Microbial community structure and ecology of Marine Group A bacteria in the oxygen minimum zone of the Northeast subarctic Pacific Ocean

by

JODY JENNIFER WRIGHT

BSc. Simon Fraser University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Microbiology and Immunology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2013

Abstract

Oxygen minimum zones (OMZs) are intrinsic water column features that arise when the respiratory oxygen (O_2) demand during microbial remineralization of organic matter exceeds O_2 supply rates in poorly ventilated regions of the ocean. Microbial processes play a key role in mediating biogeochemical cycling of nutrients and radiatively active trace gases in OMZs. Specific roles of individual microbial groups and the ecological interactions among groups that drive OMZ biogeochemistry on a global scale, however, remain poorly constrained. This dissertation focuses on describing microbial community structure in the world's largest and least studied OMZ, located in the Northeast subarctic Pacific Ocean (NESAP), with a specific emphasis on characterizing the ecology of Marine Group A, an uncultivated candidate phylum of bacteria found to be prevalent in this region. To begin, I performed a survey of microbial community structure in the NESAP at two time points and over a range of depths based on traditional ecological analyses. I applied techniques derived from network theory to identify co-occurrence patterns among microbial groups within the NESAP and determined that MGA bacteria most frequently co-occurred with other MGA bacteria, suggesting that intra-phylum interactions may play a role in governing microbial processes in this region. Through analysis of small subunit ribosomal rRNA (SSU rRNA) gene sequences affiliated with MGA, I identified 8 novel subgroups and established the phylogeny and population structure of both novel and previously detected MGA subgroups. Finally, I provided first insights into the metabolic capacity of this little-known candidate phylum through investigations of metagenomic data obtained from NESAP waters. Analysis of large-insert genomic DNA fragments derived from MGA revealed protein-coding genes associated with adaptation to oxygen deficiency and sulfur-based energy metabolism. These observations may implicate MGA bacteria in the cryptic sulfur cycle, recently discovered to play a central role in biogeochemical cycling within OMZs. This work describes the first survey of microbial community structure in the NESAP OMZ and the first application of co-occurrence networks to study the ecology of deep ocean microbial communities, in addition to the first analysis of the diversity, population structure, and metabolic capacity of the enigmatic bacterial lineage MGA.

Preface

The work presented in this dissertation would not have been possible without contributions made by collaborators, contractors, and undergraduate students, as described in the following paragraphs. As research advisor, Dr. Steven Hallam was involved in all aspects of this work, including conceptualization of the experiments, analysis of results, and manuscript writing.

In chapter 1, I generated and analysed the small subunit ribosomal rRNA (SSU rRNA) gene clone libraries from Northeast subarctic Pacific Ocean samples with the assistance of Olena Shevchuk. The construction of microbial co-occurrence networks was performed by Dr. Kishori Konwar with my input and interpretation. I performed the literature review, generated figures, and drafted the initial manuscript comprising the majority of chapter 1. The summary of archaeal community structure in oxygen minimum zones was conceived in collaboration with Dr. Osvaldo Ulloa at the Universidad de Concepcion (UdeC), Concepcion, Chile. Dr. Ulloa and I both drafted portions of the initial manuscript presented in sections 1.3.2 – 1.3.2.3 and 1.4.2 – 1.4.4. The phylogenetic analysis of archaeal SSU rRNA gene sequences presented in Figure 1.6 was performed by Dr. Lucy Belmar at UdeC, and I contributed to the development of the illustration. I performed analysis for and generated Figure 1.7 in this section. Young Song contributed to the development of Figure 1.5.

In chapter 2, I collected and prepared samples for 454-pyrotag sequencing with the assistance of Drs. Anissa Merzouk and Kendra Maas. 454-pyrotag sequencing was performed at the Department of Energy's Joint Genome Institute (JGI) in Walnut Creek, CA, USA as part of a community sequencing proposal. The construction of microbial co-occurrence networks was performed by Dr. Kishori Konwar with my input and interpretation. Drs. William Orsi and Virginia Edgcomb of the Woods Hole Oceanographic Institution provided input on taxonomic identification of eukaryotic sequences. I performed all other analyses with input from Dr. Konwar, Aria Hahn, and Sarah Perez. I generated all of the figures for chapter 2 with input from Martin Krzywinski, and wrote the manuscript.

In chapter 3, the CARD-FISH study was conceived by Drs. Elke Allers and Matthew Sullivan at the University of Arizona, Tucson, AZ, USA, and was carried out in the laboratory of Dr. Sullivan under the supervision of Dr. Allers. Dr. Allers designed the oligonucleotide probes and performed all CARD-FISH procedures, in addition to measuring chlorophyll a and DAPI counts of prokaryotic cells in seawater samples. She also made contributions to statistical analysis of data. Nutrient analysis of seawater samples was performed by Department of Fisheries and Oceans scientists aboard the *CCGS John P. Tully*. I collected seawater samples for DNA extraction with the assistance of Dr. Allers and Dr. Kendra Mitchell, performed DNA extraction, generated SSU rRNA gene clone libraries and prepared clones for sequencing, prepared DNA samples for 454-pyrotag sequencing, analysed sequence data, performed statistical analysis, and wrote the manuscript. Drs. Allers and Sullivan both provided constructive feedback on the manuscript. Sanger sequencing and 454-pyrotag sequencing was performed at the JGI.

In chapter 4, fosmid libraries were constructed by Drs. David Walsh and Kendra Maas at the University of British Columbia (UBC), and full-length sequencing and sequence assembly of several fosmids was performed by Keith Mewis at UBC. Sequencing of fosmid ends and remaining full-length fosmids was performed at the JGI, and at Canada's Michael Smith Genome Sciences Centre (GSC), Vancouver, BC. I collected seawater samples for DNA extraction with the assistance of Dr. Maas, performed all sequence analyses, generated figures, and wrote the manuscript. I received input on the figures from Martin Krzywinski.

In chapters 1 – 4, I performed all other experimental work, data analysis, and manuscript writing, with the input of my advisor, Dr. Steven Hallam.

Elements of chapter 1 have been published in two locations (in one review paper and one book chapter):

Wright, J. J., Konwar, K.M., & Hallam, S.J. (2012). Microbial ecology of expanding oxygen minimum zones. *Nature Reviews Microbiology* (Vol. 10, pp. 381-394). doi:10.1038/nrmicro277

Ulloa, O., **Wright, J.J.,** Belmar, L., & Hallam, S.J. Pelagic Oxygen Minimum Zone Microbial Communities. In E. Rosenberg et al. (eds.), *The Prokaryotes – Prokaryotic Communities and Ecophysiology,* DOI 10.1007/978-3-642-30123-0_45, © Springer-Verlag Berlin Heidelberg 2013

Data presented in chapter 1 was also included in a third publication:
Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., **Wright, J.J.**, Tringe, S.G., Tortell, P.D., and Hallam, S.J. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science 326*, 578-582

Methods for DNA extraction applied in chapters 1 – 4 have been published:
Wright, J.J., Lee S., Zaikova, E., Walsh, D.A., Hallam, S.J. (2009). DNA extraction from 0.22 microM Sterivex filters and cesium chloride density gradient centrifugation. *J Vis Exp*. 31: pii 1352

A version of chapter 3 has been published:
Allers, E.*, **Wright, J.J.***, Konwar, K.M., Howes, C.G., Beneze, E., Hallam, S.J., & Sullivan, M.B. (2012). Diversity and population structure of Marine Group A bacteria in the northeast subarctic pacific ocean. *The ISME Journal*, *7*, 256-268. doi:10.1038/ismej.2012.10
*Co-first authors

A version of chapter 4 has been accepted for publication:
Wright, J.J., Mewis, K., Hanson, N.W., Konwar, K.M., Maas, K.R., Hallam, S.J. (2013) Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *The ISME Journal. Accepted July 1st, 2013.*

A version of chapter 2 is in preparation for submission to a peer-reviewed journal.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

16S	SSU rRNA in bacteria and archaea
18S	SSU rRNA in eukaryotes
µM	micromolar
nM	nanomolar
AOA	Ammonia oxidizing archaea
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BB	Bacterial biomass
BGR	Bacterial growth rate
bp	base pair
BP	Bacterial production
CARD-FISH	Catalyzed Reporter Deposition Fluoresence In Situ Hybridization
DAPI	4',6-diamidino-2-phenylindole, a fluorescent DNA-binding stain
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNRA	Dissimilatory nitrate reduction to ammonia
DOM	Dissolved organic matter
ENSO	El Niño Southern Oscillation
GC	Guanine-cytosine
HCA	Hierarchical cluster analysis
HISH-SIMS	Halogen In Situ Hybridization-Secondary Ion Mass Spectroscopy
HNLC	High nutrient low chlorophyll
HOT	Hawaii Ocean Time-series
LSA	Local similarity analysis
M	Molar
MAR-CARD-FISH	Microautoradiography catalysed reporter deposition fluorescence *in situ* hybridization
MGA	Marine Group A
NADH	Nicotinamide adenine dinucleotide
NESAP	Northeast subarctic Pacific Ocean
NPIW	North Pacific Intermediate Waters
OMZ	Oxygen minimum zone
ORF	Open reading frame
PCR	Polymerase chain reaction
PDO	Pacific Decadal Oscillation
POM	Particulate organic matter
QIIME	Quantitative insights into microbial ecology
rRNA	Ribosomal ribonucleic acid
SI	Saanich Inlet

SPOTS	San Pedro Ocean Time Series
SSU	Small subunit
SSW	Subtropical Subsurface Waters
tRNA	Transfer ribonucleic acid

Acknowledgements

I would like to acknowledge my advisor, Dr. Steven Hallam, who accepted me as the first doctoral student in his young laboratory to work on this challenging and interesting project. I learned a great deal about science and life from our interactions, and I thank him for his guidance, expertise, and support. In particular I am grateful to Steven for his many generous investments in my scientific and personal development, even when the benefits of these investments may have been more tangible for me personally than for the lab.

To my wonderful lab-mates, past and present, I am grateful for all of your support, assistance, and constructive feedback over the past years, and I am proud to have been a part of the development of such a tightly knit and collegial group. I am infinitely grateful to the captains, crews, and scientists aboard the *CCGS John P. Tully* and to chief scientist Marie Robert, without whose assistance this work would not have been possible. I offer many thanks to all of my friends and peers, the staff, and the faculty in the Department of Microbiology and Immunology (in particular Darlene Birkenhead, Dr. Michael Gold, Dr. Michael Murphy, and Susan Palichuk) for creating such a positive working environment. I am grateful to my supervisory committee members Dr. Leonard Foster, Dr. Carl Hansen, and Dr. Philippe Tortell for their constructive comments, sincere encouragement, and insightful criticism, which has made me a better scientist and helped me to grow as a person. I am grateful to my mentors, Dr. Joanne Fox, Dr. Joanne Nakonechney, Dr. Lynne Quarmby, Dr. Rosie Redfield, and Dr. Lacey Samuels, for inspiring me to continue pursuing research, and for their generous listening and advice during times of struggle. I was fortunate to have received financial support throughout the majority of my doctoral program, and gratefully acknowledge the aid provided by the University of British Columbia and the Natural Sciences and Engineering Research Council.

Finally, I could not have made it to this stage of my life (or through this degree) without the enduring love and support provided by my family. To my parents, David and Wendy, my parents-in-law, Stratis and Carol, and my brother and sister-in law, Sandy and Kari, thank you for always believing in me, for encouraging me to pursue my dreams, and for reminding me to relax and not take life so seriously. To my partner and best friend, Costa, thank you so much for your love and support, and for teaching me how to live and enjoy life to its fullest every day.

For Josh and Kim

Chapter 1: Introduction to microbial ecology of expanding oxygen minimum zones[1]

1.1 Synopsis

Dissolved oxygen concentration is a crucial organizing principle in marine ecosystems. As oxygen levels decline, energy is increasingly diverted away from higher trophic levels into microbial metabolism, leading to loss of fixed nitrogen and to production of greenhouse gases, including nitrous oxide and methane. In this Chapter, I describe current efforts to explore the fundamental factors that control the ecological and microbial diversity in oxygen-starved regions of the ocean, termed oxygen minimum zones. I also discuss how recent advances in microbial ecology have provided information about the potential interactions in distributed co-occurrence and metabolic networks (graphic visualizations of potential interactions among microbes and their metabolisms) in oxygen minimum zones, and provide new insights into coupled biogeochemical processes in the ocean.

1.2 Oxygen minimum zones: important bellwethers for global change

Over geological time the ocean has evolved from being an anaerobic incubator of early cellular existence into a solar-powered emitter of molecular oxygen (O_2), a transformation that has been punctuated by catastrophic extinctions followed by the iterative re-emergence of biological diversity (Kasting and Siefert, 2002; Falkowski et al., 2008). Today, the ocean is being transformed in response to human activities. Indeed, the fourth assessment report of the Intergovernmental Panel on Climate Change observed that the ocean is becoming substantially warmer and more acidic (Kumar, 2007). As these changes intensify, marine ecosystems will experience disturbances in the structure and dynamics of food webs, with resulting feedback on

[1] Sections of this chapter have been published in:
Wright, J.J., Konwar, K.M., & Hallam, S.J. (2012). Microbial ecology of expanding oxygen minimum zones. *Nature Reviews Microbiology*. 10, 381-394

Ulloa, O., **Wright, J.J.,** Belmar, L., & Hallam, S.J. (2013) Pelagic Oxygen Minimum Zone Microbial Communities. In E. Rosenberg et al. (eds.), *The Prokaryotes – Prokaryotic Communities and Ecophysiology,* Springer-Verlag Berlin Heidelberg

the climate system (Doney, 2010). Oxygen-starved regions of the ocean, known as oxygen minimum zones (OMZs), are important bellwethers for these changes (Falkowski et al., 2011). OMZs are an intrinsic feature of water columns that arise when the respiratory O_2 demand during the remineralization of organic matter exceeds O_2 supply rates in poorly ventilated regions of the ocean (Paulmier et al., 2008; Helm et al., 2011; Karstensen et al., 2008). Increases in ocean temperature drive decreases in O_2 solubility and reduced ventilation owing to thermal stratification of the water column (Helm et al., 2011; Keeling et al., 2010), resulting in OMZ expansion. Consistent with this, between 1956 and 2006 the O_2 concentrations in the OMZ of the Northeast subarctic Pacific (NESAP) declined by 22%, and the hypoxic boundary layer (defined as ~60 μmol O_2 per kg water) shoaled upwards from a depth of 400 m to 300 m (Whitney et al., 2007). Similar declines have been observed in the eastern tropical Atlantic (Stramma et al., 2008), the equatorial (Stramma et al., 2008) and northeast (Bograd et al., 2008; Emerson et al., 2004) Pacific, and in the Southern Ocean (Helm et al., 2011) during the past 50 years.

1.2.1 OMZs as habitats for marine organisms

As O_2 concentrations decline, the amount of habitat available to aerobically respiring organisms in benthic ecosystems and pelagic ecosystems reduces, changing the species composition and food web structure in these regions (Rabalais et al., 2010). Organisms that are unable to escape O_2-deficient conditions may experience direct mortality (that is, the fish in these regions die) or decreased fitness (Breitburg et al., 2009; Vaquer-Sunyer and Duarte, 2008). Even organisms that can escape to more highly oxygenated refuges are susceptible to increased predation and density-dependent reductions in population size (Ekau et al., 2010). OMZ expansion also causes changes in the cycling of trace gases such as methane (CH_4), nitrous oxide (N_2O) and carbon dioxide (CO_2), which are important for metabolism and can have an effect on climate. CH_4 and N_2O are powerful greenhouse gases with radiative forcing effects that are approximately 25 and 300 times the effect of CO_2, respectively. Although oceanic CH_4 emissions are minor (<2% of natural CH_4 emissions), the ocean accounts for at least one-third of all natural N_2O emissions, a large fraction of which are derived from OMZs via microbial respiration of nitrate (NO_3^-) and nitrite (NO_2^-) (Naqvi et al., 2010). Moreover, OMZs account for up to 50% of oceanic fixed-nitrogen loss, and their expansion has the potential to affect primary production, with resulting feedback

on carbon transport processes (Codispoti et al., 2001; Gruber and Galloway, 2008; Lam et al., 2009; Ward et al., 2009). Although OMZs are inhospitable to aerobically respiring organisms, these zones support thriving microbial communities that mediate cycling of nutrients and radiatively active trace gases (which affect the climate). Therefore, systems-level investigations of microbial communities in the OMZ-containing water column have great potential to enhance our mechanistic understanding of a pervasive ecological phenomenon that is integral to ocean productivity and climate balance.

This chapter reviews recent observations that have emerged from the intersection of taxonomic and functional gene surveys, gene expression studies and measurements of process rates to better formulate hypotheses regarding the metabolic interactions that drive OMZ ecology and biogeochemistry on a global scale. This chapter focuses on bacterial and archaeal contributions to these networks, with the understanding that microbial eukaryotes and viruses have biologically essential but as-yet physiologically uncharacterized roles in modulating matter and energy transformations in OMZs.

1.2.2 OMZ formation and expansion

OMZs are typically found on the western boundaries of continental margins, where wind-driven circulation patterns push nutrient-rich waters upwards to the surface in a process known as coastal upwelling. This process effectively fertilizes surface waters and results in high levels of photosynthetic primary production. During photosynthesis, phytoplankton fix CO_2. Much of the inorganic carbon that is fixed through photosynthesis is respired in surface and intermediate layers of the water column through microbial remineralization processes. A fraction of the product of primary production sinks as dead organisms and particles that are exported to depth. Throughout the ocean this process, called the biological carbon pump, has a large influence on the biogeochemical carbon cycle because carbon is sequestered in the interior of the ocean for long periods of time, during which it cannot influence the climate (Siegenthaler and Sarmiento, 1993). Estimates of carbon rain rates to carbon sediments in the northeast Pacific suggest that the presence of an OMZ greatly increases the amount of carbon exported to the deep ocean (Devol and Hartnett, 2001).

Persistent O_2 deficiency occurs when the amount of dissolved O_2 in the water column is consumed faster than it is resupplied through air–sea exchange, photosynthetic O_2 production, and ventilation (Rabalais et al., 2010). Global circulation patterns transport younger, more oxygenated waters throughout the deep ocean, resulting in deep oxycline formation. Thus, in profile, OMZs resemble a band of O_2-deficient water inserted between two O_2-containing water masses (Figure 1.1). The upper O_2 thresholds chosen to define OMZs have been manifold, ranging from <2 μmol O_2 per kg water to 90 μmol O_2 per kg water (Paulmier et al., 2008). This dissertation adopts the criterion of <20 μmol O_2 per kg water, to include the maximum O_2 level at which the use of alternative electron acceptors (in this case, NO_3^-) have been reported (Smethie Jr, 1987). Using this definition, OMZs currently constitute 1–7% of the volume of the global ocean, occupying approximately 102 million km^3 (Paulmier et al., 2008; Lam et al., 2009; Ulloa and Pantoja, 2009; Fuenzalida et al., 2009) (Figure 1.1).

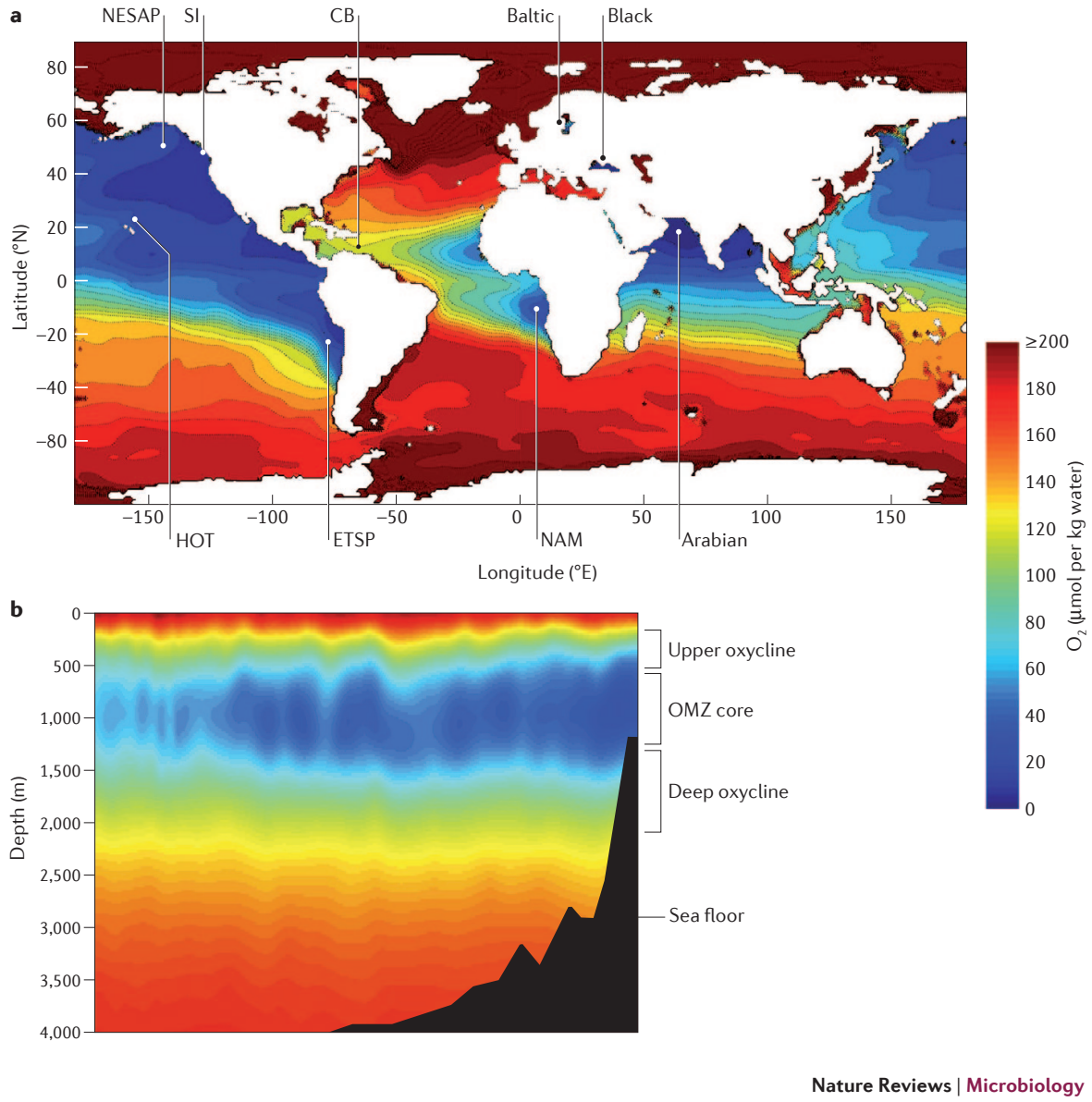


Figure 1.1 O_2 concentrations in the ocean

a. Minimum molecular oxygen (O_2) concentrations for different regions of the ocean. Locations highlighted in this chapter are indicated and comprise the Hawaii Ocean Time-series (HOT), the Northeast subarctic Pacific (NESAP), Saanich Inlet (SI), the eastern tropical South Pacific (ETSP), the Cariaco Basin (CB), the Namibian upwelling (NAM; also known as the Benguela upwelling), and the Baltic, Black, and Arabian seas. Oxygen data were derived from Garcia et al., (2010). **b.** Cross-section of the NESAP oxygen minimum zone (OMZ), showing the O_2 concentration from surface waters to the sea floor. Upper oxycline: transition from surface waters to the OMZ core. OMZ core: defined by O_2 concentrations <20 µmol per kg water. Deep oxycline: transition from the bottom of the OMZ core to abyssal waters. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.2.3 Geography of OMZs

Geographically, OMZs occur in the Pacific Ocean (in the NESAP, off western North America; the eastern tropical North Pacific (ETNP), off Mexico; and the eastern tropical South Pacific (ETSP), off Peru and Chile), the Atlantic Ocean (in the Northwest-African upwelling and the Namibian or Benguela upwelling) and the Arabian Sea (Figure 1.1). The Pacific OMZs are more voluminous than those in the Atlantic Ocean and the Arabian Sea. This is due to decreased ventilation at high latitudes in the western Pacific as well as to the length of time that these waters have been isolated from the atmosphere (a result of global circulation patterns). However, O_2 deficiency is often more intense in the Arabian Sea and in coastal Atlantic waters on the African shelf than in the Pacific Ocean owing to unusually high levels of carbon export and sub-surface respiration in these naturally eutrophic waters of the Arabian Sea and Atlantic Ocean (Estrada and Marrasé, 1987). Compounding the effects of respiratory demands for O_2, many upwelling systems experience episodic plumes of hydrogen sulfide (H_2S) that can be attributed to diffusive flux from underlying sediments (Canfield, 2006). Such sulfidic events are toxic to most O_2-respiring organisms. In addition to coastal and open-ocean OMZs, enclosed or semi-enclosed basins - including the Baltic Sea (Conley et al., 2002), Black Sea (Jørgensen, 1982), Cariaco Basin (Scranton et al., 2001), and Saanich Inlet (Anderson and Devol, 1973) - experience varying degrees of O_2 deficiency and sulfide accumulation, making them useful model ecosystems for exploring microbial community responses to OMZ expansion and intensification.

1.2.4 O_2 deficient shallow coastal and estuarine environments

Human activities exacerbate the natural O_2 deficiency in shallow coastal and estuarine environments, where nutrient run-off from agricultural and wastewater sources results in eutrophication (Diaz and Rosenberg, 2008). Moreover, changes in wind-driven circulation patterns can induce upwelling of O_2-deficient waters from coastal OMZs onto continental shelves, increasing mortality of shelf-dwelling organisms (Helly and Levin, 2004). Over the past two decades, shelf intrusions have produced ‘dead-zones’ off coastal Oregon (Grantham et al.,

2004) and in the Gulf of Mexico (Rabalais et al., 2001) (USA), and off the coast of Chile (Fuenzalida et al., 2009), Africa (Monteiro et al., 2008), and India (Naqvi et al., 1998), contributing to a drop in production from commercial fisheries (Diaz and Rosenberg, 2008). Regardless of the water body (estuary, basin, coastal waters, or open ocean), O_2 deficiency shifts energy away from pelagic macrofauna towards microorganisms, decoupling predator–prey interactions and changing the trophic exchanges that occur through existing food webs (Figure 1.2).

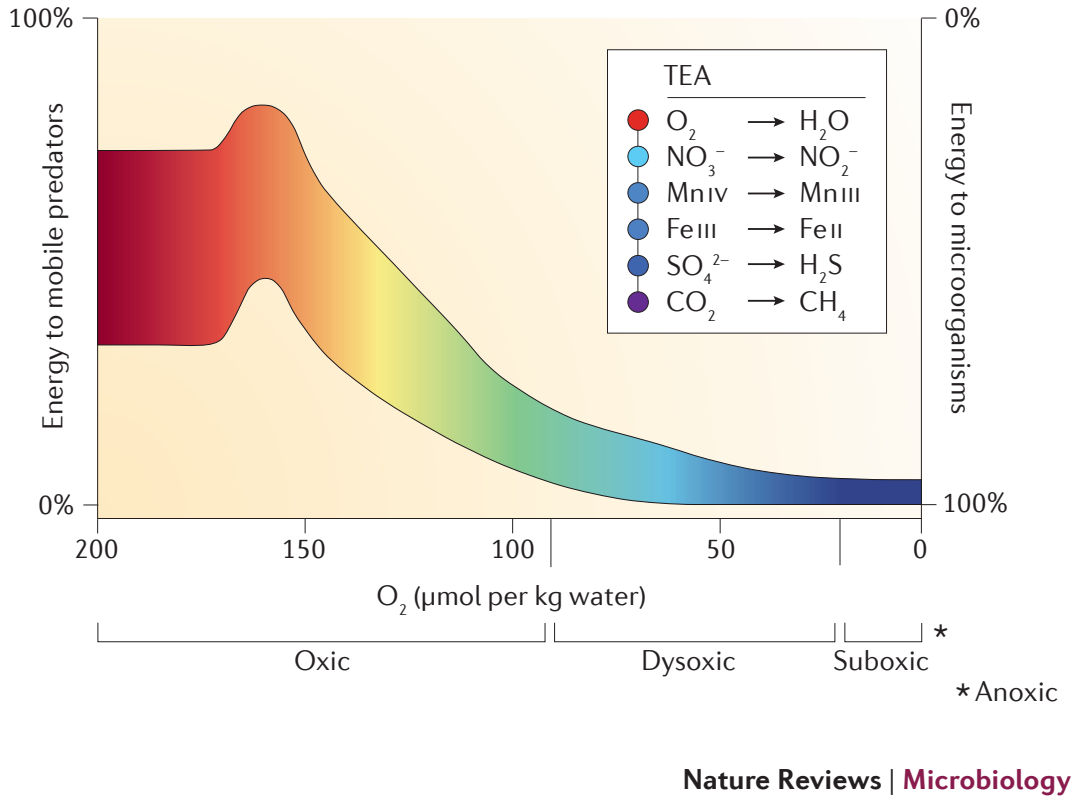


Figure 1.2 O_2 concentration affects ecosystem energy flow

Alternative states of the seawater (oxic, dysoxic, suboxic and anoxic) and corresponding molecular (O_2) concentrations are defined. The red-orange area indicates the range of energy transferred from pelagic nutrients to higher-level predators under oxic conditions. With declining O_2, higher-level predation is suspended and the proportion of energy transferred to microorganisms rapidly increases (the yellow-green-blue area). This energy is generated via microbial respiration using a defined order of terminal electron acceptors (TEAs), with O_2 as the preferred TEA, followed by nitrate (NO_3^-), manganese IV (Mn IV), iron III (Fe III), sulfate (SO_4^{2-}) and, finally, carbon dioxide (CO_2). Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.3 Microbial energetics in OMZs

Under oxic conditions (>90 μmol O_2 per kg water), 25–75% of the energy generated via oceanic primary production is transferred to mobile predators (Diaz and Rosenberg, 2008). As O_2 levels decline, aerobic organisms escape to more oxygenated refuges, resulting in habitat compression and a concomitant diversion of energy into microbial metabolism in O_2-deficient waters (Diaz and Rosenberg, 2008) (Figure 1.2). Typically, energy flows according to a well-defined sequence of reduction–oxidation (redox) reactions, the order of which is determined by the amount of free energy available through each reaction. O_2 is the most favourable electron acceptor because it provides more energy (via its reduction) than any other electron acceptor. CO_2, the electron acceptor used by methanogenic archaea, yields the least energy. Thus, the electron acceptors that are available in a given environment are reduced in a sequential order according to the free energy yield: O_2, then NO_3^- and NO_2^-, followed by manganese and iron, then sulfate (SO_4^{2-}) and, finally, CO_2 (Zehnder and Stumm, 1988) (Figure 1.2). This sequence helps define specific metabolic niches and biogeochemical potentials spanning oxic, dysoxic (20–90 μmol O_2 per kg water), suboxic (1–20 μmol O_2 per kg water), and anoxic (<1 μmol O_2 per kg water) water column conditions, under which multiple electron acceptors can be used simultaneously to maximize the free energy yield at different ecological scales (Figure 1.2).

1.3.1 The nitrogen cycle as a distributed metabolic network

Examples of biogeochemical processes in which sequential reactions are carried out by different organisms can be found in the microbial pathways that drive the nitrogen cycle (Lam et al., 2009; Gruber and Galloway, 2008). Nitrogen gas (N_2) is the most abundant form of nitrogen on earth, but few microorganisms are able to use (fix) N_2, converting it to the more amenable form of ammonia (NH_3) or its protonated species, ammonium (NH_4^+), both of which can be terminally oxidized to NO_3^- by nitrifying bacteria. Nitrification is a chemoautotrophic process that is carried out in two steps, the first by NH_3-oxidizing bacteria or archaea, which convert NH_3 to NO_2^-, and the second by NO_2^--oxidizing bacteria, which convert the NO_2^- intermediate to NO_3^-. Nitrification typically takes place under dysoxic or suboxic conditions, resulting in N_2O production (Codispoti, 1985; Santoro et al., 2011). The oxidizing nature of the modern ocean has

resulted in NO_3^- being the most abundant form of nitrogen in the ocean. However, biological fixation of N_2 to NH_3^+ and subsequent oxidation to NO_3^- must be balanced by N_2 production in order to maintain atmospheric N_2 at a constant level over geological timescales (Deutsch et al., 2007). Under suboxic or anoxic conditions, NO_3^- and NO_2^- are used as terminal electron acceptors by denitrifying bacteria in dissimilatory NO_3^- reduction (denitrification) (Zumft, 1997) and by anammox bacteria in anaerobic NH_4^+ oxidation (anammox) (Mulder et al., 1995). Both of these processes regenerate N_2, but denitrification also produces N_2O, thus contributing to a loss in fixed nitrogen and to the production of greenhouse gases. Dissimilatory NO_3^- reduction to NH_4^+ (DNRA), a process that takes place under suboxic or anoxic conditions, has the potential to moderate the loss in fixed nitrogen and to regenerate redox couples (in the form of NO_2^- and NH_4^+) for anammox (Lees and Simpson, 1957; Cole and Brown, 1980; Simon, 2002). Taken together, these microbial nitrogen transformations constitute a distributed metabolic network linking the metabolic potentials of different taxonomic groups to higher-order biogeochemical cycling of nitrogen in the environment. Recent studies also posit an essential role for sulfur cycling in OMZs, coupling the production and consumption of reduced sulfur compounds to dissimilatory NO_3^- reduction and the fixation of inorganic carbon (Walsh et al., 2009; Canfield et al., 2010). The integration of carbon, nitrogen, and sulfur cycles represents a recurring theme in the O_2-deficient water column, where electron donors and acceptors are actively recycled between lower and higher oxidation states (Figure 1.3).

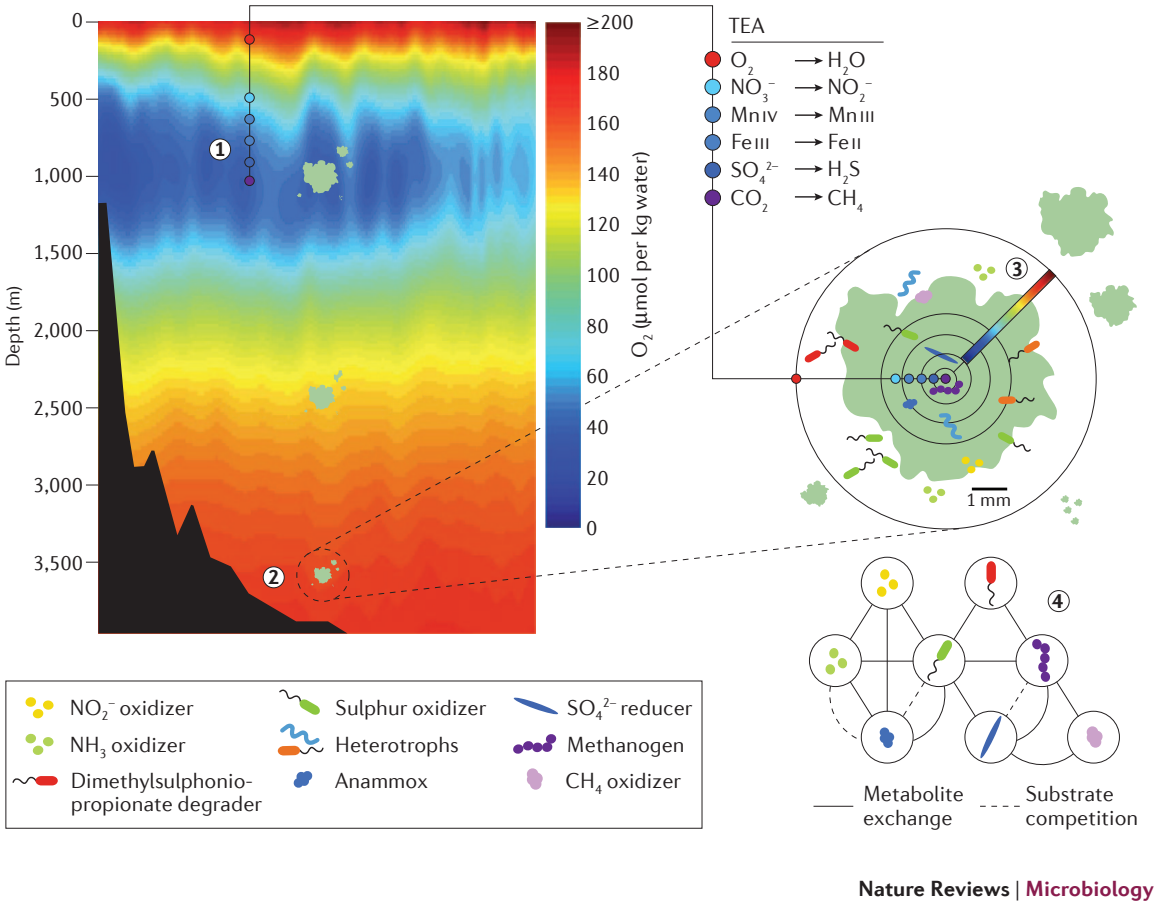


Figure 1.3 Redox-driven niche partitioning

Reduction-oxidation (redox)-driven niche partitioning in the molecular oxygen (O_2)-deficient water column selects for shared metabolic capabilities across different ecological scales. Consistent with this observation, the chemical gradients found in marine oxygen minimum zones (see the figure, label 1) also exist in interior oceanic waters in the form of sinking organic particles or 'marine snow' (Alldredge and Silver, 1988) (see the figure, label 2). Particle association provides a nucleation point for otherwise suboxic or anoxic processes in oxygenated waters owing to the formation of microscale oxyclines (Alldredge and Cohen, 1987; Karl et al., 1984; Woebken et al., 2007) (see the figure, label 3). The sulfate-reducing potential of such particles has been demonstrated (Shanks and Reeder, 1993), and a similar relationship has been identified for methane production and transport in the North Pacific Ocean (Karl and Tilbrook, 1994). The interplay between particle-associated and free-living bacteria creates distributed networks of metabolite exchange between community members with alternative or competing nutritional or energetic needs (Paerl and Pinckney, 1996) (see the figure, label 4). The recent identification of the SAR324 cluster as particle-associated bacteria with genomic potential for inorganic carbon assimilation, sulfur oxidation and methane oxidation reinforces the ecological and biogeochemical importance of microzone formation throughout the water column

(Swan et al., 2011). CH_4, methane; CO_2, carbon dioxide; H_2S, hydrogen sulfide; NH_3, ammonia; NO_2^-, nitrite; NO_3^-, nitrate; TEA, terminal electron acceptor; SO_4^{2-}, sulfate. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.4 OMZ microbiota

Taxonomic survey data based on small-subunit ribosomal RNA (SSU rRNA) gene sequences indicates that there are conserved patterns in microbial community composition between open-ocean and coastal OMZs and enclosed or semi-enclosed basins experiencing water column O_2 deficiency. The most abundant phyla in the OMZs are (in order of abundance) *Proteobacteria*, *Bacteroidetes*, Marine Group A (a candidate phylum), *Actinobacteria* and *Planctomycetes* (Figure 1.4). The phyla *Firmicutes*, *Verrucomicrobia*, *Gemmatimonadetes*, *Lentisphaerae* and *Chloroflexi*, as well as the candidate divisions TM6, WS3, ZB2, ZB3, GN0, OP11 and OD1, are also present in OMZs (Figure 1.4) (note that the taxonomy used in this chapter is taken from the Greengenes database (DeSantis et al., 2006a). The distribution of these taxa varies throughout the water column, with different subdivisions partitioning along the oxycline. These patterns reflect unique and overlapping interactions among individual microorganisms, and among populations and communities of microorganisms. In the oxic surface waters overlying OMZs, sequences affiliated with the SAR11 cluster and the order *Rhodobacterales* (class *Alphaproteobacteria*), with the order *Methylophilales* (class *Betaproteobacteria*), and with the SAR86 cluster and the clone Arctic96B-1 (class *Gammaproteobacteria*) are prevalent, as are sequences affiliated with the phylum *Cyanobacteria*, with the marine OM1 clade (phylum *Actinobacteria*), and with the clone Arctic97A-17 and the genus *Polaribacter* (phylum *Bacteroidetes*). In dysoxic and suboxic waters, prevalent sequences include those that are affiliated with the SAR11 cluster (class *Alphaproteobacteria*), with the agg47 cluster (also known as **E**SP **O**MZ **S**equence **A**ccumulation cluster II (EOSA-II) (Stevens and Ulloa, 2008)) and the clones Arctic96B-1, ZD0417, ZD0405, and ZA3412c (class *Gammaproteobacteria*), and with the SAR324 cluster and the genus *Nitrospina* (class *Deltaproteobacteria*). Sequences affiliated with *Microthrixineae* (class *Actinobacteria*), anammox bacteria (genus '*Candidatus* Scalindua', phylum *Planctomycetes*), the phylum *Chloroflexi* and various *Verrucomicrobia* are also present in dysoxic and suboxic regions. In suboxic and anoxic waters (that is, OMZs), the dominant sequences are affiliated with the SUP05 cluster (**Su**iyo Seamount hydrothermal **p**lume

group 5 (Sunamura et al., 2004); also known as EOSA-I (Stevens and Ulloa, 2008); class *Gammaproteobacteria*). In addition to these SUP05 sequences, prevalent sequences in anoxic or sulfidic waters include those affiliated with the sulfate-reducing family *Desulphobacteraceae* (class *Deltaproteobacteria*), the sulfur-oxidizing *Arcobacteraceae* (class *Epsilonproteobacteria*), the clone VC21_Bac22 (phylum Bacteroidetes), the phylum *Gemmatimonadetes* and the phylum *Lentisphaerae*. The presence of candidate divisions increases with decreasing O_2 concentrations, with most sequences affiliated with OP11 and OD1 identified in anoxic waters. In addition to bacterial SSU rRNA genes, sequences affiliated with NH_3-oxidizing marine group I (MGI) archaea in the phylum *Thaumarchaeota* have been identified in several locations, where they are most prevalent in the oxycline (Damsté et al., 2002; Lin et al., 2006; Zaikova et al., 2010; Labrenz et al., 2007; Belmar et al., 2011).

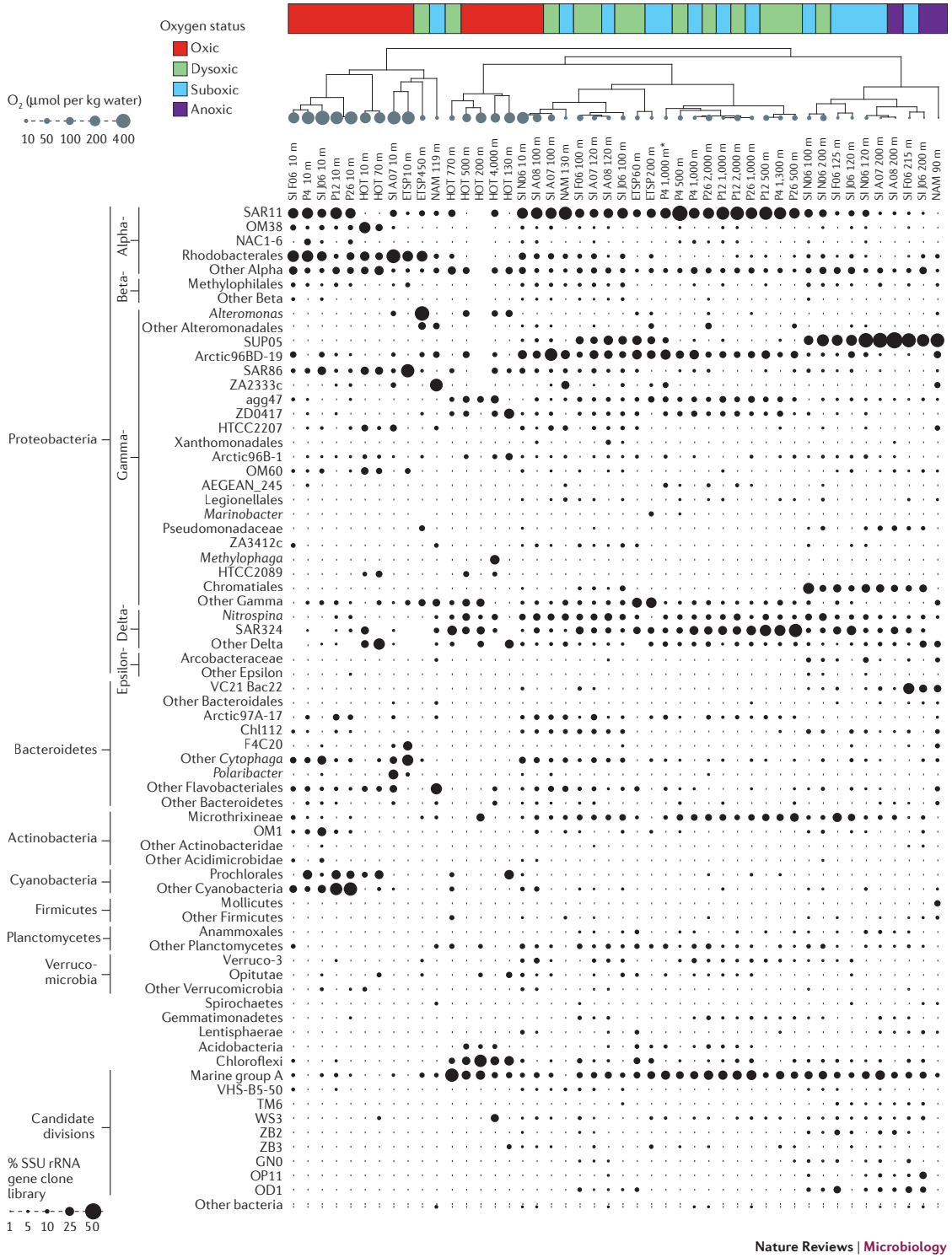


Figure 1.4 Bacterial diversity in the oceanic OMZs

Dot plot of the diversity of bacterial taxa at various sample points and depths in Saanich Inlet (SI), the Northeast subarctic Pacific (NESAP; labeled P4, P12 and P26), the Hawaii Ocean Time-series (HOT), the eastern tropical

South Pacific (ETSP) and the Namibian upwelling (NAM; also known as the Benguela upwelling), based on small-subunit ribosomal RNA (SSU rRNA) gene sequence profiles. '*' indicates a sample taken from P4 1000 m in June 2008; all other NESAP samples were taken in February 2009. Samples are organized according to the similarity of their community composition, as revealed by hierarchical clustering of the distribution of taxonomic groups across environmental samples. The molecular oxygen (O_2) concentration is shown for each oceanic sample, and the classification of the environment as oxic, dysoxic, suboxic, or anoxic is also indicated in the colour bar. Names for identifying bacterial groups were selected according to the taxonomic level at which the most relevant information was available. Data used to generate the dot plot were derived from sequences deposited in Genbank. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.4.1 Bacterial taxa of emerging interest in OMZs

Although many of the taxa identified in O_2-deficient waters are ubiquitous throughout the ocean, patterns of endemism emerge at the level of operational taxonomic units (OTUs) obtained by clustering SSU rRNA gene sequences together at specific identity thresholds. These OTU distribution patterns reinforce a model of ecological type (ecotype) selection in which genetically cohesive populations manifest distinct ecological or biogeochemical roles (Koeppel et al., 2008). These distinct roles in turn form the basis of distributed interaction networks that integrate the phenotypes of different taxonomic groups. The following section focuses on the four most abundant bacterial taxonomic groups identified in surveys of OMZs (the SUP05–Arctic96BD-19 group, the SAR11 and SAR324 clusters, and the candidate phylum Marine Group A) for OTU distribution analysis (Figure 1.5).

