on a fluorimager or transilluminator (e.g. Typhoon 9410).

III. Community proteomics of aquatic microbial communities

1. Prefiltrate 20 L seawater using SmartWater Household Filtration system (cartridge size: 5-10 μm), and collect the filtrate.
2. Add Chloramphenicol (0.2 g) and 2 ml Protease inhibitor cocktail (CALBIOCHEM, Cat. 539132) to the filtrate, mix well.
3. Further filter the filtrate through 3 μm pore-size filter (pressure tank Alloy Products Corp. cat #: 105193-015; 144 mm filtration apparatus cat #: 0919003A) tubing and container), collect the filtrate. Note: change the filter when the flow rate is too slow (usually ~5 L in the Chesapeake Bay).
4. Concentrate the filtrate (about 20 L) using tangential flow ultrafiltration system (Millipore S1Y30 (30K) Amicon spiral cartridge cat #: 540640, 30, 000 MW cutoff) into ~150 ml.
5. Pellet the cells (13,000 rpm for 10 min).
6. Rinse the cells using TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 7.4), resuspend and recentrifuge.
7. Prepare the 2D ReadyPrep Rehydration/Sample Buffer (Bio-Rad, 163-2106), freshly add DNase and RNase (100× stock, DNase 10 mg/ml and RNase 2.5 mg/ml), TBP (reducing agent, 100×, 200 mM), Protease inhibitor cocktail (CALBIOCHEM, Cat. 539132, ~100×).

8. Apply the extraction buffer to the cells and lyse for 30 min on ice, vortex every 10 min.
9. French pressure cell (20,000 PSI) to crack the cells open at low temperature (4°C).
10. Centrifuge at 10,000 rpm for 3 min and collect the supernatant.
11. TCA precipitation: add 100 μl TCA (100%) to each 1 ml supernatant and precipitate at -20°C for 20 min).
12. Centrifuge at 13,000 rpm for 5 min and discard the supernatant.
13. Apply the exactly same buffer to the pellet, resuspend and redissolve the proteins.
14. Measure the protein concentration (A280 or Lowry method by Bio-Rad DC protein assay).
15. Load sample to IEF system (Bio-Rad Protean IEF System cat. #: 165-4000) (for 11 cm IPG strips, 200 μl and 250 μg protein preferred) and start the IEF program.

Rehydration and IEF program:

Step 1	50 V	12 hr (11-16 hr)		
Step 2	250 V	20 min	--	Ramp (linear)
Step 3	8,000 V	2.5 hr	--	Ramp (linear)
Step 4	8,000 V	--	40,000 V-hr	Ramp (rapid)

16. Second dimension: SDS-PAGE. 200 V, ~1 hr (Bio-Rad Criterion system with precasting gel).
17. SYPRO-Ruby staining (Bio-Rad cat #: 170-3125) and imaging (Typhoon 9410).
18. Gel images analysis (Z3 2-D gel image analysis system or Amersham ImageMaster software).
19. Protein characterization (N terminal sequence or mass spectrometry).

Bibliography

Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature. 430:551-554.

Agawin, N. S. R., C. M. Duarte, and S. Agusti. 1998. Growth and abundance of *Synechococcus* sp. in a Mediterranean Bay: seasonality and relationship with temperature. Mar. Ecol. Prog. Ser. 170:45-53.

Akkermans, A. D. L., J. D. van Elsas, F. J. de Bruijn. 1999. Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht.

Alavi, M., T. Miller, K. Erlandson, R. Shneider, and R. Belas. 2001. Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. Environ. Microbiol. 3:380-396.

Alverca, E., I. C. Biegala, G. M. Kennaway, J. Lewis, and S. Franca. 2002. *In situ* identification and localization of bacteria associated with *Gyrodinium instriatum* (Gymnodiniales, Dinophyceae) by electron and confocal microscopy. Eur. J. Phycol. 37:523-530.

Amann R. I., W. Ludwig, and K. Schleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143-169.

Anderson, N. L., and N. G. Anderson. 1998. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 19:1853-1861.

Anderson, A., I. Haecky, and A. Haraström. 1994. Effect of temperature and light on the growth of micro-, nano-, and picoplankton: impact on algal succession. Mar. Biol. 120:511-520.

Aubel, D., F. N. R. Renaud, and J. Freney. 1997. Genomic diversity of several Corynebacterium species identified by amplification of the 16S-23S rRNA gene spacer regions. Int. J. Syst. Bacteriol. 47:767-772.

Azam F. 1998. Microbial control of oceanic carbon flux: the plot thickens. Science. 280:694-696.

Bano N. J., and T. Hollibaugh. 2002. Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. Appl. Environ. Microbiol. 68:505-518.

Barcina, I., P. Lebaron, and J. Vives-Rego. 1997. Survival of allochthonous bacteria in aquatic systems: a biological approach. FEMS Microbiol. Ecol. 23:1-9.

Béjà, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong. 2000a. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science. 289:1902-1906.

Béjà, O., E. N., Spudich, J. L. Spudich, M. Leclerc, and E. F. DeLong. 2001. Proteorhodopsin phototrophy in the ocean. Nature. 411:786-789.

Béjà, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong. 2000b. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ. Microbiol. 2:516-529.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2005. GenBank. Nucleic. Acids. Res. 33:D34-38.

Bergh, O., K. Y. Borsheim, G. Brathak, and M. Heldal. 1989. High abundance of virus found in aquatic environments. Nature. 340:467-468.

Bernard, L., H. Schaefer, F. Joux, C. Courties, G. Muyzer, and P. Lebaron. 2000. Genetic diversity of total, active and culturable marine bacteria in costal seawater. Aquat. Microb. Ecol. 23:1-11.

Biddanda, B., M. Ogdahl, and J. Cotner. 2001. Dominance of bacterial metabolism in oligotrophic relative to eutrophic waters. Limnol. Oceanogr. 46: 730-739.

Bidle, K. D., and M. Fletcher. 1995. Comparison of free-living and particle-associated bacterial communities in the Chesapeake Bay by stable low-molecular-weight RNA analysis. Appl. Environ. Microbiol. 61:944-952.

Biegala, I. C., G. Kennaway, E. Alverca, J. Lennon, D. Vaulot, and N. Simon. 2002. Identification of bacteria associated with dinoflagellates (Dinophyceae) *Alexandrium* spp. using tyramide signal amplification-fluorescent *in situ* hybridization and confocal microscopy. J. Phycol. 38:404-411.

