

ABSTRACT

Title of Dissertation: SPATIAL AND TEMPORAL
VARIABILITY OF
BACTERIOPLANKTON COMMUNITIES
ACROSS RIVER TO OCEAN
ENVIRONMENTAL GRADIENTS.

Caroline Sara Fortunato, Doctor of
Philosophy, 2012

Dissertation Directed By: Dr. Byron C. Crump
Associate Professor of Oceanography
University of Maryland Center for
Environmental Science

Bacterioplankton communities are deeply diverse within and across environments, yet also display repeatable patterns over seasonal and annual time scales. I assessed patterns of bacterioplankton community variability across the Columbia River coastal margin over space and time. Coastal zones encompass a complex spectrum of environmental gradients, which impact the composition of bacterioplankton communities. Few studies have attempted to address these gradients comprehensively, especially across large spatial and long temporal scales. I generated a 16S rRNA gene-based bacterioplankton community profile of a coastal zone from water samples collected from the Columbia River, estuary, plume, and along coastal transects covering 360 km of the Oregon and Washington coasts and extending to the deep ocean (>2000 m). I collected nearly 600 water samples during four consecutive

years and eleven research cruises. Spatially, bacterioplankton communities separated into seven environments across the coastal zone (ANOSIM, $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom (depth < 350 m), and slope bottom (depth > 850 m). Communities correlated strongly with the structuring physical factors of salinity, temperature, and depth. Within each environment, community variability correlated with factors important to primary and secondary production. In the freshwater-influenced environments of the Columbia River, estuary, and plume, communities varied seasonally and reassembled annually. Freshwater SAR11, Oceanospirillales, and Flavobacteria taxa were indicators of changing seasonal conditions in these environments. In contrast, seasonal change in communities was not detected in the coastal ocean but instead varied spatially with environmental conditions. Each coastal ocean environment had distinct taxa including SAR406 and SUP05 taxa in the deep ocean and *Prochlorococcus* and SAR11 taxa in the upper water column. A survey of metabolic potential (metagenomics) and gene expression (metatranscriptomics) across the salinity gradient showed that although communities were taxonomically distinct, the metabolic potential of these communities was highly similar. Additionally, gene expression patterns were extremely different and reflected the short-time scales on which microbial processes persist in an environment. Across the coastal zone, bacterioplankton communities were taxonomically distinct but metabolically similar, structured by physical factors, and predictable across seasons from river to ocean.

SPATIAL AND TEMPORAL VARIABILITY OF BACTERIOPLANKTON
COMMUNITIES ACROSS RIVER TO OCEAN ENVIRONMENTAL
GRADIENTS

By

Caroline Sara Fortunato

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
2012

Advisory Committee:

Dr. Byron C. Crump, Chair

Dr. Raleigh R. Hood

Dr. Judith M. O'Neil

Dr. David L. Kirchman

Dr. Peter Zuber

Dr. Maile C. Neel, Dean's Representative

© Copyright by
Caroline Sara Fortunato
2012

Acknowledgements

The entirety of this work was made possible through National Science Foundation funding of the Center for Coastal Margin Observation & Prediction (CMOP), granted to Antonio Baptista. I am extremely lucky to be part of this large, long-term project, and to have the opportunity to carry out my research with this generous funding source.

This research included an extensive amount of ship-time in the Pacific Ocean, and our sampling efforts would not have been possible without the work of the captain and crews of the *R/V Wecoma*, *R/V New Horizon*, *R/V Barnes*, and *R/V Point Sur*. In addition, the planning and execution of cruises by the many chief scientists of our thirteen research cruises was invaluable. Sample collection would have not been possible without the hard work of the many graduate and undergraduate students, interns, and technicians. Specifically from the CMOP team I would like to thank Joe Needoba and Tawnya Peterson for providing me with environmental data, sampling assistance, and most importantly advice. I would like to give special thanks to Lydie Herfort who has been a helpful co-author, chief-scientist, and mentor these past five years. I would also like to thank the CMOP Saturn observatory team, especially Charles Seaton and Alex Jaramillo, who helped to compile our extensive environmental dataset, worked with us to develop modeling tools for our data, and helped to answer my many questions. Finally, I would like to thank Suzanne DeLorenzo, my bunk-mate, fellow water-filterer, confidant, and good friend without whom I would not have made it through many weeks at sea.

This research would not have been possible without the help and support of our sequencing centers at Engencore at the University of South Carolina, specifically Joe Jones and John Busch, and Integrated Genomic Services (IGS) at the University of Maryland, specifically Brandi Cantarel, Jerry Liu, and Luke Tallon. Patrick Schloss and Rob Knight provided additional help and advice with analysis. Mary Ann Moran, Brandon Satinsky, and Christa Smith from the University of Georgia provided invaluable guidance and advice for our genomic work.

I also want to thank the Crump lab, all past and present members, especially Jen Nannen, Mary Doherty, Colleen Kellogg, and Michelle Stuart who helped with lab work, gave me invaluable advice, let me complain, and made me laugh more times than I can count. I would also like to thank my fellow Horn Pointers, who are my family away from home.

A huge thanks to my committee, Raleigh Hood, David Kirchman, Peter Zuber, and Judy O'Neil who have been supportive and flexible and given me invaluable advice and guidance. I also want to thank my family and friends, especially my fiancé Vince and my parents, who have listened to countless hours of science and given me support and advice even without fully understanding what I was talking about. And finally I would like to my advisor, Byron Crump, who from day one has been available to help with anything and everything, from random chats about data, to long discussions about metacommunity theory, to help with lab work, and to editing papers at lightning speed, he was there. Byron has provided countless hours of guidance, advice, editing, support, and most importantly patience, and for that I am forever grateful.

Table of Contents:

Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables.....	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
Community variation across environmental gradients.....	3
Spatial and temporal community variability.....	6
Ecological roles of individual taxa.....	9
Linking structure and function.....	12
The Columbia River coastal margin.....	17
Chapter questions and hypotheses.....	19
Chapter 2: Bacterioplankton community variation across river to ocean environmental gradients.....	21
Introduction.....	22
Methods.....	25
Results.....	30
Discussion.....	33
Tables.....	39
Figures.....	41
Chapter 3: Spatial variability overwhelms seasonal patterns bacterioplankton communities across a river to ocean gradient.....	44
Introduction.....	45
Methods.....	49
Results.....	53
Discussion.....	57
Tables.....	64
Figures.....	65
Supplemental Tables.....	71
Supplemental Figures.....	72
Chapter 4: Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin.....	74
Introduction.....	76
Methods.....	79
Results.....	84
Discussion.....	88
Tables.....	95
Figures.....	97
Supplemental Tables.....	103
Supplemental Figures.....	105
Chapter 5: Comparison of gene abundance and expression patterns across an estuarine salinity gradient.....	107
Introduction.....	109
Methods.....	112

Results.....	116
Discussion.....	122
Tables.....	131
Figures.....	133
Chapter 6: Final conclusions.....	141
Complete reference list.....	149

List of Tables

Chapter 2

Table 2.1: Percent similarity values within and between groups.....	39
Table 2.2: BV-STEP and BIO-ENV Spearman rank coefficients showing correlation between location-specific groups and environmental variables	40

Chapter 3

Table 3.1: Percent similarity values within and between groups.....	64
Table 3.S1: Most abundant OTUs within each group based on average relative abundance	71

Chapter 4

Table 4.1: BV-STEP and BIO-ENV Spearman rank coefficients showing correlation between location-specific groups and environmental variables	95
Table 4.2: BV-STEP and BIO-ENV Spearman rank coefficients showing correlation between taxonomic groups and environmental variables	96
Table 4.S1: Top ten indicator taxa for each of the seasonal groups of the river, estuary, and plume	103
Table 4.S2: Top ten indicator taxa for nearshore and offshore communities in the epipelagic	104

Chapter 5

Table 5.1: Environmental and sequence information for each sample.....	131
Table 5.2: Gene function diversity and richness.....	132

List of Figures

Chapter 2

Figure 2.1: Map of Oregon and Washington coasts	41
Figure 2.2: MDS diagram of percent similarities for all 71 samples.....	42
Figure 2.3: MDS diagram of epipelagic (A), shelf bottom (B), and estuary (C).....	43

Chapter 3

Figure 3.1: Map of Oregon and Washington coasts	65
Figure 3.2: MDS diagram of percent similarities for all 300 samples.....	66
Figure 3.3: Correlation of Dimension 1 for the 300 samples from Figure 3.2 and salinity.....	67
Figure 3.4: Seasonal multidimensional scaling diagram of river, estuary, and plume.....	68
Figure 3.5: Percentage of sequences in OTUs classified by location.....	69
Figure 3.6: Average Chao1 index per group.....	70
Figure 3.S1: Percentage of sequences for each spatial group classified by taxonomic group	72
Figure 3.S2: Spatial distribution of the most abundant OTUs based on average relative abundance	73

Chapter 4

Figure 4.1: Seasonal multidimensional scaling diagram of river samples.....	97
Figure 4.2: Seasonal multidimensional scaling diagram of estuary samples.....	98
Figure 4.3: Seasonal multidimensional scaling diagram of plume samples	99
Figure 4.4: Taxonomic composition of each environment	100
Figure 4.5: Bubble plot of top indicator OTUs in each environment	101
Figure 4.6: Correlations of the relative abundance of top indicator OTUs and environmental variables for each environment.....	102
Figure 4.S1: Map of Oregon and Washington coasts	105
Figure 4.S2: Multidimensional scaling diagram of all 596 samples.....	106

Chapter 5

Figure 5.1: Map of Columbia River, estuary and plume sampling locations	133
Figure 5.2: Shared gene functions between the metagenome and metatranscriptome for each sample	134
Figure 5.3: Shared gene functions between samples for metagenomes and metatranscriptomes	135
Figure 5.4: Taxonomy of 16S rRNA genes identified in each metagenome and corresponding 16S amplicon sample	136
Figure 5.5: Relative abundance of sequenced genomes in each metagenome and metatranscriptome.....	137

Figure 5.6: Relatedness of metagenomes and metatranscriptomes	138
Figure 5.7: Abundance of functional genes present and expressed across all samples.....	139

CHAPTER 1:
INTRODUCTION

Bacteria and other microbes control most of the biogeochemical cycles that shape all ecosystems, and they do this with a wide range of metabolic abilities that are not available to macroscopic organisms. Taxonomically, microbial communities vary across habitats, and diversity within and between these habitats changes across scales of millimeters to thousands of kilometers (Fierer 2008). However, the patterns of this diversity and the mechanisms that create and sustain this diversity remain largely unknown. One way to explore these mechanisms is with biogeographical surveys. Biogeography is the study of organism diversity across space and time and shows where organisms live, at what abundance, and why (Martiny et al. 2006). Microbial biogeography has been studied on small scales for decades, but recent advancements in molecular methods have given us the ability to study microbial diversity patterns more deeply and more widely across environments and over time (Martiny et al. 2006). The objective of microbial biogeography is not only to understand the mechanisms that control microbial diversity patterns, but to also understand the functional role of specific organisms, helping to link community structure with community function. Moreover, a combination of microbial biogeographical surveys and measurements of the metabolic potential using genomics provides a complete assessment of functional and taxonomic microbial diversity and allows for the prediction of shifts in microbial diversity and function with environmental change.

This dissertation aims to describe and assess microbial biogeographical patterns over space and time across a large and complex coastal zone, where there exist strong environmental gradients from river to ocean. The addition of data assessing metabolic diversity as well as taxonomic diversity across these gradients

provides a link between microbial structure and function. With this taxonomic and metabolic analysis we gain a better understanding of microbial diversity patterns, the mechanisms behind these patterns, and role these patterns play in ecosystem function.

Community variation across environmental gradients:

Environmental gradients strongly influence bacterioplankton diversity and community composition. Globally, over many different environment types including marine, freshwater, and soil, salinity was found to strongly correlate with changes in bacterial community composition (Lozupone and Knight 2007). Thus estuaries are important environments where freshwater and marine bacterioplankton communities mix and composition can change drastically over short time and spatial scales.

Previous studies along estuarine salinity gradients describe distinct freshwater and marine taxa (Bernhard et al. 2005, Bouvier and Giorgio 2002, Crump et al. 2004, Hewson and Fuhrman 2004, Selje and Simon 2003, Troussellier et al. 2002), although some studies identify taxa that appear to be ubiquitous from river to ocean (Hewson and Fuhrman 2004). In Crump et al. (2004), freshwater and marine bacteria in Plum Island Sound, MA co-occurred with a distinct estuarine community, which formed during summer when estuary residence times exceeded bacterial doubling times. A similar result was seen in Weser estuary, Germany, where well-separated freshwater, brackish, and marine communities were observed (Selje and Simon 2003). Bernhard et al. (2005) demonstrated a switch from freshwater to marine taxa across an estuarine gradient in Tillamook Bay, OR and identified freshwater populations from the rivers that drain into the estuary. Salinity, however, is not necessarily the only factor

influencing bacterial populations along river to ocean gradients, as other factors that co-vary with salinity including temperature, silica and other dissolved inorganic nutrients, and concentrations of dissolved and particulate organic matter also contribute to variation among communities (Bernhard et al. 2005, Bouvier and Giorgio 2002, Stepanauskas et al. 2003).

Across coastal zones, where freshwater flushes into the ocean, salinity influences bacterioplankton community composition but in offshore regions, depth is also a major factor controlling community composition (Lee and Fuhrman 1991), as many specific phylogenetic groups show vertically variable distributions including the SAR11 (Field et al. 1997), Bacteroidetes (Blumel et al. 2007), and Proteobacteria groups (Acinas et al. 1999, Wright et al. 1997). Bacterial diversity also is influenced by depth, as Brown et al. (2009) showed a decrease in bacterial diversity with depth and little overlap in community composition along a depth gradient in the North Pacific.

Brown et al. (2009) measured bacterial diversity as the abundance of bacterial species within each sample, usually referred to as a community. This type of diversity measure is defined as alpha diversity, and it is a measure of the taxonomic diversity (i.e., number of different taxa) and evenness (i.e., relative abundances of different taxa) within a single community. Many studies of bacterial community composition across environmental gradients also measure beta diversity. Beta diversity is the measure of species diversity between communities, and it provides a measure of how species change within communities over time and across environments. In bacterial diversity studies, diversity is often measured as the abundance Operational

Taxonomic Units (OTU) in a community because there is no clear definition of a bacterial species. When two or more bacterial 16S rRNA gene sequences are greater than 97% similar, they are clustered together to form an individual OTU. The 16S rRNA gene is a highly conserved gene that is used as a genetic marker in studies of microbial communities.

Although salinity and depth are both major determinants of bacterioplankton community composition on a large spatial scale, other biotic and abiotic factors also influence the structuring of bacterial communities including pH, oxygen, and nutrient concentrations. In soil, community composition was found to vary with pH (Lauber et al. 2009) as well as with nitrogen concentrations (Fierer et al. 2012), with shifts in the most abundant taxonomic groups across these environmental gradients. Along an oxygen gradient across the oxygen minimum zone off the Chilean coast, Stewart et al. (2012) found a shift in both bacterial and archaeal community composition as oxygen decreased and nitrogen processes shifted from ammonium oxidation to anammox and denitrification. Bacterioplankton community composition was also shown to shift along a nutrient gradient in a eutrophic tropical estuary, with lower alpha diversity and distinct taxonomic composition at nutrient-rich stations (Vieira et al. 2008). Although these studies lend insight into changes in community composition across gradients, many focus on variability across one environment or with a singular factor, like oxygen, and use different methods for measuring diversity, composition, and correlation to the environment. Thus there is a need for more comprehensive studies of microbial community variability across complex environments and gradients using one standardized approach, which would allow for comparison between environments

and further our understanding of the complex nature of bacterioplankton communities.

Coastal zones are complex aquatic systems that present bacterioplankton communities with many different and overlapping seasonal and spatial environmental gradients. Chapter two of this dissertation focuses on the spatial variability of bacterioplankton communities across the complicated environmental gradients encountered from freshwater to the deep ocean. I generated a synoptic, 16S rRNA gene-based bacterioplankton community profile of a complex coastal zone using water samples collected across the environmental gradients of the Columbia River, estuary, and plume, and along coastal transects covering 360 km of the Oregon and Washington coasts and extending to the deep ocean. With this work I identified key physical and biological factors correlating with community composition across the complex mixing zones of a river-influenced coastal margin and described how bacterioplankton communities varied with salinity, temperature, depth, nutrients gradients.

Spatial and temporal community variability:

Spatial variability in bacterioplankton communities has been explored on scales that range from millimeters (Long and Azam 2001) to kilometers (Hewson et al. 2006b) to global (Fuhrman et al. 2008, Pommier et al. 2007). This variability is often attributed to a combination of biological factors that influence the rate of growth of individual taxa, and physical parameters that prevent communities from interacting across different spatial scales. A broad range of environmental conditions

can potentially influence bacterial communities including pH, salinity, temperature, and organic matter, and biotic factors such as bacterial production, phytoplankton interactions, and grazing. Varying sets of these factors have been shown to explain variation in microbial communities within soils, lakes, rivers, estuaries, and the oligotrophic surface ocean (Crump et al. 2003, Crump and Hobbie 2005, Fierer et al. 2009, Hewson et al. 2006a, Kent et al. 2007, Lauber et al. 2009). Dispersal processes have also been shown to influence bacterial biogeography (Martiny et al. 2006). Although it is known that environmental conditions and dispersal processes govern community diversity and taxonomic composition, the scale and the degree to which these two factors influence bacterial spatial variability is still relatively unknown.

Globally, marine bacterioplankton, like metazoans, follow a latitudinal diversity gradient, with increasing diversity towards the equator (Fuhrman et al. 2008). Although, diversity measures may be similar among communities at similar latitudes, taxonomic composition has been found to be greatly variable between locations. In a study of coastal bacterioplankton diversity across ocean basins, Pommier et al. (2007) found many bacterial taxa to be endemic to one location and only a small number of taxa to be present at more than one location, with these cosmopolitan taxa being highly abundant compared to endemic taxa. Although many individual taxa were only found at one location, the same major taxonomic groups were represented at all locations (Pommier et al. 2007). Thus the structure of communities was similar across all sites, with each community comprised of many endemic and a few dominant taxa of varying taxonomic affiliations (Pommier et al. 2007). Although a community may comprise many rare taxa, many of these taxa are

metabolically active and may become dominant when environmental conditions change (Campbell et al. 2011). Crump et al (2012), found that dispersal processes also influence community composition, showing that in a stream-fed arctic lake more than half of the dominant lake taxa were rare taxa in upstream environments and demonstrating the adaptability of rare taxa to different niches. Thus, bacterioplankton communities are highly dynamic and community composition is strongly influenced by both local environmental conditions and dispersal processes.

Bacterioplankton communities vary spatially with environmental conditions, but patterns of seasonal succession among communities have also been shown from freshwater lakes to the open ocean. Seasonal shifts in microbial community composition have been demonstrated in the Sargasso and Baltic Seas and the English Channel, where succession of microbial communities correlated with changes in seasonally influenced environmental factors such as mixed layer depth, temperature, and nutrient concentrations (Andersson et al. 2010, Carlson et al. 2009, Gilbert et al. 2009, Morris et al. 2005). Kent et al (2007) attributed seasonal succession of phytoplankton in freshwater lakes to similar environmental factors but found that succession of bacterial communities was driven by biological interactions with phytoplankton species.

Seasonal succession of bacterioplankton communities is repeatable over an annual cycle (Andersson et al. 2010, Carlson et al. 2009, Crump et al. 2009, Fuhrman et al. 2006, Morris et al. 2005). Crump et al. (2009) showed seasonal synchrony of bacterioplankton communities in six large arctic rivers but also annual reassembly, indicating that communities shift in predictable patterns each year. Coastal

bacterioplankton communities also reassemble annually. Fuhrman et al. (2006) identified seasonal patterns in communities off the coast of California and showed that these same patterns reoccur each year and are highly predictable based on a combination of biotic and abiotic factors. Thus, bacterioplankton communities are variable across space and time with environmental conditions and taxonomic composition can potentially be predictable based on these specific conditions.

Although many studies have determined the spatial or temporal patterns of bacterioplankton communities, most of these studies were restricted to single dimensions, focusing either on long-term time series, depth profiles, or horizontal surveys across environmental gradients. Chapter three of this dissertation presents a study that compared bacterioplankton community composition in all three of these dimensions: spatially from river to surface ocean, by depth from surface to deep ocean, and through time seasonally over an annual cycle. In this study I characterized bacterioplankton community composition from three hundred samples, which enabled a fine-scale resolution of community variability and led to the discovery of robust spatial patterns from river to ocean, and seasonal shifts that may not have been observed if fewer samples were analyzed. Determining both the spatial and temporal variability of bacterioplankton communities provided a framework for modeling these communities across environmental gradients.

Ecological role of individual taxa:

Several studies have described broad-scale spatial and temporal changes in microbial communities in freshwater, coastal, and open ocean (e.g. (Crump et al.

2012, Fuhrman et al. 2006, Galand et al. 2010, Gilbert et al. 2009, Kirchman et al. 2010), and although these studies discuss changes in community composition and structure, little is mentioned about the specific taxa that distinguish one community from another and potentially serve as the drivers of change in these environments. Recently, a few studies described how specific taxonomic groups and individual taxa vary with environmental conditions as well as with each other (Barberan et al. 2012, Eiler et al. 2012, Gilbert et al. 2012). In one long-term study of the English Channel, Gilbert et al. (2012) showed strong repeatable seasonal patterns in the general bacterioplankton community and demonstrated that seasonal variability in specific taxonomic groups was best explained by length of day, temperature, nutrients, and photosynthetically active radiation (i.e., “bottom-up” variables), rather than biomass of zooplankton species (i.e., “top-down” variables). Eiler et al. (2012) found complex interdependences in bacterioplankton communities and synchrony over the temporal scale for specific taxonomic groups in a freshwater lake system. Taxonomic interdependencies have also been observed in soil communities, where non-random co-occurrence patterns were found to occur more than expected by chance indicating strong inter-taxa relationships among specific soil taxa (Barberan et al. 2012). Although recent microbial biogeographical studies have been more focused on the variability of specific taxonomic groups instead of overall beta diversity, there is still much to be learned about how taxa interact and change with environmental conditions and their importance in ecosystem function.

The spatial and temporal distributions of ecologically important marine bacterial taxa such as SAR11, SAR86, and *Prochlorococcus* have been recently

described, although most of these studies focus on single bacterial types within a specific environment (Carlson et al. 2009, Dupont et al. 2012, Malmstrom et al. 2010, Morris et al. 2012, Treusch et al. 2009). Malmstrom et al. (2010) showed consistency in the depth distribution of ecotypes of *Prochlorococcus* over a five-year period in the Pacific and also showed that these distributions, although disturbed by seasonal mixing in the Atlantic, would always reestablish during periods of stratification. The depth distribution of SAR11 ecotypes has also been described, with distinct populations observed in the epipelagic and upper mesopelagic at the long-term Bermuda Atlantic Time Series (BATS) (Carlson et al. 2009), and across nutrient, chlorophyll, and organic carbon gradients in the South Atlantic (Morris et al. 2012). Spatially separated phylotypes have also been observed in the common surface ocean bacteria SAR86, where SAR86-C and D phylotypes were found in colder coastal environments while SAR86-B was found in warmer waters and SAR86-A was found to have a ubiquitous distribution (Dupont et al. 2012).

The recent studies describing patterns of different taxonomic groups looked specifically at variation of taxa over time or with environmental variables only at fixed stations. The challenge in understanding taxonomic patterns then is to define key taxa in a community for different environments over both spatial and temporal scales. Additionally, there is also a need to determine how ecologically important taxonomic groups, like SAR11 or SAR86, affect overall patterns of community distribution in different environments. Chapter four of this dissertation builds on these past studies by examining both the variation of specific taxa as well as whole communities and by identifying key taxa that define specific environments. I

characterized the biogeography of specific bacterioplankton populations from 600 water samples collected across the Columbia River coastal margin over a four-year period. The focus of this study was on population-level analyses to determine the key taxa in each environment and how abundances of these taxa were shaped by changing environmental conditions. The goal of this study was to develop a set of key taxa that were representative of specific conditions in an environment and were indicators of change in bacterioplankton community composition.

Linking community structure and function:

Microbes strongly influence biogeochemical cycling and thus the transfer of organic matter and energy within and across ecological systems. Metagenomic and metatranscriptomic data provide powerful insight into the metabolic potential and expression in an environment without previous knowledge of taxonomic or functional characteristics (Poretsky et al. 2005). A metagenome captures all the genetic material of a microbial community in an environment, while a metatranscriptome captures the community gene expression of this environment. Many recent studies have used metagenomic or metatranscriptomic approaches to answer important ecological questions about the biogeochemical function of microbes in soil, in the surface coastal and open ocean, as well as the deep sea (Eloe et al. 2011, Fierer et al. 2012, Gifford et al. 2011, Hewson et al. 2010, Poretsky et al. 2005). Together, metagenomic and metatranscriptomic studies further solidify the importance of microbes to ecosystem function, and provide greater detail about the genetic capabilities of these organisms.

The goal of many early metagenomic studies was to describe the functional processes occurring in different environments and to ascribe a metabolic fingerprint in addition to a taxonomic one in order to better link structure and function in bacterial communities (Biddle et al. 2008, DeLong et al. 2006, Rusch et al. 2007, Tringe et al. 2005). DeLong et al. (2006) found there was a distinct vertical distribution of taxonomy and function genes, with genes associated with carbon metabolism, motility, and viral interactions all having variable abundances with depth. Variable abundances of functional genes have also been observed across habitat types, where different environments have been shown to have habitat-specific functional fingerprints based on variability of important processes like photosynthesis and starch metabolism (Tringe et al. 2005). Similar to metagenomic studies, early metatranscriptomic studies described the dominant microbially mediated biogeochemical processes occurring in an environment. Early studies analyzed microbial community mRNA transcripts of the open and coastal ocean and within a salt marsh creek (Frias-Lopez et al. 2008, Gilbert et al. 2008, Poretsky et al. 2005). These studies determined key functional and metabolic processes active within each system, including genes involved in sulfur oxidation, photosynthesis, carbon fixation, and nitrogen acquisition. Determining which microbial genes are being expressed within an environment lends greater insight into the critical metabolic processes of the system. For example, Frias-Lopez et al. (2008) found that some of the most highly expressed genes in open ocean surface waters are proteorhodopsins, light-driven proton pumps used for the creation of chemical energy. Corresponding analysis of community DNA found that these genes were not only highly expressed but were also

found in a diverse group of microbial taxa, indicating proteorhodopsins were important genes for driving cellular and metabolic processes of microbes in the open ocean.

With the advancement of sequencing technology and thus an increase in the size of metagenomic and metatranscriptomic libraries, there has been a shift from descriptive to more hypothesis-based research. In a comparison of day/night transcripts from the open ocean, Poretsky et al. (2011) found that microbial communities have metabolic and biogeochemical responses to changes in solar forcing, as day transcripts were mostly for photosynthesis and oxidative phosphorylation, while night transcripts were mostly for housekeeping activities, such as DNA repair, amino acid biosynthesis, protein export, and other processes important to cell maintenance. Hewson et al. (2010) looked at spatial patterns across metatranscriptomes from ocean surface waters around the globe and found that the spatial differences between these metatranscriptomes were driven by expression of genes involved in nutrient acquisition and transport as well as cell metabolism and growth. These spatial patterns in gene expression could be a consequence of differences in nutrient concentrations across ocean basins but also a difference in abundance of cyanobacteria, specifically *Prochlorococcus marinus*, which dominated transcript libraries (Hewson et al. 2010). A shift in taxonomy was also shown in a experiment-based study describing changes in gene expression in DOM-amended bottle experiments where the addition of high-molecular weight DOM to surface seawater resulted in shifts in taxonomy and gene expression as different groups took advantage of the added carbon source (McCarren et al. 2010). From these studies, we

see that marine bacterioplankton community gene expression changed on both temporal and spatial scales and that taxonomic composition of the community played an important role in gene expression patterns.

Recent metagenomic and metatranscriptomic studies have found coupled taxonomic and metabolic shifts across ecologically relevant biogeochemical gradients in marine and soil environments. Soil metagenomes sampled across a nitrogen gradient revealed increases in genes associated with replication, electron transport, and protein metabolism as well as a shift from oligotrophic to more copiotrophic taxa with increasing nitrogen concentrations (Fierer et al. 2012). Oxygen gradients across oxygen minimum zones have also been the focus of metagenomic and metatranscriptomic studies (Canfield et al. 2010, Stewart et al. 2012). Stewart et al. (2012) compared both metagenomic and metatranscriptomic data across an oxygen gradient along the Chilean coast. Again, changes in gene abundance, gene expression, and taxonomic composition occurred across the oxygen gradient as nitrogen processes changed from oxidative to reductive and the community shifted from nitrifying taxa to those taxa associated with anammox and denitrification (Stewart et al. 2012). Looking at the same oxygen gradient, metagenomic data also revealed the first evidence of an active sulfur cycle in an oxygen minimum zone and coupling between sulfur and nitrogen processes (Canfield et al. 2010). These metagenomic studies across both nutrient and oxygen gradients have provided gene-level information concerning important ecological process and microbial community dynamics in these biogeochemically important environments.

Many metagenomic and metatranscriptomic studies focus on the marine environment, but few have looked at metagenomic data from river systems. Ghai et al. (2011) looked at a metagenome taken from the largest river system in the world, the Amazon. Results from this study found the presence of common freshwater lineages including the Actinobacteria AC1 clade as well as the Betaproteobacteria taxa, *Polynucleobacter* among metagenomic sequences. In addition, this metagenome had a disproportionate amount of heterotrophic carbon processing genes when compared to marine metagenomes, indicating these abundant freshwater microbes played an important role in the breakdown and transformation of terrestrial carbon in the river, which has important implications for the coastal ocean (Ghai et al. 2011).

Although gradients have recently become a focus of metagenomic and metatranscriptomic studies, to date there have not been any of these types of studies focusing on gene abundance and expression across a salinity gradient. Chapter five of this dissertation describes changes in metabolic potential and gene expression across the salinity gradient of the Columbia River coastal margin. Previous work on bacterioplankton community composition has shown spatially and taxonomically distinct communities in the river, estuary, plume, and surface ocean. The addition of gene abundance and expression data to already established patterns of taxonomic composition creates a link between community structure and community function, and provides information to explore how and why specific populations are distributed across from river to ocean.

The Columbia River coastal margin:

The Columbia River coastal margin is a productive ecosystem due to nutrients from seasonal upwelling and the Juan de Fuca strait and the Columbia River (Hickey and Banas 2003). The biological and physical processes of these waters are complicated by variable winds, remote wind forcing, shelf width, and submarine canyons (Hickey and Banas 2003, Hickey and Banas 2008, Hickey et al. 2010), which may in turn affect the composition of bacterioplankton communities along the coast. The Pacific Northwest coast is highly productive, with greater productivity occurring along the Washington coastline compared to Oregon due to increased nutrient delivery from the Juan de Fuca strait and the Columbia River as well as nutrient entrainment from a wider shelf and submarine canyons (Hickey and Banas 2003, Hickey and Banas 2008, Hickey et al. 2010). The Columbia River is the second largest river in the United States with a mean annual discharge of $7300 \text{ m}^3 \text{ s}^{-1}$ (Hickey et al. 1998) and this significant release of freshwater has a large impact on the chemical, physical, and biological characteristics of the adjacent Oregon and Washington coasts. The Columbia River supplies large amounts of iron and silica to the coastal ocean (Hickey et al. 2010), which can influence primary and thereby secondary production along the coast. This nutrient supply when coupled with seasonal upwelling of nitrate during the summer months can make the Columbia River plume region highly productive (Hickey et al. 2010).

Previous microbial biogeographical studies of the Columbia River coastal margin system focused on the characterization and distribution of particle-attached and free-living bacteria and archaea in the river and estuary (Crump et al. 1998,

Crump et al. 1999, Crump and Baross 2000). Additionally, a recent study looking at patterns of carbon and nitrogen gene expression across the estuarine salinity gradient using microarray data has added insight into how important biochemical processes shift from river to ocean (Smith et al. 2010). Smith et al. (2010) showed strong seasonal shifts in gene expression patterns but variable spatial differences with some seasons showing little difference in gene expression across the river to ocean salinity gradient. This dissertation expands on these previous studies and describes patterns in bacterioplankton diversity and taxonomic composition over broad spatial and temporal scales and across the many environmental gradients that define coastal zones. This dissertation also presents metagenomic and metatranscriptomic information that describes metabolic potential and gene expression patterns across the Columbia River coastal margin without any assumptions of metabolic processes.

Chapter questions and hypotheses:

1) How is bacterioplankton community variability potentially influenced by environmental gradients? Which environmental factors are most important in shaping communities?

On a large scale, variation across spatial gradients will be influenced by structuring environmental factors including salinity, temperature, and depth as fresh and marine, warm and cold, surface and deep waters mix across a complex coastal zone. On a smaller scale, however, more location-specific factors, including nutrient and chlorophyll concentrations, will be influential in determining the spatial patterns of bacterioplankton communities.

2) How are bacterioplankton communities spatially and seasonally distributed?

Bacterioplankton communities will separate by location from fresh, to estuarine, to coastal waters, as large salinity and depth gradients physically separate water masses and associated bacterioplankton communities. Mixing of communities will be apparent at interfaces between environments. With seasonal shifts come changes in influential variables such as temperature and nutrient concentrations, so it is expected that community composition will change across seasons. Seasonal shifts, however, will be overwhelmed by the spatial patterns of communities as the sampling area encompasses great spatial distances.

3) What are the ecologically important dominant taxonomic groups of the Columbia River coastal margin? How are these taxa potentially influenced by environmental variables both spatially and seasonally?

Due to the spatial complexity of the coastal zone, different dominant taxonomic groups will be present from river to ocean and from surface to deep. Additionally, variability of these key taxa will be affected differentially across the coastal margin, depending on the physical and biological factors affecting each environment.

4) What is the spatial pattern of community metabolic potential and gene expression across a salinity gradient from river to coastal ocean?

Gene abundance and availability across a river to ocean gradient will shift dramatically, but that taxonomically distinct populations would perform similar metabolic activities regardless of the salinity of the water. Thus, the typical phylogenetic shift from fresh, to estuarine, to marine bacteria over the salinity gradient would be apparent in metagenomic data, but would not be reflected in metatranscriptomic gene expression patterns.