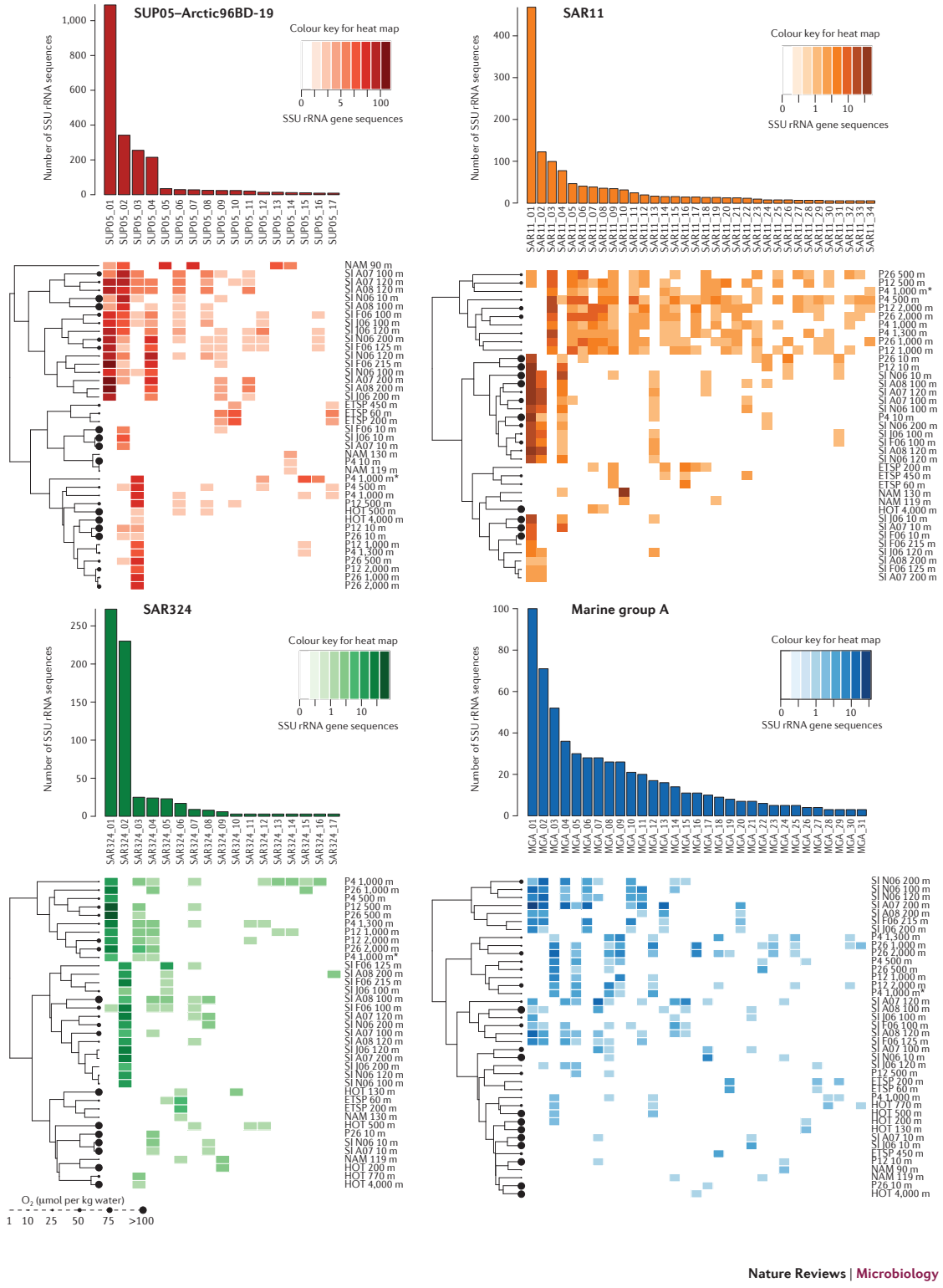


Figure 1.5 Diversity in the four most abundant bacterial groups identified in OMZs

The four most abundant groups that were identified in small-subunit ribosomal RNA (SSU rRNA) gene surveys of oxygen minimum zones (OMZs) are the SUP05- Arctic96BD-19 group, the SAR11 and SAR324 clusters, and Marine Group A (MGA). Each histogram bar represents a cluster of SSU rRNA gene sequences, or an operational taxonomic unit (OTU), generated at a 97% identity cutoff (clustered using the furthest-neighbour algorithm). The height of the bar is equivalent to the sum of all sequences belonging to a specific OTU across all environments surveyed: the eastern tropical South Pacific (ETSP), the Hawaii Ocean Time-series (HOT), the Northeast subarctic Pacific (NESAP; labeled as P4, P12 and P26), Saanich Inlet (SI) and the Namibian upwelling (NAM; also known as the Benguela upwelling). '*' indicates a sample taken from P4 1000 m in June 2008; all other NESAP samples were taken in February 2009. Heat maps below the histograms represent the distribution of sequences in each OTU across all environments surveyed. Heat maps were clustered by row using Euclidean distance and the furthest-neighbour algorithm to highlight patterns of diversity among samples. Inset colour scales depict the colour code for the number of SSU rRNA gene sequences in heat maps. Data were derived from sequences deposited in Genbank. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.4.1.1 SUP05 Gammaproteobacteria

SSU rRNA gene sequences affiliated with chemoautotrophic, sulfur-oxidizing gill symbionts of deep-sea clams and mussels were first identified in open-ocean OMZs in the Arabian Sea, the ETSP and the Namibian upwelling (Stevens and Ulloa, 2008; Fuchs et al., 2005; Lavik et al., 2009). Phylogenetic analysis indicates that these symbionts are part of a larger group of free-living relatives of symbionts (also referred to as the gammaproteobacterial sulfur-oxidizing cluster (or GSO)) consisting of two closely related, co-occurring and currently uncultivated lineages, SUP05 (Sunamura et al., 2004) (encompassing the clam and mussel symbionts) and Arctic96BD-19 (Bano and Hollibaugh, 2002) (Figure 1.5). SUP05 and Arctic96BD-19 exhibit overlapping but not identical distribution patterns, consistent with redox-driven niche partitioning. SUP05 is most abundant in the slightly to moderately sulfidic waters at the base of the sulfide–NO_3^- transition zone, and the organisms in this clade derive energy from the oxidation of reduced sulfur compounds using NO_3^- as a terminal electron acceptor (Walsh et al., 2009; Lavik et al., 2009). Arctic96BD-19 is most abundant in dysoxic and suboxic waters, as its members derive energy from reduced sulfur compounds using O_2 as a terminal electron acceptor (Walsh and Hallam, 2011; Swan et al., 2011). Both SUP05 and Arctic96BD-19 members have the potential to use the energy gained from the oxidation of reduced sulfur compounds to fix inorganic carbon via Rubisco (Walsh et al., 2009; Swan et al., 2011). Under more sulfidic water

column conditions, SUP05 members are replaced by Epsilonproteobacteria that exhibit similar metabolic capabilities (Lin et al., 2008; Grote et al., 2008; Grote et al., 2012). The presence of SUP05 species in non-sulfidic OMZs serves as a biomarker for changing ecosystem dynamics, indicating an increased potential for toxic sulfur blooms and periodic or persistent anoxia (Zaikova et al., 2010; Lavik et al., 2009; Walsh and Hallam, 2011).

1.4.1.2 SAR11 Alphaproteobacteria

SAR11 is the most abundant and ubiquitous clade of Alphaproteobacteria in the ocean, often constituting 30% of surface bacterioplankton communities (Morris, 2002). The dominant SAR11 OTUs observed in OMZs (Figure 1.5) are closely related to *Pelagibacter ubique*, a cultivated member of SAR11 in subgroup Ia, and there is also a minority representation for subgroups Ib and II. SSU rRNA gene surveys support the existence of several SAR11 ecotypes that exhibit geographical and depth-specific water column distributions (Field et al., 1997). Ecotype selection is particularly apparent in the NESAP OMZ, where more than 30 different SAR11 OTUs have been identified (Figure 1.5). Current studies of cultivated SAR11 strains and free-living populations indicate a genomic repertoire that is streamlined for rapid heterotrophic growth (Conley et al., 2002). Comparative genomics analyses suggest that there are differences in the glycolytic potential of coastal and open-ocean SAR11 populations. Consistent with this observation, carbon use and gene expression assays have measured preferential glucose use by coastal isolates that are associated with a gene cluster encoding a variant form of the Entner–Doudoroff pathway (Schwalbach et al., 2010). However, the specific metabolic capabilities that enable SAR11 members to thrive in O_2-deficient waters remain unknown. The role of surface water SAR11 populations in mediating demethylation of dimethysulfoniopropionate (DMSP) to methylmercaptopropionate (MMPA) may indicate a role for this group in OMZ sulfur cycling. Although this phenotype is shared between a number of different pelagic bacteria (Howard et al., 2006; González et al., 2003), SAR11 members are unable to perform dissimilatory sulfate reduction, making them dependent on exogenous sources of reduced sulfur for growth and further reinforcing a model of distributed metabolite exchange (Tripp et al., 2008).

1.4.1.3 Marine Group A bacteria

Marine Group A (MGA) bacteria were first identified in SSU rRNA gene clone libraries generated from surface waters of the Atlantic and Pacific Oceans (Fuhrman et al., 1993; Fuhrman and Davis, 1997; Gordon and Giovannoni, 1996). MGA, originally referred to as the "SAR406 gene lineage", represents a deeply branching lineage of bacteria related to the genus *Fibrobacter* and the green sulfur bacterial (GSB) phylum, which includes the genus *Chlorobium* (Gordon and Giovannoni, 1996). To date, MGA remains a candidate phylum with no cultured representatives. Modern phylogenetic analyses indicate that the closest cultivated relatives of MGA are *Caldithrix abyssi* and *Caldithrix palaeochoryensis*, both belonging to the phylum *Caldithrix*. These isolates are anaerobic, mixotrophic, thermophiles obtained from hydrothermal vent and sediment environments, respectively (Miroshnichenko et al., 2003; Miroshnichenko et al., 2010). While ubiquitous in the dark ocean, MGA appear to be most prevalent and diverse in OMZs and permanent or seasonally stratified anoxic basins (Madrid et al., 2001; Fuchs et al., 2005; Schattenhofer et al., 2009; Stevens and Ulloa, 2008; Zaikova et al., 2010). The dominant MGA OTUs observed in OMZs (Figure 1.5) are closely related to subgroups SAR406, Arctic95A-2, Arctic96B-7 and ZA3648c, with an additional minority representation for subgroup ZA3312c. Six or more additional subgroups seem to be endemic to the NESAP, Saanich Inlet and the ETSP. These results are consistent with previous observations indicating that there is strong habitat selection for different MGA subgroups (Rappé and Giovannoni, 2003). Despite these organisms being widespread in the ocean, their metabolic capabilities remain unknown.

1.4.1.4 SAR324 Deltaproteobacteria

Similarly to MGA, the SAR324 clade (also known as Marine Group B) of the class Deltaproteobacteria is also prevalent in the dark ocean (Fuhrman and Davis, 1997; Wright et al., 1997; Brown and Donachie, 2007). The most common SAR324 OTUs observed in OMZs (Figure 1.5) are closely related to Marine Group B–SAR324 clade II, and there is also a minority representation for Marine Group B–SAR324 clade I; two additional clades are endemic to Saanich Inlet and the ETSP. SAR324 members have the potential to oxidize one-carbon (C1) compounds and reduced sulfur compounds, using the resulting energy to fix inorganic carbon via Rubisco (Swan et al., 2011). Consistent with a functional role for Rubisco, microautoradiography

linked with catalysed reporter deposition fluorescence *in situ* hybridization (MAR–CARD–FISH) demonstrated that SAR324 members fix inorganic carbon and undergo particle association in oxygenated waters of the North Atlantic (Swan et al., 2011).

1.4.2 Archaeal taxa of emerging interest in OMZs

In contrast to the bacterial domain, less is known about archaeal community composition along the O_2 gradients of OMZs and euxinic basins. This section uses published SSU rRNA sequence data from the ETSP (Belmar et al., 2011), the Black Sea (Vetriani et al., 2003; Coolen et al., 2007), the Cariaco Basin (Madrid et al., 2001; Jeon et al., 2008), the Namibian upwelling (Woebken et al., 2007), and the Baltic Sea (Labrenz et al., 2010), as well as new data from the NESAP, Saanich Inlet (SI), and eastern subtropical South Pacific (ESP) to highlight the major groups present in these systems. Figure 1.6 illustrates the distribution of the respective phylotypes within the general archaeal phylogenetic tree. Most phylotypes are affiliated with well-recognized pelagic marine clades such as Group I.1a (G-I.1a, (DeLong, 1998)), pSL12-related group (Mincer et al., 2007), Marine Group II (MG II, (DeLong, 1992)), and Marine Group III (MG III, (Fuhrman and Davis, 1997)). However, a significant number of phylotypes cluster within clades originally found in sediments (Marine Benthic Group A and E; (Vetriani et al., 1999)) or deep-hydrothermal vents environments (DHVE-4 and DHVE-5, (Takai and Horikoshi, 1999)). Archaeal community structure mirrors trends for bacterial denizens of OMZs, including close taxonomic affiliation with archaeal groups from diverse seafloor environments (e.g. subseafloor sediments, deep-sea hydrothermal vents and cold seep). Although these similarities likely reflect recurring patterns of niche selection based on convergent environmental conditions (e.g. O_2 depletion), the precise ecological and biogeochemical roles of archaea in OMZs and other seafloor environments remain poorly constrained.

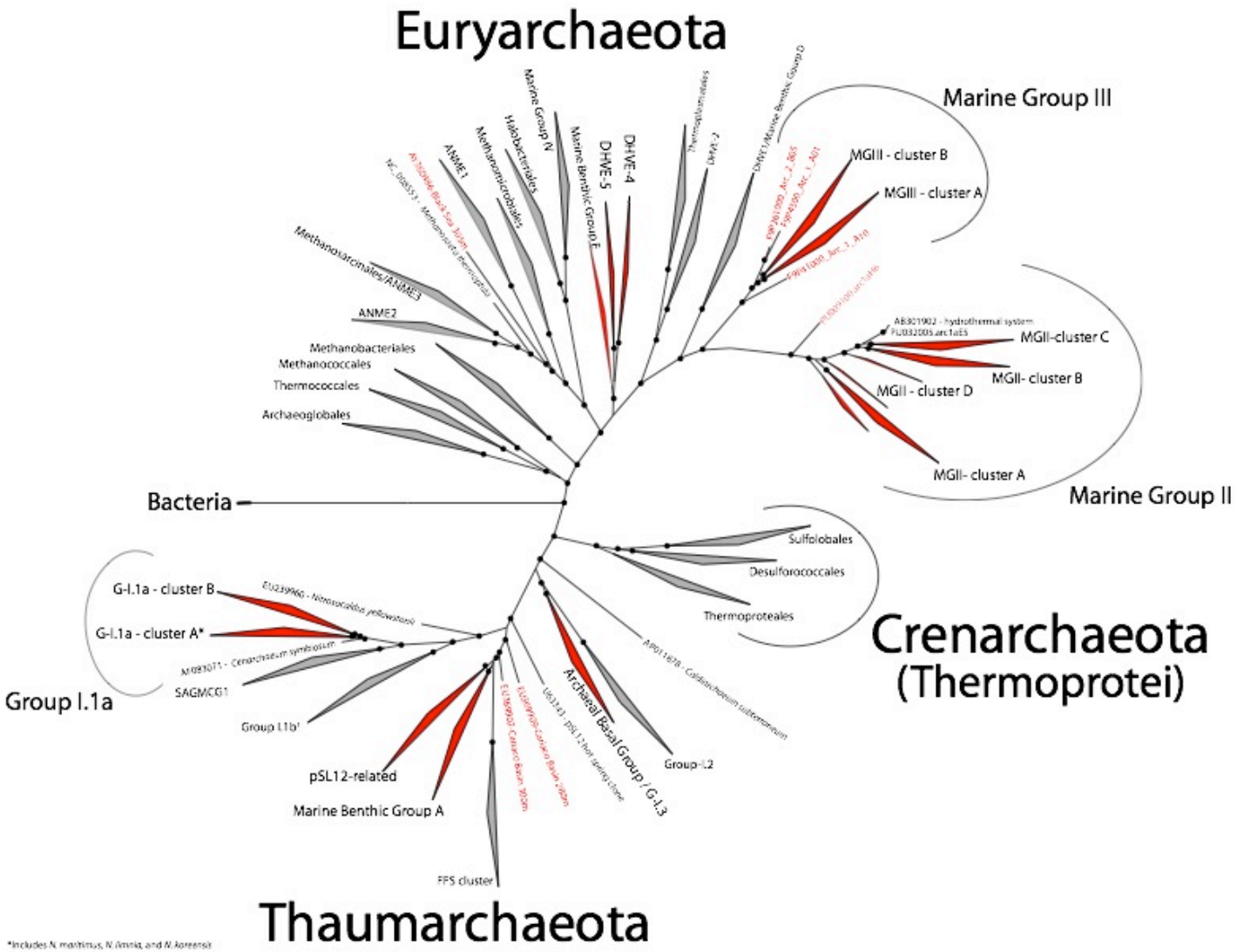

Euryarchaeota
Marine Group III
MGIII - cluster B
MGIII - cluster A
Marine Group IV
Marine Benthic Group E
DHVE-5
DHVE-4
Thermoplasmatales
DHVE-2
DHVE1/Marine Benthic Group D
ANME1
Halobacteriales
Methanomicrobiales
Methanosarcinales/ANME3
ANME2
Methanobacteriales
Methanococcales
Thermococcales
Archaeoglobales
MGII-cluster C
MGII- cluster B
MGII - cluster D
MGII- cluster A
Marine Group II
Bacteria
Sulfolobales
Desulfurococcales
Thermoproteales
Crenarchaeota
(Thermoprotei)
G-I.1a - cluster B
G-I.1a - cluster A*
SAGMCG1
Group I.1a
Group I.1b†
pSL12-related
Marine Benthic Group A
FFS cluster
Archaeal Basal Group / G-I.3
Group I.2
Thaumarchaeota
*Includes N. maritimus, N. limnia, and N. koreensis
†Includes N. gargensis

Figure 1.6 Maximum-likelihood phylogenetic tree of archaeal SSU-rRNA gene sequences
Representative sequences of OMZ phylotypes (>97% similarity, using UCLUST (Edgar, 2010)) together with other sequences from Genbank were aligned with Mafft (Katoh et al., 2002). The phylogenetic tree was built with Bosque (Ramírez-Flandes and Ulloa, 2008), using FastTree (Price et al., 2010) and applying the general time-reversible DNA model. Red boxes represent phylogenetic clusters containing OMZ phylotypes. Dots at nodes represent branches with support values of ≥70%. The scale bar indicates the expected changes per sequence position (note: the scale only applies to branches of the tree; boxes are not scaled). Figure used with permission from Springer (Ulloa et al., 2013).

1.4.2.1 Thaumarchaeotal G-I.1a

A remarkably high proportion of archaeal sequences recovered from OMZs affiliate with the thaumarchaeotal G-I.1a, a group well represented in all of the considered systems (Figure 1.7). This group, initially referred to as Marine Group I (DeLong, 1992) is ubiquitous and abundant in the global ocean (Francis et al., 2005; Hallam et al., 2006). Group I.1a contains two statistically supported clusters, designated as A and B (Belmar et al., 2011), although some authors have divided this group into additional clusters (Massana et al., 2000). With the exception of *Cenarchaeum symbiosum*, which appears outside of the A and B subdivisions, the G-I.1a-A cluster comprises all marine thaumarchaeotal species that have been fully sequenced thus far (i.e., *Nitrosopumumilus maritimus* and *Nitrosoarchaeum limnia*). The G-I.1a-A cluster also includes sequences retrieved from diverse terrestrial and marine environments including surface waters, deep-ocean sediments and agricultural soils. In contrast, cluster G-I.1a-B includes very few phylotypes from oxic surface waters, and is mainly composed of sequences from deep waters, marine hydrothermal vents, and O_2-deficient waters. Since many representatives of the G-I.1a archaeal group are known ammonium-oxidizers (AOA), and given the correlation between phylogenetic markers for AOA and the functional marker ammonia monooxygenase subunit alpha (*amoA)*, OMZ representatives of this group are considered presumptive nitrifiers (Molina et al., 2010).

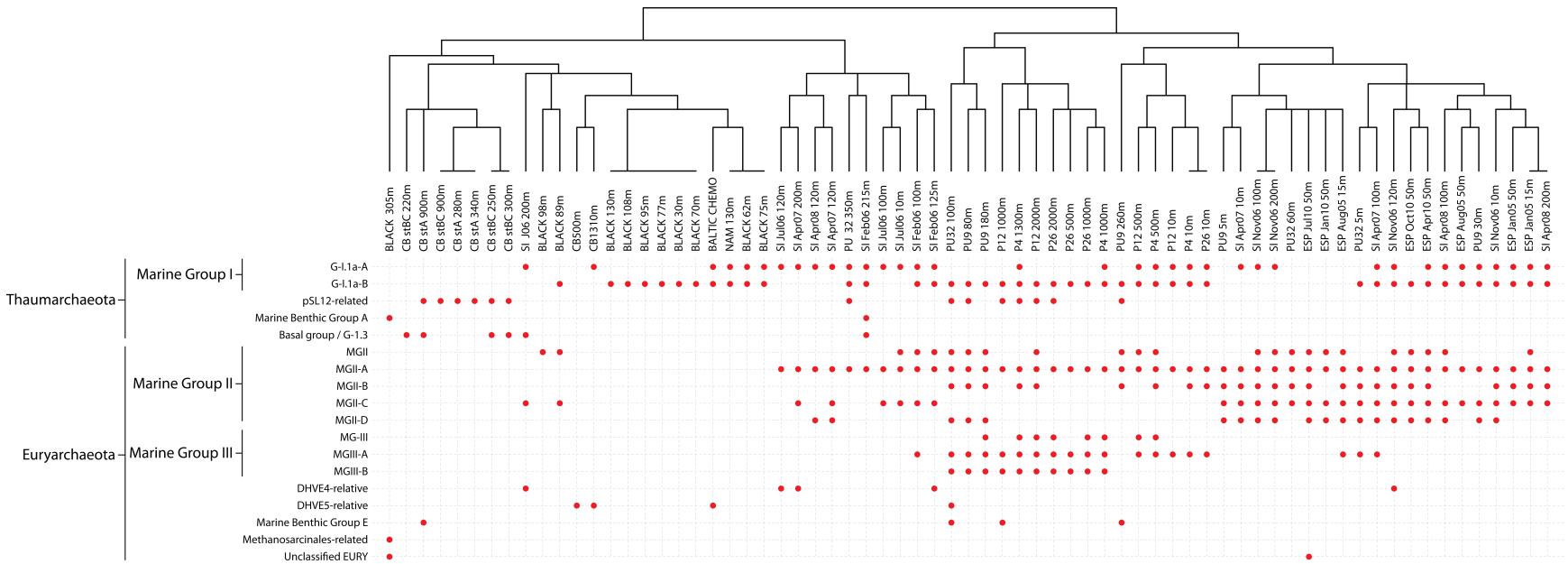


Figure 1.7 Presence/absence dot plot of archaeal taxa at various sample points and depths in the ocean based on SSU rRNA gene sequence profiles

Samples are organized according to the similarity of their community composition, as revealed by hierarchical clustering of the distribution of taxonomic groups across all environments surveyed: the Northeast subarctic Pacific (NESAP; labeled as P4, P12 and P26), the eastern subtropical South Pacific (ESP), the Peru Upwelling (PU); Saanich Inlet (SI), the Black Sea (BLACK), the Baltic Sea (BALTIC), and the Cariaco Basin. Names for identifying archaeal groups were selected according to the taxonomic level at which the most relevant information was available (data used to generate the dot plot were derived from sequences in Genbank). Figure used with permission from Springer (Ulloa et al., 2013).

1.4.2.2 Thaumarchaeotal G-I.3

Some thaumarchaeal phylotypes found in anoxic or euxinic waters classify as being part of the major branch that includes the pSL12-related group (Mincer et al., 2007), the marine benthic group A (Vetriani et al., 1999), and the FFS cluster, which contains sequences retrieved from forest soil (Jürgens et al., 1997). This major branch is a sister group of the branch joining terrestrial Group I.1b and marine Group I.1a, and is related to the extremophile representative pSL12 (Mincer et al., 2007). Additional phylotypes recovered from below the chemocline in the Cariaco Basin (Jeon et al., 2008) appear at the base of the *Thaumarchaeota* (Figure 1.6). Interestingly, these sequences were generated using primers designed for eukaryotes from anoxic/euxinic waters in the Cariaco Basin and composed a group with phylotypes from freshwater sediments, rice roots and soil (Jürgens et al., 2000), sediments near deep hydrothermal vents (Takai et al., 2001), and sub-seafloor sediments (Inagaki et al., 2003; Sørensen and Teske, 2006). This group has been named "Group I.3" (Jürgens et al., 2000) or Miscellanous Crenarchaeotic group (Inagaki et al., 2003) (Figure 1.6).

1.4.2.3 Euryarchaeota

Other groups prevalent in oxygen-deficient systems are the euryarchaeal MGII and MGIII (Figures 1.6 and 1.7). MGII is a cosmopolitan group, and the majority of sequences observed in coastal and open-ocean OMZs are affiliated with the MGII-A cluster. MGIII is a less prominent group in the global ocean, but appears to be important in OMZs (Belmar et al., 2011). Finally, some euryarchaeal phylotypes from OMZs and euxinic waters associate with Marine Benthic Group E and DHVE-4 and DHVE-5 groups.

1.5 The symbiotic ocean

Recent advances in microbial ecology that combine cultivation-independent molecular methods with process rate measurements are beginning to reveal previously unknown metabolic interactions in the O_2-deficient water column. Parallel advances in exploring the dark ocean have identified similar metabolic interactions at different ecological scales. The following case studies touch on these observations, with particular emphasis on the integration of carbon, nitrogen and

sulfur cycles.

1.5.1 Unraveling a cryptic sulfur cycle in OMZs

The identification of the SUP05–Arctic96BD-19 clade of Gammaproteobacteria suggested an important role for sulfur cycling in the ecology and biogeochemistry of OMZs (Stevens and Ulloa, 2008; Zaikova et al., 2010; Lavik et al., 2009). Metabolic reconstruction of the SUP05 metagenome identified numerous genes encoding components of the sulfide oxidation and nitrate reduction pathways. Principal components of both pathways were found to be clustered in a 52 kb 'metabolic island' that also contains a gene encoding the large subunit of form II Rubisco, consistent with coordinated regulation of carbon and energy metabolism (Walsh et al., 2009). More recently, single-cell techniques were used to assemble genomic scaffolds for Arctic96BD-19 from North Atlantic and Pacific waters, uncovering sulfur oxidation and CO_2 fixation genes (Swan et al., 2011). The discovery of potential sulfur oxidizers in non-sulfidic waters is enigmatic, bringing into question the source of the reducing equivalents that are needed to fix inorganic carbon.

1.5.1.1 Enigmatic sulfate reduction in OMZs

Sinking particles have been proposed to be sources of reduced compounds such as sulfide in oxic waters (Shanks and Reeder, 1993; Alldredge and Cohen, 1987). However, SO_4^{2-} reduction in the water column is difficult to measure because sulfide rapidly auto-oxidizes in the presence of even trace amounts of O_2. Typically, when sulfide is detected in OMZs, it originates in rare pockets of NO_3^-- and NO_2^--depleted water (Dugdale et al., 1977) or is released by diffusive flux from sediments (Canfield, 2006; Brüchert et al., 2003). Thus, the *in situ* component of the sulfur cycle in non-sulfidic waters has been described as cryptic because it lacks obvious chemical expression in the water column. To resolve this enigma, researchers conducted process rate measurements of SO_4^{2-} reduction in the ETSP OMZ using radiolabelled SO_4^{2-} ($^{35}SO_4^{2-}$) after a pulse of unlabelled sulfide to capture the formation of radiolabelled sulfide (Canfield et al., 2010). High rates of SO_4^{2-} reduction were detected (between 0.28 and 1.0 nmol per m^2 per day) coupled to the production of NO_2^-, N_2O and N_2. Consistent with process rate measurements, metagenome sequencing recovered a limited number of genes originating from canonical sulfate-

reducing bacteria, including *Desulphatibacillum*, *Desulphobacterium*, *Desulphococcus*, *Syntrophobacter*, and *Desulphovibrio* spp.

1.5.1.2 Sulfur oxidation in OMZs

Sulfur oxidation coupled to NO_3^- reduction in the ETSP OMZ was supported by metatranscriptomic analyses, which revealed that transcripts for dissimilatory sulfite reductase (*dsr*) genes, sulfur oxidation (*sox*) genes (encoding proteins that mediate thiosulfate oxidation), adenosine-5′-phosphosulfate (APS) reductase (*apr*) genes (encoding proteins that mediate the conversion of sulfur to SO_4^{2-}) and the gene encoding the catalytic subunit of respiratory NO_3^- reductase (*narG*) were highly expressed (Stewart et al., 2012). Although most of these transcripts originated from sulfur-oxidizing organisms, including members of the SUP05 clade and close relatives, a minority of *aprA* and *dsrB* transcripts affiliated with canonical SO_4^{2-}-reducing bacteria were detected, consistent with there being an active sulfur cycle in this environment. Interestingly, in ETSP metagenomes, 32% of the top hits to *aprA* were affiliated with SAR11, and cognate *aprA* transcripts were highly expressed throughout the oxycline. Although the precise role of SAR11 in sulfur cycling in OMZs remains unknown, the expression of Apr could indicate a link between DMSP demethylation and the production of reducing equivalents for sulfur oxidation in the surrounding water column. In addition to providing reducing equivalents for NO_3^- reduction, sulfur metabolism may contribute to NH_4^+ production through the process of DNRA (also termed NO_3^- or NO_2^- ammonification). Both SO_4^{2-}-reducing and sulfur-oxidizing bacteria have been shown to carry out DNRA, in some cases using the conversion of NO_3^- or NO_2^- to NH_4+ as an electron sink for substrate-level phosphorylation (Cole and Brown, 1980), and in others coupling the conversion process to generate a proton motive force (Simon, 2002). Although most of what we know about DNRA comes from sediment incubation or laboratory experiments with pure cultures (Simon, 2002; S rensen, 1987), several studies using nitrogen-15 incubations have measured potential DNRA rates in the water columns of the Baltic Sea (Samuelsson and Rönner, 1982), the ETSP (Lam et al., 2009), and the Namibian upwelling (Kartal et al., 2007). In the ETSP OMZ, DNRA is estimated to provide a substantial portion of the NH_4^+ that is required for anammox (Lam et al., 2009), with up to 22% derived from SO_4^{2-} reduction (Canfield et al., 2010). The potential contributions of sulfur-oxidizing bacteria,

including members of the SUP05–Arctic96BD-19 and SAR324 clades, to DNRA in the O_2-deficient water column remain to be determined.

1.5.1.3 The role of SUP05 bacteria in the cryptic sulfur cycle

The existence of a cryptic sulfur cycle coupling the metabolic activities of SUP05–Arctic96BD-19 bacteria and sulfate-reducing bacteria in the O_2-deficient water column or in association with sinking particles is reminiscent of the symbiotic associations found at oxic–anoxic interfaces (Brune et al., 2000). Indeed, chemoautotrophic symbioses are a common innovation at hydrothermal vent and cold seep habitats, where eukaryotic hosts provide optimal access to the redox couples that are needed to fix inorganic carbon on or near the sea floor (Stewart et al., 2005). Similarly, symbiotic sulfur-oxidizing and SO_4^{2-}-reducing bacteria have been described in association with the shallow-water sand-dwelling mouthless worm *Olavius algarvensis*; for these bacteria, metabolite exchange between different taxonomic groups balances out the fitness costs associated with resource competition in the host milieu (Dubilier et al., 2001). Other forms of syntrophy, including direct electron transfer, have been described between bacterial and archaeal cells such as acetogens and methanogens or SO_4^{2-} reducers and methane oxidizers, resulting in the production of reduced compounds that fuel subsurface metabolism and deep-sea chemolithoautotrophic communities (Schink, 2002; Stams and Plugge, 2009). Thus, the ecology and biogeochemistry of OMZs represent one manifestation of the greater symbiotic ocean and its impact on the world around us. This impact is rooted in the collective metabolic capabilities of microbial cells that drive matter and energy transformations throughout the depth continuum.

1.5.2 Photoautotrophy in anoxic OMZs

Anoxic OMZs can impinge on the photic zone, creating a unique environment for photoautotrophs, and particularly oxygenic ones adapted to low O_2 tensions. The latter could provide a local source of O_2 to feed aerobic processes (e.g., nitrification) in a typically anoxic environment. Indeed, picocyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are frequent inhabitants of low-light oceanic OMZ waters (Johnson et al., 1999; Goericke et al., 2000; Galán et al., 2009) of the Arabian Sea, ETNP and ETSP. A recent study in the eastern

tropical Pacific showed that OMZ *Prochlorococcus* communities contain novel phylotypes (Lavin 2010). The genomic characteristics of these OMZ photoautotrophs remain to be determined. They may provide new insights about the evolution of photosynthesis as the planet and the ocean became oxygenated.

1.5.3 Ammonia oxidation in OMZ boundaries

While anaerobic microorganisms performing nitrogen and sulfur transformations characterize the core of the OMZ, the oxycline and low O_2 waters of the upper OMZ are critical zones for aerobic nitrifying microorganisms, particularly the AOA. Early studies pointed to a significant role for the process of ammonia oxidation in OMZs, particularly at the upper boundaries (e.g., (Ward and Zafiriou, 1988; Ward et al., 1989; Lipschultz et al., 1990)). Catalyzed by the ammonia monooxygenase (Amo) enzyme, the ability to oxidize ammonia was originally thought to be restricted to a few groups within the γ- and β-proteobacteria. However, metagenomic studies performed in the last decade revealed the existence of unique *amoA* genes derived from uncultivated, nonextremophilic Crenarchaeota (Venter et al., 2004; Hallam et al., 2006; Treusch et al., 2005), now recognized as a separate phylum, the Thaumarchaeota (Figure 1.6).

In addition, an isolate of the marine thaumarchaeon *Nitrosopumilus maritimus* demonstrated a capacity for growth using ammonia oxidation as an energy source, resulting in stoichiometric production of nitrite (Konneke et al., 2005). Subsequently, high abundances of archaeal *amoA* genes have been detected in a variety of O_2-deficient marine environments including the OMZs of the ETNP and ETSP, and the suboxic zones of the Black Sea, the Gulf of California, and the Baltic Sea (Francis et al., 2005; Coolen et al., 2007; Lam et al., 2007; Beman et al., 2008; Molina et al., 2010). Metatranscriptomic analysis in the ETSP showed that up to 20% of all protein-coding transcripts matched *N. maritimus* in the upper OMZ and that thaumarchaeotal *amo* genes were highly transcribed in this zone (Stewart et al., 2012). These results reinforce the emerging perspective that thaumarchaeotal ammonia-oxidation contributes substantially to nitrogen cycling in diverse marine environments (Wuchter et al., 2006; Prosser and Nicol, 2008).

1.5.4 Carbon fixation in OMZs

In addition to playing key roles in nitrogen and sulfur cycling, OMZ microorganisms may contribute with a substantial proportion of fixed organic carbon. Sulfur-oxidizers like SUP05, for example, harbour genes for inorganic carbon fixation through the Calvin-Benson-Bassham cycle (Walsh et al., 2009), while anammox bacteria can make use of the acetyl-coenzyme A (CoA) pathway for carbon fixation (Strous et al., 2006). Isolation of the ammonia-oxidizing thaumarchaeon *N. maritimus* also revealed a capacity for chemolithoautotrophic growth on ammonia as a sole energy source and bicarbonate as a sole carbon source (Konneke et al., 2005). Subsequent sequencing of the *N. maritimus* genome confirmed that it contains genes for the 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) pathway of autotrophic carbon fixation (Walker et al., 2010). The actual contribution of these groups (and of others) to the carbon economy of OMZs remains to be determined.

1.6 Microbial co-occurrence networks

Just as cellular complexity arises through networks of genes, proteins, and metabolites interacting across multiple hierarchical levels (Jeong et al., 2000; Ravasz et al., 2002; Barabási and Oltvai, 2004), so ecological and biogeochemical phenotypes arise from complex interactions between microbial community members. As stated by Chisholm and Cary, "No single organism contains all the genes necessary to perform the diverse biogeochemical reactions that make up ecological community function. Yet, distributed among the community, are all the functions necessary to define that community's interaction with its environment" (Chisholm et al., 2001). These interactions form the basis of distributed networks in which nodes are taxa and links are the correlations between taxa. Microbial networks continuously evolve by the arrival and departure of new nodes and links through mutation, gene transfer or habitat selection, creating functionally redundant modules that are separated in space and time. The application of network theory to discover and define co-occurrence patterns among so-called 'free-living' microorganisms represents a new frontier in microbial ecology (Raes and Bork, 2008; Faust and Raes, 2012) (Figure 1.8).

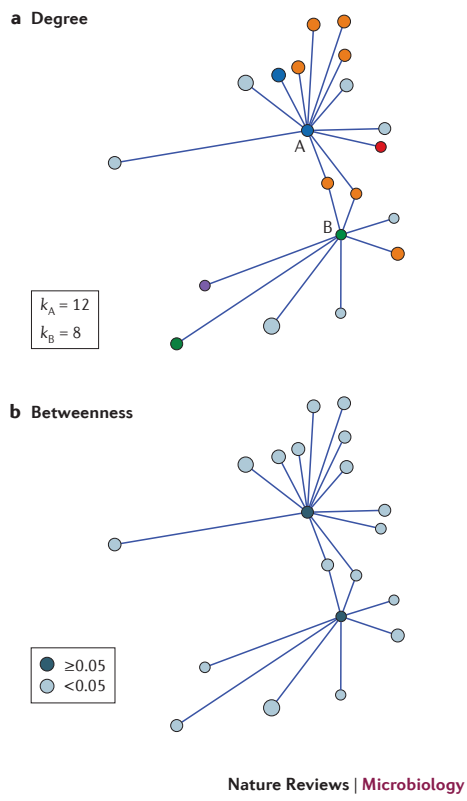


Figure 1.8 Network analysis

The properties of many complex systems, including the cell, the brain, and the internet, are the result of numerous pairwise interactions of individual components. Such a system can be represented by a set of nodes (components or subunits) connected by links (interactions between nodes) to form a network (see the figure). In the co-occurrence network shown in Figure 1.9 of this chapter, nodes represent operational taxonomic units (OTUs) and links represent Pearson correlation coefficients greater than 0.4. In order for a network model to bolster understanding of a complex system such as a microbial community, it is necessary to quantify the topological features or properties of the network. Important properties include:

• *Degree*: the number of neighbours of a node, also known as connectivity. Nodes with high degrees (many links to other nodes) are referred to as hubs. In the example network (see the figure, part a), node A has a degree of 12 and node B has a degree of 8. The degree distribution, P(*k*), of a network is the probability that a given node has exactly *k* links (Barabasi and Albert, 1999).

• *Betweenness*: the frequency at which a node is present on the shortest path between all other nodes; in other words, a measure of how central a node is within a network. Nodes with high betweenness have been shown to control the flow of information across a network (Newman, 2010). In the example network (see the figure, part b), the dark-blue nodes have high betweenness relative to the light-blue nodes.

Quantification of the properties of a network is the basis for distinguishing the network type, from which we can

infer certain biological properties (Barabási and Oltvai, 2004). Many biological networks reported in the literature (including metabolic and protein networks) are scale-free networks (Jeong et al., 2000; Jeong et al., 2001), meaning that they exhibit power-law degree distributions (that is, $P(k) \sim k^{-b}$, in which b is the degree exponent, an experimentally observed quantity that typically ranges between 2 and 3) (Barabási and Oltvai, 2004). Scale-free distribution implies that a network consists of a small number of hubs in addition to numerous nodes with fewer links (Barabasi and Bonabeau, 2003). In scale-free networks (including the co-occurrence network shown in Figure 1.9), the hubs display a high betweenness, suggesting that these nodes have important roles in regulating network interactions. Although co-occurrence networks do not directly implicate specific modes of metabolite exchange, they provide an excellent framework for generating hypotheses regarding potential metabolic interactions that can be further tested using environmental parameter, *in situ* process rate, and functional gene data. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.6.1 Patterns of microbial co-occurrence in OMZs

Recently, local similarity analysis (LSA) (Ruan et al., 2006) was used to calculate co-occurrence patterns between bacterial OTUs and environmental parameter data recovered from the chlorophyll maximum at the San Pedro Ocean Time Series (SPOTS) (Fuhrman and Steele, 2008). The resulting networks revealed both positive and negative correlations between specific OTUs and water column conditions, and these correlations were either direct or time-lagged in nature (Fuhrman and Steele, 2008). From an interpretive perspective, positive correlations could represent cooperative activities, including distributed metabolism, cross-feeding or overlapping habitat preference, whereas negative correlations could represent resource competition, predation or alternative habitat preference (Fuhrman and Steele, 2008). For example, ten SAR11 OTUs participated in different subnetworks over time (some correlating with other bacterial OTUs and others with environmental parameter data), consistent with ecotype selection and succession. The LSA approach has been extended to include three-domain interactions (between Archaea, Bacteria and protists (from Eukaryota)) occurring at SPOTS and in the English Channel, revealing progressive changes in microbial co-occurrence patterns over time and uncovering potential symbiotic associations (Steele et al., 2011; Gilbert et al., 2012). On a global scale, co-occurrence analysis of 298 591 publicly available SSU rRNA gene sequences was used to define nonrandom and recurring co-occurrence networks, consistent with habitat preference between specific taxonomic groups (Chaffron et al., 2010).

Although networks based on phylogenetic information alone cannot explain underlying

mechanisms of metabolite exchange, they can help define putative metabolic interactions and enable more direct hypothesis testing when combined with data about environmental parameters, process rates and functional genes. In this chapter, co-occurrence network analysis is applied to publicly available SSU rRNA gene sequences from the Hawaii Ocean Time-series, the NESAP, Saanich Inlet, and the ETSP, and nonrandom patterns of co-occurrence between microbial taxa associated with oxic, dysoxic, suboxic, or anoxic water column conditions are evident (Figure 1.9). In this analysis, dysoxic, suboxic, and anoxic subnetworks are dominated by OTUs representing the SUP05–Arctic96BD-19, SAR11, Marine Group A and SAR324 clades, consistent with members of these clades having overlapping habitat preferences and the potential for metabolite exchange. Moreover, OTUs representing these taxa were identified as hubs in the larger network on the basis of the number of connections running through them, consistent with previous reports of 'keystone' connectivity in marine ecosystems (Steele et al., 2011; Gilbert et al., 2012). Interestingly, the most abundant bacterial OTUs in the network are not typically the most connected, and different OTUs for several taxonomic groups participated in multiple subnetworks, suggesting that the overall network consists of many functionally redundant modules with the potential to change over time.

