Patterns of bacterial communities in aquatic ecosystems

Rachel K. Lange

A thesis submitted in partial fulfillment

of the requirements for the degree of:

Master of Science

University of Washington

2013

M. Claire Horner-Devine, Chair

Steven Roberts

Julian Olden

Program authorized to Offer Degree:

Aquatic & Fishery Sciences

Table of Contents

General introduction 3

Part 1: Patterns of bacterial alpha and beta diversity in aquatic ecosystems 5

Part 2: Challenges in describing bacterial beta diversity 11

Chapter 1: Patterns of bacterial richness and community composition in Hood Canal, WA, a seasonally oxygen-deficient estuary

Abstract 21

Introduction 22

Methods 24

Results 28

Discussion 34

Conclusion 39

References 41

Tables & Figures 49

Supplementary Tables & Figures 56

Appendix A: Assessing the performance of multiple community similarity measures for bacterial communities in the face of sampling biases

Introduction 59

Methods 61

Results and Discussion 65

Conclusion 70

References 71

Figures 74

Acknowledgements

This thesis was made possible by the generous funding of the NSF-GRFP, NSF-GK12 program, and the UW School of Aquatic and Fishery Sciences. I wish a sincere thank-you to my advisor, Claire Horner-Devine. Your dedicated mentorship, unwavering support, and always-positive encouragement are so inspiring. To my committee members, Steven Roberts and Julian Olden, I am truly grateful for your thoughtful and constructive advice throughout my research. Lastly, a special thanks to my many friends at SAFS for being constant sources of support through the ups and downs of the past three years.

I wish to dedicate this work to my family Ric, Cheryl, Joey, Alex, and Gram and Gramps who've selflessly supported me, pushed me to never give up, inspired me to appreciate nature and think creatively, and showed me how to enjoy the good things in life without dwelling on the negative.

General introduction

Understanding drivers of species diversity and composition is a central goal of ecology as it provides a window into how communities form and why diversity varies across ecosystems (Rosenzweig 1995). The presence and persistence of a species in a given location or habitat is dependent on a suite of factors, including resource availability and biotic and abiotic interactions (Rosenzweig 1995). Just as the ecosystem influences the distribution of species, species diversity and composition can also impact the ecosystem by altering the physical environment and nutrient cycling (Tilman et al. 1997). Therefore, understanding what may impact the geographic distribution, or biogeography, of species is essential for understanding ecology as well as understanding the functioning of ecosystems, especially in the face of increasing anthropogenic stress.

Microbial ecologists have long been characterizing and describing the distributional patterns of microbes; however, our understanding of the biogeography of microbes is less advanced than that for macroorganisms (Hanson et al. 2012; Martiny et al. 2006; Horner-Devine et al. 2007). Such an understanding for patterns of microbial biogeography and the processes driving these patterns are incredibly important because of the ecosystem services provided by the highly abundant and diverse organisms. Microbes are ubiquitous on Earth, inhabiting lakes, soils, oceans, and other organisms as well as more extreme environments including hot springs (Brock 1978), deep sea hydrothermal vents (McCollom & Shock 1997), saline pools (Anton et al. 2000), glacial ice (Boyd et al. 2011), acid mine tailings (Baker & Banfield 2003), and kilometers beneath Earth’s surface (White et al. 1998). Not only are microbes abundant and ubiquitous but also they are involved in cycling essential nutrients and energy for upper-trophic level organisms as well as removing harmful toxins and pollutants from ecosystems. Although much remains to

be understood, some clear patterns of microbial biogeography are beginning to unfold. In the following introduction, I discuss our current state of knowledge of bacterial taxa richness (alpha diversity) and community composition (beta diversity).

Understanding patterns of bacterial community diversity and composition associated with environmental gradients is crucial. Once a knowledge base exists of patterns of bacterial communities across environmental gradients, we will then be able to better understand the processes underlying microbial biogeography and potentially formulate predictions of how microbial communities might respond to changing environmental conditions. It is clear that we are beginning to understand some of the patterns of diversity across space and time; however, a strong need exists for further characterizing microbial communities across ecologically important gradients such as depleted oxygen concentrations in aquatic ecosystems. Low oxygen zones continue to rapidly expand globally threatening ecologically and economically important aquatic habitats and the organisms within (Diaz & Rosenberg 2008). Though upper trophic level organisms are excluded from these low oxygen environments, microbes continue to thrive and cycle energy and nutrients. Thus, understanding the changes in microbial communities associated with gradients of environmental factors in low-oxygen zones is crucial to be able to predict and potentially rehabilitate these deteriorating ecosystems. In Chapter One, I examine changes in bacterial communities in Hood Canal, WA, which is an ideal natural laboratory to examine how bacterial communities shift across space, time, and environmental gradients. Particularly, I discuss the relationship between changes in bacterial communities with decreasing dissolved oxygen, which is a point of concern in Hood Canal where hypoxia conditions commonly occur. Appendix A contains an expansion on the selection of bacterial community similarity indices. Multiple biases exist in quantifying differences in bacterial community

composition, and, in Appendix A, I examine the impact of these biases on numerous commonly used community similarity indices.

Part I: Bacterial alpha and beta diversity in aquatic systems

In order to compare communities to one another it is necessary to quantify various aspects of the community. These quantified properties include measures of alpha diversity (the number of taxa present in a single community or location), beta diversity (variation in the composition of communities), and gamma diversity (the total diversity across a landscape) (Magurran 2003). In bacterial communities, alpha and beta diversity are two measures commonly used to describe, compare, and examine changes in community structure across environmental gradients, space, and time. Four processes—selection, drift, dispersal, and mutation—structure microbial communities, and characterizing patterns of microbial community richness and composition across space and time can provide insight into these processes that underlie them (Hanson et al. 2012).

A. Bacterial alpha diversity

Maintaining bacterial alpha diversity in natural and managed ecosystems is crucial for both function and stability of ecosystem processes (Bell et al. 2005; Eisenhauer et al. 2012). A decrease in bacterial alpha diversity would likely result in the loss of crucial ecosystem services. Due to the importance of bacterial alpha diversity, microbial ecologists have been interested in describing patterns of bacterial richness across space and time (Shaw et al. 2008) and uncovering the processes leading to maintenance and generation of bacterial diversity (Fierer & Lennon 2011). Challenges of accurately quantifying bacterial diversity exist, however, because bacteria

are so miniscule in size yet incredibly abundant in nature (Hughes et al. 2001; Chao et al. 2006) (for further discussion on these methodological challenges, see Appendix A).

Bacterial alpha diversity has been shown to change across environmental gradients, including across gradients of primary productivity (Horner-Devine et al. 2003). Interestingly, the diversity-productivity relationship is not consistent for all bacterial taxa, which suggests that patterns of alpha diversity may not be universal for all taxonomic groups of bacteria. In the marine environment, studies have shown that bacterial richness is associated with microscale habitat heterogeneity, which is likely a response to the increased number of niches provided by patchiness of available nutrients and resources (Long & Azam 2001). Further investigations into the influence of heterogeneity on larger spatial and temporal scales in marine environments could provide support for the processes generating and maintaining bacterial alpha diversity in these ecosystems.

Patterns of bacterial richness within individual habitats have been frequently studied; however, broad distributions of bacterial taxa across Earth's major habitat types have only recently been described with the advances of molecular phylogenetic approaches (Nemergut et al. 2011; Lozupone & Knight 2007). Though soils are often regarded as being one of the most diverse microbial environments on Earth, a global study revealed that soils had lower bacterial alpha diversity than expected and that saline sediments, or sediments beneath diverse saline environments including sea ice, marine, coastal, thermal features, and meromictic lakes, were sites of highest bacterial alpha diversity (Lozupone & Knight 2007). This high diversity can be explained by the complex and fine-scale chemical gradients that establish in saline sediments, which provide exceedingly high numbers of ecological niches thus supporting high bacterial diversity (Torsvik et al. 2002). Compared to the sediment habitats, aquatic habitats are typically

hypothesized to host fewer bacterial taxa likely as a result of less heterogeneity than in sediments (Curtis & Sloan 2004; Torsvik et al. 2002; Horner-Devine et al. 2004).

The processes generating and maintaining bacterial alpha diversity are emerging as patterns of diversity are being uncovered (Fierer & Lennon 2011). Given the relatively short generation time and rapid rates of mutation and lateral gene transfer, evolution of bacterial taxa has the potential to give rise to bacterial taxa with novel phenotypic or ecological functions thus increasing biodiversity (Kassen & Rainey 2004). Similar to the seed bank hypothesis for plants, Jones and Lennon (2010) hypothesize that some bacteria are able to remain dormant within their environment until conditions are ripe for the proliferation and success. Because assessments of bacterial diversity are typically based on DNA sequences, the capture of dormant bacterial taxa can dramatically increase measured taxa richness in an environment. Methodological advances in the detection and quantification of bacterial diversity will be necessary to entirely understand the underlying processes supporting and maintaining diversity of bacteria.

B. Bacterial beta diversity

It was long thought that because bacteria are microscopic and easily dispersed, that geographic separation could not influence diversity but rather environmental filtering structured bacterial communities (Baas-Becking). However, patterns of bacterial community composition across space and time have emerged and it is evident that taxa are distributed non-randomly across ecosystems (Horner-Devine et al. 2004). Characterizing bacterial community composition and drivers of change in beta diversity has provided insight into some process that lead to assembly and structure. Furthermore, consideration of spatial and temporal scales in measuring and comparing beta diversity is crucial, as these scales are known to affect ecological studies in macroorganisms as well as bacteria (Levin 1992; Martiny et al. 2011; Hatosy et al.

2013). In order to predict changes in bacterial community structure, elucidating the processes that underlie structure and assembly of the communities are essential and consideration to the spatial and temporal scales are likely important in clearly understanding these mechanisms.

Selection of superior organisms best fit for a habitat and its conditions has long been accepted as one fundamental mechanism of variation in community composition. Likewise, environmental gradients in aquatic ecosystems drive the turnover of bacterial species across space and time (Langenheder & Ragnarsson 2007). Vertical depth profiles in the open ocean show that bacterial beta diversity is driven by temperature, light attenuation, and the availability of organic matter as these factors change from sea surface to the deep ocean (Giovannoni & Stingl 2005; DeLong et al. 2006). Across horizontal space, changes in bacterial diversity have been shown to correlate with a temperature gradient from the tropics to polar latitudes (Fuhrman et al. 2008).

However, selection of bacterial taxa along environmental gradients alone cannot completely explain variation in bacterial communities, as evidence for both cosmopolitan (widely distributed) taxa and provincial (narrowly distributed) taxa exists. The presence of cosmopolitan taxa, such as SAR11 in the marine photic zone, demonstrates that some taxa have the ability to escape selection pressures placed by environmental gradients at this taxonomic scale. Further, uncountable bacterial taxa are detected only once in global studies, often referred to as "singletons." It should also be noted that it is possible and likely that current methodology limits the detection of these provincial taxa (see section below for further discussion).

Mechanisms of community assembly and structure have been proposed to support the patterns of bacterial community composition recently uncovered (Hanson et al. 2012). Aside from selection, historical events such as mutation, genetic drift, and migration occurring within a

habitat can help to explain the unaccounted variation in bacterial communities. Stochastic genetic mutations within populations can become fixed and result in genotypic variations across communities. Genetic drift occurs via births and deaths in a population and is also known to affect the composition of bacterial communities through grazing and predation (Jürgens & Matz 2002) as well as viral infection of select bacterial populations (Fuhrman & Schwalbach 2003). Finally, migration of bacterial taxa is thought to be a strong contributor to community composition as a result of the ease of dispersal of the microscopic organisms over long distances. The role of these stochastic processes support the neutral community model and likely affect bacterial communities at the local scale (Curtis & Sloan 2004).

The spatial scale at which bacterial communities are examined has the potential to influence the interpretation of the relative contribution of selection, mutation, drift, or migration to the community composition (Martiny et al. 2011). In terrestrial systems, spatial scale is easily quantified using geographic distance. However, in aquatic ecosystems hydrologic regimes can quickly move water parcels thus transporting bacteria and other planktonic organisms with the water mass. A pyrosequencing study comparing the composition of bacterial communities at three depths throughout the water column at multiple stations in the North Atlantic Ocean reported that bacterial community composition was most similar within the same water mass, defined by depth, regardless of horizontal space (Agogué et al. 2011). A recent study of global marine bacterial beta diversity showed that in pelagic waters geographic distance between communities accounted for less than 10% of the total variation in community composition that could be explain by measured variables (Zinger et al. 2011). When comparing benthic marine communities, Zinger and colleagues found that over 27% of the total explained variation in bacterial community composition could be attributed to geographic distance. These results may

suggest that while local factors may strongly influence benthic or terrestrial bacterial communities, the pelagic communities may experience increased mixing or migration, which decreases the importance of geographic distance on bacterial beta diversity.

Long term monitoring of marine ecosystems has provided the data necessary for showing that, in addition to spatial scale, bacterial communities also change over time (Gilbert et al. 2012; Fuhrman et al. 2006; Bouskill et al. 2011; Newton et al. 2011), which may also affect the interpretation of the relative contribution of the underlying processes of bacterial community composition. These changes in beta diversity over time are typically attributed to environmental variation as a result of seasonal or annual changes within the water mass. Furthermore, the results from a recent study by Hatosy et al (2013) suggest that temporal beta diversity is dependent upon the scale at which the communities are examined. The authors conclude that the rate of change of beta diversity is dependent on temporal scale with the highest rate of change occurring at the shortest time scale. They suggest that bacterial communities quickly track environmental variation. Additionally, it is possible that at different temporal scales distinct stochastic events further drive changes in beta diversity. For example, on short time scales, ecological drift arising from stochastic events such as births and deaths contribute to heterogeneity in diversity. At longer time scales, stochastic mutation genetic processes can lead to evolutionary drift (Martiny et al. 2011).

In chapter one, I uncover patterns of bacterial richness and community composition along an ecologically and economically important gradient of dissolved oxygen in Hood Canal, WA. These patterns are consistent along the dissolved oxygen gradient even though multiple other environmental variables were simultaneously measured. Tracking changes in bacterial communities across ecologically important gradients will help scientists to further understand the

processes structuring bacterial communities and potentially aid in prediction of changing ecosystems functions and health in the face of a rapidly changing planet.

Part II: Challenges in describing bacterial diversity

As is true for the quantification of many ecological properties, numerous indices exist for alpha and beta diversity, and the underlying assumptions and calculations for each index may not be appropriate for all domains of life (Hughes et al. 2001). These various indices attempt to quantify the diversity within (alpha diversity) and between (beta diversity) communities; however, it is infeasible to exhaustively sample most communities thus making diversity indices mere estimates of actual community diversity (Chao et al. 2006; Colwell & Coddington 1994). Special consideration to indices used to quantify and compare diversity in hyper-diverse communities, such as bacterial communities, must be made in order to avoid false conclusions regarding the patterns and processes underlying bacterial biodiversity (Chao et al. 2006). Specifically for bacteria, much higher abundances (Whitman et al. 1998) and number of taxa than macroorganisms and whether or not bacteria follow similar patterns of community assembly and structure (Haegeman et al. 2013) may influence indices of diversity, though it remains somewhat unclear how the indices are affected and whether it matters (Lozupone et al. 2007; Haegeman et al. 2013; Chao et al. 2006). Below, I briefly discuss our current understanding of beta diversity indices used to quantify and compare bacterial community composition and suggest the need for further research into assumptions made and biases influencing the calculation of these indices.

Though undersampling is common throughout community ecology, some properties of bacterial communities give rise to specific properties and biases that need to be addressed. Next-

generation sequencing and improved clustering and error-proofing methods have aided in more complete sampling of bacterial communities and better detection of low abundance taxa (Sogin et al. 2006; Huse et al. 2010). Even with these advances, detection of every bacterial taxon present in a sample is not yet feasible. Furthermore, biases introduced in the sequencing methods can inflate the proportion of abundant taxa relative to the lower-abundance types (Pinto & Raskin 2012). For example, the chances of detecting abundant taxa in the PCR-step of most DNA sequencing methods are much higher than detection of low abundance, or rare taxa. Therefore, abundant taxa are better represented than low abundance types. This bias might artificially lower the observed taxonomic richness of highly uneven communities as the abundant taxa are overly detected relative to low-abundant types.

Undersampling can also affect the patterns in bacterial community composition. The problem of undetected taxa in a sample can mean that the taxon is actually absent from the sample or that the taxon was present but not detected. To compensate for the likely detection bias in current sampling and sequencing methods, the use of community similarity indices, such as the Chao or Sorensen index, use the probability of undetected taxa and under sampling to estimate the compositional similarity; however, poorly sampled communities can inflate community dissimilarity (Chao et al. 2006). It is clear that a need exists to explore the influences of these biases on commonly-used bacterial beta diversity indices.