CHAPTER 2:
BACTERIOPLANKTON COMMUNITY VARIATION ACROSS RIVER TO
OCEAN ENVIRONMENTAL GRADIENTS

Abstract:

Coastal zones encompass a complex spectrum of environmental gradients that each impact the composition of bacterioplankton communities. Few studies have attempted to address these gradients comprehensively. I generated a synoptic, 16S rRNA gene-based bacterioplankton community profile of a coastal zone by applying the fingerprinting technique denaturing gradient gel electrophoresis (DGGE) to water samples collected from the Columbia River, estuary, and plume, and along coastal transects covering 360 km of the Oregon and Washington coasts and extending to the deep ocean (>2000m). Communities were found to cluster into five distinct environments based on location in the system (ANOSIM, $p < 0.003$): estuary, plume, epipelagic, shelf bottom (depth < 150 m), and slope bottom (depth > 650 m). Across all environments, communities varied strongly with abiotic factors (salinity, temperature, depth, $\rho = 0.734$). Within each coastal environment, communities varied more with biotic factors. Thus, structuring physical factors in coastal zones, such as salinity and temperature, define the boundaries of many distinct microbial habitats, but within these habitats variability in microbial communities is driven by biological gradients in primary and secondary productivity.

Introduction:

Microbial communities are abundant and diverse and vary over space, time, and across environmental gradients. Key to understanding the distribution of microbial communities is determining the factors that both create and sustain these distribution patterns. Microbial biogeographical patterns are shaped by both

environmental variables and dispersal processes (Green et al. 2008, Horner-Devine et al. 2004, Martiny et al. 2006). Horner-Devine et al. (2004) showed that there is a taxa-area relationship for bacteria in which the number of taxa increases with increasing sampling area, and that this relationship was driven primarily by environmental heterogeneity. More specifically, abiotic factors such as pH, salinity, temperature, and organic matter, and biotic factors such as bacterial production, phytoplankton interactions, and grazing have been shown to explain variation in microbial communities within soils, lakes, rivers, estuaries, and the oligotrophic surface ocean (Crump et al. 2003, Crump et al. 2004, Crump and Hobbie 2005, Fierer et al. 2009, Hewson et al. 2006b, Kent et al. 2007, Lauber et al. 2009, Simek et al. 2001). These studies provide insight into varying environmental tolerances and ecological strategies of microbial taxa (Lauber et al. 2009).

Variability in bacterioplankton communities has been explored on spatial scales that range from millimeters (Long and Azam 2001) to kilometers (Hewson et al. 2006b) to global (Fuhrman et al. 2008, Pommier et al. 2007). Many studies describe community variability in aquatic systems, but most focus on specific environments including estuaries, coastal regions, and ocean gyres. A handful describe variability across one or two of the many gradient regions that separate these environments, including the freshwater-saltwater interface, the estuary-coastal ocean interface, nearshore-offshore interface, or across depth gradients in the coastal and open ocean (e.g. Alonso-Saez et al. 2007, Celussi et al. 2009, Crump et al. 2004, Ghiglione et al. 2008, Hewson et al. 2006a, Kataoka et al. 2009, Vieira et al. 2008). Very few studies have characterized microbial community distribution across a large

river to ocean gradient as seen in the Columbia River system. The Columbia River is the second largest river in the United States with a mean annual discharge of $7300 \text{ m}^3 \text{ s}^{-1}$ (Hickey et al. 1998). This significant release of freshwater has a large impact on the chemical, physical, and biological characteristics of the adjacent Oregon and Washington coasts. Previous microbial community studies of the Columbia River system focused on the characterization and distribution of particle-attached and free-living bacteria and archaea in the river and estuary (Crump et al. 1998, Crump et al. 1999, Crump and Baross 2000). Here I expand on previous work and assess differences in the spatial distribution of bacterioplankton communities across the river to ocean gradient from freshwater to the deep ocean, and identify environmental factors that explain the variability seen among these communities.

For this study I examined a bacterioplankton community profile using the community fingerprinting technique denaturing gradient gel electrophoresis (DGGE) to water samples collected from the Columbia River, estuary, and along the Oregon and Washington coasts. Environmental variables were also measured in order to link variation in microbial community composition with environmental factors. I hypothesized that bacterioplankton communities would differ over the large spatial scale of the Columbia River system and that variation across spatial gradients would vary with key environmental factors including, but not limited to, salinity and depth. Our results indicate that communities separated into five distinct groups based on location, and varied with abiotic factors (salinity, temperature, depth). Within each group, communities varied with biotic factors involved in primary and secondary production.

Methods:

Samples were collected from the Oregon and Washington coasts and the Columbia River and estuary (latitude 44.652° and 47.917°, longitude -123.874° and -125.929°) during a Center for Coastal Margin Observation and Prediction (CMOP) cruise on August 13-31, 2007 aboard the *R/V Wecoma*. Water samples were taken from Columbia River estuary, plume, and four coastal ocean lines (Figure 2.1). For coastal lines, samples were taken at three depths per station (surface, within thermocline, and bottom). Plume samples were taken at 2 depths (surface and bottom) and estuary samples were collected across a salinity gradient of 0 to 30. Samples were collected with 10-liter Niskin bottles, and depth profiles of salinity, temperature (°C), turbidity (NTU), oxygen (mgL⁻¹), and chlorophyll fluorescence were recorded. For all samples, surface was defined as 1 m depth and bottom was defined as 5 m above sediment. Depth for samples taken within the thermocline was determined by CTD temperature sensor.

For collection of photosynthetic pigments (chlorophyll *a*, phaeophytin), 0.5-2 L of water was filtered through 25 mm GF/F (Whatman) filters. Filters were snap frozen in liquid nitrogen. Concentrations were determined using high performance liquid chromatography (HPLC) (Wright et al. 1991). Pigment samples were cold-extracted (-15°C) in polypropylene centrifuge tubes using a fixed volume of 90% acetone in water (v/v) and chromatographic separations were made using a reverse-phase column and diode array detection at 436 nm. Pigments were quantified by comparison of integrated peak area with response factors for authentic commercial standards.

Filtrate from ashed 25 mm GF/F (Whatman) filters was collected for dissolved organic carbon (DOC), total dissolved nitrogen (TDN) and phosphorus (TDP), and nutrients. For DOC, 20 ml filtrate were stored at -20°C in polypropylene vials and analyzed by Horn Point Laboratory (HPL) analytical services (Sugimura and Suzuki 1988), using a Shimadzu TOC-5000 total organic carbon analyzer. Filtrate (20 ml) for TDN and TDP was collected in reagent-conditioned 30 ml polypropylene bottles, stored at -20°C, and analyzed by HPL analytical services (Valderrama 1981). Dissolved organic nitrogen (DON) and phosphorus (DOP) concentrations were calculated from TDN and TDP values by subtracting dissolved inorganic nitrogen and phosphorus concentrations. For dissolved inorganic nutrient measurements, filtrate (25 ml) was collected in acid-washed polyethylene vials, stored at -20°C, and analyzed using standard continuous segmented flow autoanalyzer techniques for ammonium, nitrate+nitrite, nitrite, dissolved silica (DSi), and soluble reactive phosphorus (SRP) (Gordon et al. 1994).

Bacterial production was measured as the rate of incorporation of L-[³H] leucine (20 nM final concentration) into cold trichloroacetic acid (TCA; 5% final concentration)-insoluble fraction of macromolecules in four 1.5-ml subsamples, including one killed control, incubated for 1 h at *in situ* temperatures in the dark in 2-ml microcentrifuge tubes (Fisher Scientific). TCA-precipitated macromolecules were centrifuged at 14,000xG for 10 min and washed once with 1.5 ml cold 5% TCA, flooded with scintillation cocktail and counted in a Packard Tri-Carb liquid scintillation counter. Bacterial carbon production (BP) was calculated from leucine incorporation used a ratio of cellular carbon to protein of 0.86, a fraction of leucine in

protein of 0.073, and an intracellular leucine isotope dilution of 2 (Kirchman et al. 1993).

DNA samples (1-6 L per sample) were pumped through 0.2 µm Sterivex-GP filters (Millipore). 1 mL DNA extraction buffer (DEB: 0.1 M Tris-HCl (pH 8), 0.1 M Na-EDTA (pH 8), 0.1 M Na₂H₂PO₄ (pH 8), 1.5 M NaCl, 5% CTAB) (Zhou et al. 1996) was added to each filter, which were subsequently sealed and stored at -80°C. DNA was extracted using methods adapted from Zhou et al. (1996) and Crump et al. (2003).

Extracted DNA was PCR-amplified using 16S ribosomal RNA gene primers for the V3 region in 50 µl reactions (1X PCR buffer (Promega), 2.5 mM MgCl₂, 0.2 mM deoxynucleoside triphosphates (dNTPs), 0.25 µM primers, 0.5 U *Taq* polymerase (Promega)). The primers were bacteria-specific 357f with GC-clamp attached to the 5' end (5'-CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCGCCCC-CCTACGGGAGGCAGCAG-3') and universal primer 519r (5'-ACCGCGGCTGCTGGCAC-3'). PCR conditions, which followed Crump et al. (2004) were 30 cycles with a final one-hour extension at 72°C (Janse et al. 2004). For DGGE, PCR products were loaded onto 29-lane acrylamide (8%) gels prepared with 30% acrylamide-bis-acrylamide (37.5:1; Bio-Rad), 0.5X TAE buffer (1X TAE is 40 mM Tris [pH 8.0], 20 mM acetic acid, 1 mM EDTA), and gradients of 35 to 60% denaturants (100% equals 7 M urea and 40% formamide). The amount of PCR product loaded into each lane was adjusted to ensure a similar amount of DNA was run for each sample. Gels were run for 24 hrs at 75 V. A DGGE ladder, previously constructed from clone libraries,

was run every 6 lanes in order to accurately assess the position of bands across each gel and to allow for comparison between gels. DGGE gels were photographed using a Chemi-Doc gel documentation system (Bio-Rad) with exposure times set for photographs to be undersaturated and analyzed using the GelcomparII software package (Applied Maths) as described in Crump et al. (2007). Briefly, banding patterns were subjected to ‘rolling ball’ background subtraction (disk width, 50) and bands were considered present when peak height exceeded 5% of the peak height of the darkest band in the sample.

To assess similarity between bacterioplankton communities, pairwise similarity matrices were calculated from relative band height (i.e. band intensity) using the Bray-Curtis similarity coefficient (Legendre and Legendre 1998). Relative band height was determined by dividing the height of each band by the sum of all band heights in the sample. Community patterns using relative band height were very similar to those determined using presence/absence data, and previous work has shown that community patterns are the same regardless of the method used to analyze DGGE banding patterns (Crump et al. 2007). Similarity matrices were visualized using multiple dimensional scaling (MDS) diagrams, a form of ordination. Analysis of similarity statistics (ANOSIM) were run to test significant differences between a priori sampling groups identified visually in MDS diagrams. Similarity matrices, MDS diagrams, and ANOSIM statistics were carried out using PRIMER v6 for Windows (PRIMER-E Ltd, Plymouth, UK).

Environmental data were compiled and tested for normality (Shapiro-Wilkes test, $p > 0.05$). Variables were divided into abiotic and biotic categories, and analyzed

separately. Abiotic variables included in analyses were: depth, salinity, temperature, turbidity, dissolved oxygen (DO), nitrate, nitrite, ammonium, ortho-phosphate, silicic acid (DSi), dissolved organic nitrogen, phosphorus and carbon (DON, DOP, DOC), chlorophyll *a* (Chl *a*), bacterial production rate (BP). Variables not normally distributed were transformed to be as close to normality as possible, and 65 of the 71 samples were used for environmental analyses due to gaps in environmental data. Results were analyzed with Primer v6 (PRIMER-E Ltd, Plymouth, UK) (Clarke and Ainsworth 1993). BV-STEP analysis was used to identify sets of variables that best explained community variability ($\rho > 0.95$, $\Delta\rho < 0.001$, 10 random starting variables, 100 restarts). BIO-ENV was used to rank each individual environmental variable by degree of association with community variability. Analyses were run separately for abiotic and biotic variables. BV-STEP and BIO-ENV use the Spearman rank correlation coefficient (ρ) to determine the degree of association between similarity matrices of DGGE (Bray-Curtis similarity) and environmental data (Euclidean distances), where a coefficient of one denotes all variability explained. To ensure variables did not correlate with each other, pairwise correlation coefficients were determined for the overall dataset as well as for specific sample groups using Primer v6. This reduced the size of the environmental dataset by removing highly correlated variables ($\rho > 0.90$), and allowed for more interpretable results. Although a part of experimental design, depth was included in the environmental analysis and treated as a proxy variable for a number of environmental conditions that change in the vertical, including light. Depth did not correlate with any other variable in the dataset ($\rho < 0.55$).

Results:

Bacterioplankton communities separated into five distinct environments based on location in the system (ANOSIM, $p < 0.003$): estuary, plume, epipelagic, shelf bottom (depth < 150 m), slope bottom (depth > 650 m). The plume community consisted of coastal surface samples with salinity less than 31. The epipelagic group consisted of coastal samples collected at the surface and within the thermocline (maximum depth = 50 m). Coastal bottom samples were divided into those taken on the continental shelf (depth < 150 m) and deeper samples taken on the continental slope (depth > 650 m). The MDS diagram of all samples (Figure 2.2) shows little overlap among these five groups, indicating large spatial differences among bacterioplankton communities. The percent similarity among samples ranged broadly from 4% to 91% (average $36\% \pm 14\%$ s.d.), and similarity values were higher within groups than between groups (Table 2.1).

Within the epipelagic group (Figure 2.3A) there was a clear distinction between near-shore and offshore communities. In samples taken close to shore (< 35 km), surface and within thermocline communities clustered together by location along the coast. In contrast, samples collected offshore (> 60 km), surface and within thermocline samples clustered separately, independent of the coastal line sampled. Within the shelf bottom group (depth < 150 m), clustering was less defined (Figure 2.3B). These communities separated into three groups based on coastal location: Oregon coast (Newport hydroline), Columbia River line, and Washington coast (La Push, Willapa Bay lines). Shelf bottom samples taken below the plume region clustered with nearby Columbia River line or Washington coast samples. Within the

estuary, bacterioplankton communities separated by salinity (Figure 2.3C), irrespective of location of the sampling site, forming three clusters: oligo/mesohaline, mesohaline, and meso/polyhaline. Although estuarine sample size is small ($n=10$), the salinity trend is readily apparent. Unlike the previous groups, plume communities did not separate by salinity or coastal location and a pattern of separation was not easily seen (data not shown).

Comparison of all samples ($n=65$) to the environmental dataset showed that communities varied strongly with salinity, temperature, and depth across the entire sample set, together giving a correlation coefficient of $\rho=0.734$. With the addition of turbidity and inorganic nutrients (ortho-phosphate, nitrate), the correlation coefficient value increased slightly to $\rho=0.755$ (Table 2.2). When salinity, temperature, and depth were removed from the analysis, community variability was correlated with turbidity, inorganic nutrients (ortho-phosphate, nitrate), and DSi ($\rho=0.603$). Results indicate that salinity, temperature, and depth appear to be the most important variables, with a combination of other abiotic environmental factors contributing slightly to bacterioplankton community variability. Biotic factors had little influence on variability in the overall bacterioplankton community, with phaeophytin and bacterial production combining for a correlation value of $\rho=0.297$.

Within the separate coastal environments (epipelagic, shelf bottom, plume), biotic factors played a larger role in the variability of communities, as gradients of primary and secondary production were evident from environmental variables. Chlorophyll *a* and bacterial production were the top two factors correlating with community variability in the epipelagic ($n=23$), shelf bottom ($n=16$), and plume

(n=13), where $\rho=0.471$, 0.494 , and 0.515 respectively. Abiotic factors played a lesser role within the coastal environments. In the epipelagic, turbidity was the abiotic factor with the highest correlation coefficient ($\rho=0.333$), while in the shelf bottom a combination of several abiotic variables, including temperature, turbidity, orthophosphate, nitrite, ammonium, DOP, and DOC produced a much lower correlation coefficient when compared to the biotic variables ($\rho=0.296$). Similarly in the plume, a combination of abiotic factors had little influence on community variability. Looking at all three coastal environments, it is clear that bacterioplankton communities are sensitive to changes in secondary and primary production. However, although communities varied strongly with biotic factors, these biotic factors (e.g., bacterial production, chlorophyll *a*) are dependent upon abiotic factors such as nutrients, turbidity, and temperature. Thus it is likely that abiotic and biotic factors function in combination to control bacterioplankton community composition within each coastal environment.

Unlike the three coastal environments, in the estuary biotic factors were not as important in describing the community variability. Variation of bacterioplankton communities in the estuary (n=10) was strongly correlated with salinity ($\rho=0.628$, Figure 2.3C). DOP and DO were also important factors, with correlation coefficients of 0.452 and 0.440 respectively.

It is also important to note that DGGE does introduce certain biases that must be taken into account when interpreting our results. DGGE has been shown to be unable to detect populations that make up less than 1% of the total bacterial community (Muyzer et al. 1993), but this limit may be less depending on the bacterial

group (Casamayor et al. 2000, Kan et al. 2006a). Additionally, some DGGE bands may not represent bacterial populations, but chloroplasts, as primers amplifying the V3 region of the 16S gene have been found to amplify chloroplasts sequences (Ferrari and Hollibaugh 1999).

Discussion:

Bacterioplankton communities varied dramatically across all major environment types in the coastal region of Oregon and Washington, and the environmental conditions potentially influencing this variability differed depending on the spatial scale of analysis. In coastal zones, communities can vary across gradients created by physical properties like upwelling and currents (Alonso-Saez et al. 2007, Kataoka et al. 2009), but also by those created by influences from land, including freshwater input. With the larger spatial scope of this study, I was able to assess both land and ocean influences on coastal zone bacterioplankton communities, which gives a unique perspective on how communities adapt across river to ocean gradients. One study of similar spatial scale was carried out in Moreton Bay, Australia (Hewson and Fuhrman 2004). This study found that bacterioplankton communities were significantly different across gradients from river to bay to open ocean (Hewson and Fuhrman 2004). However, this study also notes that although there was heterogeneity between environment types, bacterioplankton communities within each environment were relatively homogenous (Hewson and Fuhrman 2004). This is in contrast to our results in which bacterioplankton communities within each environment could be further clustered across finer spatial scales. Also, low values of

percent similarity within each group (range: 44-68%) relative to typical similarity values for replicate samples (>90%, data not shown) indicate heterogeneity of communities within each environment type, although communities were not significantly different.

Across the entire dataset, salinity, temperature, and depth appeared to be the three most important environmental factors describing variability. This makes sense because these three factors define the differences among the environments hosting the five major clusters of bacterioplankton communities seen in this study. Salinity and temperature vary between the estuary and the rest of the coastal zone, and temperature and depth vary between the surface and bottom ocean. Salinity has been shown in many studies to be an important factor for the separation of bacterial communities (e.g. Bernhard et al. 2005, Bouvier and Giorgio 2002, Crump et al. 2004, Hewson and Fuhrman 2004, Selje and Simon 2003, Troussellier et al. 2002). Depth has also been shown to be an important factor in determining bacterial community composition (e.g. Ghiglione et al. 2008, Riemann et al. 1999). Although depth was part of the experimental design in this study, it can be considered a proxy variable for many biological and physical factors that vary in the vertical dimension but were not directly measured, including photosynthetically available radiation (PAR). Once bacterioplankton communities were separated by salinity, temperature, and depth gradients, community composition within environments varied with biotic factors, specifically variables contributing to both primary and secondary production. Thus, by using only community fingerprinting data and a suite of environmental variables, the most important variables predicting bacterioplankton community

composition were elucidated for a dataset that encompassed a large spatial distance, from river to ocean.

Within the epipelagic environment, communities separated based on coastal proximity. At distances greater than 60 km offshore, fingerprinting data showed communities are laterally homogeneous across a large sampling area, but were variable by depth with surface (<2.5 m) communities differentiating from communities within the thermocline (15-50 m). In contrast, the near-shore coastal ocean (<35 km offshore) was more heterogeneous, with different bacterioplankton communities developing at different locations along the coast, and with surface (<2.5m) and thermocline (7-15 m) samples clustering together (Figure 2.3A). Differing patterns with depth in the epipelagic may be explained by the fact that the thermocline is relatively compressed close to shore, which could facilitate vertical mixing and homogenization of epipelagic communities. Offshore the thermocline reaches deeper depths, and this physical separation appears to differentiate the surface and deeper communities within the epipelagic zone.

The separation of the near- and offshore communities can be explained by the physical processes affecting the Pacific Northwest. Fluctuations of local winds can cause near-shore coastal waters to shift between upwelling to downwelling conditions over short time scales (Hickey 1989). Across the shelf, however, shifting winds affect waters differently. The mid to outer shelf is not affected by changing winds on these short time scales, and along-shore surface currents remain intact (Kosro 2005). On the inner shelf (less than 10 km), however, a shift in wind direction causes currents to change almost immediately (Hickey 1989), and thus causes primary productivity to

increase along the inner shelf, as nutrients are brought to the surface. The productivity gradient from inner to outer shelf is one likely explanation for the separation of epipelagic communities between near shore and off shore. Different bacteria taxa may be present depending on availability and quality of organic matter from primary production, which could result in different community composition (Cottrell and Kirchman 2000). This explanation is supported by environmental data, where chlorophyll *a* and bacterial production were the two most important factors influencing community variability. Separation of near-shore communities laterally along the coast can also be explained by productivity gradients. The Washington coastline has been found to be more productive than Oregon (Hickey and Banas 2003, Hickey and Banas 2008, Hickey et al. 2010), and this difference may explain the separation of near-shore communities by location along the coast. Near-shore communities may each be considered distinct water masses, with different physical processes occurring in each region, in turn affecting productivity and influencing the composition of bacterioplankton communities.

Chlorophyll *a* and bacterial production were also important in shaping communities in the plume and shelf bottom. In the shelf bottom environment, bacterioplankton communities separated loosely into three groups based on location along the coast. This result is similar to the pattern seen in the epipelagic group. Productivity in the surface ocean influences conditions along the bottom, especially in shallower shelf environments. The depth range for the shelf bottom group was 42 to 145 m, and the shallower, near shore sites generally had higher concentrations of chlorophyll *a* and phaeophytin, and higher bacterial production rates when compared

to the deeper, offshore sites. This difference in depth, although not dramatic, created a gradient of productivity across the shelf bottom that was reflected in the composition of communities such that those close to shore, influenced by light and surface water productivity, had different community composition from the deeper offshore sites.

In the plume environment, there was no spatial pattern for communities, but chlorophyll *a* and bacterial production were again the two top factors potentially influencing variability. Plume sites were sampled both north and south of the Columbia River mouth. The Columbia River plume supplies large amounts of iron and silica to the coastal ocean (Hickey et al. 2010), which can influence primary and thereby secondary production along the coast. This nutrient supply coupled with seasonal upwelling of nitrate can make the plume region highly productive (Hickey et al. 2010). The importance of both chlorophyll *a* and bacterial production in shaping plume bacterioplankton communities further illustrates the importance of plume waters on coastal productivity. Along the coast, both chlorophyll *a* and bacterial production played an important role in influencing community composition. Within the estuary, however, salinity was the most important factor potentially influencing bacterioplankton community variation. Dissolved nutrients, specifically DOC and DOP, were also factors potentially influencing community variability. Estuaries are sites of high productivity where river and ocean communities mix along both salinity and nutrient gradients. In the Columbia River estuary, short residence times, and thus rapid movement of water through the estuary, results in the differentiation of bacterioplankton communities by the degree of mixing between river and coastal ocean waters (Crump et al. 1999).

It is often difficult to identify environmental factors that potentially influence community variability (Lozupone and Knight 2007). Factors affecting bacterioplankton community composition are complex and consequently it is the combination of many that influence community variability. Overall, communities across the Columbia River region appeared to be shaped by three abiotic factors: salinity, temperature, and depth. Differences among these factors helped to explain the separation of communities into five distinct spatial environments. Within the coastal environments, differences in productivity due to variable physical processes appeared to shape bacterioplankton communities. Thus, communities were first separated by location along salinity, temperature, and depth gradients, and then varied further across coastal productivity gradients. With community fingerprint data and a suite of environmental variables, I was able to identify potential drivers of bacterioplankton community variability across a river to ocean gradient. Understanding how and why bacterioplankton communities are distributed across environmental gradients allows for better prediction of how these communities, and ecosystems as a whole, might be shaped by future environmental change.

Table 2.1: Percent similarity values within and between groups \pm standard deviation (ANOSIM: $p < 0.003$) as determined from relative band height data using the Bray-Curtis similarity coefficient.

	Estuary	Plume	Epipelagic	Shelf bottom	Slope bottom
Estuary	67.6 \pm 10.2				
Plume	31.0 \pm 8.4	56.8 \pm 10.0			
Epipelagic	26.8 \pm 7.1	42.0 \pm 10.8	42.4 \pm 10.0		
Shelf bottom	19.8 \pm 7.8	28.4 \pm 7.7	35.4 \pm 9.7	52.6 \pm 12.4	
Slope bottom	19.9 \pm 6.9	21.8 \pm 5.3	23.7 \pm 6.3	25.0 \pm 7.7	44.1 \pm 2.5

Table 2.2: Spearman rank correlation coefficients (ρ) showing correlation between bacterioplankton communities and both abiotic and biotic environmental variables. BV-STEP ρ values represent maximum values when all environmental variables were included. For BIO-ENV, the maximum number of variables was set to one to determine the degree of association of each environmental variable individually, ρ values for the top two variables are shown.

Environment	Abiotic BV-STEP factors	ρ	Abiotic BIO-ENV factors	ρ
All	Salinity, Temperature, Turbidity, Nitrate, Depth	0.755	Temperature Salinity	0.565 0.508
Estuary	Salinity, DOP, DOC	0.628	Salinity DOP	0.628 0.452
Plume	Salinity, Temperature, DO, Ortho-phosphate	0.382	DO Turbidity	0.362 0.261
Epipelagic	Temperature, Turbidity, DOC, Depth	0.425	Turbidity Temp	0.333 0.281
Shelf bottom	Temperature, Turbidity, Ortho-phosphate, Nitrite, Nitrate, DOP, DOC	0.286	Nitrite DOP	0.226 0.151
	Biotic BV-STEP factors	ρ	Biotic BIO-ENV factors	ρ
All	Phaeophytin, BP	0.297	Phaeophytin BP	0.274 0.172
Estuary	Chl <i>a</i> , BP	0.313	Phaeophytin BP	0.233 0.226
Plume	Chl <i>a</i> , BP	0.515	Chl <i>a</i> BP	0.437 0.424
Epipelagic	Chl <i>a</i> , BP	0.471	Chl <i>a</i> BP	0.422 0.413
Shelf bottom	Chl <i>a</i> , BP	0.494	Chl <i>a</i> BP	0.452 0.490

Figure 2.1: Map of Oregon and Washington coasts. Inset depicts Columbia River estuary and plume region. Dotted line denotes approximate location of shelf break.

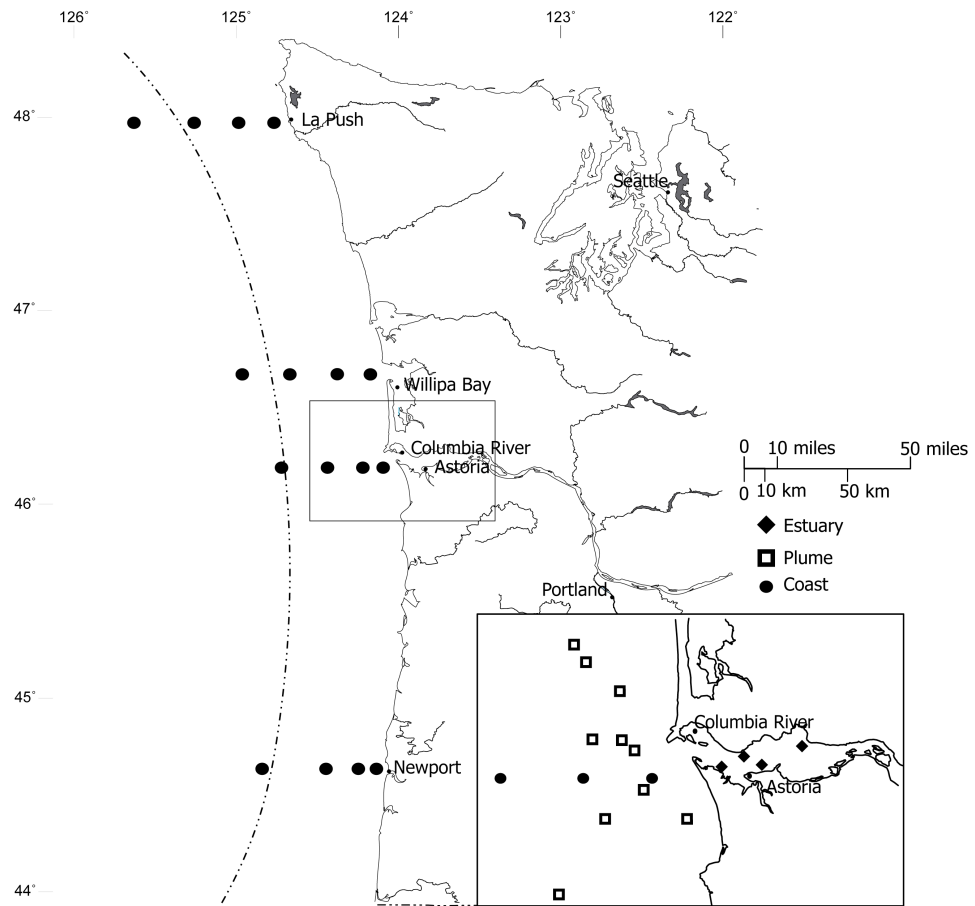


Figure 2.2: Multiple dimensional scaling (MDS) diagram of percent similarities for all 71 communities in August 2007. Bacterioplankton communities separated into five groups based on location within the study site (ANOSIM: $p < 0.003$, Stress: 0.18).

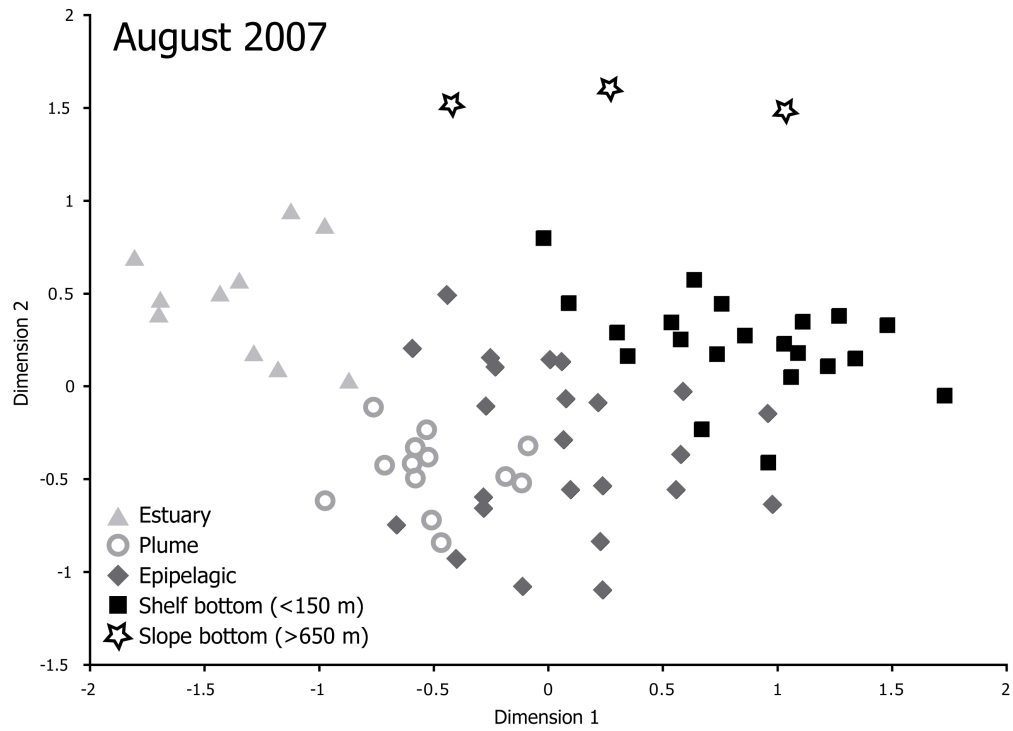
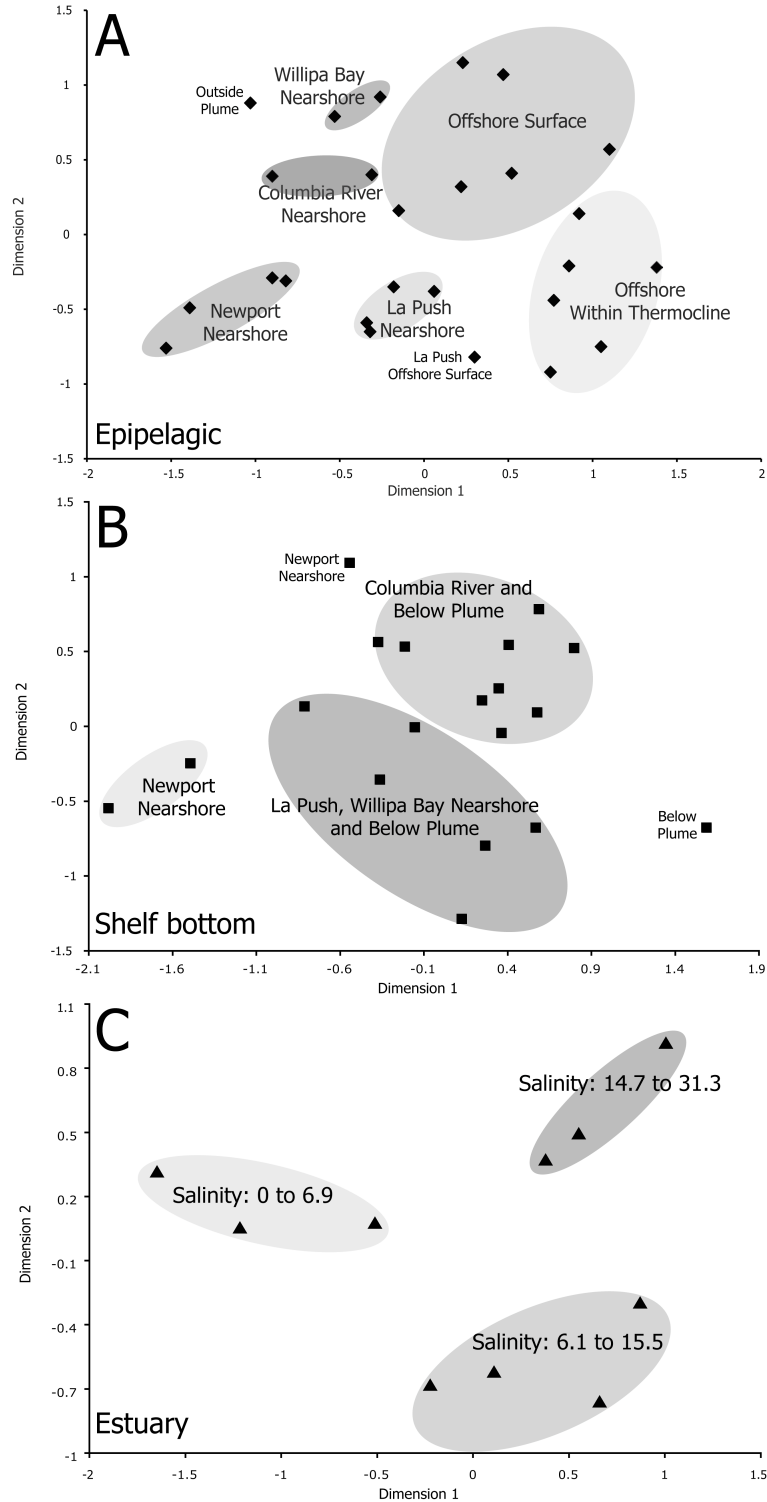


Figure 2.3: MDS diagram of epipelagic (A), shelf bottom (B), and estuary (C) bacterioplankton communities. Epipelagic and shelf bottom group communities clustered by location, estuary communities clustered by salinity. Stress values for diagrams are 0.2, 0.17, and 0.08 respectively.