Bird, D. and J. Kalff. 1984. Empirical relationship between bacterial abundance and chlorophyll concentrations in fresh and marine waters. Can. J. Fish. Aquat. Sci. 41:1015-1023.

Blackstock, W. P., and M. P. Weir. 1999. Proteomics: quantitative and physical mapping of cellular proteins. Trends. Biotechnol. 17:121-127.

Boesch, D. F., R. B. Brinsfield, R. E. Magnien. 2001. Chesapeake Bay eutrophication: scientific understanding, ecosystem restoration, and challenges for agriculture. J. Environ. Qual. 30:303-320.

Boicourt, W. C. 1992. Influences of circulation processes on dissolved oxygen in the Chesapeake Bay. In D. E. Smith, M. Leffler and G. Mackiernan [*eds*.], Oxygen dynamics in the Chesapeake Bay. Maryland Sea Grant.

Bouvier, T. C., and P. A. del Giorgio. 2002. Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. Limnol. Oceanogr. 47: 453-470.

Brett, M. T., F. S. Lubnow, M. Villar-Argaiz, A. Muller-Solger, and C. R. Goldman. 1999. Nutrient control of bacterioplankton and phytoplankton dynamics. Aqua. Ecol. 33:135-145.

Brinkmeyer, R., K. Knittel, J. Jurgens, H. Weyland, R. Amann, and E. Helmke. 2003. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. Appl. Environ. Microbiol. 69:6610-6619.

Britschgi, T. B., and S. J. Giovannoni. 1991. Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. Appl. Environ. Microbiol. 57:1707-1713.

Brown, M. V., and J. A. Fuhrman. 2005. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. Aquat. Microb. Ecol. 41:15-23.

Buchan, A., J. M. Gonzalez, and M. A. Moran. 2005. Overview of the marine *Roseobacter* lineage. Appl. Environ. Microbiol. 71:5665-5677.

Button, D. K., F. Schut, P. Quang, R. Martin, and B. R. Robertson. 1993. Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. Appl. Environ. Microbiol. 59:881-891.

Campbell, B. J., and S. C. Cary. 2001. Characterization of a novel Spirochete associated with the hydrothermal vent polychaete annelid, Alvinella pompejana. Appl. Environ. Microbiol. 67:110-117.

Campbell L, H. A. Nolla, and D. Vaulot. 1994. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. Limnol. Oceanogr. 39:954-961.

Caron, D. A., E. L. Lim, R. W. Sanders, M. R. Dennett, and U. -G. Berninger. 2000. Responses of bacterioplankton and phytoplankton to organic carbon and inorganic nutrient additions in contrasting oceanic ecosystems. Aquat. Microb. Ecol. 22:175-184.

Cary, S. C., M. T. Cottrell, J. L. Stein, F. Camacho, and D. Desbruyères. 1997. Molecular identification and localization of a filamentous symbiotic bacteria associated with the hydrothermal vent annelid, *Alvinella pompejana*. Appl. Environ. Microbiol. 63:1124-1130.

Casamayor, E. O., H. Schafer, L. Baneras, C. Pedros-Alio, and G. Muyzer. 2000. Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 66:499-508.

Castle, D., and D. L. Kirchman. 2004. Composition of estuarine bacterial communities assessed by denaturing gradient gel electrophoresis and fluorescence in situ hybridization. Limnol. Oceanogr.: Methods. 2:303-314.

Chao, A., and T. -J. Shen. 2003-2005. Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide published at http://chao.stat.nthu.edu.tw.

Chen, F., C. A., Suttle, and S. M. Short. 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. 62:2869-2874.

Chen, F., J. Lu, B. J. Binder, Y. Liu, and R. E. Hodson. 2001. Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR Gold. Appl. Environ. Microbiol. 67:539-545

Chen, F., K. Wang, J. Kan, M. T. Suzuki, and K. E. Wommack. 2006. Diverse and unique picocyanobacteria in the Chesapeake Bay revealed by 16S-23S rRNA internal transcribed spacer sequences. Appl. Environ. Microbiol. 72:2239-43.

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

Cho, J. C., and S. J. Giovannoni. 2004b. Cultivation and growth characteristics of a diverse group of oligotrophic marine γ-*Proteobacteria*. Appl. Environ. Microbiol. 70:432-440.

Clarke, K. R. 1993. Nonparametric multivariate analyses of changes in community structure. Austra. J. Ecol. 18: 117-143.

Cole, J. J., G. E. Likens, and D. L. Strayer. 1982. Photosynthetically produced dissolved organic carbon: an important carbon source for planktonic bacteria. Limnol. Oceanogr. 27:1080-1090.

Cole, J. J., S. Findlay, and M. L. Pace. 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Mar. Ecol. Prog. Ser. 43:1-10.

Coleman, M. L., M. B. Sullivan, A. C. Martiny, C. Steglich, K. Barry, E. F. DeLong, S. W. Chisholm. 2006. Genomic islands and the ecology and evolution of *Prochlorococcus*. Science. 311:1768-1770.

Conway, T, and G. K. Schoolnik. 2003. Microarray expression profiling: capturing a genome-wide protrait of the transcriptome. Mol. Microbiol. 47:879-889.

Corno, G., and K. Jürgens. 2006. Direct and indirect effects of protist predation on population size structure of a bacterial strain with high phenotypic plasticity. Appl. Environ. Microbiol. 72:78-86.

Cottrell, M. T. and D. L. Kirchman. 2000a. Natural assemblages of marine Proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. Appl. Environ. Microbiol. 66:1692-1697.

Cottrell, M. T., and D. L. Kirchman. 2000b. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Appl. Environ. Microbiol. 66:5116-5122.

Cottrell, M. T., L. A. Waidner, L. Yu, and D. L. Kirchman. 2005. Bacterial diversity of metagenomic and PCR libraries from the Delaware River. Environ. Microbiol. 7:1883-1895.

Crosby, L. D, and C. S. Criddle. 2003. Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. Biotechniques. 34:790-802.

Crump, B. C., C. S. Hopkinson, M. L. Sogin, and J. E. Hobbie. 2004. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. Appl. Environ. Microbiol. 70:1494-1505.

Crump, B. C., E. V. Armbrust, and J. A. Baross. 1999. Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. Appl. Environ. Microbiol. 65: 3192-3204.

Crump, B. C., G. W. Kling, M. Bahr, and J. E. Hobbie. 2003. Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. Appl. Environ. Microbiol. 69: 2253-2268.

Currie, D. J. 1990. Large-scale variability and interactions among phytoplankton, bacterioplankton, and phosphorus. Limnol. Oceanogr. 35:1437-1455.