Conclusion

In what follows, I address two knowledge gaps in microbial ecology. First, our understanding of spatial and temporal distributions of bacteria in disturbed ecosystems lags that for macroorganisms. I use next-generation DNA sequencing technology deeply characterize

patterns of bacterial richness and composition across space and time in a seasonally oxygen-deficient estuary, Hood Canal, WA. The results of this study suggest that bacterial community composition shifts as dissolved oxygen levels drop suggesting changes in nutrient cycling processes. This study lays the groundwork for further investigation into the functional response of microbial communities to rapidly depleting dissolved oxygen. As low oxygen zones expand globally, it will be increasingly important to understand the potential changes in microbial community diversity and function as energy is shifted from upper trophic levels to microbes in these oxygen-depleted waters (Diaz & Rosenberg 2008).

Secondly, with rapidly evolving sequencing technology and computational power, it is crucial for microbial ecologists to understand the impacts of various assumptions being made regarding bacterial community properties in order to accurately describe bacterial community composition. To address this issue, I used large datasets of bacterial community data to explore numerous indices currently used to characterize bacterial community composition in order to determine the most appropriate and robust descriptors in the face of multiple challenges that remain in describing bacterial beta-diversity.

References

Agogué H, Lamy D, Neal PR, Sogin ML, Herndl GJ. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. Molecular ecology 20:258–74.

Anton J, Rosello-Mora R, Rodriques-Valera F, Amann R. (2000). Extremely Halophilic Bacteria in Crystallizer Ponds from Solar Salterns. Applied and environmental microbiology 66:3052–3057.

Baas-Becking L. Geobiologie of Inleiding Tot de Milieukunde. In:The Hague, Stockkum, V & Zoon, eds (ed)., The Netherlands.

Baker BJ, Banfield JF. (2003). Microbial communities in acid mine drainage. FEMS microbiology ecology 44:139–52.

Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. (2005). The contribution of species richness and composition to bacterial services. Nature 436:1157–60.

Bouskill NJ, Eveillard D, O'Mullan G, Jackson GA, Ward BB. (2011). Seasonal and annual reoccurrence in betaproteobacterial ammonia-oxidizing bacterial population structure. Environmental microbiology 13:872–86.

Boyd ES, Lange RK, Mitchell AC, Havig JR, Hamilton TL, Lafrenière MJ, et al. (2011). Diversity, abundance, and potential activity of nitrifying and nitrate-reducing microbial assemblages in a subglacial ecosystem. Applied and environmental microbiology 77:4778–87.

Brock T. (1978). Thermophilic microorganisms and life at high temperatures. Springer: New York.

Chao A, Chazdon RL, Colwell RK, Shen T-J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361–71.

Colwell RK, Coddington JA. (1994). Estimating terrestrial biodiversity through extrapolation. Philosophical transactions of the Royal Society 345:101–118.

Curtis TP, Sloan WT. (2004). Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. Current opinion in microbiology 7:221–6.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. Science (New York, N.Y.) 311:496–503.

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

Eisenhauer N, Scheu S, Jousset A. (2012). Bacterial diversity stabilizes community productivity. PloS one 7:e34517.

Fierer N, Lennon JT. (2011). The generation and maintenance of diversity in microbial communities. American journal of botany 98:439–48.

Fuhrman JA, Schwalbach M. (2003). Viral influence on aquatic bacterial communities. The Biological bulletin 204:192–5.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown M V, 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, Hewson I, Schwalbach MS, Brown M V, Green JL, et al. (2008). A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences 105:7774–7778.

Gilbert J, Steele J, Caporaso JG, Steinbrück L, Reeder J, Temperton B, et al. (2012). Defining seasonal marine microbial community dynamics. The ISME Journal 6:298–308.

Giovannoni SJ, Stingl U. (2005). Molecular diversity and ecology of microbial plankton. Nature 437:343–8.

Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. (2013). Robust estimation of microbial diversity in theory and in practice. The ISME journal 1092–1101.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. Nature reviews. Microbiology 1–10.

Hatosy S, Martiny JBH, Sachdeva R, Steele J, Fuhrman JA, Martiny AC. (2013). Beta-diversity of marine bacteria depends on temporal scale. Ecology.

Horner-Devine MC, Leibold MA, Smith VH, Bohannan BJM. (2003). Bacterial diversity patterns along a gradient of primary productivity. Ecology Letters 6:613–622.

Horner-Devine MC, Carney KM, Bohannan BJM. (2004). An ecological perspective on bacterial biodiversity. Proceedings of the Royal Society of London, Biological Sciences 271:113–22.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM. (2004). A taxa-area relationship for bacteria. Nature 432:750–753.

Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJM, Colwell RK, Fuhrman JA, et al. (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. Ecology 88:1345–53.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. (2001). Counting the Uncountable□: Statistical Approaches to Estimating Microbial Diversity. Applied and environmental microbiology 67:4399–4406.

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

Jones SE, Lennon JT. (2010). Dormancy contributes to the maintenance of microbial diversity. Proceedings of the National Academy of Sciences of the United States of America 107:5881–6.

Jürgens K, Matz C. (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. Antonie van Leeuwenhoek 81:413–34.

Kassen R, Rainey PB. (2004). The ecology and genetics of microbial diversity. Annual review of microbiology 58:207–31.

Langenheder S, Ragnarsson H. (2007). The role of environmental and spatial factors for the composition of aquatic bacterial communities. Ecology 88:2154–61.

Levin SA. (1992). The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Long RA, Azam F. (2001). Microscale patchiness of bacterioplankton assemblage richness in seawater. Aquatic Microbial Ecology 26:103–113.

Lozupone CA, Hamady M, Kelley ST, Knight R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology 73:1576–85.

Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences of the United States of America 104:11436–40.

Magurran A. (2003). Measuring Biological Diversity. Blackwell Publishing: Malden, MA.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. (2006). Microbial biogeography: putting microorganisms on the map. Nature reviews. Microbiology 4:102–12.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108:7850–7854.

McCollom TM, Shock EL. (1997). Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. Geochimica et cosmochimica acta 61:4375–91.

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. Microbiology and molecular biology reviews : MMBR 75:14–49.

Pinto AJ, Raskin L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PloS one 7:e43093.

Rosenzweig M. (1995). Species diversity in space and time. Cambridge University Press: New York.

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 10:2200–10.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” Proceedings of the National Academy of Sciences 103:12115–20.

Tilman D, Knops J, Wedin D, Reich P, Ritchie M, Siemann E. (1997). The Influence of Functional Diversity and Composition on Ecosystem Processes. Science 277:1300–1302.

Torsvik V, Øvreås L, Thingstad TF. (2002). Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science (New York, N.Y.) 296:1064–6.

White DC, Phelps TJ, Onstott TC. (1998). What's up down there? Current opinion in microbiology 1:286–90.

Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes□: The unseen majority. Proceedings of the National Academy of Sciences 95:6578–6583.

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

Chapter 1: Patterns of bacterial richness and community composition in Hood Canal, WA, a seasonally oxygen-deficient estuary

Abstract

When ocean waters are depleted of dissolved oxygen, energy shifts from macroorganisms to microorganisms, which dominate in these hypoxic zones. Patterns of microbial communities associated with dissolved oxygen gradients are thus of particular interest as coastal oxygen minimum zones continue to expand. We examined factors associated with the distribution of bacteria in Hood Canal, WA - a glacial fjord-like water body that experiences low dissolved oxygen levels detrimental to fish and other marine organisms. Pyrosequencing techniques were applied to sixteen surface and deep (near bottom) water samples from a north-south transect along Hood Canal collected at three different time points from Spring to Autumn 2007. We obtained a total of 165 022 bacterial 16S-rRNA gene sequences and identified 3 807 unique operational taxonomic units based on 97% sequence identity, across all samples. We found that bacterial richness and evenness are significantly higher in deep waters than in surface waters, and richness values show a strong association with dissolved oxygen content. Bacterial community composition grouped strongly by sample water depth, and surface water bacterial communities exhibited a weak but discernable seasonal shift. Dissolved oxygen was most correlated with observed variation in community composition. Our results suggest that bacterial communities vary throughout Hood Canal and, specifically, that changes in the bacterial community are highly associated with levels and variability in dissolved oxygen concentration. These patterns are remarkably similar to patterns in bacterial communities in oxygen minimum zones across the globe, supporting recent findings that common taxa are widely distributed among these similar environments.

Introduction

Globally, over a quarter-million square kilometers of marine ecosystems are threatened by low dissolved oxygen (DO) levels, or hypoxia, which can result in the exclusion or death of resident macroorganisms creating so-called ‘dead-zones’ (Diaz & Rosenberg 2008). Indeed, over the past half century, the extent of dead zones has grown at an alarming rate as a result of eutrophication and climate change. The hypoxic waters in these dead zones may exclude fish and benthic organisms and eliminate habitat that is important both for commercial fishing and recreation (Newton et al. 2007); however, bacterial and archaeal populations remain active in mineralization of organic matter and other biogeochemical cycles and are, therefore, important players in the functioning of this oxygen-deficient ecosystem (Diaz & Rosenberg 2008; Urakawa et al. 2010).

Uncovering patterns of microbial communities across space and time is crucial for our understanding of ecosystem function, yet patterns of microbial diversity and function remain understudied relative to those of macro-organisms, especially in reference to dissolved oxygen gradients. Abiotic factors, such as temperature, pH, carbon and nutrient availability, and sunlight have been shown to play important roles in structuring marine microbial communities due to the diverse multitude of niches occupied by various groups of microbial taxa (Fuhrman et al. 2008; Koskinen et al. 2011; Gilbert et al. 2012). Of particular concern in coastal waters, decreased DO concentration has both direct as well as indirect effects on bacterial community structure and function, which in turn can impact energy flow and food web structure as well as global greenhouse gas cycling (Rabalais et al. 2010). For example, nitrogen cycling processes such as denitrification and anammox have been examined extensively in oxygen-deficient water bodies because as biologically available nitrogen (or fixed nitrogen) is consumed, nitrous oxide (N_2O),

an atmospherically active trace gas, is released. In fact, nearly 50% of fixed-N is released as N_2O from OMZs globally (Codispoti et al. 2001).

Recent surveys of microbial communities in the OMZs across the globe including the eastern tropical South Pacific (Stevens & Ulloa 2008), Arabian Sea (Fuchs et al. 2005; Zubkov et al. 2006), Cariaco Basin (Madrid et al. 2001; Lin et al. 2006), Baltic Sea (Labrenz et al. 2007), Black Sea (Vetriani et al. 2003; Wakeham et al. 2007; Fuchsman et al. 2011), and Saanich Inlet (Zaikova et al. 2010) reveal consistent patterns of increased microbial diversity and distinct shifts in community composition associated with lowered DO. Due to the growing prevalence of OMZs, a recent meta-analysis (Wright et al. 2012) examined global patterns of microbial community composition in OMZs found in the open ocean, coastal areas, and enclosed basins. They concluded that particular bacterial and archaeal taxa are common in suboxic waters and that these taxa tend to co-occur in a non-random pattern associated with particular levels of DO (Wright et al. 2012). *Proteobacteria*, *Bacteroidetes*, marine group A, *Actinobacteria*, and *Firmicutes* comprise the majority of taxa found in oxygen-deficient waters, and these dominant taxa tend to be distributed in regular patterns throughout the water column along the oxycline (Wright et al. 2012).

To further understand global patterns and the functional response of microbial communities to expanding oxygen minimum zones, we examined spatial and temporal variation in bacterial communities in Hood Canal, WA between April and October of 2007. Hood Canal, a long and narrow glacial fjord located 80 miles west of Seattle, WA, seasonally experiences periods of low and even completely depleted DO concentrations as a result of naturally occurring physical and hydrographical conditions (Newton et al. 2007). This system is unique compared to other OMZs because of its relatively small-scale, enclosed circulation patterns, and rapidly

changing abiotic conditions, and therefore, interesting to compare to patterns found globally in other oxygen-depleted environments. Throughout the year, the long narrow fjord remains highly stratified and has little internal circulation, which limits the movement of oxygenated waters (Gregg & Pratt 2010). Furthermore, human activity over the past few decades has exacerbated nutrient loading particularly in the southern reaches of Hood Canal (Steinberg et al. 2010), in turn increasing the frequency and duration of recent hypoxia events (Newton et al. 2009). We thus examined how environmental parameters change over space and time in Hood Canal and examined the bacterial communities associated with these biotic and abiotic factors, including depleted DO concentration. We hypothesized that bacterial communities demonstrate non-random changes in diversity, structure, and composition related to variation in abiotic factors, and these community patterns are associated particularly with changing levels of DO experienced in Hood Canal.

Methods

Sample collection.

Hood Canal, WA is a narrow, glacial-fjord west of Seattle (Figure 1A). High resolution depth profiles for dissolved oxygen and chlorophyll a for two mid-canal stations, Hama Hama and Sister's Point, were collected via Oceanic Remote Chemical Analyzer (ORCA) buoys that are maintained by the Northwest Association of Networked Ocean Observing Systems (http://www.nanoos.org) (Devol et al. 2007). Water samples were collected during April, June, and October of 2007 using a Niskin rosette while simultaneously measuring physical and chemical water properties with a CTD (conductivity, temperature, and depth) sensor. April and October samples were collected from two stations near the middle of Hood Canal: Hama Hama

and Sister's Point. During the June sample collection, two additional stations (Bangor and Lynch Cove) were added in order to survey a full north to south transect of the Canal. At each sampling station and time, samples were collected both at 5 meters depth (herein referred to as surface samples) as well as within 10 meters of the bottom substrate (herein referred to as deep samples). These so-called deep samples varied in depth from 145 m at Hama Hama to 12 m at Lynch Cove as the depth of the water column decreases near the southern reaches of the canal. The timing of the collection points were intended to coincide with the late spring algal bloom (April), subsequent die-off (June), and a smaller autumn algal bloom (October) which directly impact dissolved oxygen concentration and nutrient availability.

Water samples were divided up to measure various biotic and abiotic parameters. Dissolved oxygen was measured using the Winkler titration method. To measure inorganic nutrients, 50 ml of each water sample were passed through a 0.45-micron filter and immediately frozen until being analyzed at the University of Washington's Oceanography Marine Chemistry Lab for nitrate, nitrite, ammonium, phosphate, silicate, and salinity. To measure bacterial abundance, 5ml of formalin-preserved water from each sample were diluted with sterile water and 0.5-1.0 ml aliquots were filtered onto 0.22um polycarbonate filters and stained with 4',6-diamidino-2-phenylindole (DAPI) and viewed under a Nikon Eclipse 80i with UV light from an X-cite Series EXFO to count the number of cells present in fifteen randomly selected fields using NIS-Elements BR 3.0 software then averaged to quantify the total bacterial abundance. Statistical analyses of environmental factors were tested using principal components analyses conducted in *R* using the *vegan* package.

DNA extraction and processing.

Following water sample collection, 300 ml of each water sample were filtered onto 0.22 um Supor filters and preserved with 3 ml of NaCl/Tris/EDTA solution then immediately frozen at -80°C until being processed for bacterial characterization. Whole DNA was extracted in duplicate from filters collected from each sample using Qiagen DNeasy Tissue Kit following the manufacturer's instructions for extraction from gram-positive bacteria (Qiagen Valencia, CA). The DNA extracts were prepped and sent to MBL's Keck Facility for PCR amplification and sequencing on GS-FLX Titanium 454 platform. The V6 region of the 16S rRNA was amplified via polymerase chain reaction (PCR) using bacterial specific primers (see Sogin et al. 2006 for previously described methods), excluding amplification of archaea. Sequencing was performed in conjunction with the International Census of Marine Microbes (ICoMM) project (http://icomm.mbl.edu), which uses the same methodology to sequence marine samples from across the globe.

Data analysis.

As part of the ICoMM project, the sequence data were cleaned and initially analyzed using a pipeline that effectively cleans, clusters, and assigns taxonomy to the sequences (see Sogin et al. 2006, Huse et al. 2007, 2010). Through this pipeline, sequences are pre-clustered using a single-linkage algorithm to smooth sequencing errors and reduce noise in the data. Sequences are then clustered using average-linkage into operational taxonomic units (OTUs) based on 97% sequence identity. Prior to community analyses, sequences that were highly similar to chloroplast sequences were removed from the dataset.

Due to the 454 methodology, the number of sequences obtained from each sample varies, which can affect the observed and estimated richness values because the chance of detecting more or less abundant taxa can increase when more sequences are captured. To minimize the

impact of sequencing effort on richness estimates, all samples were randomly subsampled down to the number of sequences of the sample with the fewest number of sequences, the richness measures were calculated, then the process was repeated and averaged over 1000 iterations. Observed taxa richness, or the number of OTUs identified at the pre-defined sequence identity, as well as two richness estimators ACE and Chao were calculated using the subsampling method and averaged over 1000 iterations. Two measures of diversity, Shannon and Simpson, were also calculated on the full dataset using the subsampling method and averaged over 1000 iterations. To examine how the taxa richness of bacterial communities in Hood Canal compared to that in other aquatic environments, we obtained sequence data from other pyrosequenced marine samples from the ICoMM project. Observed bacterial richness in those samples was calculated using the same methods as Hood Canal samples.