CHAPTER 3:
SPATIAL VARIABILITY OVERWHELMS SEASONAL PATTERNS IN
BACTERIOPLANKTON COMMUNITIES ACROSS A RIVER TO OCEAN
GRADIENT

Abstract:

Few studies of microbial biogeography address variability across both multiple habitats and multiple seasons. Here I examine the spatial and temporal variability of bacterioplankton community composition of the Columbia River coastal margin using 16S rRNA gene amplicon pyrosequencing of 300 water samples collected in 2007 and 2008. Communities separated into seven environments (ANOSIM, $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom (depth < 350 m), and slope bottom (depth > 850 m). The ordination of these samples was correlated with salinity ($\rho = -0.83$) and depth ($\rho = -0.62$). Temporal patterns were obscured by spatial variability among the coastal environments, and could only be detected within individual groups. Thus, structuring environmental factors (e.g., salinity, depth) dominate over seasonal changes in determining community composition. Seasonal variability was detected across an annual cycle in the river, estuary, and plume where communities separated into two groups, early year (April-July) and late year (August-Nov), demonstrating annual reassembly of communities over time. Determining both the spatial and temporal variability of bacterioplankton communities provides a framework for modeling these communities across environmental gradients from the river to the deep ocean.

Introduction:

Over the past few decades, aquatic microbial communities have been shown to be abundant, deeply diverse, and variable across space and time. Yet several recent studies demonstrate repeatable and predictable patterns in the composition of these

communities. Spatial variability in aquatic microbial communities has been explored on scales that range from millimeters (Long and Azam 2001) to kilometers (Hewson et al. 2006b) to global (Fuhrman et al. 2008, Pommier et al. 2007). This variability is often attributed to a combination of environmental factors that influence the rate of growth of individual taxa and physical parameters that prevent different communities from interacting (Crump et al. 2004, Fuhrman et al. 2006, Fuhrman et al. 2008, Lozupone and Knight 2007, Nemergut et al. 2011). Of these factors, salinity, temperature, and depth appear to be the most important in distinguishing aquatic communities over large spatial scales, in part because many environmental factors vary with salinity, temperature, and depth (i.e., light, nutrients, pressure) which results in separation of water masses and thereby communities (Carlson et al. 2009, Fortunato and Crump 2011, Fuhrman et al. 2008, Morris et al. 2005, Treusch et al. 2009). On a global scale, Lozupone and Knight (2007) showed that the primary determinant of aquatic microbial community composition was salinity while Fuhrman et al (2008) found that changes in diversity of marine bacteria across a latitudinal gradient was highly correlated to temperature.

Temporal variability in marine and freshwater microbial communities is also predictable within individual environments. Seasonal shifts in microbial community composition have been demonstrated in marine environments such as the Sargasso and Baltic Seas and the English Channel, where succession of microbial communities correlated with changes in mixed layer depth, temperature, and nutrient concentrations through the year (Andersson et al. 2010, Carlson et al. 2009, Gilbert et al. 2009, Morris et al. 2005). Mixing, temperature, and nutrient concentrations are

important factors influencing communities in freshwater systems as well (Berdjeb et al. 2011, Kent et al. 2007, Nelson 2009, Shade et al. 2008). Shade et al. (2008) found distinct communities in layers of a stratified lake, where gradients of temperature, dissolved oxygen, and nutrients were present. Seasonal succession in both marine and freshwater has also been shown to be repeatable (Andersson et al. 2010, Carlson et al. 2009, Crump et al. 2009, Fuhrman et al. 2006, Morris et al. 2005, Nelson 2009). Crump et al. (2009) showed synchronous shifts in communities in six arctic rivers strongly correlated with seasonal changes in the environment, suggesting microbial communities may shift in predictable patterns from season to season.

Microbial communities are highly diverse, but the extent and the variability of this diversity in freshwater and marine systems are uncertain. High throughput pyrosequencing of PCR-amplified 16S rRNA genes is beginning to resolve the deep diversity of these systems. Due to the large number of sequences per run (~1 million reads), 16S rRNA gene amplicon pyrosequencing provides better resolution of microbial biogeographical patterns because the depth of diversity captured with each sample is greater when compared to classical community fingerprinting techniques (e.g. DGGE, T-RFLP, ARISA), which only capture the most dominant species in an environment (Sogin et al. 2006). Recent studies have used 16S amplicon pyrosequencing to determine the microbial diversity of many different environments including deep sea, arctic, soil, and estuarine communities (Andersson et al. 2010, Galand et al. 2009, Gilbert et al. 2009, Lauber et al. 2009, Sogin et al. 2006).

Microbial community composition and diversity have been characterized spatially and temporally in various environments, but rarely have they been assessed

over both spatial and seasonal scales. Using 16S amplicon pyrosequencing, I characterized bacterioplankton communities from 300 water samples collected across the Columbia River coastal margin over an annual cycle. The coastal waters of the Pacific Northwest are highly productive due to nutrient delivery from seasonal upwelling and from the Juan de Fuca strait and Columbia River (Hickey and Banas 2003). The biological and physical processes of these coastal waters are complex due to variable winds, remote wind forcing, shelf width, and submarine canyons (Hickey and Banas 2003, Hickey and Banas 2008, Hickey et al. 2010), which in turn may differentially affect the composition of bacterioplankton communities along the Oregon and Washington coasts (Fortunato and Crump 2011). The Columbia River is the second largest river in the United States with a mean annual discharge of $7300 \text{ m}^3\text{s}^{-1}$ (Hickey et al. 1998). This significant release of freshwater strongly impacts the chemical, physical, and biological characteristics of the coastal ocean including primary and secondary production within the river plume and differentially along the Oregon and Washington coasts (Hickey et al. 2010).

In a previous study from August 2007, the community fingerprinting technique DGGE was used to broadly characterize spatial variation of microbial communities in the Columbia River coastal margin (Fortunato and Crump 2011). Here I used 16S rRNA gene amplicon pyrosequencing to expand on this earlier dataset by increasing the sample size four fold and characterizing communities across multiple seasons using a more resolved spatial scale from the river to the deep ocean. I hypothesized that due to the large spatial scale of this study, bacterioplankton communities would separate from river to ocean, across salinity, depth, and other

environmental gradients that vary from fresh to marine waters. Results indicate that spatial variability overwhelmed seasonal trends across the entire sample set, and temporal variability could only be resolved within single environment types.

Methods:

Water samples were collected from the Oregon and Washington coasts and the Columbia River and estuary (latitude 44.652 and 47.917, longitude -123.874 and -125.929) as part of the NSF-funded Center for Coastal Margin Observation and Prediction (CMOP) science and technology center. Samples were collected between 2007-08 on eight cruises aboard the *R/V Wecoma* and *R/V Barnes*. Aboard the *R/V Wecoma*, water samples were collected from the Columbia River, estuary, plume, and two coastal ocean lines (Columbia River line, Newport Hydroline) in August and November of 2007 and April, June, July, and September of 2008 (Figure 3.1). For coastal lines, samples were taken at three depths per station (surface, within thermocline, and bottom). Plume samples were taken at 2 depths (surface and bottom) in 2007 and 4 depths in 2008 (surface, below plume, within thermocline, bottom). In the estuary, samples were collected based on the location of the salt gradient in both the north and south channels of the river. Samples were collected across the salt gradient from 0 to 30. Samples were collected using a conductivity-temperature-depth (CTD) rosette water sampler with 10-liter Niskin bottles. With each CTD cast, depth profiles of salinity, temperature ($^{\circ}\text{C}$), turbidity (NTU), oxygen (mgL^{-1}), and chlorophyll fluorescence were recorded. Water samples aboard the *R/V Barnes* were collected using a high volume low-pressure pump over salinity gradients in the

estuary in August 2007 and July 2008. For all samples, surface was defined as being between 1 and 2 m depth and bottom was defined as being between 1 and 5 m above sediment. Data from CTD fluorescence and temperature sensors were used to determine exact sampling depths for water collected at the chlorophyll maximum and within the thermocline.

DNA samples (1-6 L per sample) were collected, preserved and extracted as described previously (Fortunato and Crump 2011) using methods adapted from Zhou et al (1996) and Crump et al (2003). Extracted DNA was PCR-amplified using primers targeting bacterial 16S ribosomal RNA genes. Each sample was assigned a uniquely barcoded reverse primer and amplified in four replicate 20 µl reactions (Hamady et al. 2008). Primers used for amplification were bacteria-specific primers focusing on the V2 region, 27F with 454B FLX linker (GCCTTGCCAGCCCGCTCAG TC AGRGTTTGATYMTGGCTCAG) and 338R with 454A linker and unique 8 basepair barcode, denoted by N in primer sequence (GCCTCCCTCGCGCCATCAG NNNNNNN CA TGCWGCCWCCCGTAGGGWT) (Modified from (Hamady et al. 2008)). Replicate amplifications were combined, purified, and normalized using Invitrogen SequelPrep normalization plates (Invitrogen, Carlsbad, CA, USA). Five µl from each sample were combined into a single tube and sent for pyrosequencing on a Roche-454 FLX pyrosequencer at Engencore at the University of South Carolina (<http://engencore.sc.edu/>).

Sequence data was processed using two different methods: 1) Manual global alignment and removal of pyrosequencing errors using ARB (Ludwig et al. 2004) and

MOTHUR software (Schloss et al. 2009), and 2) denoising and pairwise alignment using the QIIME (v1.2.0) software package (Caporaso et al. 2010).

For the first method, raw sequences were sorted and quality controlled (min length 150 bp, no ambiguous bases) using the Ribosomal Database Project (RDP) Pyro tools (Cole et al. 2005). A reference sequence database was created using the community analysis program MOTHRUR (Schloss et al. 2009) consisting of unique sequences from the overall dataset. These unique sequences were imported into ARB and manually aligned. Extra bases commonly added in pyrosequencing (i.e., pyronoise) were placed in gaps added to the alignment. Once the manual alignment was completed, sequences were trimmed to *E. coli* basepair positions 136-335 and were exported using a 3% basepair frequency filter to mask insertions, but include variable bases. This reference dataset of manually aligned unique sequences was then used to align the entire dataset using MOTHRUR. Our approach removed insertions from pyrosequencing, but did not repair deletions of bases, which were included in downstream analyses. OTUs were determined based on 97% sequence similarity using MOTHRUR.

For the second method using QIIME, sequences were quality controlled using the Split_Libraries.py script with default settings (min length 200, max length 1000, min mean quality score 25, max ambiguous bases 0, max homopolymer length 6, max primer mismatch 0). To account for pyronoise, the remaining sequences were denoised using the denoiser.py script with the 'fast' method and default settings. Sequences were then clustered using the pick_otus.py script with the uclust method (97% sequence similarity). Potentially chimeric sequences were identified among

representative sequences from each OTU with ChimeraSlayer, and a total of 3,952 sequences composing 196 OTUs were eliminated from the dataset.

For both methods, relative abundance was calculated for the OTUs in each sample and used to calculate pairwise similarities among samples using the Bray-Curtis similarity coefficient (Legendre and Legendre 1998). I also calculated pairwise similarities among samples using both weighted and unweighted UNIFRAC metrics (Lozupone et al. 2006), but the results were nearly identical to those based on Bray-Curtis, and so are not presented. Bray-Curtis similarity matrices were visualized using multiple dimensional scaling (MDS) diagrams, a form of ordination. Analysis of similarity statistics (ANOSIM) was calculated to test the significance of differences among a priori sampling groups based on environmental parameters. Similarity matrices, MDS diagrams, and analysis of similarity (ANOSIM) statistics were carried out using PRIMER v6 for Windows (PRIMER-E Ltd, Plymouth, UK).

Alpha diversity for samples was calculated using MOTHUR. The number of sequences was normalized before calculation by randomly selecting the same number of sequences per sample, based upon the sample with the least number of sequences (n=209 sequences). The taxonomy of OTUs identified was determined using the RDP Classifier tool. Taxonomic assignments with less than 80% confidence were marked as unknown. A total of 306 samples were analyzed overall. This number was reduced to 300 as samples with a low number of sequences were removed.

All sequences can be downloaded from the NCBI Sequence Read Archive (SRA) database under the accession number SRP006412.

Results:

Comparison of the two sequence analysis methods showed that the overall patterns of microbial community structure for this study are highly robust, as both spatial and temporal patterns in beta-diversity were the same for both methods. The number of OTUs identified by the QIIME analysis (8039) was slightly lower than the ARB/MOTHUR analysis (9389), but this was because fewer sequences passed the initial QIIME quality control step due to different quality control parameters including maximum homopolymers length and primer mismatches. Since the patterns of community variability were comparable, the results presented are based on the QIIME sequence analysis protocol.

Bacterioplankton communities separated into seven distinct environments (ANOSIM, $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom, and slope bottom. The plume consisted of coastal surface samples with salinity less than 31, the epipelagic included coastal surface and chlorophyll maximum samples (average depth=8 m), the mesopelagic consisted of coastal samples within and below the thermocline (average depth=44 m), the shelf bottom of bottom samples with depth less than 350 m, and the slope bottom of bottom samples deeper than 850 m. Percent similarity for all samples was 22.9% ($\pm 15.3\%$) with a range from 0 to 74.8% similarity. Similarity values were higher within groups than between groups (Table 3.1).

A multidimensional scaling plot of all 300 samples based on Bray-Curtis similarity values (Figure 3.2) depicts the seven environments based on location in the system. Environments separated along two axes that form a V-shaped arrangement of

microbial communities. The first axis is clearly related to salinity, and the second is related to depth. There was a strong correlation between Dimension 1 and salinity, with a Spearman's rho (ρ) value of -0.83 ($p < 0.001$, Figure 3.3). A weaker relationship was observed between Dimension 2 and sample depth ($\rho = -0.62$, $p < 0.001$ for Dimension 2 axis and depth), although this relationship improved when river and estuary samples were omitted ($\rho = -0.76$, $p < 0.001$).

Spatial variation in communities based on sampling location is readily apparent in Figure 3.2. Temporal variation, however, appears to be overwhelmed by the strong spatial gradients of salinity and depth. Temporal variation was only detectable when each spatial group was analyzed separately. For river, estuary, and plume samples, a seasonal trend is apparent from river to ocean (Figure 3.4). In the river, three communities are visible, spring, freshet-early summer, and late summer-fall. In the estuary, seasonal clustering of communities was not as clear, although communities did split into two significant clusters (ANOSIM, $p < 0.001$), an early year community, encompassing samples from April to July, and a late year community encompassing samples from August to November. These same two communities, early and late, are also present in the plume (ANOSIM, $p < 0.001$). The seasonal pattern in the other groups is less discernable. There was significant seasonal variation in the shelf bottom and epipelagic environments according to ANOSIM statistics, but these patterns could not be discerned in the individual MDS plots due to the large amount of variability within each group. There was no significant temporal pattern in the slope bottom or mesopelagic.

Most sequences in the dataset were related to the phyla Proteobacteria (44.7%) and Bacteroidetes (33.6%). Within the Proteobacteria, Alpha (21.2%), Gamma (17.0%), Beta (2.6%), and Delta (0.4%) were present. In the Bacteroidetes, Flavobacteria was the largest group, with 55,915 sequences, making up 28% of the total dataset. The most abundant OTU belonged to the SAR11 clade and consisted of 16,635 sequences. Overall, SAR11 made up 11.3% of the dataset with a total of 22,454 sequences belonging to 208 OTUs. The second largest OTU was a Gammaproteobacteria with 13,137 sequences. Cyanobacteria were a small percentage of the total dataset, only 1.8%, but constituted as much as 19% of sequences in epipelagic samples collected off the shelf. More specific taxonomic information for each of the seven spatial groups can be found in the supplemental material (Table 3.S1, Figure 3.S1).

To better understand community composition, I classified each of the 8039 OTUs in this study based on the location in the system where they exhibited their maximum average relative abundance in pooled sequences (Figure 3.5). For example, if OTU-1 was most abundant in the plume (based on its relative abundance within each pool of sequences from the seven groups), it was classified as a plume OTU. Results suggest mixing of water masses and microbial communities from estuary to the shelf bottom. The river and slope bottom groups appear to be end members in the system, as most of the river and slope bottom sequences are found only in their respective locations. The estuary community is primarily a mix of sequences belonging to river and estuarine OTUs, with some addition from the plume and epipelagic. In the plume, however, plume sequences are mostly classified as being

from plume, epipelagic, and mesopelagic OTUs with few sequences coming from river or estuary OTUs.

I mapped the relative abundance of the top OTU from each of the seven spatial groups (based on average relative abundance per group) using the ordination of Figure 3.2. These bubble plots show that the top OTUs for each group are most abundant in samples from their location and less abundant in neighboring locations (Figure 3.S2). The top OTUs for the estuary and the river display some seasonality, with the largest abundances occurring in only one or two seasons (e.g., June and July 2008 for the estuary).

Alpha-diversity varied across the spatial groups (Figure 3.6). The river and slope bottom groups had the highest and third highest average diversity (Chao1=1104 and 868, respectively), indicating the presence of many more endemic taxa within these two environments, and showing further that freshwater and deep ocean represent end members in this study. As water mixes from the river to the coastal surface ocean, diversity measurements decrease to the lowest diversity in the epipelagic group (Chao1=380). Diversity then increased from surface to the deep ocean, with the mesopelagic, shelf bottom, slope bottom groups each having a higher diversity than the previous. Diversity measurements show that when water mixes from fresh to salt and from deep to surface, taxa are reduced in abundance beyond our limit of detection and thus community composition becomes more streamlined in the coastal surface.

Discussion:

Previous studies of variability and diversity in bacterioplankton communities are restricted to single dimensions, focusing either on long-term time series, depth profiles, or horizontal surveys across environmental gradients (Andersson et al. 2010, Fuhrman and Steele 2008, Gilbert et al. 2009, Hewson et al. 2006a, Lozupone and Knight 2007, Morris et al. 2005, Nemergut et al. 2011, Pommier et al. 2007, Treusch et al. 2009). Here I present a dataset that compares bacterioplankton community composition in all three of these dimensions: spatially from river to surface ocean, by depth from surface to deep ocean, and through time seasonally over an annual cycle. This larger scale biogeographical analysis was enabled by the use of 16S amplicon pyrosequencing, which assesses diversity through DNA sequencing of hundreds of thousands of PCR-amplified gene copies. Previous 16S amplicon pyrosequencing studies focused on deep sampling of small numbers of samples, allowing for characterization of the “rare biosphere” but only at limited spatial and temporal scales (Andersson et al. 2010, Galand et al. 2009, Gilbert et al. 2009, Kirchman et al. 2010). In this study I took a different approach to characterizing bacterioplankton communities by applying 16S amplicon pyrosequencing to ten times the number of samples seen in previous studies. Sequencing more samples produces fewer sequences per sample and limits the resolution of the rare biosphere. However, the greater number of samples in this study (n=300) led to the discovery of robust spatial patterns from river to ocean, and seasonal shifts that may not have been observed if fewer samples were sequenced. Based on a previous community fingerprinting study of 71 samples from August 2007 using DGGE, I found that communities separated

into just five environments defined by location across salinity and depth gradients (Fortunato and Crump 2011). With the addition of over four times the number of samples, in this study I was able to further resolve the spatial patterns of bacterioplankton communities into seven distinct environments across steep salinity and depth gradients in addition to determining temporal variability.

Salinity and depth change dramatically from the Columbia River to the deep ocean, and these factors appear to strongly influence the composition of bacterioplankton communities. In contrast, temporal variability in bacterioplankton communities was relatively small, and was obscured by the spatial variability in communities across environments in the coastal zone. Several studies of coastal zone bacterioplankton identify time as the principle axis of community variability (Fuhrman et al. 2006, Gilbert et al. 2009, Kan et al. 2006b, Stepanauskas et al. 2003), but these studies were restricted to one environment type (e.g., estuaries or a fixed coastal station) within which spatial variability of bacterioplankton communities was limited. Few studies address temporal variability across many different habitats, so it was difficult to compare our results to other studies. However, one study by Kirchman et al. (2010) identified a similar pattern among 11 surface water samples in which winter/summer differences in Arctic Ocean bacterioplankton communities was minimal compared to spatial variability across their sampling range. Thus, while temporal variability may occur within many marine habitats, it is clear that structuring environmental factors (e.g., salinity, depth) dominate over seasonal changes in determining community composition.

Spatial differentiation among samples was highly correlated with salinity, confirming the observations of two global meta-analyses of microbial diversity based on 16S rRNA gene sequences (Lozupone and Knight 2007, Tamames et al. 2010). In one of these studies, Lozupone and Knight (2007) found that salinity was the primary environmental determinant for community composition across marine, freshwater, sediment, and soil environments, more so than temperature, pH, or other environmental factors. In the coastal marine environment, salinity contributes to density gradients that physically separate water masses and their resident microbial communities. However, the degree to which these water masses are separated depends on the magnitude of mixing via river flow, tides, upwelling, surface winds, etc. This mixing from fresh to marine or from surface to deep leads to the formation of communities in mixing zones that are comprised of bacterioplankton populations from multiple water masses. For example in the Columbia River estuary, the flushing rate exceeds the doubling time of bacterioplankton populations, thus a distinct free-living estuarine community is unable to form (Crump et al. 1999). Our study confirmed this observation, demonstrating that estuarine bacterioplankton communities are composed of populations from the river and the coastal ocean. I also identified significant overlap in communities across environmental gradients in the coastal ocean including the plume, epipelagic, mesopelagic and shelf bottom environments, although it is unclear whether this is the result of mixing or the presence of generalist organisms that thrive in different environments.

Coastal bacterioplankton communities correlated with depth from surface to the deep ocean despite the fact that samples were collected over multiple seasons and

at sampling sites as much as 150 km apart. Salinity varies with depth, as do many other environmental parameters including temperature, light, and nutrients. I therefore am treating depth here as a proxy for many factors that vary in the vertical dimension. The vertical structuring of bacterioplankton communities in the ocean has been demonstrated in many studies and has been linked to changes in hydrostatic pressure as well as water mass properties (Blumel et al. 2007, Carlson et al. 2009, Lee and Fuhrman 1991, Morris et al. 2005, Treusch et al. 2009). For example, Treusch et al (2009) found that Sargasso Sea bacterial communities separated into surface (upper 40 m), deep chlorophyll maximum, and upper mesopelagic communities. I also observed a separation of the epipelagic and upper mesopelagic communities, but not between surface and chlorophyll maximum samples, possibly because the mixed layer depth (5 to 56 m) was, in general, shallower than in the Sargasso Sea (<50 m to 350 m) (Carlson et al. 2009, Treusch et al. 2009). Treusch et al (2009) attributed separation of these communities to stratification and seasonal mixing in the upper water column. The coastal zone of the Pacific Northwest experiences seasonal upwelling, and thus a mixing of communities from bottom to surface. The degree of mixing is evident in Figure 3.5, where the mesopelagic is actually a mix of populations from the bottom and surface. In July 2008 during strong upwelling, near-shore surface samples from the Newport Hydroline contained a higher proportion (23%) of sequences belonging to shelf bottom and slope bottom OTUs than during other times in 2008 (5%). Also during that month, the most abundant estuary-classified OTU was found in some shelf bottom samples, indicating a possible exchange between these two environments.

Temporal variability could only be resolved within some environments. Seasonal changes were observed in the river, estuary, and plume environments. In the river, there were three separate groups, spring, freshet-early summer, and late summer-fall, corresponding to seasonal changes in Columbia River discharge, where maximum discharge occurs in late spring and is minimum in late summer to early fall (Prahl et al. 1998). In the plume, seasonal upwelling strongly influences temperature and nutrient concentrations and thereby production in the plume (Hickey et al. 2010). Thus, plume community composition is tightly linked to the physical processes occurring along the coast. The seasonality of the estuary community then can be attributed to a combination of both river and coastal processes. The periods of maximum and minimum discharge of the river correspond to the two seasonal bacterioplankton groups seen in the estuary, early (April to July) and late (August to November). During times of high river flow, the estuarine community is shaped by the river and when river flow is at a minimum, community composition is influenced more by the plume and coastal ocean.

River and deep ocean (slope bottom) appear to be end members in this system in that they contribute populations to nearby environments, but receive little to no contributions themselves. In the other five environments there was tremendous overlap in community composition from estuary to shelf bottom suggesting dynamic exchange of communities through advection and mixing. Within each environment there also appeared to be environment-specific communities, based on maximum relative abundance. In the plume, 37% of plume sequences were classified as belonging to plume OTUs, indicating the presence of a plume-specific community.

Additionally, only 5% of plume sequences were from the river and estuary while 36% came from epipelagic and mesopelagic OTUs, indicating the plume community is comprised more of coastal populations than bacteria flushed from the estuary. As mentioned previously, the plume is highly productive due to nutrient delivery from the river and coastal upwelling (Hickey et al. 2010) and as primary production increases in the plume, different epipelagic taxa could increase depending on availability and quality of organic matter. This would result in a different combination of bacteria populations and a clear distinction between the plume and epipelagic communities. I speculate then that each spatial group, from estuary to shelf bottom, contains bacterioplankton populations that are broadly distributed across environments, but each group supports a different combination of these bacteria, creating distinct communities within each environment.

16S amplicon pyrosequencing, like any molecular technique is prone to errors and it is important to analyze sequences in a way that accurately assesses community patterns. Analyzing 16S amplicon pyrosequencing data is difficult due to sequencing errors termed “pyronoise”, which may artificially increase the number of OTUs observed. In Kunin et al. (2009), the authors PCR-amplified a 300bp region of the 16S rRNA gene from a known cultured *E. coli* strain and then pyrosequenced it. The results returned a largely inflated number of OTUs, showing that pyrosequencing errors may lead to a gross overestimation of the number of OTUs in a sample. An increase in the number of OTUs leads to inflated alpha diversity within samples, and greater beta diversity between samples. I found that global alignment combined with manual removal of pyronoise insertions was comparable in total OTU number, alpha-

diversity, and beta-diversity patterns to analysis using a QIIME analysis pipeline that includes denoising (denoiser.py) and pairwise sequence alignment (uclust). I also found that removing the pyronoise is crucial for minimizing the total number of OTUs and overall sequencing errors. To demonstrate this I globally aligned our sequences using a reference database from SILVA (Pruesse et al. 2007) and found that although our beta-diversity patterns were comparable, the OTU number and alpha diversity estimates were nearly twice that of our previous methods (data not shown). It is important then that pyrosequencing datasets be subjected to rigorous quality checking and denoising in order to accurately assess both the overall community patterns and the rare biosphere.

Table 3.1: Percent similarity values within and between groups \pm standard deviation (ANOSIM: $p < 0.001$) as determined by Bray-Curtis similarity coefficient.

	River	Estuary	Plume	Epipelagic	Mesopelagic	Shelf Bottom	Slope Bottom
River	33.6 \pm 11.2						
Estuary	17.8 \pm 11.7	24.9 \pm 11.1					
Plume	2.6 \pm 4.1	19.2 \pm 11.4	36.5 \pm 10.5				
Epipelagic	0.1 \pm 0.2	15.4 \pm 11.2	35.0 \pm 9.2	38.4 \pm 9.9			
Mesopelagic	0.4 \pm 1.5	13.1 \pm 10.0	27.3 \pm 11.1	33.4 \pm 11.8	37.2 \pm 11.5		
Shelf Bottom	0.2 \pm 0.3	10.3 \pm 8.6	17.9 \pm 9.5	22.8 \pm 10.9	32.2 \pm 12.1	41.8 \pm 12.6	
Slope Bottom	0.2 \pm 0.3	4.3 \pm 4.0	6.2 \pm 5.8	8.7 \pm 6.3	16.4 \pm 9.6	24.7 \pm 9.0	50.3 \pm 6.6

Figure 3.1: Map of Oregon and Washington coasts. Inset depicts Columbia River estuary and plume region. Dotted line denotes approximate location of shelf break.

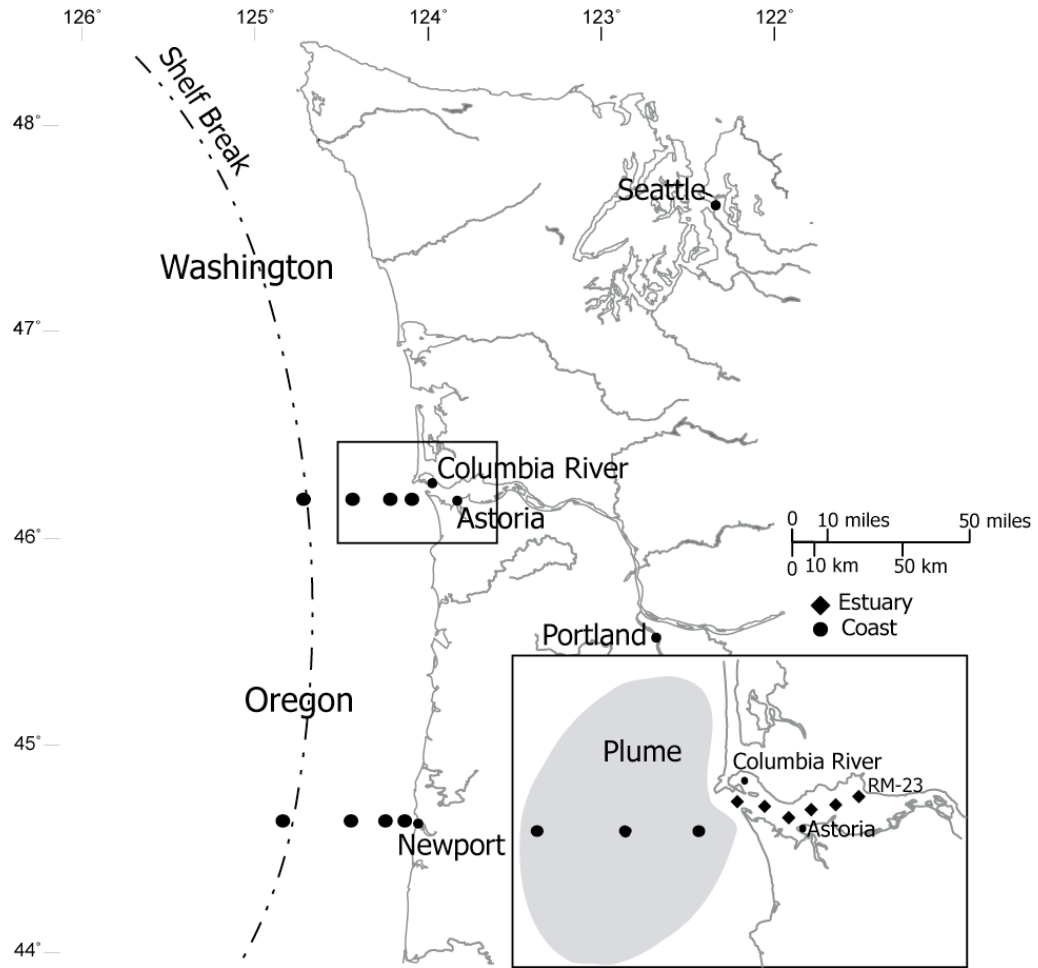


Figure 3.2: Multiple dimensional scaling diagram of percent similarities for all 300 samples. Bacterioplankton communities separated into seven groups based on location, across salinity and depth gradients (ANOSIM: $p < 0.001$, Stress: 0.12).

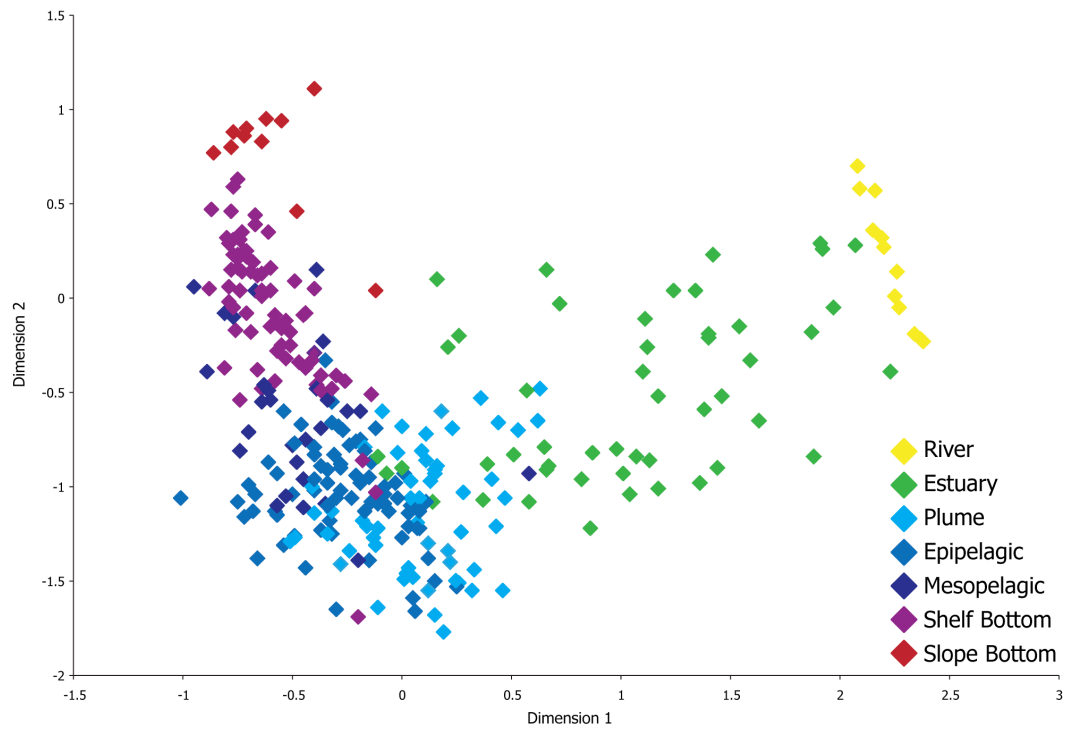


Figure 3.3: Correlation of Dimension 1 for the 300 samples from Figure 3.2 and salinity. A Spearman's rho (ρ) value of -0.83 ($p < 0.001$) indicates a strong relationship between salinity and bacterial community variation.