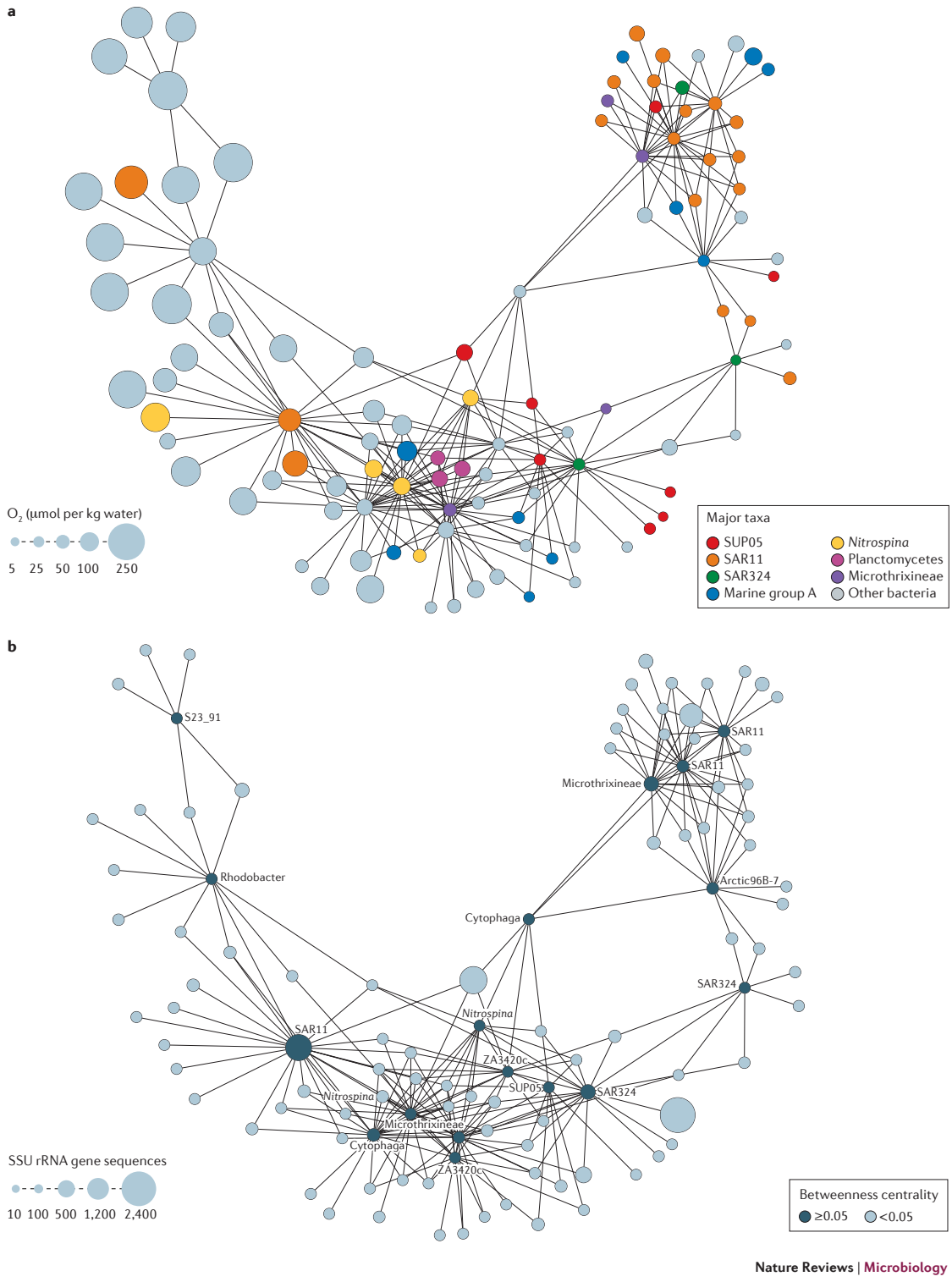


Figure 1.9 Co-occurrence networks: correlations among bacterial OTUs in different OMZs

a. The network of interactions between the operational taxonomic units (OTUs) identified in Figure 1.4 and found in various oceanic oxygen minimum zones (OMZs) described in Figures 1.4 and 1.5. Dominant bacterial OTUs are shown as per the key. Nodes are sized according to the weighted average O_2 concentration across all samples where that OTU is found. Each node represents a different OTU, although multiple OTUs can belong to the same taxa. The left side of the network consists of oxic subnetworks; dysoxic and suboxic subnetworks are present in the centre; and anoxic subnetworks are in the upper right corner. **b.** The betweenness data for this network. Nodes exhibiting a betweenness centrality of ≥0.05 (that is, those that are statistically likely to be central to the network) are highlighted. Node sizes are based on the total number of small-subunit ribosomal RNA (SSU rRNA) gene sequences belonging to that OTU summed across all OMZ samples. Data for this figure were derived from sequences deposited in Genbank. Figure used with permission from Nature Reviews Microbiology (Wright et al., 2012).

1.7 Future directions in OMZ microbiology research

We live on an ocean-dominated planet, and the collective metabolic expression of cellular life in the ocean has a profound influence on the evolution of the biosphere. Cellular life in the ocean is in turn dominated by microbial communities that form interaction networks, which are both resilient and responsive to environmental perturbation. Over geological timescales, recurring changes in the oxygenation status of the ocean have resulted in multiple biotic crises with concomitant changes in marine ecosystems and climate balance (Falkowski et al., 2011). Available monitoring data suggest that OMZ expansion in the modern ocean is consistent with a renewed period of change (Keeling et al., 2010). When viewed from an Earth systems perspective, these observations take on immediate significance as we consider the potential impacts of OMZ expansion on marine resources and global warming trends. These impacts include reduced biodiversity and food security and increased production and transport of radiatively active trace gases owing to changes in microbial interaction networks (Rabalais et al., 2010; Naqvi et al., 2010). Determining how these interaction networks form, function, and change over time reveals otherwise hidden links between microbial community structure and higher-order ecological and biogeochemical processes. Indeed, over the past few years, plurality sequencing combined with process rate analyses and targeted gene surveys in coastal and open-ocean OMZs has identified conserved patterns of microbial community structure and function, and has uncovered novel modes of metabolic integration that couple carbon, nitrogen and sulfur cycles. These findings have important implications for our understanding of the nutrient and energy flow patterns in expanding marine OMZs. Looking forwards, comparative studies are needed to define the shared

or specialized metabolic subsystems that mediate microbial community responses to changing levels of O_2 deficiency in the water column in different oceanic provinces. Additional time series monitoring studies combining gene expression and process rate measurements are also needed to validate pathway predictions and provide parameters for regulatory and network dynamics for more effective ecosystem modelling. An effective human adaptation and response to OMZ expansion, ranging from our environmental management to our policy towards Earth systems engineering, may depend on our collective capacity to understand and mimic the problem-solving power of the symbiotic ocean.

1.8 Dissertation study site: the Line P transect of the Northeast subarctic Pacific Ocean

The Northeast subarctic Pacific Ocean (NESAP) contains the largest and least studied suboxic OMZ in the global ocean (Paulmier et al., 2008). The study site selected for this dissertation, Line P, is a 1425 km survey line of the NESAP, originating in the coastal fjord Saanich Inlet, B.C. (SI; 48°N, 123°W), and terminating at Ocean Station Papa (OSP, also known as station P26; 50°N, 145°W), on the southeast edge of the Alaskan Gyre (Peña and Bograd, 2007; Peña and Varela, 2007) (Figure 1.10). For over 50 years, hydrographic data have been collected along the Line P transect, making it one of the longest running time-series in the global ocean (Whitney and Freeland, 1999; Freeland, 2007). The datasets collected at the 28 stations along this transect suggest that it is representative of the NESAP as a whole (Freeland, 2007).

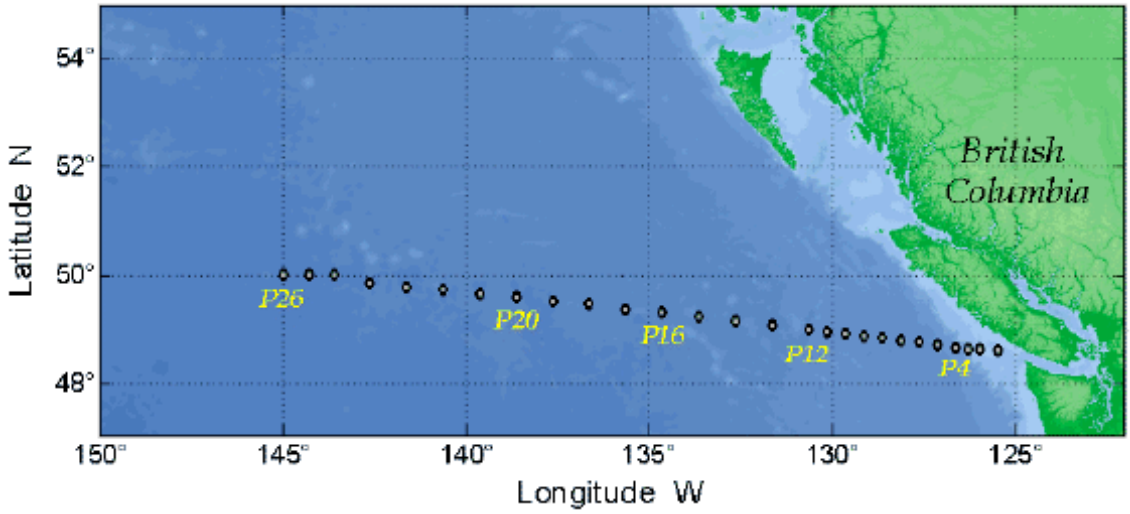


Figure 1.10 Major stations highlighted on the Line P transect

Figure obtained from the public domain at www.pac.dfo-mpo.gc.ca

1.8.1 Biological features of Line P

An important feature of Line P is that it traverses three distinct oceanic regions that can be differentiated based on macronutrient supply and utilization: (i) *coastal waters*, which extend ~75 km across the continental shelf and where productivity is stimulated during summer by periods of upwelling, (ii) a *'transition'* area unaffected by coastal upwelling and which may experience nitrate (NO_3^-) depletion in summer, and (iii) an *open ocean* region characterized by high macronutrients and low chlorophyll *a* (HNLC) (Whitney et al., 1998) (Figure 1.11). Coastal waters are rich in iron and support a high biomass of predominantly large (> 50 µm) centric diatoms (e.g. *Chaetoceros* spp. and *Thalassosira* spp.) whose productivity is seasonally limited by nitrogen availability (Taylor and Haigh, 1996; Harris et al., 2009). In contrast, phytoplankton productivity in the HNLC region is limited by the availability of iron (Martin and Fitzwater, 1988; Boyd et al., 2004; Harrison et al., 1999), and in winter can be co-limited by iron and irradiance (Maldonado et al., 1999). The resulting phytoplankton assemblage in the HNLC region is low in biomass and dominated by small-celled (<5 µm) phytoplankton (e.g. the cyanobacterial *Synechococcus* spp.; (Ribalet et al., 2010)), which are continuously grazed by rapidly growing microzooplankton (Miller et al., 1991; Frost, 1991). Higher diatom abundances are occasionally observed in the HNLC region associated with episodic iron supply (Boyd et al.,

1996; Boyd and Harrison, 1999). The transition zone is characterized by relatively high phytoplankton biomass and production over the year, and phytoplankton assemblages in summer typically resemble iron-enriched HNLC communities dominated by small-cells (Boyd and Harrison, 1999; Ribalet et al., 2010).

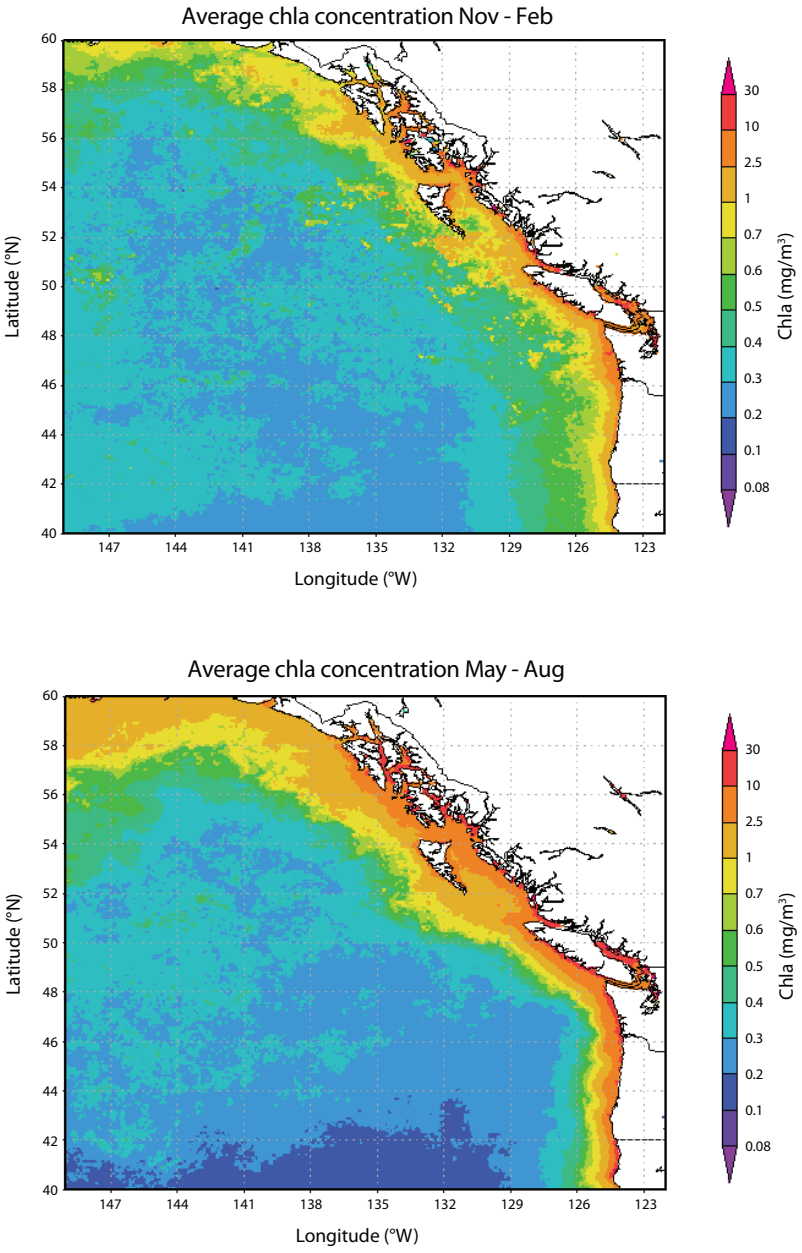


Figure 1.11 2D spatial maps of satellite derived chlorophyll a concentrations in the Northeast subarctic Pacific Ocean

(a) November to February, and (b) May to August, both averaged between 1997 and 2013. Satellite data were obtained from the NASA Aqua-Modis sensor with 4x4 km resolution. Stations P4, P12, and P26 along the Line P oceanographic transect are highlighted.

Previous studies of seasonal variability in nutrient and phytoplankton dynamics in NESAP surface waters have reported that coastal regions of the Line P transect are characterized by the classical seasonal cycle of spring and summer blooms (primary production >3 g C m^{-2} d^{-1}) whereas HNLC regions display relatively low seasonality in biomass and primary production (mean winter production 0.3 g C m^{-2} d^{-1}, mean spring/summer production 0.85 g C m^{-2} d^{-1}) (Boyd and Harrison, 1999; Peña and Varela, 2007). Relative to classic sub-polar regions such as the North Atlantic Ocean, seasonal variability in primary production in the NESAP is low, primarily due to iron limitation (Parsons and Lalli, 1988). The fate of phytoplankton biomass in the coastal region is likely sedimentation by diatom-dominated spring blooms, with recycling via the microbial food web predominating at other times of year (Boyd and Harrison, 1999). Phytoplankton fate in the HNLC region is most likely via recycling through the microbial food web, with relatively low sedimentation compared to the coastal region (Boyd and Harrison, 1999). Although sedimentation is lower in the HNLC region, fluxes of all biogenic materials (including carbon, nitrogen, and silica) exported to the deep ocean (>3800 m) in HNLC waters at station P26 do show distinct seasonality, with a winter minimum in total mass flux of 38 mg m^{-2} d^{-1} in February and a summer maximum of 150 mg m^{-2} d^{-1} in May/June and in August (Wong et al., 1999). Fluxes are dominated by *in situ* biological sources, with little influence from terrigenous or aeloean sources (Wong et al., 1999).

Growth rates of heterotrophic bacteria measured in the euphotic zone across the Line P transect are reported to be low (<0.1 d^{-1}) compared to growth rates of phytoplankton (0.1 – 0.8 d^{-1}) (Kirchman et al., 1993; Sherry et al., 1999). Some studies have concluded that bacterial growth rates (BGR) are limited by low temperatures and relatively low supply of dissolved organic matter (DOM) (Kirchman et al., 1993; Sherry et al., 1999), while other studies provide evidence that low iron availability may influence bacterial metabolism, leading to a reduction in growth efficiency (Tortell et al., 1996). Bacterial biomass (BB) and bacterial production (BP) in the euphotic zone appear to vary little across the Line P transect in winter (~12 μg C L^{-1} and ~0.5 μg C L^{-1} d^{-1}, respectively), while these parameters are much more variable between coastal and

HNLC stations in spring and summer, averaging approximately double winter amounts (up to ~34 µg C L^{-1} and ~6 µg C L^{-1} d^{-1}, respectively) (Sherry et al., 1999; Doherty, 1995). In comparison to other regions of the global ocean, winter BB and BP in the NESAP are similar to the Equatorial Pacific and roughly 4 – 5 fold greater than in the Sargasso Sea (Kirchman et al., 1995; Carlson and Ducklow, 1996; Sherry et al., 1999). During summer BB and BP in the NESAP increase by ~ 2-fold relative to winter, whereas these parameters show less seasonality in the Sargasso Sea and Equatorial Pacific (Kirchman et al., 1995; Carlson and Ducklow, 1996; Sherry et al., 1999).

1.8.2 Physical features of Line P

The ocean along Line P is an area of relatively weak currents, with the terminal station (P26) usually located within the Alaska Gyre (Freeland, 2007) (Figure 1.12). The North Pacific Current flows approximately parallel to Line P towards the west coast of North America, where it bifurcates near Vancouver Island into a northward branch (forming the Alaskan Stream) and a southward branch (forming the California Current) (Freeland, 2007).

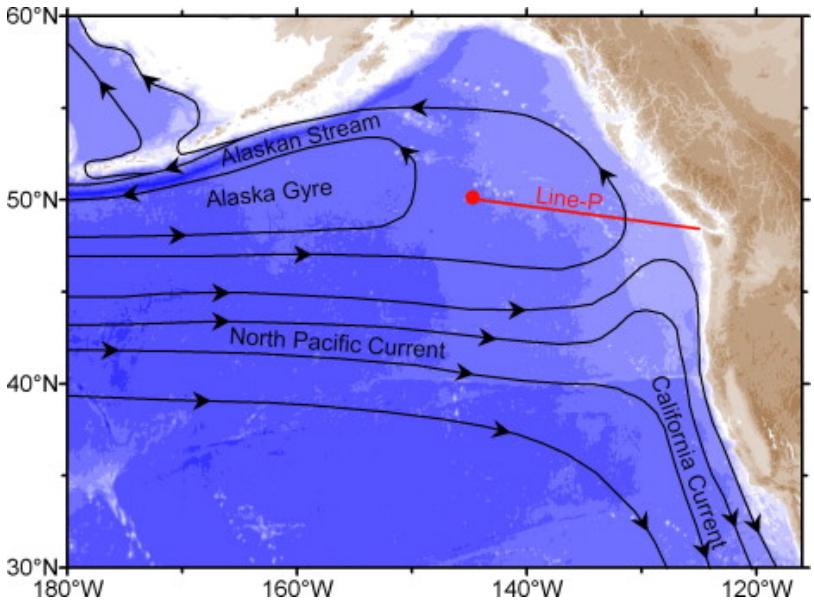


Figure 1.12 The relationship of Line P to the dominant ocean current systems

Figure used with permission from Progress in Oceanography (Freeland, 2007).

Near-surface regions of the NESAP are characterized by strong density stratification due to low salinity surface waters that are mixed to a maximum depth of 125 – 150 m during winter months, with a minimum mixing depth of ~40 m in summer months (Freeland et al., 1997; Whitney et al., 1998). As such, the interior regions of the NESAP are insulated from the atmosphere, creating a vast OMZ centered at 1000 m with oxyclines extending from ~400 – 2000 m with O_2 concentrations ranging between ~9 – 60 μmol kg^{-1}. These O_2-deficient interior waters are sourced in the Sea of Okhotsk located in the western Pacific Ocean north of Japan, where well ventilated winter waters submerge into the interior of the Pacific and travel eastward, becoming isolated from the atmosphere and forming the North Pacific Intermediate Waters (NPIW) (Whitney et al., 2007). Further east, the NPIW mix with the O_2-deficient subtropical subsurface waters (SSW), and O_2 is further depleted by microbial remineralization of organic matter sinking down from productive surface waters along the North American continental margin, creating a west-east gradient of declining O_2 concentrations (Whitney et al., 2007). Thus, the strength of the OMZ is affected by both the biological demand for O_2 imposed by microbial respiration, as well as by the relative inputs of O_2 by the NPIW and the SSW.

The major source of inter-annual variability along Line P arises during and following El Niño Southern Oscillation (ENSO) events (Whitney et al., 1998; Wong et al., 1995; Whitney and Freeland, 1999). El Niño events are characterized in this region by strong southerly winds bringing anomalously warm and fresh waters to the British Columbia and Oregon coasts (DFO 2011). Biogenic fluxes to the deep ocean have been observed to increase by up to 49% during warm El Niño years, associated with increases in primary production (Wong and Matear, 1999). On longer time scales, variability along Line P is affected by the Pacific Decadal Oscillation (PDO), defined as the leading principal component of North Pacific monthly sea-surface temperature variability pole-ward of 20ºN (Mantua et al., 1997; Peña and Varela, 2007).

Longer-term variability has also been observed in the NESAP associated with climate forcing. During the past 50 years, coastal and open ocean surface waters of the NESAP have both warmed and freshened as a consequence of global climate change, resulting in an increased water-column density gradient and strengthening stratification (Whitney et al., 2007). In addition, changes in ocean circulation and ventilation have resulted in a slowing of NPIW formation, decreasing the amount of O_2 reaching interior waters of the NESAP and causing an

expansion of the hypoxic boundary layer (Whitney et al., 2007). Specifically, from 1956 – 2006 O_2 concentrations within the OMZ declined by 22% and the hypoxic boundary layer (defined as ~60 μmol kg^{-1}) shoaled from 400 m to 300 m in depth (Whitney et al., 2007). In coastal waters west of Vancouver Island, wind patterns and divergence of surface waters to the north and south create an upwelling regime that brings up nutrient-rich subsurface waters (Whitney and Freeland, 1999). Under present conditions, this coastal upwelling draws oxygenated waters from depths of 100 m to 250 m, but continued expansion of the OMZ may transport O_2-depleted waters, with major consequences for coastal ecosystems including fisheries (Whitney et al., 2007). Indeed, hypoxia-induced fish and crab kills have already been observed along the Oregon, Washington, and British Columbia coastlines (Grantham et al., 2004). Beyond these impacts on higher trophic levels, much more significant changes may also occur due to the expansion of microbial communities populating coastal and open ocean OMZs, with concomitant changes in the flow of carbon and other nutrients between trophic levels. In order to study responses of microbial community structure and function to declining O_2 in the NESAP, baseline assessments of the structure and function of extant communities must be established.

1.8.3 Microbial ecology of the oxygen minimum zone in the Northeast subarctic Pacific Ocean

Over the last decade, taxonomic and functional gene surveys coupled with gene expression studies and measurements of process rates have begun to shed light on the microbial diversity and metabolism of dominant microbial groups residing in OMZs and driving vital biogeochemical cycles. However, microbial community structure and function within the world's largest permanent suboxic OMZ, located in the NESAP, has yet to be characterized. Preliminary surveys of microbial community structure in the NESAP OMZ highlighted patterns of microbial community composition that are conserved in other well-studied OMZ systems. These results provide further evidence to suggest that the NESAP OMZ is a suitable model system for studying microbial community responses to declining O_2, and that results of such studies will be extensible to other OMZ systems. Dominant and conserved microbial groups identified in the NESAP OMZ include but are not limited to the Alphaproteobacterial cluster SAR11, the candidate bacterial phylum Marine Group A (MGA), the Deltaprotebacterial cluster SAR324,

and Marine Group I Thaumarchaeota. Of the dominant bacterial groups, MGA was identified as being the most diverse, most abundant, and least well understood, and was highlighted as a target for further study. Prior to this dissertation, little to nothing was known about the diversity, distribution, population structure, or metabolism of MGA bacteria in OMZs or in the ocean at large.

1.9 Thesis objectives

The overall goal of this dissertation was to describe the microbial community structure of the Northeast subarctic Pacific Ocean (NESAP), with a specific emphasis on characterizing the ecology of Marine Group A (MGA), the most abundant bacterial group present in O_2-deficient interior waters of the NESAP.

This dissertation had the following objectives:

1. Synthesize current knowledge regarding microbial community structure and function in oxygen minimum zones in a global and comparative context
2. Describe the overall microbial community structure within surface and O_2-deficient interior waters of the NESAP and identify co-occurrence patterns among extant microbial groups (in particular, patterns involving MGA)
3. Establish the diversity, distribution, and population structure of MGA bacteria in the NESAP
4. Assess the metabolic capacity of MGA bacteria

To address these objectives, the dissertation is structured as follows:

In Chapter 1, I reviewed what is known about the microbial ecology and biogeochemistry of oxygen minimum zones on a global scale.

In Chapter 2, I present a detailed survey of microbial community structure in the NESAP at two time points and over a range of depths, based on traditional ecological analyses and a novel microbial co-occurrence network analysis.

In Chapter 3, I describe the diversity, distribution, and population structure of MGA bacteria in the NESAP as determined by small subunit ribosomal rRNA (SSU rRNA) gene sequencing of clone libraries and 454-pyrotags, and by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH).

In Chapter 4, I describe the genomic analysis of large-insert DNA fragments derived from MGA bacteria living in the NESAP and other North Pacific Ocean environments, with an emphasis on insights into MGA energy metabolism.

In Chapter 5, I integrate the results of Chapters 1 – 4, discuss the significance of this work, and suggest avenues for future research.

Chapter 2: Microbial community structure and co-occurrence network architecture in the Northeast subarctic Pacific Ocean

2.1 Synopsis

This chapter presents the first comprehensive survey of microbial community structure in the Northeast subarctic Pacific Ocean (NESAP), and applies 454-pyrotag sequencing of small subunit ribosomal RNA (SSU rRNA) genes to document the abundance, distribution, and patterns of co-occurrence of rare through abundant microbes affiliated with all three domains of life (*Archaea, Bacteria, Eukaryota*) in this region. Standard ecological metrics are applied to compare and contrast patterns in abundance and distribution of microbes across datasets collected at two time points (August 2007 and February 2010) throughout surface (10 m) and mesopelagic, O_2-deficient waters (500 – 2000 m) located within the oxygen minimum zone (OMZ). These analyses are followed by the application of novel techniques derived from network theory to identify co-occurrence patterns among microbial groups that might be indicative of ecological interactions occurring within NESAP microbial communities. In this chapter, the candidate phylum Marine Group A is identified as the most abundant bacterial group residing in OMZ waters of the NESAP. Marine Group A bacteria are documented to play a central role in structuring the microbial co-occurrence network, suggesting an important function for these little-known organisms in this region.

2.2 Materials and Methods

2.2.1 Sample collection and processing

Sampling was conducted via multiple hydrocasts using a rosette water sampler, with an attached Conductivity, Temperature, Depth (CTD) sensor aboard the *CCGS John P. Tully* during Line P cruises 2007-15 and 2010-01 in the NESAP in August 2007. Major stations sampled include: P4 [48º39.0N, 126º4.0W] – August 16th, P12 [48º58.2N, 130º40.0W] – August 18th, and P26 [50ºN, 145ºW] – August 22nd) and in February 2010 (Major stations: P4 [48º39.0N, 126º4.0W] – February 4th, P12 [48º58.2N, 130º40.0W] – February 11th, P16 [49º17.0N, 134º4.0W] – February 7th, and P20 [49º34.0N, 138º40.0W] – February 8th and 9th). Station P26 was not sampled in

February 2010 due to poor weather conditions. At the aforementioned major stations, 20 L samples for DNA isolation were collected from the surface (10 m), while 120 L samples were taken from three depths spanning the OMZ core (1000 m) and upper (500 m) and deep (1300 m at P4, 2000 m at all other stations) oxyclines. Sample collection and filtration protocols can be viewed as visualized experiments at http://www.jove.com/video/1159/ (Zaikova et al., 2009) and http://www.jove.com/video/1161/ (Walsh et al., 2009), respectively. The CTD-mounted O_2 probe (Model SBE 43, Sea-Bird Electronics, Bellevue, WA) reported O_2 concentrations in μmol kg^{-1}. Seawater samples for nutrient analysis were collected in 16 x 125 mm polystyrene test tubes and analyzed at sea (stored at 4 °C and in the dark for < 12 hrs prior to analysis) using an Astoria Analyzer (Astoria-Pacific, Clackamas, OR) as described by (Barwell-Clarke and Whitney, 1996).

2.2.2 Enumeration of cells by flow cytometry

Cells were enumerated by flow cytometry using samples fixed with formaldehyde (final concentration of 4% wt/vol) and stored at 4 ºC for 7 to 14 days until analysis using SYBR Green I (Invitrogen, Carlsbad, CA) on a FACS LSRII (Becton Dickonson, Franklin Lakes, NJ) (Zaikova et al., 2010). For flow cytometric analysis, a 500 ml sample was incubated with 5 ml of a 10 000-fold dilution of SYBR Green I (nucleic acid stain; Invitrogen, Carlsbad, CA) overnight at 4 °C in the dark. Cells were counted with a FACS LSRII (Becton Dickonson, Franklin Lakes, NJ) equipped with an air-cooled argon laser (488 nm, 15 mW). Stained cells, excited at 488 nm, were identified and enumerated according to their right angle scatter (SSC) and green fluorescence (FL1) emission measured at 530 nm ± 30 nm. The exact volume analysed and subsequent estimation of cell concentrations were calculated by the addition of a know concentration of 6 mm fluorescent beads (Invitrogen).

2.2.3 Environmental DNA extraction

DNA was extracted from sterivex filters as described in (Zaikova et al., 2010) and (DeLong et al., 2006). To concentrate microbial biomass for downstream environmental DNA (eDNA) extraction, the 20 – 120 L samples were filtered through 47-mm Whatman GF/D prefilters (2.7

μm nominal cut-off) in-line with 0.22 μm Sterivex-GV filters (Millipore, Billerica, MA) using a Masterflex L/S 7553-70 peristaltic pump (Cole-Parmer, Montreal, QC). After filtration, 1.8 ml of storage/lysis buffer (40 mM EDTA pH 8.0, 50 mM Tris pH 8.3, 0.75 M Sucrose) was added to each filter prior to storage at -80 °C. Sterivex filters were allowed to thaw on ice prior to the addition of 50 ml of lysozyme (0.125 mg ml^{-1}). Filters were then incubated for 1 h at 37 °C with intermittent mixing followed by the addition of 50 ml Proteinase K (Qiagen, Germantown, MD) and 100 ml 20% SDS. Samples were then incubated at 55 °C for 1 h with intermittent mixing. At the end of the incubation period cell lysate was removed using a 3 ml syringe. Filters were rinsed with 1 ml of lysis buffer that was then added to the original lysate. An equal volume of phenol:chloroform : IAA (25:24:1, pH 8.0) was added to the lysate and gently mixed. The aqueous layer was collected and an equal volume of chloroform:IAA (24:1) was added. After gentle mixing the aqueous layer was collected and loaded onto Microcon Centrifugal filter devices (0.5 ml) (Millipore), washed three times with 2 ml TE buffer (pH 8.0), and concentrated to a final volume of 180 ml. Total DNA concentrations were quantified on a Nano-Drop Spectrophotometer (NanoDrop, Wilmington, DE) using ~2 ml of sample. DNA quality was determined by running 5 ml of each sample along with 100, 250 and 500 ng of HindIII ladder (New England Biolabs, Ipswich, MA) on 1% agarose gels in 1X TBE overnight at 15 V. The DNA extraction protocol can be viewed as a visualized experiment at http://www.jove.com/video/1352/ (Wright et al., 2009).

2.2.4 454-pyrotag amplification, sequencing, and analysis

PCR amplification of SSU rRNA gene for pyrotag sequencing

The V6-V8 region of the SSU rRNA gene (from *Archaea*, *Bacteria*, and *Eukaryota*) was amplified from August 2007 and February 2010 DNA samples using primers 926F [with addition of extra wobble in position 47] (5'-cct atc ccc tgt gtg cct tgg cag tct cag AAA CTY AAA KGA ATT GRC GG-3') and 1392R (5'-cca tct cat ccc tgc gtg tct ccg act cag-<**XXXXX**>-ACG GGC GGT GTG TRC-3'). Primer sequences were modified by the addition of 454 A or B adapter sequences (lower case). In addition, the reverse primer included a 5 bp barcode designated <**XXXXX**> for multiplexing of samples during sequencing. Twenty-microlitre PCR reactions were performed in duplicate and pooled to minimize PCR bias using 0.4 μL Advantage

GC 2 Polymerase Mix (Advantage-2 GC PCR Kit, Clonetech, Mountainview, CA), 4 µL 5X GC PCR buffer, 2 µL 5M GC Melt Solution, 0.4 µL 10mM dNTP mix (MBI Fermentas, Glen Burnie, MA), 1.0 µL of each 25 nM primer, and 10 ng sample DNA. The thermal cycler protocol was 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 50 °C for 45 s, and 68 °C for 90 s, and a final 10-min extension at 68 °C. PCR amplicons were purified using SPRI Beads and quantified using a Qubit fluorometer (Invitrogen). Samples were diluted to 10 ng/µL and mixed in equal concentrations. Emulsion PCR and sequencing of the PCR amplicons were performed at the Department of Energy Joint Genome Institute (Walnut Creek, CA) following the Roche 454 GS FLX Titanium (454 Life Sciences, Branford, CT) technology according to the manufacturer's instructions.

Processing of pyrotag sequences

A total of 378 796 pyrotag sequences were generated from 26 discrete samples from the NESAP water column (using the sampling and sequencing protocols mentioned above) and analysed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (Caporaso et al., 2010). Reads with length shorter than 200 bases, ambiguous bases, and homopolymer runs were removed prior to chimera detection. Chimeras were detected using the chimera slayer provided in the QIIME software package and removed prior to taxonomic analysis. All non-chimeric sequences were phylogenetically identified in QIIME using a BLAST-based assignment method and clustered at 97% identity against the SILVA taxonomic database (www.arb-silva.de; (Pruesse et al., 2007)). Singleton OTUs (OTUs represented by one read) were omitted from downstream analyses, as recommended by Kunin and colleagues (Kunin et al., 2010), Tedersoo and colleagues (Tedersoo et al., 2010), and Gihring and colleagues (Gihring et al., 2012), leaving 14 567 OTUs (containing 327 555 sequences) for downstream analyses. OTU abundance information was normalized to the total number of sequences per sample. OTUs were divided into frequency classes termed abundant, intermediate, or rare. Abundant OTUs were arbitrarily defined as having a frequency >1% in at least one sample, intermediate OTUs as having a frequency ≤1% and ≥0.1% in at least one sample, and rare OTUs as having a frequency <0.1% in all samples (equal to the frequency of detecting 1 sequence in the smallest pyrotag library [1179 sequences]) (Galand et al., 2009).

2.2.5 Hierarchical cluster analysis

Hierarchical cluster analysis was performed on OTU abundance data generated in QIIME and normalized to the number of pyrotag sequences in each sample. Input data were transformed to create a Bray Curtis similarity matrix and hierarchically clustered using a Group Average algorithm in PRIMER v6.0 (Clarke, 1993; Clarke and Gorley, 2006). Dendrograms were also produced in PRIMER.

2.2.6 Indicator species analysis

Indicator species analysis (ISA) was performed in R (http://www.r-project.org/) using the indval command present in the labdsv package, with default settings and 1000 iterations (http://cran.r-project.org/web/packages/labdsv/index.html). Sample groups for ISA were defined based on fine-scale clusters identified in hierarchical cluster analysis as follows: 1. A07 10 m, 2. F10 10 m, 3. A07 500 m, 4. A07 1000+ m, 5. F10 500 m, 6. F10 1000+ m.

2.2.7 Microbial co-occurrence network construction & analysis

To identify patterns of co-occurrence among OTUs across and within domains and frequency classes, pairwise Pearson's correlation coefficients were calculated between OTUs across all 26 samples and a microbial co-occurrence network was constructed from the resulting correlation matrix. Only OTUs present in at least 25% of samples were included for calculating correlations (leaving a total of 2 727 out of 14 567 OTUs), and only interactions with strong and significant correlation coefficients ($R>0.8$, $p<0.001$) were depicted in the network. This network contained a total of 2 005 nodes (OTUs) connected by 18 905 edges with significant correlations to one another, and was visualized using the Edge-Weighted Spring Embedded layout in Cytoscape (Shannon et al., 2003). Several global properties of the network were calculated, including the degree distribution and clustering coefficient, using Network Analyzer in the software package Cytoscape (Assenov et al. 2008). The equation for calculating the global clustering coefficient (the average of the local clustering coefficients of all n vertices) for the network defined by Watts and Strogatz (1998) is:

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

where *Ci* is the local clustering coefficient for all undirected graphs in the network defined as:

$$C_i = \frac{2\left|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}\right|}{k_i(k_i - 1)}$$

where e_{jk} is an edge between vertices v_i and v_j, E is the set of edges between a set of vertices, and k_i is the number of vertices $|N_i|$ in the neighbourhood N_i of a vertex.

The equation for estimating the clustering coefficient of a random network of similar size (that is, with a similar number of nodes and edges) is defined by Watts and Strogatz (1998) as:

$$C = \frac{2|e|}{n(n-1)}$$

2.3 Results

2.3.1 Physiocochemical characteristics of the study site

Relevant physicochemical data measured along the Line P transect and related to the present study are described below. Salinity gradients ranging from 32.2 – 32.5 PSU (August 2007) and 32.2 – 32.4 (February 2010) at the surface (10 m) and 34.1 – 34.6 PSU (August 2007 and February 2010) in the ocean's interior generated a stratified water column across the Line P transect (Table 2.1). Sea surface temperatures ranged from 12.2°C – 15.3°C in August 2007 and from 6.9°C – 9.8°C in February 2010. Sea surface temperatures were slightly higher than average in February 2010, likely due to the occurrence of a strong El Niño event during this sampling period. Average O_2 concentrations were 278.2 µmol kg^{-1} (August 2007) and 283.7 µmol kg^{-1} (February 2010) at the surface, reaching a minimum of 10.3 µmol kg^{-1} (August 2007) and 8.7 µmol kg^{-1} (February 2010) between 1000 and 1100 m across the transect. Nutrient concentrations were higher in the OMZ core and the upper (500 m) and deep (2000 m) oxyclines

than at the surface at both time points. In 10 m samples, nitrate and phosphate concentrations were highest at P26 in August 2007 (10.3 µmol L^{-1} and 1.08 µmol L^{-1}, respectively) and at P20 in February 2010 (10.8 µmol L^{-1} and 1.08 µmol L^{-1}, respectively). At 1000 m, nitrate concentration was highest at P26 in August 2007 (46 µmol L^{-1}) and at P20 in February 2010 (45.6 µmol L^{-1}), while phosphate concentration was highest at P4 and P12 in August 2007 (3.31 µmol L^{-1}) and at P12 in February 2010 (3.27 µmol L^{-1}). Cruise and sample IDs referred to in the text as well as environmental parameters measured at sampling sites are listed in Table 2.1. All contextual data is available through the Canadian Department of Fisheries and Oceans (url: http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/line-p/).

Cruise ID	Station	Depth [m]	Sample ID	Sampling date [m/d/y]	Latitude [°N]	Longtitude [°W]	Temperature [°C]	Oxygen [µmol/kg]	Salinity [psu]	Nitrate[a] [µmol/l]	Phosphate [µmol/l]	Silicate [µmol/l]	Microbial cell abundance [cells/ml]
A07	P4	500	A07.P4.500m	8/16/07	48.65	126.67	5.2251	28	34.1142	ND[b]	ND	ND	2.49E+04
A07	P4	1000	A07.P4.1000m	8/16/07	48.65	126.66	3.6032	10.3	34.3705	45.6	3.31	128.1	7.84E+04
A07	P12	10	A07.P12.10m	8/18/07	48.97	130.67	15.6345	275.4	32.202	3.3	0.64	9.1	9.37E+05
A07	P12	500	A07.P12.500m	8/18/07	48.97	130.67	4.2559	38.7	34.0498	ND	ND	ND	2.46E+04
A07	P12	1000	A07.P12.1000m	8/18/07	48.97	130.67	3.1655	12.3	34.3568	45.9	3.31	143.6	4.93E+04
A07	P12	2000	A07.P12.2000m	8/18/07	48.97	130.67	1.9368	63.1	34.5913	43.7	3.09	181.1	3.69E+04
A07	P26	10	A07.P26.10m	8/22/07	50.00	145.00	12.2996	281	32.467	10.3	1.08	18.1	2.70E+05
A07	P26	500	A07.P26.500m	8/22/07	50.00	145.00	3.9197	36.4	34.0985	ND	ND	ND	2.77E+04
A07	P26	1000	A07.P26.1000m	8/22/07	50.00	145.00	2.914	15.3	34.3883	46	3.28	149.8	5.28E+04
A07	P26	2000	A07.P26.2000m	8/22/07	50.00	145.00	1.9254	58.8	34.5894	43.3	3.09	171.5	1.43E+04
F10	P4	10	F10.P4.10m	2/04/10	48.65	126.67	9.8109	275.9	32.4485	6.8	0.79	8.6	3.67E+03
F10	P4	500	F10.P4.500m	2/04/10	48.65	126.67	5.6251	40.7	34.1904	39.9	2.93	73.1	1.10E+03
F10	P4	1000	F10.P4.1000m	2/04/10	48.65	126.67	3.5949	11.9	34.3924	44	3.26	128.9	5.01E+02
F10	P4	1300	F10.P4.1300m	2/04/10	48.65	126.67	2.8773	22.7	34.6468	44.2	3.23	148.6	1.03E+03
F10	P12	10	F10.P12.10m	2/11/10	48.97	130.64	8.4139	288.9	32.3522	7.1	0.87	10.6	3.95E+03
F10	P12	500	F10.P12.500m	2/11/10	48.98	130.66	4.4952	32.2	34.0677	42.2	3.02	89.8	2.49E+04
F10	P12	1000	F10.P12.1000m	2/11/10	48.97	130.67	3.3367	11.3	34.3924	45.1	3.27	133.9	1.52E+03
F10	P12	2000	F10.P12.2000m	2/11/10	48.97	130.67	1.9434	59.5	34.5917	43.7	3.01	172.2	7.81E+03
F10	P16	10	F10.P16.10m	2/07/10	49.28	134.67	8.1619	286.2	32.4803	8.1	0.92	11.3	2.10E+03
F10	P16	500	F10.P16.500m	2/07/10	49.28	134.67	4.267	47.9	34.0497	41.3	2.94	93.7	1.40E+03
F10	P16	1000	F10.P16.1000m	2/07/10	49.28	134.66	3.1128	12	34.3521	45.2	3.25	139.5	7.42E+02
F10	P16	2000	F10.P16.2000m	2/07/10	49.28	134.66	1.9609	55.3	34.5879	43.3	3.05	172.1	1.20E+03
F10	P20	10	F10.P20.10m	2/08/10	49.57	138.67	6.8901	ND	32.4478	10.8	1.08	15.4	ND
F10	P20	500	F10.P20.500m	2/09/10	49.57	138.67	4.0585	40.2	34.0849	41.8	3	99.7	ND
F10	P20	1000	F10.P20.1000m	2/09/10	49.57	138.67	3.0297	8.7	34.3597	45.6	3.21	140.3	1.13E+03
F10	P20	2000	F10.P20.2000m	2/09/10	49.57	138.67	1.9471	53	34.5883	43.6	3.03	171.2	1.11E+03

[a]Nitrate + Nitrite
[b]Not determined

Table 2.1 Environmental variables in the NESAP during August 2007 and February 2010

2.3.2 Sampling scheme and initial pyrotag processing

To investigate the microbial community structure of the NESAP water column, 454-pyrotag sequencing was performed on small subunit ribosomal RNA (SSU rRNA) genes amplified from 26 environmental genomic DNA samples collected in August 2007 and February 2010 (see Materials and Methods). Samples were collected from four depth intervals spanning surface and O_2-deficient mesopelagic waters at three to four stations along the coastal to open ocean Line P transect, a 1425 km survey line of the NESAP originating in Saanich Inlet, British Columbia (SI; 48°N, 123°W), and terminating at Ocean Station Papa (P26; 50°N, 145°W). Samples derived from O_2-deficient regions of the water column are referred to as the upper and deep oxycline (500 m and 1300 m [at P4] or 2000 m [at all other stations], respectively) and the OMZ core (1000 m). For pyrotag sequencing, primers able to amplify the V6-V8 region of SSU rRNA genes derived from all three domains of life were used. A total of 327 609 high quality sequences were generated: 112 292 archaeal, 187 584 bacterial, 27 679 eukaryotic, and 54 that were not taxonomically identifiable (termed 'unclassified') and thus were removed from the dataset for downstream analyses (Table 2.2). The remaining 327 555 sequences were clustered into operational taxonomic units (OTUs) at a 97% similarity threshold, and 14 567 OTUs were taxonomically identified. As this method of massively parallel 454-pyrotag sequencing allows for the amplification of many low-abundance OTUs that are often not detected in traditional molecular studies (Sogin et al., 2006), differentiation between patterns of diversity and distribution exhibited by rare (low-abundance) through abundant microbial groups was enabled. As such, all 14 567 OTUs were divided into frequency classes termed 'abundant', 'intermediate', or 'rare' (Table 2.2). The breakdown of all OTUs by domain (*Archaea*, *Bacteria*, *Eukaryota*) and frequency class (abundant, intermediate, rare) is shown in Table 2.3. Proportionately, the dataset contained 10% archaeal OTUs, 76% bacterial OTUs and 14% eukaryotic OTUs; of which 1% were abundant OTUs, 5% were intermediate OTUs, and 94% were rare OTUs. As water samples in the current survey were filtered through a 2.7 μm pre-filter before reaching the 0.22 μm filter (from which DNA was extracted for pyrotag sequencing), the sampling of eukaryotic OTUs is expected to be biased against larger (>2.7 μm) protists.