Curtis, T. P., W. T. Sloan, and J. W. Scannell. 2002. Estimating prokaryotic diversity and its limits. Proc. Natl. Acad. Sci. USA. 99:10494-10499.

Daley, R. J., and J. E. Hobbie. 1975. Direct counts of aquatic bacteria by a modified epifluorescent technique. Limnol. Oceanogr. 20:875-882.

de la Torre, J. R., L. M. Christianson, O. Béjà, M. T. Suzuki, D. M. Karl, J. Heidelberg, E. F. DeLong. 2003. Proteorhodopsin genes are distributed among divergent marine bacterial taxa. Proc. Natl. Acad. Sci. USA. 100:12830-12835.

DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. -U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science. 311:496-503.

DeLong, E. F., D. G. Franks, and A. L. Alldredge. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. 38:924-934.

del Giorgio, P. A., and T. C. Bouvier. 2002. Linking the physiological and phylogenetic succession in free-living bacterial communities along an estuarine salinity gradient. Limnol. Oceanogr. 47:471-486.

Diez, B., C. Pedros-Alio, T. L. Marsh, and R. Massana. 2001. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Appl. Environ. Microbiol. 67:2942-2951.

Doucette, G. J., M. Kodama, S. Franca, and S. Gallacher. 1998. Bacterial interactions with harmful algal bloom species: bloom ecology, toxigenesis and cytology. In D. M. Andersen, A. D. Cembella & G. M. Hallegraeff (*eds*.) Physiological Ecology of Harmful Algal Blooms. Springer-Verlag, Berlin, p 619-647.

Ducklow, H. W. and F. K. Shiah. 1993. Estuarine bacterial production. In: T. Ford (*ed*), Aquatic Microbiology: An ecological approach. Blackwell, London, pp. 261-284.

Ducklow, H. W., G. Schultz, P. Raymond, J. Bauer, and F. K. Shiah. 1999. Bacterial dynamics in large and small estuaries. Proceedings of the 8th International Symposium on Microbial Ecology, Halifax, Canada.

Dutton, C. M., C. Paynton, and S. Sommer. 1993. General method for amplifying regions of very high G+C content. Nucleic. Acids. Res. 21:2953-2954.

Edmond, J. M., and K. L. von Damm. 1985. Chemistry of ridge crest hot springs. Bull. Biol. Soc. Wash. 6:43-47.

Eymann, C, G. Homuth, C. Scharf, and M. Hecker. 2002. *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J. Bacteriol. 184:2500-2520.

Fandino, L. B., L. Riemann, G. F. Steward, R. A. Long, and F. Azam. 2001. Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. Aquat. Microb. Ecol. 23:119-130.

Farrelly, V., F. A. Rainey, and E. Stachebrandt. 1995. Effects of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. 61:2798-2801.

Faulkner, D. J. 2002. Marine natural products. Nat. Prod. Rep. 19:1-48.

Ferris, M. J., G. Muyzer, and D. M. Ward. 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. 62:340-346.

Fichmann, J., and R. Westermeier. 1999. 2-D protein gel electrophoresis. In: A. J. Link (*ed.*), 2-D proteome analysis protocols. Humana Press, New Jersey, p 1-8.

Findlay, S., M. L. Pace, D. Lints, J. J. Cole, N. F. Caraco, and B. Peierls. 1991. Weak coupling of bacterial and algal production in a heterotrophic ecosystem: the Hudson River estuary. Limnol. Oceanogr. 36:268-278.

Fisher, M. M., and E. W. Triplett. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl. Environ. Microbiol. 65:4630-4636.

Fuhrman, J. A., J. W. Ammerman, and F. Azam. 1980. Bacterioplankton in the coastal eutrophic zone: distribution, activity and possible relationships with phytoplankton. Mar. Biol. 60:201-207.

Fuhrman, J. A., and C. A. Suttle. 1993. Viruses in marine planktonic systems. Oceanogr. 6:51-63.

Fuhrman, J. A., I. Hewson, M. Brown, M. S. Schwalbach, J. A. Steele, and A. Patel. 2004. Microbial biogeography and temporal patterns in marine plankton. 10th International Symposium on Microbial Ecology, Cancun, Mexico.

Fuhrman, J. A., K. McCallum, and A. A. Davis. 1993. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. Appl. Environ. Microbiol. 59:1294-1302.

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

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature. 345: 60-63.

Giovannoni, S. J., and M. S. Rappé. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes. In: D. L. Kirchman (*ed*.), Microbial ecology of the oceans. Wiley-Liss, New York, p 47-84.

Giovannoni, S. J., H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, D. Baptista, L. Bibbs, J. Eads, T. H. Richardson, M. Noordewier, M. S. Rappé, J. M. Short, J. C. Carrington, and E. J. Mathur. 2005. Genome Streamlining in a cosmopolitan oceanic bacterium. Science. 309:1242-1245.

Glibert, P. M., D. J. Conley, T. R. Fisher, L. W. J. Harding, and T. C. Malone. 1995. Dynamics of the 1990 Winter/Spring bloom in Chesapeake Bay. Mar. Ecol. Prog. Ser. 122:27-43.

Glöckner, F. O., B. M. Fuchs, and R. Amann. 1999. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. Appl. Environ. Microbiol. 65:3721-3726.

Glöckner, F. O., E. Zaichikov, N. Belkova, L. Denissova, J. Pernthaler, A. Pernthaler, and R. Amann. 2000. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of *Actinobacteria*. Appl. Environ. Microbiol. 66:5053-65.

González, J. M., E. B. Sherr, and B. F. Sherr. 1990. Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. Appl. Environ. Microbiol. 56:583-589.

González, J. M., and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class *Proteobacteria* in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

González, J. M., R. Simo, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedros-Alio, and M. A. Moran. 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl. Environ. Microbiol. 66:4237-4246.

Goodchild, A., N. F. W. Saunders, H. Ertan, M. Raftery, M. Guilhaus, P. M. G. Curmi, and R. Cavicchioli. 2004. A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii*. Mol. Microbiol. 53:309-321.

Guindon, S., and O. Gascuel. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52:696-704.

Habermann, B., J. Oegeman, S. Sunyaev, and A. J. Shevchenko. 2004. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Mol. Cell. Proteomics. 3:238-249.

Haddad, M. A., F. Camacho, P. Durand, and S. C. Cary. 1995. Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. Appl. Environ. Microbiol. 61:1679-1687.