Because bias is introduced during the PCR step of pyrosequencing, using the relative abundances of OTUs in measures of beta diversity is preferred over traditional presence-absence measures (Pinto & Raskin 2012). Therefore, we assessed how bacterial community composition, as measured by relative abundance of taxa, varies by environmental factors. Numerous measures of dissimilarity exist and their performances in undersampled conditions or uneven sample sizes have yet to be understood for bacterial communities. Prior to examining changes in Hood Canal bacterial community composition, we first assessed the performance of six similarity indices commonly used in microbial studies under common sampling biases: undersampling, uneven sampling, and differing taxa evenness (see Appendix A for details). Similarly to the alpha diversity calculations, the bacterial community matrix was randomly subsampled to the number of sequences of the smallest sample then the Sorensen abundance-based similarity was calculated between samples to account for sequencing effort bias (Horner-Devine et al. 2004, Martiny et al.

2011). The subsampling and similarity calculation was repeated 1000 times and the averaged values were used for subsequent analysis. The Sorensen abundance-based estimator outperformed the other similarity indices we assessed, and thus we used this similarity index in subsequent analyses.

Analysis of similarity (ANOSIM) tests were used to test the effects of depth, site, or season on bacterial community composition of each sample and significance was tested using 999 random permutations of the community composition dataset. A non-parametric iterative approach was employed to select the set of abiotic variables were most correlated with changes in bacterial community composition (Clarke & Ainsworth 1993). Calculations of alpha and beta diversity, ANOVA, NMDS ordinations, and ANOSIM were conducted in *R* using *vegan* package.

Results

Environmental context.

In the spring of 2007 samples were collected from Hama Hama and Sister's Point (Figure 1A) directly followed the major spring phytoplankton bloom as indicated by the heightened level of chlorophyll in the surface water layer at these two stations (Figure 1B). In June 2007 all four stations were sampled when much of the phytoplankton bloom had died off, and DO levels were beginning to drop in the mid-reaches of the canal. The middle two stations were sampled again in October following a small autumnal phytoplankton bloom, indicated by the slightly heightened chlorophyll concentration in the surface waters at Sister's Point (Figure 1B). Because of the long residence time of water in Hood Canal, DO level south of Hama Hama tends to decrease more rapidly than the northern reaches of the Canal where occasional fresh,

oxygenated water enters from Puget Sound over a shallow sill (Gregg & Pratt 2010). Major flushing and recirculating events in Hood canal typically occur in autumn but can be variable in terms of timing as a result of upwelling events off the coast of Washington (Warner et al. 2000).

Using the variables measured at the time of sample collection, a principal components analysis (PCA) was conducted to visualize abiotic factors driving the differences between samples (Figure 2). Axis one explained 57% of the total measured variation in abiotic variables and appeared to separate surface water samples from deep water samples (Figure 2). Nitrate, DO, salinity, and phosphate were most correlated with axis one and thus were likely driving the separation between the two depths. An additional 19% of the total variation in abiotic parameters was explained by axis two, which was most correlated with ammonium concentration. In addition to separating samples by water depth, abiotic factors were associated with significant separation of samples by season. Surface samples separated by season generally along axis one, while deep water samples separated by season along axis two. During the April sample collection, a significant phytoplankton bloom was beginning in the Southern end of the canal. This event was marked by high DO concentrations (12.32-13.18 mg/L) and extremely low nutrient concentrations in the surface waters (NO_3^-, NO_2^-, NH_4^+, PO_4^{3-}, SiO_4), whereas deep waters were characterized by very low DO concentrations (3.59-3.94 mg/L) and high nutrient concentrations. June samples were collected at the end of the phytoplankton bloom, and thus the surface waters were characterized by mid-range DO (8.66-10.25 mg/L) and nutrient concentrations. At the same time, the deep waters contained low dissolved oxygen (2.95-7.12 mg/L) and high nutrient concentrations. October samples were collected after a lesser fall phytoplankton bloom. Dissolved oxygen, nutrient concentrations, and dissolved silicate are in

the mid to high range in both water depths, which appear to be less stratified than earlier in the year (SI Figure 1, SI Table 1).

Bacterial alpha diversity.

Across all samples, over 57% of the sequences clustered into only 42 unique OTUs (defined by 97% sequence similarity), while less than 2% of the sequences clustered into over 2300 unique OTUs. The abundant OTUs were typically widespread across all the samples, whereas each of the less abundant OTUs was detected in one or two samples. Uneven distribution of individuals into taxa is not unusual in bacterial communities (Sogin et al. 2006). When sequences were grouped by class, the relative proportion of the sample comprised by a given class varied significantly by sampling depth (G-test; G=37.585, p=0.0006), but not by season or horizontal space (Figure 3). The relative proportions of dominant groups appear to change across a gradient of DO, and relative proportions of the dominant groups are more even in lower DO environments.

Of these fifteen bacterial classes, the majority of sequences recovered from surface waters belonged to three classes: *Flavobacteria* (40.53%±11.32%), *Alphaproteobacteria* (28.02%±8.27%), and *Gammaproteobacteria* (19.04%±5.56%). Similarly, in deep waters the highest proportion of sequences fell into the *Alphaproteobacteria* (31.77%±5.96%), *Gammaproteobacteria* (20.67%±3.58%), and *Flavobacteria* classes (16.22%±9.58%). The proportions of *Alphaproteobacteria* and *Gammaproteobacteria* were not significantly different between surface and deep waters; however, the proportion of *Flavobacteria* sequences was considerably higher in the surface water compared to deep ($F_{1,14}$=21.463; p=0.0004). Though the same three classes of bacterial sequences dominate in both water layers, in surface waters these

classes comprised 88.5% of sequences, whereas in deep waters they comprised only 68.7% of sequences.

Observed bacterial taxa richness, or the number of OTUs recovered, in Hood Canal was significantly higher in the deep water mass than in the surface water by an average factor of two ($F_{1,14}$=22.55; p=0.0003). Neither station (horizontal space) nor season had a significant effect on bacterial richness (results not shown). Of the eight abiotic variables measured in this study, DO concentration alone was most correlated with variation in observed bacterial richness across samples (adjusted R^2=0.7379, p<0.0001). Bacterial richness peaked at DO concentration of approximately 4 mg/L (Figure 4). When comparing Hood Canal bacterial richness estimates to environments around the globe, we found that samples from Cariaco Basin (an anoxic basin off the northern central coast of Venezuela), the Gulf of Aqaba, and Blanes Bay had estimates of bacterial richness very similar to those of Hood Canal. Compared to the open ocean waters of the Pacific Ocean, Hood Canal samples demonstrated considerably higher bacterial richness. However, compared to samples from the northeastern coast of the Atlantic Ocean, Hood Canal richness appeared to be lower, though not significantly (Figure 5).

The relative abundance of each taxon in a community provides an additional window into the diversity of a community; therefore, we examined taxa evenness in Hood Canal samples by calculating Pielou's J. Evenness of taxa was higher in deep waters than in the surface water ($F_{1,14}$=9.517, p=0.008071), while neither season nor sampling site had an effect on the evenness of the bacterial community (results not shown). Of all possible combinations of abiotic factors, increasing phosphate concentration alone was most associated with increasing evenness of bacterial communities ($F_{1,14}$=17.52, p=0.0009151).

Bacterial beta diversity

There was no significant effect of sampling station on the composition of bacterial communities in Hood Canal (Figure 6, NMDS, stress=0.0725, ANOSIM R=-0.1456, p=0.9148). In contrast, there were clear differences in community composition between surface and deep-water communities (Figure 6, ANOSIM: R_{depth}=0.649, p_{depth}=0.007) and among seasons (R_{season}=0.519, p_{season}=0.003). A post-hoc test showed that the surface water communities were most influenced by a seasonal effect (R=0.8625, p=0.0051), while deep-water communities did not change significantly across seasons (R=0.0375, p=0.3874). Of all possible combinations of the eight abiotic variables, DO concentration alone was most correlated with dissimilarities in the composition of bacterial communities among samples (Spearman's rho =0.7320).

Patterns of individual classes.

Patterns of taxonomic richness within the three most abundant classes were very similar to those of the full bacterial domain, which is unsurprising given that these three classes comprised the majority of OTUs in the full dataset. Taxonomic richness within each of the three most abundant classes was significantly higher in deep waters than in surface waters and sampling station or season of sampling had no effect on richness for any group (Table 2). As was true for the full bacterial dataset, *Flavoproteobacteria* and *Gammaproteobacteria* richness was negatively associated with DO; however, *Alphaproteobacteria* taxa richness increased significantly as salinity increased but had no significant association with the DO gradient. Abiotic associations with taxa evenness, measured as Pielous' J, also varied across the three dominant classes and showed strong correlations with abiotic factors that were different for each class (Table 1).

Changes in the community composition of the top three bacterial classes were also associated with factors similar to the full bacterial domain. Composition within each of the three

dominant classes differed significantly between surface waters and deep waters and by season (Table 2). Interestingly, although taxa richness these groups did not all respond similarly to DO, changes in community composition were strongly correlated with the DO gradient for all three classes (Table 2). These findings further emphasize the strong role of DO in shaping several aspects of the bacterial communities in Hood Canal, though the specific impact may differ across various taxonomic groups.

In order to evaluate potential functional changes in the community composition, we examined changes in the relative abundances of several groups of bacteria in our dataset known to be important in specific biogeochemical processes. Two groups of ammonia oxidizing bacteria (AOB), *Nitrosomonas* and *Nitrosococcus,* were detected in Hood Canal. The relative abundance of AOB was higher in deep water than surface water ($F_{1,14}$=18.98, p=0.0006573) and also demonstrated a significant negative relationship with dissolved oxygen concentration ($F_{1,14}$=21.42, p=0.0003912). Given that we used bacterial-specific 16S primers, we did not detect ammonia-oxidizing archaea. The relative abundance of two groups of nitrite-oxidizing bacteria (NOB) in Hood Canal, *Nitrospira* and *Nitrospina*, was also significantly higher in the deep water than in the surface water ($F_{1,14}$=24.9, p=0.0002). Additionally, the relative abundance of this group was not only strongly associated with a decrease in DO but was also correlated with decreased NH_4^+ concentration and temperature ($F_{3,12}$=69.04, p=7.721x10^{-08}). There were no significant patterns of increased *Cyanobacteria* abundance by water depth, sampling season, or sample site; however, the relative abundance of this photosynthetic phylum increased with higher temperatures ($F_{1,14}$=112.3, p=4.45x10^{-08}). The relative abundance of methylotrophic bacteria in Hood Canal, including *Methylobacteriaceae*, *Methylococcaceae*, *Methylocystaceae*, and *Methylophilaceae*, also demonstrated no significant patterns across water depths, season, or

sampling site but their abundances did increase with decreasing DO and higher temperature ($F_{2,13}$=12.17, p=0.001049).

Discussion

Hood Canal, WA is an ecologically, economically, and recreationally important ecosystem in Western Washington that seasonally experiences hypoxia and anoxia, which has led to fish kills and substantial changes in ecological communities (Newton et al. 2007). While phytoplankton are considered to be biological drivers for seasonal, episodic hypoxia events as the detrital material sinks to deep waters thus increasing the biological oxygen demand (Newton et al. 2011), bacterial communities are, in fact, directly responsible for the consumption of oxygen and cycling of essential nutrients in Hood Canal, yet are understudied relative to physical, chemical, and planktonic processes (see Gregg & Pratt 2010; Newton et al. 2011; Warner et al. 2000; Steinberg et al. 2010; Keister & Tuttle 2013). To our knowledge, this is the first study describing patterns of bacterial richness and community composition in Hood Canal, WA.

We captured a range of abiotic conditions over a seven-month time period in the sixteen Hood Canal samples collected for this study. This range of abiotic conditions was similar to observations in prior years, although DO did not reach extreme low levels in 2007 as observed in 2004 or 2006 (Newton et al. 2011). Abiotic factors measured in the surface and deep samples differed significantly, indicative of the strong stratification and stability of water masses resulting from the hydrography and hydrology of Hood Canal (Gregg & Pratt 2010). The environmental separation between surface and deep water masses was most pronounced during

April and June; by October the surface and deep water samples were more similar to one another in terms of the abiotic conditions. In autumn, offshore upwelling of more saline water can displace deep waters in Hood Canal, resulting in weaker stratification and increased oxygen levels (Warner et al. 2000).

Bacterial taxonomic richness, evenness, as well as community composition varied significantly between surface waters and deep water masses. Given the dramatic stratification and abiotic differences between the two water masses, we expected that the bacterial communities within each water mass would differ. On a global scale, different physical, chemical, and biological processes in surface waters versus deep waters structure bacterial communities, which gives rise to very distinct bacterial community composition in each water mass (Ghiglione et al. 2012). Indeed, distinction between surface water and deep water bacterial community composition was strongest during sampling points of heightened stratification (April and June), mirroring patterns of abiotic conditions. In October, the bacterial community composition differed between surface and deep water masses, but the degree of separation was much lower. Potential episodic mixing between water masses and deep water renewal prior to October could result in mixing of the bacterial communities, thus decreasing the distinction between surface and deep water masses.

Bacterial communities are distinct between separate water masses in marine environments with bacterial richness and evenness increasing from surface to deeper water layers (Agogué et al. 2011; Friedline et al. 2012; Pommier et al. 2010). Though the water column of Hood Canal is shallow relative to open-ocean studies, we uncovered a similar pattern of higher bacterial species richness and evenness in deep-waters than in surface waters. Surface waters, compared to deep waters, experience increased variability in abiotic factors on daily to seasonal

timescales (Figures 1B and 2), which may limit the number of bacterial taxa that can persist while allowing those well-adapted to fluctuating environments to flourish, thus reducing the overall bacterial richness and evenness. As suggested previously, ecosystems that experience intermediate levels of abiotic variability may allow for higher diversity than those with high variability or disturbance (Pommier et al. 2010; Petraitis et al. 1989; Connell 1978; Hutchinson 1961), which is evident in the deep water samples of this study. Additionally, the sinking of organic matter substrates from surface detritus into deep waters may provide a variety of carbon sources to heterotrophic bacteria (Cho & Azam 1988), giving rise to increased environmental heterogeneity in these zones (Long & Azam 2001). Habitat heterogeneity, or the number of niches in a habitat, has long been suggested as a strong determinant of bacterial community composition and diversity (Martiny et al. 2006; Fierer & Jackson 2006; Horner-Devine et al. 2004; Shade et al. 2008; Tilman 1999), and this relationship is further supported by the community patterns demonstrated in this study. Furthermore, sediment underlying the water column typically houses higher bacterial richness and could be a potential source of bacterial richness for overlying deep water layers (Feng et al. 2009). Finally, it is possible that methodology restricted detection of low abundance types of bacterial taxa. A combination of low evenness and the steep slope of rarefaction curves in surface water samples indicate that few well-adapted taxa exhibit prolific growth and may inhibit our detection of rare taxa (Amend et al. 2010). Whether or not these low-abundance bacterial taxa are active or important players in the bacterial community is unclear (Bowen et al. 2012; Jones & Lennon 2010; Huse et al. 2010).

Within each water mass (surface and deep), the geographic location of sampling or station did not affect the bacterial communities observed in Hood Canal. Water movement and circulation within each water mass of Hood Canal is complex as a result of wind, tides, and

seiches (Newton et al. 2011), which could prohibit local factors at the scale sampled from structuring bacterial communities within each layer. Though we did not detect changes in the bacterial communities across stations, we did detect a weak seasonal effect on the structure and composition of the surface bacterial communities. At latitudes similar to Hood Canal phytoplankton demonstrate strong seasonal cycles (Pinckney et al. 1998). Bacterial communities are also known to show predictable seasonal patterns especially in surface waters where sunlight and external factors, such as the weather, may shape the bacterial communities (Gilbert et al. 2012; Fuhrman et al. 2006). Though we expect bacterial communities in Hood Canal to exhibit similar dynamics, increased sampling frequency is necessary to investigate seasonal cycles.

Across all samples in this study, DO consistently explained the most variation in bacterial richness and composition. As DO concentration dropped, the number of bacterial taxa in a sample increased until a peak in richness was reached near 4 mg/L of DO, below which bacterial richness decreased. Similarly, observations of microbial communities in a seasonally anoxic fjord documented an increase in bacterial richness as DO levels reach hypoxic levels and then a sharp dropoff of richness and diversity as DO levels continued to drop in deeper water samples (Zaikova et al. 2010). Interestingly, this level of DO (2-4 mg/L) is a threshold for sublethal and lethal affects on macrofauna (Vaquer-Sunyer & Duarte 2008), and below this threshold benthic macrofauna experience decreased abundance, biomass, and diversity (Seitz et al. 2009; Conley et al. 2007; Montagna & Ritter 2006). The stress induced upon oxygen-dependent organisms may help explain the peak in number of bacterial taxa co-existing as benthic energy is transferred from macrofauna to microbes (Diaz & Rosenberg 2008). It is possible that an intermediate DO concentration where aerobic and facultative anaerobic lifestyles can coexist and this DO concentration may be near 4 mg/L. For example, above this concentration anaerobic taxa are

less competitive, and conversely aerobic taxa are less competitive below the optimum concentration. Therefore, the highest number of different types of bacteria persists at the intersection of the two fitness curves.