Sample ID	Total # of pyrotags	# Archaeal pyrotags	# Bacterial pyrotags	# Eukaryotic pyrotags	# Unclassified pyrotags	OTU Richness
A07.P4.500m	3397	1177	2165	54	1	1202
A07.P4.1000m	1179	774	394	11	0	398
A07.P12.10m	4640	130	3715	792	3	1074
A07.P12.500m	5685	2529	3081	75	0	1588
A07.P12.1000m	7253	3196	3883	173	1	1956
A07.P12.2000m	4257	1920	2264	72	1	1222
A07.P26.10m	4583	3	2099	2478	3	982
A07.P26.500m	5626	2599	2942	85	0	1440
A07.P26.1000m	4548	1894	2420	232	2	1345
A07.P26.2000m	2941	1379	1479	81	2	858
F10.P4.10m	20610	5937	11666	3007	0	2043
F10.P4.500m	18391	5541	12414	433	3	2663
F10.P4.1000m	3045	1996	1034	15	0	511
F10.P4.1300m	10610	4917	5556	136	1	1511
F10.P12.10m	18572	2551	11792	4227	2	1981
F10.P12.500m	28174	13670	13917	587	0	2954
F10.P12.1000m	10871	6511	4345	15	0	1544
F10.P12.2000m	17393	6082	11023	287	1	2470
F10.P16.10m	18768	2165	12462	4135	6	2011
F10.P16.500m	19449	9003	9930	508	8	2347
F10.P16.1000m	15763	5977	9461	325	0	2058
F10.P16.2000m	15876	6891	8680	303	2	2062
F10.P20.10m	12829	724	8513	3592	0	1628
F10.P20.500m	25211	10581	10384	4238	8	2495
F10.P20.1000m	27380	9334	17184	862	0	3522
F10.P20.2000m	20558	4811	14781	956	10	2889
TOTAL	**327,609**	**112,292**	**187,584**	**27,679**	**54**	

[a]Number of tags after quality control and filtering steps, see Materials and Methods

Table 2.2 Number of pyrotags and OTU richness per sample

	Abundant	Medium	Rare	TOTAL
Archaea	26	116	1,284	1,426
Bacteria	39	535	10,468	11,042
Eukarya	13	132	1,954	2,099
TOTAL	**78**	**783**	**13,706**	**14,567**

Table 2.3 Frequency distribution of archaeal, bacterial, and eukaryotic OTUs classified as abundant, intermediate, and rare

2.3.3 Major taxonomic lineages identified in NESAP waters

To describe the major taxonomic groups present in the NESAP, lineages present at average abundances >1% in surface and O_2-deficient mesopelagic waters regardless of frequency class or time of sampling were identified. Bacterial subgroups with the largest proportion of affiliated pyrotags in surface samples of the NESAP included the Alphaproteobacterial cluster SAR11 (12.8±3.1% of total pyrotag sequences in a given sample), the Cyanobacterial genus *Synechococcus* (7.7±3.3%), the candidate phylum Marine Group A (MGA) (5.8±3.3%), and the Gammaproteobacterial cluster SAR86 (4.2±4.4%) (Figure 2.1). The most abundant archaeal subgroups identified in surface waters of the NESAP were affiliated with Marine Group II (MGII) Euryarchaeota of the class *Thermoplasmata* (7.9±6.3%%) and Marine Group I (MGI) Thaumarchaeota (2.5±4.7%). Abundant eukaryotic subgroups were represented by Haptophytes of the genus *Phaeocystis* (6.1±5.4%) and the genus *Chrysochromulina* (1.43±5.4%), Alveolates affiliated with the genus *Dinophysis* (2.9±1.0%), and Stramenopiles affiliated with the order *Florenciellales* (2.6±2.1%), and the genus *Aureococcus* (2.8±1.4%). Other dominant microbial subgroups present in NESAP surface waters included several bacterial subgroups affiliated with the phylum Bacteroidetes (order *Flavobacteriales*), and several Proteobacterial subgroups affiliated with the Alpha, Gamma, and Beta classes (Figure 2.1).

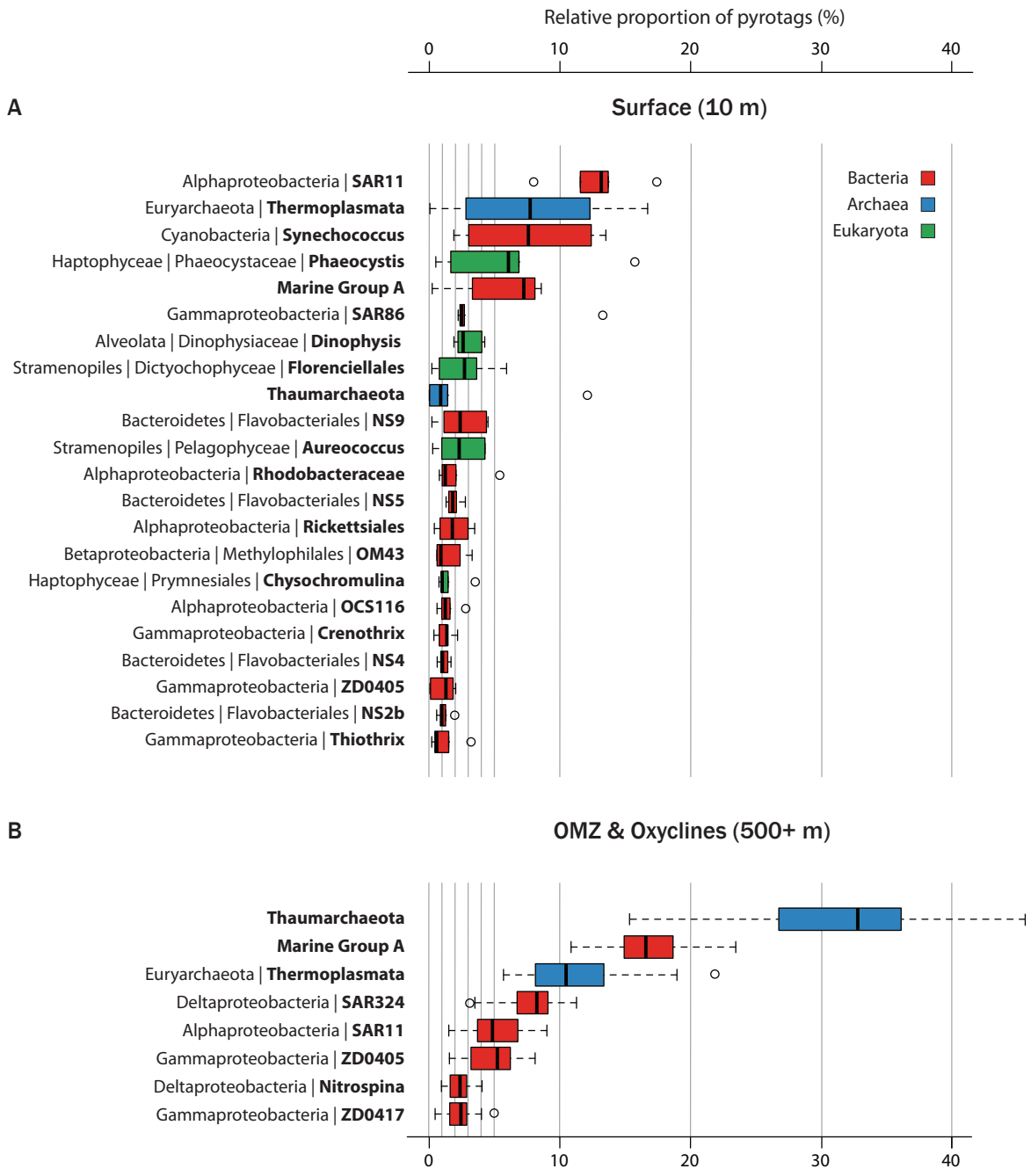


Figure 2.1 Taxonomic affiliation and distribution of abundant microbial subgroups in the NESAP
(a) surface (10 m) and (b) OMZ & Oxycline (500 m, 1000 m, 2000 m) regions of the NESAP water column averaged over all August 2007 and February 2010 samples. Left hand edge of each box denotes first quartile and right hand edge of each box denotes third quartile; data points between the first and third quartile (inside the box) fall within the interquartile range (IQR). The band inside each box denotes the median; whiskers denote data within 1.5 x IQR; hollow circles denote outlying data points.

Oxygen minimum zone (1000 m) and oxycline (500 m, 1300 m, and 2000 m) waters were dominated by archaeal subgroups affiliated with MGI Thaumarchaeota (32.2±8.3%) and MGII Euryarchaeota of the class *Thermoplasmata* (11.4±4.2%) and bacterial subgroups affiliated with MGA (16.9±3.3%), the Deltaproteobacterial cluster SAR324 (7.8±2.3%) and genus *Nitrospina* (2.5±0.9%), the Alphaproteobacterial cluster SAR11 (5.1±2.0%), and Gammaproteobacterial clusters ZD0405 (4.9±1.8%) and ZD0417 (2.4±1.2%) (Figure 2.1). Eukaryotic subgroups identified in OMZ and oxycline waters were present at abundances <1% when averaged across all mesopelagic samples.

The taxonomic structure of NESAP waters defined here based on the results of 454-pyrotag sequencing overall supports previous assessments of the bacterial and archaeal community structure of the NESAP surface and O_2-deficient regions based on small subunit ribosomal RNA (SSU rRNA) gene clone libraries, as described in Wright et al., (2012) and Ulloa et al., (2013), respectively.

2.3.4 Microbial community structure across domains and frequency classes

To compare proportions of rare through abundant components of the microbial communities in the NESAP, the distribution of all pyrotag sequences across domains and frequency classes in each sample was plotted. The majority of surface (10 m) samples from both A07 and F10 were dominated by sequences affiliated with *Bacteria* (63.1±11.4%), with the exception of the A07.P26.10m sample which was dominated by eukaryotic sequences (54.1%) (Figure 2.2). Sequences affiliated with *Eukaryota* were also well represented in remaining surface samples (20.9±5.2%, not including A07.P26.10m). Within the domain *Archaea*, F10 surface samples contained a sizeable proportion (14.9±9.9%) of sequences when compared to A07 surface samples (1.4±1.9%). There appeared to be an inverse relationship between the presences of eukaryotic and archaeal sequences from coastal (P4) to open ocean (P20) surface samples in F10, whereby the proportion of eukaryotic sequences increased (from 14.5% to 28.0%) while archaeal sequences decreased (from 28.8% to 5.6%) towards the open ocean. The proportion of bacterial sequences in all surface samples was slightly skewed towards abundant OTUs (25.9±5.8%) compared to intermediate (21.9±5.2%) and rare (15.3±2.8%) bacterial OTUs. This was also the case for archaeal and eukaryotic OTUs in most samples.

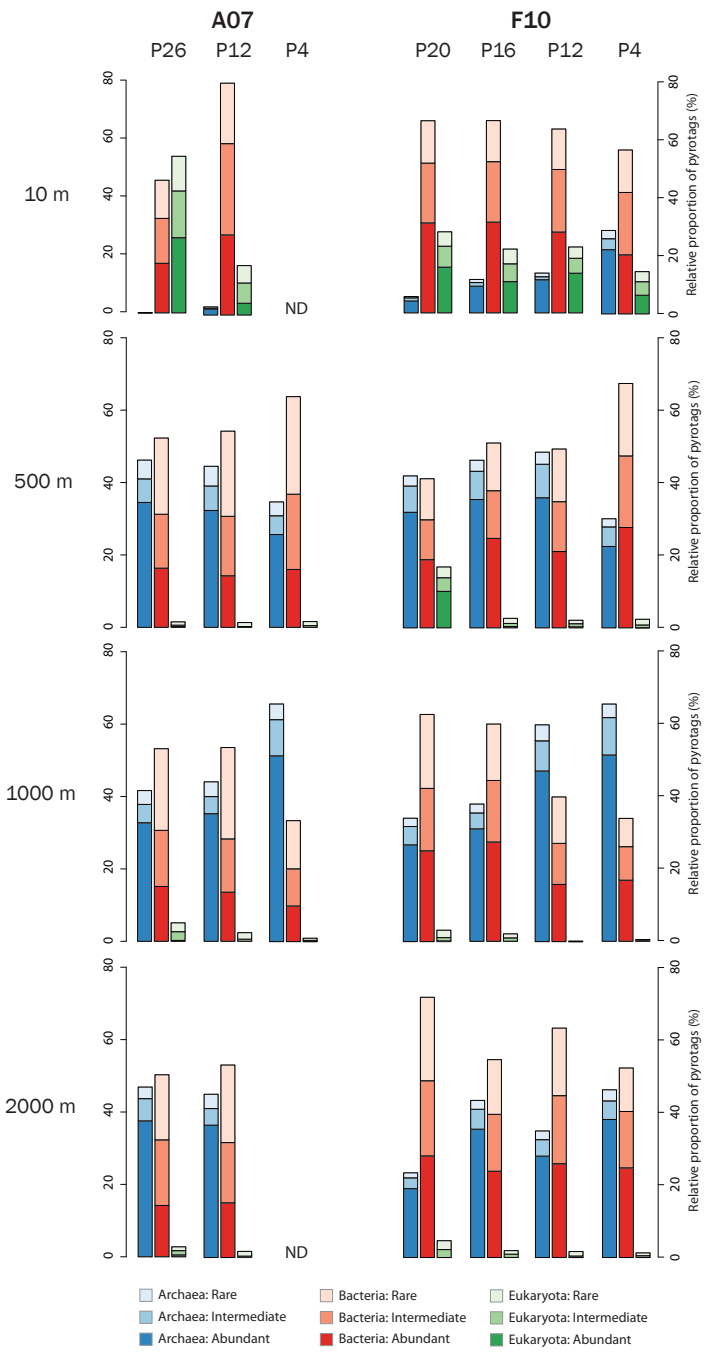


Figure 2.2 Percentage of archaeal, bacterial, and eukaryotic pyrotag sequences associated with Abundant, Intermediate, and Rare OTUs in the NESAP

Abundant (>1% frequency), intermediate (≤0.1 and ≥1%) and rare (<0.1%); from August 2007 (A07) and February 2010 (F10)

In general, the largest proportion of sequences identified in OMZ and oxycline regions of the water column were bacterial (53.1±10.3%), although a number of samples contained an

almost even distribution of bacteria and archaea and several samples contained a greater proportion of archaea (e.g. A07.P4.1000m, F10.P12.1000m, F10.P4.1000m). The average proportion of archaeal sequences across all mesopelagic samples was 44.1±10.7%). The most abundant domain / frequency class combination of OTUs belonged to abundant archaeal OTUs, which comprised ~19.0 – 51.3% of all mesopelagic seqeunces. Abundant archaeal OTUs tended to dominate the contribution of archaeal sequence space in mesopelagic samples (34.4±8.4% for abundant vs. 6.3±2.0% and 3.4±1.1% for intermediate and rare OTUs, respectively), while bacterial sequences tended to be more evenly distributed across abundant (19.7±5.7%), intermediate (15.5±3.3%), and rare (17.8±5.2%) frequency classes. The contribution of eukaryotic sequences to the taxonomic composition of most mesopelagic samples was quite small (2.8±3.5%), although an increased proportion of eukaryotic sequences (16.8%) in the F10.P20.500m sample was observed. Clear coastal to open ocean patterns in domain or frequency class structure were not observed in mesopelagic samples from either time point.

2.3.5 Hierarchical cluster analysis of community profiles

In order to define relationships among microbial communities within the NESAP occurring across time and space, hierarchical cluster analysis (HCA) of all OTUs present in all 26 samples was performed (Figure 2.3). Samples grouped into 3 broad-scale clusters (10 m, A07 mesopelagic, and F10 mesopelagic) and 6 fine-scale clusters (A07 10m, F10 10m, A07 500 m, A07 1000+ m, F10 500 m, and F10 1000+ m). From a coarse-scale perspective, mesopelagic samples obtained from varying depths (500 m – 2000 m) but at the same time point were more similar to each other (~50% identical) than samples from the same depth at different time points (~40% identical). Surface (10 m) samples were much less similar to each other across time points (~20% identical). Within each time point, A07 surface samples from transition and open ocean waters were also highly dissimilar (~28%), while F10 surface samples were relatively more similar to one another across the transect (~50%). Within most fine-scale clusters, coastal (P4) samples were least similar to transition (P12) and open ocean (P20, P26) samples.

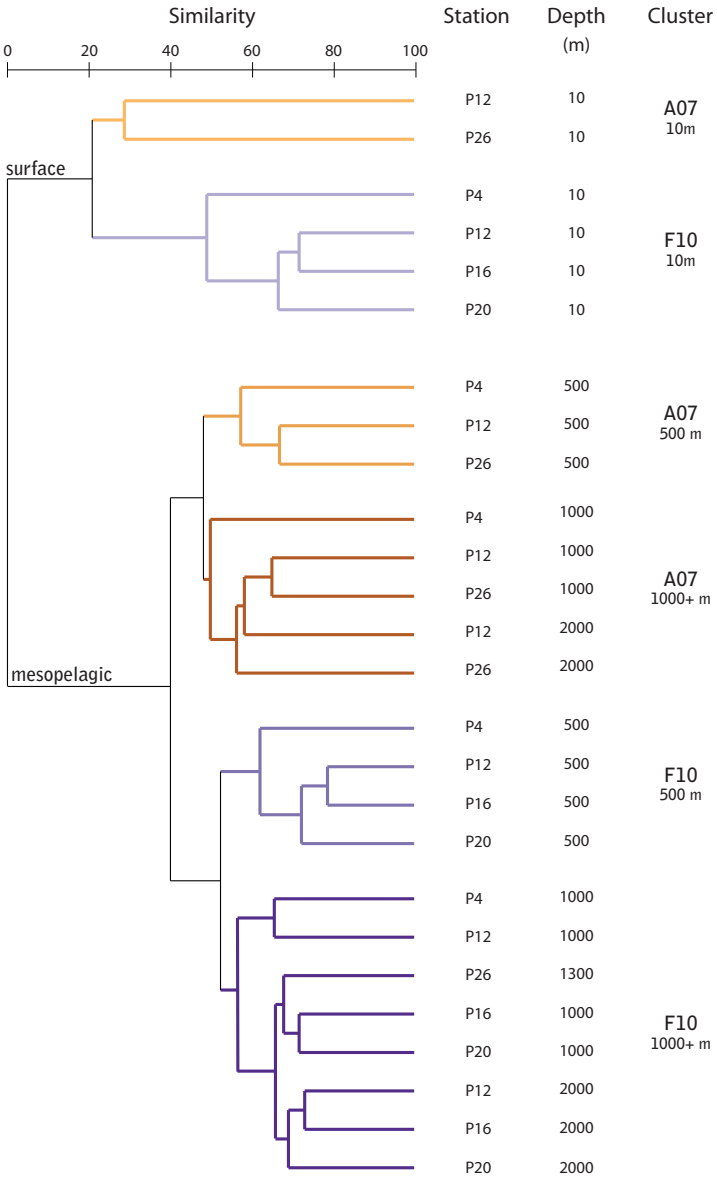


Figure 2.3 Dendrogram generated by hierarchical cluster analysis showing similarity in composition of 26 microbial communities from the NESAP

Colours highlight fine-scale clusters identified in this analysis (A07 10 m, F10 10 m, A07 500 m, A07 1000+ m, F10 500 m, F10 1000+ m). Clustering is based on a distance matrix computed with Bray-Curtis similarity and the dendrograms were inferred with the Group Average algorithm in PRIMER.

To tease apart the contribution of OTUs affiliated with each domain and frequency class to overall community structure, independent HCAs were performed for every possible combination of domain and frequency class (Figure 2.4). Including all domains, hierarchical

clustering of OTUs affiliated with each of the three frequency classes (Figure 2.4, top row) recovered the same general structure as hierarchical clustering of all OTUs. As with the clustering of all OTUs, mesopelagic samples were more similar to one another than surface samples were to one another across all frequency classes. Community structure of mesopelagic samples was most similar for OTUs of the abundant class (>80% identical for most samples) and least similar for OTUs of the rare class (~8-45% identical).

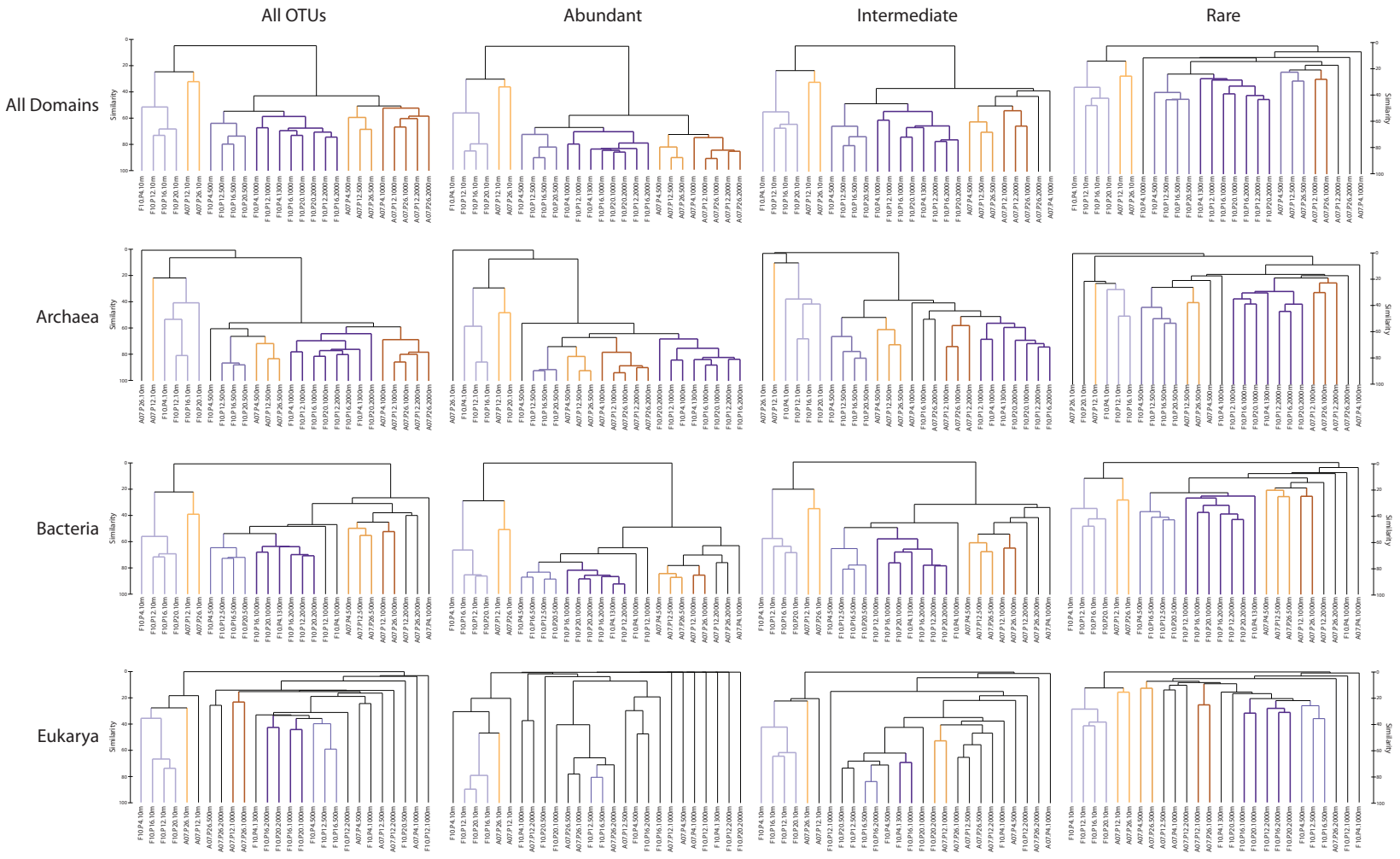


Figure 2.4 Dendrograms generated by hierarchical cluster analysis showing similarity in composition of 26 microbial communities from the NESAP. Colours highlight fine-scale clusters identified in this analysis (A07 10 m, F10 10 m, A07 500 m, A07 1000+ m, F10 500 m, F10 1000+ m); see Figure 2.3. Clustering is based on a distance matrix computed with Bray-Curtis similarity and the dendrograms were inferred with the Group Average algorithm in PRIMER.

In contrast to the clustering of all domains simultaneously, clustering of archaeal OTUs independently (Figure 2.4, second row) indicated that samples from the same compartment of the water column (i.e. 500 m vs. 1000+ m) were more closely related to one another than mesopelagic samples from the same time point. Hierarchical clustering of OTUs affiliated with the domain *Bacteria* repeated patterns of clustering observed for all OTUs across all domains in that mesopelagic samples were more closely related based on time of sampling than on depth of sampling (Figure 2.4, third row). Hierarchical clustering of OTUs affiliated with the domain *Eukaryota* indicated eukaryotic communities were highly distinct from one another in all samples (Figure 2.4, fourth row).

2.3.6 Indicator Species Analysis

In order to identify microbial OTUs indicative of fine-scale clusters defined by HCA and thus potentially driving observed patterns of community partitioning, Indicator Species Analysis (ISA) was performed. ISA permits the identification of species (referred to here as indicator OTUs, because microbial OTUs do not necessarily represent true "species") associated with or indicative of groups of samples or environments, often with a goal of identifying species as bioindicators of specific ecosystems (Dufrêne and Legendre, 1997; Bakker, 2008). A perfect indicator for a group of samples is present in all of the samples in the group, and not present in any sample outside of the group, thus receives an indicator value of 1. For the purposes of this study, the 6 fine-scale clusters defined in the HCA of all OTUs across all domains were used as indicator groups (Materials and Methods).

The complete distribution of indicator values and *p*-values calculated for all 14 567 OTUs is depicted in Figure 2.5a. In order to focus on indicator OTUs most likely to have an impact on community clustering, all indicator OTUs with indicator value >0.5 and p-value <0.01 were extracted, generating 949 significant indicator OTUs for downstream analyses (Figure 2.5a, black points). Of these significant indicators, 646 belonged to the rare frequency class, 270 to intermediate, and 32 to abundant (Figure 2.5b). The majority of significant indicator OTUs were representative of F10 clusters while fewer indicator OTUs were specific to A07 clusters. Within each time point, the number of significant indicators decreased with depth. The majority of

significant indicators for each cluster belonged to the rare class of OTUs, with the exception of indicators for the A07 1000+ m cluster that predominantly belonged to the intermediate class.

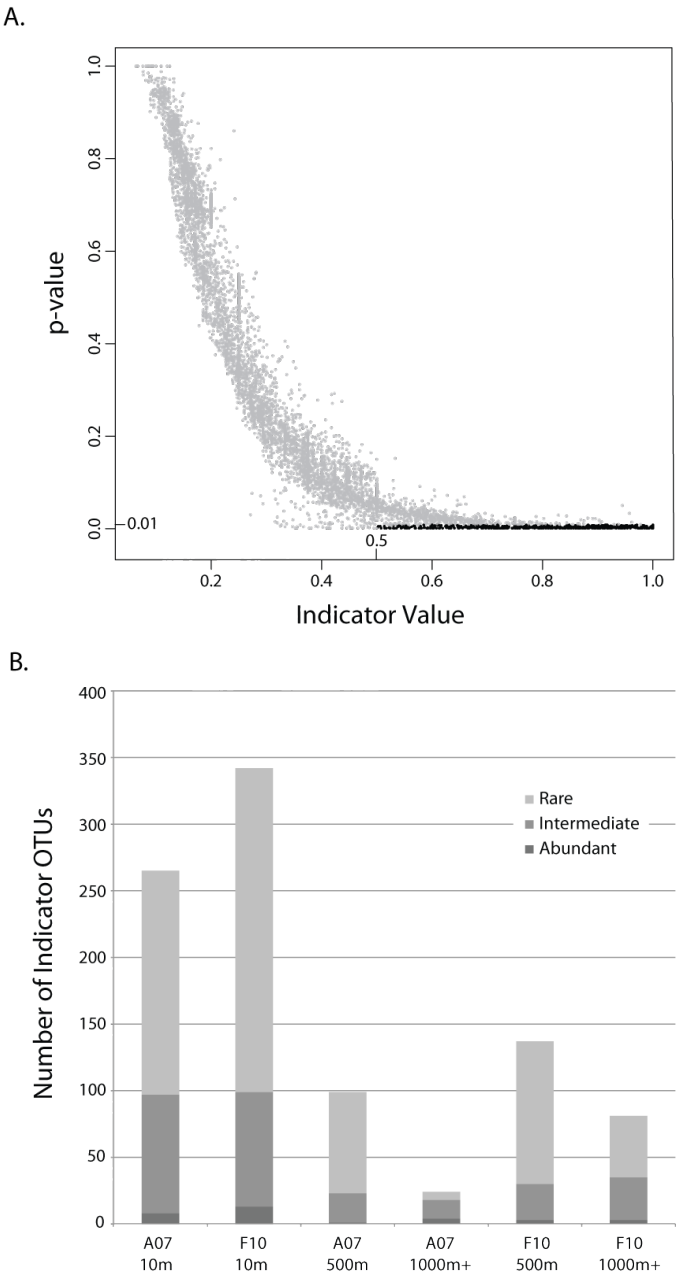


Figure 2.5 Results of indicator species analysis

(a) Distribution of indicator values with associated p-values calculated for all OTUs identified in NESAP microbial community profiles from August 2007 and February 2010. Black points indicate OTUs with indicator value >0.5 and p-value <0.01. (b) Distribution of significant indicator OTUs (indicator value >0.5, p-value <0.01) by frequency class across fine-scale clusters.

To more effectively interpret the potential ecological relevance of indicator OTUs, the taxonomic identity of significant indicator OTUs was determined (Figure 2.6). Indicator OTUs for surface samples were most frequency affiliated with Alphaproteobacteria. Indicator OTUs for surface samples affiliated with MGII Euryarchaeota and MGA were overrepresented in F10 samples (compared to A07), whereas indicator OTUs affiliated with Bacteroidetes, Beta- and Gammaproteobacteria, and eukaryotic groups were overrepresented in A07 samples (compared to F10). Indicator OTUs for mesopelagic samples were most frequently affiliated with MGA. In mesopelagic samples across all depths, indicator OTUs affiliated with Chloroflexi, MGA, and Alphaproteobacteria were overrepresented in F10 samples, whereas indicator OTUs affiliated with MGII Euryarchaeota, Bacteroidetes, and Delta- and Gammaproteobacteria were overrepresented in A07 samples. Indicator OTUs affiliated with MGA and Alpha- and Gammaproteobacteria were identified in all samples throughout the water column.

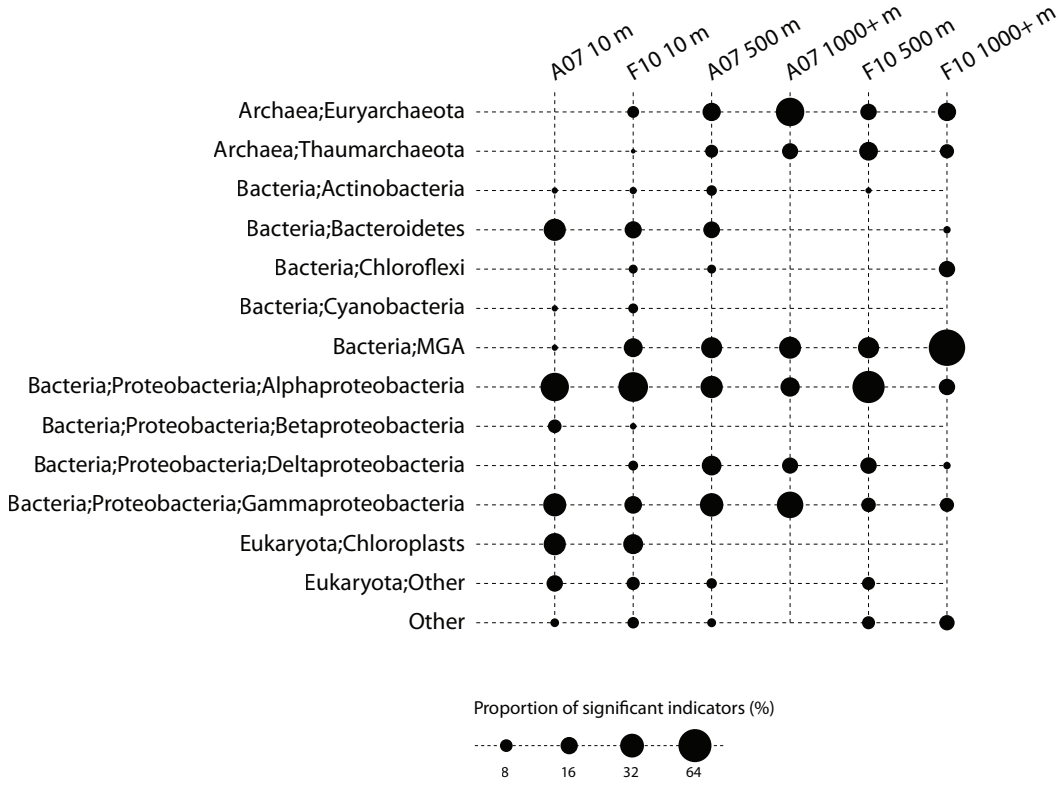


Figure 2.6 Dot plot indicating taxonomic distribution of significant indicator OTUs affiliated with each indicator group identified in NESAP microbial community profiles from August 2007 and February 2010

2.3.7 Microbial co-occurrence network analysis

It is possible to use microbial abundance data to predict microbial interactions in the environment under the premise that strongly nonrandom distribution patterns occur mostly as a result of ecological associations (Faust and Raes, 2012). To identify patterns of co-occurrence among OTUs across and within domains and frequency classes detected in the NESAP, pair-wise Pearson's correlation coefficients were calculated between OTUs across all 26 samples and a microbial co-occurrence network was generated from the resulting correlation matrix (Figure 2.7, see Materials and Methods). Only OTUs present in at least 25% of samples were included for calculating correlations, and only interactions with strong and significant correlation coefficients ($R>0.8$, $p<0.001$) were depicted in the network. This network contained a total of 2005 nodes (OTUs) connected by 18 905 edges (significant correlations). All correlations in the network were positive except for 2 (between OTUs 14934 [*Thaumarchaetota*] and 30119 [*Rhodobacteraceae*], and between OTUs 16729 [*Thaumarchaeota*] and 30119 [*Rhodobacteraceae*]). These two correlations (links) connected the two main clusters identified in the network (Figure 2.7: large cluster top right, and small cluster bottom left). Several global properties of the network were computed, including the degree distribution and clustering coefficient, using Network Analyzer in the software package Cytoscape (Assenov et al., 2008). The degree distribution indicated a scale-free nature, meaning the probability that a node has k links follows the power-law distribution $P(k) \sim k^{-b}$, where b is a degree exponent (Barabasi and Oltvai, 2004). The estimated value of b in the network was 1.154. The clustering coefficient of the network was 0.468 as opposed to 0.009 for a random graph with a similar number of nodes and edges.

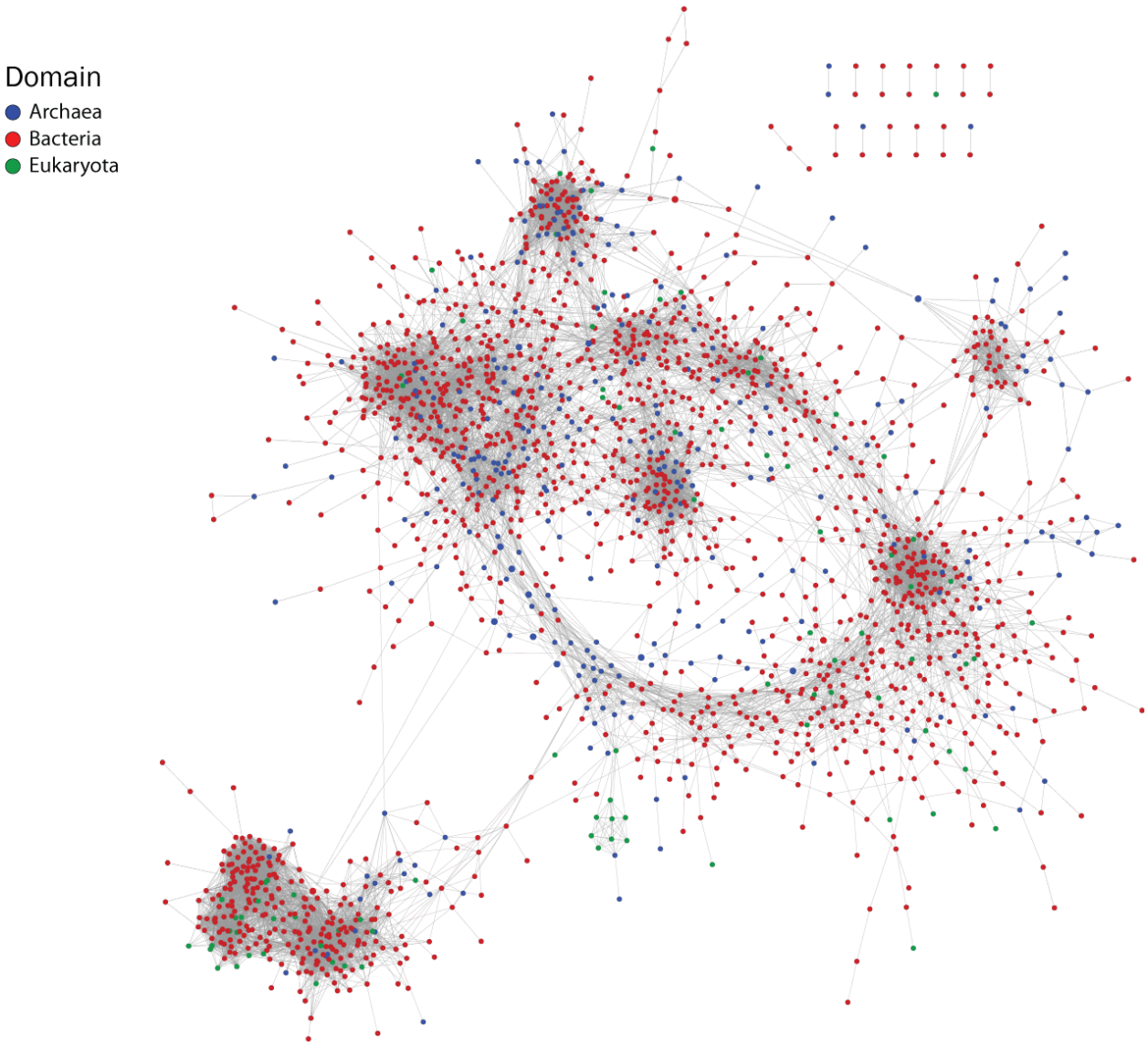


Figure 2.7 Microbial co-occurrence network, depicting all OTUs present in at least 25% of samples with significant correlation coefficients (p<0.001, R>0.8)

Nodes are coloured by domain (*Archaea*: blue; *Bacteria*: red; *Eukaryota*: green).

The network was composed of 15.0% archaeal, 79.0% bacterial and 6.0% eukaryotic nodes, which is relatively enriched in *Archaea* and depleted in *Eukaryota* compared to the distribution of domains present in the original dataset (10% archaeal, 76% bacterial; 14% eukaryotic) (Table 2.2, Table 2.4).

	Abundant	Intermediate	Rare	TOTAL
Archaea	24	100	207	331
Bacteria	36	350	1355	1741
Eukaryota	11	46	76	133
TOTAL	71	496	1638	**2205**

Table 2.4 Network node distribution by domain & frequency class

Connections between nodes affiliated with each domain were identified and the most common connections were between bacterial nodes, followed by connections between bacterial and archaeal nodes, and connections between bacterial and eukaryotic nodes (Table 2.5). With respect to frequency class, the network was composed of 3.2% abundant, 22.5% intermediate, and 74.3% rare nodes. The most common connections were between rare nodes, followed by connections between intermediate and rare nodes and connections between intermediate nodes.

	Archaea	Bacteria	Eukaryota
Archaea	591	3191	112
Bacteria	-	12503	2076
Eukaryota	-	-	432

	Abundant	Intermediate	Rare
Abundant	136	1166	749
Intermediate	-	2904	5892
Rare	-	-	8057

Table 2.5 Network link distribution by domain & frequency class

To search for relationships among domain, frequency class, and degree of connectedness in the network, the degree distribution was plotted for nodes affiliated with each domain and frequency class overall, in addition to each domain / frequency class combination of nodes present in the network. While the difference in average degree across domains was not statistically significant, eukaryotic nodes displayed the highest mean degree (25.2±20.8) of all three domains compared to bacterial (19.1±18.4) and archaeal (15.1±15.8) nodes (Figure 2.8).

On average, abundant nodes (mean degree of 31.7±24.0) displayed a slightly higher degree than intermediate (27.1±20.1) and rare (15.6±15.9) nodes, although the range of possible degrees was greater for intermediate (1 to 82) and rare (1 to 84) nodes (Figure 2.9). Again, the difference in average degree across frequency classes was not statistically significant, although this may be the result of an inadequate sample size. The node with the highest degree distribution in the network was a rare node affiliated with the domain *Bacteria* (a Gammaproteobacterium of the genus *Crenothrix*). Of all 9 domain / frequency class combinations of nodes, abundant bacterial nodes displayed the highest average degree (41.4±24.6), and rare archaea displayed the lowest (13.1±14.2) (Figure 2.10).

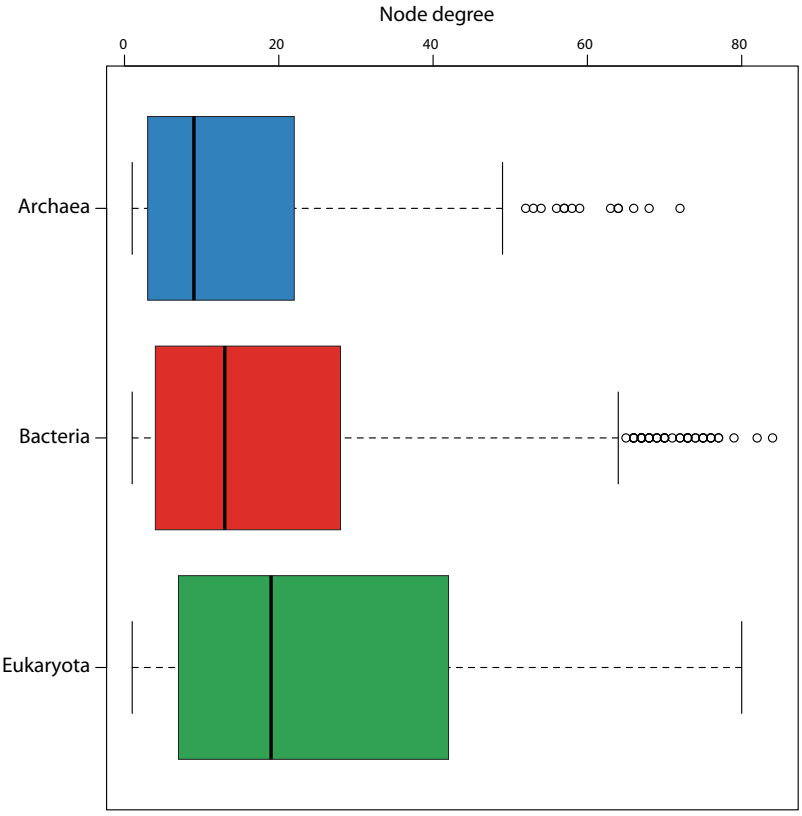


Figure 2.8 Node degree distribution by domain

(*Archaea*: blue; *Bacteria*: red; *Eukaryota*: green). Left hand edge of each box denotes first quartile and right hand edge of each box denotes third quartile; data points between the first and third quartile (inside the box) fall within the interquartile range (IQR). The band inside each box denotes the median; whiskers denote data within 1.5 x IQR; hollow circles denote outlying data points.

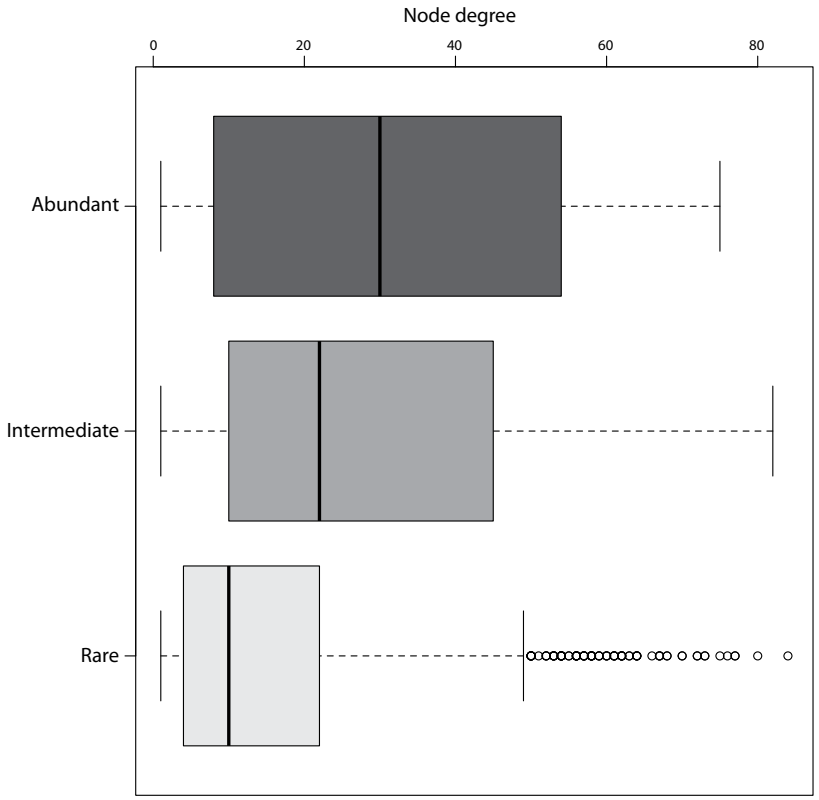


Figure 2.9 Node degree distribution by frequency class

Left hand edge of each box denotes first quartile and right hand edge of each box denotes third quartile; data points between the first and third quartile (inside the box) fall within the interquartile range (IQR). The band inside each box denotes the median; whiskers denote data within 1.5 x IQR; hollow circles denote outlying data points.