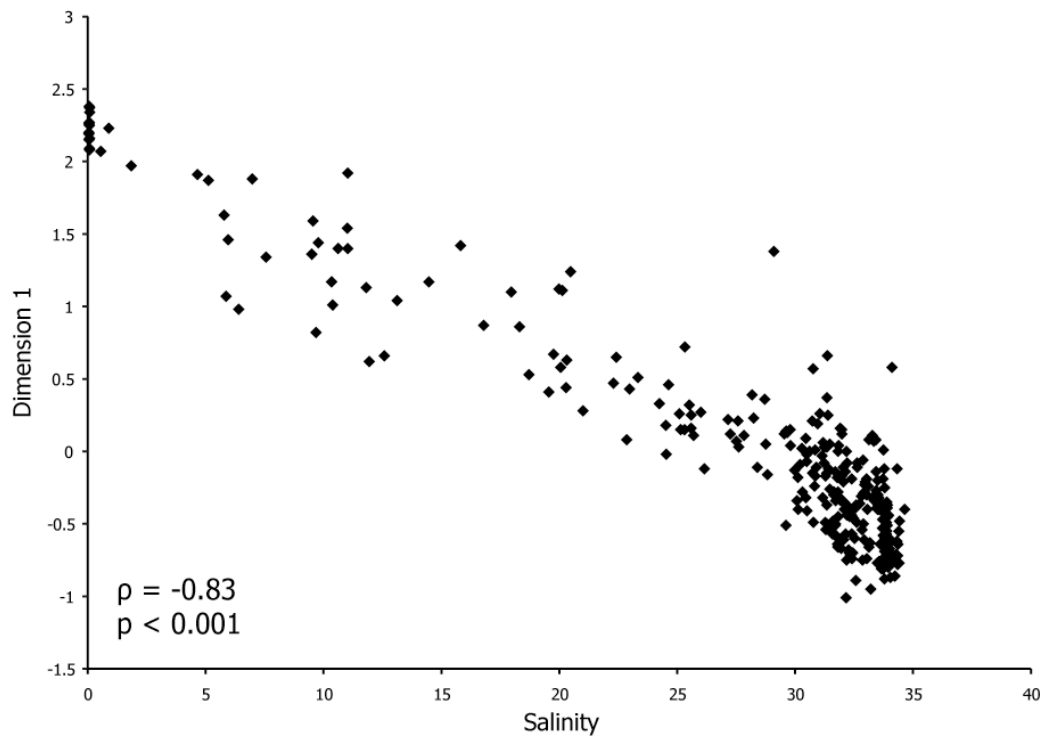


Figure 3.4: Seasonal multidimensional scaling diagram of river, estuary, and plume. River displays three seasonal communities, which cluster into two communities, early (April to July) and late (August to November), in the estuary and plume. Stress = 0.04, 0.15 and 0.17 for river, estuary, and plume, respectively.

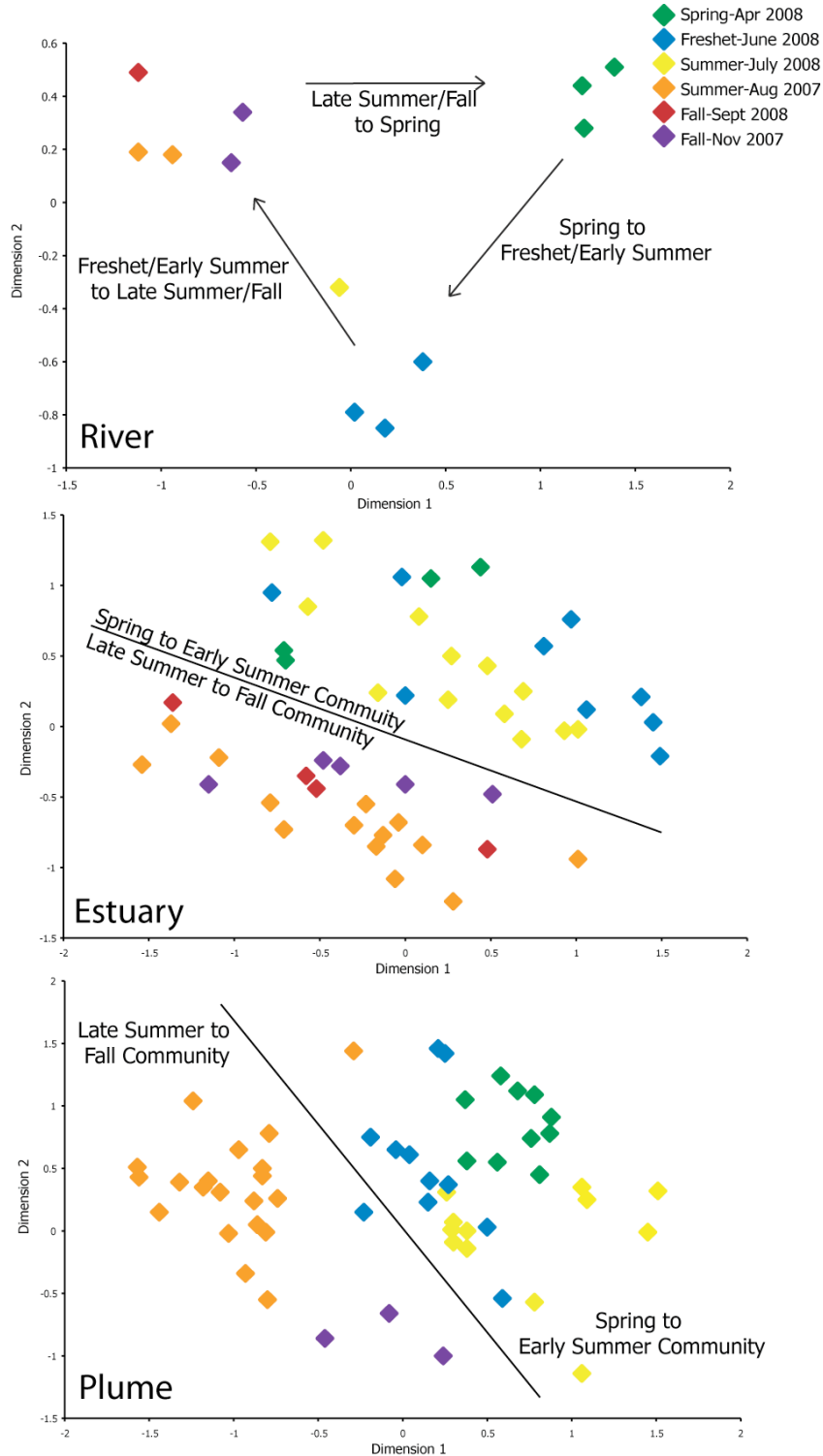


Figure 3.5: Percentage of sequences in OTUs classified by location. Slope bottom and river groups represent end members in the system. Rare category represents sequences belonging to OTU that make up less than 0.1% of the total number of sequences from each corresponding location.

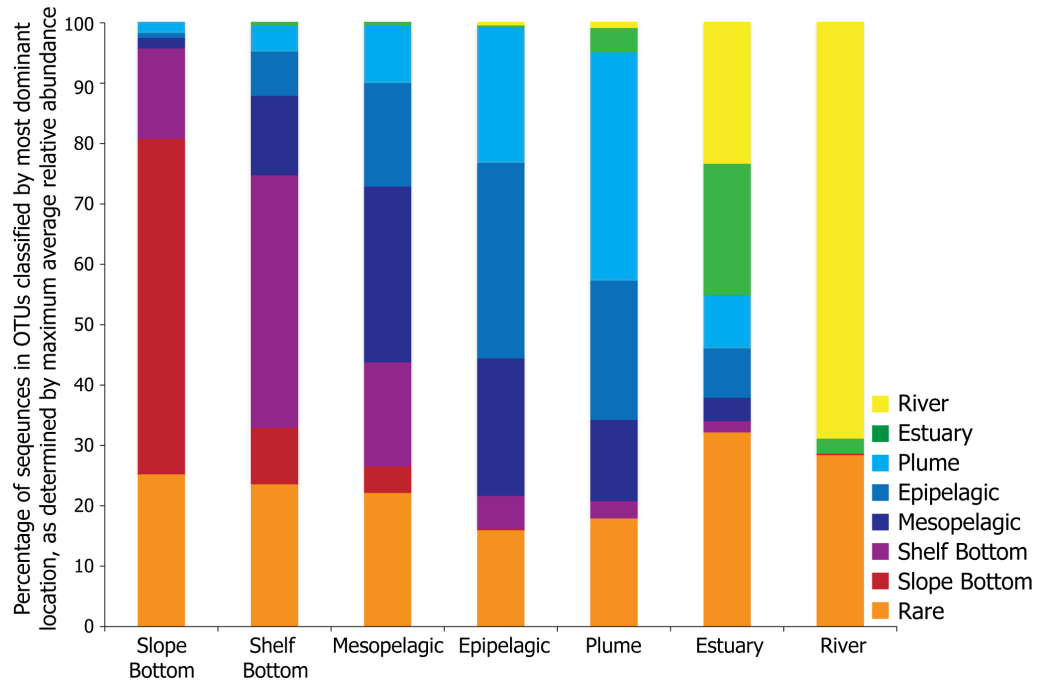
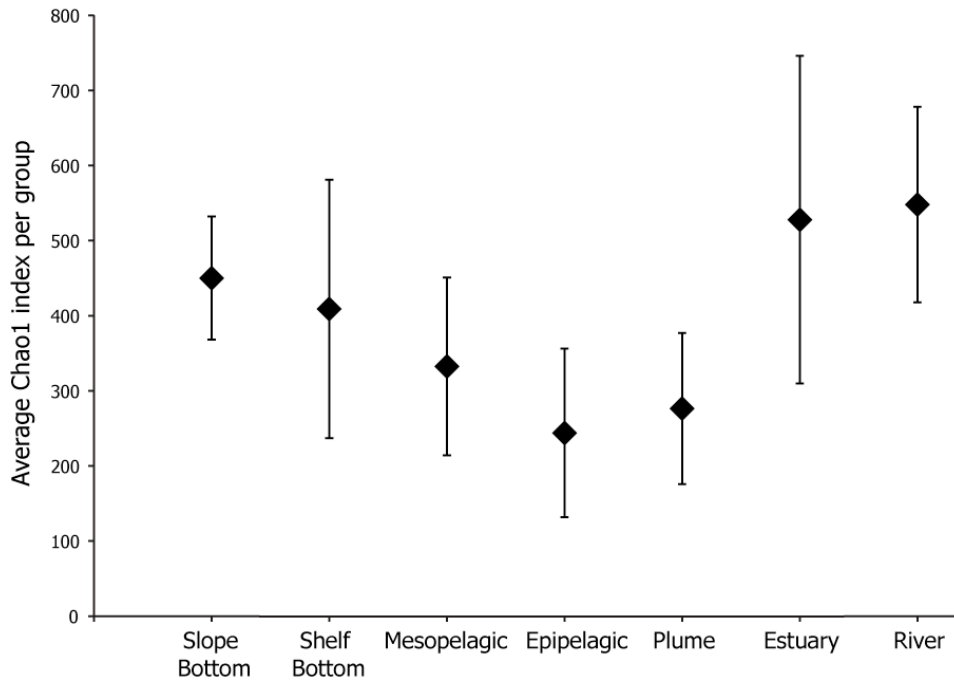


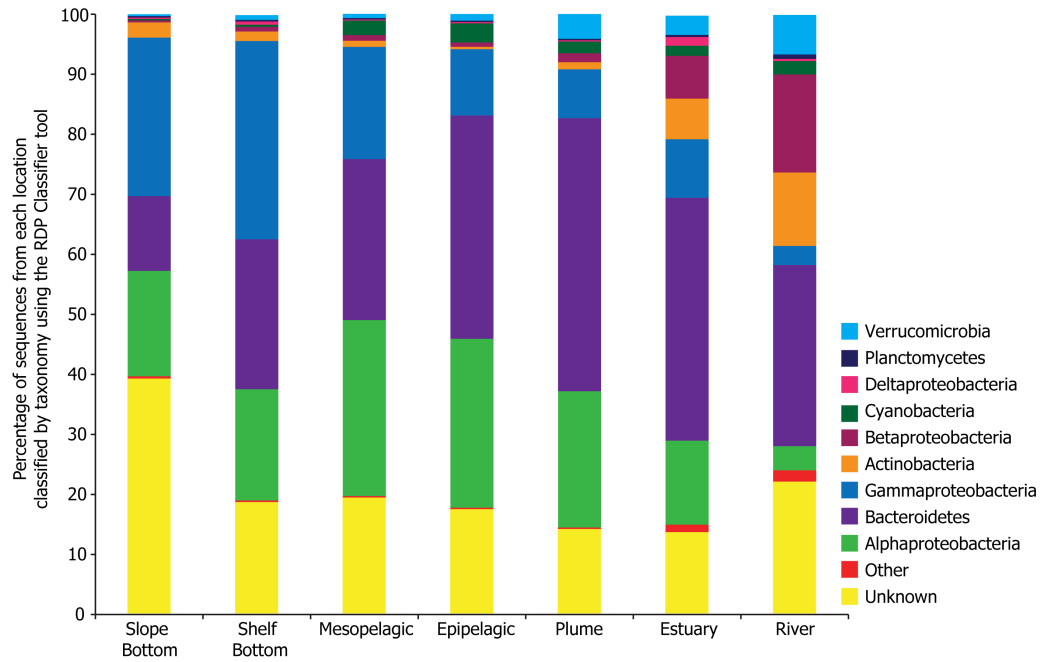
Figure 3.6: Average Chao1 index per group \pm standard deviation as determined using MOTHUR (v.1.15.0). OTU number was normalized to the sample with the smallest number of sequences (n=209 sequences).



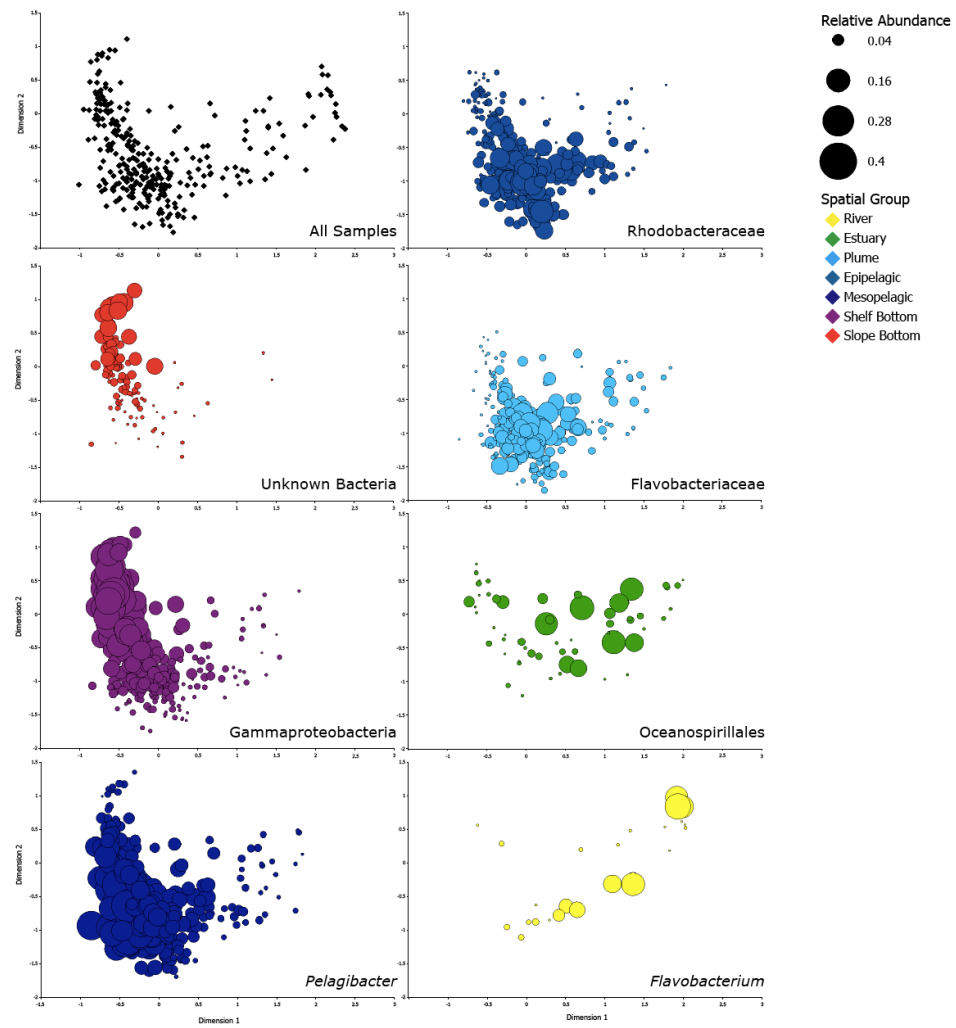
Supplemental Table 3.1: Most abundant OTUs within each group based on average relative abundance. Taxonomy was determined by the RDP Classifier (80% confidence threshold).

Environment	Taxonomy (Phylum, Class, Order, Family, Genus)	Average relative abundance within group (%)	
River	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>	4.1	
	Proteobacteria, Betaproteobacteria, Burkholderiales, Comamonadaceae	4.0	
	Proteobacteria	3.4	
	Unknown Bacteria	3.3	
	Actinobacteria, Actinobacteridae, Actinomycetales	2.7	
	Proteobacteria, Betaproteobacteria, Methylophilales, Methylophilaceae	2.4	
	Actinobacteria, Actinobacteridae, Actinomycetales	2.4	
	Verrucomicrobia	1.8	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>	1.6	
	Actinobacteria	1.3	
		Sum: 27.0	
	Estuary	Proteobacteria, Gammaproteobacteria, Oceanospirillales	2.5
		Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	2.4
Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae		1.2	
Bacteroidetes, Flavobacteria, Flavobacteriales, Cryomorphaceae, <i>Fluvicola</i>		1.2	
Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae		1.2	
Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae		1.1	
Proteobacteria, Betaproteobacteria, Burkholderiales, Comamonadaceae		1.1	
Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>		0.9	
Bacteroidetes		0.8	
		0.7	
		Sum: 13.0	
Plume	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	3.9	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Ulvibacter</i>	3.5	
	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae	2.7	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>	2.0	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.8	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Cryomorphaceae	1.7	
	Unknown Bacteria	1.5	
	Bacteroidetes, Flavobacteria, Flavobacteriales	1.3	
	Verrucomicrobia, Opitutae	1.2	
	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae	1.2	
		Sum: 20.8	
Epipelagic	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae	5.6	
	Unknown Bacteria	3.0	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	2.0	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	2.0	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.8	
	Cyanobacteria, Cyanobacteria, Family II, GpIIa	1.7	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.6	
	Proteobacteria, Gammaproteobacteria	1.5	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.2	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.2	
		Sum: 21.7	
Mesopelagic	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>	14.0	
	Proteobacteria	1.5	
	Proteobacteria, Alphaproteobacteria	1.2	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae,	1.1	
	Proteobacteria, Gammaproteobacteria	1.0	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>	0.9	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	0.9	
	Unknown Bacteria	0.7	
	Bacteroidetes, Flavobacteria, Flavobacteriales	0.6	
	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>	0.4	
		Sum: 22.2	
Shelf Bottom	Proteobacteria, Gammaproteobacteria	16.4	
	Proteobacteria, Gammaproteobacteria	3.2	
	Proteobacteria, Gammaproteobacteria	2.2	
	Proteobacteria, Gammaproteobacteria	2.2	
	Unknown Bacteria	1.1	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.0	
	Proteobacteria, Gammaproteobacteria	1.0	
	Actinobacteria	0.8	
	Bacteroidetes	0.7	
	Bacteroidetes	0.7	
		Sum: 29.3	
Slope Bottom	Unknown Bacteria	8.4	
	Unknown Bacteria	4.2	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	2.7	
	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>	1.8	
	Proteobacteria, Gammaproteobacteria	1.8	
	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>	1.7	
	Proteobacteria, Gammaproteobacteria	1.3	
	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae	1.2	
	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>	1.1	
	Unknown Bacteria	0.9	
		Sum: 25.1	

Supplemental Figure 3.1: Percentage of sequences for each spatial group classified using the RDP Classifier tool with an 80% confidence threshold.



Supplemental Figure 3.2: Spatial distribution of the most abundant OTUs based on average relative abundance in each spatial group across the ordination of Figure 2. Taxonomic identification for each OTU is listed to the furthest classification possible as determined by the RDP Classifier tool with an 80% confidence threshold. Stress = 0.12



CHAPTER 4:
INDICATOR TAXA ACROSS SPATIAL AND SEASONAL GRADIENTS IN THE
COLUMBIA RIVER COASTAL MARGIN

Abstract:

Bacterioplankton communities are deeply diverse and highly variable across space and time, but several recent studies demonstrate repeatable and predictable patterns in this diversity. I expanded upon previous studies by determining patterns of variability in both individual taxa and overall bacterial communities across coastal aquatic environmental gradients. I surveyed bacterioplankton diversity across the river-to-ocean gradient in the Columbia River coastal margin, USA, using amplicon pyrosequencing of 16S rRNA genes from 596 water samples collected between 2007 and 2010. Our results showed seasonal shifts and annual reassembly of bacterioplankton communities in the freshwater influenced environments of the Columbia River, estuary, and plume. I identified indicator taxa, including species from freshwater SAR11, Oceanospirillales, and Flavobacteria groups that characterize the changing seasonal conditions in these environments. In the river and estuary, Actinobacteria and Betaproteobacteria indicator taxa correlated strongly with seasonal fluctuations in particulate organic carbon ($\rho=-0.664$) and estuarine residence time ($\rho=0.512$) respectively. In contrast, seasonal change in communities was not detected in the coastal ocean but instead varied more with the spatial variability of environmental factors including temperature and dissolved oxygen. Indicator taxa of coastal ocean environments included SAR406 and SUP05 taxa from the deep ocean and *Prochlorococcus* and SAR11 taxa from the upper water column. Overall, I found that in the Columbia River coastal margin, freshwater influenced environments were relatively consistent and predictable whereas variability in the physical conditions of the coastal ocean complicated community composition. This study moved beyond

beta-diversity patterns to focus on the occurrence of specific taxa and lends insight into the potential ecological roles these abundant taxa play in coastal ocean environments.

Introduction:

Microbial biogeographical studies describe spatial and temporal distribution patterns of microbes, and shed light on the degree of diversity, patterns of dispersal, and levels of species interactions in different environments (Martiny et al. 2006). We know from these studies that microbes are not ubiquitously distributed. Microbial communities display a high degree of endemism and contain few cosmopolitan taxa (Nemergut et al. 2011, Pommier et al. 2007). In aquatic systems, these communities vary spatially with water mass characteristics including depth, salinity, temperature, pH, hydrological conditions, organic matter, phytoplankton interactions, and other biological and physical factors (Crump et al. 2003, Field et al. 1997, Fortunato and Crump 2011, Fuhrman et al. 2008, Galand et al. 2010, Ghiglione et al. 2008, Kent et al. 2007, Lauber et al. 2009, Lozupone and Knight 2007). On temporal scales, seasonal succession and annual reassembly of communities have been shown in both fresh and marine waters (Andersson et al. 2010, Carlson et al. 2009, Crump et al. 2009, Fuhrman et al. 2006, Morris et al. 2005, Nelson 2009).

Biogeography of microbial communities has been well described across spatial and seasonal scales (Crump et al. 2012, Fortunato et al. 2012, Galand et al. 2010, Gilbert et al. 2009, Kirchman et al. 2010), and although most studies discuss changes in community composition and structure, there is little focus on the specific

taxa that distinguish one community from another. In one long-term study of the English Channel, Gilbert et al. (2012) showed strong repeatable seasonal patterns in the general bacterioplankton community and in specific taxonomic groups, and found that seasonal variability was influenced by length of day, temperature, nutrients, and photosynthetically active radiation, rather than biomass of zooplankton species. Eiler et al. (2012) found complex interdependences in bacterioplankton communities and synchronous temporal changes for specific taxonomic groups in a freshwater lake system. These studies described variation in specific taxa over time relative to variation in environmental conditions, but both studies were conducted at one or a handful of fixed locations, and did not consider the effect of immigration.

Several studies of common marine bacterial taxa, such as SAR11, SAR86, and *Prochlorococcus* describe spatial and temporal distributions of these groups over different conditions (Carlson et al. 2009, Malmstrom et al. 2010, Morris et al. 2012, Treusch et al. 2009). Malmstrom et al. (2010) showed consistency in the depth distribution of ecotypes of *Prochlorococcus* over a five-year period in the Pacific. The depth distribution of SAR11 ecotypes has also been well described, with distinct populations observed in the epipelagic and upper mesopelagic at the long-term Bermuda Atlantic Time Series (BATS) (Carlson et al. 2009), and across gradients in nutrient, chlorophyll, and organic carbon concentrations in the South Atlantic (Morris et al. 2012). What is not fully described in these studies, however, is how patterns in the abundance of these taxonomic groups are related to overall patterns of community distribution in these environments.

I expanded on these studies by describing patterns of variability in both individual taxa and overall bacterial communities across a broad range of coastal aquatic environments, and by identifying key taxa that define specific environments. Using 16S rRNA gene amplicon pyrosequencing, I characterized bacterioplankton communities from 596 water samples collected across the Columbia River coastal margin over a four-year period from 2007-2010. The coastal waters of the Pacific Northwest are highly productive due to nutrients supplied by seasonal upwelling and freshwater inputs (Hickey and Banas 2003). The biological and physical processes of these waters are complex due to variable winds, remote wind forcing, shelf width, and submarine canyons (Hickey and Banas 2003, Hickey and Banas 2008, Hickey et al. 2010), which in turn may differentially affect the composition of bacterioplankton communities along the Oregon and Washington coasts (Fortunato and Crump 2011). The Columbia River is the second largest river in the United States with a mean annual discharge of $7300 \text{ m}^3\text{s}^{-1}$ (Hickey et al. 1998). This significant release of freshwater strongly impacts the chemical, physical, and biological characteristics of the river plume and coastal ocean (Hickey et al. 2010).

An earlier study of 300 water samples collected in 2007 and 2008 showed bacterioplankton communities of the Columbia River coastal margin separated into seven environments: river, estuary, plume, epipelagic, mesopelagic, shelf bottom, and slope bottom (Fortunato et al. 2012). Here I expand on this work and focus on population-level analyses to determine the key taxa in each environment and describe how the relative abundance of these taxa shifts with changing environmental conditions. I hypothesized that there exists seasonal shifts of dominant taxa in each

environment and reassembly of the same environment-specific communities each year. Results confirmed both these characteristics of bacterioplankton communities in freshwater influenced environments, and I identified indicator taxa that characterize each season. In contrast, the coastal ocean communities varied with environmental conditions that were not linked to seasonality but instead tended to show strong spatial variability.

Methods:

Water samples were collected from the Oregon and Washington coasts and the Columbia River and estuary (latitude 44.652 to 47.917, longitude -123.874 to -125.929) as part of the NSF-funded Center for Coastal Margin Observation and Prediction (CMOP). Samples were collected between 2007 and 2010 on 14 cruises aboard the *R/V Wecoma*, *R/V New Horizon*, and *R/V Barnes*. Aboard the *R/V Wecoma*, water samples were collected from the Columbia River, estuary, plume, and three coastal lines (Columbia River line, Newport Hydroline, La Push Line) in April, August, and November of 2007, April, June, July, and September of 2008, and May and July of 2010. In 2009, the same set of samples was collected aboard the *R/V New Horizon* in May and September (Figure 4.S1). Samples were collected using an instrumented 12 or 24-bottle rosette equipped with a Seabird 911+ conductivity-temperature-depth (CTD) sensor or a high volume low-pressure pump as described previously (Fortunato et al. 2012).

DNA samples (1-6 L per sample) were collected from 10 L Niskin bottles, preserved and extracted as described previously (Fortunato and Crump 2011), and

PCR-amplified for amplicon pyrosequencing using primers targeting the V2 region of the bacterial 16S ribosomal RNA gene (Hamady et al. 2008) as described previously (Fortunato et al. 2012) . Pyrosequencing was completed in two runs at Engencore at the University of South Carolina (<http://engencore.sc.edu/>). The first run of 2007-08 samples was performed on a Roche-454 FLX pyrosequencer and has been published (Fortunato et al. 2012). The second run of 2009-10 samples was completed using Titanium chemistry.

Water was also collected and analyzed for a suite of environmental variables as described previously (Fortunato and Crump 2011) or as described below. Particulate organic carbon (POC) and particulate nitrogen (PN) was collected by filtering between 250-1000 mL of water onto 25 mm diameter combusted (4.5 h at 500°C) GF/F filters (Whatman). Filters were folded, placed in plastic collection bags, and stored at -20°C. POC and PN content of the suspended particulate matter on acid-fumed filters (Hedges and Stern 1984) was determined at University of California, Davis using a Carlo Erba NA-1500 Elemental Analyzer system as described by Verardo et al (1990). For each day that samples were collected I determined several basic physical characteristics of the Columbia River Coastal margin. Columbia River flow ($\text{m}^3 \text{s}^{-1}$) at Bonneville Dam (upstream of furthest river sampling site) was measured by the U.S. Army Corps of Engineers (USACE, <http://www.nwd-wc.usace.army.mil>). Columbia River estuary residence time (in days) was calculated from river discharge records by USACE (<http://www.nwd-wc.usace.army.mil>). Upwelling index at latitudes 45° and 48° ($\text{m}^3/\text{s}/100\text{m}$) as

determined by Schwing et al. (1997) was reported by U.S. National Oceanographic and Atmospheric Administration (NOAA, <http://las.pfeg.noaa.gov/>).

Pyrosequences were quality controlled using the AmpliconNoise pipeline (Quince et al. 2011) with recommended procedures for FLX (Clean204.pl, PyroNoise, SeqDist, SeqNoise) and Titanium (CleanMinMax.pl, PyroNoiseM, SeqDistM, SeqNoiseM) chemistry. Maximum sequence length was set to 250 basepairs (Parse.pl), and chimera were identified and removed (PerseusD). Sequences were clustered into operational taxonomic units (OTU) using QIIME (Caporaso et al. 2010). Sequences from each sample were unweighted (unweight_fasta.py), concatenated, and primers were removed. OTUs were identified using uclust (pick_otus.py), and representative sequences were selected (pick_rep_set.py). The taxonomy of OTUs was determined in MOTHUR (v.1.21.0, Schloss et al. 2009) using the improved Greengenes 2011 taxonomic database (McDonald et al. 2012). Taxonomic assignments with less than 80% confidence were marked as unknown. A total of 608 samples were analyzed but was reduced to 596 because samples with low number of sequences were removed.

Relative abundances were calculated for OTUs in each sample and pairwise similarities among samples were calculated using the Bray-Curtis similarity coefficient (Legendre and Legendre 1998). Similarity matrices were visualized using multiple dimensional scaling (MDS) ordination. Analysis of similarity statistics (ANOSIM) was calculated to test the significance of differences among a priori sampling groups based on environmental parameters. These analyses were carried out using PRIMER v6 for Windows (PRIMER-E Ltd, Plymouth, UK).

As found in Fortunato et al (2012), bacterioplankton communities separated into seven environments (ANOSIM, $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom, and slope bottom. The plume group consisted of surface samples with salinity < 31 , the epipelagic group included surface and chlorophyll maximum samples (average depth = 9 m), the mesopelagic group consisted of samples within and below the thermocline (average depth = 56 m), the shelf bottom group of bottom samples with depth less than 350 m, and the slope bottom group of bottom samples deeper than 850 m. These seven environmental categories were used for further analyses. To better determine the variability within each environment, OTUs were assigned to a specific location based on the maximum average relative abundance of each OTU. For example, if OTU #1 was most abundant in the river, then it was a river OTU.

To identify the specific OTUs that characterize each of the environments, I used Indicator Species Analysis run in *R* (R-Development-Core-Team 2011), using the package *labdsv* (<http://ecology.msu.montana.edu/labdsv/R>) and test *indval* (Dufrene and Legendre 1997). Indicator values (IV) range from 0 to 1, with higher values for stronger indicators. Only OTUs with $IV > 0.3$ and $p < 0.05$ were considered good indicators (Dufrene and Legendre 1997). This approach was also used to identify key OTUs for seasonal groups within river, estuary, and plume communities and for spatial groups within the epipelagic community.

Environmental data were compiled from all years and tested for normality (Shapiro-Wilkes test, $p > 0.05$). Variables not normally distributed were transformed to as close to normality as possible. Variables included were: depth, salinity,

temperature, dissolved oxygen (DO), nitrate, nitrite, ammonium, ortho-phosphate, silicic acid (DSi), dissolved organic nitrogen, phosphorus and carbon (DON, DOP, DOC), particulate organic carbon (POC), and particulate nitrogen (PN), chlorophyll *a* (Chl *a*), bacterial production rate (BP), Columbia River flow, Columbia River estuary residence time, and coastal upwelling index at latitudes 45° and 48°N. Analyses were completed with a reduced set of 312 samples due to missing environmental data. The number of samples in each environment was: slope bottom = 16, shelf bottom = 72, mesopelagic = 23, epipelagic = 95, plume = 52, estuary = 46, river = 8. I also removed OTUs that were only present in one sample because environmental analyses may be skewed by rare taxa, which resulted in a reduced set of 5038 OTUs.

Our analysis used a two-step approach. First, in Primer v6 (PRIMER-E Ltd, Plymouth, UK), BV-STEP analysis was used to identify sets of variables that influenced community variability ($\rho > 0.95$, $\Delta\rho < 0.001$, 10 random starting variables, 100 restarts) (Clarke and Ainsworth 1993). BIO-ENV was used to rank each individual environmental variable by degree of association with community variability. BV-STEP and BIO-ENV use the Spearman rank coefficient (ρ) to determine the degree of association between similarity matrices of 16S amplicon sequences (Bray-Curtis similarity) and environmental data (Euclidean distances). Then, using the environmental variables identified by BV-STEP, I performed a Canonical Correspondence Analysis (CCA) to determine the percent of community variability explained by environmental variables (ter Braak 1986). When community data varied linearly along environmental gradients (instead of unimodal), I ran Redundancy analysis (RDA) instead of CCA (Legendre and Anderson 1999).