Though examining the overall patterns of richness and composition of all bacteria in Hood Canal provides valuable insight into the biological dynamics of the system, we do not expect all groups of bacteria to exhibit similar responses to environmental change and whole-domain patterns can mask changes on a finer taxonomic scale. Furthermore, undersampling of community diversity is common in bacterial studies given the incredible abundance and richness of bacteria in the environment. Thus, though detection of every bacterial taxon present in a sample is not yet feasible (SI Figure 2), we expect the most abundant classes of bacteria to be sampled deeper thus examining class-level patterns is of interest. The rise in *Flavobacteria* taxa richness associated with a decreased in DO suggests a greater metabolic diversity of *Flavobacteria* is present in the low oxygen samples, where high rates of decomposition are likely consuming oxygen. *Flavobacteria* are crucial in the uptake and degradation of high molecular weight dissolved organic carbon, such as chitin and cellulose, especially following periods of intense enrichment from a phytoplankton bloom (Fandino et al. 2001). The *Gammaproteobacteria* class is also metabolically diverse; therefore, it is unsurprising that we found increased *Gammaproteobacteria* richness in samples of lower dissolved oxygen. Consistent with our results, Wright et al. (2012) observed an increased number of *Gammaproteobacteria* taxa in suboxic zones relative to oxic zones across multiple regions. Contrary to patterns of other dominant groups, the taxa richness of *Alphaproteobacteria* was strongly associated with salinity rather than with DO. *Alphaproteobacteria* are predominantly found in marine environments; therefore, it is not surprising that we detected more taxa in high

salinity water samples that are more similar to marine environments than to the freshwater inputs of Hood Canal.

The rapid expansion of oxygen minimum zones and coastal hypoxia has huge implications for the global nitrogen and carbon cycles; therefore, understanding the changes in bacterial diversity and composition of particular functional groups involved in these cycles is essential. While the approach used in this study does not allow a detailed examination of specific bacterial activity, it does offer a window into the dynamics of important functional groups. We detected significant relationships between DO concentration and the abundances of major players in the nitrogen and carbon cycles. Shifts in the relative proportion and compositions of particular functional groups suggest that the potential exists for decreasing DO concentration to alter the fate of nitrogen and carbon. Further exploration into the taxonomic and functional responses of individual bacterial groups in Hood Canal is crucial for us to better understand how depleted DO affects multiple biogeochemical cycles on both regional and global scales.

Conclusion

Here we report shifts in bacterial richness and community composition across a gradient of abiotic factors in Hood Canal, WA, a seasonally oxygen-deficient estuary. The estuary remains highly stratified year-round and supports distinct bacterial communities in surface and deep water masses. Changes in DO concentration were highly correlated with changes in the bacterial communities at multiple taxonomic scales. We uncovered a hump-shaped relationship between bacterial richness and DO concentration, which peaked at the same concentration at which biological stress occurs in macro-organisms. The hydrography of Hood Canal establishes

a wide gradient of abiotic factors, including DO that changes rapidly on an annual scale, which establishes a unique natural laboratory in which these questions can be explored. We are only beginning to understand how the microbial communities are responding to rapidly decreasing DO, but further questions into the impacts on individual functional groups of bacteria will be crucial in elucidating the global consequences of rapidly expanding coastal hypoxia.

Acknowledgements

The authors wish to thank Allan Devol, Jan Newton, and Wendi Ruef for supplying oxygen and chlorophyll data from ORCA buoys in Hood Canal, which are operated by the University of Washington and supported by and maintained as part of the Northwest Association of Networked Ocean Observing Systems (NANOOS), part of the U.S. Integrated Ocean Observing System (IOOS), awarded by NOAA (NA07NOS4730203). Sampling was accomplished through a Hood Canal Dissolved Oxygen Program Integrated Assessment and Modeling (HCDOP-IAM) program award to GR and MCHD funded through the U.S. Navy, Naval Sea Systems Command (N0024-02-D-6602/50). Sequencing was performed as part of the International Census of Marine Microbes (ICoMM) supported by a W.M. Keck Foundation award to the Marine Biological Laboratory at Woods Hole. We thank Michael Dyen for assistance in sample processing and data collection and KathiJo Jankowski for helpful and insightful comments on the manuscript. RKL acknowledges support from the National Science Foundation Graduate Research Fellowship Program.

References:

Agogué H, Lamy D, Neal Phillip R, Sogin ML, Herndl GJ. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. Molecular ecology 20:258–74.

Amend AS, Seifert KA, Bruns TD. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Molecular ecology 19:5555–65.

Bowen JL, Morrison HG, Hobbie JE, Sogin ML. (2012). Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. The ISME Journal 6:2014–2023.

Chao A, Chazdon RL, Colwell RK, Shen T-J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361–71.

Cho BC, Azam F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. Nature 332:441–443.

Clarke KR, Ainsworth M. (1993). A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series 92:205–219.

Codispoti LA, Brandes JA, Christensen JP, Devol A.H., Naqvi SWA, Paerl Hans W., et al. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene?*. Scientia Marina 65:85–105.

Conley DJ, Carstensen J, Aertburg G, Christensen PB, Dalsgaard T, Hansen JLS, et al. (2007). Long-term changes and impacts of hypoxia in Danish coastal waters. Ecological Applications 17:S165–S184.

Connell JH. (1978). Diversity in Tropical Rain Forests and Coral Reefs. Science 199:1302–1310.

Devol Allan H, Ruef Wendi, Emerson S, Newton Jan. (2007). In Situ and Remote Monitoring of Water Quality in Puget Sound□: the ORCA Time-Series. Environmental Research, Engineering and Management 1:19–23.

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

Feng B-W, Li X-R, Wang J-H, Hu Z-Y, Meng H, Xiang L-Y, et al. (2009). Bacterial diversity of water and sediment in the Changjiang estuary and coastal area of the East China Sea. FEMS Microbiology Ecology 70:236–248.

Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences 103:626–31.

Friedline CJ, Franklin RB, McCallister SL, Rivera MC. (2012). Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. Biogeosciences 9:2177–2193.

Fuchs BM, Woebken D, Zubkov M V, Burkill P, Amann Rudolf. (2005). Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. Aquatic Microbial Ecology 39:145–157.

Fuchsman CA, Kirkpatrick JB, Brazelton WJ, Murray JW, Staley JT. (2011). Metabolic strategies of free-living and aggregate-associated bacterial communities inferred from biologic and chemical profiles in the Black Sea suboxic zone. FEMS Microbiology Ecology 78:586–603.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown M V, 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, Hewson I, Schwalbach MS, Brown M V, Green JL, et al. (2008). A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences 105:7774–7778.

Ghiglione J-F, Galand PE, Pommier T, Pedrós-Alió C, Maas EW, Bakker K, et al. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. Proceedings of the National Academy of Sciences 109:17633–8.

Gilbert J, Steele J, Caporaso JG, Steinbrück L, Reeder J, Temperton B, et al. (2012). Defining seasonal marine microbial community dynamics. The ISME Journal 6:298–308.

Gregg MC, Pratt LJ. (2010). Flow and Hydraulics near the Sill of Hood Canal, a Strongly Sheared, Continuously Stratified Fjord. Journal of Physical Oceanography 40:1087–1105.

Horner-Devine MC, Carney KM, Bohannan BJM. (2004). An ecological perspective on bacterial biodiversity. Proceedings of the Royal Society of London, Biological Sciences 271:113–22.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM. (2004). A taxa-area relationship for bacteria. Nature 432:750–753.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 8:R143.

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

Hutchinson GE. (1961). The Paradox of the Plankton. The American Naturalist 95:137–145.

Jones SE, Lennon JT. (2010). Dormancy contributes to the maintenance of microbial diversity. Proceedings of the National Academy of Sciences of the United States of America 107:5881–6.

Keister JE, Tuttle LB. (2013). Effects of bottom-layer hypoxia on spatial distributions and community structure of mesozooplankton in a sub-estuary of Puget Sound, Washington, U.S.A. Limnology and Oceanography 58:667–680.

Koskinen K, Hultman J, Paulin L, Auvinen P, Kankaanpää H. (2011). Spatially differing bacterial communities in water columns of the northern Baltic Sea. FEMS Microbiology Ecology 75:99–110.

Labrenz M, Jost G, Jürgens K. (2007). Distribution of abundant prokaryotic organisms in the water column of the central Baltic Sea with an oxic – anoxic interface. Aquatic Microbial Ecology 46:177–190.

Lin X, Wakeham SG, Putnam IF, Astor YM, Scranton MI, Chistoserdov AY, et al. (2006). Comparison of Vertical Distributions of Prokaryotic Assemblages in the Anoxic Cariaco Basin and Black Sea by Use of Fluorescence In Situ Hybridization. Applied and Environmental Microbiology 72:2679–2690.

Long RA, Azam F. (2001). Microscale patchiness of bacterioplankton assemblage richness in seawater. Aquatic Microbial Ecology 26:103–113.

Madrid VM, Taylor GT, Scranton MI, Chistoserdov AY. (2001). Phylogenetic Diversity of Bacterial and Archaeal Communities in the Anoxic Zone of the Cariaco Basin. Applied and Environmental Microbiology 67:1663–1674.

Martiny Jennifer B H, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108:7850–7854.

Martiny Jennifer B Hughes, Bohannan BJM, Brown JH, Colwell RK, Fuhrman J a, Green JL, et al. (2006). Microbial biogeography: putting microorganisms on the map. Nature reviews. Microbiology 4:102–12.

Montagna PA, Ritter C. (2006). Direct and indirect effects of hypoxia on benthos in Corpus Christi Bay, Texas, U.S.A. Journal of Experimental Marine Biology and Ecology 330:119–131.

Newton J, Devol A, Ruef W. (2009). Measuring nitrate fluxes to assess estuarine eutrophication. Decision Analysis 2:5–6.

Newton JA, Bassin C, Devol Al, Kawase M, Ruef Wendi, Warner M, et al. (2007). Hypoxia in Hood Canal□: An overview of status and contributing factors. In:Georgia Basin Puget Sound Research Conference.

Newton Jan, Bassin C, Devol Allan H, Richey JE, Kawase M, Warner MJ. (2011). Hood Canal Dissolved Oxygen Program Integrated Assessment and Modeling Report 2011.

Petraitis PS, Latham RE, Niesenbaum RA. (1989). The Maintenance of Species Diversity by Disturbance. The Quarterly Review of Biology 64:393–418.

Pinckney JL, Paerl H. W., Harrington MB, Howe KE. (1998). Annual cycles of phytoplankton community-structure and bloom dynamics in the Neuse River Estuary, North Carolina. Marine Biology 131:371–381.

Pinto AJ, Raskin L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PloS one 7:e43093.

Pommier T, Neal P. R., Gasol JM, Coll M, Acinas SG, Pedrós-Alió C. (2010). Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. Aquatic Microbial Ecology 61:221–233.

Rabalais NN, Díaz RJ, Levin LA, Turner RE, Gilbert D, Zhang J. (2010). Dynamics and distribution of natural and human-caused hypoxia. Biogeosciences 7:585–619.

Seitz RD, Dauer DM, Llansó RJ, Long WC. (2009). Broad-scale effects of hypoxia on benthic community structure in Chesapeake Bay, USA. Journal of Experimental Marine Biology and Ecology 381:S4–S12.

Shade A, Jones SE, McMahon KD. (2008). The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. Environmental Microbiology 10:1057–67.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal Phillip R, et al. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proceedings of the National Academy of Sciences 103:12115–20.

Steinberg PD, Brett MT, Bechtold JS, Richey JE, Porensky LM, Smith SN. (2011). The influence of watershed characteristics on nitrogen export to and marine fate in Hood Canal, Washington, USA. Biogeochemistry 106:415–433.

Steinberg PD, Brett MT, Porensky LM, Smith SN. (2010). The influence of watershed characteristics on nitrogen export to and marine fate in Hood Canal , Washington , USA.

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

Tilman D. (1999). The Ecological Consequences of Changes in Biodiversity: A Search for General Principles. Ecology 80:1455–1474.

Urakawa H, Martens-Habbena W, Stahl DA. (2010). High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. Applied and Environmental Microbiology 76:2129–35.

Vaquer-Sunyer R, Duarte CM. (2008). Thresholds of hypoxia for marine biodiversity. Proceedings of the National Academy of Sciences of the United States of America 105:15452–7.

Vetriani C, Tran H V, Kerkhof LJ. (2003). Fingerprinting Microbial Assemblages from the Oxic/Anoxic Chemocline of the Black Sea. Applied and Environmental Microbiology 69:6481–6488.

Wakeham SG, Amann Rudi, Freeman KH, Hopmans EC, Jorgensen BB, Putnam IF, et al. (2007). Microbial ecology of the stratified water column of the Black Sea as revealed by a comprehensive biomarker study. Organic Geochemistry 38:2070–2097.

Warner MJ, Kawase M, Newton JA. (2001). Recent Studies of the Overturning Circulation in Hood Canal. Puget Sound Research.

Wright JJ, Konwar KM, Hallam SJ. (2012). Microbial ecology of expanding oxygen minimum zones. Nature Reviews Microbiology 10:381–94.

Zaikova E, Walsh D a, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Environmental microbiology 12:172–91.

Zubkov M V, Tarran GA, Burkill PH. (2006). Bacterioplankton of low and high DNA content in the suboxic waters of the Arabian Sea and the Gulf of Oman□: abundance and amino acid uptake. Aquatic Microbial Ecology 43:23–32.

Tables

Table 1. Abiotic factors affect the diversity and composition of bacterial communities in Hood Canal, WA. A (**) indicates a significant effect, p<0.01; (+) indicates a positive significant association; (-) indicates a negative significant association. Relationships were considered significant at alpha=0.05 level.

	Depth	Season	Site	DO	Temp	NH_4^+	NO_3^-	NO_2^-	PO_4^{3-}	SiO_4	Sal
Richness											
Whole domain	**										
Alpha-proteo	**										+
Flavo	**			-							
Gamma-proteo	**			-							
Evenness											
Whole domain	**								+		
Alpha-proteo	**						-			-	
Flavo	**								+		+
Gamma-proteo	**								+	+	
Composition											
Whole domain	**	**		**							
Alpha-proteo	**	**		**					**		
Flavo	**	**		**							
Gamma-proteo	**	**		**							

Table 2. Abiotic factors affect the relative abundances of various bacterial functional groups. A (+) sign indicates a significant positive relationship and a (-) sign indicates a significant negative relationship. Relationships were considered significant at alpha=0.05 level.

Functional group	Total # of reads	Average proportion	DO	Temp	NH4	NO3	NO2	PO4	SiO4	Sal
Nitrite-oxidizing	2484	0.0136	-		-					
Ammonia-oxidizing	300	0.0012				+				
Cyanobacteria	701	0.0044		+						
Methylotrophs	2241	0.0136	-	+						

Figures

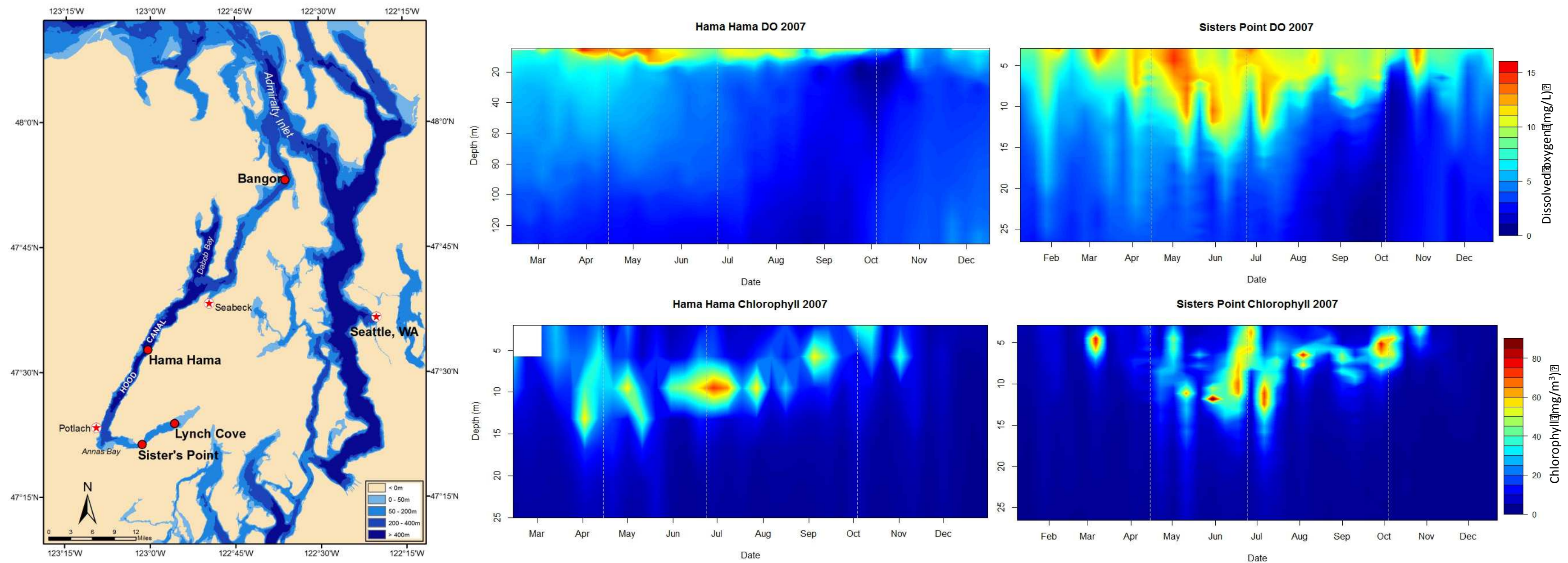


Figure 1. (A) Map of sampling stations within Hood Canal, WA. (B) Contour plots showing range of dissolved oxygen and chlorophyll at Hama Hama and Sister's Point stations in Hood Canal, WA. Data for high resolution depth profiles collected by ORCA buoys maintained by NANOOS (http://www.nanoos.org). Note differences in scale on both the x and y axes. Vertical dashed lines indicate sampling time points.