Hahn, M. W., E. R. B. Moore, and M. G. Höfle. 1999. Bacterial filament formation, a defense mechanism against flagellate grazing, is growth rate controlled in bacteria of different phyla. Appl. Environ. Microbiol. 65:25-35.

Hahn, M. W., E. R. B. Moore, and M. G. Höfle. 2000. Role of microcolony formation in the protistan grazing defense of the aquatic bacterium *Pseudomonas* sp. MWH1. Microb. Ecol. 39:175-185.

Hahn, M. W., and M. G. Höfle. 2001. Grazing of protozoa and its effect on populations of aquatic bacteria. FEMS Microbiol. Ecol. 35:113-121.

Hahn, M. W., H. Lunsdorf, Q. Wu, M. Schauer, M. G. Höfle, J. Boenigk, and P. Stadler. 2003. Isolation of novel ultramicrobacterial classified as *Actinobacteria* from five freshwater habitats in Europe and Asia. Appl. Environ. Microbiol. 69:1442-1451.

Hahn, M. W., and M. Pöckl. 2005. Ecotypes of planktonic *Actinobacteria* with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. Appl. Environ. Microbiol. 71:766-773.

Harding, L. W. 1994. Long-term trends in the distribution of phytoplankton in Chesapeake Bay - roles of light, nutrients and streamflow. Mar. Ecol. Prog. Ser. 104:267-291.

Harding, L. W., M. E. Mallonee, and E. S. Perry. 2002. Toward a predictive understanding of primary productivity in a temperate, partially stratified estuary. Estuar. Coast. Shelf. S. 55:437-463.

Haygood, M. G., E. W. Schmidt, S. K. Davidson, and D. J. Faulkner. 1999. Microbial symbionts of marine invertebrates: opportunities for microbial biotechnology. J. Mol. Microbiol. Biotechnol. 1:33-43.

Heck, K. L., Jr., G. V. Belle, and D. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecol. 56:1459-1461.

Heidelberg, J. F., K. B. Heidelberg, and R. R. Colwell. 2002. Seasonality of Chesapeake Bay bacterioplankton species. Appl. Environ. Microbiol. 68: 5488-5497.

Henriques, I. S., M. A. Almeida, M. A. Cunha, and A. Correia. 2004. Molecular sequence analysis of prokaryotic diversity in the middle and outer sections of the Portuguese estuary Ria de Aveiro. FEMS Microbiol. Ecol. 49:269-279.

Henriques, I. S., A. Alves, M. Tacao, A. Almeida, A. Cunha, and A. Correia. 2006. Seasonal and spatial variability of free-living bacterial community composition along an estuarine gradient (Ria de Aveiro, Portugal). Estuar. Coast. Shelf S. 68:139-148.

Hewson, I., and J. A. Fuhrman. 2004. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. Appl. Environ. Microbiol. 70:3425-33.

Hewson, I., D. M. Winger, K. E. Williamson, J. A. Fuhrman, and K. E. Wommack. 2006. Viral and bacterial assemblage covariance in oligotrophic waters of the West Florida Shelf (Gulf of Mexico). J. Mar. Biol. Ass. U. K. 86:591-603.

Hill, R. T. 2004. Microbes from marine sponges: a treasure trove of biodiversity for natural products discovery. In: A. T. Bull (ed.), Microbial Diversity and Bioprospecting. ASM Press, Washington, DC, pp. 177-190.

Hobbie, J. E., R. J. Daley, and J. Jasper. 1977. Use of Nuclepore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. 33:1225-1228.

Höfle, M. G., H. Haas, and K. Dominik. 1999. Seasonal dynamics of bacterioplankton community structure in a eutrophic lake as determined by 5S rRNA analysis. Appl. Environ. Microbiol. 65: 3164-3174.

Holland, S. 1998. aRarefactWin. University of Georgia, Athens.

Hollibaugh, J. T., P. S. Wong, and M. C. Murrell. 2000. Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. Aquat. Microb. Ecol. 21:103-109.

Hoppe, H. G. 1976. Determination and properties of actively metabolizing heterotrophic bacteria in the sea, investigated by the means of micro-autoradiography. Mar. Biol. 36:291-302.

Humayoun, S. B., N. Bano, and J. T. Hollibaugh. 2003. Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. Appl. Environ. Microbiol. 69:1030-1042.

Hurlbert, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecol. 52:577-586.

Jannasch, H. W., and G. E. Jones. 1959. Bacterial populations in sea water as determined by different methods of enumeration. Limnol. Oceanogr. 4:128-139.

Janssen, P. H., P. S. Yates, B. E. Grinton, P. M. Taylor, and M. Sait. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. Appl. Environ. Microbiol. 68:2391-2396.

Jensen, M. A., J. A. Webster, and N. Straus. 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Appl. Environ. Microbiol. 59:945-952.

Johnson, Z. I., E. R. Zinser, a. Coe, N. P. McNulty, E. Malcolm, S. Woodward, and S. W. Chisholm. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. Science. 311:1737-1740.

Jonas, R. B., and J. H. Tuttle. 1990. Bacterioplankton and organic carbon dynamics in the lower mesohaline Chesapeake Bay. Appl. Environ. Microbiol. 56:747-757.

Jürgens, K., and C. Matz. 2002. Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. Antonie Leeuwenhoek. 81:413-434.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. In: H. N. Munro (*ed*.) Mammalian protein metabolism. Academic Press, New York, p 21–132

Kan, J., and F. Chen. 2004. Co-monitoring bacterial and dinoflagellates communities by denaturing gradient gel electrophoresis (DGGE) and SSU rDNA sequencing during a dinoflagellate bloom. Acta. Oceanol. Sin. 23: 483-492.

Kan, J., B. C. Crump, K. Wang, and F. Chen. 2006a. Bacterioplankton community in the Chesapeake Bay: predictable or random assemblages. Limnol. Oceanogr. 51:2157-2169.

Kan, J., K. Wang, and F. Chen. 2006b. Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). Aquat. Microb. Ecol. 42: 7-18.

Kent, A. D., S. E. Jones, A. C. Yannarell, J. M. Graham, G. H. Lauster, T. K. Kratz, and E. W. Triplett. 2004. Annual patterns in bacterioplankton community variability in a humic lake. Microb. Ecol. 48: 550-560.

Kirchman, D. L., and J. H. Rich. 1997. Regulation of bacterial growth rates by dissolved organic carbon and temperature in the equatorial Pacific Ocean. Microb. Ecol. 33:11-20.