CCA/RDA analyses were run in *R*, using the *vegan* package (<http://vegan.r-forge.r-project.org/>), for each environment and for abundant taxonomic groups within each environment. I reduced the size of the environmental dataset by removing highly correlated variables ($\rho > 0.90$). Depth was included as a proxy variable for conditions that change in the vertical, including light and pressure.

Sequences are deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP006412 for 2007-08 and SRA058065 for 2009-10.

Results:

Sequence analysis of the full dataset ($n=596$) yielded 11,082 OTUs comprising 428,372 sequences. Bacterioplankton communities separated into the seven environments described previously (Fortunato et al. 2012) (ANOSIM $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom, and slope bottom along gradients of salinity and depth (Figure 4.S2).

Within each of the seven environments, bacterioplankton communities were variable across space and over time. Seasonal variability and annual reassembly of bacterioplankton communities in the Columbia River, estuary, and plume were evident in the MDS diagrams (Figures 4.1, 4.2, 4.3). For analysis of seasonal patterns, only OTUs specific for each environment were used to create MDS plots in order to specifically address the variability of the dominant members of the local communities. In the river, communities shifted along a seasonal continuum from April to November, and separated into two significant groups: early year (April-July) and late year (August-November) (ANOSIM, $p < 0.001$). Actinobacteria made up a

larger percentage (32%) of the late year community than the early year community (12%). A freshwater SAR11 group was present in the late year community (7%), but was absent from early year samples (Figure 4.1, Table 4.S1). Comparing river community variability with environmental conditions, I found that the concentration of POC and PN, varied with river community composition ($\rho=0.686$, Table 4.1). The correlation coefficient increased to $\rho=0.857$ with inclusion of ammonium, DOC, Chl *a*, and DO (Table 4.1). Early in the year, during spring and freshet periods, the river had higher flow and higher concentrations of POC, PN, DOC, and some inorganic nutrients.

In the estuary, there was a significant difference between early and late year communities (ANOSIM, $p<0.001$, Figure 4.2), and community composition correlated with seasonality in river flow and residence time ($\rho=0.550$, Table 4.1). Including temperature, DO, ortho-phosphate, and PN increased the coefficient to $\rho=0.638$. The early community had a larger percentage of Oceanospirillales, while the late community was dominated by Flavobacteria (Figure 4.2). Indicator analysis showed the top indicator for the early estuary was an Oceanospirillales OTU, which in some samples comprised up to 34% of the community (Figure 4.2, Table 4.S1). This same OTU was found in the estuary in a previous study (Crump et al. 1999). The top late year indicator was a Rhodobacteraceae OTU, which made up to 9.6% of the community.

The plume community also appeared to be influenced by river inputs, but did not correlate as strongly with environmental variables. The plume community varied seasonally and correlated strongly with temperature ($\rho=0.465$) and bacterial

production ($\rho=0.317$) (Table 4.1). These variables were elevated in the late summer months (August and September) compared to the rest of the year with average temperature increasing from 7.1 to 8.6°C and average bacterial production from 0.18 to 0.23 $\mu\text{gCL}^{-1}\text{h}^{-1}$. Changes in environmental conditions corresponded to the development of a significantly distinct summer bacterioplankton community (August-September) that assembled each year in the plume (ANOSIM, $p<0.001$, Figure 4.3). In this community, 31% of sequences were from *Polaribacter* and Rhodobacteraceae, which made up only 2% of non-summer communities. In non-summer months, 72% of the community was represented by unknown Flavobacteria. Indicator analysis showed that the top indicator for summer plume was a Rhodobacteraceae OTU (Figure 4.3, Table 4.S1).

In the slope bottom environment, in deep waters far off shore, DO ($\rho=0.606$) and temperature ($\rho=0.587$, Table 4.1) were most important to community variability. In the shallower shelf bottom environment, which includes samples from nearshore (<35 km to shore) and offshore (>35 km from shore) and a depth range of 18 m to 350 m, the community varied with differences in depth, temperature, DO, and rates of bacterial production ($\rho=0.618$, Table 4.1). In the epipelagic there was little variation in the depth of samples, but this group extended over large longitudinal and latitudinal scales. The MDS plot of the epipelagic community showed no significant difference across the latitudinal gradient but there was significant difference between nearshore and offshore communities (ANOSIM, $p<0.001$). The top indicators were a Flavobacteria OTU for the nearshore community, and a *Pelagibacter* OTU for the

offshore community (Table 4.S2). Community variability did not correlate with any measured environmental variables (Table 4.1).

Within each of the seven environments, I found distinct taxonomic assemblages. The upper panel of Figure 4.4 shows the complete taxonomy of the seven environments, which each displayed a unique taxonomic fingerprint consisting of different percentages of major taxonomic groups. The phylum SAR406 and the Gammaproteobacteria family SUP05 were most prevalent in the deep ocean environment, and mixed out in shallower water. The opposite can be said for the common surface bacteria groups SAR86 and Rhodobacteraceae, which were prevalent in the estuary, plume and epipelagic. The ubiquity of SAR11 was evident, as it made up a large percentage of the community in each of the seven environments.

The lower panel of Figure 4.4 depicts the taxonomy of OTUs specific to each environment. Focusing on environment-specific OTUs, it is evident that one or two taxonomic groups dominate each environment. In the mesopelagic, SAR11 sequences accounted for over 60% of the community, while Flavobacteria dominated the plume, specifically *Polaribacter* sequences which made up 15% of the community. Many of the top indicators corresponded to the dominant taxonomic group in each environment. In the slope bottom environment, top indicators included a SAR406 (IV: 0.94) and a gammaproteobacterium (IV: 0.91) OTU, both large percentages of the deep ocean community (Figure 4.4, Figure 4.5). In the mesopelagic, indicators included a SUP05 and a SAR11 OTU while in the epipelagic, surface ocean taxa SAR86 and *Prochlorococcus* were top indicators (Figure 4.5). In the plume, a *Polaribacter* OTU was the top indicator (IV: 0.42), and comprised up to 30% of the

sequences in some plume samples. Top estuary indicators included a Flavobacteria and Rhodobacteraceae OTU while in the river, common freshwater taxa from the class Betaproteobacteria were top indicators (Figure 4.5).

The environmental variables that correlated highly with variability of the dominant taxonomic groups in their respective environments were, with differing degrees, the same variables correlating with overall community variability in that environment (Table 4.2, Figure 4.6). In the slope bottom environment, the SAR406 community was correlated with changes in DO ($\rho=0.472$) while in the shelf bottom the SUP05 community was correlated with temperature and DO ($\rho=0.432$, Table 4.2, Figure 4.6). For the family Rhodobacteraceae and Flavobacteria in the estuary, river flow and residence time were the top factors correlated with these taxonomic groups ($\rho=0.660$ and 0.550 respectively, Table 4.2). In the river, variability of the freshwater SAR11 community was correlated with inorganic and organic nutrients, particularly POC and PN ($\rho=0.760$, Table 4.2).

Discussion:

Recent biogeographical studies comparing abundance of bacterial taxa with environmental conditions have either focused on the temporal variation of one common taxonomic group (Carlson et al. 2009, Malmstrom et al. 2010) or have studied multiple taxa but on a limited spatial scale (Eiler et al. 2012, Gilbert et al. 2012). In this study I assessed the variability of communities over four years, in seven environments spanning broad environmental gradients, and identified key taxa that characterize each of these environments. Results showed seasonal reassembly of

bacterioplankton communities in the river, estuary, and plume, with identification of taxa that characterize specific seasonal conditions in each environment, and correlation of community variability with seasonally influenced variables such as river flow and residence time. Seasonality was not detected in the coastal ocean, where communities were more variable on a spatial scale, and community composition was complicated by seasonal and annual fluctuations in physical factors like mixing and upwelling.

The physical mixing of communities was evident throughout the Columbia River coastal margin as taxonomic groups increased and decreased across salinity and depth gradients (Figure 4.4). In each of the seven environments, specific niches were created which lead to the prevalence of dominant taxonomic groups in each environment. This is evident in Figure 4.4, in which the upper panel of the overall community shows gradual shifts in these communities across space, but in contrast, the lower panel of environment-specific communities depicts sharp differences among environments and shows there are certain taxa that dominate each environment. The OTUs from these dominant taxonomic groups were also the top indicators for each environment, suggesting these dominant groups were key to shaping community composition, and indicating that the variability within each environment was driven by shifts in the most abundant taxa.

Seasonality of communities in the river, estuary, and plume environments was linked to the environment of the Columbia River. In the river community, seasonal fluxes of inorganic and organic nutrients, especially POC and PN, appeared to shape the changes in the bacterioplankton community over the annual cycle. Higher river

flow generally brings more particulate matter downstream, with high POC concentrations during the spring freshet, and low concentrations during summer months (Prahl et al. 1998). The relative abundances of OTU 11923, a member of Actinobacteria and OTU 2500, a freshwater SAR11, both negatively correlated with POC concentrations ($\rho=-0.664$, -0.616 respectively, Figures 4.5, 4.6), such that these taxa were more abundant in low flow summer months. The opposite was found for OTU 7389, a Burkholderiales OTU, where periods of higher POC concentrations resulted in higher abundance ($\rho=0.527$). The sequences of these three river OTUs match, with 100% identity, clones from freshwater environments around the globe (unpublished data from NCBI). OTU 7389 also matched a clone from low salinity bottom waters of the Chesapeake Bay estuary (Shaw et al. 2008), suggesting that this OTU favored a particle rich, high nutrient environment.

Community composition in the estuary also was influenced by upstream river conditions, specifically river flow and residence time. In a river-dominated system, like the Columbia River estuary, the amount and composition of river water and the time that water spends in the estuary greatly affects the development of the bacterioplankton community. In this study, the early estuary community (April-July) was more river-dominated because river flow was higher (average= $3.82 \text{ m}^3\text{s}^{-1}$) and residence time was shorter (average=0.68 d) compared to the late community (August-November), when flow was lower (average= $3.47 \text{ m}^3\text{s}^{-1}$) and residence time was longer (average=1.03 d). Indicator analysis showed that the top indicator for the early estuary was an Oceanospirillales OTU (OTU 627), which made up over one third of the sequences in some estuarine samples. A BLAST search of this OTU

sequence revealed it was identical to both free-living and particle-attached clones from a previous study of the Columbia River estuary (Crump et al. 1999). Crump et al. (1999) sampled the estuary in May when abundance for OTU 627 was highest, indicating this Oceanospirillales taxon is a well-established population in the early months of the year. BLAST results also showed this sequence was 100% identical only to clones from the Columbia River estuary. The indicator for the late year, and more coastal ocean influenced community, was a Rhodobacteraceae OTU (OTU 8288). This sequence was identical to clones from the Delaware Bay (Shaw et al. 2008) and was nearly identical (99%) to *Roseobacter* sp. clones from coastal waters of the Gulf of Mexico (Pinhassi et al. 2005), demonstrating that this taxon is prevalent in coastal ocean influenced environments.

In the plume, the community composition was less correlated with river-specific variables like river flow and residence time and more correlated with seasonal variables like temperature, leading to a distinct bacterioplankton community in the summer months. Of the top ten indicators for the plume environment, seven were Flavobacteria OTUs, with Flavobacteria sequences making up 46% and 72% of sequences in summer and non-summer months respectively. This class of bacteria was previously shown to be prevalent in productive environments like phytoplankton blooms (Simon et al. 1999) and upwelling zones (Alonso-Saez et al. 2007). The Columbia River plume is highly productive with seasonal upwelling supplying nutrients to the surface waters to fuel production (Hickey et al. 1998, Hickey et al. 2010). The top plume indicator was OTU 9443, a *Polaribacter* taxon, which is a genus of class Flavobacteria. One study of bacterioplankton communities along a

coast to open ocean transect in the North Atlantic Ocean found *Polaribacter* taxa to be the most prevalent of all Flavobacteria, with higher abundances in coastal samples (Gomez-Pereira et al. 2010). A comparison of all *Polaribacter* OTUs in the plume to environmental variables showed correlation with temperature, DO, BP, DOC, and upwelling ($\rho=0.576$, Table 4.2). Additionally, *Polaribacter* OTU 9443 closely matched sequences of clones from coastal upwelling systems, including the Brazilian (Cury et al. 2011) and Chilean coasts (Pommier et al. 2007), as well as along the Oregon coast during a diatom bloom (Morris et al. 2006), further demonstrating the prevalence of this taxon in highly productive coastal environments like the Columbia river plume.

Seasonality of environmental conditions clearly shaped communities across the river to ocean gradient, but in the coastal ocean, the variability within each environment varied strongly with spatial differences in environment and less by seasonal changes. In the slope bottom, where samples ranged in depth from 600 m to 2900 m, community composition correlated strongly with DO concentration ($\rho=0.607$, Table 4.1). DO ranged from 0.26 to 2.67 mg/L, with lower concentrations found in shallower DO minimum zone samples (~1000 m) and higher concentrations found in deeper waters (~2900 m). The top indicator for the slope bottom environment was a SAR406 OTU (OTU 4871), which identically matched sequences from oxygen minimum zones in the eastern North Pacific (Walsh et al. 2009) and the Hawaii Ocean Timeseries (HOT) Station ALOHA (Swan et al. 2011). In the shallower shelf bottom (18 m to 350 m), community composition varied with proximity to shore and location along the coast, and with temperature, DO, and

bacterial production rates. One of the top indicators for this environment was a very abundant OTU (up to 38% of sequences) taxonomically identified as *SUP05* (OTU 14135). However, a BLAST search showed this OTU closely matched sequences identified as the ARCTIC96BD-19 group, a closely related Gammaproteobacteria group to SUP05 (Walsh et al. 2009). Both of these groups are important to chemolithoautrophy in oxygen minimum zones (Walsh et al. 2009). Swan et al (2011), showed single amplified genomes of ARCTIC96BD-19 from HOT Station ALOHA have both Rubisco and sulfur oxidation genes, including an isolate that was 99% identical to OTU 14135. This similarity, coupled with the abundance of OTU 14135, suggests chemolithoautotrophic processes may be important in low oxygen bottom environments on the Oregon and Washington shelf.

In the epipelagic there was little correlation to any of the environmental variables that I measured and no seasonal patterns were evident. The heterogeneity of the epipelagic caused by mixing, upwelling, river flow, and other physical and biological factors made it difficult to define a set of environmental variables that best correlated with community composition. This contrasts with the river, estuary, and plume environments where seasonal patterns of community composition were more clearly defined and communities appeared to be driven by seasonally fluctuating environmental conditions. Throughout the year, the river, in general, is inoculated by the same upstream sources of bacteria and flows into the estuary, which then influences the plume, creating a consistent, predictable bacterioplankton community that changes seasonally and reassembles annually. In the coastal ocean, winds, upwelling, runoff from land, and mixing fluctuate both seasonally and annually.

Thus, the epipelagic community was inoculated by different sources depending on various physical conditions, which created a more complicated community where taxa increased and decreased frequently.

In conclusion, this study demonstrates that bacterioplankton communities are consistent and predictable across river to plume environments, with specific taxa defining each season. Community composition in the epipelagic was complicated by seasonal and annual fluctuations in physical factors like mixing and upwelling and thus correlation with environmental variables was low. In the bottom environments, DO strongly correlated with community composition, and prevalence of taxonomic groups like *SUP05* and *ARCTIC96BD-19* suggests chemolithoautotrophic processes are important in the carbon cycle of these low oxygen environments. The next step is to further develop a set of key taxa that can be used to model specific conditions in an environment and indicate change in bacterioplankton community composition.

Table 4.1: Spearman rank correlation coefficients (ρ) showing correlation between location-specific bacterioplankton communities and environmental variables. Top panel includes analyses with all OTUs and bottom panel with just OTUs specific to each location. BV-STEP ρ values represent maximum values when all environmental variables were included. For BIO-ENV, the maximum number of variables was set to one to determine the degree of association of each environmental variable individually, ρ values for the top two variables are shown. Environmental variables from BV-STEP results were used in Canonical Correspondence Analysis (CCA)/Redundancy Analysis (RDA) to determine the percent of community variability explained.

Environments with all OTUs					
<i>Environment</i>	<i>BV-STEP factors</i>	ρ	<i>BIO-ENV factors</i>	ρ	<i>Variability explained</i>
Slope Bottom	Temperature, DO, BP	0.707	DO	0.606	24.2%
			Temperature	0.587	
Shelf Bottom	Depth, Temperature, DO, BP	0.618	Temperature	0.482	12.5%
			DO	0.440	
Mesopelagic	Depth, Salinity, Temperature	0.599	Temperature	0.513	19.9%
			Depth	0.429	
Epipelagic	Salinity, Temperature, DO, Chl a	0.376	Temperature	0.248	10.2%
			Salinity	0.240	
Plume	Salinity, Temperature, DSi, DOC, BP	0.517	Temperature	0.465	18.7%
			BP	0.317	
Estuary	Salinity, Temperature, DO, Ortho-phosphate, PN, River flow, Residence time	0.638	River flow	0.550	22.9%
			Residence time	0.550	
River	DO, Chl a, Ammonium, DOC, POC, PN	0.857	POC	0.686	80.9%
			PN	0.686	

Environments with only location-specific OTUs					
<i>Environment</i>	<i>BV-STEP factors</i>	ρ	<i>BIO-ENV factors</i>	ρ	<i>Variability explained</i>
Slope Bottom	DO, BP	0.649	DO	0.546	17.0%
			Temperature	0.473	
Shelf Bottom	Temperature, DO, BP	0.611	DO	0.484	7.6%
			Temperature	0.434	
Mesopelagic	Depth, Salinity, BP	0.484	Depth	0.398	16.5%
			Salinity	0.251	
Epipelagic	Salinity, Temperature, Chl a	0.283	Temperature	0.222	8.7%
			Chl a	0.185	
Plume	Salinity, Temperature, DSi, DOC, BP	0.517	Temperature	0.394	20.9%
			BP	0.364	
Estuary	Salinity, Nitrate, POC, BP, River flow, Residence time	0.736	River flow	0.642	19.3%
			Residence time	0.642	
River	DO, Chl a, Ammonium, DOC, POC, PN	0.855	POC	0.689	81.5%
			PN	0.689	

Table 4.2: Spearman rank correlation coefficients (ρ) showing correlation between abundant taxonomic groups within each environment and environmental variables. Variables from BV-STEP results were used in CCA/RDA analysis to determine the percent of community variability explained.

Slope Bottom	BV-STEP factors	ρ	BIO-ENV factors	ρ	Variability explained
SAR406	DO, BP	0.641	DO	0.472	18.4%
			Temperature	0.371	
Deltaproteobacteria	DO, BP	0.466	BP	0.356	23.9%
			Temperature	0.333	
SAR11	Temperature, DO, Chl a, Nitrite, BP, Upwelling	0.366	DO	0.189	41.7%
			Upwelling	0.183	
Shelf Bottom					
Gammaproteobacteria	Temperature, DO, BP	0.533	DO	0.440	44.6%
			Temperature	0.413	
SUP05	Temperature, DO	0.432	Temperature	0.392	40.1%
			DO	0.376	
Mesopelagic					
SAR11	Depth, Salinity, Chl a, DSi, River flow, Residence time	0.441	Depth	0.287	52.4%
			Salinity	0.259	
Epipelagic					
<i>Prochlorococcus</i>	Salinity, DO	0.324	Salinity	0.289	2.5%
			DO	0.239	
SAR11	Salinity, DO	0.397	Salinity	0.334	3.6%
			DO	0.313	
Plume					
Rhodobacteraceae	Depth, Temperature, DO, Nitrite, Org P, BP, River flow, Residence Time	0.318	BP	0.242	49.4%
			Temperature	0.215	
<i>Polaribacter</i>	Temperature, DO, DOC, BP, Upwelling	0.505	BP	0.457	24.0%
			Temperature	0.368	
Estuary					
Rhodobacteraceae	Ammonium, River flow, Residence time	0.610	River flow	0.550	15.2%
			Residence time	0.550	
Flavobacteria	Salinity, NO ₃ , BP, River flow, Residence time	0.684	River flow	0.660	18.2%
			Residence time	0.660	
River					
Actinobacteria	DO, Chl a Ammonium, DOC, POC, PN	0.811	POC	0.698	84.8%
			PN	0.698	
Betaproteobacteria	Depth, DO, POC, PN, BP	0.580	POC	0.571	81.4%
			PN	0.571	
SAR11	Nitrate, Org N, POC, PN, River flow, Residence time	0.802	POC	0.760	96.8%
			PN	0.760	

Figure 4.1: Seasonal multidimensional scaling diagram of river samples, produced with Bray-curtis similarity values for communities that only included river-specific OTUs (Stress = 0.11). Samples separated into two significantly different communities, early (April to July) and late (August to November) (ANOSIM, $p < 0.001$). Taxonomy of the two groups is depicted in the upper right panel. Two bottom panels depict the spatial distribution of a top indicator for Early (IV: 0.74) and Late (IV: 0.94), by average relative abundance, across the ordination of the river MDS (IV = Indicator value).

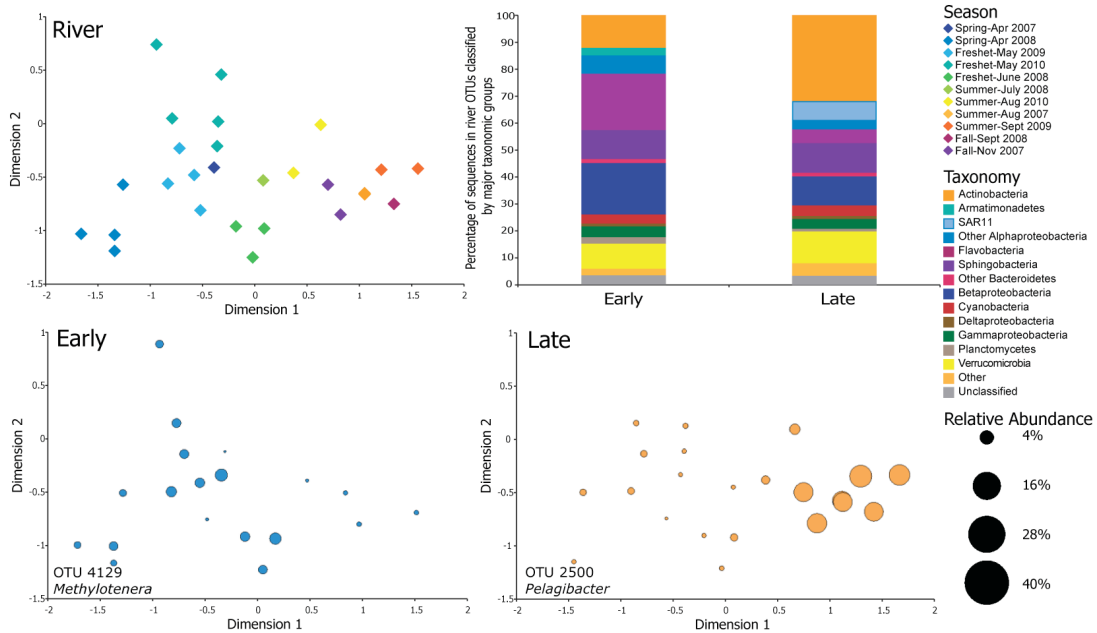


Figure 4.2: Seasonal multidimensional scaling diagram of estuary samples, produced with Bray-curtis similarity values for communities that only included estuary-specific OTUs (Stress = 0.19). Samples separated into two significantly different communities, early (April to July) and late (August to November) (ANOSIM, $p < 0.001$). Taxonomy of the two groups is depicted in the upper right panel. Two bottom panels depict the spatial distribution of a top indicator for Early (IV: 0.96) and Late (IV: 0.87), by average relative abundance, across the ordination of the estuary MDS.

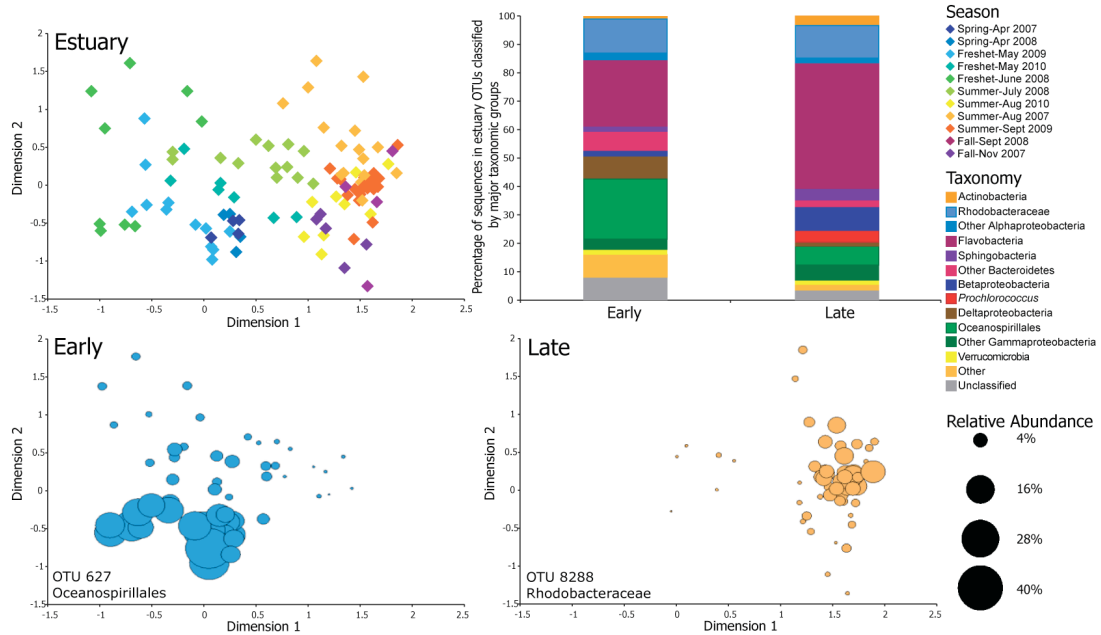


Figure 4.3: Seasonal multidimensional scaling diagram of plume samples, produced with Bray-Curtis similarity values for communities that only included plume-specific OTUs (Stress = 0.17). Samples separated into two significantly different communities, non-summer (April-July, Nov.) and summer (August to September) (ANOSIM, $p < 0.001$). Taxonomy of the two groups is depicted in the upper right panel. Two bottom panels depict the spatial distribution of the top indicator for non-summer (IV: 0.76) and summer (IV: 0.67), by average relative abundance, across the ordination of the plume MDS.

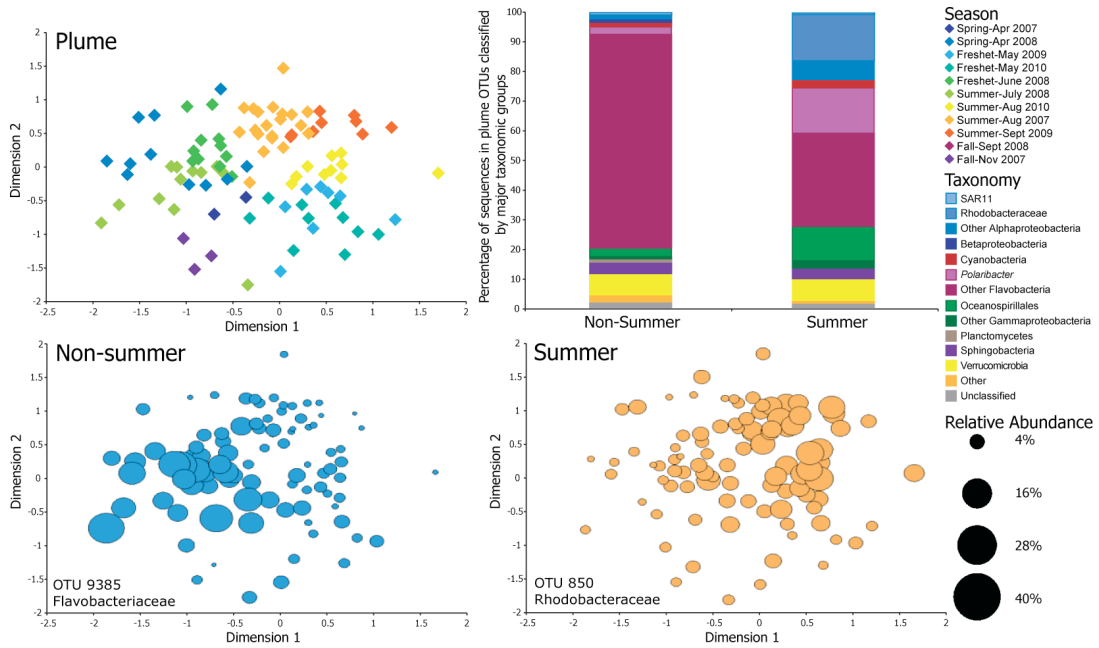


Figure 4.4: Taxonomic composition of each environment. Top panel depicts taxonomy for all OTUs in each environment. Bottom panel depicts taxonomy only for OTUs that were assigned to each environment based on the maximum average relative abundance.

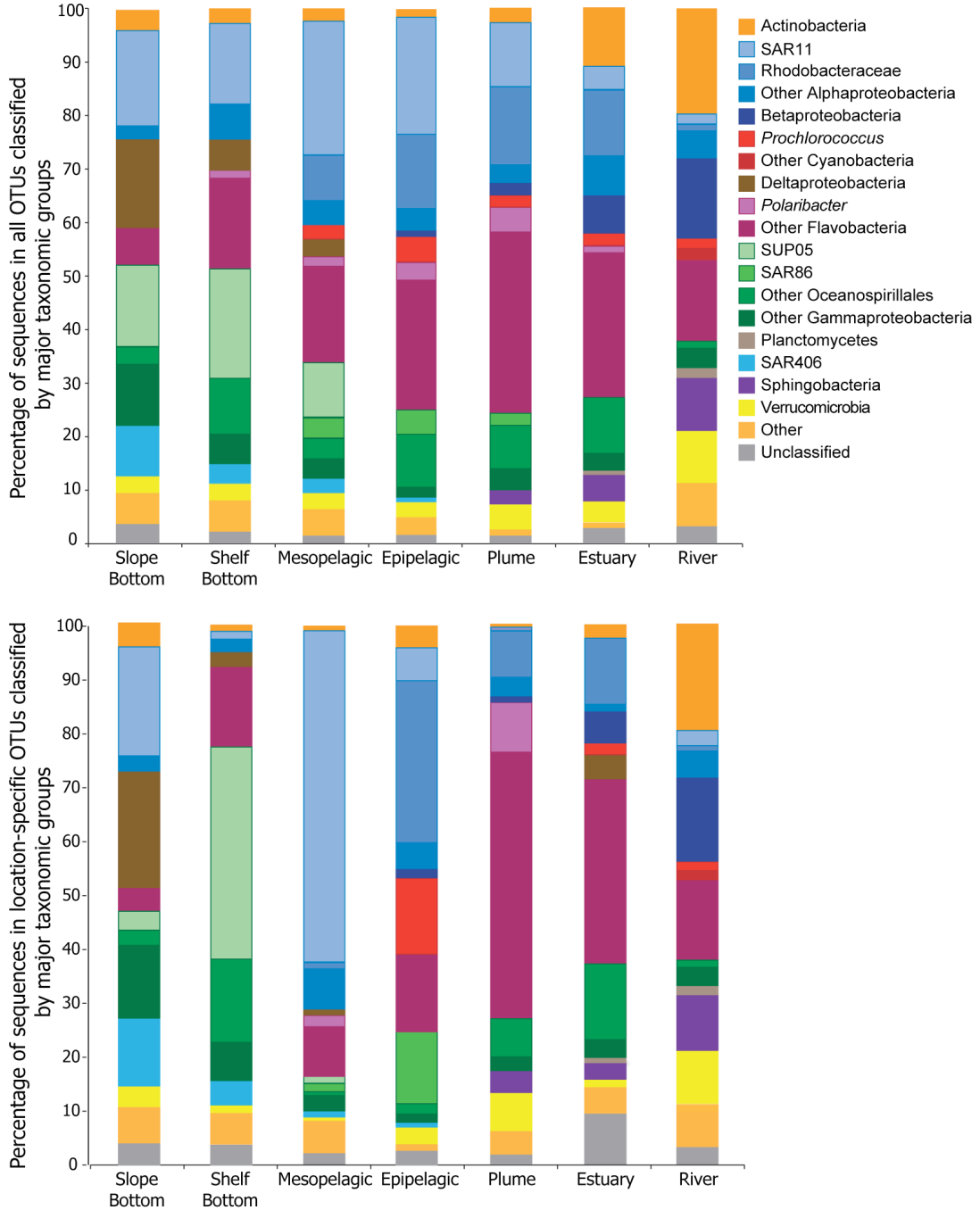


Figure 4.5: Bubble plot of top indicator OTUs in each environment. The size of the bubble indicates the average relative abundance (%) of each OTU in each of the seven environments. Black shaded bubbles show the environment for which each OTU is an indicator. Indicator values are displayed next to each OTU.