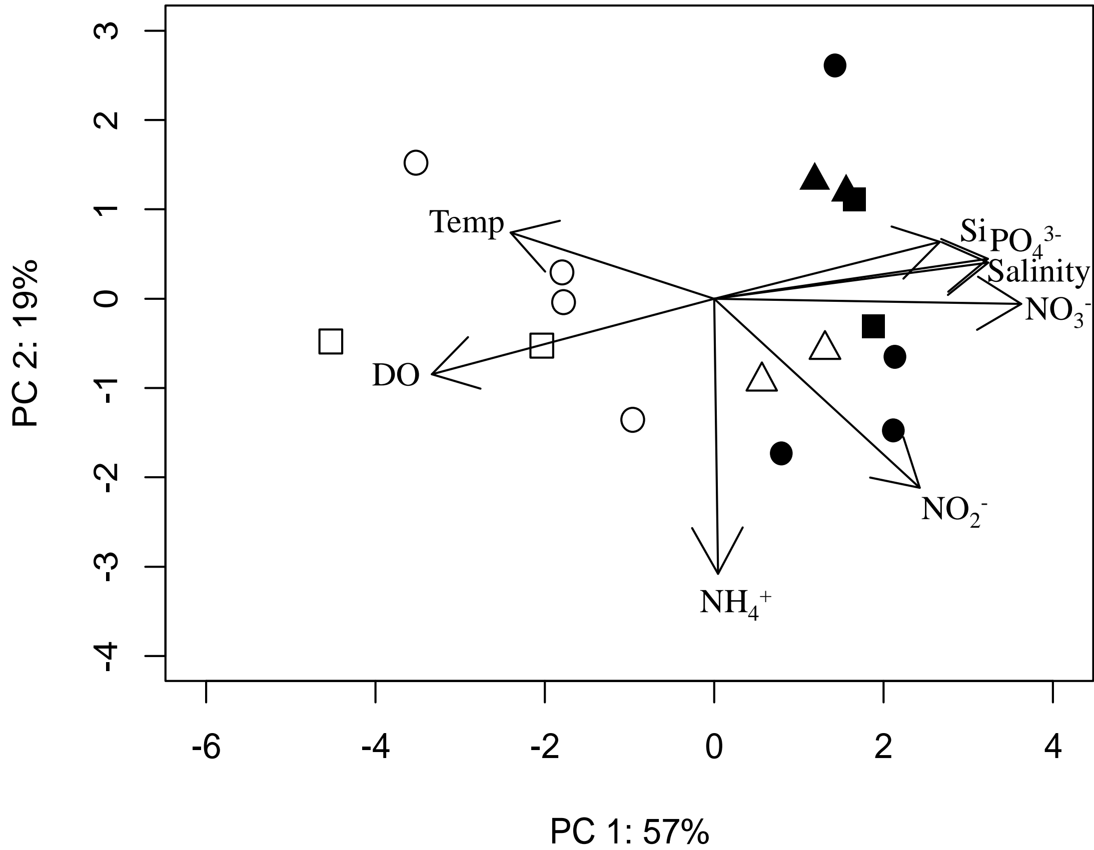


Figure 2. Principal components analysis of environmental characteristics of the sixteen samples ordinated based on Euclidean distance calculated from environmental factors. Points are coded by water depth as either deep (filled symbols) or surface (open symbols) and by season as April (square), June (circle), or October (triangle). Nitrate, phosphate, salinity, and dissolved oxygen best explain the separation of samples along axis 1, and ammonium alone explains variation along axis 2 based on the variable loadings on each principal component.

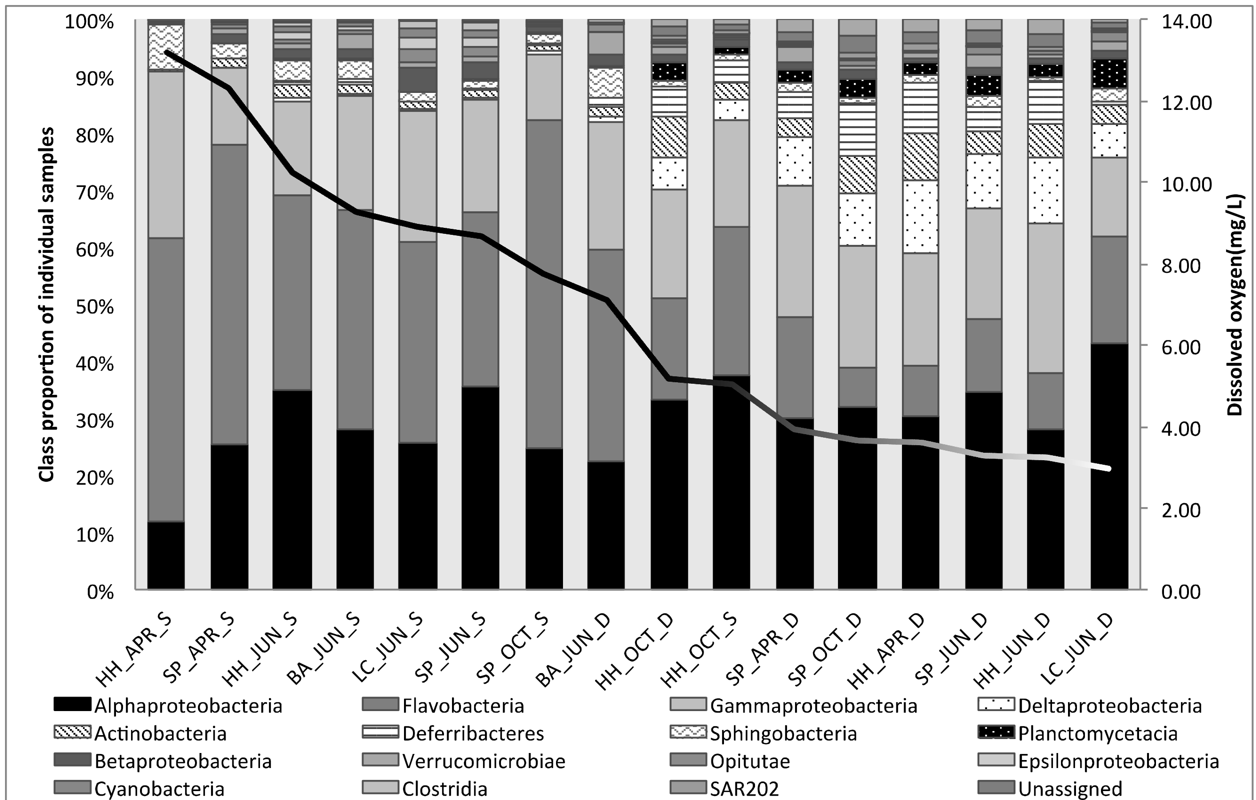


Figure 3. Relative proportions of the most abundant classes change across a dissolved oxygen gradient and are significantly different when grouped by sampling depth. The samples are ordered by dissolved oxygen concentration (solid line). Abbreviations: Hama Hama (HH), Sister's Point (SP), Bangor (BA), Lynch Cove (LC), April (APR), June (JUN), October (OCT), Surface (S), Deep (D). Classes that comprised less than 1% of the sample's total community were condensed into the "other" category.

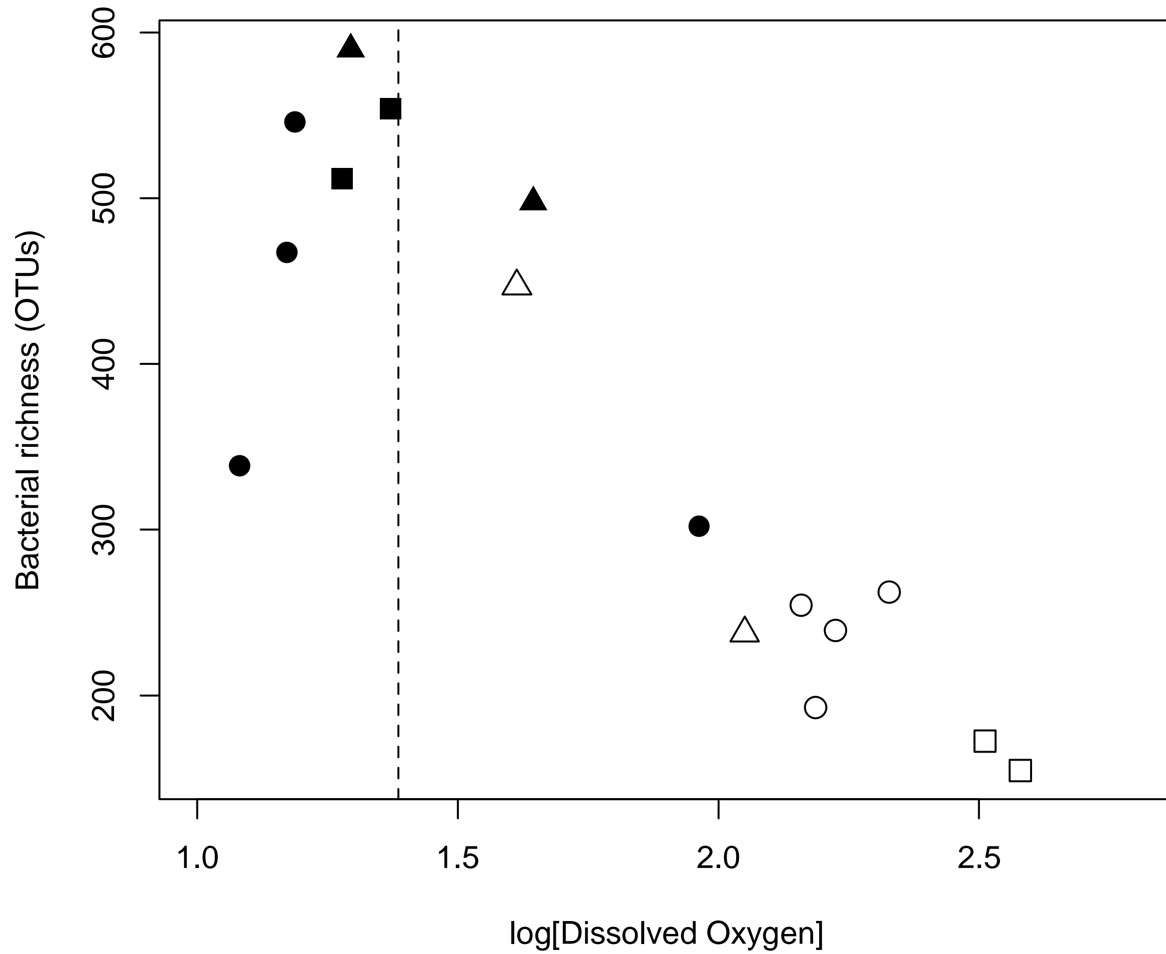


Figure 4. A hump-shaped relationship exists between dissolved oxygen concentration and bacterial taxa richness. Points are coded by water depth as either deep (filled symbols) or surface (open symbols) and by season as April (square), June (circle), or October (triangle). Vertical dashed line represents the equivalent dissolved oxygen concentration of 4 mg/L, which is the threshold for hypoxia-induced biologic stress in macro organisms.

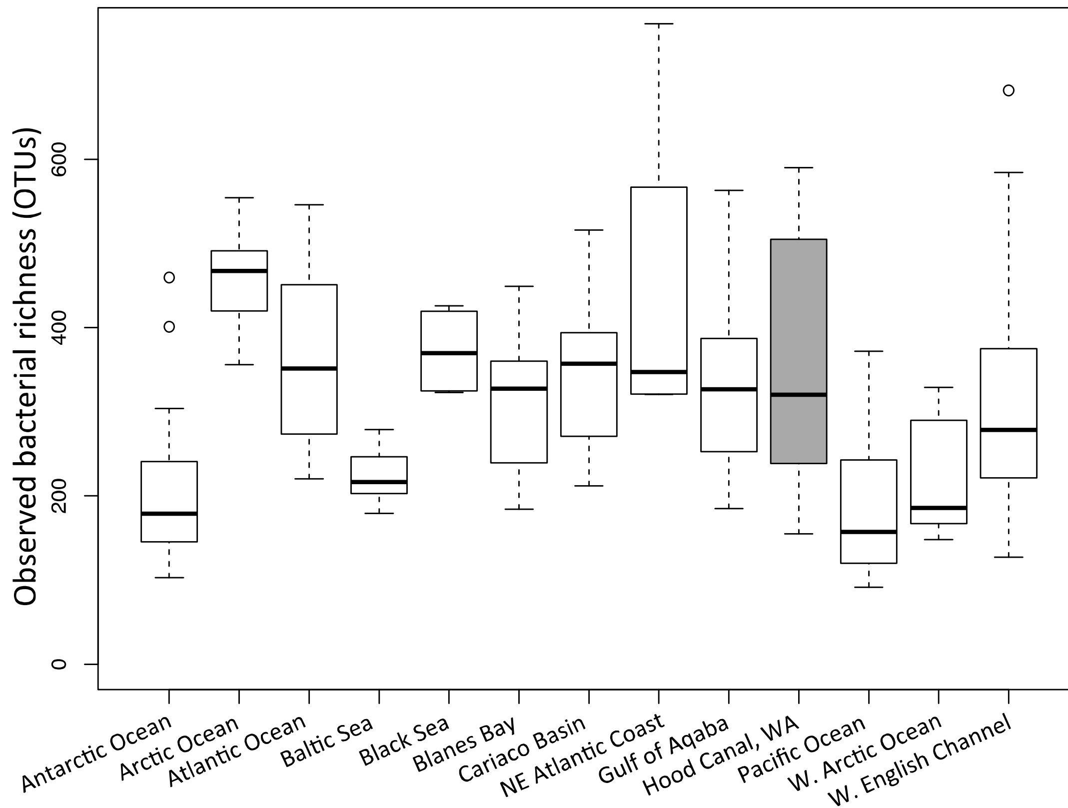


Figure 5. Bacterial richness varies across biogeographic regions. Hood Canal samples from this study are shown in grey. Sequence data from other locations were collected from the ICoMM database and processed through the same analysis pipeline.