Kirchman, D. L., K. A. Hoffman, R. Weaver, and D. A. Hutchins. 2003a. Regulation of growth and energetic of a marine bacterium by nitrogen source and iron availability. Mar. Ecol. Prog. Ser. 250: 291-296.

Kirchman, D. L., L. Yu, and M. T. Cottrell. 2003b. Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware estuary. Appl. Environ. Microbiol. 69:6587-6596.

Kisand, V., and J. Wikner. 2003. Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. J. Microbiol. Methods. 54:183-191.

Klappenbach, J. A., P. R. Saxman, J. R. Cole, and T. M. Schmidt. 2001. rrndb: the ribosomal RNA operon copy number database. Nucleic. Acids. Res. 29:181-184.

Kopczynski, E. D., M. M. Bateson, and D. M. Ward. 1994. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. 60:746-748.

Lane, D. J. 1991. 16S/23S rRNA sequencing, In E. Stackebrandt and M. Goodfellow (*eds*.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons. p. 115-175.

Le, J., J. D. Wehr, and L. Campbell. 1994. Uncoupling of bacterioplankton and phytoplankton production in fresh waters is affected by inorganic nutrient limitation. Appl. Environ. Microbiol. 60:2086-2093.

Lee, S., and J. A. Fuhrman. 1991. Spatial and temporal variation of natural bacterioplankton assemblages studied by total genomic DNA cross-hybridization. Limnol. Oceanogr. 36:1277-1287.

Lee, Y. K., J. –H. Lee, and H. K. Lee. 2001. Microbial symbiosis in marine sponges. J. Microbiol. 39:254-264.

Li, W. K. W. 1994. Primary productivity of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. Limnol. Oceanogr. 39:169-175.

Linley, E. A. S., R. C. Newell, and M. I. Lucas. 1983. Quantitative relationships between phytoplankton and bacteria and heterotrophic microflagellates in shelf waters. Mar. Ecol. Prog. Ser. 12:77-89.

Liska, A. J., and A. J. Shevchenko. 2003. Combining mass spectrometry with database interrogation strategies in proteomics. Trends. Anal. Chem. 22:291-299.

Liu, W. T., T. L. Marsh, H. Cheng, and L. J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63:4516-4522.

Long, E. O., and I. B. Dawid. 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. 49:727-764.

Lopez, M. F. 1999. Proteome analysis I. Gene products are where the biological action is. J. Chromatogr. 722:191-202.

Lopez-Garcia, P., F. Rodriguez-Valera, C. Pedros-Alio, and D. Moreira. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature. 409:603-607.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Y. Kumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lupmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. P. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363-1371.

Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese. 1997. The RDP (Ribosomal Database Project). Nucleic Acids Res. 25:109-111.

Malone, T. C., W. M. Kemp, H. W. Ducklow, W. R. Boynton, J. H. Tuttle, and R. B. Jonas. 1986. Lateral variation in the production and fate of phytoplankton in a partially stratified estuary. Mar. Ecol. Prog. Ser. 32:149-160.

Malone, T. C., H. W. Ducklow, E. R. Peele, and S. E. Pike. 1991. Picoplankton carbon flux in Chesapeake Bay. Mar. Ecol. Prog. Ser. 78: 11-22.

Mann, M., A. Shevchenko, M. Wilm, and O. Vorm. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68:850-858.

Marvalin, O., L. Aleya, and H. J. Hartman. 1989. Coupling of the seasonal patterns of bacterioplankton and phytoplankton in a eutrophic lake. Can. J. Microbiol. 35:706-712.

McDaniel, L., L. Houchin, S. Williamson, and J. H. Paul. 2002. Lysogeny in natural populations of marine *Synechococcus*. Nature. 415:496

Methe, B. A., K. E. Nelson, J. W. Deming, B. Momen, E. Melamud, X. Zhang, J. Mouldt, R. Madupu, W. C. Nelson, R. J. Dodson, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, R. T. Deboy, J. F. Kolonay, S. A. Sullivan, L. Zhou, T. M. Davidsen, M. Wu, A. L. Huston, M. Lewis, B. Weaver, J. F. Weidman, H. Khouri, T. R. Utterback, T. V. Feldblyum, and C. M. Fraser. 2005. The phschrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analysis. Proc. Natl. Acad. Sci. USA. 102:10913-10918.

Methe, B. A., W. D. Hiorns, and J. P. Zehr. 1998. Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George and six other Adirondack lakes. Limnol. Oceanogr. 43:368-374.

Middelboe, M., and P. G. Lyck. 2002. Regeneration of dissolved organic matter by viral lysis in marine microbial communities. Aquat. Microb. Ecol. 27:187-194.

Momen, B., J. P. Zehr, C. W. Boylen, and J. W. Sutherland. 1999. Determinants of summer nitrate concentration in a set of Adirondack lakes, New York. Water. Air. Soil. Poll. 111:19-28.

Monsen, N. E., J. E. Cloern, L. V. Lucas, and S. G. Monismith. 2002. A comment on the use of flushing time, residence time, and age as transport time scales. Limnol. Oceanogr. 47:1545-1553.

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

Moran, M. A., A. Buchan, J. M. Gonzalez, J. F. Heidelberg, W. B. Whitman, R. P. Kiene, J. R. Henriksen, G. M. King, R. Belas, C. Fuqua, L. Brinkac, M. Lewis, S. Johri, B. Weaver, G. Pai, J. A. Eisen, E. Rahe, W. M. Sheldon, W. Y. Ye, T. R. Miller, J. Carlton, D. A. Rasko, I. T. Paulsen, Q. H. Ren, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, M. J. Rosovitz, D. H. Haft, J. Selengut, and N. Ward. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature. 432:910-913.

Morris, R. M., K. L. Vergin, J. C. Cho, M. S. Rappé, C. A. Carlson, and S. J. Giovannoni. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. Limnol. Oceanogr. 50:1687-1696.

Morris, R. M., M. S. Rappe, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A. Carlson, and S. J. Giovannoni. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature. 420:806-810.

Muller, D. G., M. Sengco, S. Wolf, M. Brautigam, C. E. Schmid, M. Kapp, and R. Knippers. 1996. Comparison of two DNA viruses infecting the marine brown algae *Ectocarpus siliculosus* and *E. fasciculatus*. J. Gen. Virol. 77:2329-2333.

Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceanogr. 40:148-158.

Murray, A. E., C. M. Preston, R. Massana, L. T. Taylor, A. Blakis, K. Wu, and E. F. DeLong. 1998. Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. Appl. Environ. Microbiol. 64:2585-2595.