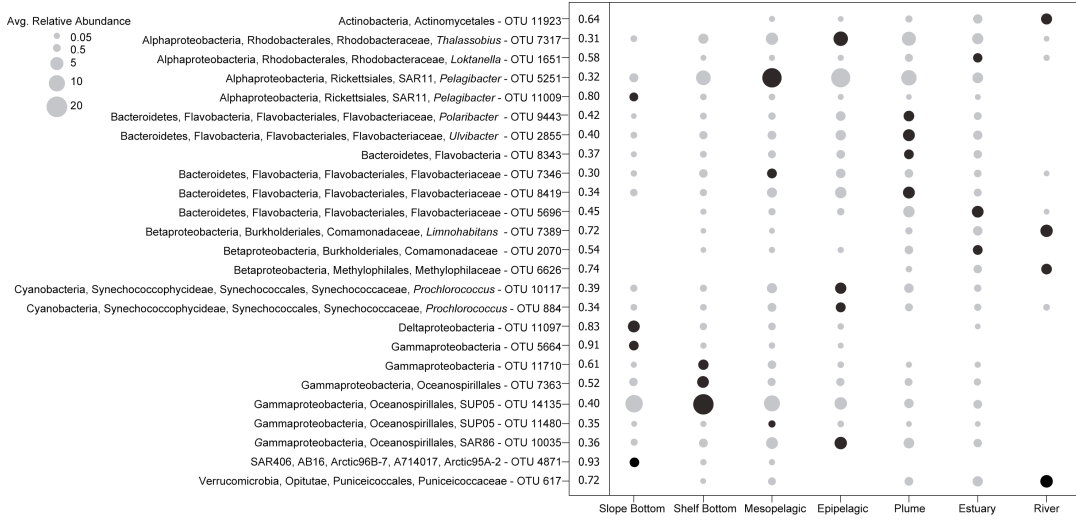
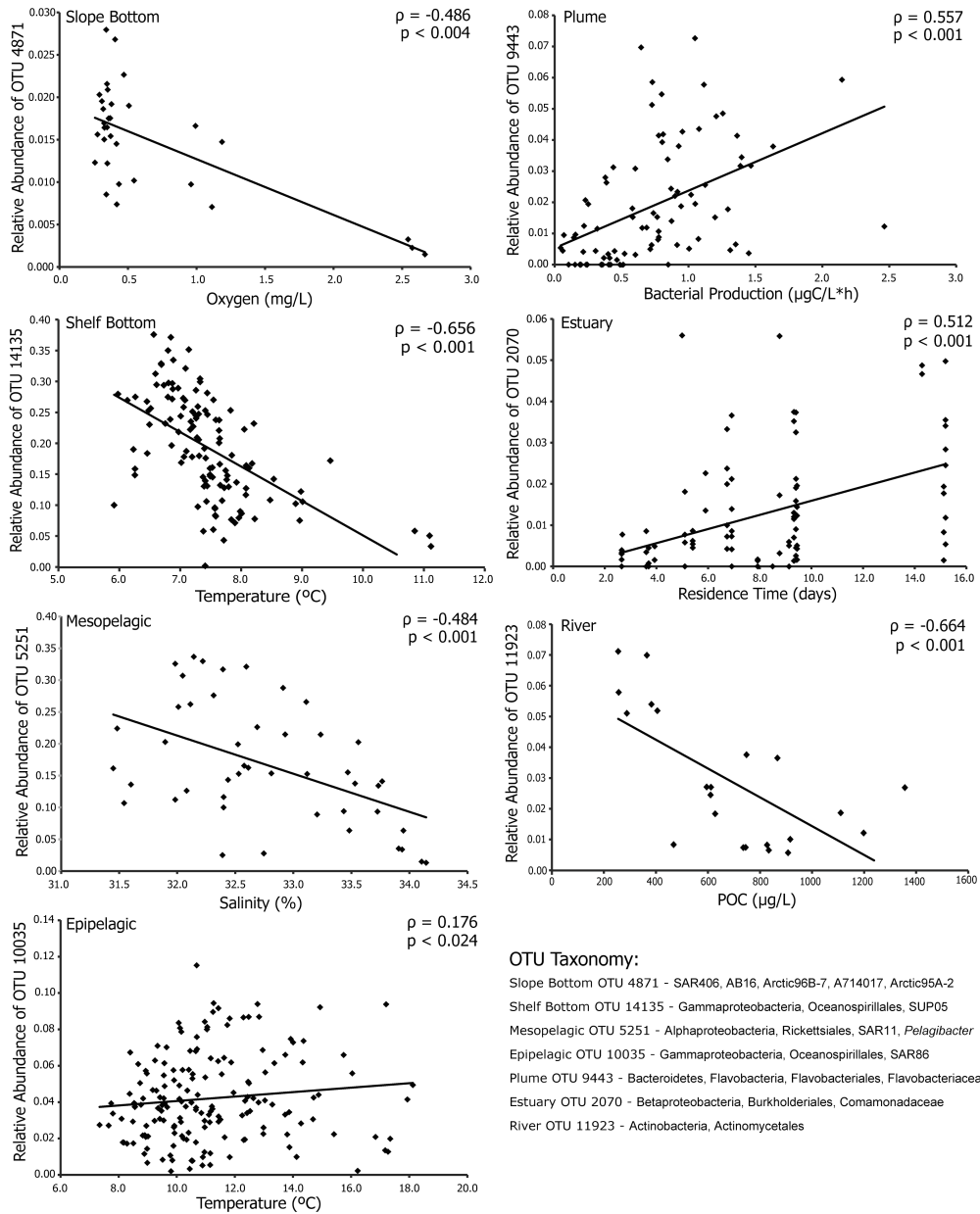


Figure 4.6: Correlations of the relative abundance of top indicator OTUs and environmental variables for each of the seven environments. Spearman correlation coefficient (ρ) and p-values are indicated for each relationship.



OTU Taxonomy:

- Slope Bottom OTU 4871 - SAR406, AB16, Arctic96B-7, A714017, Arctic95A-2
- Shelf Bottom OTU 14135 - Gammaproteobacteria, Oceanospirillales, SUP05
- Mesopelagic OTU 5251 - Alphaproteobacteria, Rickettsiales, SAR11, *Pelagibacter*
- Epipelagic OTU 10035 - Gammaproteobacteria, Oceanospirillales, SAR86
- Plume OTU 9443 - Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, *Polaribacter*
- Estuary OTU 2070 - Betaproteobacteria, Burkholderiales, Comamonadaceae
- River OTU 11923 - Actinobacteria, Actinomycetales

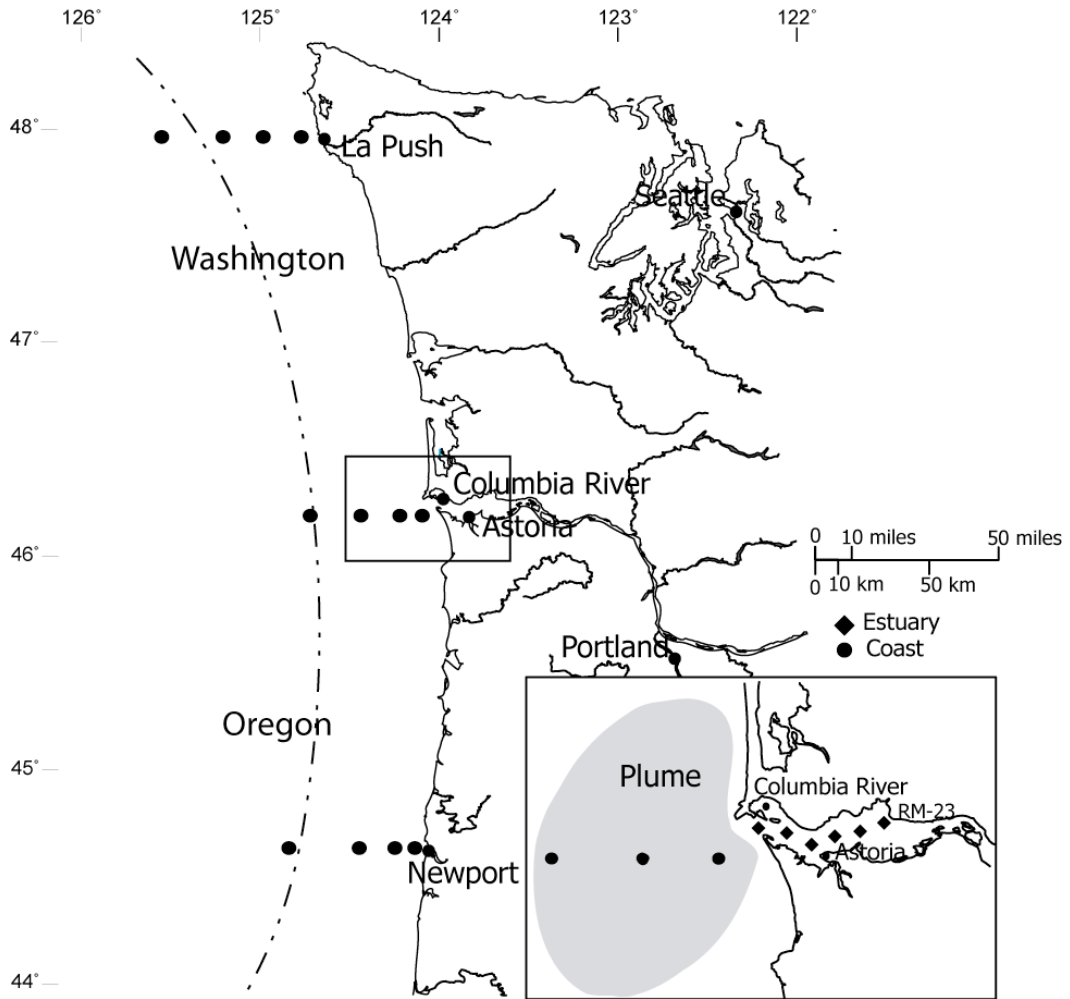
Supplemental Table 4.1: Top ten indicator taxa for each of the seasonal groups of the river, estuary, and plume. Taxonomy is listed to the furthest classification.

Environment, Season	OTU	Indicator Value	Taxonomy: Phylum, Class, Order, Family, Genus, Species
River, Early	10165	0.80	Armatimonadetes, Armatimonadia, Armatimonadales, Armatimonadaceae
	6790	0.77	Verrucomicrobia, Spartobacteria, Spartobacteriales, Spartobacteriaceae, <i>Xiphinematobacter</i>
	11327	0.76	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
	4129	0.74	Proteobacteria, Betaproteobacteria, Methylophilales, Methylophilaceae, <i>Methylotenera</i>
	13463	0.73	Proteobacteria, Alphaproteobacteria, Sphingomonadales, Sphingomonadaceae
	3721	0.72	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	13717	0.72	Cyanobacteria, Synechococcophycideae, Pseudanabaenales, Pseudanabaenaceae, <i>Pseudanabaena</i>
	11240	0.72	Bacteroidetes, Sphingobacteria, Sphingobacteriales, Cyclobacteriaceae, <i>Algoriphagus</i>
	4144	0.69	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
	8328	0.67	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
River, Late	3480	0.99	Actinobacteria, Actinobacteria, Actinomycetales, ACK-M1
	11011	0.94	Actinobacteria, Actinobacteria, Acidimicrobiales
	2500	0.94	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>
	8752	0.88	Cyanobacteria, Synechococcophycideae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
	10530	0.84	Bacteroidetes, Flavobacteria, Cryomorphaeae, Fluvicola
	9600	0.82	Cyanobacteria, Synechococcophycideae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
	13388	0.80	Proteobacteria, Gammaproteobacteria, Chromatiales, Sinobacteraceae
	11923	0.80	Actinobacteria, Actinobacteria, Actinomycetales, ACK-M1
	5354	0.78	Proteobacteria, Alphaproteobacteria, Rhodospirillales, Acetobacteraceae, Roseomonas
	10937	0.78	Cyanobacteria, Synechococcophycideae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
Estuary, Early	627	0.96	Proteobacteria, Gammaproteobacteria, Oceanospirillales
	10165	0.90	Armatimonadetes, Armatimonadia, Armatimonadales, Armatimonadaceae
	11327	0.81	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
	12554	0.79	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	7389	0.77	Proteobacteria, Betaproteobacteria, Burkholderiales, Comamonadaceae, <i>Limnohabitus</i>
	12427	0.70	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
	8328	0.69	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Flavobacterium</i>
	617	0.68	Verrucomicrobia, Opiritae, Puniceicoceales, Puniceococcaceae
	7356	0.68	Actinobacteria, Actinobacteria, Actinomycetales, ACK-M1
	7717	0.66	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
Estuary, Late	850	0.88	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	8288	0.87	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	3480	0.83	Actinobacteria, Actinobacteria, Actinomycetales, ACK-M1
	9914	0.81	Bacteroidetes, Flavobacteria, Cryomorphaeae
	2500	0.78	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>
	10937	0.77	Cyanobacteria, Synechococcophycideae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
	11487	0.76	Proteobacteria, Gammaproteobacteria, Oceanospirillales, HTCC2188
	7317	0.75	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae, <i>Thalassobius</i>
	6664	0.73	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	11742	0.73	Proteobacteria, Gammaproteobacteria, Oceanospirillales, Alteromonadaceae, HTCC2207
Plume, Non-summer	9385	0.76	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	2855	0.71	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Ulvibacter</i>
	41	0.63	Bacteroidetes, Flavobacteria, Cryomorphaeae, Fluvicola
	5251	0.61	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter ubique</i>
	2439	0.60	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter ubique</i>
	2763	0.55	Verrucomicrobia, Verrucomicrobiae, Verrucomicrobiaceae
	7363	0.53	Proteobacteria, Gammaproteobacteria, Oceanospirillales
	14135	0.51	Proteobacteria, Oceanospirillales, SUP05
	10461	0.50	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	781	0.49	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>
Plume, Summer	850	0.67	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	13028	0.67	Cyanobacteria
	11667	0.64	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	8648	0.63	Proteobacteria, Alphaproteobacteria, Rhodospirillales, Rhodospirillaceae
	11487	0.62	Proteobacteria, Gammaproteobacteria, Oceanospirillales, HTCC2188
	1623	0.62	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>
	3050	0.61	Verrucomicrobia, Opiritae, Puniceicoceales, Puniceococcaceae
	9443	0.59	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>
	2745	0.58	Proteobacteria, Gammaproteobacteria, Oceanospirillales, Alteromonadaceae, HTCC2207
	9065	0.57	Verrucomicrobia, Opiritae, Puniceicoceales, Puniceococcaceae

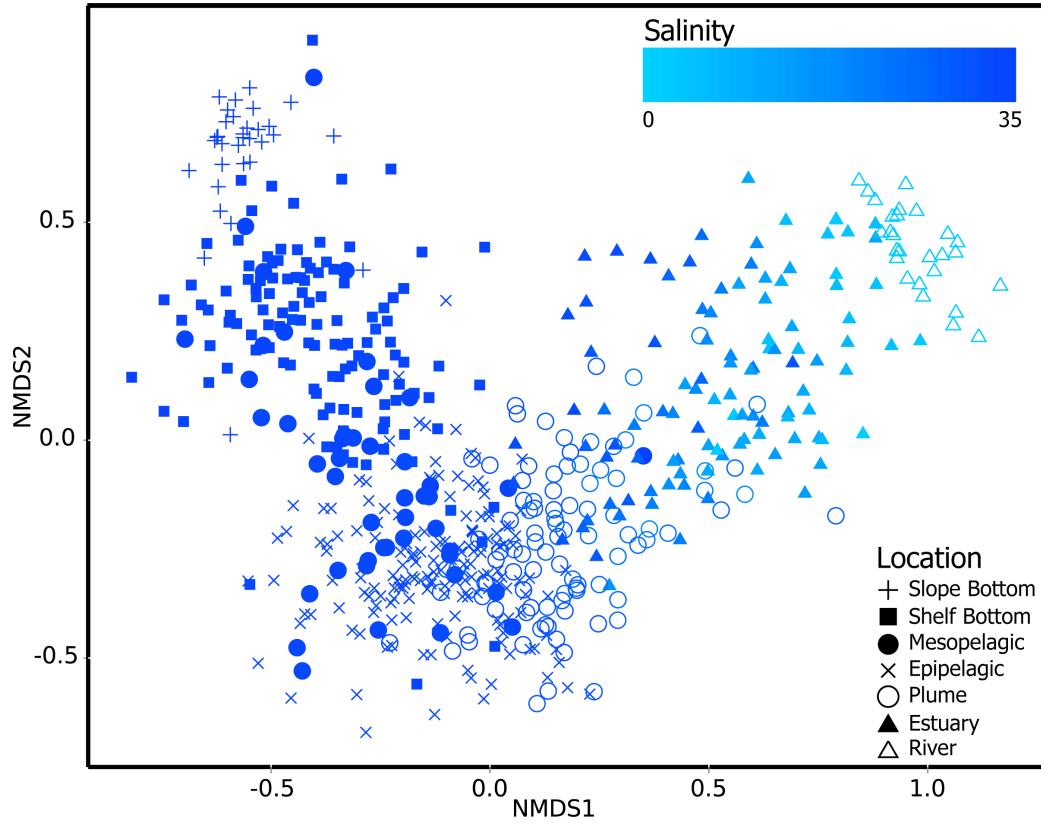
Supplemental Table 4.2: Top ten indicator taxa for nearshore and offshore communities in the epipelagic. Taxonomy is listed to the furthest classification.

Season	OTU	Indicator Value	Taxonomy: Phylum, Class, Order, Family, Genus, Species
<i>Nearshore</i>	13095	0.74	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	1263	0.64	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	2855	0.62	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Ulvibacter</i>
	7317	0.60	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae, <i>Thalassobius</i>
	9443	0.60	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>
	14654	0.59	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	850	0.58	Proteobacteria, Alphaproteobacteria, Rhodobacterales, Rhodobacteraceae
	13530	0.57	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	10719	0.57	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	10571	0.54	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae, <i>Polaribacter</i>
<i>Offshore</i>	781	0.74	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>
	14347	0.73	Proteobacteria, Alphaproteobacteria
	884	0.72	Cyanobacteria, Synechococcophycidae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
	2439	0.70	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter ubique</i>
	10117	0.66	Cyanobacteria, Synechococcophycidae, Synechococcales, Synechococcaceae, <i>Prochlorococcus</i>
	11501	0.65	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter</i>
	5251	0.62	Proteobacteria, Alphaproteobacteria, Rickettsiales, SAR11, <i>Pelagibacter ubique</i>
	5300	0.59	Bacteroidetes, Flavobacteria, Flavobacteriales, Flavobacteriaceae
	5804	0.57	Actinobacteria, Actinobacteria, Koll13
	9632	0.56	Proteobacteria, Alphaproteobacteria

Supplemental Figure 4.1: Map of Oregon and Washington coasts. Inset depicts Columbia River estuary and plume region. Dotted line denotes approximate location of shelf break



Supplemental Figure 4.2: Multidimensional scaling diagram of all 596 samples. Samples separated into seven distinct groups (ANOSIM, $p < 0.001$): river, estuary, plume, epipelagic, mesopelagic, shelf bottom, and slope bottom. Samples are colored by salinity.



CHAPTER 5:
GENE ABUNDANCE AND EXPRESSION PATTERNS ACROSS AN
ESTUARINE SALINITY GRADIENT

Abstract:

Microbial communities mediate the biogeochemical cycles that drive ecosystems, and it is important to understand how these communities are affected across changing environmental conditions, especially in complex coastal zones. From river to ocean, environmental conditions shift with salinity, and microbial communities must adjust to these changing biogeochemical conditions. I collected five metagenomic and metatranscriptomic samples across the salinity gradient of the Columbia River coastal margin, from the Columbia River, estuary, plume, and coastal ocean in August 2010. Results from metagenomic 16S rRNA gene sequences showed there were taxonomically distinct communities from river to ocean, and that these communities mixed in the estuary. Mapping metagenomic sequences to genomes showed that freshwater Actinobacteria genomes were highly abundant in metagenomic samples, while *Pelagibacter* genomes were dominant in higher salinity samples. Despite the strong salinity gradient (0 to 33), the metabolic potential of microbial communities was very similar from river to ocean, with metagenomic samples showing an average of 78% similarity. In contrast, the metatranscriptomes were only 23% similar, and showed large differences in functional genes among samples. In the metagenomes, there was a relative increase in sulfur metabolism and stress genes and a decrease in phosphorus metabolism genes with increasing salinity. In the plume there was a higher proportion of photosynthesis genes compared to other samples. Gene expression patterns were highly variable in the metatranscriptomes with no apparent relationship with salinity. Across the salinity gradient, the microbial community followed three different patterns of diversity. The phylogeny of the

community varied strongly with salinity, while metabolic potential was similar, with only small differences in gene abundance from river to ocean, and gene expression was highly variable and generally was independent of changes in salinity.

Introduction:

The key to understanding the ecological role of microbial communities in an environment is determining the link between community structure and community function. Metagenomic and metatranscriptomic data provide powerful insight into the metabolic potential and gene expression in an environment without previous knowledge of taxonomic or functional characteristics (Poretsky et al. 2005). Thus, many recent studies have used these approaches to answer ecological questions about the biogeochemical function of microbes in soil, in the coastal and open ocean, and in the deep sea (Eloe et al. 2011, Fierer et al. 2012, Gifford et al. 2011, Hewson et al. 2010, Poretsky et al. 2005). Together, these studies further solidify the importance of microbes to ecosystem function.

Initially, the goal of metagenomic and metatranscriptomic studies was to describe the dominant microbially mediated biogeochemical processes occurring in an environment (Frias-Lopez et al. 2008, Poretsky et al. 2005). With the advancement of sequencing technology and the increase in size of metagenomic and metatranscriptomic libraries, there has been a shift from descriptive to more hypothesis-based research. A comparative study of metatranscriptomes taken from the surface ocean at the Hawaiian Ocean Time Series (HOT) Station Aloha looked found an overrepresentation of photosynthesis and C1 metabolism during the day and

an abundance of housekeeping related processes such as membrane repair and amino acid biosynthesis at night (Poretsky et al. 2009b). Hewson et al. (2010) found that differences between metatranscriptomes were driven by differential expression of genes involved in nutrient acquisition, cell metabolism, and growth, which could be a consequence of variable nutrient and Cyanobacteria concentrations across ocean basins. A difference in gene expression was also seen in DOM-amended bottle experiments where the addition of high-molecular weight DOM resulted in changes in gene expression as specific taxa took advantage of the added carbon source (McCarren et al. 2010). Environmental conditions then play a large role in determining the taxonomic composition of communities and consequently gene expression.

More recently, genomic studies have started to examine gene abundance and expression across ecologically relevant gradients of various environments. Soil metagenomes sampled across a nitrogen gradient revealed increases in genes associated with replication, electron transport, and protein metabolism and a shift from oligotrophic to copiotrophic taxa with increasing nitrogen concentrations (Fierer et al. 2012). Stewart et al. (2012) compared both metagenomic and metatranscriptomic data across an oxygen gradient in the oxygen minimum zone off the Chilean coast. Again, changes in gene abundance and expression as well as a shift in taxonomic composition occurred across the oxygen gradient as nitrogen processes changed from oxidative to reductive and the community shifted from nitrifying taxa to taxa associated with anammox and denitrification (Stewart et al. 2012). Looking at the same oxygen gradient, metagenomic data also revealed the first evidence of an

active sulfur cycle in an oxygen minimum zone and a coupling between sulfur and nitrogen processes (Canfield et al. 2010). These metagenomic studies across both nutrient and oxygen gradients have provided a genetic basis for important ecological process and microbial community dynamics in these biogeochemically important environments.

Although gradients have recently become a focus of metagenomic and metatranscriptomic studies, to date there have not been any studies focusing on gene abundance and expression across a salinity gradient. In fact, there have few studies looking at metagenomic data from river systems (e.g., Ghai et al. 2011). In this study, I looked at changes in metabolic potential and gene expression across the salinity gradient of the Columbia River coastal margin. The Columbia River is the second largest river in the United States with a mean annual discharge of $7300 \text{ m}^3\text{s}^{-1}$ (Hickey et al. 1998). This release of freshwater has a large impact on the chemical, physical, and biological characteristics of the Oregon and Washington coasts (Hickey et al. 2010). Using microarray analysis to characterize gene expression from the Columbia River to the coastal ocean, Smith et al. (2010) showed strong seasonal shifts in gene expression but variable spatial differences as some seasons showed little difference in gene expression across the salinity gradient. In addition, my work on 16S rRNA genes identified spatially and taxonomically distinct bacterioplankton communities in the river, estuary, plume, and surface ocean (Fortunato et al. 2012).

Building on patterns of community structure and previous microarray data, the addition of gene abundance and expression data has the potential to create a link between community structure and function and provides information to explore how

and why specific populations are distributed across the salinity gradient. I hypothesized that gene abundance and availability across a river to ocean gradient will shift dramatically, but that taxonomically distinct populations would perform similar metabolic activities regardless of the salinity of the water. Thus I anticipated that the typical phylogenetic shift from fresh, to estuarine, to marine bacteria over the salinity gradient would be apparent in metagenomic data but would not be reflected in metatranscriptomic gene expression patterns. Results, however, show that in fact the metagenomes from river to ocean are very similar with little change in metabolic potential across the salinity gradient. Patterns of gene expression, however, are extremely variable, with each metatranscriptome dominated by a handful of different functional genes.

Methods:

Metagenomic and metatranscriptomic samples were collected from the Columbia River, estuary, and plume (latitude 46.184 and 46.239, longitude -124.161 and -123.182) as part of the NSF-funded Center for Coastal Margin Observation and Prediction (CMOP) science and technology center. Samples were collected aboard the *R/V Wecoma* between August 1 and 8, 2010. Water was collected using a high volume low-pressure pump with an attached Seabird 911+ conductivity-temperature-depth (CTD) sensor. With each CTD cast, depth profiles of salinity, temperature ($^{\circ}\text{C}$), turbidity (NTU), oxygen (mgL^{-1}), and chlorophyll fluorescence were recorded. Surface samples (0.5-1 m) were collected in the river, estuary, and plume across a salinity gradient of 0-25. A sample was also taken below plume waters (16 m) at a

salinity of 33 (Table 5.1, Figure 5.1). Water was collected in 25 L carboys and immediately filtered. Filtering time was stopped at 30 min to ensure preservation of mRNA. Samples were pre-filtered through a 3.0 μm polycarbonate filter and then collected onto a 0.2 μm polycarbonate filter, preserved in RNAlater and stored at -80°C until extraction. For extraction, the 0.2 μm filters were thawed and rinsed with sterile PBS or water, depending on salinity of the sample, to remove RNAlater. Cells that became dislodged after rinsing were captured on a 0.2 μm Sterivex-GP (Millipore) filter and extracted with the polycarbonate filter. One quarter of the polycarbonate filter was used to extract DNA for a metagenome and the remaining was used for RNA extraction. Additional sample water was filtered onto Sterivex-GP (Millipore) filters for DNA to be used in 16S rRNA gene amplicon pyrosequencing. These filters were extracted, amplified using 16S rRNA gene specific primers, pyrosequenced, and analyzed as described in Fortunato et al. (2012) and Fortunato et al. (in prep).

RNA was extracted using a modified protocol from the RNeasy kit (Qiagen) as described in Poretsky et al. (2009a). Briefly, the thawed, rinsed, 0.2 μm filter was cut into pieces using a sterile blade and placed in a 50 mL tube with RLT buffer (Qiagen) and RNA beads (MoBio). The tube was vortexed for 10 min and centrifuged twice at 5000 rpm. The lysate was placed in a new tube with one-volume 100% ethanol and pushed through a 20-gauge needle several times to shear the RNA. The extraction then followed the RNeasy kit (Qiagen) according to manufacturer instructions. DNA was removed from the extraction using a Turbo-DNase kit (Ambion). A final volume of 100 μl was extracted. Ribosomal RNA was removed

from RNA samples using subtractive hybridization as described in Stewart et al. (2010). Sample specific rRNA probes were constructed from DNA collected on 0.2 µm Sterivex filters. After rRNA removal, RNA final concentration was determined on an Agilent 2100 Bioanalyzer. DNA was extracted using a phenol-chloroform extraction method adapted from Zhou et al. (1996) and Crump et al (2003). DNA and RNA aliquots (~1 µg) were then sent to Integrated Genomic Services (IGS) at the University of Maryland for sequencing on an Illumina Hi-Seq 1000 system. Resulting libraries were paired-end, non-overlapping. The average single read length was 100 basepairs (bp).

Raw metatranscriptomic reads were aligned to SILVA SSU and LSU databases (Pruesse et al. 2007) using Bowtie2 (v 2.0.0-beta5, Langmead and Salzberg 2012) with a local alignment and default settings to remove additional rRNA sequences. Ribosomal RNA from the metagenomes was also identified but not removed using the above method. Once metagenomic rRNA sequences were identified, I then identified potential 16S rRNA sequences from the overall rRNA pool by aligning to the Greengenes 2011 taxonomic database (McDonald et al. 2012) and to a manually curated freshwater taxonomic database (Newton et al. 2011 and K.D. McMahon pers. comm.) using Bowtie2. Potential 16S rRNA sequences were then taxonomically identified using MOTHUR (v 1.21, Schloss et al. 2009) using the Greengenes and freshwater taxonomic databases. Taxonomic assignments with less than 80% confidence were marked as unknown.

Metagenomes and metatranscriptomes were assembled prior to annotation due to the large number of reads in each library. Assemblies were carried out using CLC

Genomics Workbench (v 4.9) with a minimum length fraction of 80% and a similarity of at least 95%. The minimum contig length was set to 250 bp. Assembled contigs for each library were uploaded to the MG-RAST web-server (Meyer et al. 2008). Open Reading Frame (ORF) annotation was done using SEED subsystem categories with a BLAST threshold of $1e-5$, a minimum of 60% identity, and a minimum alignment length of 50 bp. Sequences were then mapped to all ORFs with Bowtie2 (Langmead and Salzberg 2012) using an end-to-end alignment and default settings to determine the number of sequences per ORF.

In order to compare between samples and account for differences in sequencing depth, libraries were randomly downsampled to match the number of subsystem-annotated sequences in the smallest sample with MOTHUR (Schloss et al. 2009). For each metatranscriptome $n=168,743$ sequences and $n=6,525,399$ for each metagenome. When comparisons were done between metatranscriptomes and metagenomes, each library was downsampled to the smallest metatranscriptome ($n=168,743$). Comparative functional gene analysis between samples was completed using Statistical Analysis of Metagenomic Profiles (STAMP) software (Parks and Beiko 2010). STAMP settings for pairwise comparisons between metagenomic and metatranscriptomic samples used a Fisher's exact test (one-sided) with Newcombe-Wilson calculations of Confidence Interval (0.95) and Bonferroni multiple test correction (Eloe et al. 2011). Pairwise Pearson correlations (r), gene diversity indices, and hierarchical clustering of metagenomic and metatranscriptomic samples were performed using *R* (R-Development-Core-Team 2011) using the package *vegan* (<http://vegan.r-forge.r-project.org/>).

Additionally, metagenomes and metatranscriptomes were mapped to published genomes with Bowtie2 using a local alignment and default settings. Our dataset of genomes consisted of 182 marine genomes assembled and annotated by the J. Craig Venter Institute (JCVI) as part of the Marine Microbial Genome Sequencing project (MMGSP). In addition, 60 marine genomes and 740 non-marine genomes from the dataset described in Yooseph et al (2010) were also used. Since our data included freshwater libraries, I also added five freshwater genomes from Actinobacteria (AC1 lineage) and Alphaproteobacteria (LD12 lineage) and a genome from a coastal Crenarchaeota to our genome dataset (unpublished data).

Results:

Sequencing of the five salinity gradient samples produced paired-end sequence reads with an average of 66 million fragments per metagenome and 67.5 million fragments per metatranscriptome (Table 5.1). I consider a fragment to be an intact paired-read, with sequences covering both ends. Despite efforts to remove rRNA from the total RNA pool prior to sequencing, an average of 81% rRNA was identified and removed from each metatranscriptomic library prior to analysis. A comparison between the metagenome and metatranscriptome for each sample showed most of the expressed annotated genes identified in the metatranscriptomes were also found in the metagenomes (Figure 5.2). The Venn diagrams in Figure 5.2, which depict the presence/absence of annotated genes in each sample, showed an almost complete inclusion of the number of expressed genes within the functional genes

across the salinity gradient, indicating the metatranscriptomes accurately represented expression of the metabolic potential found in each sample.

Comparisons between metagenomes showed that 60% of the annotated genes were shared between freshwater, estuary, and below plume samples, with unique genes making up 13% or less of the total (Figure 5.3). A similar pattern was found when comparing metagenomes from the freshwater, new plume, and below plume samples, with 63% of genes being shared. The metatranscriptomes showed less overlap between samples. A comparison of freshwater, new plume, and below plume metatranscriptomes showed only 22% of expressed annotated genes were shared, with uniquely expressed genes making up 25%, 27%, and 28% of the total in each sample, respectively (Figure 5.3). When comparing the estuary (salinity=3.9) to the freshwater and below plume samples, only 18% of expressed genes were seen in all three samples, but almost all of the expressed genes seen in the freshwater (89%) and below the plume (86%) were also expressed in the estuary (Figure 5.3).

Although the metagenomes appeared to share many functional genes across the salinity gradient, these genes came from taxonomically distinct communities. Our previous study of 16S rRNA gene amplicon sequences showed taxonomically distinct bacterial communities in the river, estuary, plume, and coastal ocean (Fortunato et al. 2012). The taxonomy of sequences from the two ends of the paired-end Illumina sequences were nearly identical and were similar to the taxonomy of the amplicon sequences (Figure 5.4). For each metagenome I identified an average of 108,000 16S rRNA gene sequences. For each amplicon sample there was an average of 321 sequences, not including the estuary sample which only had 50 sequences.

Actinobacteria, Betaproteobacteria, and Verrucomicrobia taxa were abundant in the freshwater and estuary samples, but decreased in the plume as salinity increased. In contrast SAR11, Flavobacteria, and Rhodobacteraceae taxa increased with salinity. In the old plume sample there was a greater abundance of *Polaribacter* (2.5%) compared to the other plume samples. The below plume sample had a greater abundance of Gammaproteobacteria (25.5%), which was perhaps a reflection of the higher ocean salinity and deeper sampling depth.

This shift in taxonomic composition across the salinity gradient was also observed in our comparison of metagenomes and metatranscriptomes to sequenced microbial genomes (Figure 5.5). I found an over-representation of archaeal transcripts in our metatranscriptomes compared to the metagenomes, where transcripts matching archaeal genomes made up to 66% of all transcripts in some samples while making up less than 1% on average in metagenomic samples. Further analysis of this metatranscriptomic over-representation suggested that perhaps there was incomplete removal of archaeal rRNA, which subsequently mapped to archaeal genomes. For this reason, results presented only included metagenomic and metatranscriptomic sequences that mapped to bacterial genomes. Figure 5.5 includes the ten most abundant genomes to which sequences mapped for metagenomes and metatranscriptomes from each of the five samples.

As with Figure 5.4, Figure 5.5 shows a shift in the abundance of different taxonomic groups across the salinity gradient for both the metagenomic and metatranscriptomic samples. In the freshwater metagenome, freshwater *Actinobacteria* genomes made up 47% of mapped sequences. This percentage

decreased as salinity in the samples increased, although Actinobacteria genomes still made up 26% of the new plume metagenome (Figure 5.5). Transcripts matching Actinobacteria genomes were expressed in the freshwater metatranscriptome (15%), but a much larger percentage of Actinobacteria genomes (46%) were expressed in the estuary metatranscriptome. This is surprising since Actinobacteria genomes were not prevalent in the estuarine metagenome. The most abundant genomes found in the estuary metagenome were Cyanobacteria species (e.g. *Synechococcus*, *Cyanobium*), but these organisms were not well represented in the transcriptome, indicating they were present, but not active (Figure 5.5). In both the freshwater and estuary samples, the most abundant genomes found in the metagenomes, in general, were not the same as the most abundant genomes found in the metatranscriptomes. This, however, was not the case in the plume samples, where the most abundant genomes were found in both the metagenomes and metatranscriptomes, with the exception of a *Roseobacter sp.* genome that was highly expressed in the plume, but made up a small percentage of plume metagenomes (Figure 5.5). Abundant genomes in the new, old, and below plume samples included a Rhodobacteraceae genome (Alphaproteobacteria HTCC2255) and various SAR11 genomes. The deeper below plume sample also had high abundance of two Gammaproteobacteria genomes, similar to what was seen in the 16S data (Figures 5.4, 5.5).