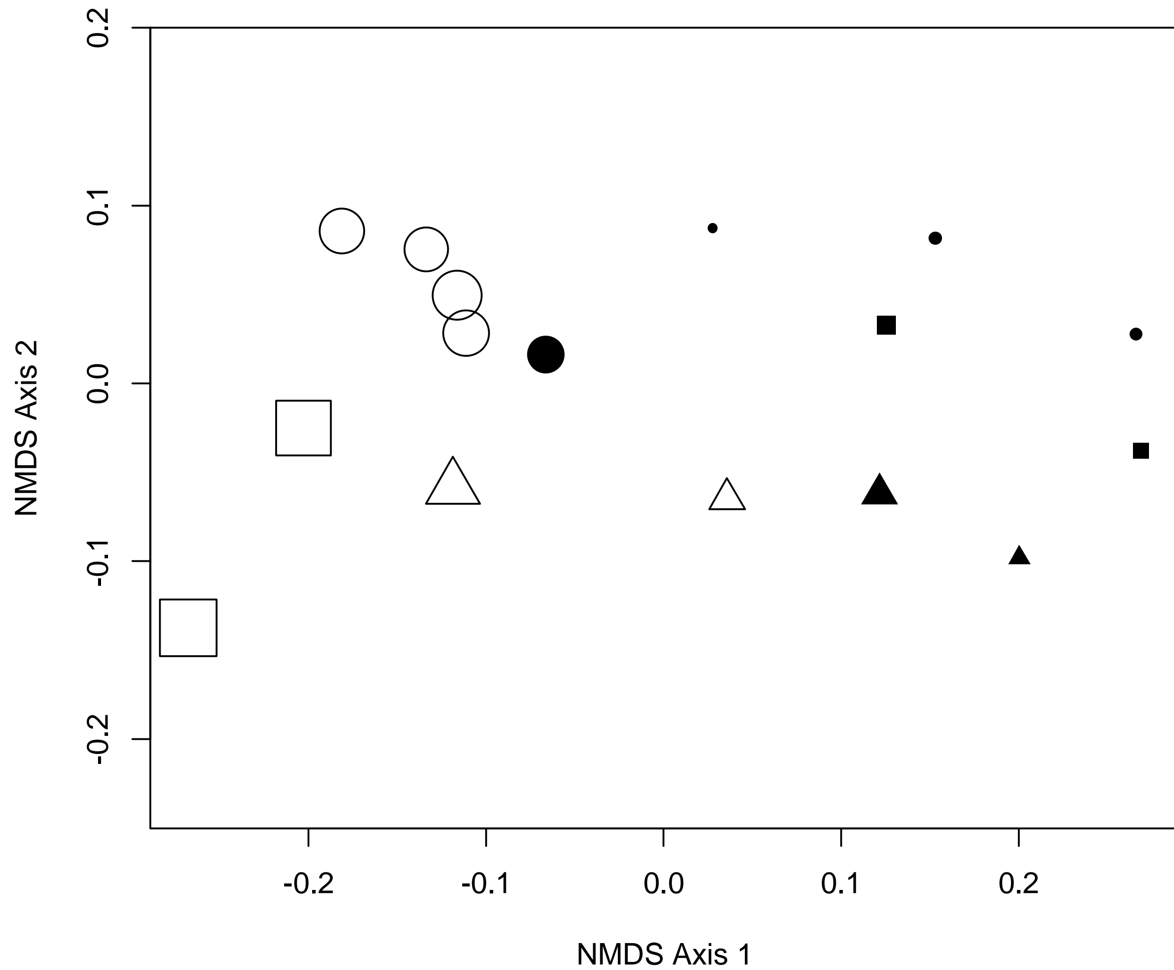
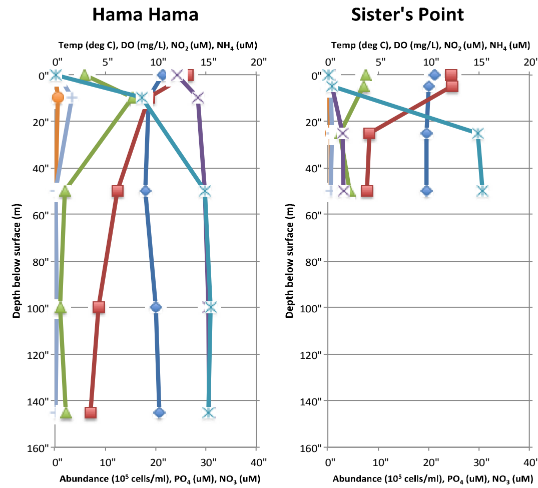


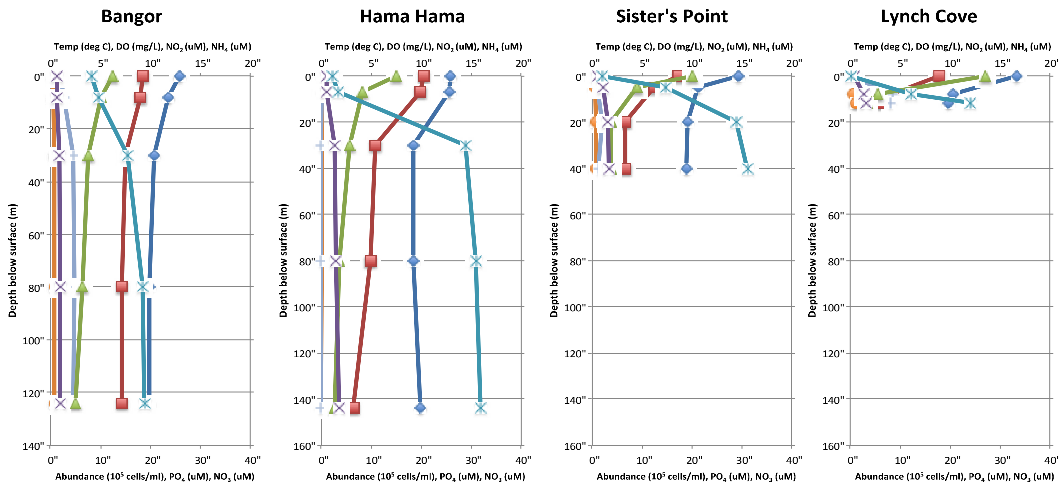
Figure 6. Non-metric multidimensional scaling ordination shows bacterial community composition of samples are more similar to those from the same water depth, season, or dissolved oxygen concentration. Position of samples is calculated using the Sorensen abundance-based similarity index of bacterial community composition and the absolute dissimilarities between samples is condensed onto two-dimensions with a resulting stress value of 0.0679. Points are coded by water depth as either deep (filled symbols) or surface (open symbols) and by season as April (square), June (circle), or October (triangle). The size of the sample point is proportional to dissolved oxygen content with the highest dissolved concentration represented by the largest sample point.

Supplementary figures

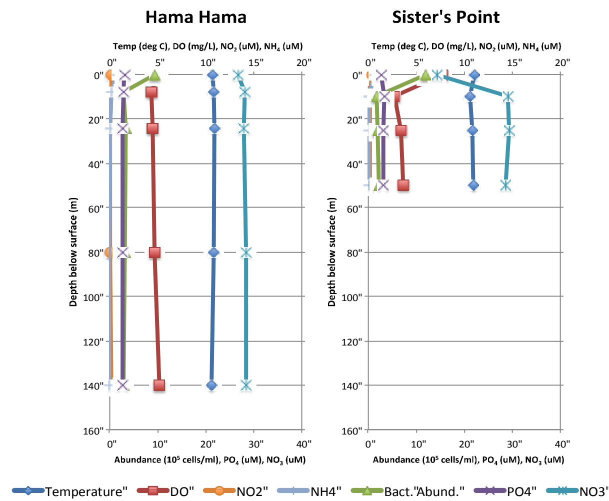
April 2007



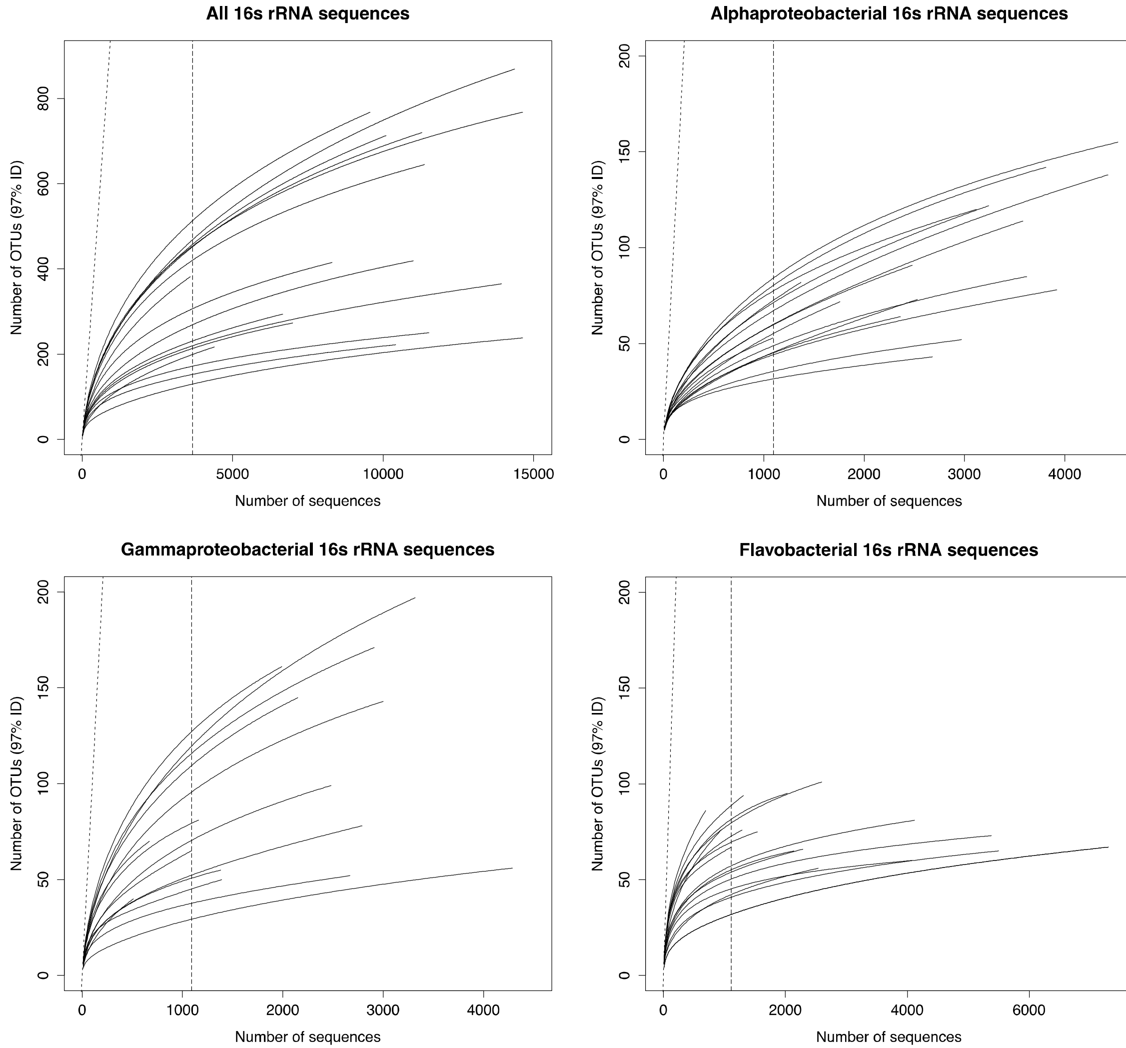
June 2007



October 2007



SI Figure 1. Depth profiles of physical and chemical data recorded during (A) April, (B) June, and (C) October sampling at Bangor, Hama Hama Sister's Point, and Lynch Cove.



SI Figure 2. Rarefaction curves of A) All 16S sequences, B) Alphaproteobacterial sequences only, C) Gammaproteobacterial sequences only, and D) Flavobacterial sequences only. Each solid line represents the sequence accumulation curve for each sample in the dataset. The dotted line represents a 1:1 line, which is the maximum taxa accumulation per sequence. The dashed, vertical line in each plot represents the number of sequences to which each set was subsampled. Note that plot A is on a different scale from plots B-D.

SI Table 1. Abiotic data collected to accompany each bacterial pyrosequencing sample and used in PCA and regression analyses.

	Depth	*Nitrate [µM]*			*Nitrite [µM]*			*Ammonium [µM]*			*Phosphate [µM]*			*Silicate [µM]*			*Salinity [PSU]*			*Temperature [°C]*			*Oxygen [mg L^{-1}]*		
Location	*(m)*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*	*Apr*	*June*	*Oct*
	0		8.18			0.30			0.96			1.18			27.54			28.56			12.91			9.25	
Bangor	124		18.79			0.38			2.31			1.91			41.85			30.53			9.86			7.12	
	0	0.73	2.32	26.77	0.00	0.08	0.21	0.46	0.10	0.77	0.01	1.01	3.03	36.52	27.41	79.97	24.29	27.72	28.93	10.71	12.96	10.74	13.18	10.25	5.02
Hama- Hama	140	30.76	31.97	28.36	0.05	0.00	0.14	0.06	0.00	0.00	3.46	3.55	2.71	77.90	80.30	58.60	30.62	30.32	30.66	10.39	9.95	10.62	3.59	3.23	5.18
	0	0.03	2.00	14.24	0.04	0.05	0.37	0.30	0.16	0.67	0.73	1.00	2.62	56.23	46.75	91.08	23.59	26.18	26.49	10.09	14.63	11.03	12.32	8.66	7.77
Sisters Point	45	30.59	31.08	28.69	0.32	0.38	0.17	0.28	0.62	0.01	3.12	3.34	3.10	73.79	75.04	65.39	29.84	29.71	30.49	9.78	9.51	10.97	3.94	3.28	3.65
	0		0.08			0.00			0.04			0.77			41.47			24.80			16.62			8.90	
Lynch Cove	11.7		23.94			0.50			4.01			3.21			73.64			29.53			9.80			2.95	

SI Table 2. Observed, Chao, and ACE estimates of taxa richness at various samples collected in ICoMM project.

Location	*Observed richness*	*Chao*	*ACE*
Hood Canal, WA	**361**	**742**	**732**
Pacific Ocean	188	328	334
Western Arctic Ocean	223	381	368
Atlantic Ocean	362	639	641
Baltic Sea	224	363	362
Cariaco Basin	350	681	723
NE Atlantic Ocean Coast	444	1055	1059
Gulf of Aqaba	332	568	572
Western English Channel	307	574	570

Appendix A: Assessing the performance of multiple community similarity measures for bacterial communities in the face of sampling biases

Introduction

Understanding changes in the incidences and abundances of species over space and time is crucial for ecologists (Whittaker 1962, 1972). These patterns in community composition, or beta diversity, can inform ecologists about the scale at which communities vary over a range of spatial and temporal gradients. Understanding how communities are structured and organized offers valuable insight in the forces that underlie and maintain diversity – a central goal of ecology. Changes in beta diversity over space and time may also impact important ecosystem functions and health, as different species are responsible for different ecosystem processes (Downing & Leibold 2002; Hooper et al. 2005). Further, such an understanding is crucial for predicting how communities respond to anthropogenic impacts as well as to natural perturbations and variability. Understanding patterns of beta diversity for microbial communities is no exception, though our understanding of community similarity indices used to measure and compare bacterial community composition lags behind that of macroorganisms and may be impacted by multiple biases – some unique to microbial communities.

Multiple similarity (or dissimilarity) indices have been proposed and used by ecologists to compare the species composition of two biological communities (see reviews of indices in Magurran 2003); however, no consensus currently exists as to which measure of beta-diversity is preferred (Whittaker et al. 2001; Koleff et al. 2003; Cardoso et al. 2009; Beck et al. 2013). These indices of beta-diversity can be divided into two groups. First, qualitative indices consider only the presence or absence of taxa without regard to the relative abundances of those taxa. For example, the widely-used classic Jaccard (Jaccard 1908) and classic Sørensen (Sørensen 1948)

similarity indices depend on the total number of taxa present in each community and the number of taxa shared between the two communities (Koleff et al. 2003). Using only incidence data, two pairs of communities may produce the same similarity indices even though the abundances of shared species could differ considerably between the two pairs. To detect these differences, abundance based similarity indices, referred to as quantitative indices, comprise the second group of indices (Chao et al. 2005). The classic Jaccard and Sørensen indices were revised to include relative abundances of shared species better compare two communities (Chao et al. 2005). Additional indices, including but limited to Bray-Curtis and Morista-Horn, have also been developed to quantify community similarity using abundance-based data (Magurran 2004).

The extent of sampling effort can often limit accurate quantification and comparison of community composition in diverse communities (Cardoso et al. 2009), and it is common to not detect all individuals residing in the community. Low sampling effort commonly leads to underestimation of beta diversity due to lack of detection of shared species in both communities (Colwell & Coddington 1994; Chao et al. 2005). In an attempt to overcome this detection issue, probabilistic approaches have been applied to abundance-based similarity indices to estimate the effect of unseen individuals in a sample of community composition (Chao et al. 2005, 2006). The problem of undetection of some taxa is an issue faced by ecologists studying any diverse community, and it has been shown that for diverse insect and plant communities incorporation of the effect of unseen shared species can not only increase accuracy, but can also change the interpretation of results (Chao et al. 2005).

The probability of not detecting all species is more pronounced for rich microbial communities than even in diverse insect and plant communities (Hughes et al. 2001). Characteristics of microbial communities including high abundances and hyper-diversity likely

affect the measurements of beta diversity in bacterial communities as well (Lemos et al. 2012). Many studies of bacterial community composition use similarity indices and approaches originally designed for use in ecological studies of macroorganisms that may not perform well in the circumstances of undersampling and uneven sample sizes known to be the case for many microbial communities.

Similarity indices exist that include a correction value to estimate for the unseen taxa from the sample (Chao et al. 2006). The performance of these estimators have been assessed in diverse and under-sampled insect and tree sapling datasets (Chao et al 2005, Cardoso et al 2009, Beck et al 2013), but never before for hyper-diverse bacterial communities. In the current study, I investigated the performance of commonly used qualitative and quantitative community similarity indices in the face of three biases common in bacterial assessments: undersampling, uneven sampling effort, the effect of highly skewed distributions of taxa abundances. I used pyrosequencing datasets previously collected from Hood Canal, WA water samples to evaluate the performance of these similarity indices.

Methods

Similarity indices.

Two classic qualitative community similarity indices were tested. The classic Jaccard index (1908) is given by:

$$S_{CJ}=\frac{a}{a+b+c} \qquad \text{(Equation 1)}$$

Where *a* is the total number of species present in both samples, *b* equals the number of species in sample one, and *c* is equal to the number of species in sample two.

The second qualitative index, the classic Sørensen index (1948), is similar to the Jaccard but weights presences in both samples more heavily than absences in one sample in order to avoid underestimation of similarity when unseen species may be an issue. The index is given by:

$$S_{CS}=\frac{2a}{2a+b+c} \quad \text{(Equation 2)}$$

The performance of six commonly used quantitative community similarity indices were examined in this study. The Bray-Curtis (Bray & Curtis 1957) similarity index is described by the equation

$$S_{BC}=\frac{2jN}{(N_a+N_b)} \quad \text{(Equation 3)}$$

where N_a= the total individuals in site A, N_b= the total individuals in site B, and 2jN= the sum of the lower of the two abundances for species found in both sites.