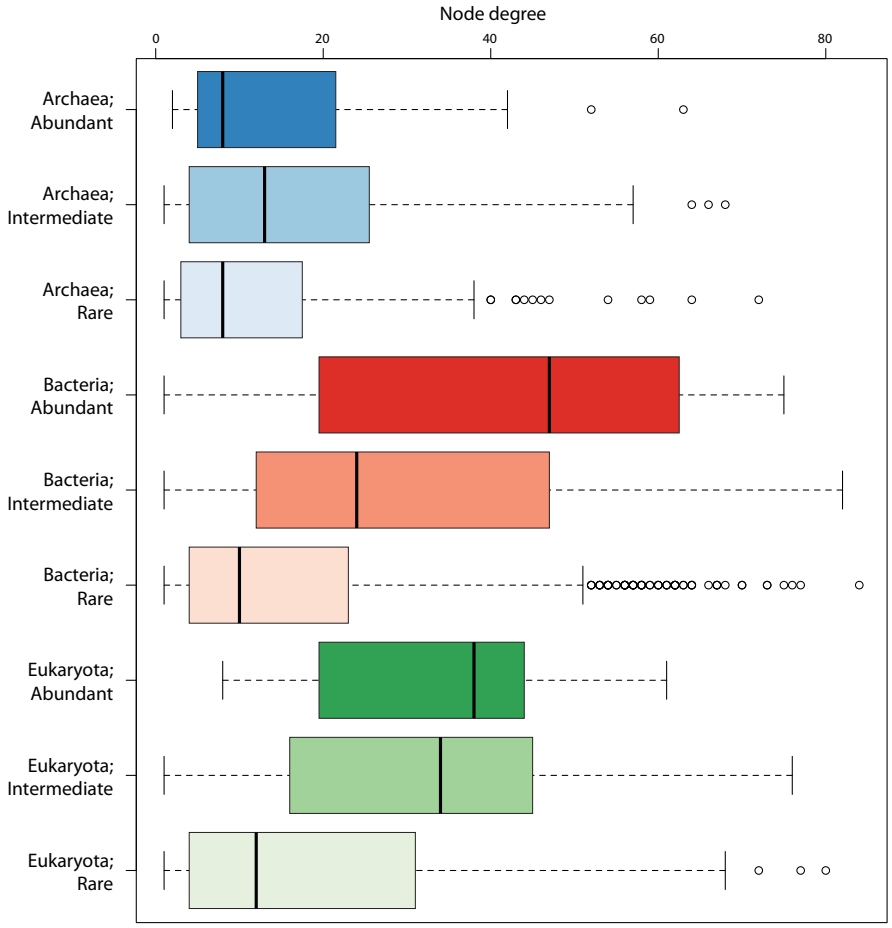


Figure 2.10 Node degree distribution by domain / frequency class combination

Left hand edge of each box denotes first quartile and right hand edge of each box denotes third quartile; data points between the first and third quartile (inside the box) fall within the interquartile range (IQR). The band inside each box denotes the median; whiskers denote data within 1.5 x IQR; hollow circles denote outlying data points.

Clusters within the network representing fine-scale clusters of samples in the NESAP (determined by HCA) were identified. To do this, significant indicator OTUs present in the network were highlighted (389 indicator OTUs out of a total of 949 [41%] were present in the network). Clusters representing the 6 indicator groups of samples employed in Indicator Species Analysis (see Materials and Methods) were evident in the network (Figure 2.11). Surface clusters F10 10 m and A07 10 m grouped together in a dense cluster separated from and negatively correlated with the larger mesopelagic cluster (Figure 2.11, bottom left). Within the large mesopelagic cluster, clusters were temporally segregated with F10 500 m and F10 1000+ m

present near one another in space (Figure 2.11, right) and further away from clusters representing A07 500 m and A07 1000+ m (Figure 2.11, left). Surface and mesopelagic clusters were connected by two negatively correlated links between two mesopelagic Thaumarchaeotal OTUs and one surface OTU affiliated with *Rhodobacteraeae*.

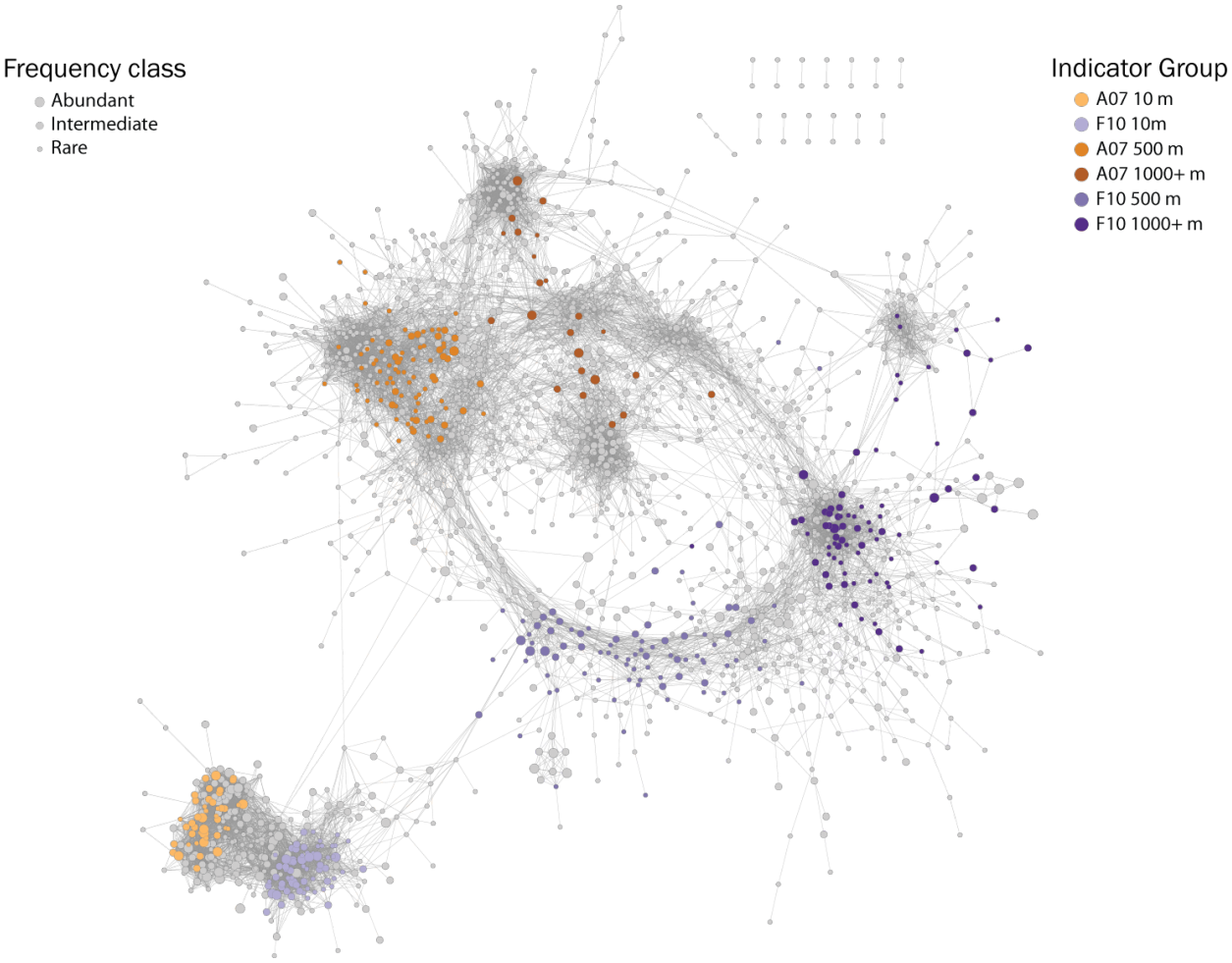


Figure 2.11 Distribution of significant indicator OTUs in microbial co-occurrence network

Significant indicator OTUs (nodes) for each indicator group are shown in colour (see legend). Nodes are sized based on frequency class (abundant [>1%], intermediate [0.1-1%) and rare [<0.1%]).

As MGA was the most abundant bacterial group detected in mesopelagic waters of the NESAP (Figure 2.1), the distribution of MGA-affiliated nodes was mapped onto the network to observe the distribution and connectedness of MGA within the network (Figure 2.12). MGA comprised 459 out of 2005 nodes (23% of total nodes) in the network. All connections to MGA nodes were highlighted (Figure 2.13) as a test case of identifying putative ecological interactions

represented in the network. MGA nodes displayed a similar degree distribution to other bacterial nodes (average degree of 18.3±18.0, range from 1 to 76), and formed a total of 6 995 connections with other nodes in the network (37% of all 18 905 connections). MGA nodes were most frequently connected with other MGA nodes, followed by SAR11, Thaumarchaeotal, and Euryarchaeotal nodes (Figure 2.13). Connections between MGA nodes and several other dominant bacterial groups (including Gammaproteobacterial clusters ZD0405 and ZD0417, and Deltaproteobacterial SAR324) were identified, in addition to connections with other less abundant microbial groups.

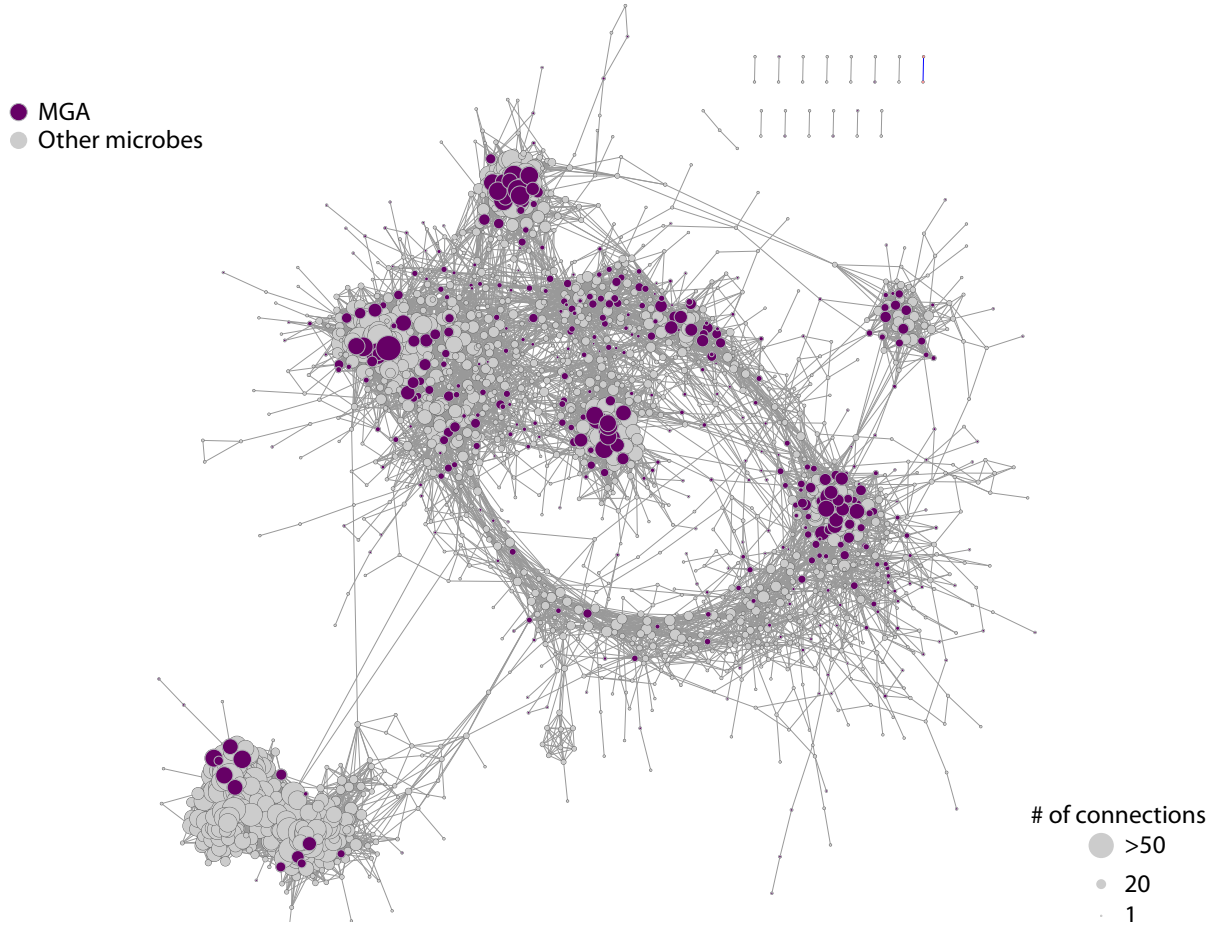


Figure 2.12 Distribution and degree of MGA bacteria as nodes in microbial co-occurrence network

Nodes affiliated with the candidate bacterial phylum Marine Group A (MGA) are highlighted in purple. Nodes are sized based on degree (number of connections to other nodes).

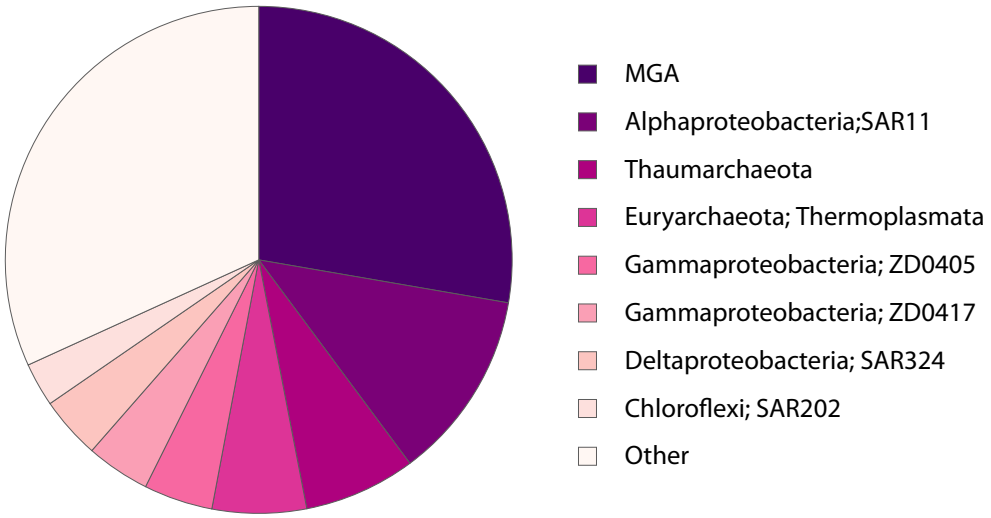


Figure 2.13 Correlations involving MGA nodes in microbial co-occurrence network

2.4 Discussion

Until recently, molecular studies of microbial community structure in the environment were limited to assessing the abundance and distribution of one domain or group of organisms, and within that group, to the most prevalent members. However, microorganisms affiliated with different domains of life do not exist in isolation; they form complex webs of interaction, affecting the abundance and distribution of each other (Faust and Raes, 2012; Raes and Bork, 2008). In addition, abundant species represent only a small portion of microbial diversity, while the "rare biosphere" (the long tail of rare microbes in an abundance distribution) comprises a very high number of rare groups that contain most of the diversity and can play important functional roles in microbial systems (e.g. in the case of nitrogen fixation) (Pedrós-Alió, 2006; Sogin et al., 2006; Galand et al., 2009). New technologies have made it possible to study rare and abundant organisms affiliated with all three domains of life (*Archaea, Bacteria, Eukaryota*) simultaneously. In order to develop a more accurate and holistic understanding of ecosystem structure and function, a quantitative understanding of the diversity and distribution of, and perhaps most importantly, the interactions among, rare through abundant microbial members affiliated with all three domains of life is essential. This chapter surveys the diversity and distribution of rare through abundant microbes affiliated with *Archaea*, *Bacteria*, and *Eukaryota*

in the Northeast subarctic Pacific Ocean (NESAP) and describes trends in co-occurrence patterns among microbial groups that may represent ecological interactions in this environment.

Surface waters of the NESAP in August 2007 and February 2010 were characterized by a dominance of *Bacteria* in most samples. Major bacterial lineages identified in surface waters included the Alphaproteobacterial cluster SAR11, the Cyanobacterial genus *Synechococcus*, and the candidate phylum Marine Group A (MGA). *Archaea* were nearly absent from August 2007 surface samples, while *Archaea* affiliated with MGI Thaumarchaeota and MGII Euryarchaeota (class *Thermoplasmatales*) were present in high proportions in February 2010 coastal surface samples (~17% and 12% of pyrotag sequences per sample, respectively) with declines to ~5% and 0.5% in the open ocean at this time point. Similar seasonal observations in archaeal distribution have been reported in studies of the Southern Ocean, where *Archaea* were virtually absent from summer surface waters but comprised a significant portion of the microbial community in winter (Murray et al., 1998; Grzymski et al., 2012). A study of archaeal community structure in an Arctic shelf ecosystem documented that MGI Thamarchaeota were relatively more abundant in open ocean surface waters while MGII Euryarchaeota dominated the archaeal assemblages in coastal waters (Galand et al., 2008); in contrast, both archaeal groups appeared to be more prevalent in NESAP coastal waters versus open ocean waters. Eukaryotic groups typically comprised ~25% of pyrotag sequences in surface samples, but an increased proportion of eukaryotic groups (54.1%) was detected in August 2007 at the open ocean station P26, predominantly composed of *Phaeocystis*, *Dinophysis* and *Florenciellales*. Small-celled haptophytes (including *Phaeocystis*) and dinoflagellates have been previously documented as abundant members of picoeukaryotic assemblages in NESAP surface waters, particularly in spring and summer (Booth et al., 1993; Wong et al., 2006; Royer et al., 2010). While in many oceanic regions picoeukaryotes are typically 1 to 3 orders of magnitude less abundant than bacteria, the abundance of picoeukaryotes has been documented to equal or even exceed the abundance of bacteria in some high-nutrient marine environments, such as are found in the NESAP (Biegala et al., 2005).

Oxygen minimum zone and oxycline waters of the NESAP were consistently dominated by Thaumarchaeota with significant representation of bacteria affiliated with MGA, Deltaproteobaterial SAR324 and *Nitrospina*, SAR11, and Gammaproteobacterial clusters

ZD0405 and ZD0417. These groups have been observed in other OMZs and seasonally anoxic basins including the Eastern Tropical South Pacific, Saanich Inlet, Namibian Upwelling, (Stevens and Ulloa, 2008; Zaikova et al., 2010; Woebken et al., 2008; Wright et al., 2012). It is interesting to note that anoxic-OMZ (AMZ) affiliated organisms such as the SUP05 clade of Gammaproteobacteria (Sunamura et al., 2004), in addition to certain groups of Epsilonproteobacteria and Planctomycetes were rare or absent in O_2-deficient regions of the NESAP (where O_2 concentrations have not been observed to drop below ~9 μmol kg^{-1}), consistent with what is known about the preferred niches of these apparently O_2-sensitive organisms (Walsh et al., 2009; Lavik et al., 2009; Stevens and Ulloa, 2008; Grote et al., 2008; Lin et al., 2008; Grote et al., 2012; Woebken et al., 2008; Ulloa et al., 2012). Presence of these groups in future studies of NESAP microbial community structure could indicate a shift to more anoxic conditions. Eukaryotic subgroups were present in small proportions in most OMZ and oxycline samples and further work on taxonomic identification of eukaryotic OTUs is required to adequately determine whether the identity of observed subgroups detected in the NESAP is consistent with previous studies of eukaryotic diversity in other O_2-deficient systems, (Orsi et al., 2011; Edgcomb et al., 2011; Orsi et al., 2012). It is important to note that water samples in the current survey were filtered through a 2.7 μm pre-filter before reaching the 0.22 μm filter, from which DNA was extracted for pyrotag sequencing. Thus, we would not expect to obtain sequences affiliated with protists known to be larger than 2.7 μm.

Hierarchical cluster analysis (HCA) of NESAP microbial community profiles suggested that surface communities display large temporal variability (Figure 2.3). This variability could most easily be explained by the large variability in sea surface temperatures observed in the NESAP between February and August. Temperature has been reported in numerous studies of as a key driver of microbial community structure in the environment (Wang et al., 2012; Wang et al., 2013; Norris et al., 2002). Related to seasonal changes in sea surface temperature, previous studies of seasonal variability in nutrient and phytoplankton dynamics in NESAP surface waters have reported that inshore stations (P4 – P16) of the Line P transect are characterized by spring and summer blooms (primary production >3 g C m^{-2} d^{-1}) whereas open ocean stations (P20 and P26) display low seasonality in biomass and primary production (Boyd and Harrison, 1999; Peña and Varela, 2007). However, hierarchical cluster analysis of microbial community structure

showed temporal variability across the entire transect, suggesting that temporal changes in community structure may be correlated with factors other than biomass and primary production. In O_2-deficient mesopelagic regions of the NESAP water column, microbial communities from various depths sampled at the same time were more similar to one another than communities present at the same depth at different times. Related to seasonal changes in surface water biology of the NESAP, fluxes of all biogenic materials exported to the deep ocean (>3800 m) do show distinct seasonality, with a winter minimum in total mass flux of 38 mg m^{-2} d^{-1} in February and a summer maxima of 150 mg m^{-2} d^{-1} in May/June and in August (Wong et al., 1999). Given that heterotrophic microbial growth in the interior ocean is thought to be mainly supported by organic matter exported from the euphotic zone (del Giorgio and Duarte, 2002; Reinthaler et al., 2010), the documented seasonality in export production and efficiency of the biological pump could explain the observed temporal variability in microbial community structure in mesopelagic regions of the NESAP sampled in February and August. It is also worth noting that February 2010 sampling took place during a strong El Niño event; such events have previously been associated with drastic increases in export flux compared to normal winter fluxes (Wong et al., 1999). Extrapolating these prior observations would suggest that export fluxes in February 2010 may have been more similar to average summer fluxes, potentially nullifying the hypothesis that August 2007 and February 2010 mesopelagic microbial communities differ due to variability in the quantity of export flux. Direct evidence coupling variability in particulate organic matter (POM) quality and quantity with related microbial community structure and metabolism would be required to substantiate the hypothesis that variability in the microbial community is correlated with variability in export flux.

While hierarchical clustering of bacterial OTUs mirrored clustering patterns of all microbial OTUs, clustering of archaeal OTUs revealed that archaeal assemblages from the same depth were more similar to one another at both time points than communities sampled at different depths at the same time. This suggests that robust populations of archaea exist in each depth compartment that are persistent at both time points. For clustering of eukaryotic OTUs alone, communities were almost completely unique across samples. This could partly be due to the very small sample size of eukaryotic sequences potentially resulting from filtration biases imposed by pre-filtering, but it could also suggest that mesopelagic eukaryotes are particularly

specialized to local environmental conditions and thus the eukaryotic community across time and space is less uniform (Orsi et al., 2011; Edgcomb et al., 2011; Orsi et al., 2012).

The relatively high degree of similarity among microbial community profiles at the level of abundant OTUs compared to the lower similarity among communities at the level of rare OTUs suggests that the rare biosphere of the NESAP has a more distinct biogeography than the abundant biosphere. These observations suggest that the rare biosphere of the NESAP is not cosmopolitan in distribution, a result also documented in deep-sequencing studies of Arctic Ocean and Mediterranean Sea microbial communities (Galand et al., 2009; Hugoni et al., 2013). This implies that the distribution of rare microbial groups is subject to ecological mechanisms including speciation, extinction, and dispersal as opposed to being governed solely by random dispersal (Galand et al., 2009).

Indicator species analysis identified 949 significant indicator OTUs associated with fine-scale clusters (indicator groups of samples) within the NESAP. These indicators were primarily affiliated with the rare frequency class, with a declining number affiliated with intermediate and abundant frequency classes. This observation could be interpreted as follows: if abundant OTUs are present more uniformly and thus have less distinct biogeography (as suggested by HCA), these OTUs are less likely to be significant indicators as they are less likely to be unique to a specific group of samples. Rare OTUs, with a distinct biogeography and variable presence across samples, are more likely to be indicative of specific groups of samples. The number of significant indicators decreased with depth in both sampling months, an observation reflected in HCA by the fact that surface samples are least similar to one another (increasing the probability of finding significant indicators). The number of significant indicators in August 2007 was also greater than in February 2010. This could be a real biological phenomenon, but could also possibly be explained by the variability in sequencing effort between February 2010 and August 2007. A07 pyrotag libraries contain significantly fewer sequences than F10 libraries (see Table 2.2) due to improvements in sequencing efficacy achieved between the times that these samples were sequenced. Assuming that a relatively larger proportion of rare OTUs are missing in August 2007 samples due to decreased sequencing effort, it is also possible that a proportion of rare indicator OTUs for A07 samples were not detected.

In order to move from a taxonomic inventory of microbial community structure in the NESAP towards an understanding of potential ecological interactions in the microbial system, a microbial co-occurrence or association network was constructed using pairwise Pearson correlation coefficients calculated between abundance profiles of individual OTUs. From a broader perspective, calculating global properties of microbial co-occurrence networks can provide useful insights into the organization of microbial communities, in addition to placing microbial networks in context with other ecological, biological, and non-biological networks (Faust and Raes, 2012; Steele et al., 2011). The degree distribution of the network followed a scale-free or power law distribution ($P(k) = k^{-b}$) with a degree exponent b of 1.154. Scale-free implies the presence of many nodes with only a few links and a smaller number of highly connected nodes (hubs). Many other types of networks have been found to follow a scale-free distribution, including protein interaction networks (Uetz et al., 2000; Ito et al., 2001; Yook et al., 2004), metabolic networks (Jeong et al., 2000; Wagner and Fell, 2001), genetic regulatory networks (Featherstone and Broadie, 2002; Agrawal, 2002), human social networks (Wasserman and Galaskiewicz, 1994), airline networks (Guimera and Amaral, 2005), and the World Wide Web (Albert et al., 1999), in addition to microbial co-occurrence networks generated from soil (Barberán et al., 2012; Zhou et al., 2011) and marine (Steele et al., 2011; Gilbert et al., 2012) environments. The value of b calculated for the microbial co-occurrence network presented here is lower than that often calculated for biological and non-biological networks, where b typically ranges between 2 and 3 (Barabasi and Albert, 1999). A value of $b \leq 2$ indicates a heightened importance of hubs in the network, wherein the main hubs are in contact with a large fraction of all nodes (Barabási and Oltvai, 2004). Thus, the low value of b calculated for the NESAP network is indicative of a degree distribution that is left-skewed towards several very high degree hubs, with a long tail of nodes maintaining few connections. The global clustering coefficient of the network was 0.468, compared to 0.009 for a random network of equivalent size and degree distribution. The fact that the observed clustering coefficient is greater than the estimate of the clustering coefficient for an equivalently proportioned random network indicates that the network is modular, that is, it contains dense modules (or clusters) of highly interconnected nodes. Mapping of significant indicator OTUs onto the network allowed for identification of modules associated with the 6 indicator groups, representing densely interconnected groups of

nodes (clusters) co-occurring at different locations and times in the NESAP water column. High clustering appears to be a generic feature of many biological networks, wherein many clusters appear to represent specific patterns of interconnection associated with distinct functional roles (Hartwell et al., 1999; Barabási and Oltvai, 2004). Whether modules detected in the NESAP network represent functionally unique or functionally redundant communities remains to be uncovered in future studies of spatiotemporal variation in metabolic capacity of microbial communities residing within various regions of the water column at different times.

Focusing on the properties of nodes themselves can provide insight into the relative importance of certain nodes in maintaining network structure (Guimera and Amaral, 2005). With respect to observed properties of nodes in the NESAP network, a breakdown of node degree distribution by domain revealed that eukaryotic nodes were, on average, more connected than bacterial and archaeal nodes, although the observed differences in average node degree across domains were not statistically significant. Highly connected nodes (hubs) were detected within each domain. Although the difference in average degree across domains was not determined to be statistically significant in this study, perhaps due to an inadequate sample size, these results highlight interesting questions regarding the potential role of different domains of microorganisms within a microbial ecosystem: are free-living archaea truly less likely to form ecological interactions than bacteria or protists? Why might this be? Could this pattern reflect differences in metabolic dependence on other organisms to perform key metabolic steps in shared (distributed) pathways? Are eukaryotic microbes truly more likely to interact with other organisms than archaea or bacteria? It is possible that the high number of links associated with eukaryotic nodes could be representative of protist grazing activity or interactions between protists and prokaryotic symbionts, as has been suggested by other recent co-occurrence network studies where a high number of eukaryotic-prokaryotic correlations were detected (Gilbert et al., 2012; Steele et al., 2011; Martinez-Garcia et al., 2011). Future studies involving larger datasets and comparing the degree distribution of nodes in microbial co-occurrence networks from different environments will help to further characterize patterns in degree distribution across and within domains.

A breakdown of node degree distribution by frequency class revealed that abundant nodes were, on average, more connected than intermediate and rare nodes, although these

differences were also not shown to be statistically significant. It is tempting to speculate that nodes that are both abundant and highly connected might play a particularly important role in regulating microbially-mediated processes. Indeed, abundant groups are thought to be well adapted to their environments and to contribute most to biomass production in microbial communities (Cottrell and Kirchman, 2003; Galand et al., 2009). In protein interaction networks, there is a strong relationship between the hub status of a molecule and its role in maintaining the viability and growth of a cell (Jeong et al., 2001). Protein hubs also tend to be conserved over evolutionary time-scales, further corroborating their importance in maintaining cellular integrity (Fraser et al., 2002). One study has claimed that hub nodes in microbial co-occurrence networks might be analogous to "keystone species", whose presence has a disproportionately large effect on its environment relative to its abundance (Steele et al., 2011). With additional genomic information illuminating the metabolic capacity and lifestyle preferences of hub microbes, it will be possible to test the hypothesis that these organisms play a central role in microbial networks, for example by involvement in multiple metabolic conversions that are essential to the maintenance of distributed pathways of community metabolism.

Moving beyond topology, we can characterize patterns in the links between nodes, keeping in mind that co-occurrence networks can be useful for generating hypotheses regarding potential ecological interactions, but cannot independently distinguish between true interactions (i.e. cross-feeding and syntrophy) and other non-random processes (i.e. niche overlap) (Barabási and Oltvai, 2004; Faust and Raes, 2012). While correlations between bacterial nodes were most common in the co-occurrence network (potentially because the dataset was enriched in bacterial OTUs to begin with), correlations between nodes affiliated with all domains were indeed present throughout the network. A microbial co-occurrence network generated from bacterial, archaeal, and eukaryotic OTUs detected at San Pedro Ocean Time Series (SPOTS) also contained numerous correlations between domains (Steele et al., 2011), as did a network generated from bacterial and eukaryotic abundance information collected in the English Channel (Gilbert et al., 2012). As a test case of identifying co-occurrence patterns involving a specific group of organisms, the distribution of links to MGA bacteria, which were highly represented in the network (22% of all nodes), was characterized. MGA nodes were disproportionately well-connected, being involved in 37% of all links in the network. MGA nodes were most frequently

correlated with other MGA nodes, suggesting the potential for intra-phylum interactions within this dominant group of organisms. There is empirical evidence to suggest that adjacent nodes in a variety of complex networks often show significant correlations in their properties, a phenomenon referred to as "assortative mixing" (Newman and Banfield, 2002; Park and Barabási, 2007). While correlated MGA nodes share one obvious property, their phylogenetic affiliation, it will be of interest to further document shared properties among correlated MGA nodes, for example in terms of their metabolic properties.

Regarding correlations within and between nodes of varying frequency classes, correlations between rare OTUs (enriched in the original dataset) were most common, but correlations between nodes affiliated with all frequency classes were detected. As this is the first study to parse a microbial co-occurrence network by frequency class, it is not yet possible to ascertain whether these trends are common in other environments.

2.5 Conclusions

Identifying strong patterns of co-occurrence among pairs or groups of microbes can help generate hypotheses regarding potential ecological interactions occurring in microbial communities. These hypotheses can inform further *in situ* studies (i.e. fluorescence *in situ* hybridization) to confirm physical interactions among microbes, or analyses of metabolic activities (i.e. through meta'omics approaches or activity profiling) that can confirm syntrophy or patterns of distributed metabolism by revealing complementary pathways (Faust and Raes, 2012; Chaffron et al., 2010). Analysis of the distribution and connectivity of specific taxa within co-occurrence networks can also provide a rich source of new hypotheses for identifying distinct microbial populations and targeting key microbial groups for further study (Zhou et al., 2011; Chaffron et al., 2010). In this chapter, application of both traditional ecological analyses and more novel co-occurrence network analyses highlighted the candidate bacterial phylum MGA as a key microbial group in the NESAP demanding further study. MGA was the most abundant bacterial group detected at all depths and time points within mesopelagic waters. Significant indicator OTUs affiliated with MGA were present in all depth- and time-resolved groups of samples identified in the NESAP, suggesting that OTUs affiliated with MGA play an important role in defining community structure throughout the water column in this region. In addition,

nodes affiliated with MGA were highly represented and highly connected in the co-occurrence network, indicating that MGA subgroups co-occur frequently and consistently with other microbial groups and potentially form ecological interactions with these groups. MGA nodes were most frequently connected to other MGA nodes, suggesting the potential for intra-phylum interactions among MGA bacteria in this region. These results highlight the need to further characterize the diversity and ecological roles of MGA bacteria.

Chapter 3: Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean[2]

3.1 Synopsis

Marine Group A (MGA) is a candidate phylum of *Bacteria* that is ubiquitous and abundant in the ocean. Despite being prevalent, the structural and functional properties of MGA populations remain poorly constrained (see Chapter 1, section 1.4.1.3). In this chapter, MGA diversity and population structure was quantified in relation to nutrients and O_2 concentrations in the oxygen minimum zone (OMZ) of the Northeast subarctic Pacific Ocean using a combination of CARD-FISH and 16S rRNA gene sequencing (clone libraries and 454-pyrotags). Estimates of MGA abundance as a proportion of total bacteria were similar across all three methods although estimates based on CARD-FISH were consistently lower in the OMZ (5.6%±1.9%) compared to estimates based on 16S rRNA gene clone libraries (11.0%±3.9%) or pyrotags (9.9%±1.8%). Five previously defined MGA subgroups were recovered in 16S rRNA gene clone libraries and five novel subgroups were defined (HF770D10, P262000D03, P41300E03, P262000N21, and A714018). Rarefaction analysis of pyrotag data indicated that the ultimate richness of MGA was very nearly sampled. Spearman's rank correlation analysis of MGA abundances by CARD-FISH and O_2 concentrations provided statistical support for vertical partitioning of MGA subgroups in the NESAP water column. Analyzed in more detail by 16S rRNA pyrotag sequencing, MGA OTUs affiliated with subgroups Arctic95A-2 and A714018 comprised 0.3 to 2.4% of total bacterial sequences and displayed strong correlations with decreasing O_2 concentration. This study is the first comprehensive description of MGA diversity using complementary techniques. These results provide a phylogenetic framework for interpreting future studies on ecotype selection among MGA subgroups, and suggest a potentially important role for MGA in the ecology and biogeochemistry of OMZs.

---

[2] A version of this chapter has been published: Allers, E.*, **Wright, J.J**.*, Konwar, K.M., Howes, C.G., Beneze, E., Hallam, S.J., & Sullivan, M.B. (2012). Diversity and population structure of marine group A bacteria in the northeast subarctic pacific ocean. *The ISME Journal*, *7*, 256-268. doi:10.1038/ismej.2012.10
*Co-first authors

3.2 Materials and methods

3.2.1 Sample collection and processing

Sampling was conducted via multiple hydrocasts using a rosette water sampler, with an attached Conductivity, Temperature, Depth (CTD) sensor aboard the *CCGS John P. Tully* during Line P cruise 2009-09 in the NESAP in June 2009. Major stations sampled include: P4 [48º39.0N, 126º4.0W] – June 7th, P12 [48º58.2N, 130º40.0W] – June 9th, and P26 [50ºN, 145ºW] – June 14th). At these 3 stations, 20 L samples for DNA isolation were collected from the surface (10 m), while 120 L samples were taken from three depths spanning the OMZ core and upper and deep oxyclines (500 m, 1000 m, 1300 m at station P4 and 500 m, 1000 m, 2000 m at stations P12 and P26). Sample collection and filtration protocols can be viewed as visualized experiments at http://www.jove.com/video/1159/ (Zaikova et al., 2009) and http://www.jove.com/video/1161/ (Walsh et al., 2009), respectively.

For small-volume sampling (for CARD-FISH), the water from Niskin bottles was transferred into pre-rinsed 1-L plastic bottles, filtered through a 10 mm nylon mesh filter, and processed immediately. The CTD-mounted O_2 probe (Model SBE 43, Sea-Bird Electronics, Bellevue, WA) reported O_2 concentrations in μmol kg^{-1}. Seawater samples for nutrient analysis were collected in 16 x 125 mm polystyrene test tubes and analyzed at sea (stored at 4 °C and in the dark for < 12 hrs prior to analysis) using an Astoria Analyzer (Astoria-Pacific, Clackamas, OR) as described by (Barwell-Clarke and Whitney, 1996).

3.2.2 Chlorophyll a

Chlorophyll *a* (Chl*a*) was measured *in situ* with a Seapoint chlorophyll fluorometer (Seapoint Sensors, Exeter, New Hampshire) and calibrated with 109 selected reference samples collected on 47 mm GF/F filters (Whatman International, Maidstone, UK) for Chl*a* extraction (Holm-Hansen et al., 1965). The linear regression between reference sample fluorescence and Chl*a* data was used to transform depth corrected fluorescence units to Chl*a* (Cuttelod and Claustre, 2010) (R^2=0.90).

3.2.3 Enumeration of cells by flow cytometry

Cells were enumerated by flow cytometry using samples fixed with formaldehyde (final concentration of 4% wt/vol) and stored at 4 ºC for 7 to 14 days until analysis at the University of British Columbia (Zaikova et al., 2010). For flow cytometric analysis, a 500 ml sample was incubated with 5 ml of a 10 000-fold dilution of SYBR Green I (nucleic acid stain; Invitrogen, Carlsbad, CA) overnight at 4 °C in the dark. Cells were counted with a FACS LSRII (Becton Dickonson, Franklin Lakes, NJ) equipped with an air-cooled argon laser (488 nm, 15 mW). Stained cells, excited at 488 nm, were identified and enumerated according to their right angle scatter (SSC) and green fluorescence (FL1) emission measured at 530 nm ± 30 nm. The exact volume analysed and subsequent estimation of cell concentrations were calculated by the addition of a know concentration of 6 mm fluorescent beads (Invitrogen).

3.2.4 CARD-FISH

Pre-filtered (10 μm) seawater samples were fixed with Formaldehyde (16%, Polysciences, Warrington, PA) at a final concentration of 1-2% at 4 °C for 12-24 h. Subsamples were filtered onto 47 mm 0.2 μm membrane filters (GTTP, Millipore, Billerica, MA) and rinsed with Milli-Q water. Filters were left to air dry and then stored at -80 °C until analysis by CARD-FISH as described by Pernthaler *et al.* (2004). In brief, cells were fixed to the filter membrane by agarose embedding in 0.1% (w/v) low gelling point agarose prepared with MilliQ water. Endogenous peroxidases were inactivated by HCl treatment of filter membranes using 50 ml of 0.01 M HCl incubated for 10 minutes at room temperature. Cells embedded on filters were then permeabilized in lysozyme solution for 60 minutes at 37 °C. Additional subsampled filters were incubated in a combination of lysozyme and achromopeptidase solution or HCl to test for optimization of permeabilization. For hybridization, horseradish peroxidase (HRP)-labeled probes EUBI-III (Amann et al., 1990; Daims et al., 1999) and NON338 (Wallner et al., 1993) were added to hydridiztion buffers containing 35% formamide (Fisher, Pittsburg, PA), while HRP-labeled probe SAR406-97 (Fuchs et al., 2005) was added to a hybridization buffer containing 40% formamide. Hybridizations were performed on a rotation shaker at 35 °C for 2 to 15 hours and followed by washing steps to remove unspecifically bound probe. For cytochemical probe detection (CARD step), filters were incubated in PBS for 15 minutes at room temperature,

followed by addition of 1000 uL of amplification buffer and 4 uL of Alexa Fluor® 488 dye (Invitrogen Molecular Probes, Carlsbad, CA) and incubation for 15 minutes at 46 °C in the dark. The fraction of FISH-stained bacteria was quantified microscopically at 1000x magnification in at least 1000 DAPI-stained cells in 10 or more fields of vision per sample using an AxioImager (Zeiss, Germany).

3.2.5 Environmental DNA extraction for 16S rRNA gene clone library construction

DNA was extracted from sterivex filters as described in (Zaikova et al., 2010) and (DeLong et al., 2006) – see Chapter 2 section 2.2.3 for a detailed explanation of DNA extraction methods. The DNA extraction protocol can be viewed as a visualized experiment at http://www.jove.com/video/1352/ (Wright et al., 2009).

3.2.6 Phylogenetic & population structure analysis

PCR amplification of 16S rRNA gene, clone library construction and sequencing

A total of 12 DNA extracts from samples collected from four depths at stations P4, P12, and P26 in February 2009 (using the same sampling plan and protocols described above) were amplified using small subunit ribosomal DNA (16S rRNA gene) primers targeting the bacterial domain: B27F (5'-AGAGTTTGATCCTGGCTCAG) and U1492R (5'-GGTTACCTTATGTACGACTT) under the following PCR conditions: 3 min at 94 ºC followed by 35 cycles of 94 ºC for 40s, 55 ºC for 1.5 min, 72 ºC for 2 min and a final extension of 10 min at 72 ºC. Each 50 µL reaction contained 1 µL of DNA, 1 µL each 10 mM forward and reverse primer, 2.5U Taq (Qiagen, Germantown, MD), 5 µL 10 mM deoxynucleotides, and 41.5 µL 1x Qiagen PCR Buffer. 16S rRNA gene amplicons were purified, transformed and cloned as described previously (Zaikova et al., 2010) with the following modifications: one 384-well plate per depth interval was picked and sent for Sanger sequencing at the Michael Smith Genome Sciences Centre (GSC, Vancouver, BC). Sequence data was collected on an AB 3730xls (Applied Biosystems, Carslbad, CA). Plasmids were sequenced bidirectionally with M13F (5'-GTAAAACGACGGCCAG) and M13R (5'-CAGGAAACAGCTATGAC) primers. Bidirectional sequence reads were assembled using Sequencher v4.8 (Gene Codes Corporation, Ann Arbor, MI) and manually edited for base-calling errors. The resulting datasets were checked for chimeras with the open source application

Bellerophon (Huber et al., 2004) (using default settings) and 745 chimeric sequences were removed.

Phylogenetic analysis and tree construction using MGA 16S rRNA gene sequences

A total of 3 164 non-chimeric 16S rRNA gene sequences were imported into the ARB software package (Release 106; (Ludwig et al., 2004)). Sequences were added to the full-length SILVA database (www.arb-silva.de; (Pruesse et al., 2007)), aligned to the closest relative, and added to an existing tree of sequences from the ARB database by using the ARB parsimony tool (using default parameters).

A maximum likelihood phylogenetic tree of MGA 16S rRNA gene sequences exported from ARB was inferred by PHYML (Guindon et al., 2005) using an HKY + 4G + I model of nucleotide evolution where the parameter of the gamma distribution, the proportion of invariable sites, and the transition/transversion ratio were estimated for each dataset. The confidence of each node was determined by assembling a consensus tree of 100 bootstrap replicates. Bacterial 16S rRNA gene sequences (including 170 previously published sequences generated from the Line P transect in June 2008 (Station P4 1000 m) (Walsh et al., 2009) were also placed in taxonomic hierarchy for downstream analysis using the NAST aligner (DeSantis et al., 2006b) and blast using default parameters against the 2008 Greengenes database (DeSantis et al., 2006a), and 290 sequences were identified as belonging to MGA. These 290 sequences were clustered at 97% identity using mothur (v.1.19.0; (Schloss et al., 2009)). Representative sequences from each of these clusters were identified using the get.oturep command in mothur and were included in the phylogenetic tree.

PCR amplification of 16S rRNA gene for pyrotag sequencing

To more directly compare the quantitative distribution of MGA in relation to CARD-FISH counts, the V6-V8 region of 16S rRNA was amplified from June 2009 DNA samples using primers 926F (5'-cct atc ccc tgt gtg cct tgg cag tct cag AAA CTY AAA KGA ATT GRC GG-3') and 1392R (5'-cca tct cat ccc tgc gtg tct ccg act cag-<**XXXXX**>-ACG GGC GGT GTG TRC-3'). Primer sequences were modified by the addition of 454 A or B adapter sequences (lower case). In addition, the reverse primer included a 5 bp barcode designated <**XXXXX**> for

multiplexing of samples during sequencing. Twenty-microlitre PCR reactions were performed in duplicate and pooled to minimize PCR bias using 0.4 μL Advantage GC 2 Polymerase Mix (Advantage-2 GC PCR Kit, Clonetech, Mountainview, CA), 4 μL 5X GC PCR buffer, 2 μL 5M GC Melt Solution, 0.4 μL 10mM dNTP mix (MBI Fermentas, Glen Burnie, MA), 1.0 μL of each 25 nM primer, and 10 ng sample DNA. The thermal cycler protocol was 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 50 °C for 45 s, and 68 °C for 90 s, and a final 10-min extension at 68 °C. PCR amplicons were purified using SPRI Beads and quantified using a Qubit fluorometer (Invitrogen). Samples were diluted to 10 ng/μL and mixed in equal concentrations. Emulsion PCR and sequencing of the PCR amplicons were performed at the Department of Energy Joint Genome Institute (Walnut Creek, CA) following the Roche 454 GS FLX Titanium (454 Life Sciences, Branford, CT) technology according to the manufacturer's instructions.