To further compare the metagenomes and metatranscriptomes I clustered sequences hierarchically based on the distribution of sequences matching different SEED subsystem categories using pairwise Pearson correlations (Figure 5.6). Figure 5.6 depicts the clustering of metagenomes and metatranscriptomes at different

subsystem category scales, from basic metabolic functional categories (Level 1) to specific genes (Function level), and shows heatmaps depicting the distribution of these categories across samples. At every functional level, the metagenomes clustered separately from four of the five metatranscriptomes indicating that, in general, metagenome samples were more similar to one another than to their corresponding metatranscriptome (Figure 5.6). The estuary metatranscriptome, however, did cluster more closely with the metagenomes. Correlations between metagenomes were higher than between metatranscriptomes, indicating that the variation in gene function among samples was lower than variation in gene expression. This result was also seen in the corresponding heatmaps, particularly in Level 1, where the relative abundance of each functional category was similar across the five metagenomes but more variable across the metatranscriptomes (Figure 5.6).

Dendrograms of the metatranscriptomes showed that the freshwater and new plume samples as well as the old plume and below plume clustered together at each level. The estuary transcriptome, as mentioned, clustered separately from the other four (Figure 5.6). In addition to pairwise Pearson correlations, diversity and evenness of genes was also assessed (Table 5.2). Using the inverse Simpson index as a measure of gene diversity, I found that all metagenomes displayed much higher functional diversity than the metatranscriptomes and the abundance of these genes was more evenly distributed. The metatranscriptomes, however, had very low evenness, as all metatranscriptomes were dominated by a small number of highly expressed genes (Figure 5.6, Table 5.2).

To look more closely at functional differences among and between the metagenomes and metatranscriptomes, I compared the relative abundance of genes corresponding to different functional processes across the salinity gradient. These processes were nitrogen, sulfur, and phosphorus metabolism, carbon fixation, photosynthesis, and stress (Figure 5.7). In general, the metagenomes varied more strongly with the salinity gradient compared to the metatranscriptomes, with genes associated with sulfur metabolism and stress proportionally increasing from river to ocean. Across the salinity gradient there was a relative increase in genes for sulfur oxidation, DMSP degradation, and taurine utilization, and a decrease in thioredoxin-disulfide reductase genes, which is an enzyme important to cell reduction-oxidation reactions (Figure 5.7). There was a relative decrease in genes associated with phosphorus metabolism across the salinity gradient, with the freshwater metagenome having proportionally more genes for high-affinity phosphate uptake and alkylphosphonate utilization compared to below the plume. There was not a clear pattern with salinity in nitrogen metabolism genes. However, as the below plume sample had nearly twice the proportion of genes associated with nitrogen metabolism than the other four metagenomes with a large percentage of nitric oxide synthase genes present. There were very few nitrogen fixation genes, but denitrification genes were present in the freshwater and below plume samples, perhaps associated with anaerobic processes occurring within suspended particles. The relative abundance of carbon fixation genes was even among the metagenomes, with a slightly higher relative abundance in the estuary. Photosynthesis genes were abundant in the estuary

as well as the old plume, with the old plume containing proportionally more genes for proteorhodopsins compared to the estuary (Figure 5.7).

As for the metatranscriptomes, there was no clear pattern for expression of genes across the salinity gradient, but there were differences between samples. Similar to the metagenomes, the relative abundance of nitrogen metabolism genes was greatest in the below plume sample, with a proportionally higher expression of nitric oxide synthase genes, the same gene that made up a large percentage of the metagenome (Figure 5.7). Expression of sulfur metabolism genes was variable across samples, with the old plume having the highest relative gene expression, especially of genes associated with using glutathione, an organic sulfur compound. The relative expression of stress genes, both oxidative and osmotic, was also highest in the old plume metatranscriptome (Figure 5.7). Looking at genes expressed for carbon fixation and photosynthesis, the freshwater and new plume samples had the highest relative expression of carbon fixation genes with the lowest expression of photosynthesis genes, while the estuary, new plume, and old plume showed the opposite pattern (Figure 5.7). Although there were differences in relative gene expression patterns for important metabolic processes among metatranscriptomic samples, many of the differences were variable and difficult to interpret.

Discussion:

Most metagenomic and metatranscriptomic studies of aquatic microbial communities have focused on the metabolic potential and gene expression patterns of the surface ocean, with few studies describing freshwater systems (Ghai et al. 2011,

Hewson et al. 2010, Poretsky et al. 2009b, Rusch et al. 2007). This study represents the first to connect fresh and marine environments in order to understand the changes in metabolic function across a salinity gradient. Our results suggest that although taxonomically distinct communities were present from river to ocean, the metabolic potential of these communities was highly similar. Gene expression patterns across the salinity gradient, however, were extremely different and reflect the short-time scales on which microbial processes occurred and persisted in an environment.

Previous work of 16S amplicon pyrosequences has shown taxonomically different communities in the river, estuary, plume, and coastal ocean (Fortunato et al. 2012). This pattern was also reflected in our metagenomic samples, with a clear shift from freshwater taxa such as Actinobacteria and Betaproteobacteria to marine Alphaproteobacteria, Flavobacteria and Gammaproteobacteria. Despite the large difference in the number of 16S sequences between the metagenomes (average=108,000 sequences) and the 16S amplicons (average=321 sequences) for each sample, results from the taxonomic analysis were very similar, with the exception of the estuary, where the number of amplicon sequences was very low and probably did not accurately reflect community composition. This similarity despite the large difference in sequencing depth validates shallow 16S amplicon surveys as accurate depictions of the dominant microbial community. In addition, similar results from metagenomic sequencing strengthen the patterns of abundance of these taxonomic groups and the large differences between freshwater, estuary, and plume microbial community composition.

This taxonomic shift was also observed in the abundance of metagenomic and metatranscriptomic sequences mapping to freshwater and marine genomes. Across the salinity gradient the freshwater Actinobacteria AC1 clade was extremely abundant in the freshwater, and decreased with increasing salinity. These freshwater bacteria were also present in the new plume metagenome and showed expression in the metatranscriptome. This showed that perhaps the new plume sample was not well mixed, but instead consisted of several thin layers of water integrated during sampling, including a layer of freshwater and higher salinity coastal water. Freshwater discharge onto the coast forms a surface lens on top of higher salinity coastal waters (Hill and Wheeler 2002) and thus the timing of the sample may have captured this layered structure before substantial vertical mixing occurred, especially since measured wind speed was low at the time of sampling (3 m/s).

As freshwater bacteria were diluted out along the salinity gradient, dominant marine bacteria appeared to mix into the plume, including many *Pelagibacter* and Gammaproteobacteria genomes, which were highly abundant in both metagenomes and metatranscriptomes. The most abundant genome detected in the new and old plume metagenomes was an alphaproteobacterium belonging to the Rhodobacteraceae family. In a previous study of 16S amplicons, a Rhodobacteraceae taxon was found to be highly abundant in the plume during the summer (Aug-Sept) and was found to be a key indicator of the summer plume microbial community (Fortunato et al, in prep). Thus the high abundance and expression of a Rhodobacteraceae genome in the plume sample further solidified the importance of these taxa to plume metabolic processes.

Although the taxonomic composition of the microbial community shifted across the salinity gradient, the metabolic potential from fresh to marine waters was highly similar. Similarity between metagenomes was evident in Venn diagrams, which showed almost complete overlap of gene functions across the salinity gradient (Figure 5.3). Gene functions were on average 78% similar between metagenome samples (Pairwise Bray-Curtis). Evident from the dendrograms of Figure 5.6, metagenome samples clustered separately from the metatranscriptomes and had high pairwise Pearson correlations at all subsystem category levels with average correlations of 0.99, 0.98, and 0.88 at Level 1, Level 2, and the functional level respectively. In fact, the Pearson correlation between the freshwater and below plume metagenomes at the finest-scale gene function level was 0.86, indicating these two metagenomes share a high number of genes in the same ratios despite the large difference in salinity. Visually, the heatmaps of Figure 5.6 also were similar between the relative abundance of genes in each sample, especially at Level 1, where the heatmap of gene categories was nearly identical from freshwater to below plume.

Contrary to this study, other metagenomic studies across environmental gradients saw large differences in metabolic potential. Fierer et al (2012) found significant changes in metagenomes across a nitrogen gradient in soil and Stewart et al. (2012) found a shift from aerobic to anaerobic metabolism across an oxygen minimum zone. Additionally, Ghai et al (2011) found a large difference between a freshwater metagenome from the Amazon to metagenomes from the Global Ocean Survey (GOS), specifically a high abundance of genes associated with microbial heterotrophy in fresh vs. marine waters. These studies however, used samples from

environments that are known to possess large differences in metabolic functions, e.g. aerobic vs. anaerobic waters, heterotrophic river systems vs. oligotrophic open ocean. The salinity gradient of the Columbia River coastal margin, however, does not have these large metabolic gradients. For example, the majority of organic matter in the Columbia River, unlike the Amazon, is not allochthonous material, but derived from river phytoplankton blooms (Prahl et al. 1998), and thus genes for the breakdown of terrestrial-derived organic matter may not be as prevalent as in the Amazon. All samples, even the below plume sample, were taken in the photic zone and did not span large oxygen, depth, or nutrient gradients where there would be a large shift in metabolic function. In a previous study of surface ocean, metagenomes from two different ocean basins were found to be highly similar (Rusch et al. 2007). Thus, despite salinity differences, surface metabolic potentials appeared to be similar from river to ocean.

Although metagenomes were highly similar, there were some trends in functional gene abundances across the salinity gradient. The relative abundance of genes associated with sulfur metabolism increased from freshwater to below plume, corresponding with the increase in sulfate availability in higher salinity waters. A relative increase in genes associated with DMSP breakdown was seen from estuary to plume, indicating the importance of the organic sulfur source in marine microbe metabolisms. Relative abundance of genes associated with phosphorous metabolism decreased from fresh to high salinity waters, perhaps reflective of the limited phosphorus pool in river systems due to particle bound complexes. The freshwater sample contained the most high-affinity phosphate uptake genes. Additionally, the

freshwater had the highest chlorophyll *a* concentrations (10.6 µg/L) and so these high-affinity transporters may be used to scavenge phosphorus during nutrient depleting phytoplankton blooms. The relative abundance of nitrogen genes was similar across the salinity gradient, except for a high proportion of nitric oxide synthase genes in the below plume sample. These genes are known to help cells counter oxidative stress and for transcriptional regulation (Rafferty 2011), but it is unclear why there would be more of these genes below the plume.

Gene expression determined from metatranscriptomes across the salinity gradient was more variable than gene presence. Metatranscriptomic libraries were on average only 29% similar based on Bray-Curtis similarity of expressed gene functions. Metatranscriptomes, except for the estuary sample, clustered separately from the metagenomes, and had very low Pearson correlations across all subsystem category levels, which on average were 0.70, 0.52, and 0.30 for Level 1, Level 2, and the functional gene level respectively. The lowest correlation was seen between the freshwater and below plume samples, which had a correlation of 0.04, showing almost complete separation of these two metatranscriptomes. Compared to the heatmaps of the metagenomes, gene expression in the metatranscriptome was highly variable among samples. Unlike in the metagenomes, each of metatranscriptomes was dominated by a few highly expressed genes, including a ribokinase gene in the freshwater and a phage capsid protein gene in the below plume sample. The high expression of a limited number of genes in each metatranscriptome was reflected in the low gene evenness measured. An average evenness of 0.05 was found for the metatranscriptomes compared to 0.23 for the metagenomes, where an evenness of 1

implies an equal number of sequences per gene function. Stewart et al. (2012) also observed low evenness in metatranscriptomic samples. This high variability in expressed gene proportions is perhaps reflective of the short-time scales on which many microbial processes operate and the limited view metatranscriptomic samples provide.

Shared expressed gene content across the salinity gradient was more variable than seen with metagenomes. A Venn diagram between the freshwater, estuary, and below plume samples showed a high number of expressed gene functions only present in the estuary, with almost all of the freshwater and below plume expressed genes also expressed in the estuary, suggesting a combination of gene expression patterns from the river, ocean, and uniquely within the estuary. The estuary metatranscriptome had the highest expressed gene diversity and evenness compared to the other metatranscriptomes, and thus may be why the estuary clustered with the metagenomes at all subsystems levels in Figure 5.6. The estuary, thus, can be seen as a highly diverse environment hosting a mixture of marine and freshwater microbial metabolic processes. The freshwater and new plume metatranscriptomes clustered together at subsystem category Level 1 and 2 and closely at the functional level. The similarity between these samples was seen in the abundance of typical freshwater genomes in both the freshwater and new plume samples and also in expression of genes involved in carbon fixation, photosynthesis, nitrogen and sulfur metabolism, and stress, which further supports the idea that the new plume sample was a stratified sample of freshly discharged river water and higher salinity water beneath.

A study of ocean surface water metatranscriptomes from different ocean basins found that gene transcription patterns were mostly shared between geographically diverse locations (Hewson et al. 2010), unlike what was seen along a salinity gradient in this study. Thus, although the metagenomic samples shared similar metabolic potential similar to what was seen in the surface ocean metagenomes, gene expression across the salinity gradient does not compare to patterns seen in surface ocean metatranscriptomes. There were no functional processes that exhibited a clear pattern of gene expression with salinity, as all were highly variable between metatranscriptomes. Relative expression of nitrogen metabolism genes was highest in the below plume sample, with a high expression of nitric oxide synthase genes, which were also abundant in the metagenome. Although these genes combat UV and oxidative stress (Rafferty 2011), stress gene expression was relatively low in the below plume sample. Nitric oxide synthase is also important for regulation of transcription (Rafferty 2011), but it unclear why the below plume sample would have higher expression of this gene. Gene expression patterns in the old plume sample showed signs of productivity, with high relative expression of phosphate metabolism genes, photosystem and light harvesting complex genes, and genes involved in DMSP and other organic sulfur compound breakdown. Although chlorophyll *a* concentrations were lower (6.3 $\mu\text{g/L}$), expression of these genes suggested a highly productive plume sample. Relative expression of stress genes, especially genes for osmotic stress, was also highest in the old plume sample, suggesting the presence of a stressed river community in the plume. Gene expression patterns in the new plume sample were very different than the old plume and were

more similar to the freshwater sample. Both the freshwater and new plume sample expressed proportionally high amounts of carbon fixation genes, but had low relative expression of photosynthesis genes. The opposite pattern was seen for the estuary, old plume and below plume samples, where there was low carbon fixation but high relative expression of photosynthesis genes. Although samples were taken on different days, they were all taken at a similar time of day, well after sunrise and well before sunset, and thus the mechanism behind these expression patterns remains unclear.

In conclusion, I found that, like previous studies, taxonomically different communities were present across the salinity gradient as evident from 16S rRNA genes as well as the abundance of fresh and marine genomes. Conversely, metabolic potential across the salinity gradient was similar among all environments, indicating that taxonomically distinct communities have similar metabolic functions from river to ocean. Gene expression patterns, however, were highly variable and showed no pattern with salinity, demonstrating that expression changed rapidly and transcript turnover was high as microbes were constantly adapting to their environment, especially in a complex river-influenced coastal zone.

Table 5.1: Environmental and sequence information for each sample metagenome and metatranscriptome. A fragment is an intact paired-read, with sequences covering both ends. The number of annotated reads is the total of both fragments and single reads assigned to SEED subsystem functional categories.

Sample	Salinity	Depth (m)	Nucleic Acid	No. of fragments (x10⁶)	%rRNA	No. annotated
Freshwater	0	0.5	DNA	60	0.5	6,525,399
			mRNA	86	85.2	1,567,407
Estuary	3.9	0.5	DNA	60	2.0	6,924,410
			mRNA	87.5	87.5	222,737
New Plume	15.5	1	DNA	84	0.4	15,834,963
			mRNA	67.5	69.8	312,123
Old Plume	25.3	1	DNA	67.5	1.0	13,161,525
			mRNA	63	81.3	829,226
Below Plume	33	16	DNA	61	1.3	11,768,111
			mRNA	33.5	80.8	168,743

Table 5.2: Gene function diversity and richness for each metagenome and metatranscriptome. Reciprocal Simpsons diversity ($1/D$) was used for diversity calculations, where $D = \sum P_i^2$ and P is the proportion of the total number of sequences represented by the i th unique gene function. Evenness was calculated by $(1/D)/S$, where S is gene richness (number of gene functions) for each sample. Evenness ranges from 0 to 1, where an evenness of 1 implies an equal number of sequences per gene function.

	DNA		mRNA	
	<i>Diversity</i>	<i>Evenness</i>	<i>Diversity</i>	<i>Evenness</i>
Freshwater	888.10	0.24	13.93	0.02
Estuary	886.66	0.24	284.20	0.12
New Plume	780.30	0.23	33.13	0.03
Old Plume	794.98	0.22	107.25	0.05
Below Plume	890.51	0.22	32.19	0.03

Figure 5.1: Map Columbia River, estuary and plume sampling locations. Plume samples were collected at the same sampling location on consecutive days.

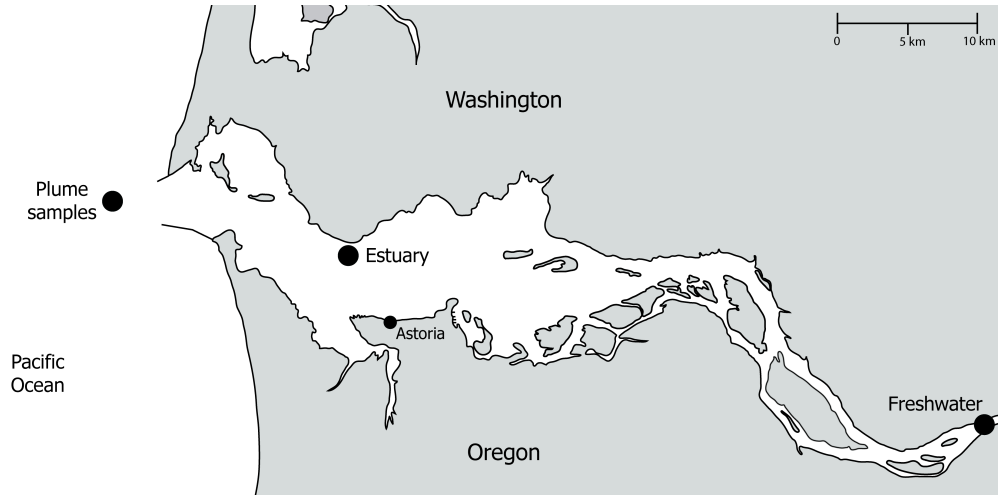


Figure 5.2: Shared gene functions between the metagenome and metatranscriptome for each sample and overall. All metagenomes and metatranscriptomes were normalized to 168,743 sequences for accurate comparison.

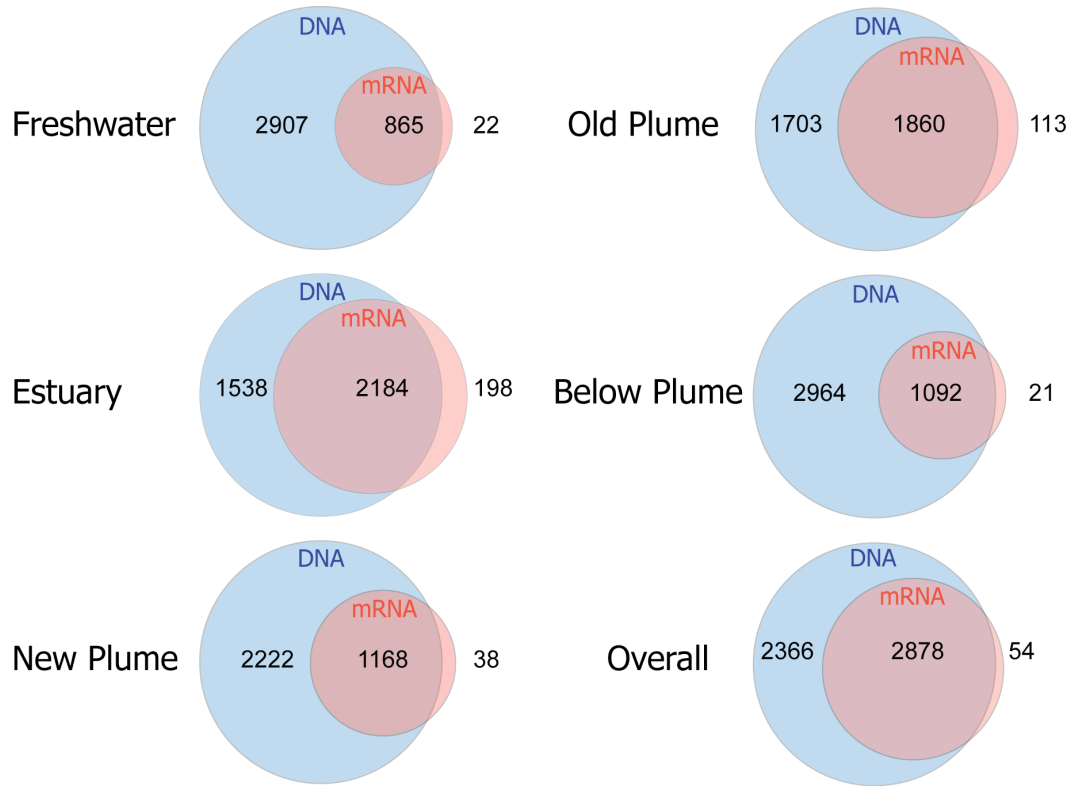


Figure 5.3: Shared gene functions between Freshwater, Estuary, New Plume, and Below Plume metagenomes and metatranscriptomes. Metagenomes were normalized to 6,525,399 sequences. Metatranscriptomes were normalized to 168,743 sequences.

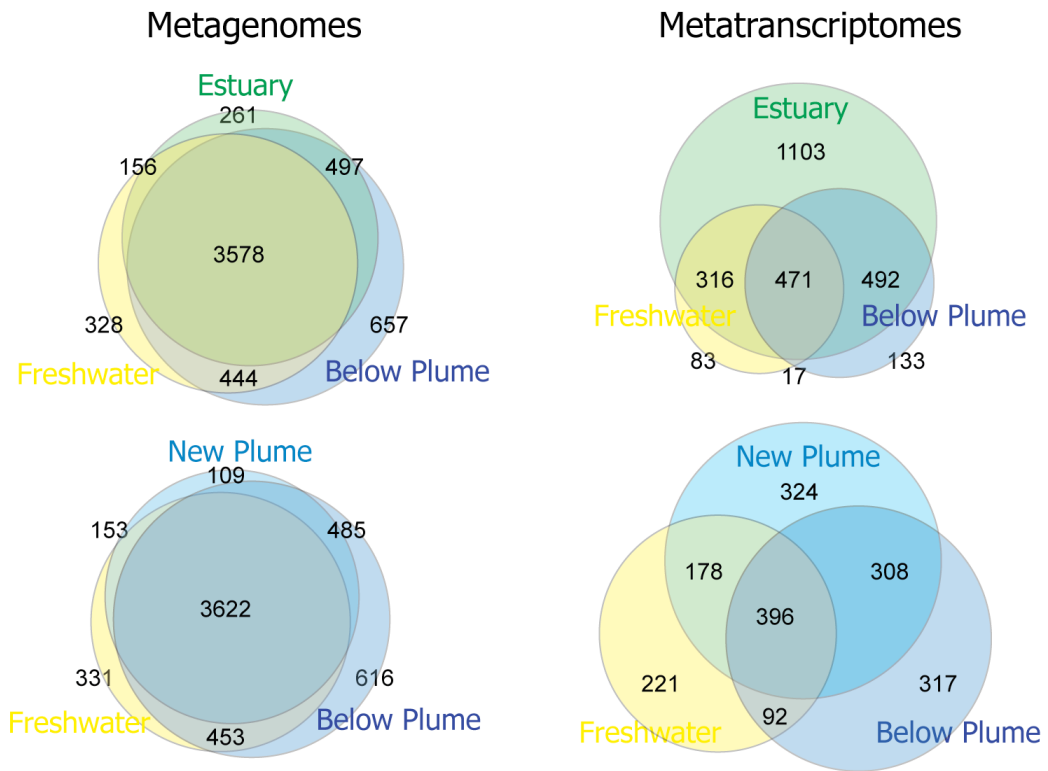


Figure 5.4: Taxonomy of 16S rRNA genes identified in each metagenome and corresponding 16S rRNA gene amplicon sample. There are two bars for each metagenome, which represent 16S sequences indentified from each end of the paired-end fragment.

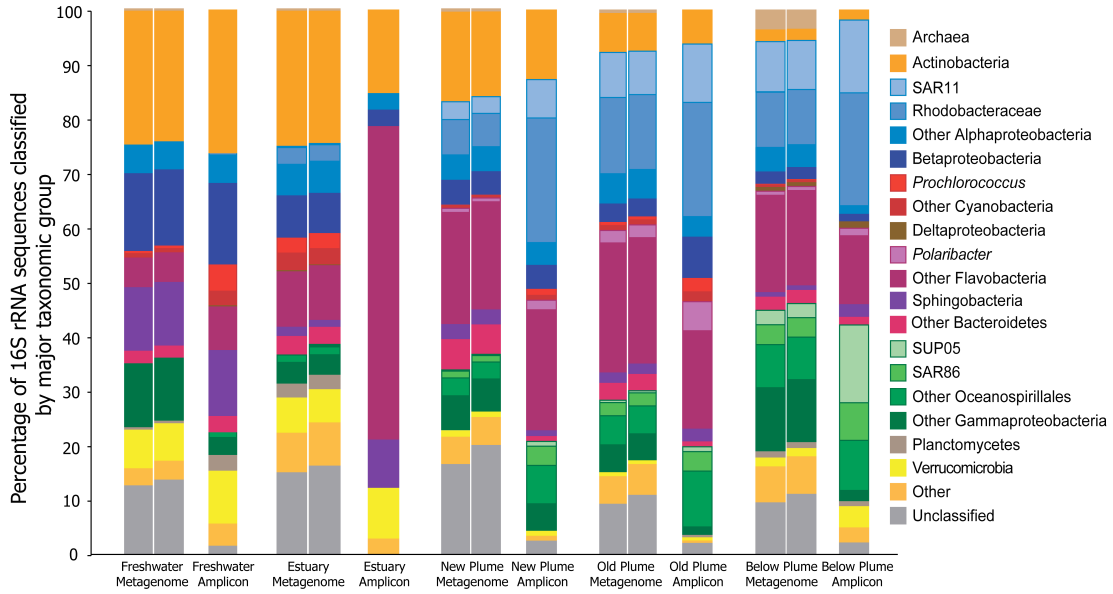


Figure 5.5: Relative abundance of sequenced genomes in each metagenome and metatranscriptome. Blue bubbles represent metagenomes, red represent metatranscriptomes.

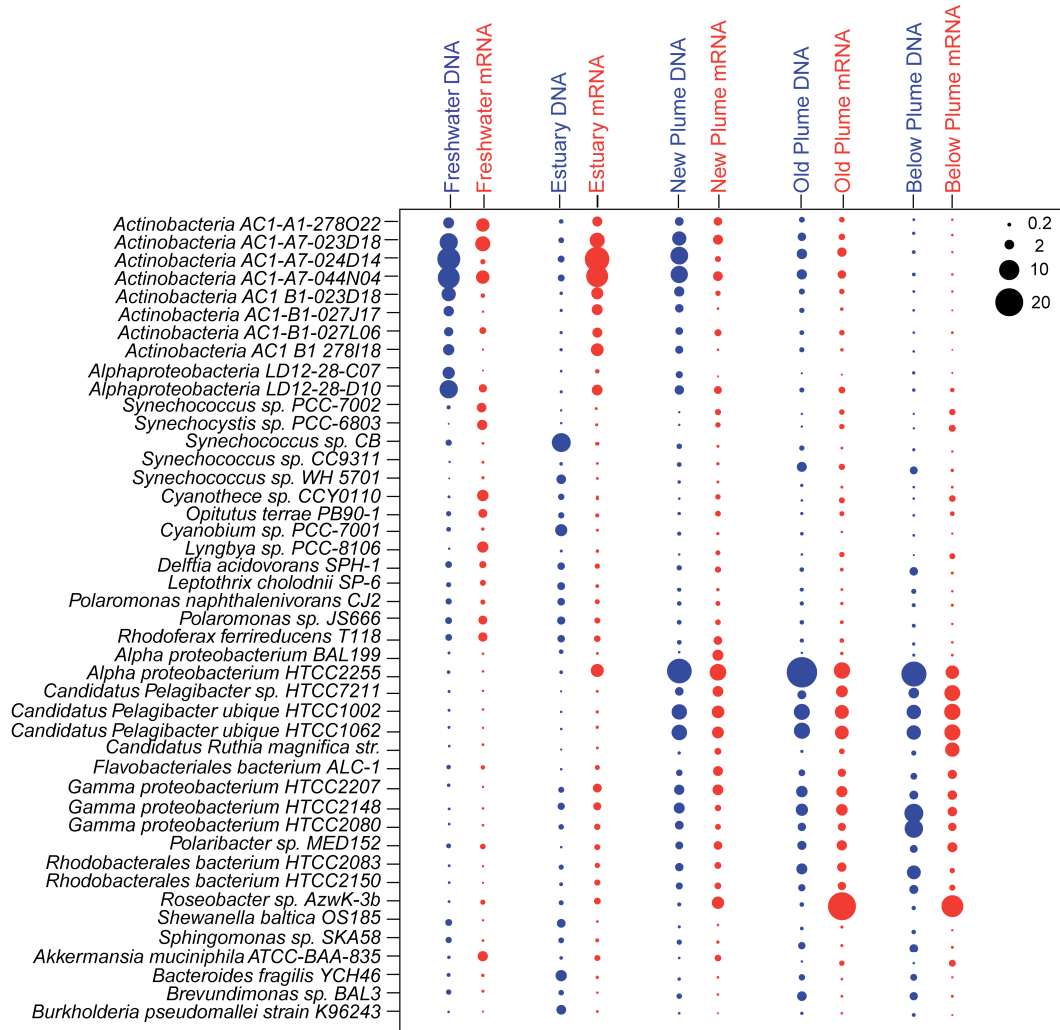
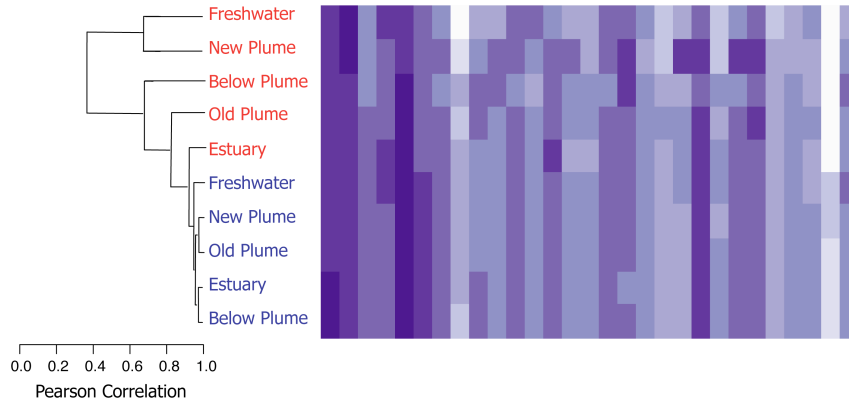
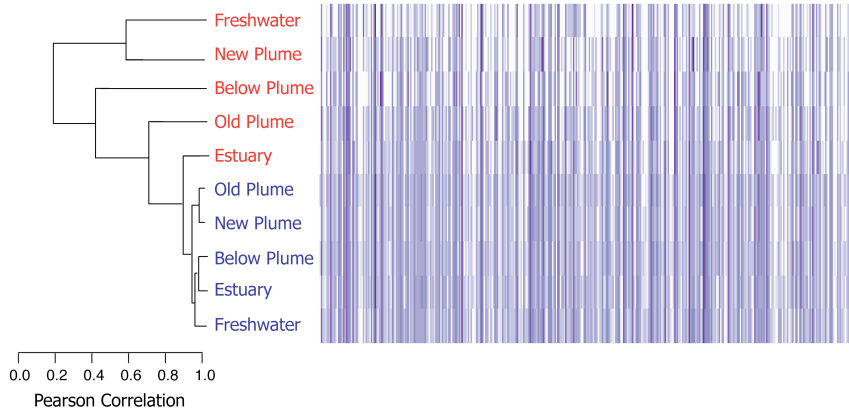


Figure 5.6: Relatedness of metagenomes and metatranscriptomes. Heatmaps show the relative abundance of sequences matching metabolic function categories at three levels of the SEED subsystem functional hierarchy. Dendrograms are based on hierarchical clustering of pairwise Pearson correlation coefficients. Blue and red represent metagenomes and metatranscriptomes respectively. All samples were normalized to 168,743 sequences.

Subsystem Level 1: DNA and mRNA



Subsystem Level 2: DNA and mRNA



Subsystem Function: DNA and mRNA

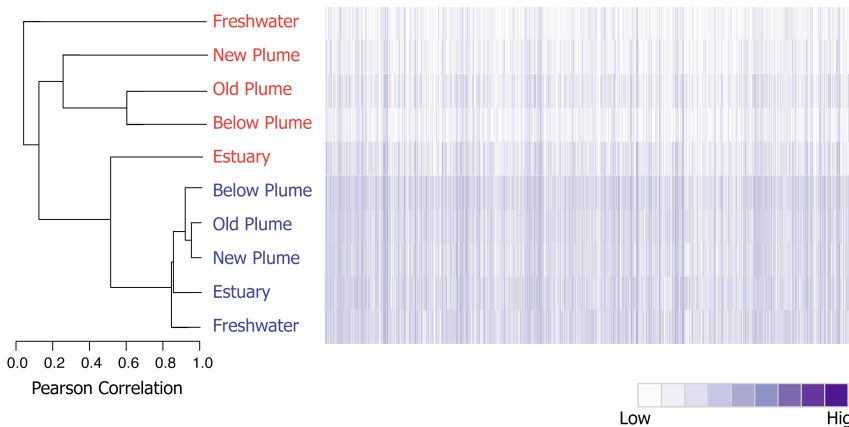


Figure 5.7: Abundance of functional genes present and expressed across all samples. Functional assignments are based on SEED subsystem categories. Metagenomic data is on the left, metatranscriptomic data on the right.