The Morista-Horn similarity index, unlike other indices, is not strongly affected by taxa richness and sample size. However, it has been noted to be sensitive to the abundances of the most abundant species (Magurran 2004). The equation for the Morista-Horn index is

$$S_{MH}=\frac{2\Sigma(a_i+b_i)}{(d_a+d_b)*(N_a*N_b)} \quad \text{(Equation 4)}$$

where N_a remains as the total number of individuals at site A and N_b the total number of individuals at site B. Additional terms a_i and b_i are the number of individuals in the ith species in site A and B, respectively. The terms d_a and d_b are described by

$$d_a=\frac{\Sigma a_i^2}{N_i^2}$$

$$d_b=\frac{\Sigma b_i^2}{N_i^2}$$

The incidence-based Jaccard similarity index was revised in 2005 by Chao and colleagues to incorporate the relative abundances of taxa to give a quantitative, rather than qualitative, measure of similarity. The raw Jaccard abundance-based similarity index is given by

$$S_J = \frac{UV}{U+V-UV} \quad \text{(Equation 5)}$$

Where U and V give the total relative abundance, or proportion, of individuals belonging to species shared between two sites found in site U or site V, respectively. However, in samples of community composition there are typically species that are unseen or undetected but exist in the community. To estimate for the unseen species, an estimated Jaccard index is described by

$$S_{EJ} = \frac{\hat{U}\hat{V}}{\hat{U}+\hat{V}-\hat{U}\hat{V}} \quad \text{(Equation 6)}$$

Where the $\hat{U}$ and $\hat{V}$ terms are similar to U and V, but include adjustments for the estimated proportion of shared species that are unseen in a sample. See Chao et al 2005 for equations of the estimator terms.

Finally, the Sørensen abundance-based index is a revision of the classic Sørensen similarity index based on incidence data. The Sørensen abundance-based similarity index down weights unshared species thus may be preferred in under sampled communities. Using the same variables as the Jaccard index, the raw Sørensen index is given by

$$S_S = \frac{2UV}{U+V} \quad \text{(Equation 7)}$$

Similar to the Jaccard similarity, the Sørensen index has been revised to include estimates for unseen species and is described by

$$S_{ES} = \frac{2\hat{U}\hat{V}}{\hat{U}+\hat{V}} \quad \text{(Equation 8)}$$

Where $\hat{U}$ and $\hat{V}$ have the same definition as in Equation 4.

Datasets.

Sequence datasets were obtained from a previous study in which pyrosequencing techniques were applied to water samples from Hood Canal, WA collected from Spring to Autumn 2007. For details on sequencing and sequence processing please refer to Lange et al., *in review*.

Analyses.

I conducted performance tests on two separate large, diverse samples from the Hood Canal dataset. Community A, collected from surface waters collected in April 2007 at station Hama Hama, contained the second most sequences and rarefaction curves (Figure 5) suggested this community was most exhaustively sampled from the entire dataset, thus allowing us to assume it to be the best representation of a full and complete community library. The library contained 14,743 sequences, or individuals, and 341 operational taxonomic units (OTUs) defined at 97% sequence identity, of which 198 were singletons and 36 were doubletons. Community B was from the deep water mass collected at the same time in April 2007 at station Hama Hama as Community A. Community B contained 14,898 sequences and 1,020 OTUs thus suggesting that this community has more even distribution of relative abundances of taxa than Community A. A rarefaction curve for Community B shows that the diversity of the sample is high with community sampling near saturation and is thus slightly undersampled at 14,898 individuals (Figure 5).

To examine the effect of undersampling, pairs of equal-sized samples from the full library were compared. I randomly sampled individuals, with replacement, from the full library to produce pairs of samples containing the same number of individuals as the full library. This was repeated to obtain pairs of samples with one-half the total number of individuals in the full library, and then I calculated the eight similarity indices between each pair. I repeated this process for 1/4, 1/8, 1/16, 1/32, and 1/64 the number of individuals of the full library. The process was repeated 1000 times for each sample size, and the mean similarity index was calculated over the 1000 replicates. Because I sampled with replacement, some taxa may not be

present in each sample, and likewise the abundances of other taxa may be higher than the original abundance in the full library.

First used in a study of bacterial taxa-area relationships by Horner-Devine et al (2004), bacterial community datasets are often rarefied to equilibrate sample sizes. This rarefaction subsampling process is then repeated multiple times and the indices are averaged in order to avoid stochastic events that arise from a single subsample and a more accurate representation of the community can be achieved. Using this approach, I followed the methods used by Chao et al (2005, 2006) to assess the performance of similarity indices when samples of uneven sizes are compared (i.e. differing numbers of sequences or individuals). Individuals were randomly sampled, with replacement, to 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64 of the size of the full library. Each of lower-sized samples was compared to the full library and similarity indices were calculated for each pair of sample sizes. This process was also repeated 1000 times and each similarity index was averaged for each sample size pair.

Finally, I assessed the influence of community structure, particularly differing evenness of taxa abundances, on the performance of eight similarity indices. To examine this effect, the first two tests of undersampling and uneven sample sizes were repeated with Community B, which was comprised of taxa with more even abundances than Community A.

Results and discussion

Numerous similarity indices exist to compare the species composition between two communities. In microbial ecology, it is common to see many of these different similarity indices being used in the literature. However, the performance of these different similarity indices varies when faced with undersampling and uneven sample sizes, both of which are

typical of microbial sampling methods. The performance of eight similarity indices were assessed under two different sampling conditions as well as for two communities of differing taxa abundance distributions in order to make recommendations for the use of available similarity indices for microbial community ecology.

Test 1: Undersampling

The performance of two qualitative similarity indices was tested when faced with undersampling of a community. If a similarity index were not affected by undersampling, then the value would be 1 indicating that the reduced sample and the full community are identical. The lower the similarity index between the reduced sample and the full community indicates a larger influence of undersampling on the index. It was predicted that the similarity indices between samples would decrease as the community was undersampled because two highly undersampled communities would appear to share fewer taxa than they do in reality and these two qualitative indices do not include corrections for unseen taxa. However, a response different from that seen for ant communities (Chao et al 2005) was observed for both the classic Jaccard and classic Sørensen indices (Figure 1). The similarity indices initially decreased as smaller samples were collected but both reached a threshold near a sample size of $1/8^{th}$ the total community. As sample sizes continued to decrease and undersampling intensified, the similarity indices tend to increase. An explanation for the unique response in bacterial communities is likely due to the low evenness of the communities. Bacterial communities are often dominated by few abundant taxa while most taxa appear only once or twice (Sogin et al. 2006), so the chance of detecting an individual belonging to an abundant taxon is much more likely than detecting one from a rare taxon. As the sample sizes decrease, individuals belonging to abundant taxa will be detected thus artificially increasing the similarity index between two grossly

undersampled communities. This response may be unique to bacterial communities with highly skewed species abundance distributions, and in insect communities the similarity index decreases linearly as the communities are increasingly undersampled (Chao et al 2006). Although both incidence-based indices were affected by undersampling, the classic Sørensen index appeared to perform slightly better than the classic Jaccard index as the similarity values overall were closer to one for the Sørensen index.

Abundance-based estimators of community similarity were designed to overcome the challenge of undersampling (Chao et al. 2006). Six commonly used abundance-based similarity indices were tested, including two indices corrected for unseen individuals, to assess how the indices performed when they encountered undersampling (Figure 3). The Morista-Horn index was closest to 1 for all undersampling levels. However, this index is not widely used among microbial ecology studies because it is affected mainly by the abundances of the most abundant taxa and is not strongly influenced by richness and sample size (Magurran 2004). Therefore, in hyper-diverse microbial communities that are dominated by a few taxa, the relatively rare taxa have very little affect in the index (Chao et al. 2006). Thus, although the Morista-Horn index performed well in this test of undersampling, it may not be able to distinguish between samples from two different communities with similar abundant taxa but very different rare taxa. The Sørensen abundance-based estimator with correction for unseen taxa also performed well in the undersampling test, a result consistent with a similar study using insect data (Chao et al. 2005). This Sørensen index includes information for less abundant taxa as well as uses probability to estimate for the taxa undetected in a sample (Chao et al. 2006). Unsurprisingly, the abundance-based Jaccard index with estimation for unseen taxa performed better than the raw Jaccard abundance-based index, as the correction should improve the index's performance in

undersampled communities. The Bray-Curtis index, which is commonly used in microbial community studies, performed very poorly at all levels of undersampling.

Test 2: Effect of uneven sample sizes

Uneven sample sizes are common for collecting community data and particularly in bacterial studies where DNA sequencing approaches are applied. To understand how uneven samples sizes affect similarity indices, pairs of samples with uneven numbers of sequences subsampled from a single bacterial community were compared. The two incidence-based similarity indices showed very similar curves and are both negatively affected by uneven sampling bias (Figure 2). Regardless of the substantial effects of uneven sample sizes, the classic Sørensen similarity index outperformed the classic Jaccard. These results are consistent with similar tests performed using insect data rather than bacterial data (Chao et al. 2005), though the negative effect appears to be more substantial for bacterial communities than for insect communities.

The abundance-based Sørensen index with correction for estimated unseen taxa outperformed the other abundance-based indices, though the Morista-Horn index was nearly as accurate (Figure 4). The Bray-Curtis index was highly affected by uneven sample sizes and nearly fell to zero when the full community was compared to the smallest, $1/64^{th}$ sample (Figure 4). The abundance-based Jaccard index performed better with the correction for estimated unseen taxa (Figure 4) than it did without the correction, similar to the performance of both abundance-based Sørensen indices.

Uneven sample sizes for bacterial community studies may not be avoidable in data collection but can be addressed by community rarefaction where samples are each subsampled to an even size (Horner-Devine et al. 2004). There is debate as to how samples should be rarefied

and the method of subsampling varies across microbial studies. For example, some studies randomly subsample all samples to the size of the smallest dataset using a program called DaisyChopper (Gilbert et al. 2009), and other studies perform community rarefaction in which they repeat the random subsampling approach to calculate mean diversity indices (Horner-Devine et al. 2004). A single subsample from a community has the opportunity for stochastic events to affect the similarity index. Whereas, when multiple subsamples are used to find an average similarity index the effect of these stochastic events is reduced, thus leading to a more accurate similarity value. Though either method may not accurately represent the true community composition, multiple subsampling is preferred, and it is essential to have even sample sizes to reduce the variation in similarity indices (Figure 4).

Test 3: Effect of evenness on similarity indices

Another important aspect of community structure that can differ between bacterial samples and have an effect on similarity indices is evenness in the distribution of taxa abundances (Figure 5). It was hypothesized that many of the similarity indices would perform differently for bacterial communities with different evenness of species abundances because the calculations for the indices may include or exclude rare taxa as well as estimate shared unseen taxa by the presence of rare taxa (Chao et al. 2006). The results of this study suggest that the structure of communities including the diversity and evenness of taxa can strongly influence the similarity indices both in the presence of undersampling and uneven sampling (Figures 1-4). The similarity indices calculated for Community B, which had more observed species and higher evenness than both Community A and the ant community presented in Chao et al (2005, 2006), were more strongly impacted by undersampling and uneven sample sizes than Community A and the ant community. Community structure was so important in measuring the performance of

incidence-based similarity indices for undersampling, that the shape of the curve was considerably different for each of the three communities that were examined (Figure 1).

Conclusion

Here, the performance of two qualitative and six quantitative indices of community similarity under two common biases in bacterial community sampling: undersampling and uneven sample sizes were investigated. A recently developed abundance-based estimator of Sørensen similarity performed best when faced with extremely uneven sample sizes. The robust nature of the abundance-based estimated Sørensen index in comparing uneven sample sizes suggests that it is best suited for comparing bacterial communities that are not rarefied to an even sample size. The very commonly used Bray-Curtis similarity index performed poorly in both tests of undersampling and uneven sampling, suggesting that is not suitable for use in bacterial community ecology. Further, the examination of two different bacterial community structures shows that not only do sampling biases (undersampling and uneven sampling) affect beta diversity measures, but the evenness and diversity of the community also influences the performance of these measures. Collectively, these results suggest that incidence-based measures are not preferred for comparing communities of microbes that are often undersampled or unevenly sampled. Although the use of abundance data is preferred, it is not clear whether undersampling or uneven sample sizes affect the rank-order of incidence-based similarity indices. Further studies comparing the community composition of multiple simulated bacterial communities are required to examine this question.

References:

Beck J, Holloway JD, Schwanghart W. (2013). Undersampling and the measurement of beta diversity Orme, D, ed. Methods in Ecology and Evolution 4:370–382.

Bray J, Curtis C. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs 27:325–349.

Cardoso P, Borges PA V., Veech JA. (2009). Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. Diversity and Distributions 15:1081–1090.

Chao A, Chazdon RL, Colwell RK, Shen T-J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecology Letters 8:148–159.

Chao A, Chazdon RL, Colwell RK, Shen T-J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361–71.

Colwell RK, Coddington JA. (1994). Estimating terrestrial biodiversity through extrapolation. Philosophical transactions of the Royal Society 345:101–118.

Downing AL, Leibold MA. (2002). Ecosystem consequences of species richness and composition in pond food webs. Nature 416:837–41.

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 11:3132–9.

Hooper DU, Chapin, III FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, et al. (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75:3–35.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM. (2004). A taxa-area relationship for bacteria. Nature 432:750–753.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. (2001). Counting the Uncountable□: Statistical Approaches to Estimating Microbial Diversity. Applied and environmental microbiology 67:4399–4406.

Jaccard P. (1908). Nouvelles recerches sur la distribution florale. Bulletin de la Société vaudoise des sciences naturelles 44:223–270.

Koleff P, Gaston KJ, Lennon JJ. (2003). Measuring beta diversity for presence-absence data. Journal of Animal Ecology 72:367–382.

Lange RK, Williams C, Rocap G, Horner-Devine MC. Patterns of bacterial richness and community composition in Hood Canal, WA, a seasonally oxygen-deficient estuary. ISME Journal, *In review*.

Lemos LN, Fulthorpe RR, Roesch LFW. (2012). Low sequencing efforts bias analyses of shared taxa in microbial communities. Folia microbiologica 57:409–13.

Magurran A. (2003). Measuring Biological Diversity. Blackwell Publishing: Malden, MA.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." Proceedings of the National Academy of Sciences 103:12115–20.

Sorensen T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol. Skr 5:1–34.

Whittaker RH. (1962). Classification of natural communities. Botanical Review 28:1–239.

Whittaker RH. (1972). Evolution and measurement of species diversity. Taxon 21:213–251.

Whittaker RJ, Willis KJ, Field R. (2001). Scale and species richness□: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453–470.

Figures

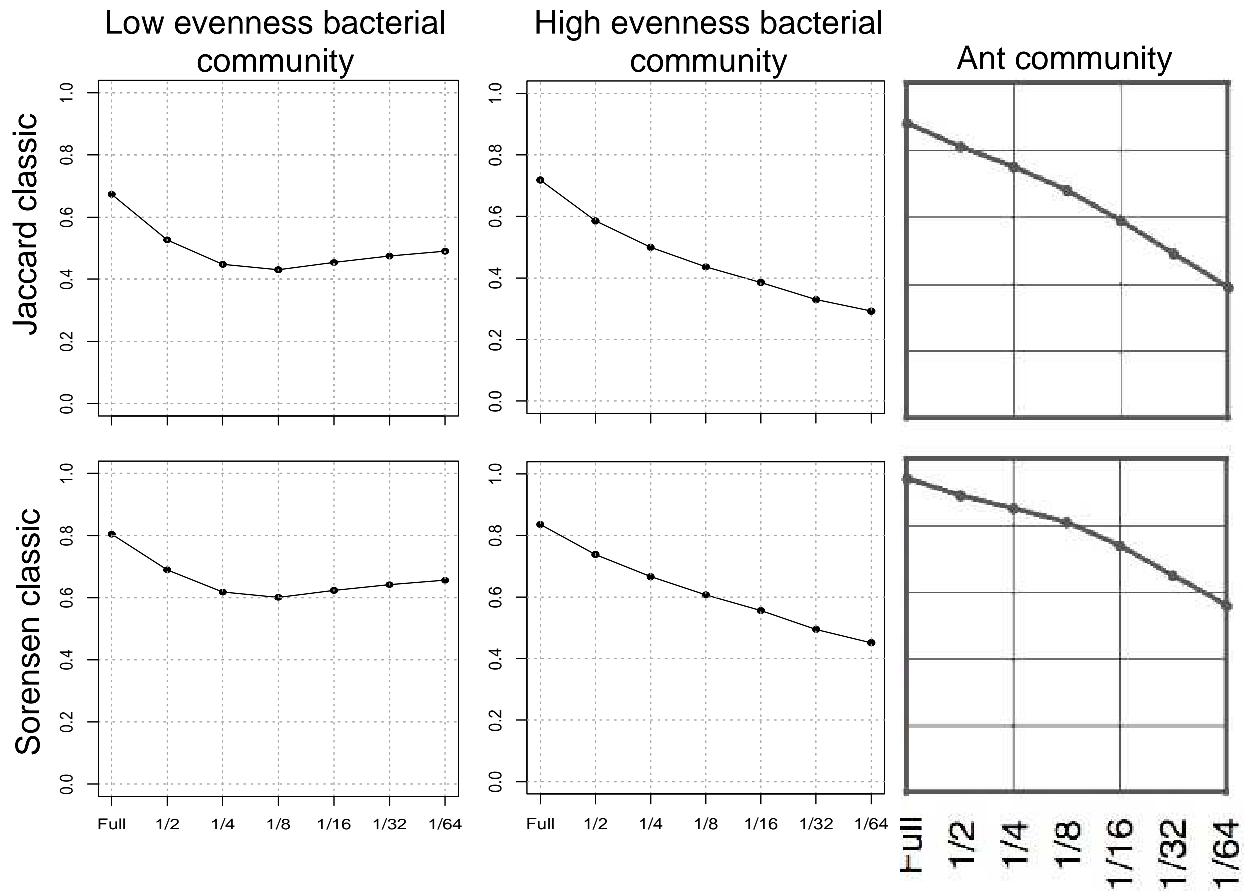


Figure 1. Performance of the Jaccard classic and Sørensen classic similarity indices in undersampling simulations, based on incidence data from three different biologic communities. The structure of the biologic community affects the performance of the incidence-based similarity indices in any undersampling situation. Overall, the Sørensen classic index outperforms the Jaccard similarity index when communities are undersampled, regardless of the community structure. Ant community plots adapted from Chao et al 2006.