Processing of pyrotag sequences

A total of 219 610 pyrotag sequences were analysed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (Caporaso et al., 2010). Reads with length shorter than 200 bases, ambiguous bases, and homopolymer runs were removed prior to chimera detection. Chimeras were detected using the chimera slayer provided in the QIIME software package and removed prior to taxonomic analysis. A total of 212 611 non-chimeric sequences were phylogenetically identified in QIIME using a BLAST-based assignment method and clustered at 97% identity against the Greengenes taxonomic database (DeSantis et al., 2006a). Singleton OTUs (OTUs represented by one read) were omitted from downstream analyses, as recommended by Kunin and colleagues (Kunin et al., 2010), Tedersoo and colleagues (Tedersoo et al., 2010) and Gihring and colleagues (Gihring et al., 2012), leaving 183 212 sequences for downstream analysis.

Clustering of pyrotags to 16S rRNA gene clone library sequences clusters

To resolve patterns of distribution among MGA clusters as a function of geographic location in the Northeast subarctic Pacific (NESAP), pyrotag sequences were recruited to MGA 16S rRNA gene clone library sequence clusters using a 97% identity cutoff in mothur. Blastn was used to query 183 212 pyrotags against a database containing 290 16S rRNA gene clone library

sequences assigned to MGA based on Greengenes taxonomy. Only hits with a perfect match across the full-length of a query sequence were retrieved, and the number of pyrotags mapping to all sequences in each cluster was summed. If a pyrotag mapped to >1 cluster, its relative contribution to each cluster was calculated by dividing by the number of clusters it mapped to and assigning the relevant fraction to each cluster. The number of pyrotags mapping to each cluster was normalized to the total number of bacterial tags in each sample and visualized as a bubble plot using bubble.pl, available for download at http://hallam.microbiology.ubc.ca/downloads/index.html. A rarefaction curve for full-length MGA 16S rRNA sequences and MGA pyrotag sequences was calculated and plotted using QIIME (Caporaso et al., 2010).

3.2.7 Estimating probe SAR406-97 detection efficiency

As no cultured standard is available for MGA cells, binding efficiency of probe SAR406-97 was estimated using sequence data. In order to test the predicted maximum binding efficiency of probe SAR406-97 (Fuchs et al., 2005) (5'-CACCCGTTCGCCAGTTTA) against MGA 16S rRNA gene clone library sequences from the NESAP, blastn (E-value=1000, word_size=7) was used to query the probe sequence against the 290 16S rRNA gene clone library sequences assigned to MGA based on Greengenes taxonomy and collect all local alignments with similarity to the probe sequence. Probe efficiency was described using the percentage of MGA sequences that contained local alignments to the probe across a range of E-value scores for each cluster.

3.3 Results

3.3.1 Physiocochemical characteristics of the study site

Relevant physicochemical data from representative coastal (P4), transition (P12), and open-ocean (P26) stations measured along the Line P transect (Figure 3.1) and related to the present study are described below. Salinity gradients ranging from 32.2-32.6 PSU at the surface (10 m) and 34.1-34.6 PSU in the ocean's interior generated a stratified water column across the Line P transect (Figure 3.2). Chlorophyll *a* (Chl*a*) was present in the top ~100 m, with deep chlorophyll maxima (DCM) ranging from 0.5 μg L^{-1} at 41 m depth at P26 to 1.1 μg L^{-1} at 25 m depth at P4 (Figure 3.3). Average O_2 concentrations were 302 μmol kg^{-1} at the surface, reaching a minimum

of 8.6–15 µmol kg^{-1} between 1000 and 1100 m across the transect (Table 3.1, Figure 3.4). The OMZ core (defined as O_2 < 20 µM [~19.5 µmol kg^{-1}]; (Helly and Levin, 2004; Paulmier and Ruiz-Pino, 2009) was 766±73 m thick and centered at 1026±63 m. Nutrient concentrations were higher in the OMZ core and the upper (500 m) and deep (2000 m) oxyclines than at the surface (Table 3.1, Figure 3.2). In 10 m samples, nitrate and phosphate concentrations were highest at P26 (9.9 µmol L^{-1} and 1.0 µmol L^{-1}, respectively). At 1000 m, nitrate concentration was highest at P26 (47.5 µmol L^{-1}), while phosphate concentration was highest at P4 (3.3 µmol L^{-1}). All contextual data is available through the Canadian Department of Fisheries and Oceans (url: http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/line-p/).

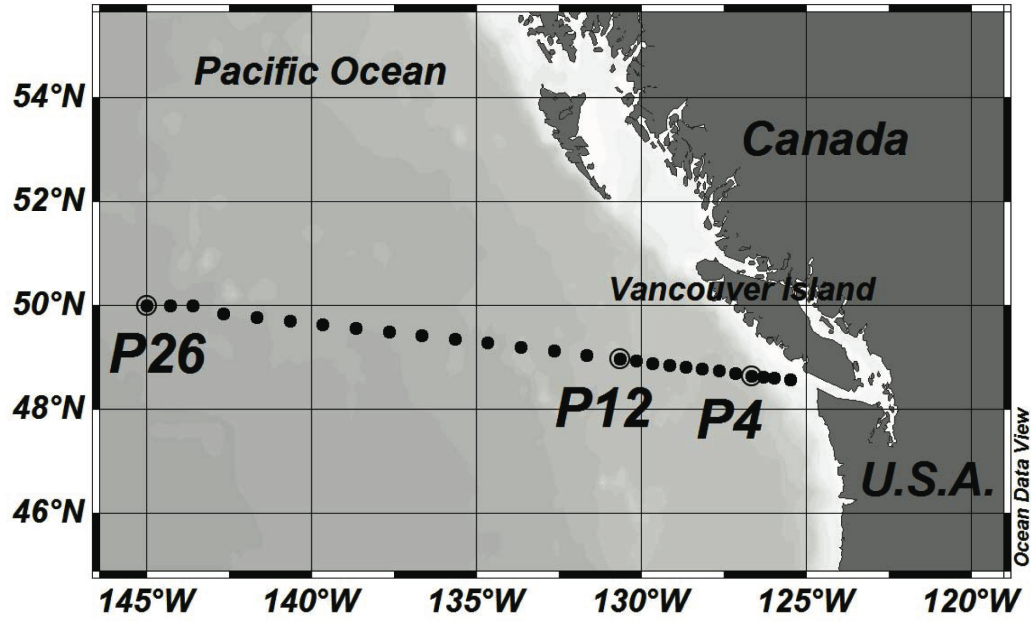


Figure 3.1 Stations P4, P12, and P26 along the Line P oceanographic transect are highlighted

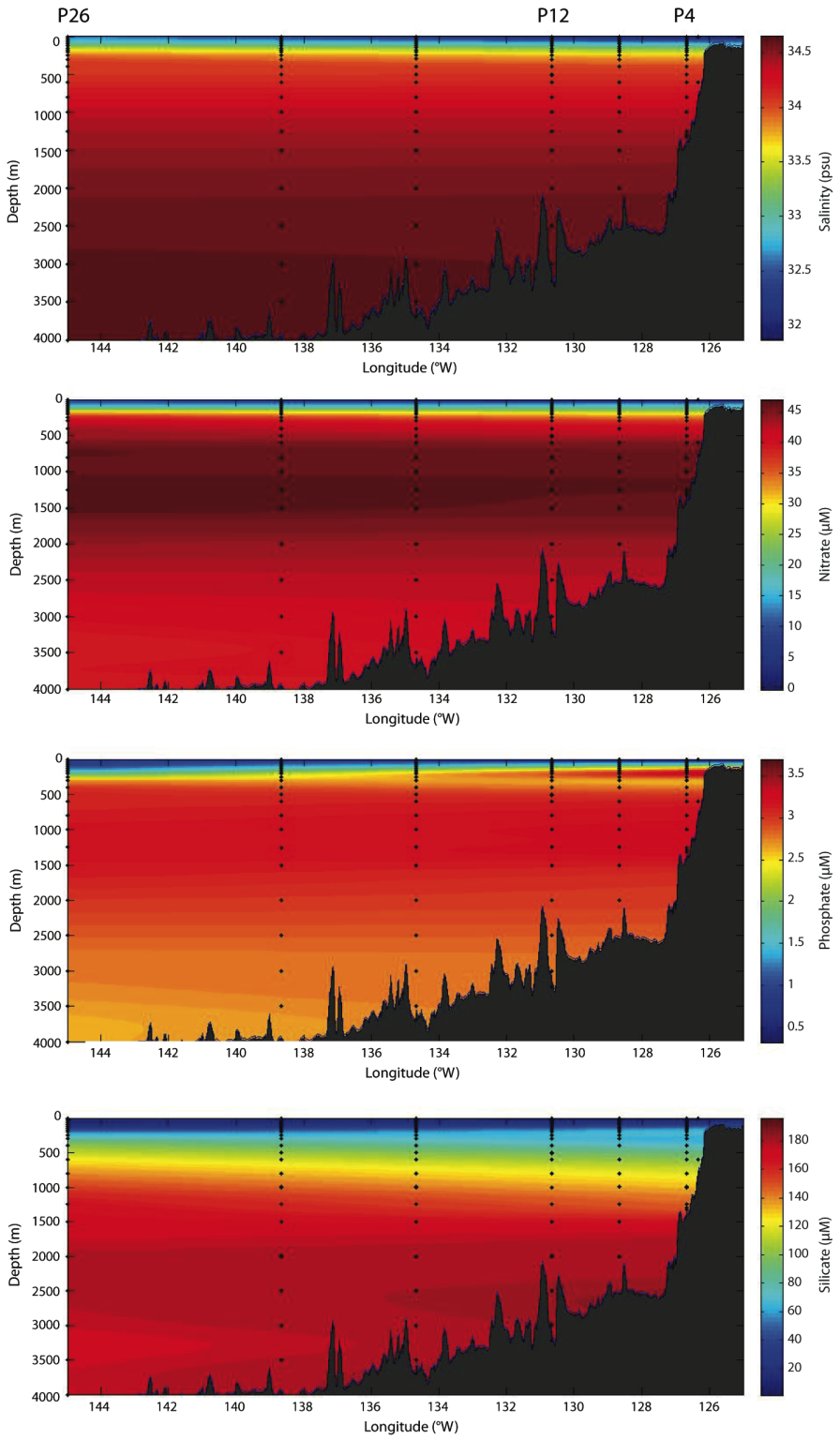


Figure 3.2 Salinity and nutrients along Line P in June 2009

(a) Salinity, (b) Nitrate, (c) Phosphate, (d) Silicate

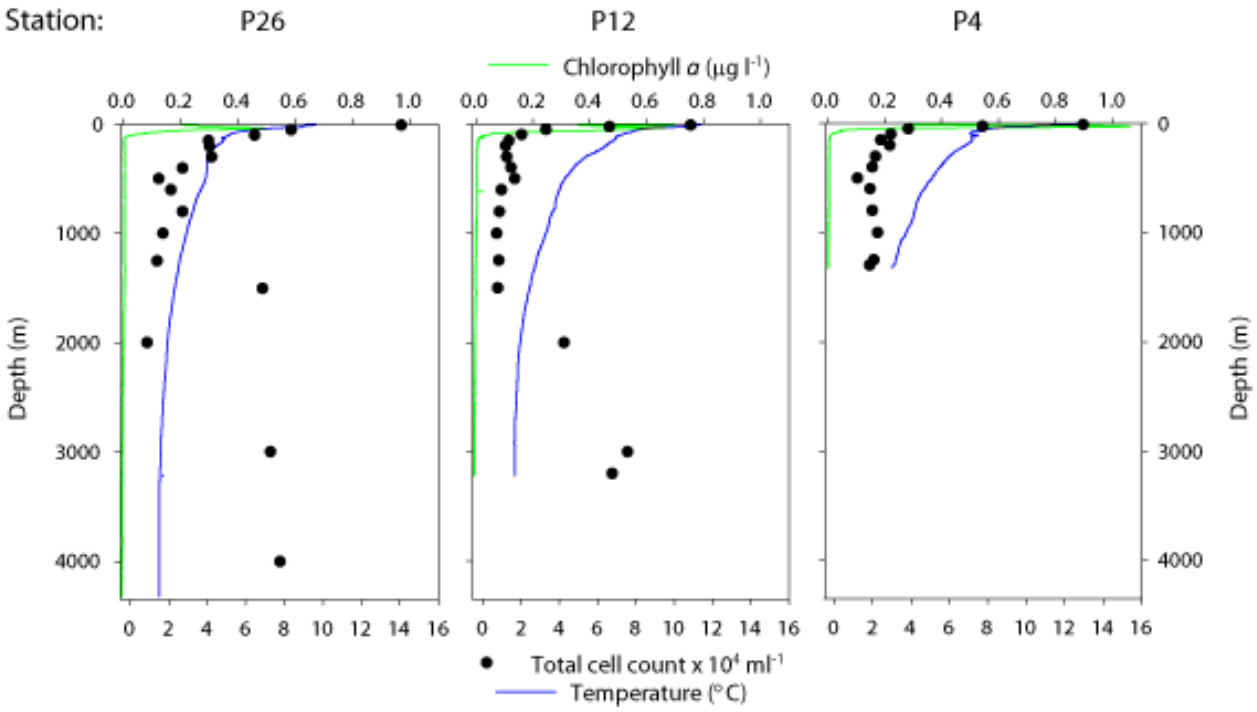


Figure 3.3 Contextual data for Line P stations P4, P12, and P26 in June 2009

Depicted are Chla, temperature, and total cell counts detected by flow cytometry.

Station	Depth [m]	Microbial cell abundance by FCM [cells ml-1]	MGA[a] [% total EUBI-III tagged cells]	MGA[a] [% total DAPI cell number]	No. of bacterial 16S rRNA clones[b]	MGA clones [% 16S rRNA clone library]	No. of bacterial pyrotags	MGA pyrotags[c] [% bacterial pyrotags in library]	Oxygen [µmol/kg]	NO_x[d] [µmol L^{-1}]
P4	10	1.25E+05	2.3	1.3	281	0.0	11 178	0.1	308.0	0.0
	500	1.23E+04	17.6	7.8	287	8.7	14 619	7.8	23.7	42.6
	1000	2.22E+04	19.0	6.7	276	9.8	6 251	11.6	8.6	45.3
	1300	1.86E+04	7.0	3.5	239	10.5	15 284	11.5	15.9	45.8
P12	10	1.21E+05	1.4	0.7	248	1.6	14 189	0.1	296.2	6.3
	500	1.66E+04	12.2	3.8	249	4.0	11 759	10.3	37.0	42.8
	1000	7.89E+03	19.4	6.7	184	13.6	14 839	8.1	9.0	46.3
	2000	7.36E+03	18.2	5.5	256	13.7	7 391	9.3	59.3	44.1
P26	10	1.41E+05	0.5	0.4	242	0.4	11 723	0.4	301.4	10.8
	500	1.47E+04	13.1	3.3	287	7.7	7 648	8.3	35.0	43.6
	1000	1.72E+04	21.4	8.2	293	16.4	7 901	9.6	14.3	45.6
	2000	8.78E+03	12.9	4.5	322	14.3	16 090	13.0	56.5	44.4

a MGA as detected by probe SAR406-97
b 16S rRNA gene clone libaries were generated from February 2009 samples
c MGA pyrotags taxonomically identified by comparison to Greengenes database
d Nitrate + Nitrite

Table 3.1 Chemical and biological parameters at Line P stations P4, P12, and P26 in June 2009

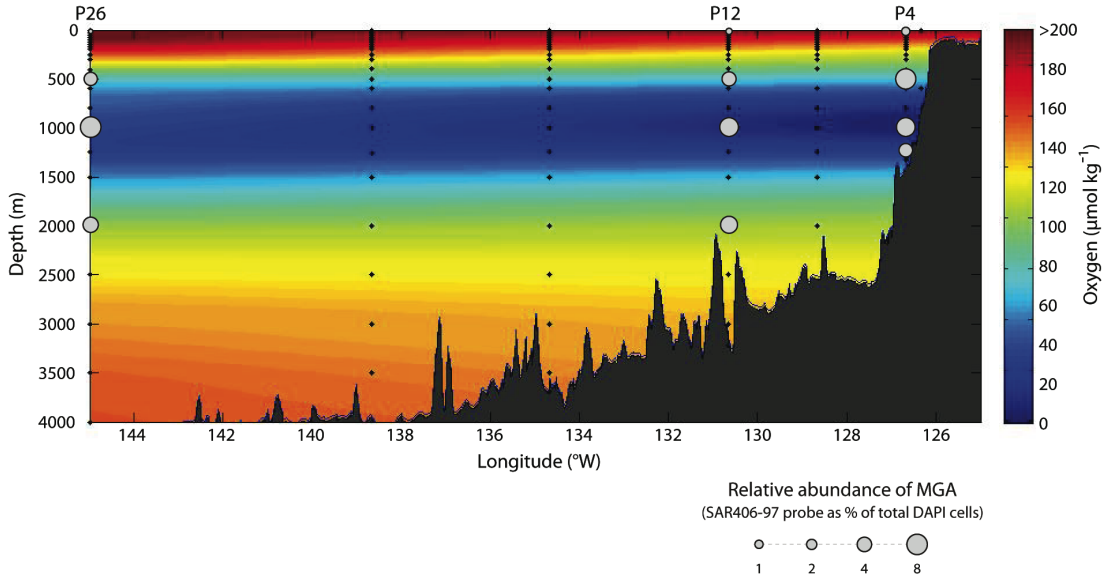


Figure 3.4 Relative abundance of MGA by CARD-FISH in the NESAP at Line P stations P4, P12, and P26 in June 2009

O_2 concentration is depicted as coloured background and MGA abundance is overlaid as gray bubbles.

3.3.2 Microbial cell numbers

Total prokaryotic cell abundance along the Line P transect was (1.3±0.1) x 10^5 ml^{-1} in surface waters and (1.39±0.2) x 10^4 ml^{-1} below 200 m as measured by flow cytometry (Table 3.1, Figure 3.3). Total prokaryotic cell abundance measured by flow cytometry was lower than DAPI counts throughout the water column (Table 3.1, Table 3.2). This discrepancy in cell abundance measured between methods could be due to the relatively long period of time (~7 – 14 days) that fixed samples were kept at 4 ºC before quantification by flow cytometry (Kamiya et al., 2007). The overall detection of *Bacteria* by probes EUBI-III ranged from 25.5%±7.6% to 79.5%±8.6% of total DAPI cell counts with higher detection rates in surface samples (Table 3.1). Low EUB detection did not appear to result from poor cell lysis, as comparison of lysozyme vs. lysozyme/achromopeptidase treatment (Pernthaler et al., 2004) revealed no significant differences (data not shown). Sequence comparison by BLAST analysis suggested that >90% of our full-length bacterial 16S rRNA gene clone library sequences were targeted by EUBI-III

probes with an E-value of 10^{-4} (corresponding to a blastn result with no mismatches and up to one missing 3' base).

	P4				P12				P26			
Depth (m)	DAPI (cells mL^{-1})	EUBI-III Detection Rate (%)			DAPI (cells mL^{-1})	EUBI-III Detection Rate (%)			DAPI (cells mL^{-1})	EUBI-III Detection Rate (%)		
10	1.49E+06	57.03	±	7.57	2.10E+06	49.2	±	6.8	1.86E+06	79.5	±	8.63
25	1.45E+06	42.1	±	4.05	-	-		-	1.55E+06	68.17	±	5.46
50	8.78E+05	41.29	±	5.9	-	-		-	1.12E+06	57.24	±	5.89
100	5.31E+05	42.6	±	7.89	-	-		-	3.45E+05	46.96	±	7.57
150	4.15E+05	39.7	±	5.02	-	-		-	2.29E+05	41.97	±	7.63
200	2.97E+05	39.32	±	6.24	-	-		-	3.38E+05	44.94	±	5.52
300	3.03E+05	42.88	±	3.03	-	-		-	2.93E+05	39.06	±	5.86
400	3.65E+05	43.6	±	5.16	-	-		-	1.47E+05	40.13	±	6.23
500	1.70E+05	44.32	±	6.75	1.70E+05	31.27	±	5.4	1.57E+05	25.48	±	7.6
600	1.58E+05	37.09	±	2.75	-	-		-	1.21E+05	34.66	±	7.29
800	1.50E+05	30.21	±	4.12	-	-		-	1.10E+05	37.59	±	6.45
1000	1.12E+05	35.22	±	5.99	8.71E+04	34.45	±	6.09	6.78E+04	38.33	±	6.33
1250	1.27E+05	50.7	±	5.37	-	-		-	6.94E+04	35.74	±	8.4
1500					-	-		-	6.64E+04	39.19	±	6.42
2000					7.18E+04	30.42	±	3.97	4.54E+04	34.64	±	9.42
3000					4.98E+04	38.2	±	6.02	2.84E+04	28.37	±	9.3
4000									3.10E+04	41.08	±	9.38
Average	4.96E+05	42.00			4.95E+05	36.71			3.87E+05	43.12		

Table 3.2 Detection efficiencies for probe set EUBI-III at Line P stations P4, P12, and P26 in June 2009

3.3.3 Diversity and population structure of MGA

Relative abundance of MGA cells as detected by probe SAR406-97 was similar at stations P4, P12 and P26, with minima in surface waters and maxima in waters ≥500 m (≤1.3% vs. ~8%, respectively; Figure 3.4). At stations P12 and P26, MGA abundance peaked in the core of the OMZ (6.7%±1.8% and 8.2%±1.6%, respectively) with lower values (3.3%-5.5%) in the upper and deep oxyclines (Table 3.1, Figure 3.4). At station P4, MGA abundance peaked in the upper oxycline (7.8%±2.3%) and decreased throughout the OMZ core and deep oxycline. Blastn-based sequence comparisons against our full-length 16S rRNA gene clone library sequences suggested that probe SAR406-97 targeted ~76% of all MGA sequences (see below) with an E-value of 10^{-4} (corresponding to a blastn result with no mismatches and up to one missing 3' base) (Appendix A).

A total of 290 MGA 16S rRNA gene sequences were recovered from 3 164 bacterial sequences traversing the water column at stations P4, P12 and P26. MGA sequences comprised an average of 0.7%±0.84% of 10 m clone libraries and 11.2%±3.9% of libraries from O_2-deficient waters (<90 μmol kg^{-1} O_2) with a maximum of 16.4% at P26 1000 m (Table 3.1). MGA

16S rRNA gene sequences clustered at 97% identity into 121 distinct operational taxonomic units (OTUs), 97 of which contained only singletons (Appendix A). Representative sequences obtained for each OTU were placed in phylogenetic context with relevant reference sequences (Figure 3.5). Five previously defined subgroups were recovered (ZA3648c and ZA3312c (Fuchs, unpublished), Arctic96B-7 and Arctic95A-2 (Bano and Hollibaugh, 2002), and SAR406 (Gordon and Giovannoni, 1996) and five additional subgroups were defined (HF770D10, P262000D03, P41300E03, P262000N21, and A714018). Branch length estimates separating these subgroups in the phylogenetic tree ranged between 3% and 25%. The most abundant OTUs present along the Line P transect comprised between 1 and 4% of at least one clone library and belonged to subgroups Arctic95A-2, HF770D10, SAR406, Arctic96B-7, and ZA3312c (Figure 3.6, Appendix A).

Figure 3.5 Unrooted phylogenetic tree based on 16S rRNA gene clone sequences showing the phylogenetic affiliation of MGA sequences identified in this study

The tree was inferred using maxiumum likelihood implemented in PhyML (Guindon et al., 2005). Reference sequences from other environments are marked with an asterisk. The bar represents 10% estimated sequence divergence.

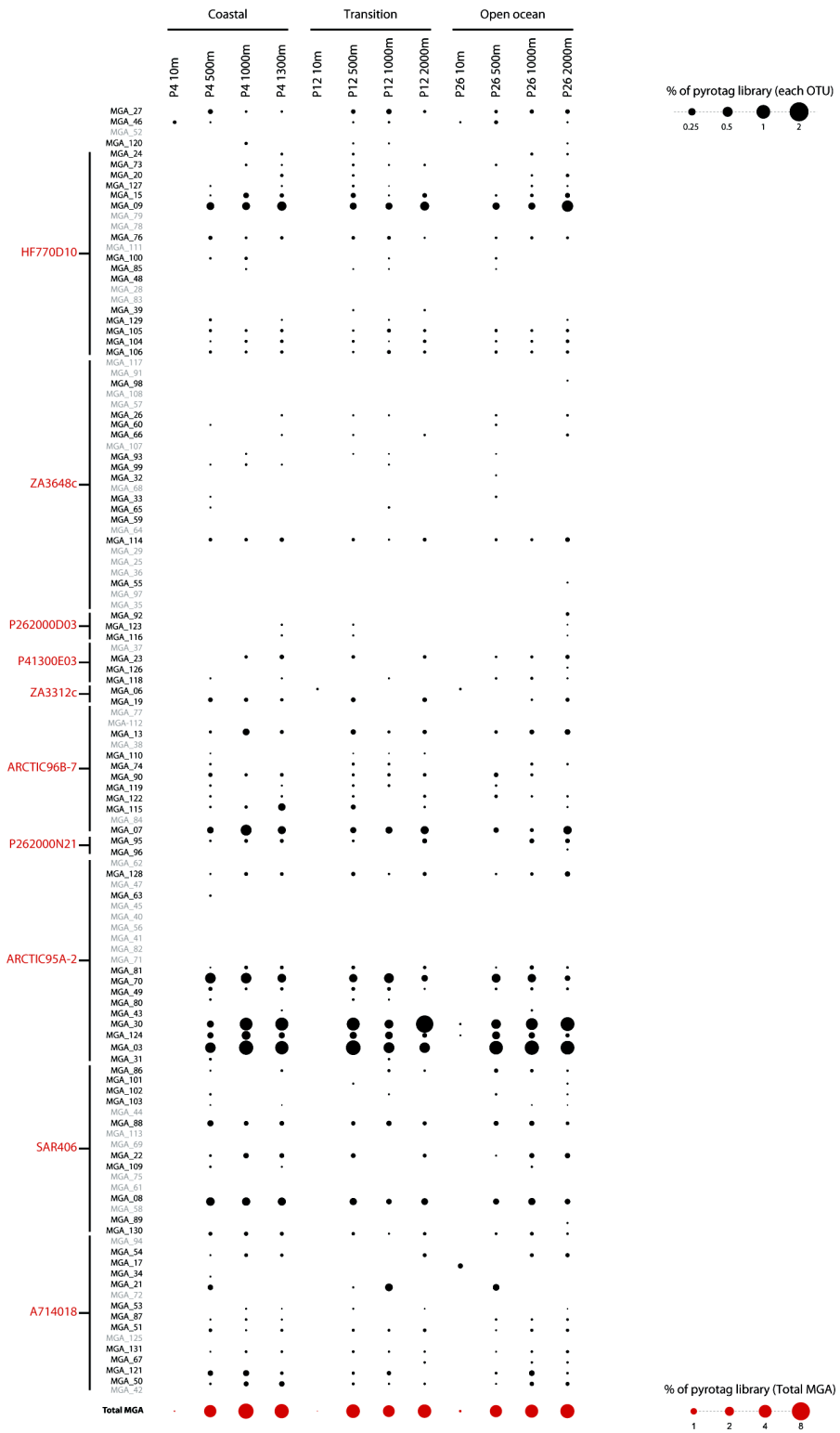


Figure 3.6 Relative abundance of MGA pyrotags affiliated with full-length MGA 16S rRNA gene clone OTUs recovered from the NESAP

Black circles represent proportion of bacterial pyrotags affiliated with each 16S rRNA OTU in each sample.

To explore the diversity and population structure of MGA subgroups with increased resolution, 454-pyrotag sequencing was performed. Pyrotags affiliated with MGA OTUs were identified using two approaches: (1) recruitment of pyrotags to full-length 16S rRNA gene sequences and (2) direct taxonomic assignment of pyrotags in blast-based queries to identify OTUs not detected in clone libraries.

In the first approach, all pyrotags were recruited to all 16S rRNA gene clone library sequences affiliated with MGA (see Materials and methods). A total of 4 403 pyrotags formed identical matches to 78 out of 121 previously defined MGA OTUs (Figure 3.6). The relative proportion of bacterial pyrotags affiliated with MGA OTUs ranged from ~0.01% in 10 m samples to a maximum of 5.7% at P4 1000 m. Within O_2-deficient waters, the average proportion of bacterial pyrotags belonging to MGA was 4.4%±0.73%. The most abundant MGA OTUs based on pyrotag recruitment were affiliated with Arctic95A-2 (~2.4%), Arctic96B-7 (0.55%), SAR406 (~0.45%), HF770D10 (0.55%) and A714018 (0.26%).

In the second approach, all non-singleton pyrotags were queried against the Greengenes database (DeSantis et al., 2006a) resulting in the identification of 10 278 sequences affiliated with MGA (Figure 3.7a). The relative proportion of bacterial pyrotags affiliated with MGA ranged from ~0.1% in 10 m samples to a maximum of 11.6% at P4 1000 m (Table 3.1). Within O_2-deficient waters, the average proportion of bacterial pyrotags belonging to MGA was 9.9%±1.8%. To identify MGA OTUs unique to pyrotags, the corresponding V6-V8 regions from the 290 16S rRNA gene clone library sequences identified as MGA were extracted and clustered with the subset of pyrotag sequences affiliated with MGA at 97% identity into 566 distinct OTUs, 491 of which were unique to pyrotags (Figure 3.7b). However, the majority of abundant OTUs (containing >200 sequences) were common between 16S rRNA gene clone libraries and pyrotag datasets (Figure 3.7c). Of the unique pyrotag OTUs, 249 were non-singleton and contained 4 253 pyrotags (40% of MGA pyrotags), with the most abundant OTU containing 1409 sequences (13.3% of MGA pyrotags) (Figure 3.7c). The slope of the rarefaction curve for MGA pyrotags became nearly asymptotic, indicating that the ultimate richness of MGA OTUs was very nearly sampled (Figure 3.8). In contrast, the rarefaction curve for MGA 16S rRNA gene clone library sequences indicated incomplete sampling.

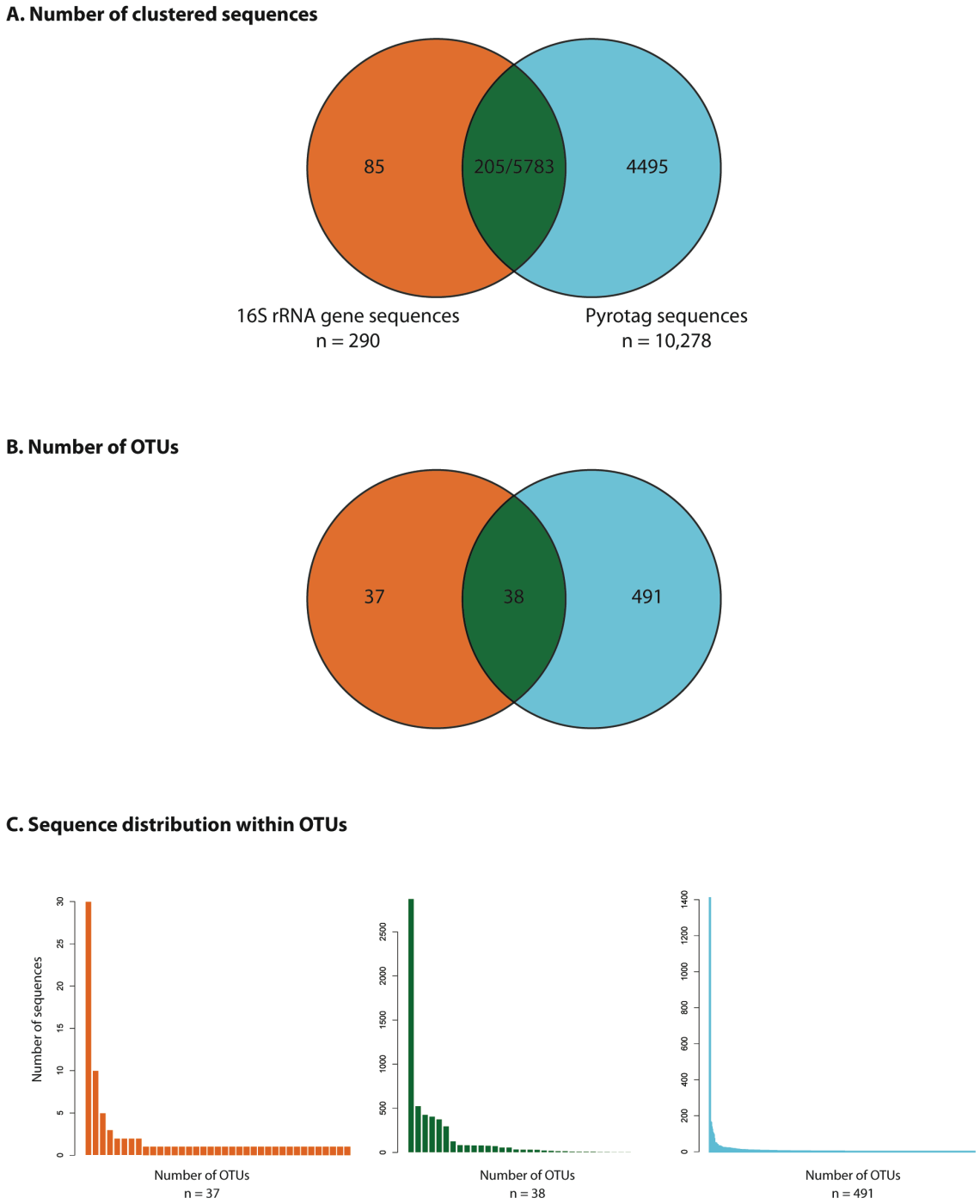


Figure 3.7 Comparison of V6-V8 region of full-length 16S rRNA gene clone sequences and pyrotags affiliated with MGA

Sequences taxonomically identified by comparison with Greengenes (DeSantis et al., 2006a). (a) Number of MGA sequences shared between and unique to 16S rRNA gene clone sequences and pyrotags. (b) Number of MGA OTUs shared between and unique to 16S rRNA gene clone libraries and pyrotags. (c) Sequence distribution within shared and unique MGA OTUs.

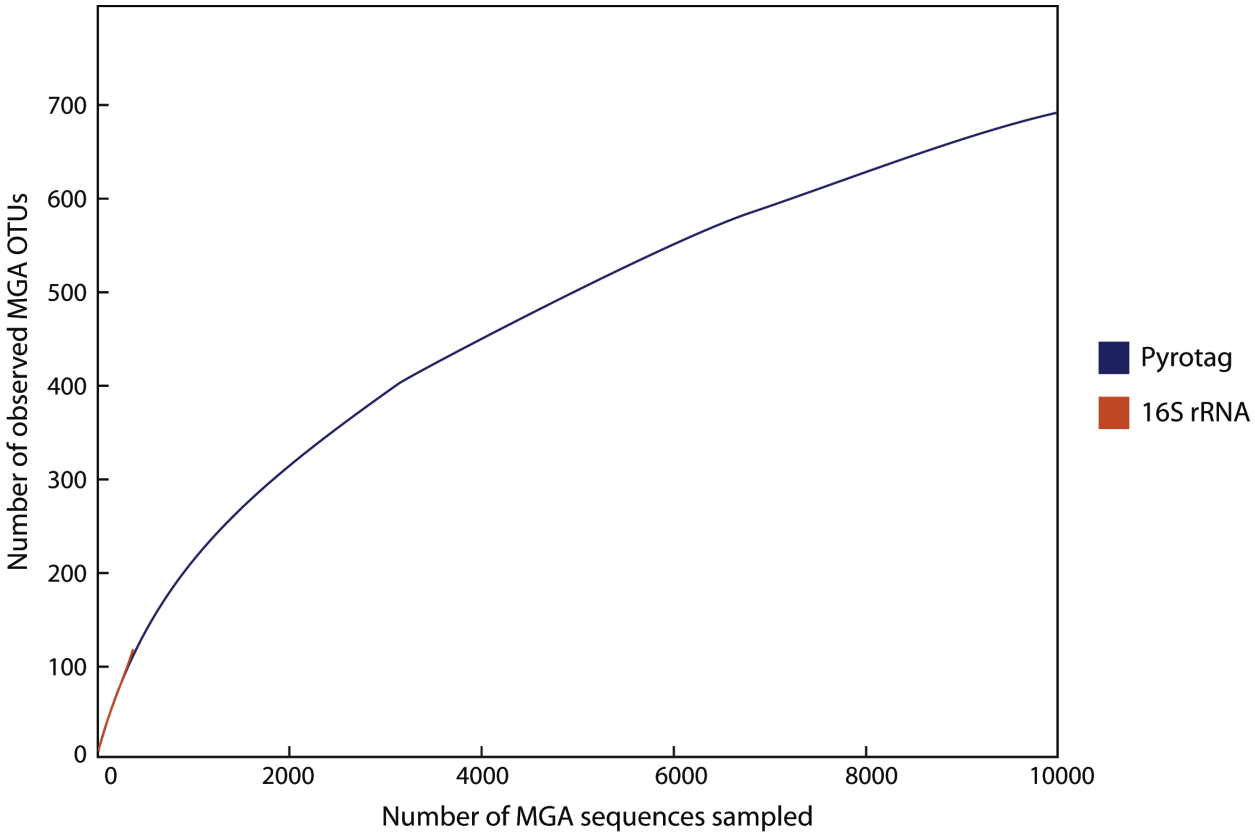


Figure 3.8 Rarefaction curves for MGA sequences in 16S rRNA gene clone libraries and pyrotags in the NESAP

3.3.4 Comparing MGA abundance across methods

To evaluate consistency in estimating MGA abundance using CARD-FISH, 16S rRNA gene clone libraries, and pyrotags, Spearman's rank correlation coefficients (ρ) were determined (3.9). CARD-FISH abundance estimates were significantly correlated ($p<0.05$) with 16S rRNA gene clone library sequence abundance (ρ=0.755) but not with pyrotag sequence abundance (ρ=0.469) (Figures 3.9a and 3.9b). 16S rRNA gene clone library and pyrotag sequence abundance were also significantly correlated (ρ=0.580, Figure 3.9c).

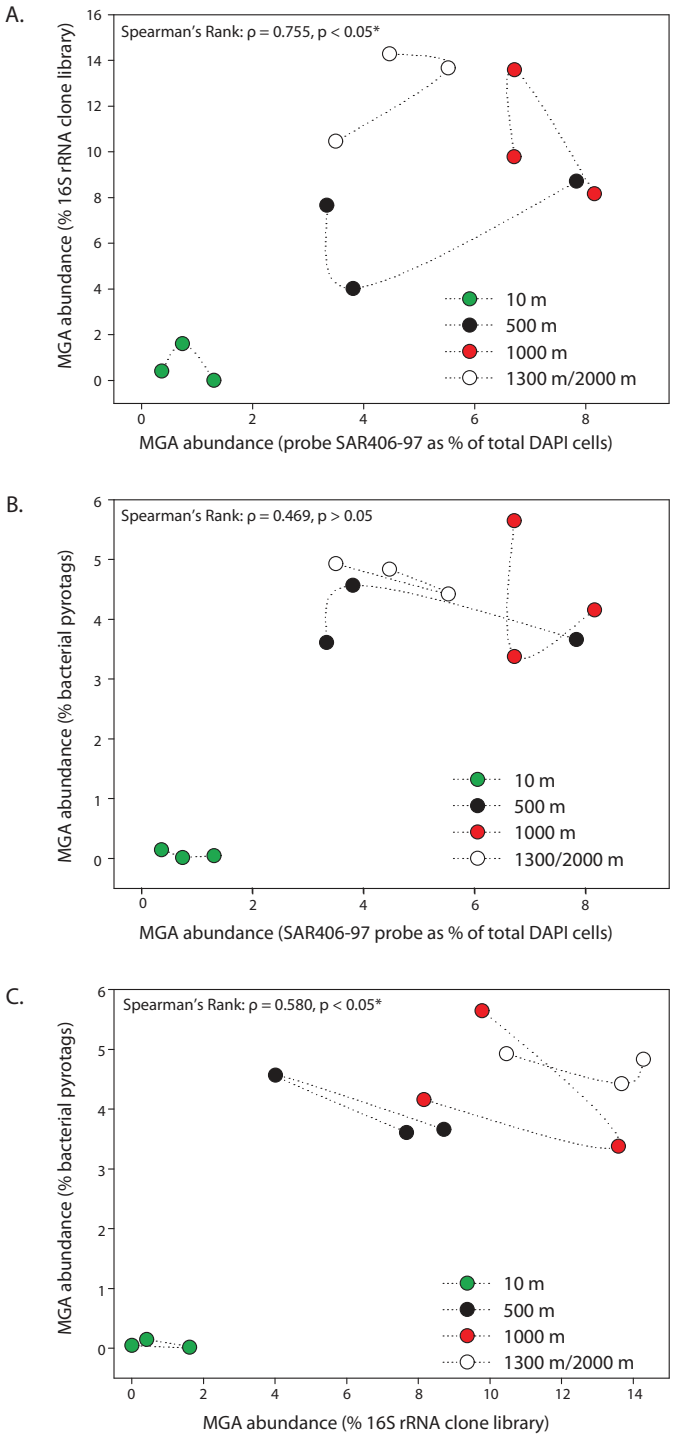


Figure 3.9 Spearman's rank correlation coefficients for estimates of relative MGA abundance

(a) 16S rRNA gene clone libraries vs. probe SAR406-97. (b) Pyrotags vs. probe SAR406-97. (c) Pyrotags vs. 16S rRNA gene clone libraries**.**

To explore potential drivers of MGA habitat selection, Spearman's rank correlation coefficients between CARD-FISH, 16S rRNA gene clone library, and pyrotag sequence abundance and environmental parameters were calculated. When calculated across the entire transect, the abundance of MGA as estimated by CARD-FISH was significantly correlated with decreasing temperature, O_2, and Chl*a*, and increasing nitrate, phosphate, and silicate (Table 3.3). However, when correlations were calculated for each station independently, statistically significant correlations were only identified at station P26 where MGA abundance was more strongly correlated with decreasing O_2 and increasing nitrate and phosphate concentrations than with temperature, Chl*a*, or silicate (Table 3.3, Figure 3.10, Figure 3.11).