Murray, A. E., J. T. Hollibaugh, and C. Orrego. 1996. Phylogenetic compositions of bacterioplankton from two California estuaries compared by denaturing gradient gel electrophoresis of 16S rDNA fragments. Appl. Environ. Microbiol. 62:2676-2680.

Muyzer, G., E. C. de Waal, and A. G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59:695-700.

Muyzer, G. 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Curr. Opin. Microbiol. 2:317-322.

Nakagawa, T, J. Ishibashi, A. Maruyama, T. Yamanaka, Y. Morimoto, H. Kimura, T. Urabe, and M. Fukui. 2004. Analysis of dissimilatory sulfite reductase and 16S

rRNA gene fragments from deep-sea hydrothermal sites of the Suiyo Seamount, Izu-Boniin Arc, Western Pacific. Appl. Environ. Microbiol. 70:393-403.

Nixon, S. W., J. W. Ammerman, L. P. Atkinson, V. M. Berounsky, G. Billen, W. C. Boicourt, W. R. Boynton, T. M. Church, D. M. Ditoro, R. Elmgren, J. H. Garber, A. E. Giblin, R. A. Jahnke, N. J. P. Owens, M. E. Q. Pilson, and S. P. Seitzinger. 1996. The fate of nitrogen and phosphorus at the land sea margin of the North Atlantic Ocean. Biogeochemistry. 35:141-180.

Noble, P. A., K. D. Bidle, and M. Fletcher. 1997. Natural microbial community compositions compared by a back-propagating neural network and cluster analysis of 5S rRNA. Appl. Environ. Microbiol. 63:1762-1770.

O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250:4007-4021.

Ogunseitan, O. A. 1997. Direct extraction of catalytic proteins from natural microbial communities. J. Microbiol. Methods. 28:55-63.

Ogunseitan, O. A. 1998. Extraction of proteins from aquatic environments. Mol. Microbial. Ecol. Mann. 4:1-12.

Oldach, D., C. Delwiche, K. Jakobsen, T. Tengs, E. Brown, J. Kempton, E. Schaefer, H. Bowers, H. J. Glasgow, J. Burkholder, K. Steidinger, and P. Rublee. 2000. Heteroduplex mobility assay-guided sequence discovery: elucidation of the small subunit (18S) rDNA sequences of *Pfiesteria piscicida* and related dinoflagellates from complex algal culture and environmental sample DNA pools. Proc. Natl. Acad. Sci. USA. 97:4303-4308.

Ostrowski, M., F. Fegatella, V. Wasinger, M. Guilhaus, G. L. Corthals, and R. Cavicchioli. 2004. Cross-species identification of proteins from proteome profiles of the marine oligotrophic ultramicrobacterium, *Sphigopyxis alaskensis*. Proteomics. 4:1779-1788.

Øvreås, L., D. Bourne, R. A. Sandaa, E. O. Casamayor, S. Benlloch, V. Goddard, G. Smerdon, M. Heldal, and T. F. Thingstad. 2003. Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms. Aqua. Microb. Ecol. 31:109-121.

Øvreås, L., L. Forney, F. L. Daae, and V. Torsvik. 1997. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl. Environ. Microbiol. 63:3367-3373.

Pace, M. L. 1988. Bacterial mortality and the fate of bacterial production. Hydrobiologia. 159:41-50.

Palumbo, A. V., R. L. Ferguson, and P. A. Rublee. 1984. Size of suspended bacterial cells and association of heterotrophic activity with size fractions of particles in estuarine and coastal waters. Appl. Environ. Microbiol. 48:157-164.

Paul, J. H. 1999. Microbial gene transfer: an ecological perspective. J. Mol. Microbiol. Biotechnol. 1:45-50.

Pernthaler, J., F. O. Glöckner, S. Unterholzner, A. Alfreider, R. Psenner, and R. Amann. 1998. Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. Appl. Environ. Microbiol. 64:4299-4306.

Petersohn, A., M. Brigulla, S. Haas, J. D. Hoheisel, U. Volker, and M. Hecker. 2001. Global analysis of the general stress response of *Bacillus subtilis*. J. Bacteriol. 183:5617-5631.

Pinhassi, J., F. Azam, J. Hempala, R. A. Long, J. Martinez, U. L. Zweifel, and A. Hagström. 1999. Coupling between bacterioplankton species composition, population dynamics, and organic matter degradation. Aquat. Microb. Ecol. 17:13-26.

Pinhassi, J., and A. Hagström. 2000. Seasonal succession in marine bacterioplankton. Aquat. Microb. Ecol. 21:245-256.

Pinhassi, M., M. Sala, H. Havskum, F. Peters, O. Guadayol, A. Malits, and C. Marrase. 2004. Changes in bacterioplankton composition under different phytoplankton regimens. Appl. Environ. Microbiol. 70:6753-6766.

Pomeroy, L. R., and W. J. Wiebe. 2001. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. Aquat. Microb. Ecol. 23:187-204.

Porter, K. G., and Y. S. Feig. 1980. The use of DAPI for identifying and counting aquatic microflora. Limnol. Oceanogr. 25:943-948.

Posada, D., and K. A. Crandell. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817-818.

Pradet-Balade, B., F. Boulme, H. Beug, E. W. Mullner, and J. A. Garcia-Sanz. 2001. Translation control: bridging the gap between genomics and proteomics? Trends. Biochem. Sci. 26:225-229.

Proctor, L. M., and J. A. Fuhrman. 1990. Viral mortality of marine bacteria and cyanobacteria. Nature. 343:60-62.

Ram, R. J., N. C. VerBerkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker, R. C. Blake II, M. Shah, R. L. Hettich, and J. F. Banfield. 2005. Community proteomics of a natural microbial biofilm. Science. 308:1915-1920.

Rappé, M. S., M. T. Suzuki, K. L. Vergin, and S. J. Giovannoni. 1998. Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. Appl. Environ. Microbiol. 64:294-303.

Rappé, M. S., P. F. Kemp, and S. J. Giovannoni. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol. Oceanogr. 42:811-826.

Rappé, M. S., S. A. Cannon, K. L. Vergin,, and S. J. Giovannoni. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature. 418:630-633.

Ravenschlag, K., K. Sahm, J. Pernthaler, and R. Amann. 1999. High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. 65:3982-3989.

Ray, R. T., L. W. Haas, and M. E. Sieracki. 1989. Autotrophic picoplankton dynamics in a Chesapeake Bay sub-estuary. Mar. Ecol. Prog. Ser. 52:273-285.