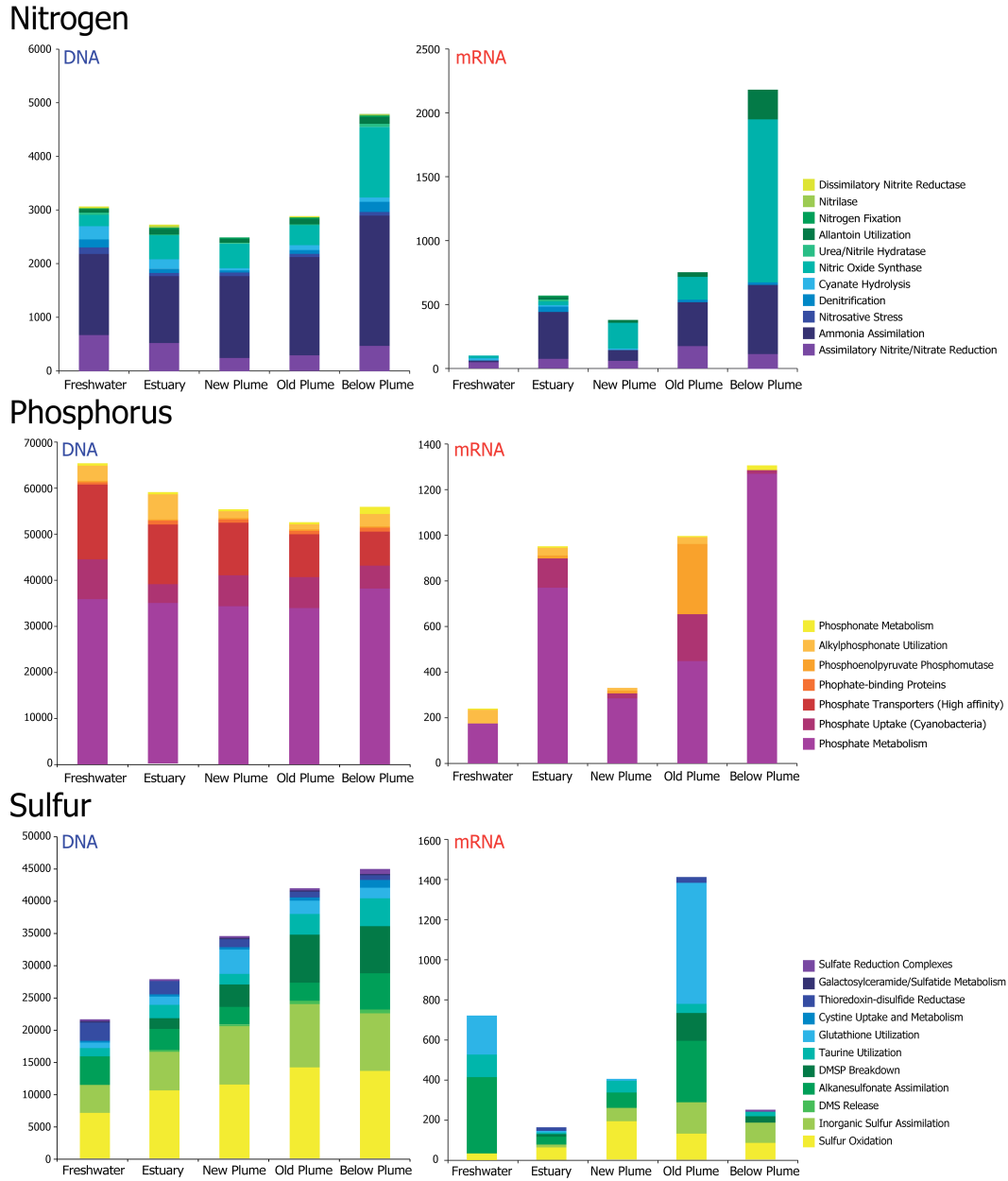
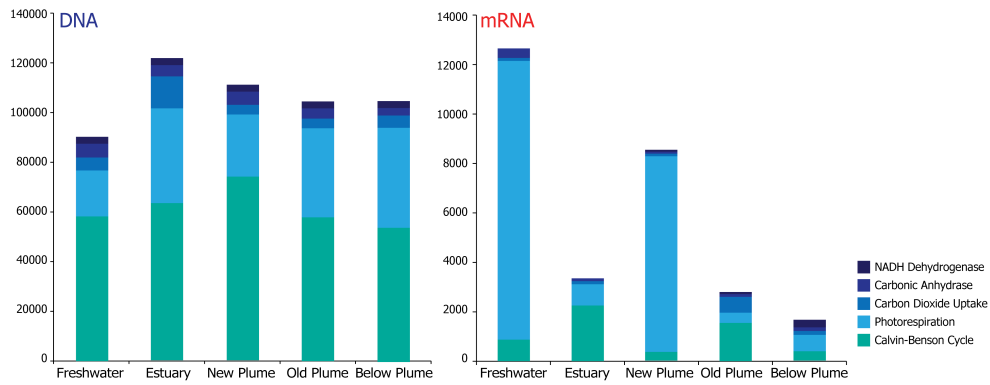
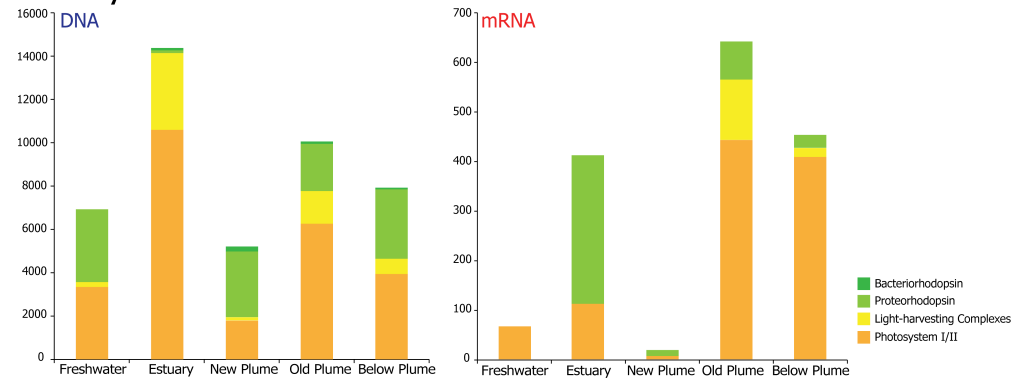


Figure 5.7: *continued*

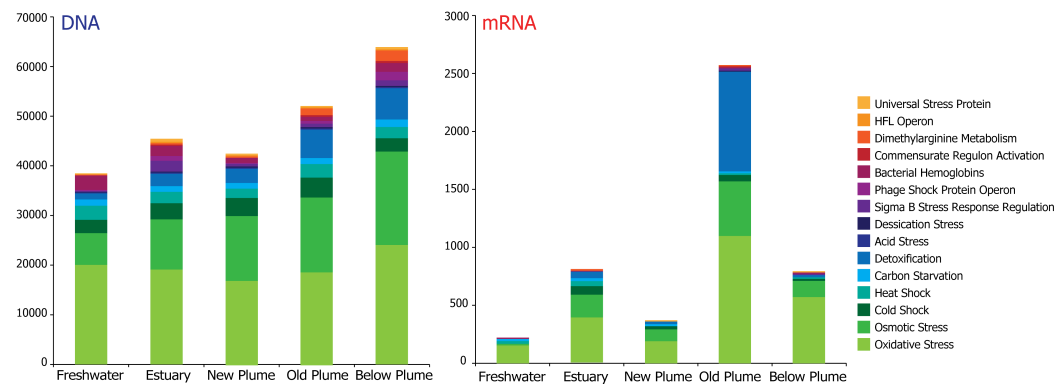
Carbon Fixation



Photosynthesis



Stress



CHAPTER 6:
FINAL CONCLUSIONS

This research is the first to characterize in detail the bacterial communities of the Columbia River coastal margin from freshwater to the deep ocean. I found that the phylogenetic composition of bacterioplankton communities separated by coastal environment based on structuring physical factors and that this great spatial variability overwhelmed any seasonal patterns across the entire coastal dataset. Taxonomically distinct communities were observed from river to the deep ocean, with different taxonomic groups characterizing each specific coastal environment. Freshwater influenced environments, from the river, to the estuary, to the plume, showed seasonal shifts in bacterioplankton community composition and reassembly of the same community each year. Seasonal shifts in communities in these environments were correlated most strongly with environmental factors that affected the rate of freshwater flow. With these seasonal shifts, I saw distinct taxa become dominant as environmental conditions changed. In the coastal ocean, however, seasonality of communities was not apparent but varied longitudinally with proximity to the coast and across environmental gradients. Combining this taxonomic survey with metagenomic and metatranscriptomic data across the salinity gradient, I found that although taxonomy changes drastically from river to ocean, bacterioplankton communities share much of the same metabolic function. Expression of these functions, however, was highly variable, with no obvious relationship with salinity.

Spatially, bacterioplankton communities separated into distinct coastal environments defined by the strong salinity, temperature, and depth gradients from river to ocean and from surface to deep. With a small dataset of 71 samples I saw communities separated into five different environments, but with the inclusion of

more samples, from 71 to 300, communities further separated into seven location-specific environments: river, estuary, plume, epipelagic, mesopelagic, shelf bottom (depth < 350 m), and slope bottom (depth > 850 m). These strong spatial patterns were even more solidified with an expansion of our dataset to nearly 600 samples across four years. Thus, this broad-scale microbial biogeographical survey gave an accurate depiction of how bacterioplankton communities were structured across the physical gradients of a coastal zone. Once bacterioplankton communities separated into environments based on physical conditions, factors related to primary and secondary production were found to influence community variation within environments, especially in the coastal ocean, where chlorophyll *a* and bacterial production appeared to be the two most important factors influencing community variability in the plume, epipelagic, and shelf bottom during our study in August 2007. In the summer months, the Pacific Northwest coast is highly productive due to upwelling of nutrients from the deep ocean, with the Washington coast being more productive than Oregon (Hickey and Banas 2003, Hickey et al. 2010). This spatial variation in primary and secondary production was reflected in bacterioplankton community composition as communities changed depending on proximity to shore and location along the coast. The spatial patterns of bacterioplankton communities were highly robust, as our extensive 16S biogeographical survey found communities were first separated by the physical parameters that define the coastal zone, and then secondarily by location-specific environmental factors.

In addition to this dataset spanning large spatial gradients from river to the deep ocean, I was also able to study temporal variability of bacterioplankton

communities across a four-year sampling period. I found temporal variability across the coastal zone to be obscured by the strong spatial variability. Several studies of coastal zone bacterioplankton identified time as the principle axis of community variability (Fuhrman et al. 2006, Gilbert et al. 2009, Kan et al. 2006b, Stepanauskas et al. 2003), but these studies were restricted to one environment type within which spatial variability of bacterioplankton communities was limited. Thus, while temporal variability may occur within many marine habitats, it was clear within our dataset that spatially variable environmental factors that structure the coastal environment (e.g., salinity, depth) were more important than temporal changes in determining bacterioplankton community composition. Temporal variability could only be resolved within some environments, specifically those influenced by freshwater flow. Our initial dataset of 300 samples over an annual cycle (2007-08) showed distinct seasonal communities in the river, estuary, and plume. These environments supported distinct early (April-July) and a late (August-November) year communities. The river also showed evidence of a unique spring community.

With the expansion of this data to include nearly 600 samples over four years, I saw that the temporal patterns became more resolved, with a strong separation of communities into early and late year in the river and estuary. In the plume there was a separation of a distinct summer community compared to the rest of the year. The seasonal separation into early and late year communities in the river and estuary corresponded to changes in Columbia River flow, where maximum flow occurs in late spring and is minimum in late summer to early fall (Prah et al. 1998). In combination with river flow, the river community varied with inorganic and organic

nutrients, especially POC and PN ($\rho=0.686$), as higher river flow generally brings more particulate matter downstream, with high POC concentrations during the spring freshet, and low concentrations during summer months. Community composition in the estuary was strongly correlated with river flow and residence time ($\rho=0.550$) as the amount and composition of river water and the time that water spent in the estuary greatly affected the development of the bacterioplankton community. Across our four-year dataset, these seasonal communities reassembled each year across the salinity gradient and thus, much like other studies showing annual reassembly, these bacterioplankton were predictable based on environmental conditions (Andersson et al. 2010, Carlson et al. 2009, Crump et al. 2009, Fuhrman et al. 2006, Morris et al. 2005).

Throughout the coastal ocean, there was no seasonal variation in community composition, and within-environment variation appeared to be driven by changing environmental conditions from near to offshore. Although shelf and slope bottom community variation correlated with dissolved oxygen and temperature from near to off shore, environmental factors in the epipelagic varied little. In both the larger survey encompassing multiple years, and the smaller study in August 2007, epipelagic communities varied from near to off shore. In August 2007, this variation also extended along the coast, presumably because epipelagic communities were influenced by variation in summer production along the coast. This along shore variability was not seen in the larger dataset. The lack of along shore variability in the larger, multi-year dataset showed the heterogeneity of the epipelagic environment as inter-annual variability of mixing, upwelling, river flow, and other physical and

biological factors made it difficult to define a set of environmental variables that best explained epipelagic community composition over multiple years and seasons.

The physical mixing of bacterioplankton communities was evident throughout the Columbia River coastal margin as taxonomic groups increased and decreased across the strong salinity and depth gradients of the dataset, with a distinct taxonomic fingerprint for each environment. In each of the seven environments across the coast, specific niches were created which led to the prevalence of different dominant taxonomic groups in each environment. The phylum SAR406 and the Gammaproteobacteria family SUP05 were most prevalent in the deep ocean and mixed out in shallower water. The opposite was seen for the common surface ocean bacteria groups SAR86 and Rhodobacteraceae, which were prevalent in the estuary, plume and epipelagic. Flavobacteria, especially *Polaribacter* taxa were extremely abundant in the plume. In the river community, Actinobacteria were dominant. The OTUs from these dominant taxonomic groups were also top indicators in their respective environments, suggesting these groups shaped community composition. Environmental conditions change seasonally, causing taxonomic composition to change as different taxa are able to dominate in different conditions. These seasonal shifts in taxonomic composition were evident in the river, estuary and plume, where shifts in the most dominant taxa appeared to drive changes in community composition. In the estuary specifically, changes in river flow and residence time also co-occurred with a shift in community composition. One Oceanospirillales OTU was more dominant during times of high flow and short residence times, and a Rhodobacteraceae OTU was dominant in low and long residence times.

From our 16S amplicon biogeographical studies, I saw changes in taxonomic composition across salinity, depth, and other environmental gradients of the coastal zone. This change in taxonomy across the salinity gradient was also evident in our metagenomic data, with a clear shift from freshwater taxa such as Actinobacteria and Betaproteobacteria to marine Alphaproteobacteria, Flavobacteria, and Gammaproteobacteria. Despite the large difference in the number of 16S sequences between the metagenomes (average=108,000 sequences per sample) and the 16S amplicons (average=321 sequences per sample), results from the taxonomic analysis were very similar. This taxonomic shift was also observed in the abundance of metagenomic and metatranscriptomic sequences mapping to freshwater and marine genomes. Unlike the taxonomy across the salinity gradient, the metabolic potential of microbial communities was similar from river to ocean. This similarity was evident in Venn diagrams, which showed almost complete overlap of gene functions and high Pearson correlation coefficients between samples, even at the finest subsystem functional category. The results were different with other metagenomic studies across oxygen and nutrient gradients (Fierer et al. 2012, Stewart et al. 2012). All of our samples were taken in the photic-zone and did not span steep depth or oxygen gradients, thus despite the salinity, the gene prevalence and abundance of these surface microbes was fairly similar. The metatranscriptomes were highly variable, although this variability was not in response to changes in salinity. I found expression patterns of genes associated with nutrient metabolism, carbon fixation, photosynthesis, and stress showed no clear pattern with salinity and overall were hard to interpret. This high variability across the metatranscriptomes demonstrates that

gene expression changes rapidly and transcript turnover is high as microbes adapt to a changing environment. Thus our analysis of the metabolic potential and gene expression across this coastal salinity gradient showed that freshwater and marine taxa were metabolically similar in these surface communities, but that the expression of these metabolic functions was highly variable.

Microbes are indicators of environmental change. The ability for microbes to quickly adapt growth rates and cell processes allows them to be sensitive to subtle shifts in biological or physical conditions. As evident by both our 16S amplicon and metagenomic work, the taxonomic composition of microbial communities shifted with spatially and seasonally fluctuating environmental factors as different taxa took advantage of specific conditions. Metabolically, these different taxa were quite similar, but the expression of these metabolisms was not yet fully understood. Together the chapters of this dissertation present a complete taxonomic and functional study of microbial community variability in a complex coastal environment, and lays the groundwork for these communities to be further explored and eventually predicted.

References:

- Acinas SG, Anton J, Rodriguez-Valera F (1999). Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl Environ Microbiol* **65**: 514-522.
- Alonso-Saez L, Aristegui J, Pinhassi J, Gomez-Consarnau L, Gonzalez JM, Vaque D *et al* (2007). Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean. *Aquat Microb Ecol* **46**: 43-53.
- Andersson AF, Riemann L, Bertilsson S (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *Isme J* **4**: 171-181.
- Barberan A, Bates ST, Casamayor EO, Fierer N (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *Isme J* **6**: 343-351.
- Berdjeb L, Ghiglione JF, Jacquet S (2011). Bottom-Up versus Top-Down Control of Hypo- and Epilimnion Free-Living Bacterial Community Structures in Two Neighboring Freshwater Lakes. *Appl Environ Microbiol* **77**: 3591-3599.
- Bernhard AE, Colbert D, McManus J, Field KG (2005). Microbial community dynamics based on 16S rRNA gene profiles in a Pacific Northwest estuary and its tributaries. *FEMS Microbiol Ecol* **52**: 115-128.
- Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH (2008). Metagenomic signatures of the Peru Margin seafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci U S A* **105**: 10583-10588.
- Blumel M, Suling J, Imhoff JF (2007). Depth-specific distribution of Bacteroidetes in the oligotrophic eastern Mediterranean Sea. *Aquat Microb Ecol* **46**: 209-224.
- Bouvier TC, Giorgio PAd (2002). Compositional Changes in Free-Living Bacterial Communities along a Salinity Gradient in Two Temperate Estuaries. *Limnol Oceanogr* **47**: 453-470.
- Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM *et al* (2009). Microbial community structure in the North Pacific ocean. *Isme J* **3**: 1374-1386.
- Campbell BJ, Yu LY, Heidelberg JF, Kirchman DL (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci U S A* **108**: 12776-12781.
- Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science* **330**: 1375-1378.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

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

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

Celussi M, Cataletto B, Umani SF, Del Negro P (2009). Depth profiles of bacterioplankton assemblages and their activities in the Ross Sea. *Deep-Sea Res Part I-Oceanogr Res Pap* **56**: 2193-2205.

Clarke KR, Ainsworth M (1993). A Method of Linking Multivariate Community Structure To Environmental Variables. *Mar Ecol-Prog Ser* **92**: 205-219.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM *et al* (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**: D294-D296.

Cottrell MT, Kirchman DL (2000). 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.

Crump BC, Baross JA, Simenstad CA (1998). Dominance of particle-attached bacteria in the Columbia River estuary, USA. *Aquat Microb Ecol* **14**: 7-18.

Crump BC, Armbrust EV, Baross JA (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 BC, Baross JA (2000). Archaeaplankton in the Columbia River, its estuary and the adjacent coastal ocean, USA. *FEMS Microbiol Ecol* **31**: 231-239.

Crump BC, Kling GW, Bahr M, Hobbie JE (2003). Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. *Appl Environ Microbiol* **69**: 2253-2268.

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

- Crump BC, Hobbie JE (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnol Oceanogr* **50**: 1718-1729.
- Crump BC, Adams HE, Hobbie JE, Kling GW (2007). Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology* **88**: 1365-1378.
- Crump BC, Peterson BJ, Raymond PA, Amon RMW, Rinehart A, McClelland JW *et al* (2009). Circumpolar synchrony in big river bacterioplankton. *Proc Natl Acad Sci U S A* **106**: 21208-21212.
- Crump BC, Amaral-Zettler LA, Kling GW (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *Isme J* **6**: 1629-1639.
- Cury JC, Araujo FV, Coelho-Souza SA, Peixoto RS, Oliveira JAL, Santos HF *et al* (2011). Microbial Diversity of a Brazilian Coastal Region Influenced by an Upwelling System and Anthropogenic Activity. *PLoS One* **6**: e16553. doi:16510.11371/journal.pone.0016553.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496-503.
- Dufrene M, Legendre P (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345-366.
- Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R *et al* (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *Isme J* **6**: 1186-1199.
- Eiler A, Heinrich F, Bertilsson S (2012). Coherent dynamics and association networks among lake bacterioplankton taxa. *Isme J* **6**: 330-342.
- Eloe EA, Fadrosch DW, Novotny M, Zeigler Allen L, Kim M, Lombardo M-J *et al* (2011). Going Deeper: Metagenome of a Hadopelagic Microbial Community. *PLoS One* **6**: e20388. doi: 20310.21371/journal.pone.0020388.
- Ferrari VC, Hollibaugh JT (1999). Distribution of microbial assemblages in the Central Arctic Ocean Basin studied by PCR/DGGE: analysis of a large data set. *Hydrobiologia* **401**: 55-68.
- Field KG, Gordon D, Wright T, Rappe M, Urbach E, Vergin K *et al* (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63**: 63-70.

Fierer N (2008). Microbial biogeography: patterns in microbial diversity across space and time. In: Zengler K (ed). *Assessing uncultivated microorganisms: from the environment to organisms and genomes and back*. ASM Press: Washington, DC. pp 95-115.

Fierer N, Carney KM, Horner-Devine MC, Megonigal JP (2009). The Biogeography of Ammonia-Oxidizing Bacterial Communities in Soil. *Microb Ecol* **58**: 435-445.

Fierer N, Nemergut D, Knight R, Craine JM (2010). Changes through time: integrating microorganisms into the study of succession. *Res Microbiol* **161**: 635-642.

Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *Isme J* **6**: 1007-1017.

Fortunato CS, Crump BC (2011). Bacterioplankton Community Variation Across River to Ocean Environmental Gradients. *Microb Ecol* **62**: 374-382.

Fortunato CS, Herfort L, Zuber P, Baptista AM, Crump BC (2012). Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *Isme J* **6**: 554-563.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**: 3805-3810.

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

Fuhrman JA, Steele JA (2008). Community structure of marine bacterioplankton: patterns, networks, and relationships to function **53**: 69-81.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al* (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A* **105**: 7774-7778.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* **106**: 22427-22432.

Galand PE, Potvin M, Casamayor EO, Lovejoy C (2010). Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *Isme J* **4**: 564-576.

Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, de Oliveira TCS *et al* (2011). Metagenomics of the Water Column in the Pristine Upper Course of the Amazon River. *PLoS One* **6**: e23785. doi:23710.21371/journal.pone.0023785.

Ghiglione JF, Palacios C, Marty JC, Mevel G, Labrune C, Conan P *et al* (2008). Role of environmental factors for the vertical distribution (0-1000 m) of marine bacterial communities in the NW Mediterranean Sea. *Biogeosciences* **5**: 1751-1764.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *Isme J* **5**: 461-472.

Gilbert JA, Field D, Huang Y, Edwards R, Li WZ, Gilna P *et al* (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS One* **3**: e3042. 3010.1371/journal.pone.0003042.

Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T *et al* (2009). The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology* **11**: 3132-3139.

Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B *et al* (2012). Defining seasonal marine microbial community dynamics. *Isme J* **6**: 298-308.

Gomez-Alvarez V, Teal TK, Schmidt TM (2009). Systematic artifacts in metagenomes from complex microbial communities. *Isme J* **3**: 1314-1317.

Gomez-Pereira PR, Fuchs BM, Alonso C, Oliver MJ, van Beusekom JEE, Amann R (2010). Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *Isme J* **4**: 472-487.

Gordon LI, Jennings JC, Ross AA, Krest JM (1994). A suggested protocol for continuous flow analysis of seawater nutrients (phosphate, nitrate, nitrite, and silicic acid) in the WOCE Hydrographic Program and the Joiny Global Ocean Fluxes Study. *WHP Office Report* **91**: Revision 1, Nov. 1994.

Green JL, Bohannan BJM, Whitaker RJ (2008). Microbial biogeography: From taxonomy to traits. *Science* **320**: 1039-1043.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235-237.

Hedges JJ, Stern JH (1984). Carbon and nitrogen determination of carbonate-containing solids. *Limnol Oceanogr* **29**: 657-663.

- Hewson I, Fuhrman JA (2004). Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Appl Environ Microbiol* **70**: 3425-3433.
- Hewson I, Capone DG, Steele JA, Fuhrman JA (2006a). Influence of Amazon and Orinoco offshore surface water plumes on oligotrophic bacterioplankton diversity in the west tropical Atlantic. *Aquat Microb Ecol* **43**: 11-22.
- Hewson I, Steele JA, Capone DG, Fuhrman JA (2006b). Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar Ecol-Prog Ser* **311**: 67-77.
- Hewson I, Poretsky RS, Tripp HJ, Joseph PM, Jonathan PZ (2010). Spatial patterns and light-driven variation of microbial population gene expression in surface waters of the oligotrophic open ocean. *Environmental Microbiology* **12**: 1940–1956.
- Hickey BM (1989). Patterns and processes of shelf ecosystems and slope circulation. *Coastal Oceanography of Washington and Oregon*: 41-115.
- Hickey BM, Pietrafesa LJ, Jay DA, Boicourt WC (1998). The Columbia River Plume Study: Subtidal variability in the velocity and salinity fields. *J Geophys Res-Oceans* **103**: 10339-10368.
- Hickey BM, Banas NS (2003). Oceanography of the US Pacific Northwest Coastal Ocean and estuaries with application to coastal ecology. *Estuaries* **26**: 1010-1031.
- Hickey BM, Banas NS (2008). Why is the Northern End of the California Current System So Productive? *Oceanography* **21**: 90-107.
- Hickey BM, Kudela RM, Nash JD, Bruland KW, Peterson WT, MacCready P *et al* (2010). River Influences on Shelf Ecosystems: Introduction and synthesis. *J Geophys Res-Oceans* **115(C00B17)**: 1-26.
- Hill JK, Wheeler PA (2002). Organic carbon and nitrogen in the northern California current system: comparison of offshore, river plume, and coastally upwelled waters. *Prog Oceanogr* **53**: 369-387.
- Horner-Devine MC, Lage M, Hughes JB, Bohannon BJM (2004). A taxa-area relationship for bacteria. *Nature* **432**: 750-753.
- Janse I, Kardinaal WEA, Meima M, Fastner J, Visser PM, Zwart G (2004). Toxic and nontoxic microcystis colonies in natural populations can be differentiated on the basis of rRNA gene internal transcribed spacer diversity. *Appl Environ Microbiol* **70**: 3979-3987.

- Kan J, Wang K, Chen F (2006a). Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat Microb Ecol* **42**: 7-18.
- Kan JJ, Crump BC, Wang K, Chen F (2006b). Bacterioplankton community in Chesapeake Bay: Predictable or random assemblages. *Limnol Oceanogr* **51**: 2157-2169.
- Kataoka T, Hodoki Y, Suzuki K, Saito H, Higashi S (2009). Tempo-spatial patterns of bacterial community composition in the western North Pacific Ocean. *J Mar Syst* **77**: 197-207.
- Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD (2007). Synchrony in aquatic microbial community dynamics. *Isme J* **1**: 38-47.
- Kirchman DL, Keil RG, Simon M, Welschmeyer NA (1993). Biomass and production of heterotrophic bacterioplankton in the oceanic sub-arctic Pacific. *Deep-Sea Res Part I-Oceanogr Res Pap* **40**: 967-988.
- Kirchman DL, Cottrell MT, Lovejoy C (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology* **12**: 1132-1143.
- Kosro PM (2005). On the spatial structure of coastal circulation off Newport, Oregon, during spring and summer 2001 in a region of varying shelf width. *J Geophys Res-Oceans* **110(C10S06)**: 1-16.
- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2009). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.
- Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**: 357-359.
- Lauber CL, Hamady M, Knight R, Fierer N (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl Environ Microbiol* **75**: 5111-5120.
- Lee SH, Fuhrman JA (1991). Spatial and temporal variation of natural bacterioplankton assemblages studied by total genomic DNA cross-hybridization. *Limnol Oceanogr* **36**: 1277-1287.
- Legendre P, Legendre L (1998). *Numerical Ecology*. Elsevier Science B.V.: Amsterdam.

- Legendre P, Anderson MJ (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* **69**: 512-512.
- Long RA, Azam F (2001). Microscale patchiness of bacterioplankton assemblage richness in seawater. *Aquat Microb Ecol* **26**: 103-113.
- Lozupone C, Hamady M, Knight R (2006). UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**.
- Lozupone CA, Knight R (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences* **104**: 11436-11440.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
- Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER *et al* (2010). Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. *Isme J* **4**: 1252-1264.
- Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102-112.
- McCarren J, Becker JW, Repeta DJ, Shi YM, Young CR, Malmstrom RR *et al* (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci U S A* **107**: 16420-16427.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A *et al* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme J* **6**: 610-618.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**.
- Morris RM, Vergin KL, Cho JC, Rappe MS, Carlson CA, Giovannoni SJ (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 RM, Longnecker K, Giovannoni SJ (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology* **8**: 1361-1370.

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

Muyzer G, Dewaal EC, Uitterlinden AG (1993). Profiling of complex microbial-populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal-RNA. *Appl Environ Microbiol* **59**: 695-700.

Nelson CE (2009). Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. *Isme J* **3**: 13-30.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al* (2011). Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology* **13**: 135-144.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiol Mol Biol Rev* **75**: 14-49.

Parks DH, Beiko RG (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715-721.

Pinhassi J, Simo R, Gonzalez JM, Vila M, Alonso-Saez L, Kiene RP *et al* (2005). Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. *Appl Environ Microbiol* **71**: 7650-7660.

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

Poretsky R, Gifford S, Rinta-Kanto J, Vila-Costa M, Moran M (2009a). Analyzing Gene Expression from Marine Microbial Communities using Environmental Transcriptomics. *JoVE* **24**: <http://www.jove.com/index/details.stp?id=1086>, doi: 10.1002/1086.

Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M *et al* (2005). Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* **71**: 4121-4126.

Poretsky RS, Hewson I, Sun SL, Allen AE, Zehr JP, Moran MA (2009b). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology* **11**: 1358-1375.

Prahl FG, Small LF, Sullivan BA, Cordell J, Simenstad CA, Crump BC *et al* (1998). Biogeochemical gradients in the lower Columbia River. *Hydrobiologia* **361**: 37-52.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J *et al* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011). Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* **12**: doi: 10.1186/1471-2105-1112-1138.

R-Development-Core-Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Rafferty S (2011). Nitric Oxide Synthases of Bacteria – and Other Unicellular Organisms. *The Open Nitric Oxide Journal* **3**: 25-32.

Riemann L, Steward GF, Fandino LB, Campbell L, Landry MR, Azam F (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 Part II-Top Stud Oceanogr* **46**: 1791-1811.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al* (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: 398-431.

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

Schwing FB, Mendelssohn R (1997). Increased coastal upwelling in the California Current System. *J Geophys Res-Oceans* **102**: 3421-3438.

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

Shade A, Jones SE, McMahon KD (2008). The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environmental Microbiology* **10**: 1057-1067.

Shaw AK, Halpern AL, Beeson K, Tran B, Venter JC, Martiny JBH (2008). It's all relative: ranking the diversity of aquatic bacterial communities. *Environmental Microbiology* **10**: 2200-2210.

- Simek K, Pernthaler J, Weinbauer MG, Hornak K, Dolan JR, Nedoma J *et al* (2001). Changes in bacterial community composition and dynamics and viral mortality rates associated with enhanced flagellate grazing in a mesoeutrophic reservoir. *Appl Environ Microbiol* **67**: 2723-2733.
- Simon M, Glockner FO, Amann R (1999). Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. *Aquat Microb Ecol* **18**: 275-284.
- Smith MW, Herfort L, Tyrol K, Suci D, Campbell V, Crump BC *et al* (2010). Seasonal Changes in Bacterial and Archaeal Gene Expression Patterns across Salinity Gradients in the Columbia River Coastal Margin. *PLoS One* **5**: e13312. doi: 10.1371/journal.pone.0013312.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103**: 12115-12120.
- Stepanuskas R, Moran MA, Bergamaschi BA, Hollibaugh JT (2003). Covariance of bacterioplankton composition and environmental variables in a temperate delta system. *Aquat Microb Ecol* **31**: 85-98.
- Stewart FJ, Ulloa O, DeLong EF (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14**: 23-40.
- Sugimura Y, Suzuki Y (1988). A high-temperature catalytic-oxidation method for the determination of non-volatile dissolved organic-carbon in seawater by direct injection of a liquid sample. *Mar Chem* **24**: 105-131.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333**: 1296-1300.
- Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A (2010). Environmental distribution of prokaryotic taxa. *BMC Microbiol* **10**: doi: 10.1186/1471-2180-1110-1185.
- ter Braak CJF (1986). Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology* **67**: 1167-1179.
- Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA *et al* (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. *Isme J* **3**: 1148-1163.
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al* (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554-557.

- Troussellier M, Schafer H, Batailler N, Bernard L, Courties C, Lebaron P *et al* (2002). Bacterial activity and genetic richness along an estuarine gradient (Rhône River plume, France). *Aquat Microb Ecol* **28**: 13-24.
- Valderrama JC (1981). The simultaneous analysis of total nitrogen and total phosphorus in natural-waters. *Mar Chem* **10**: 109-122.
- Verardo DJ, Froelich PN, McIntyre A (1990). Determination of organic carbon and nitrogen in marine sediments using the Carlo Erba NA-1500 analyzer. *Deep-Sea Research Part a-Oceanographic Research Papers* **37**: 157-165.
- Vieira RP, Gonzalez AM, Cardoso AM, Oliveira DN, Albano RM, Clementino MM *et al* (2008). Relationships between bacterial diversity and environmental variables in a tropical marine environment, Rio de Janeiro. *Environmental Microbiology* **10**: 189-199.
- Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* **326**: 578-582.
- Wright SW, Jeffrey SW, Mantoura RFC, Llewellyn CA, Bjornland T, Repeta D *et al* (1991). Improved HPLC method for the analysis of chlorophylls and carotenoids from marine-phytoplankton. *Mar Ecol-Prog Ser* **77**: 183-196.
- Wright TD, Vergin KL, Boyd PW, Giovannoni SJ (1997). A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63**: 1441-1448.
- Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60-66.
- Zhou JZ, Bruns MA, Tiedje JM (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316-322.