Station	n	Depth	Temperature	Oxygen	Salinity	Chl*a*	Nitrate	Phosphate	Silicate
P4	13	0.396	-0.385	-0.396	0.396	-0.358	0.396	0.429	0.396
P12	5	0.9	-0.9	-0.3	0.9	-0.9	0.4	0.359	0.9
P26	17	0.701*	-0.701*	-0.824**	0.699*	-0.514*	0.865**	0.853**	0.706*
all	35	0.620**	-0.577**	-0.589**	0.621**	-0.553**	0.639**	0.578**	0.623**

* $p < 0.050$; ** $p < 0.001$; n = number of samples

[a]Using probe SAR406-97

Table 3.3 Spearman's rank correlation coefficients between relative abundance of MGA estimated by CARD-FISH and environmental parameters

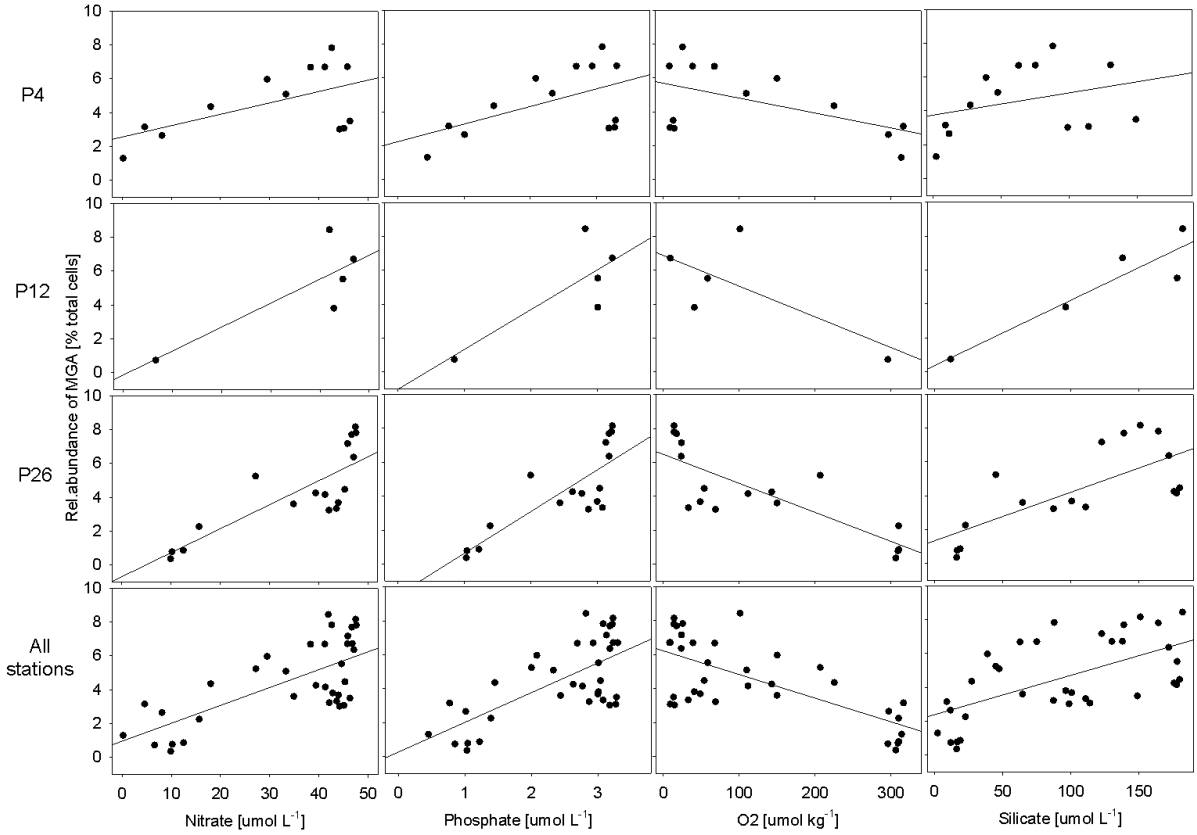


Figure 3.10 Linear regression plots for relative abundance of MGA estimated by CARD-FISH with probe SAR406-97 and depth, salinity, temperature

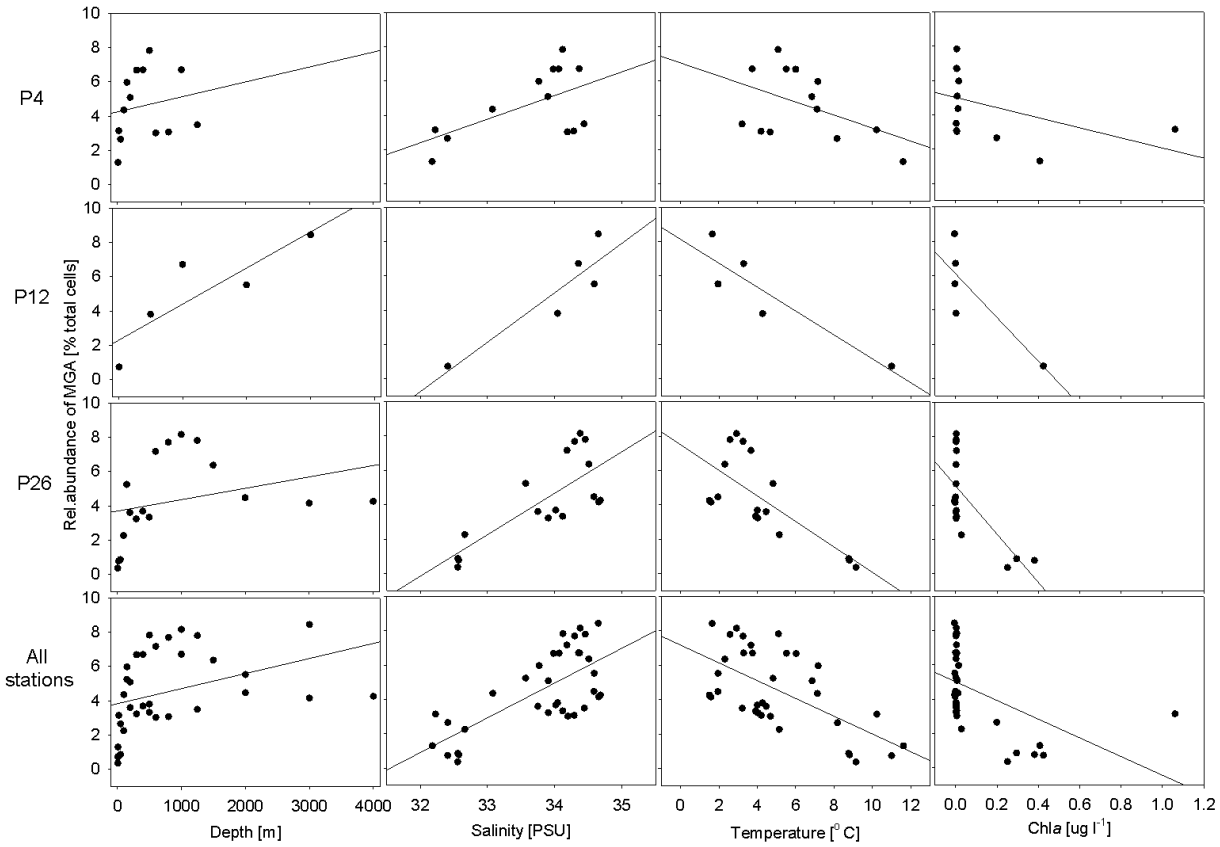


Figure 3.11 Linear regression plots for relative abundance of MGA estimated by CARD-FISH with probe SAR406-97 and nitrate, phosphate, O_2, Chla

When calculated across the entire transect and each station independently, the relative abundance of MGA OTUs based on 16S rRNA gene clone library sequences was not significantly correlated with environmental parameters (data not shown). However, the relative abundance of 4 OTUs identified in pyrotags showed significant correlations across the entire transect with decreasing O_2 after a Bonferroni correction was applied ($p<0.000079$; Table 3.4). OTUs significantly correlated with decreasing O_2 were affiliated with 2 subgroups of MGA (Arctic95A-2 and A714018), and an additional 13 OTUs affiliated with HF770D10, ZA3648c, Arctic96B-7, Arctic95A-2, SAR406 and A714018 were weakly correlated ($p<0.05$; Table 3.4). In addition, out of all 78 MGA OTUs identified by binning pyrotags to full-length 16S rRNA gene sequences, 10 displayed significant correlations ($p<0.000079$) with increasing depth, salinity and nutrients (nitrate, phosphate, silicate) or decreasing Chl*a* (Table 3.5).

Pyrotag OTU	Depth ρ	Temperature ρ	Salinity ρ	Oxygen ρ	Nitrate ρ	Phosphate ρ	Silicate ρ	Chl*a* ρ	Phylogenetic affiliation
MGA_100	0.000	0.071	0.017	-0.616 *	0.166	0.558 *	-0.071	0.067	HF770_D10
MGA_105	0.670 *	-0.648 *	0.634 *	-0.648 *	0.683 *	0.676 *	0.648 *	-0.627 *	HF770_D10
MGA_106	0.670 *	-0.648 *	0.634 *	-0.648 *	0.683 *	0.676 *	0.648 *	-0.627 *	HF770_D10
MGA_76	0.407	-0.366	0.380	-0.718 *	0.549	0.630 *	0.366	-0.500	HF770_D10
MGA_99	0.217	-0.058	0.221	-0.761 *	0.358	0.707 *	0.058	-0.096	ZA3648c
MGA_90	0.330	-0.256	0.359	-0.580 *	0.342	0.576 *	0.256	-0.313	Arctic96B-7
MGA_07	0.854 **	-0.732 *	0.810 *	-0.599 *	0.599 *	0.637 *	0.732 *	-0.718 *	Arctic96B-7
MGA_03	0.472	-0.514	0.423	-0.648 *	0.620 *	0.634 *	0.514	-0.486	Arctic95A-2
MGA_49	0.386	-0.345	0.359	-0.683 *	0.507	0.595 *	0.345	-0.486	Arctic95A-2
MGA_88	0.328	-0.331	0.359	-0.824 **	0.613 *	0.768 *	0.331	-0.331	Arctic95A-2
MGA_124	0.312	-0.326	0.291	-0.827 **	0.606 *	0.771 *	0.326	-0.396	Arctic95A-2
MGA_70	0.342	-0.275	0.359	-0.880 **	0.542	0.832 **	0.275	-0.289	Arctic95A-2
MGA_08	0.422	-0.324	0.465	-0.732 *	0.458	0.719 *	0.324	-0.317	SAR406
MGA_130	0.526	-0.444	0.570 *	-0.754 *	0.563	0.786 *	0.444	-0.338	SAR406
MGA_131	0.782 *	-0.746 *	0.725 *	-0.570 *	0.669 *	0.602 *	0.746 *	-0.697 *	A714018
MGA_50	0.724 *	-0.676 *	0.725 *	-0.725 *	0.739 *	0.786 *	0.676 *	-0.521	A714018
MGA_121	0.330	-0.317	0.349	-0.918 **	0.680 *	0.878 **	0.317	-0.235	A714018

* $p<0.05$; ** $p<0.000079$, Bonferroni corrected

Table 3.4 Pyrotag OTUs with statistically significant Spearman's rank correlations (*r*) with environmental parameters in the NESAP

Reported are all OTUs that display strong correlations ($p<0.000079$) with any of the environmental parameters.

Pyrotag OTU	Depth ρ	Temperature ρ	Salinity ρ	Oxygen ρ	Nitrate ρ	Phosphate ρ	Silicate ρ	Chl*a* ρ	Phylogenetic affiliation
MGA_09	0.887 **	-0.81 *	0.894 **	-0.486	0.585 *	0.609 *	0.81 *	-0.613 *	HF770D10
MGA_104	0.887 **	-0.838 **	0.859 **	-0.423	0.606 *	0.524	0.838 **	-0.718 *	HF770D10
MGA_114	0.822 **	-0.725 *	0.817 **	-0.394	0.451	0.485	0.725 *	-0.627 *	ZA3648c
MGA_07	0.854 **	-0.732 *	0.810 *	-0.599 *	0.599 *	0.637 *	0.732 *	-0.718 *	Arctic96B-7
MGA_95	0.816 **	-0.805 *	0.856 **	-0.276	0.573 *	0.379	0.805 *	-0.631 *	P262000N21
MGA_30	0.893 **	-0.865 **	0.865 **	-0.35	0.606 *	0.437	0.865 **	-0.837 **	Arctic95A-2
MGA_70	0.342	-0.275	0.359	-0.880 **	0.542	0.832 **	0.275	-0.289	Arctic95A-2
MGA_88	0.328	-0.331	0.359	-0.824 **	0.613 *	0.768 *	0.331	-0.331	Arctic95A-2
MGA_124	0.312	-0.326	0.291	-0.827 **	0.606 *	0.771 *	0.326	-0.396	Arctic95A-2
MGA_128	0.822 **	-0.796 *	0.754 *	-0.387	0.585 *	0.414	0.796 *	-0.831 **	Arctic95A-2
MGA_54	0.829 **	-0.81 *	0.881 **	-0.269	0.545	0.407	0.81 *	-0.519	A714018
MGA_121	0.330	-0.317	0.349	-0.918 **	0.680 *	0.878 **	0.317	-0.235	A714018

* $p<0.05$; ** $p<0.000079$, Bonferroni corrected

Table 3.5 Pyrotag OTUs with statistically significant Spearman's Rank correlations with environmental parameters in the NESAP

Reported are all OTUs that display strong correlations ($p<0.000079$) with any of the environmental parameters.

3.4 Discussion

MGA abundance estimates in the NESAP were highly correlated between CARD-FISH and 16S rRNA gene clone library, but not between CARD-FISH and pyrotag sequences, while 16S rRNA gene clone library and pyrotag sequences were correlated based on Spearman's rank correlations. Moreover, CARD-FISH-based estimates were consistently lower than 16S rRNA gene clone library or pyrotag sequence estimates for the same samples. For example, the average relative abundance of MGA sequences in O_2-deficient waters was 11.0%±3.9% based on 16S rRNA gene clone libraries, 9.9%±1.8% based on pyrotags, and 5.6%±1.9% based on CARD-FISH (Table 3.1). This suggests an under or overestimation of MGA abundance by one, some or all of the methods used. The discrepancy between methods could be purely based on primer and probe differences and the underlying methods applied. One perspective would be that CARD-FISH with probe SAR406-97 underestimated MGA abundance. Lower detection efficiency by CARD-FISH has been attributed to limited probe access to target cells when using HRP-labeled probes (Schönhuber et al., 1997), even after careful permeabilization optimization (Woebken et al., 2007). Also, the permeabilization step might cause leakage of ribosomes from target cells, which in turn could result in low-ribosome content cells dropping below the CARD-FISH detection limit (Hoshino et al., 2008). Alternatively, MGA subgroups could harbor variable copy numbers of the 16S rRNA gene, inflating PCR-based metrics (Acinas et al., 2004).

Rarefaction curves for MGA 16S rRNA gene clone library and pyrotag sequences recovered from the NESAP were consistent with known methodological limitations based on variable sample size and potential primer bias (Engelbrektson et al., 2010; Gihring et al., 2012; Schloss and Westcott, 2011). Clustering the combined data sets enabled pyrotag assignments to 78 out of 121 OTUs defined by 16S rRNA gene clone library sequences. The inability to assign pyrotags to all 121 OTUs may have resulted from the conservative nature of our clustering method: Full-length pyrotag sequences were required to match a cognate 16S rRNA gene clone library sequence with no mismatches. Alternatively, it is possible that time variable patterns in the abundance of MGA OTUs prevented assignment of all June 2009 pyrotags to OTUs identified in February 2009 16S rRNA gene clone libraries. Although ~50 to 75% of pyrotags identified as MGA in blast-based taxonomic queries were not assigned to OTUs defined by 16S rRNA gene clone library sequences, pyrotags affiliated with all ten MGA subgroups were

recovered. Indeed, comparison of 16S rRNA gene clone library and pyrotag sequence clusters revealed that the majority of MGA sequences (57%) and abundant MGA OTUs (containing >200 sequences) were identified using both methods (Figure 3.7). Unique pyrotag OTUs were generally composed of less than 50 sequences with a single abundant OTU containing 1 490 pyrotags that could not be assigned to defined MGA subgroups. Sequences in this OTU were recovered from 500, 1000, 1300 and 2000 m samples at all 3 stations indicating an environmental origin. The extent to which unique pyrotag OTUs captured components of the "rare biosphere" (Sogin et al., 2006) subject to time-variable changes in population structure remains to be determined. Despite this uncertainty, the recovery of a single abundant OTU unaffiliated with MGA subgroups defined by 16S rRNA gene clone library sequences suggests that the majority of abundant MGA subgroups in the NESAP have been identified.

Spearman's rank correlation coefficients provided statistical support for vertical partitioning of MGA subgroups in the NESAP water column. The relative abundance of MGA OTUs identified in pyrotags (affiliated with Arctic95A-2 and A714018) displayed a negative correlation with O_2 concentration consistent with habitat selection within suboxic waters (1-20 μmol kg^{-1}) of the OMZ. The extent to which patterns of vertical partitioning among and between MGA OTUs represent ecological types (ecotypes) (Koeppel et al., 2008) or class divisions remains to be determined. Environmental gradients are common drivers of selection among microorganisms at different ecological scales. For example, (Johnson et al., 2006) documented niche partitioning of *Prochlorococcus* ecotypes over ocean-basin scales across temperature (eMED4 vs. eMIT9312) and nutrient (eNATL2A or eMIT9313) gradients. Similarly, SAR11 ecotypes display depth-specific distributions with subclade Ia members more prevalent in the euphotic zone and subclade II members more abundant in deeper (mesopelagic) waters (Field et al., 1997). Such distribution patterns are associated with changes in genome composition that promote differential fitness including allelic variation (Urbach and Chisholm, 1998; Urbach et al., 1998; Wilhelm et al., 2007; Zhao and Qin, 2007) and metabolic island formation (Coleman et al., 2006; Coleman and Chisholm, 2007; Kettler et al., 2007; Rocap et al., 2003; Wilhelm et al., 2007).

Looking forward, genome-scale sequence data (i.e., single-cell and metagenomic data) representative of defined MGA subgroups will be invaluable both to more accurately assess

evolutionary relationships between MGA and thermophilic bacteria such as *Caldithrix*, as well as to attach metabolic repertoires to defined MGA subgroups (Shapiro et al., 2012; Swan et al., 2011). In turn, metabolic characterization of MGA subgroups will assist in determining whether observed 16S rRNA-based patterns of distribution across the oxycline are associated with variable forms of energy metabolism, consistent with redox-driven niche partitioning and ecotype differentiation. In addition, more extensive quantitative studies documenting the temporal dynamics of extant MGA subgroups across multiple provinces are needed to assess the stability of MGA population structure and function and better constrain the ecological and biogeochemical roles of MGA within OMZs.

3.5 Conclusions

In this chapter, MGA diversity and population structure were quantified in relation to nutrients and O_2 concentrations in the oxygen minimum zone (OMZ) of the Northeast subarctic Pacific Ocean using a combination of CARD-FISH and 16S rRNA gene sequencing (clone libraries and 454-pyrotags). Estimates of MGA abundance as a proportion of total bacteria were similar across all three methods although estimates based on CARD-FISH were consistently lower in the OMZ (5.6%±1.9%) than estimates based on 16S rRNA gene clone libraries (11.0%±3.9%) or pyrotags (9.9%±1.8%). Five previously defined MGA subgroups were recovered in 16S rRNA gene clone libraries and five novel subgroups were defined (HF770D10, P262000D03, P41300E03, P262000N21, and A714018). The relative abundance of MGA OTUs identified in pyrotags (affiliated with Arctic95A-2 and A714018) comprised 0.3 to 2.4% of total bacterial sequences and displayed a negative correlation with O_2 concentration consistent with habitat selection within suboxic waters (1-20 μmol kg^{-1} O_2) of the OMZ.

Chapter 4: Genomic analysis of large-insert DNA fragments derived from Marine Group A bacteria[3]

4.1 Synopsis

Marine Group A is a deeply-branching and uncultivated phylum of bacteria (see Chapter 1, section 1.4.1.3). Although their functional roles remain elusive, MGA subgroups are particularly abundant and diverse in oxygen minimum zones (OMZs) and permanent or seasonally stratified anoxic basins suggesting metabolic adaptation to O_2-deficiency. This chapter expands on the previous survey of MGA diversity in O_2-deficient waters of the Northeast subarctic Pacific Ocean (reported in Chapter 3) to include Saanich Inlet (SI), an anoxic fjord with seasonal O_2 gradients and periodic sulfide accumulation. Phylogenetic analysis of small subunit ribosomal RNA (16S rRNA) gene clone libraries recovered five previously described MGA subgroups and defined three novel subgroups (SHBH1141, SHBH391, and SHAN400) in SI. Determining the extent to which 16S rRNA-based patterns of MGA distribution represent ecological types (ecotypes) differentiating in response to selective environmental pressures such as O_2-deficiency requires genome-scale sequence data associated with multiple MGA subgroups to query for changes in genome composition that might promote differential fitness across the oxycline. To discern functional properties and potential niche partitioning of MGA residing along gradients of O_2 in the NESAP and SI, 14 fosmids harbouring MGA-associated 16S RNA genes were identified from a collection of 23 fosmid libraries sourced from NESAP and SI waters and sequenced to completion. Comparative analysis of these fosmids, in addition to 4 publicly available MGA-associated large-insert DNA fragments from Hawaii Ocean Time-series and Monterey Bay, revealed widespread genomic differentiation proximal to the ribosomal RNA operon that did not consistently reflect subgroup partitioning patterns observed in 16S rRNA gene clone libraries. Predicted protein-coding genes associated with adaptation to O_2-deficiency and sulfur-based energy metabolism were detected on multiple fosmids, including polysulfide reductase (*psrABC*), implicated in dissimilatory polysulfide reduction to hydrogen sulfide and

[3] A version of this chapter has been accepted for publication in the International Society for Microbial Ecology (ISME) Journal.

dissimilatory sulfur oxidation. These results posit a potential role for specific MGA subgroups in the marine sulfur cycle.

4.2 Materials and Methods

4.2.1 Sample collection and processing in the NESAP

Sampling was conducted via multiple hydrocasts using a rosette water sampler, with an attached Conductivity, Temperature, Depth (CTD) sensor aboard the *CCGS John P. Tully* during Line P cruises 2009-09 (June 2009), 2009-10 (August 2009), and 2010-01 (February 2010). Major stations sampled on cruise 2009-09 include: P4 (48º39.0N, 126º4.0W) – June 7th, P12 (48º58.2N, 130º40.0W) – June 9th, and P26 (50ºN, 145ºW) – June 14th. Major stations sampled on cruise 2009-10 include: P4 – August 21st, P12 – August 23rd, P26 – August 27th. Major stations sampled on cruise 2010-10 include: P4 – February 4th, P12 – February 11th. At each of these sampling stations, 20 L samples for DNA isolation were collected from the surface (10 m), while 120 L samples were taken from three depths spanning the OMZ core and upper and deep oxyclines (500 m, 1000 m, 1300 m at station P4; 500 m, 1000 m, 2000 m at station P12). Sampling at Saanich Inlet station S3 (48°35.30N, 123°30.22W) was performed as previously described (Zaikova et al., 2010) as part of a monthly monitoring program aboard the *MSV John Strickland*. Sample collection and filtration protocols can be viewed as visualized experiments at http://www.jove.com/video/1159/ (Zaikova et al., 2009) and http://www.jove.com/video/1161/ (Walsh et al., 2009) respectively. DNA was extracted from sterivex filters as described in Zaikova and colleagues (2010) and DeLong and colleagues (2006) – see Chapter 2 section 2.2.3 for a detailed explanation. The DNA extraction protocol can be viewed as a visualized experiment at http://www.jove.com/video/1352/ (Wright et al., 2009).

4.2.2 Phylogenetic analysis and tree construction using MGA 16S rRNA gene sequences

Full-length 16S rRNA gene clone sequences from the NESAP (3 164; (Allers et al., 2012)) and SI (6 645; (Zaikova et al., 2010)) as well as partial and full-length 16S rRNA sequences obtained from large-insert DNA fragments affiliated with MGA were imported in the ARB software package (Release 106; (Ludwig et al., 2004)), added to the SILVA database (www.arb-silva.de)

(Pruesse et al., 2007), aligned to the closest relative, and added to an existing tree of sequences from the ARB database by using the ARB parsimony tool (using default parameters).

A maximum likelihood phylogenetic tree of MGA 16S rRNA gene sequences exported from ARB was inferred by PHYML (Guindon et al., 2005) using an HKY + 4G + I model of nucleotide evolution where the parameter of the G distribution, the proportion of invariable sites, and the transition/transversion ratio were estimated for each dataset. The confidence of each node was determined by assembling a consensus tree of 100 bootstrap replicates. Non-chimeric bacterial 16S rRNA gene sequences were also placed in taxonomic hierarchy for downstream analysis using the NAST aligner (DeSantis et al., 2006b) and blast using default parameters against the 2008 Greengenes database (DeSantis et al., 2006a), and 705 sequences were identified as belonging to MGA (415 from SI in addition to 290 previously reported in by Allers, Wright and colleagues (Allers et al., 2012)). These 705 sequences were clustered at 97% identity using mothur (Schloss et al., 2009) (v.1.19.0). Representative sequences from each of these clusters were identified using the get.oturep command in mothur and were included in the phylogenetic tree. The abundance and distribution of 97% clusters was visualized in a histogram-heatmap in R.

4.2.3 Fosmid library construction, end sequencing, screening, preparation and full-length sequencing

Thirty fosmid libraries (~7,680 clones/library) were constructed from DNA samples collected from NESAP stations P4, P12, and P26 in June and August of 2009, and stations P4 and P12 during February 2010 (Table 4.2). An additional 16 fosmid libraries were constructed from DNA samples collected from SI station S3 during the 2006-2007 seasonal stratification and deep-water renewal cycle (Table 4.2) (Walsh et al., 2009). Prior to cloning, ~4 μg of environmental DNA was further purified on a CsCl density gradient as previously described (Hallam et al., 2004). Fosmid libraries were prepared using the CopyControl Fosmid Library Production Kit (Epicentre, Madison, WI). Briefly, ~1 μg of CsCl-purified DNA was blunt end repaired and separated on a 1% low melt agarose pulse-field gel O/N at 6 V/cm. The 40-50 kb fragment range was excised and gel purified using agarase, followed by concentration using an Amicon Ultracel 10K filter device (Millipore, Billerica, MA, USA). DNA was ligated into the pCC1fos vector,

packaged using the MaxPlax lambda packaging extract, and used to transfect TransforMax EPI300 *E. coli* cells (Epicentre). Transfected cells were plated on selective agar and fosmid clones picked using the QPix2 robotic colony picker (Molecular Devices, Sunnyvale, CA) and grown in selective media for DNA sequencing. The fosmid library production protocol can be viewed as a visualized experiment at http://www.jove.com/index/Details.stp?ID=1387 (Taupp et al., 2009). Bidirectional end sequencing of SI fosmids was performed with standard M13 forward (5'-GTTTTCCCAGTCACGAC) and reverse (5'-CAGGAAACAGCTATGAC) primers and the BigDye sequencing kit (Applied Biosystems, Carlsbad, CA) on a Sanger platform at the Department of Energy's Joint Genome Institute (DOE-JGI; Walnut Creek, CA). The reactions were purified by a magnetic bead protocol and run on an ABI PRISM3730 (Applied Biosystems) capillary DNA sequencer (for research protocols, see http://jgi.doe.gov). Bidirectional end sequencing of NESAP fosmids was performed with standard pCC1 forward (5'-GGATGTGCTGCAAGGCGATTAAGTTGG) and reverse (5'-CTCGTATGTTGTGTGGAATTGTGAGC) primers on a Sanger platform at Canada's Michael Smith Genome Sciences Centre (GSC; Vancouver, BC).

The 7 NESAP fosmid end sequenced libraries from February 2010 and all 16 SI fosmid end sequenced libraries were screened for the presence of 16S rRNA genes (using the NAST aligner (DeSantis et al., 2006b) and blast using default parameters against the 2008 Greengenes database (DeSantis et al., 2006a). After partial sequencing and preliminary phylogenetic analyses, 14 fosmid clones affiliated with MGA were selected for complete sequencing (Table 4.1). Sequencing of the 6 SI fosmids was carried out at the DOE-JGI as part of a Community Sequencing Proposal (CSP) on an ABI PRISM3730 (Applied Biosystems) capillary DNA sequencer (for research protocols, see http://jgi.doe.gov). Sequencing of the 8 NESAP fosmids was performed in-house using the IonTorrent PGM (Life Technologies, San Francisco, CA, USA) at the University of British Columbia (as sequencing of these fosmids was not available through the DOE-JGI CSP). Briefly, fosmid DNA was prepared using Montage Plasmid96 Miniprep kit (Millipore), and 100 ng of template was used in barcoded library construction for 200 bp read length libraries according to standard protocols provided with the IonTorrent PGM. These 8 libraries were sequenced with two Ion316 chips. Runs were combined and processed, yielding between 33 261 and 76 270 reads for each fosmid. Raw data was assembled using the

MIRA assembler (Chevreux et al., 2004), which gave outputs ranging from 2 to 77 contigs. Contigs were further processed using Sequencher 4.8 (GeneCodes Corp, Ann Arbor, MI, USA) to combine contigs using default settings (20 bp overlap, 85% similarity). Any mismatches in the overlapping regions were replaced with N. Contigs were then compared to the original end sequences to ensure proper identity, yielding one contig from each assembly that matched both original end sequences in 7 of 8 cases. In 5 of these 7 cases the vector was found in the middle of the contig, necessitating its removal. For these 5 contigs, the vector sequence was trimmed out and the resulting two contigs were joined at the opposite ends with a string of 100 Ns. One fosmid (413009-K18) produced 2 contigs (16.8 kb and 18.7 kb) with each matching either the forward or reverse end sequence. In some cases limited coverage introduced sequencing errors interrupting open reading frames. Eleven of these regions were identified and primers were designed targeting these regions for verification with Sanger sequencing. Primers to these regions are provided in Appendix B. GenBank files contain the Sanger-verified fosmid sequences.

4.2.4 Analysis of large insert DNA fragments

GC content and oligonucleotide frequency analysis

GC content of large-insert DNA fragments (14 fosmids from NESAP and SI in addition to 4 large-insert DNA fragments from other North Pacific Ocean environments; Table 4.1) was calculated using GCcontent.pl with default parameters, available for download at http://hallam.microbiology.ubc.ca/downloads/index.html. Tetranucleotide frequencies were calculated as normalized Z-scores using TETRA ((Teeling et al., 2004a; Teeling et al., 2004b); http://www.megx.net/tetra). Principal component analysis (PCA) was performed on normalized Z-score profiles for each insert using PRIMER v6.1.13 (Clarke, 1993; Clarke and Gorley, 2006). PCA was overlaid with clusters determined by Hierarchical Cluster Analysis of normalized Z-scores using a Euclidean distance matrix (also performed in PRIMER).

Global nucleotide similarity analysis

Global nucleotide similarity in large-insert DNA fragments was determined by performing pairwise blastn comparisons between all fragments using onecircos.pl with default settings for all

parameters except percent_identity (-p), which was calculated at 50%, 80%, 90%, and 95% in separate analyses. Onecircos.pl is available for download at: http://hallam.microbiology.ubc.ca/downloads/index.html and is based on Circos (http://circos.ca/ (Krzywinski et al., 2009)).

Open reading frame (ORF) prediction and gene annotation

Open-reading frames (ORFS) were predicted and annotated using the in-house MetaPathways pipeline, available for download at: http://hallam.microbiology.ubc.ca/MetaPathways/. Briefly, primary nucleotide sequences from large-insert DNA fragments were quality controlled for ambiguous bases and file-format errors. ORFs were predicted using Prodigal (Hyatt et al., 2010). ORFs shorter than 60 amino acids in length were removed and were annotated using Protein BLAST (Altschul et al., 1990) (bit-score ratio > 0.4 (Rasko et al., 2005), e-value = 1e-5) against the RefSeq (Pruitt and Maglott, 2001), KEGG (Kanehisa and Goto, 1999), COG (Tatusov et al., 2001), and MetaCyc (Karp et al., 2000) databases. Annotations were assigned to predicted ORFs based on the following five criteria: i) BLAST hit with top e-value was selected from each database; ii) each BLAST hit was assigned an 'information score' based on the sum of distinct and shared enzymatic words (prepositions, articles, and auxiliary verbs were removed) and a preference to Enzyme Commission (EC) numbers (+10 score) was assigned; iii) annotation with the highest score was selected and assigned to the respective ORF; iv) ORFs with no hits were assigned the annotation 'hypothetical protein'.

Amino acid similarity analysis

Predicted amino acid similarity of large-insert DNA fragments was plotted in Trebol (available for download at: http://bioinf.udec.cl/trebol), using tblastx with a minimum bit score cutoff of 50. COG categories present on large-insert fragments were plotted using tblastn of COG proteins against large-insert DNA fragments with a minimum e-value cutoff of 1e-4.

4.2.5 Fragment recruitment of fosmid end sequences

Coverage plots relating fosmid end sequences from individual NESAP and SI fosmid end libraries to large-insert DNA fragments were generated by using the Nucmer program

implemented in MUMmer 3.23 (Kurtz et al., 2004) using the following parameters as cited in (Hallam et al., 2006): breaklength = 60, minimum cluster length = 20, and match length = 10. Resulting delta files were converted into coordinate files using the show-coords program and visualized in graphical format (coverage plot) by using the MUMmerplot program. Also using the coordinate files, the number of fosmid end sequences recruited to each large insert DNA fragment was calculated at 60% - 80% nucleotide similarity and at >80% nucleotide similarity, ends recruiting to the 16S-23S rRNA region were subtracted, remaining ends were normalized to total number of ends per library, and the normalized proportion of sequences in each library recruited to each large-insert fragment was visualized using bubble.pl (available for download at: http://hallam.microbiology.ubc.ca/downloads/index.html). The number of fosmid end sequences recruited to the *psr* operon on fosmids FPPP_13C3 and 122006-I05 was also calculated and visualized as described above.

4.2.6 Phylogenetic analysis of polysulfide reductase (PsrABC)

Protein sequences (including predicted protein sequences for PsrA, PsrB, and PsrC identified on fosmids FPPP_13C3 and 122006-I05) were aligned using MUSCLE v3.6 with default parameters (Edgar, 2004). For the purposes of this analysis, the PsrBC fusion proteins encoded by *psrBC* on fosmid FPPP_13C3 and on certain reference sequences were divided into PsrB and PsrC subunits and analysed in separate trees. Phylogenetic analyses were performed using PHYML (Guindon et al., 2005) using a WAG model of amino acid substitution where the parameter of the G distribution and the proportion of invariable sites were estimated for each dataset. The confidence of each node was determined by assembling a consensus tree of 100 bootstrap replicates.

4.3 Results

4.3.1 Physicochemical characteristics of the NESAP and SI

This study was conducted along the Line P transect of the NESAP (Figure 4.1), beginning in Saanich Inlet, Vancouver Island, British Columbia (SI, Station S3: 48º58'N, 123º50'W) and ending at Ocean Station Papa (OSP, also referred to as station P26: 50ºN, 145ºW) (Freeland, 2007). Due to strong stratification and sluggish circulation of interior NESAP waters, a large

region of O_2-deficient (<90 μmol kg^{-1}) water spans from ~400 - 2000 m in depth resulting in a persistent OMZ (O_2 <20 μmol kg^{-1}). The OMZ is centered at 1000 m wherein dissolved O_2 concentrations typically reach minimum levels of ~ 9 μmol kg^{-1} (Whitney et al., 2007). During the past 50 years of oceanographic observation, O_2 concentrations in the OMZ of coastal to open-ocean regions of the NESAP have not been observed to reach anoxic (<1 μmol kg^{-1}) levels. However, interior and basin waters of SI typically experience seasonal periods of anoxia and sulfide accumulation on an annually recurring basis (Anderson and Devol, 1973; Lilley et al., 1982; Ward et al., 1989). Physicochemical data from basin (S3), coastal (P4), transition (P12), and open-ocean (P26) stations measured along the Line P transect relevant to the present study are provided in Table 4.1 and Table 4.2.

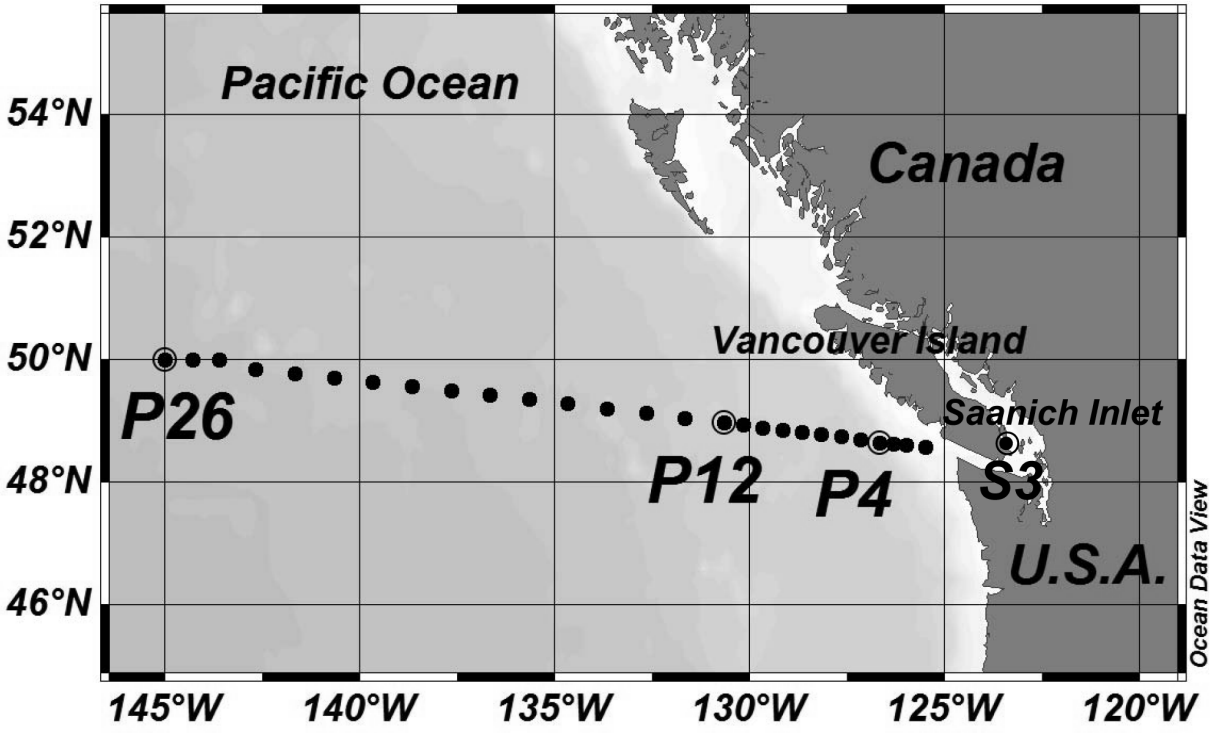


Figure 4.1 Stations P4, P12, and P26 along the Line P oceanographic transect of the NESAP, and Station S3 in SI are highlighted

	Fosmid insert size (bp)	GC content (%)	# of ORFS	# of ORFs encoding hypotheticals	Accession number	Sampling location	Sampling coordiates	Collection date	Depth (m)	[O_2] (µmol/kg)	Phylogenetic Identity	Syntenic group	Reference
Saanich Inlet													
FPPP_13C3	40,370	32.8	36	19	KF170421	Station S3	48°58'N, 123°50'W	Nov-06	10	249.0	Arctic96B-7	II	This study
FPPP_33K14	27,446	37.4	17	15	KF170417	"	"	Nov-06	10	249.0	A714018	IV	"
FGYC_13M19	32,678	43.6	16	11	KF170416	"	"	Feb-06	125	5.0	SHBH391	I	"
FPPS_57A9	32,217	37.9	21	13	KF170418	"	"	Nov-06	100	15.4	SHAN400	I	"
FPPU_33B15	32,230	37.0	14	12	KF170419	"	"	Nov-06	200	54.0	SAR406	IV	"
FPPZ_5C6*	41,899	38.2	36	26	KF170420	"	"	Apr-07	200	1.1	Arctic96B-7	I	"
Monterey Bay													
EBAC750-03B02	34,714	42.0	21	15	AY458631	Monterey Bay	36°41'N, 122°02'W	Apr-00	750	16.5	Arctic95A-2	III	Suzuki *et al.*, 2004
Hawaii Ocean Time-series													
HF0010_18O13	37,088	33.0	33	17	GU474850	Station ALOHA	22° 45'N, 158° 00'W	Oct-02	10	204.6	ZA3312c	I	DeLong *et al.*, 2006
HF0500_01L02	32,792	37.7	25	19	GU474916	Station ALOHA	"	Oct-02	500	118.0	Arctic96B-7	I	"
HF4000_22B16	31,597	43.0	17	15	GU474892	Station ALOHA	"	Dec-03	4000	147.8	P262000N21	IV	"
NESAP													
405006-B04	43,543	42.8	46	32	KF170424	Station P4	48°39'N, 126°40'W	Feb-10	500	40.4	SAR406	I	This study
4050020-J15	43,860	39.3	49	39	KF170415	"	"	"	500	40.4	*Deferribacteres*-like	V	"
413004-H17	35,813	39.8	45	29	KF170425	"	"	"	1300	22.7	Arctic96B-7	I	"
4130011-I07	34,533	41.5	39	27	KF170426	"	"	"	1300	22.7	Arctic95A-2	III	"
413009-K18	35,438	41.4	25	23	KF170413	"	"	"	1300	22.7	SAR406	IV	"
125003-E23	34,144	37.3	41	28	KF170423	Station P12	48°58'N, 130°40'W	"	500	32.2	Arctic96B-7	I	"
1250012-L08	34,444	40.6	35	26	KF170414	"	"	"	500	32.2	Arctic95A-2	III	"
122006-I05	35,932	47.7	45	26	KF170422	"	"	"	2000	59.5	P262000D03	II	"

*This fosmid was derived from an H_2S-containing sample.

Table 4.1 Characterization of large-insert DNA fragments containing MGA 16S rRNA genes

A. Saanich Inlet

Cruise ID	Sample Date (mm/dd/yy)	Station	Sample Depth (m)	[O_2] (µmol/kg)	[NO_3^-] (µmol/L)	[H_2S] (umol/L)	Bac lib ID[1]	# of 16S rRNA clones	Fosmid end lib ID[2]	Fosmid end lib accession #	# of fosmid clones	# of end reads[3]	Total Mb DNA
Feb-06	06-02-18	S3	10	212	26.7	-	SGPW	95	FGYA	LIBGSS_039102	6,144	11,926	7.4
Feb-06	06-02-18	S3	100	51.1	21	-	SGPZ	234	FGYB	LIBGSS_039103	6,528	11,473	7.1
Feb-06	06-02-18	S3	125	5	9.7	-	SGSC	373	FGYC	LIBGSS_039104	7,296	12,899	8.8
Feb-06	06-02-18	S3	215	<1	1.8	-	SGSH	555	FGYF	LIBGSS_039105	7,680	12,286	8.0
Jul-06	06-07-06	S3	10	381.6	4.6	-	SGSO	267	FGYG	LIBGSS_039106	7,680	12,871	8.7
Jul-06	06-07-06	S3	100	22.9	16.1	-	SGST	230	FGYH	LIBGSS_039107	7,296	11,388	7.2
Jul-06	06-07-06	S3	120	6.5	4.8	-	SGSX	226	FGYI	LIBGSS_039108	6,528	11,957	7.5
Jul-06	06-07-06	S3	200	<1	0.5	-	SGTA	257	FGYN	LIBGSS_039109	7,680	11,377	7.2
Nov-06	06-11-14	S3	10	249	25.5	-	SHAB	362	FPPP	LIBGSS_039110	6,528	10,631	6.0
Nov-06	06-11-14	S3	100	15.4	13	-	SHAG	346	FPPS	LIBGSS_039111	6,912	12,272	8.1
Nov-06	06-11-14	S3	120	9.8	8.9	-	SHAN	359	FPPT	LIBGSS_039112	7,680	13,128	6.8
Nov-06	06-11-14	S3	200	54	19.8	-	SHAS	316	FPPU	LIBGSS_039113	7,296	11,778	6.5
Apr-07	07-04-24	S3	10	316.1	18.2	0	SHAW	308	FPPW	LIBGSS_039114	5,760	11,354	7.0
Apr-07	07-04-24	S3	100	67.3	26.5	0	SHAZ	358	FPPX	LIBGSS_039115	5,760	11,989	6.0
Apr-07	07-04-24	S3	120	26.9	20.6	0	SHBC	720	FPPY	LIBGSS_039116	6,912	13,280	7.1
Apr-07	07-04-24	S3	200	<1	0	5.6	SHBH	690	FPPZ	LIBGSS_039117	7,296	13,149	8.7
Apr-08	08-04-09	S3	100	120	20.5	0	SHZW	298	-	-	-	-	-
Apr-08	08-04-09	S3	120	18.1	13.2	0	SHZZ	352	-	-	-	-	-
Apr-08	08-04-09	S3	200	<1	0.1	2.1	SIAC	299	-	-	-	-	-
TOTAL								**6,645**			**110,976**	**193,758**	**118.3**

B. NESAP

Cruise ID	Sample Date (mm/dd/yy)	Station	Sample Depth (m)	[O_2] (µmol/kg)	[NO_3^-] (µmol/L)	[H_2S] (umol/L)	Bac lib ID[1]	# of 16S rRNA clones	Fosmid end lib ID[2]	Fosmid end lib accession #	# of fosmid clones	# of end reads[3]	Total Mb DNA
Feb-09	09-01-29	P4	10	289.7	13.1	-	F9P410	281	-	-	-	-	-
Feb-09	09-01-29	P4	500	27.6	-	-	F9P4500	287	-	-	-	-	-
Feb-09	09-01-29	P4	1000	12.5	45.1	-	F9P41000	276	-	-	-	-	-
Feb-09	09-01-29	P4	1300	23.7	44.5	-	F9P41300	239	-	-	-	-	-
Feb-09	09-01-31	P12	10	295	9.2	-	F9P1210	248	-	-	-	-	-
Feb-09	09-01-31	P12	500	41.4	-	-	F9P12500	249	-	-	-	-	-
Feb-09	09-01-31	P12	1000	11	46.5	-	F9P121000	184	-	-	-	-	-
Feb-09	09-01-31	P12	2000	56.6	44	-	F9P122000	256	-	-	-	-	-
Feb-09	09-02-04	P26	10	9.9	47.1	-	F9P2610	242	-	-	-	-	-
Feb-09	09-02-04	P26	500	35.6	-	-	F9P26500	287	-	-	-	-	-
Feb-09	09-02-04	P26	1000	16.3	47	-	F9P261000	293	-	-	-	-	-
Feb-09	09-02-04	P26	2000	56.6	44.3	-	F9P262000	322	-	-	-	-	-
Jun-09	09-06-07	P4	10	308.0	0.0	-	-	-	GOUP	LIBGSS_039090	6,528	13,374	8.6
Jun-09	09-06-07	P4	500	23.7	42.6	-	-	-	GOUO	LIBGSS_039089	5,760	10,662	6.5
Jun-09	09-06-07	P4	1000	8.6	45.3	-	-	-	GOUN	LIBGSS_039088	6,528	12,423	7.5
Jun-09	09-06-07	P4	1300	15.9	45.8	-	-	-	GOUI	LIBGSS_039087	4,992	9,249	5.5
Jun-09	09-06-09	P12	10	296.2	6.3	-	-	-	GOUH	LIBGSS_039086	6,912	11,471	7.3
Jun-09	09-06-09	P12	500	37.0	42.8	-	-	-	GOUG	LIBGSS_039085	6,912	10,841	6.7
Jun-09	09-06-09	P12	1000	9.0	46.3	-	-	-	GOUF	LIBGSS_039084	6,912	13,132	8.4
Jun-09	09-06-09	P12	2000	59.3	44.1	-	-	-	GOUC	LIBGSS_039083	3,840	6,372	4.0
Jun-09	09-06-14	P26	10	301.4	10.8	-	-	-	GOUB	LIBGSS_039082	4,608	6,375	10.2
Jun-09	09-06-14	P26	500	35.0	43.6	-	-	-	GOUA	LIBGSS_039081	6,912	9,751	21.1
Jun-09	09-06-14	P26	1000	14.3	45.6	-	-	-	GOTZ	LIBGSS_039080	6,144	11,606	7.3
Jun-09	09-06-14	P26	2000	56.5	44.4	-	-	-	GOPP	LIBGSS_039079	6,144	12,381	8.0
Aug-09	21/8/09	P4	10	283.6	5.2	-	-	-	GTZS	LIBGSS_039091	6,912	12,158	7.5
Aug-09	21/8/09	P4	500	28	42	-	-	-	GTZT	LIBGSS_039092	8,064	8,061	22.3
Aug-09	21/8/09	P4	1000	11.6	45	-	-	-	GTZU	LIBGSS_039093	7,296	13,959	9.2
Aug-09	21/8/09	P4	1300	21.1	45.3	-	-	-	GTZW	LIBGSS_039094	6,528	7,158	15.2
Aug-09	23/8/09	P12	10	5.88	1.2	-	-	-	GTZX	LIBGSS_039095	6,528	11,063	7.1
Aug-09	23/8/09	P12	500	38.7	42.7	-	-	-	GTZY	LIBGSS_039096	5,376	5,833	13.4
Aug-09	23/8/09	P12	1000	0.27	45.9	-	-	-	GTZZ	LIBGSS_039097	6,528	11,948	7.9
Aug-09	23/8/09	P12	2000	55.9	43.6	-	-	-	GUAA	LIBGSS_039098	7,296	12,993	8.7
Aug-09	27/8/09	P26	10	269.5	7.9	-	-	-	GUAB	LIBGSS_039099	4,224	9,948	6.2
Aug-09	27/8/09	P26	500	36.4	43.2	-	-	-	GUAC	LIBGSS_039100	5,376	8,850	5.5
Aug-09	27/8/09	P26	1000	17.3	45.4	-	-	-	GUAF	LIBGSS_039101	3,456	7,244	4.3
Feb-10	10-02-04	P4	10	275.9	6.8	-	-	-	40010	LIBGSS_039075	7,680	14,275	19.8
Feb-10	10-02-04	P4	500	40.4	39.9	-	-	-	40500	LIBGSS_039076	7,680	14,705	19.8
Feb-10	10-02-04	P4	1000	11.9	44	-	-	-	41000	LIBGSS_039077	7,680	14,701	19.4
Feb-10	10-02-04	P4	1300	22.7	44.2	-	-	-	41300	LIBGSS_039078	7,680	14,488	19.4
Feb-10	10-02-11	P12	10	290.7	7.1	-	-	-	12010	LIBGSS_039072	7,680	12,477	19.6
Feb-10	10-02-11	P12	500	32.2	42.2	-	-	-	120500	LIBGSS_039074	7,680	14,886	19.4
Feb-10	10-02-11	P12	2000	59.5	43.7	-	-	-	12200	LIBGSS_039073	7,680	14,740	19.5
TOTAL								**3,164**			**193,536**	**337,124**	**345.4**

[1]ID for bacterial 16S rRNA gene clone library
[2]ID for fosmid end library
[3]Number of end reads sequenced per fosmid library
- Not determined / Not applicable

Table 4.2 Sample summary and library key

4.3.2 Taxonomic diversity of MGA in the NESAP and SI

To identify 16S rRNA genes affiliated with MGA inhabiting SI waters, 19 previously published bacterial 16S rRNA gene clone libraries (containing a total of 6 645 sequences) were screened for MGA sequences. These libraries were generated from samples traversing the water column

during the 2006-2007 seasonal stratification and deep-water renewal cycle and during the spring stratification in 2008 at Station S3 (Table 4.2; (Walsh and Hallam, 2011)). A total of 415 16S rRNA gene sequences affiliated with MGA were recovered from SI clone libraries. These sequences were added to a dataset containing 290 MGA 16S rRNA sequences previously reported from NESAP stations P4, P12, and P26 (Allers et al., 2012) and clustered at 97% identity, forming 156 distinct operational taxonomic units (OTUs), 120 of which contained only singletons. Representative sequences were obtained for each non-singleton OTU and placed in phylogenetic context with relevant reference sequences from other locations (Figure 4.1). Five out of 10 previously defined MGA subgroups were recovered in SI clone libraries (ZA3648c and ZA3312c (Fuchs, unpublished); Arctic96B-7 (Bano and Hollibaugh, 2002); SAR406 (Gordon and Giovannoni, 1996), and A714018 (Allers et al., 2012) and three novel subgroups were identified (SHBH1141, SHBH391, and SHAN400) (Figure 4.2). These novel subgroups were found exclusively in SI and contained the most abundant OTUs identified in this location (Figure 4.3).