Riemann, L., G. F. Steward, L. B. Fandino, L. Campbell, M. R. Landry, and F. Azam. 1999. Bacterial community composition during two consecutive NE monsoon periods in the Arabian Sea studied by denaturing gradient gel electrophoresis (DGGE) of rRNA genes. Deep. Sea. Res. II 46:1791-1811.

Riemann, L., G. F. Steward, and F. Azam. 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl. Environ. Microbiol. 66: 578-587.

Riesenfeld, C. S., P. D. Schloss, and J. Handelsman. 2004. Metagenomics: genomic analysis of microbial communities. Annu. Rev. Genet. 38:525-552.

Robarts, R. D., M. T. Arts, M. S. Evans, and M. J. Waiser. 1994. The coupling of heterotrophic bacterial and phytoplankton production in a hypertrophic, shallow prairie lake. Can. J. Fish. Aquat. Sci. 51:2219-2226.

Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature. 424:1042-1047.

Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisholm. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl. Environ. Microbiol. 68:1180-1191.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gulman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66:2541-2547.

Ruger, H. J., and M. G. Höfle. 1992. Marine star-shaped-aggregate-forming bacteria: *Agrobacterium atlanticum* sp. nov.; *Agrobacterium meteori* sp. nov.; *Agrobacterium ferrugineum* sp. nov., nom. rev.; *Agrobacterium gelatinovorum* sp. nov., nom. rev; and *Agrobacterium stellulatum* sp. nov., nom. rev. Int. J. Syst. Bacteriol. 42:133-143.

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

SAS Institute Inc. 1992. SAS user's guide. SAS Institute Inc.

Schleifer, K. H., and E. Stackebrandt. 1983. Molecular systematics of prokaryotes. Annu. Rev. Microbiol. 37:143-187.

Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71:1501-1506.

Schmidt, T. M., E. F. Delong, and N. R. Pace. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173:4371-4378.

Schultz, G. E. Jr., and H. Ducklow. 2000. Changes in bacterioplankton metabolic capabilities along a salinity gradient in the York River estuary, Virginia, USA. Aquat. Microb. Ecol. 22:163-174.

Schulze, W. X., G. Gleixner, K. Kaiser, G. Guggenberger, M. Mann, and E. -D. Schulze. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia. 142:335-343.

Sekiguchi, H., M. Watanabe, T. Nakahara, B. Xu, and H. Uchiyama. 2002. Succession of bacterial community structure along the Changjiang River determined by denaturing gradient gel electrophoresis and clone library analysis. Appl. Environ. Microbiol. 68:5142-5150.

Selje, N., and M. Simon. 2003. Composition and dynamics of particle-associated and free-living bacterial communities in the Weser estuary, Germany. Aquat. Microb. Ecol. 30:221-237.

Selje, N., M. Simon, and T. Brinkhoff. 2004. A newly discovered Roseobacter cluster in temperate and polar oceans. Nature. 427:445-448.

Sellner, K. G., S. G. Sellner, R. V. Lacouture, and R. E. Magnien. 2001. Excessive nutrients select for dinoflagellates in the stratified Patapsco River estuary: Margalef reigns. Mar. Eco. Prog. Ser. 220:93-102.

Shepard, J. L., B. Olsson, M. Tedengren, and B. P. Bradley. 2000. Protein expression signatures identified in *Mytilus edulis* exposed to PCBs, copper, and salinity stress. Mar. Environ. Res. 50:337-340.

Sherr, B. F., E. B. Sherr, and J. McDaniel. 1992. Effect of protistan grazing on the frequency of dividing cells in bacterioplankton assemblages. Appl. Environ. Microbiol. 58:2381-2385.

Sherr, E. B., and B. F. Sherr. 2002. Significance of predation by protists in aquatic microbial food webs. Antonie Leeuwenhoek 81:293-308.

Shevchenko, A., S. Sunyaev, A. Liska, and P. Bork. 2003. Nanoelectrospray tandem mass spectrometry and sequence similarity searching for identification of proteins from organisms with unknown genomes. Methods. Mol. Biol. 211:221-234.

Shevchenko, A., S. Sunyaev, A. Loboda, P. Bork, W. Ens, and K. G Standing. 2001. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. Anal. Chem. 73:1917-1926.

Shewan, J. M., and T. A. McMeekin. 1983. Taxonomy (and ecology) of *Flavobacterium* and related genera. Annu. Rev. Microbiol. 37:233-252.

Shiah, F. -K., and H. W. Ducklow. 1994. Temperature regulation of heterotrophic bacterioplankton abundance, production, and specific growth rate. Limnol. Oceanogr. 39:1243-1258.

Sievert, S. M., T. Brinkhoff, G. Muyzer, W. Ziebis, and J. Kuever. 1999. Spatial heterogeneity of bacterial populations along an environmental gradient at a shallow submarine hydrothermal vent near Milos Island (Greece). Appl. Environ. Microbiol. 65:3834-3842.

Silva, E. S. 1978. Endonuclear bacteria in two species of dinoflagellates. Protistologica. 14:113-119.

Smith, D. E., M. Leffler, and G. Mackiernan. 1992. Oxygen dynamics in the Chesapeake Bay: a synthesis of recent research. Maryland Sea Grant, College Park, MD.

Smith, D. C., G. F. Steward, R. A. Long, and F. Azam. 1995. Bacterial mediation of carbon fluxes during a diatom bloom in a mesocosm. Deep. Sea. Res. II 42:75-97.

Smith, E. M., and W. M. Kemp. 2003. Planktonic and bacterial respiration along an estuarine gradient: response to carbon and nutrient enrichment. Aquat. Microb. Ecol. 30:251-261.

Sobecky, P. A., M. A. Schell, M. A. Moran, and R. E. Hodson. 1996. Impact of a genetically engineered bacterium with enhanced alkaline phosphatase activity on marine phytoplankton communities. Appl. Environ. Microbiol. 62:6-12.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. USA. 103:12115-12120.

Sohn, J. H., K. K. Kwon, J. H. Kang, H. B. Jung, and S. J. Kim. 2004. *Novosphingobium pentaromativorans* sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediments. Int. J. Syst. Evol. Microbiol. 54:1483-1487.

Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44:846-849.

Staley, J. T., and A. Konopka. 1985. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. 39:321-346.

Stevenson, B. S., S. A. Eichorst, J. T. Wertz, T. M. Schmidt, and J. A. Breznak. 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl. Environ. Microbiol. 70:4748-4755.

Suttle, C. A. 1994. The significance of viruses to mortality in aquatic microbial communities. Microb. Ecol. 28:237-243.

Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625-630.

Suzuki, M. T., C. M. Preston, O. Béjà, J. R. de la Torre, G. F. Steward, and E. F. DeLong. 2004. Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. Microb. Ecol. 48:473-488.

Suzuki, M. T., L. T. Taylor, and E. F. DeLong. 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl. Environ. Microbiol. 66:4605-14.

Suzuki, M. T., M. S. Rappé, Z. W. Haimberger, H. Winfield, N. Adair, J. Strobel, and S. J. Giovannoni. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63:983-989.

Suzuki, M. T., O. Béjà, L. T. Taylor, and E. F. Delong. 2001. Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. Environ. Microbiol. 3:323-331.

Suzuki, M. T., M. S. Rappé, and S. J. Giovannoni. 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl. Environ. Microbiol. 64:4522-4529.

Suzuki, M. T. 1999. Effect of protistan bacterivory on coastal bacterioplankton diversity. Aquat. Microb. Ecol. 20:261-272.

Swofford, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Teske, A., T. Brinkhoff, G. Muyzer, D. P. Moser, J. Rethmeier, and H. W. Jannasch. 2000. Diversity of thiosulfate-oxidizing bacteria from marine sediments and hydrothermal vents. Appl. Environ. Microbiol. 66:3125-33.

Thingstad, T. F., and R. Lignell. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. 13:19-27.

Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Diestel, and M. F. Polz. 2004. Genotypic diversity within a natural coastal bacterioplankton population. Science. 307:1311-1313.

Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin. 2005. Comparative metagenomics of microbial communities. Science 308:554-557.

Troussellier, M., H. Schäfer, N. Batailler, L. Bernard, C. Courties, P. Lebaron, G. Muyzer, P. Servais, J. Vives-Rego. 2002. Bacterial activity and genetic richness along an estuarine gradient (Rhone River plume, France). Aquat. Microb. Ecol. 28:13-24.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 428:37-43.

USEPA. 1983. Method No. 350.1 in Methods for chemical analysis of water and wastes. United States Environmental Protection Agency, Environmental Monitoring and Support Laboratory. Cincinnati, Ohio. Report No. EPA-600/4-79-020.

van Hannen, E. J., G. Zwart, M. P. van Agterveld, H. J. Gons, J. Ebert, and H. J. Laanbroek. 1999. Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses. Appl. Environ. Microbiol. 65:795-801.

Venter, J.C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 304:66-74.

von Wintzingerode, F., U. B. Gobel, and E. Stachebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21:213-229.

Wang, K., and F. Chen. 2004. Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. Aquat. Microb. Ecol. 34:105-116.

Ward, B. B. 1982. Oceanic distribution of ammonium-oxidizing bacteria determined by immunofluorescent assay. J. Mar. Res. 40:1155-1172.

Ward, B. B. and G. D. O'Mullan. 2002. Worldwide distribution of *Nitrosococcus oceani*, a marine ammonia-oxidizing gamma-proteobacterium, detected by PCR

and sequencing of 16S rRNA and amoA genes. Appl. Environ. Microbiol. 68:4153-4157.

Ward, D. M., R. Weller, and M. M. Bateson. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature. 345:63-65.

Ward, D. M. 2006. A macrobiological perspective on microbial species. Microbe. 1:269-278

Warnecke, F., R. Amann, and J. Pernthaler. 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ. Microb. 6:242-253.

Wasinger, V. C., S. J. Cordwell, A. Cerpa Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith. 1995. Progress with gene product mapping of Moli cutes: *Mycoplasma genitalium*. Electrophoresis. 16:1090-1094.

Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37:463-464.

Weinbauer, M. G., and F. Rassoulzadegan. 2004. Are viruses driving microbial diversification and diversity? Environ. Microb. 6:1-11.

White, P. A., J. Kalff, J. B. Rasmussen, and J. M. Gasol. 1991. The effect of temperature and algal biomass on bacterial production and specific growth rate in freshwater and marine habitats. Microb. Ecol. 10:137-149.

Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. USA. 95:6578-6583.

Wikner, J., and A. Hagström. 1991. Annual study of bacterioplankton community dynamics. Limnol. Oceanogr. 36:1313-1324.

Wilkinson, C. R. 1978. Microbial associations in sponges. III. Ultrastructure of the in situ associations of coral reef sponges. Mar. Biol. 49:177-185.

Wilmes, P., and P. L. Bond. 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ. Microbiol. 6:911-920.

Wilmes, P., and P. L. Bond. 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends. Microbiol. 14:92-97.

Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69-114.

Wommack, K. E., J. Ravel, R. T. Hill, J. Chun, and R. R. Colwell. 1999. Population dynamics of Chesapeake Bay virioplankton: total-community analysis by pulsed-field gel electrophoresis. App. Environ. Microbiol. 65:231-240.

Yannarell, A. C., A. D. Kent, G. H. Lauster, T. K. Kratz, and E. W. Triplett. 2003. Temporal patterns in bacterial communities in three temperate lakes of different trophic status. Microb. Ecol. 46:391-405.

Yannarell, A. C., and E. W. Triplett. 2005. Geographic and environmental sources of variation in lake bacterial community composition. Appl. Environ. Microbiol. 71:227-239.

Yu, Z., and M. Morrison. 2004 Comparisons of different hypervariable regions of *rrs* genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 70:4800-4806.

Zickermann, V., M. Bostina, C. Hunte, T. Ruiz, M. Radermacher, and U. Brandt. 2003. Functional implications from an unexpected position of the 49-kDa subunit of NADH:ubiquinone oxidoreductase. J. Biol. Chem. 278:29072-29078.

ZoBell, C. E. 1946. Marine Microbiology. Chronica botanica, Waltham, Mass.

Zubkov, M. V., B. M. Fuchs, H. Eilers, P. H. Burkill, and R. Amann. 1999. Determination of total protein content of bacterial cells by SYPRO staining and flow cytometry. Appl. Environ. Microbiol. 65:3251-3257.

Zubkov, M. V., B. M. Fuchs, S. D. Archer, R. P. Kiene, R. Amann, and P. H. Burkill. 2002. A population of the alph-proteobacteria dominate the bacterioplankton and dimethylsulphoniopropionate uptake after an algal bloom in the North Sea. Deep-Sea Res. II, Top. Sud. Oceanogr. 49:3017-3038.

Zwart, G., B. C. Crump, M. P. Kamst-van Agterveld, F. Hagen, and S. Han. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat. Microb. Ecol. 28:141-155.