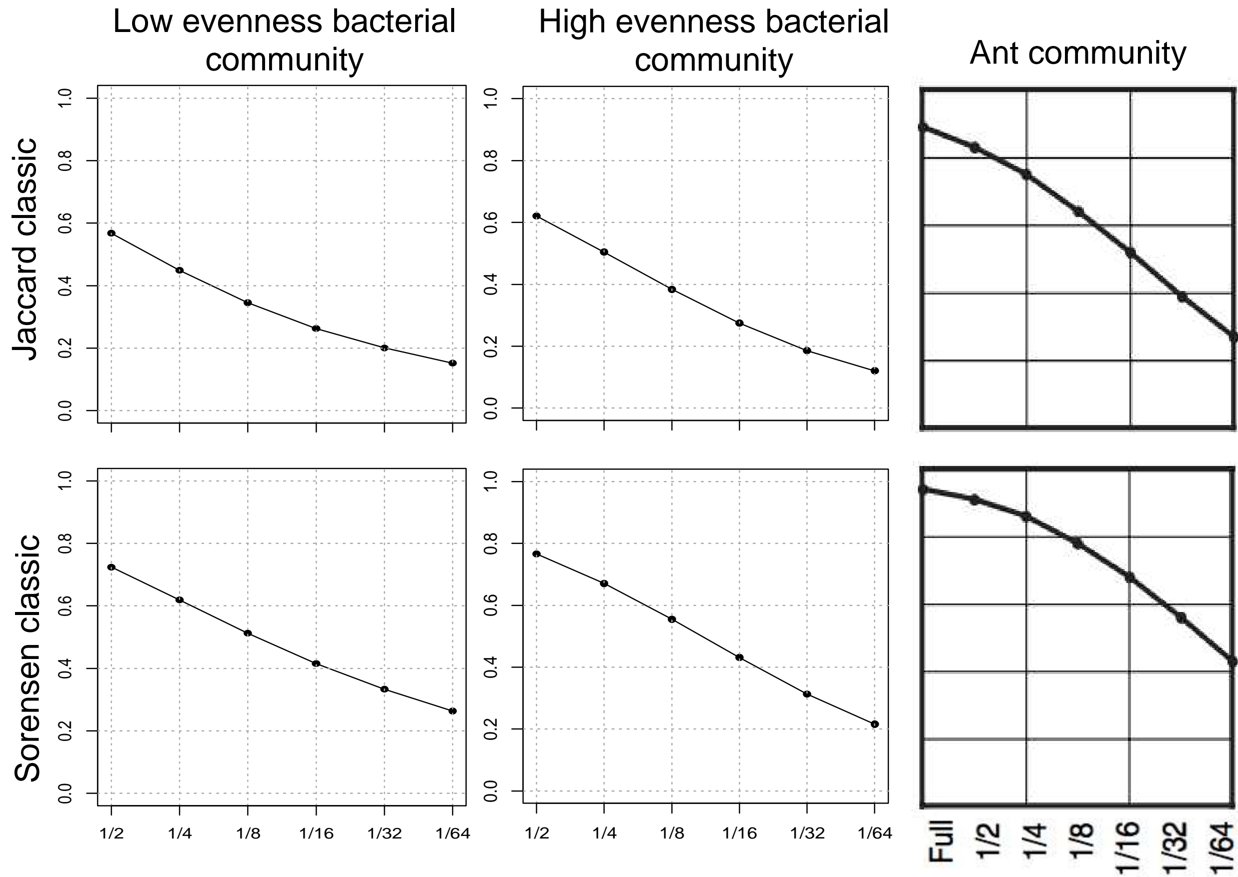


Figure 2. Performance of Jaccard classic and Sørensen classic similarity indices when sample sizes are uneven. Community structure has a weak effect on the overall performance of the indices. Both indices are largely influenced by uneven sample sizes, though the Sørensen classic index produces values closer to 1, which suggests less of an influence of uneven sample size over the Jaccard classic index. Ant community plots adapted from Chao et al 2006.

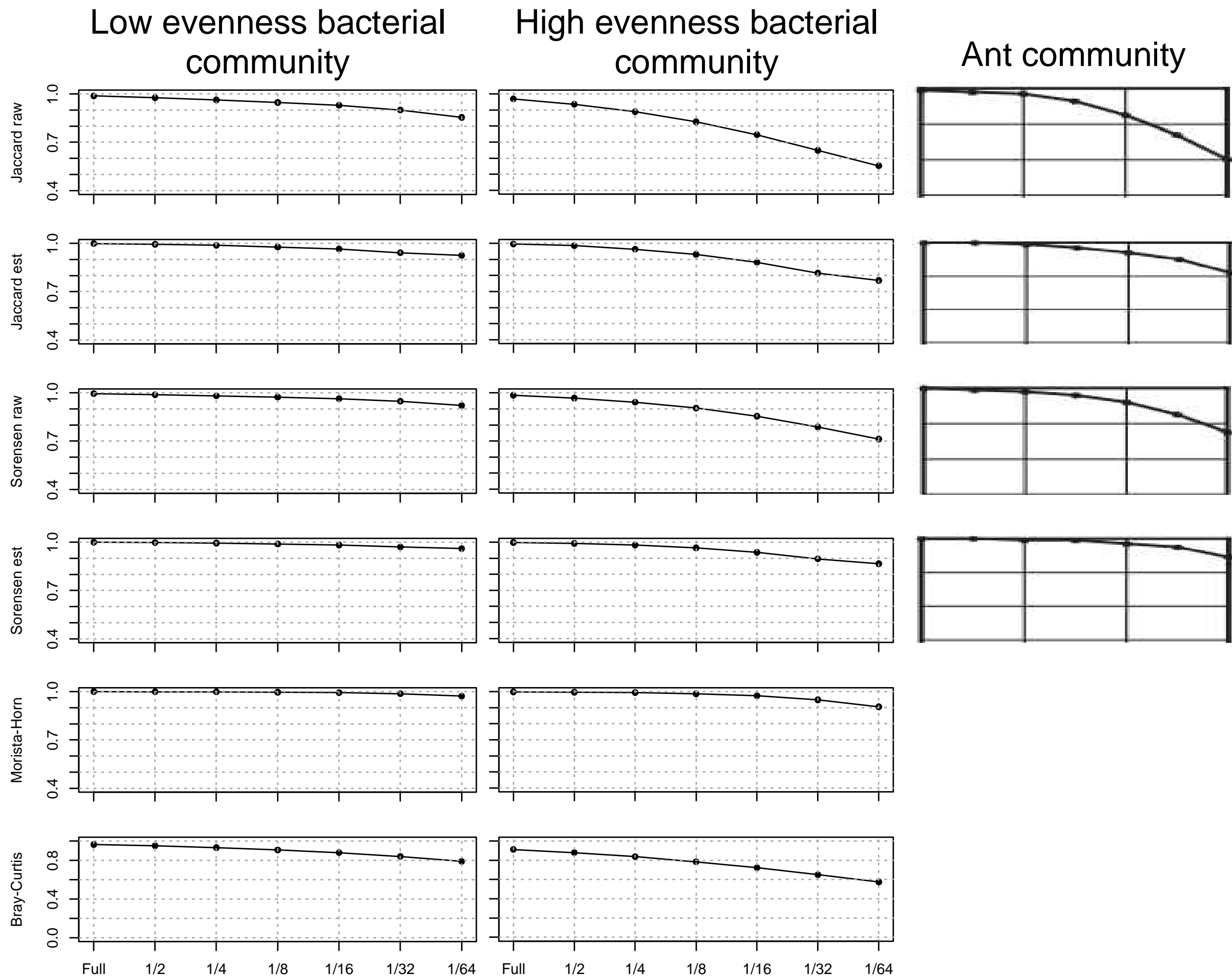


Figure 3. Performance of six abundance-based similarity indices in undersampling simulations for three different biologic communities. Community structure influences the performance of all the abundance-based indices in undersampling conditions. The abundance-based Sørensen estimator and the Morista-Horn indices both outperform the other similarity indices and are less influenced by undersampling. Ant community plots adapted from Chao et al 2006.

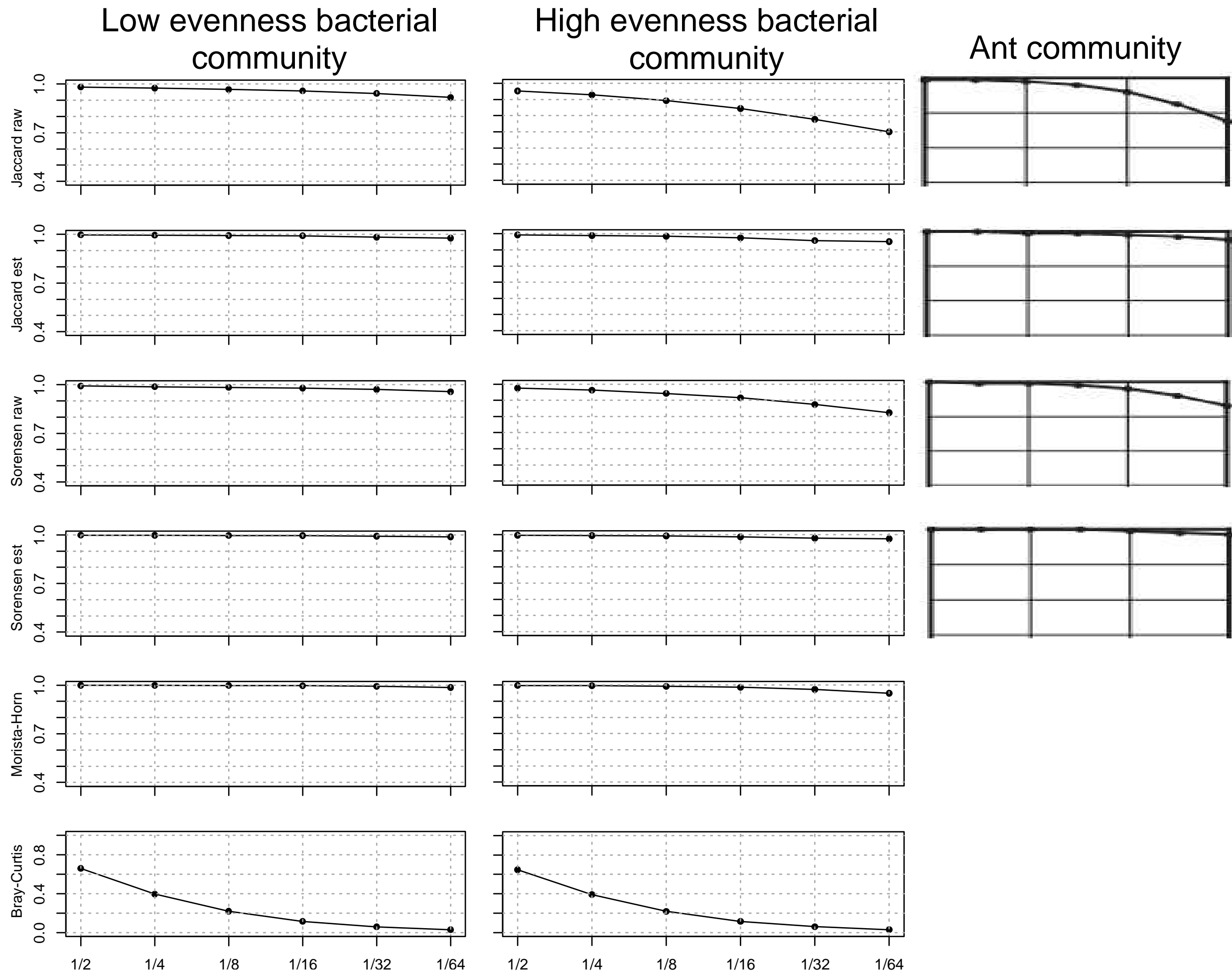


Figure 4. Performance of six abundance-based similarity indices in simulations with uneven sample sizes for three different biologic communities. Community structure also affects the abundance-based similarity indices when sample sizes are uneven. Sørensen abundance-based estimator and Jaccard abundance-based estimator outperform the other similarity indices when sample sizes are uneven. Ant community plots adapted from Chao et al 2006.

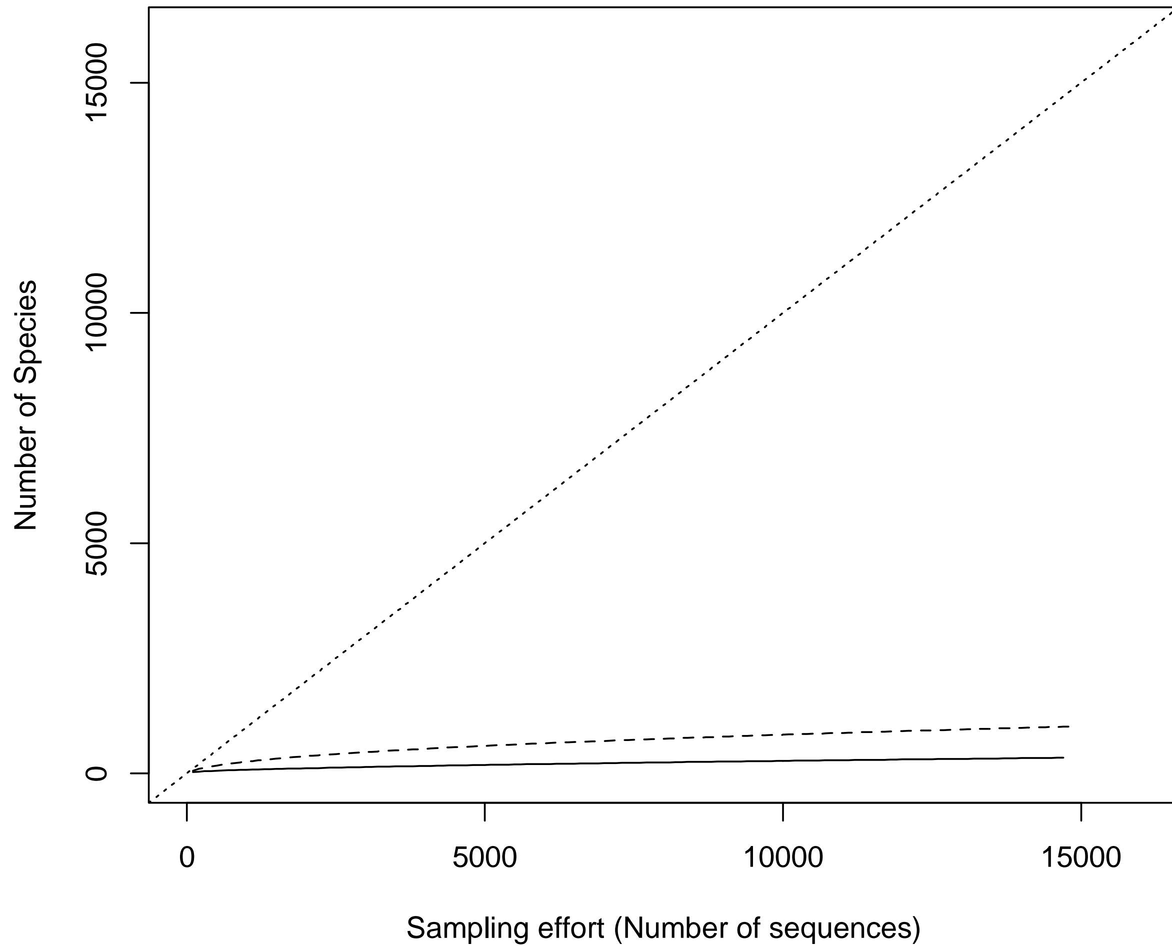


Figure 5. Species accumulation curves for the two bacterial communities used for the simulation studies. The solid line represents Community A, which had higher evenness of species abundances. The dashed line represents Community B, which had low evenness of species abundances. The dotted line is a reference for 1:1 sampling effort to species accumulation.