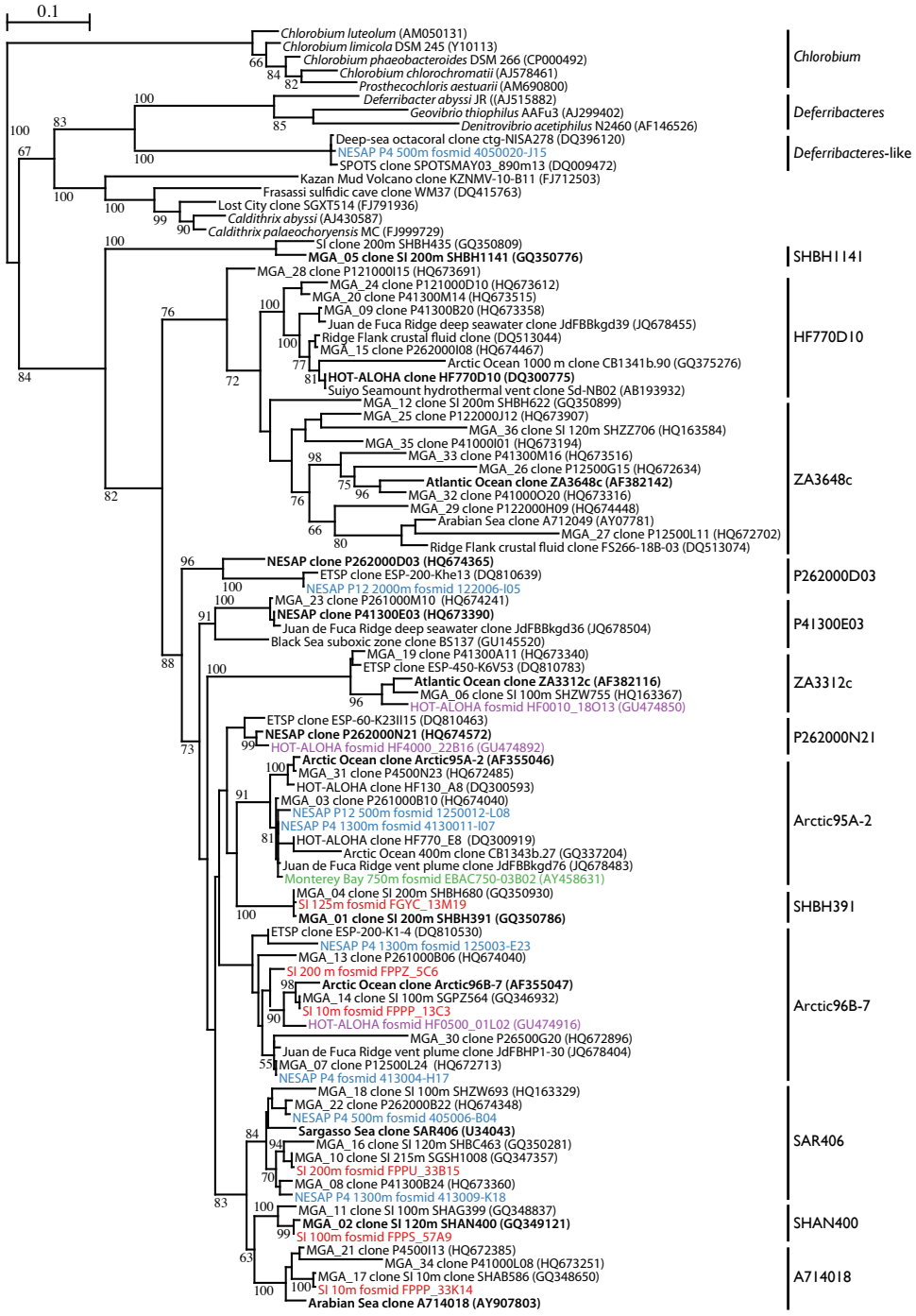


Figure 4.2 Unrooted phylogenetic tree based on 16S rRNA gene clone and large-insert DNA fragment sequences showing the phylogenetic affiliation of MGA sequences

The tree was inferred using maximum likelihood implemented in PhyML. Representative 16S rRNA gene sequences for each subgroup are shown in bold, and 16S rRNA gene sequences derived from large-insert DNA fragments described in this study are shown in colour; blue: NESAP; purple: HOT Station ALOHA; green: Monterey Bay; red: SI. The bar represents 10% estimated sequence divergence. Bootstrap values below 50% are not shown.

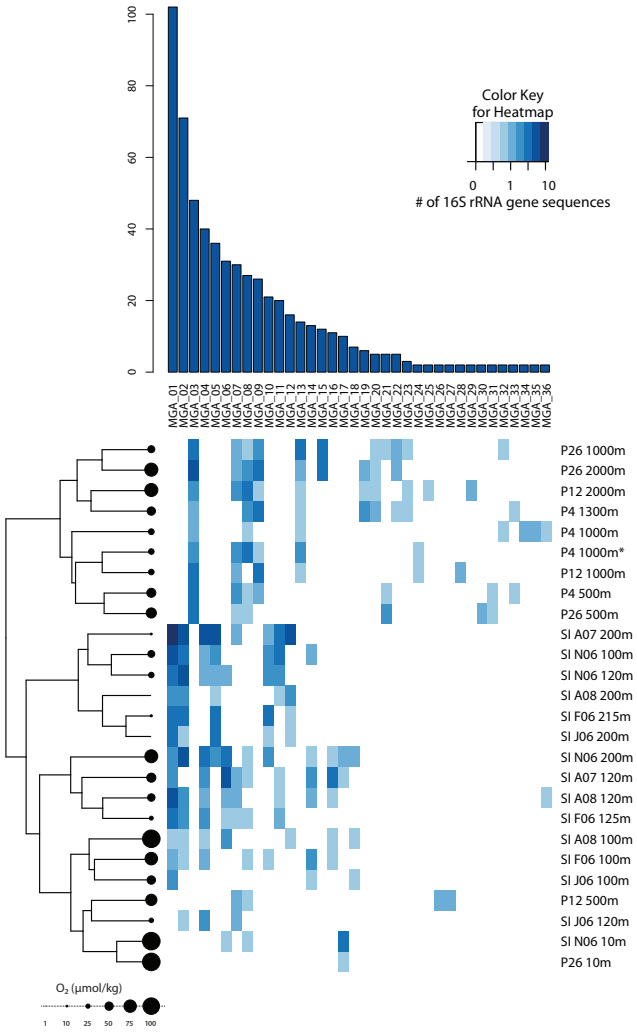


Figure 4.3 Distribution of 16S rRNA sequences affiliated with MGA identified in NESAP and Saanich Inlet clone libraries

Each histogram bar represents a cluster of 16S rRNA gene sequences, or an operational taxonomic unit (OTU), generated at a 97% identity cutoff (clustered using the average-neighbor algorithm). The height of the bar is equivalent to the sum of all sequences belonging to a specific OTU across all environments surveyed. Only non-singleton OTUs (36) are shown; an additional 120 OTUs were singletons. '*' indicates a sample taken from P4 1000m in June 2008; all other NESAP samples were taken in 2009. Heat maps below the histograms represent the distribution of sequences in each OTU within the NESAP and Saanich Inlet. Heat map rows were clustered by row using Euclidean distance and the average-neighbor algorithm to highlight patterns of diversity among samples. Inset colour scale depicts the colour code for the number of 16S rRNA gene sequences in heat maps.

As described in Allers, Wright *et al.* (2012), MGA sequences identified in coastal and open ocean waters of the NESAP comprised 0.7±0.84% of 10 m clone libraries and 11.2±3.9% of clone libraries from O_2-deficient waters, with a maximum of 16.4% at P26 1000 m. The most abundant MGA OTUs present in these locations comprised between 1 and 4% of clone libraries and belonged to subgroups Arctic95A-2, ZA3312c, Arctic96B-7, SAR406, and HF770D10, in order of decreasing OTU abundance (Figure 4.2, Figure 4.3). In comparison, MGA OTUs identified in SI comprised 1.6±0.81% of 10 m clone libraries and 7.1±3.6% of clone libraries from O_2-deficient waters. The most abundant OTUs present in SI comprised between 1 and 5% of clone libraries, and belonged to subgroups SHBH391, SHAN400, SHBH1141, ZA3312c, SAR406, and Arctic96B-7, in order of decreasing OTU abundance (Figure 4.2, Figure 4.3).

4.3.3 Characterization and phylogenetic assignment of large-insert DNA fragments

To connect 16S rRNA-based patterns of distribution across the oxycline in the NESAP and SI to genomic information associated with specific MGA subgroups, 23 previously constructed and end sequenced fosmid libraries generated from NESAP and SI samples were screened for the presence of clones containing 16S rRNA gene sequences. Collectively, fosmid end libraries contained a total of 164 736 genomic clones representing 255.3 Mb of environmental genomic DNA (Table 4.2). Screening of fosmid end sequences for 16S rRNA genes uncovered 14 fosmid inserts containing partial or full-length 16S rRNA gene sequences affiliated with MGA (Table 4.1). These 14 fosmid inserts were fully sequenced (Materials and Methods) for downstream analyses. In addition, 4 large-insert DNA fragments from Hawaii Ocean Time-series (HOT) Station ALOHA (DeLong et al., 2006; Rich et al., 2011) and Monterey Bay (Suzuki et al., 2004) harboring MGA 16S rRNA gene sequences were identified in public databases and used in comparative analyses (Table 4.1).

To identify subgroup affiliations, all 18 MGA 16S rRNA gene sequences identified on large-insert fragments from North Pacific Ocean environments were placed into the MGA reference tree described above (Figure 4.2). Seventeen out of 18 16S rRNA gene sequences identified on large-inserts grouped with 10 defined MGA subgroups. The remaining 16S rRNA gene (on fosmid 4050020-J15) appeared to group outside of MGA and was most closely

affiliated with sequences in the phylum *Deferribacteres.* This fosmid was included in downstream analyses to represent a close relative of MGA.

4.3.4 Genomic content and organization of large-insert DNA fragments derived from MGA

Four criteria were used to determine the extent to which large-insert DNA fragments partitioned with O_2 deficiency, including GC content, tetranucleotide frequency, global nucleotide similarity, and amino acid similarity of predicted open reading frames (ORFs). The size of the large-insert fragments containing MGA 16S rRNA genes ranged from 27.4 kb to 43.5 kb with a GC content ranging from 32.8% to 47.7% (Table 4.1). Pronounced phylogenetic signals have been associated with GC content in addition to the tetranucleotide usage patterns of nucleotide sequences (Teeling et al., 2004a). However, large-insert fragments affiliated with MGA did not differentiate into discrete groups based on similar GC content (Table 4.1) or tetranucleotide frequency (Figure 4.4). To further investigate potential similarities among fragments associated with nucleotide arrangement, pairwise blastn analyses were performed between all fragments (Figure 4.5). Bit scores for pairwise blastn analyses ranged between 0 and 4.5×10^4 for non-identical fragments. Large-insert fragments from Monterey Bay (EBAC750-03B02) and the NESAP (1250012-L08 and 4130011-I07), affiliated with subgroup Arctic95A-2, were most similar to one another and formed a distinct group based on global nucleotide similarity (Figure 4.5). The remaining inserts did not form distinct groups based on global nucleotide similarity, but displayed a gradient of similarity, with bit scores for pairwise blastn analyses averaging $(2.2 \pm 1.5) \times 10^3$. Fosmid 122006-I05, affiliated with subgroup P262000D03, was most unique at the nucleotide level.

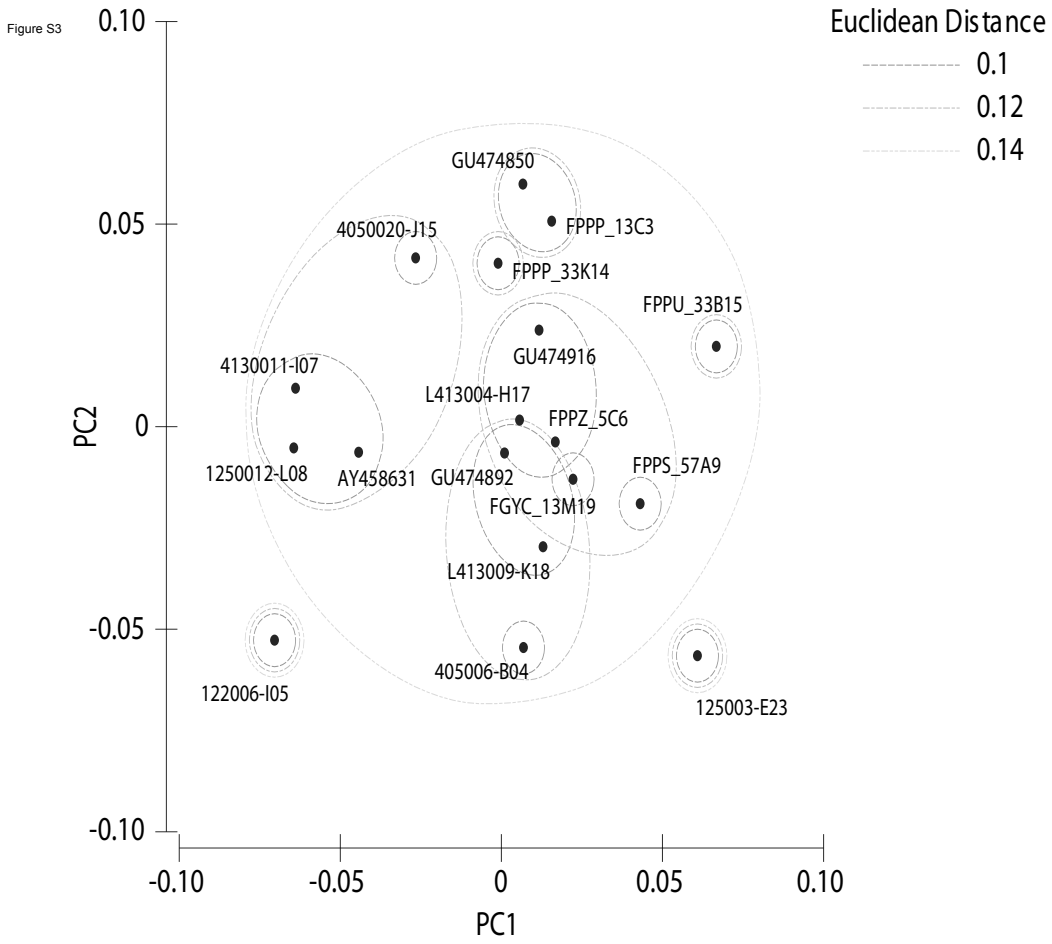


Figure 4.4 Principal component analysis performed on normalized Z-score profiles of tetranucleotide frequency for large-insert DNA fragments derived from MGA

PCA is overlaid with clusters (dotted lines) determined by hierarchical cluster analysis of normalized Z-scores using a Euclidean distance matrix.

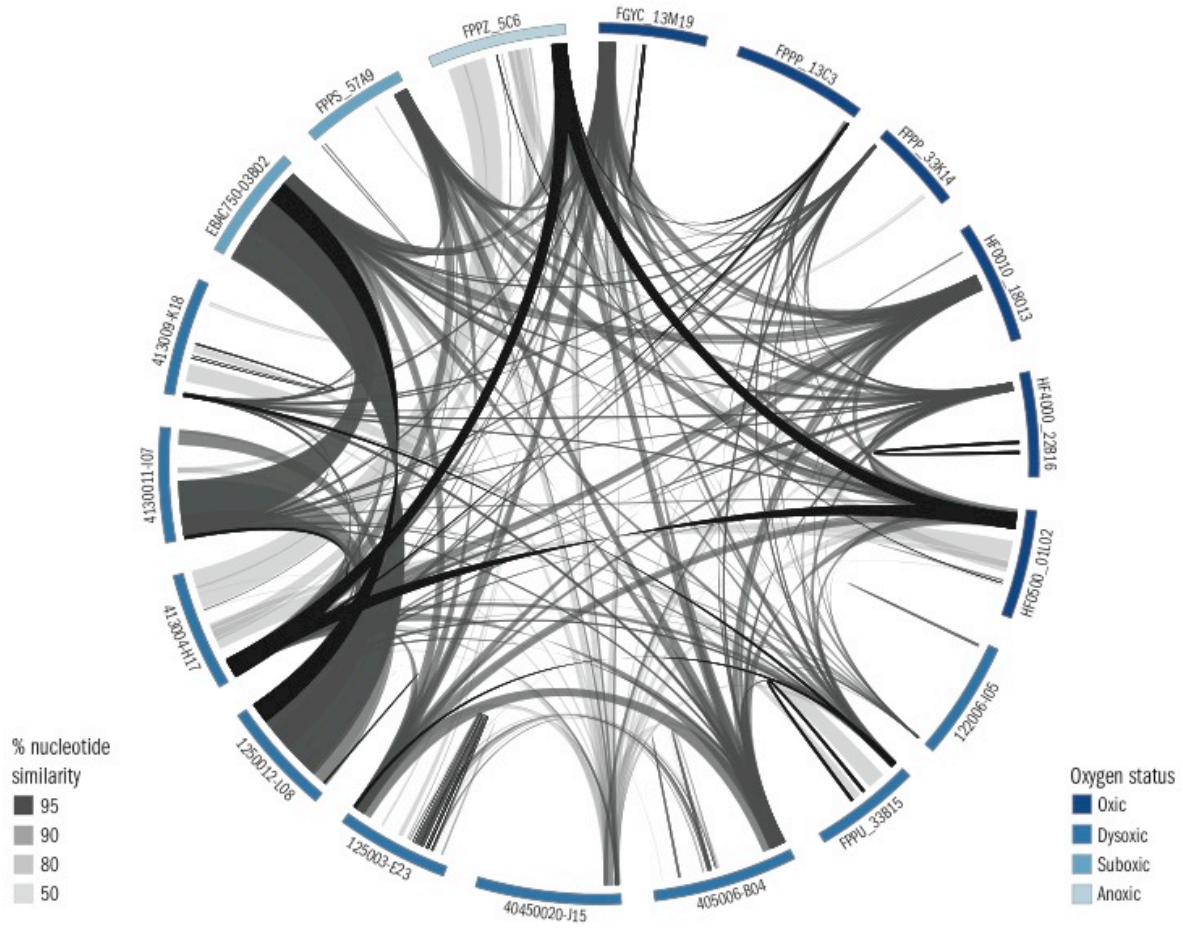


Figure 4.5 Global nucleotide similarity among 17 MGA-affiliated and 1 *Deferribacteres*-affiliated large-insert DNA fragments

Depicted at 50%, 80%, 90%, and 95% similarity.

To investigate potential similarities among large-insert fragments at the protein-coding level, ORFs were predicted and annotated (Materials and Methods). The number of predicted ORFs per insert ranged from 17 to 50, and the number of ORFs on each fragment annotated as groups with shared but not identical amino acid sequences of predicted ORFs surrounding the 16S rRNA gene were identified (groups I – IV), while the *Deferribacteres*-affiliated fosmid (4050020-J15) did not show significant similarity to any other fragments at the protein-coding level and was placed in its own group (group V) (Figure 4.6, Figure 4.7). These groups did not uniformly correlate with shared environmental origin or 16S rRNA sequence identity at the level of defined subgroups (Table 4.1, Figure 4.2). In some cases it was clear that fosmid groups represented different flanking regions of the rRNA operon (i.e. groups I and II; Figure 4.6, Figure 4.7).

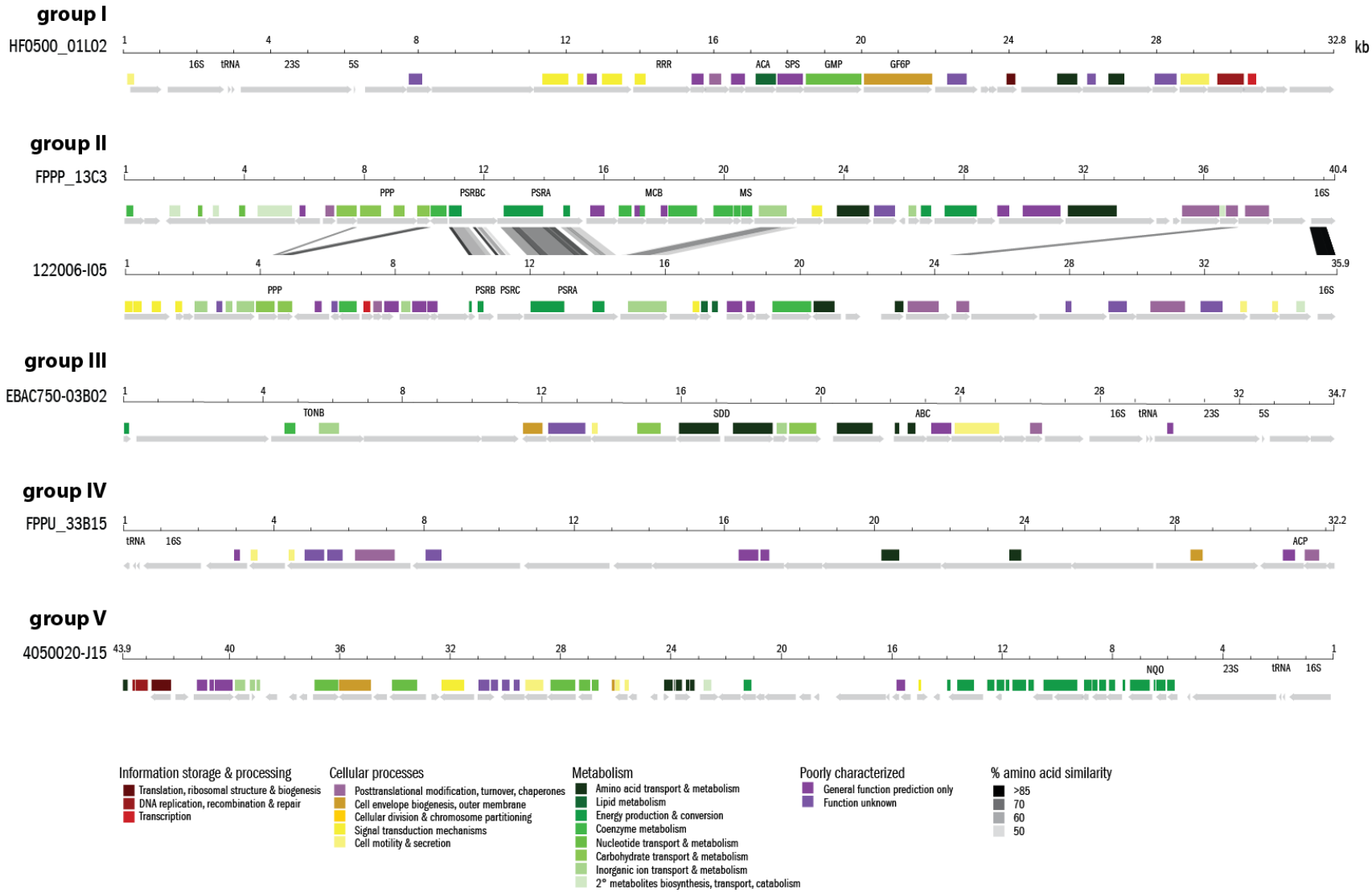

group I
HF0500_01L02
1
4
8
12
16
20
24
28
32.8
kb
16S
tRNA
23S
5S
RRR
ACA
SPS
GMP
GF6P
group II
FPPP_13C3
1
4
8
12
16
20
24
28
32
36
40.4
PPP
PSRBC
PSRA
MCB
MS
16S
122006-I05
1
4
8
12
16
20
24
28
32
35.9
PPP
PSRB
PSRC
PSRA
16S
group III
EBAC750-03B02
1
4
8
12
16
20
24
28
32
34.7
TONB
SDD
ABC
16S
tRNA
23S
5S
group IV
FPPU_33B15
1
4
8
12
16
20
24
28
32.2
tRNA
16S
ACP
group V
4050020-J15
43.9
40
36
32
28
24
20
16
12
8
4
1
NQO
23S
tRNA
16S
Information storage & processing
Translation, ribosomal structure & biogenesis
DNA replication, recombination & repair
Transcription
Cellular processes
Posttranslational modification, turnover, chaperones
Cell envelope biogenesis, outer membrane
Cellular division & chromosome partitioning
Signal transduction mechanisms
Cell motility & secretion
Metabolism
Amino acid transport & metabolism
Lipid metabolism
Energy production & conversion
Coenzyme metabolism
Nucleotide transport & metabolism
Carbohydrate transport & metabolism
Inorganic ion transport & metabolism
2° metabolites biosynthesis, transport, catabolism
Poorly characterized
General function prediction only
Function unknown
% amino acid similarity
>85
70
60
50

Figure 4.6 Genes and similarity comparison of large-insert DNA fragments containing MGA 16S rRNA genes representative of syntenic groups I - V

COG categories detected on large-insert fragments are shown in colour. **5S**, 5S rRNA; **16S**: 16S rRNA; **23S**, 23S rRNA; **ABC**, ABC-type multidrug transport system; **ACA**, acetyl-CoA carboxylase carboxyl transferase; **ACP**, ATP-dependent CLP protease; **GF6P**, glucosamine-fructose-6-phosphate aminotransferase; **GMP**, GMP synthase; **MCB**, molybdenum cofactor biosynthesis; **MS**, molybdopterin synthase; **NQO**, NADH quinone oxidoreductase; **PPP**, pentose phosphate pathway enzymes; **PSRA**, polysulfide reducase subunit A; **PSRB**, polysulfide reductase subunit B; **PSRC**, polysulfide reductase subunit C; **PSRBC**, polysulfide reductase subunit BC gene fusion; **RRR**, response regulator receiver protein; **SDD**, succinyl-diaminopimelate desuccinylase; **SPS**, stationary-phase survival protein; **TONB**, TonB dependent receptor; **tRNA**, transfer RNA.

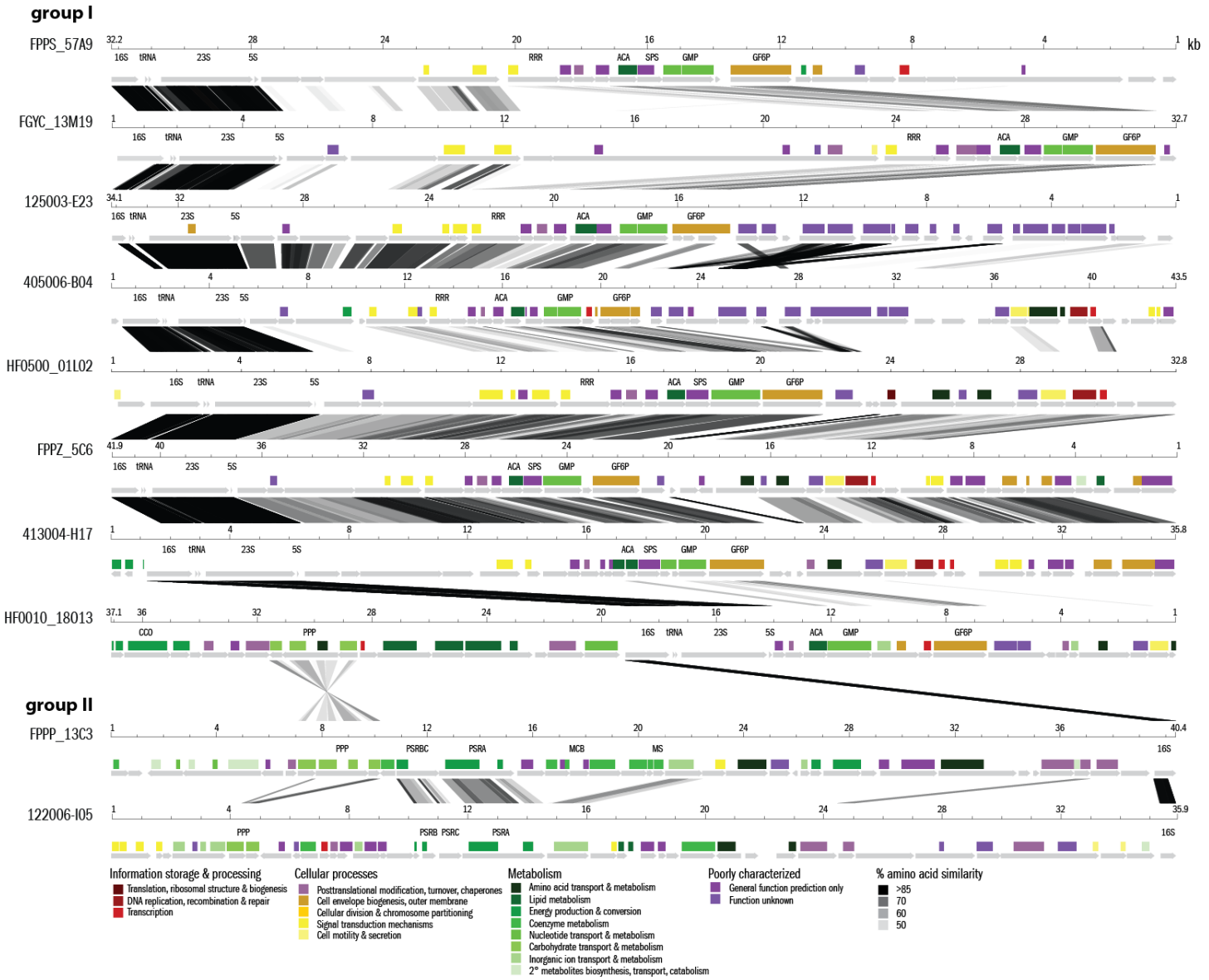

group I
FPPS_57A9
FGYC_13M19
125003-E23
405006-B04
HF0500_01L02
FPPZ_5C6
413004-H17
HF0010_18013
group II
FPPP_13C3
122006-I05
kb
16S
tRNA
23S
5S
RRR
ACA
SPS
GMP
GF6P
CCO
PPP
PSRBC
PSRA
MCB
MS
PSRB
PSRC
Information storage & processing
Translation, ribosomal structure & biogenesis
DNA replication, recombination & repair
Transcription
Cellular processes
Posttranslational modification, turnover, chaperones
Cell envelope biogenesis, outer membrane
Cellular division & chromosome partitioning
Signal transduction mechanisms
Cell motility & secretion
Metabolism
Amino acid transport & metabolism
Lipid metabolism
Energy production & conversion
Coenzyme metabolism
Nucleotide transport & metabolism
Carbohydrate transport & metabolism
Inorganic ion transport & metabolism
2° metabolites biosynthesis, transport, catabolism
Poorly characterized
General function prediction only
Function unknown
% amino acid similarity
>85
70
60
50

Figure 4.7 Genes and similarity comparison of large-insert DNA fragments in syntenic groups I and II
COG categories detected on large-insert fragments are shown in colour. **5S**, 5S rRNA; **16S**: 16S rRNA; **23S**, 23S rRNA; **ABC**, ABC-type multidrug transport system; **ACA**, acetyl-CoA carboxylase carboxyl transferase; **ACP**, ATP-dependent CLP protease; **GF6P**, glucosamine-fructose-6-phosphate aminotransferase; **GMP**, GMP synthase; **MCB**, molybdenum cofactor biosynthesis; **MS**, molybdopterin synthase; **NQO**, NADH quinone oxidoreductase; **PPP**, pentose phosphate pathway enzymes; **PSRA**, polysulfide reducase subunit A; **PSRB**, polysulfide reductase subunit B; **PSRC**, polysulfide reductase subunit C; **PSRBC**, polysulfide reductase subunit BC gene fusion; **RRR**, response regulator receiver protein; **SDD**, succinyl-diaminopimelate desuccinylase; **SPS**, stationary-phase survival protein; **TONB**, TonB dependent receptor; **tRNA**, transfer RNA.

Four out of 8 fosmids in group I were affiliated with the Arctic96B-7 subgroup, while the remaining 4 fosmids were affiliated with ZA3312c, SHBH391, SAR406, and SHAN400. Fosmids in group I contained a conserved gene cluster with genes encoding glucosamine-fructose-6-phosphate aminotransferase (involved in glucosamine biosynthesis), GMP synthase (involved in purine nucleotide biosynthesis), and acetyl-coenzyme A carboxylase carboxyl transferase subunits alpha and beta (potentially involved in fatty acid biosynthesis or CO_2 fixation) (Figure 4.6, Figure 4.7). Saanich Inlet fosmid FPPZ_5C6 also contained a gene encoding RNA polymerase sigma-70 factor (*rpoE*), known to have a role in high temperature and oxidative stress response (Hild et al., 2000). Fosmid HF0010_18O13 contained the conserved cluster of genes found in group I fosmids as well as a cluster of cytochrome c oxidase subunit genes present in Fe(II) oxidation, and 3 pentose phosphate pathway genes also found in group III fosmids (ribulose-phosphate 3-epimerase, ribose-5-phosphate isomerase b, and transkelotase).

Group II fosmids were affiliated with Arctic96B-7 and P262000D03. Both fosmids in this group (FPPP_13C3 and 122006-I05) contained a cluster of genes encoding enzymes involved in the pentose phosphate pathway of carbon metabolism, including ribulose-phosphate 3-epimerase, ribose-5-phosphate isomerase b, and in one case, transkelotase. Both fosmids also contained an operon encoding an enzyme complex related to polysulfide reductase (Psr). The operon on fosmid 122006-I05 contained three genes encoding homologues of the three Psr subunits: PsrA, a molybdopterin oxidoreductase; PsrB, a [4Fe-4S]-binding subunit; and PsrC, a membrane anchor subunit carrying the site of quinol oxidation, while the operon on fosmid FPPP_13C3 contained two genes encoding PsrA and a PsrBC fusion protein. Fosmid

FPPP_13C3 contained additional neighboring genes encoding molybdenum cofactor and molybdopterin biosynthesis proteins potentially associated with the assembly of the molybdenum and molybdopterin guanine dinucleotide-containing subunit PsrA. Fosmid FPPP_13C3 also contained a gene for glutamate synthase, often involved in nitrogen assimilation (Vanoni and Curti, 2008), and a gene for rubrerythrin, involved in oxidative stress protection in some anaerobic bacteria and archaea (deMaré et al., 1996; Sztukowska et al., 2002). Fosmid 122006-I05 contained a gene encoding a rhodanese-like protein, belonging to a superfamily of sulfur transferases (Cipollone et al., 2007), upstream of the Psr operon.

All 3 genomic inserts belonging to group III were affiliated with subgroup Arctic95A-2 and were derived from Monterey Bay and the NESAP (Table 4.1, Figure 4.6, Figure 4.8). These three fosmids also formed a discrete group based on global nucleotide similarity analysis (Figure 4.2). The main organizational feature shared by these inserts was a set of genes encoding transporters, including an ABC-type multidrug transporter, ATPase component, ABC-2 permease, and a Tonb dependent receptor. Group III inserts also contained genes encoding succinyl-diaminopimelate desuccinylase, involved in lysine biosynthesis. Monterey Bay insert EBAC750-03B02 contained a gene affiliated with methionine sulfoxide reductase (*msrB*). In *E. coli,* MsrB has been shown to have sulfoxide and dimethyl sulfoxide (DMSO) reductase activity (Grimaud et al., 2001). This insert also contained a gene encoding a rhodanese-like protein.

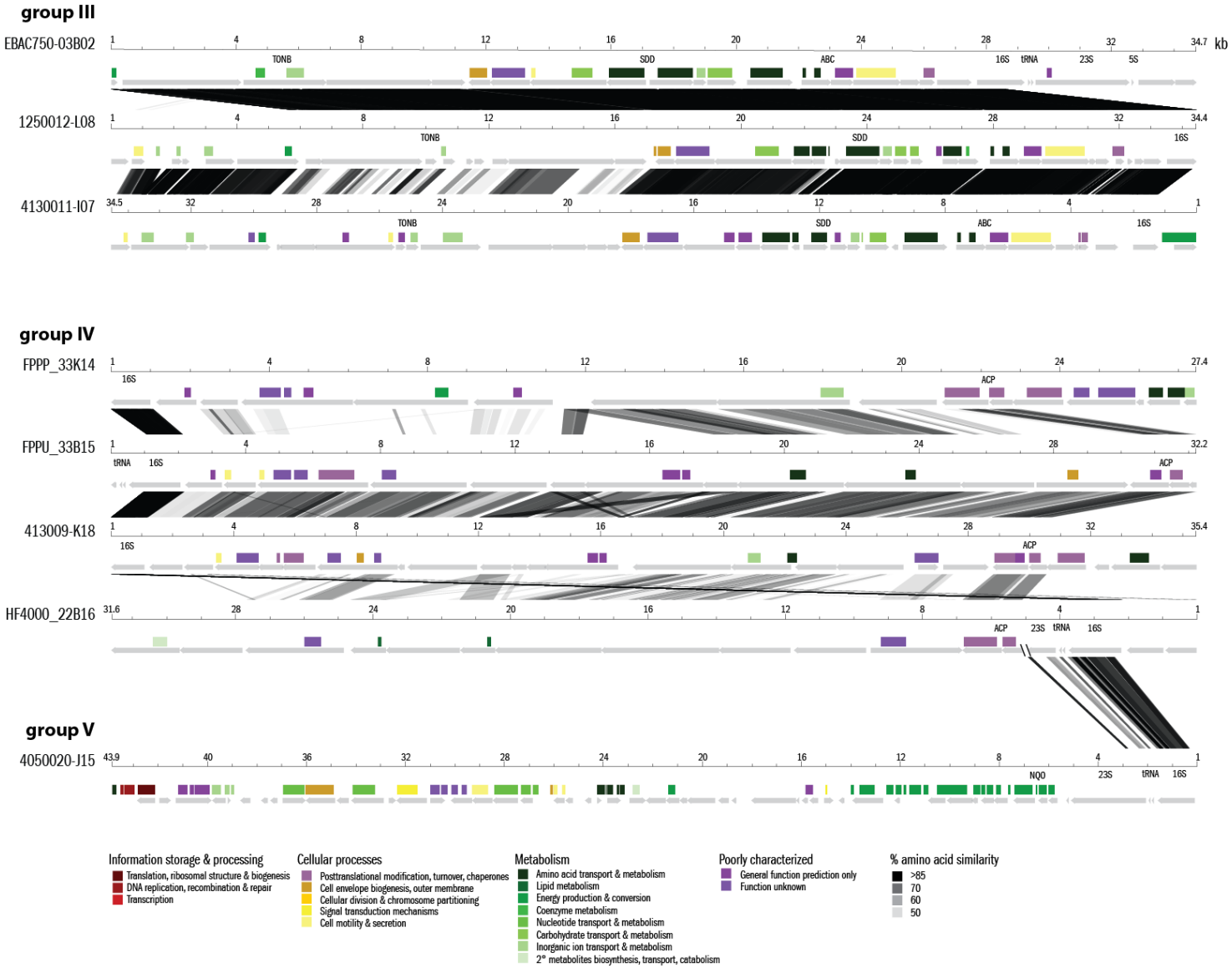

group III
EBAC750-03B02
1250012-L08
4130011-I07
TONB
SDD
ABC
16S
tRNA
23S
5S
kb
group IV
FPPP_33K14
FPPU_33B15
413009-K18
HF4000_22B16
ACP
group V
4050020-J15
NQO
Information storage & processing
Translation, ribosomal structure & biogenesis
DNA replication, recombination & repair
Transcription
Cellular processes
Posttranslational modification, turnover, chaperones
Cell envelope biogenesis, outer membrane
Cellular division & chromosome partitioning
Signal transduction mechanisms
Cell motility & secretion
Metabolism
Amino acid transport & metabolism
Lipid metabolism
Energy production & conversion
Coenzyme metabolism
Nucleotide transport & metabolism
Carbohydrate transport & metabolism
Inorganic ion transport & metabolism
2° metabolites biosynthesis, transport, catabolism
Poorly characterized
General function prediction only
Function unknown
% amino acid similarity
>85
70
60
50

Figure 4.8 Genes and similarity comparison of large-insert DNA fragments in syntenic groups III, IV, and V
COG categories detected on large-insert fragments are shown in colour. **5S**, 5S rRNA; **16S**: 16S rRNA; **23S**, 23S rRNA; **ABC**, ABC-type multidrug transport system; **ACA**, acetyl-CoA carboxylase carboxyl transferase; **ACP**, ATP-dependent CLP protease; **GF6P**, glucosamine-fructose-6-phosphate aminotransferase; **GMP**, GMP synthase; **MCB**, molybdenum cofactor biosynthesis; **MS**, molybdopterin synthase; **NQO**, NADH quinone oxidoreductase; **PPP**, pentose phosphate pathway enzymes; **PSRA**, polysulfide reducase subunit A; **PSRB**, polysulfide reductase subunit B; **PSRC**, polysulfide reductase subunit C; **PSRBC**, polysulfide reductase subunit BC gene fusion; **RRR**, response regulator receiver protein; **SDD**, succinyl-diaminopimelate desuccinylase; **SPS**, stationary-phase survival protein; **TONB**, TonB dependent receptor; **tRNA**, transfer RNA. \\ indicates the break point in fosmid HF4000_22B16 where 2 unordered contigs were connected.

Fosmids in group IV were affiliated with subgroups P262000N21, SAR406, and A714018, and primarily contained genes encoding hypothetical proteins except for 2 conserved genes encoding an ATP-dependent protease Clp ATPase subunit and protease subunit. Group IV fosmid HF4000_22B16 was assembled as 2 unordered pieces, as such, it contained a break point within the 23S rRNA gene (Figure 4.8).

The only fosmid in group V (4050020-J15; most closely related at the 16S rRNA gene sequence level to members of the phylum *Deferribacteres*) did not exhibit much protein similarity to any of the MGA-affiliated fosmids. This fosmid contained genes for NADH-ubiquinone and quinone oxidoreductase involved in energy metabolism, a major facilitator superfamily (MFS) transporter, a dihydroorotate dehydrogenase, a cell wall associated hydrolase, and a tRNA nucleotidyltransferase, in addition to genes encoding a number of hypothetical proteins.

4.3.5 Population structure of MGA syntenic groups

To determine the prevalence and distribution of MGA subgroups represented by large-insert DNA fragments detected in this study, the proportion of fosmid end sequences from each NESAP and SI library recruiting to large-insert fragments was determined (Figure 4.9). The largest proportions of sequences recruiting to large-insert fragments were derived from depths ≥500 m in the NESAP and ≥100 m in SI. A very small proportion of end sequences were recruited from Aug-09 P26 libraries, which could be due to the relatively small size of these libraries (Table 4.2). End sequences from NESAP libraries generally recruited to large-insert

fragments in larger numbers and with a higher degree of nucleotide similarity than end sequences from SI libraries, even for large-insert fragments derived from SI (Figure 4.9). End sequences similar to group III fragments were most highly and consistently represented in NESAP fosmid end libraries, followed by end sequences similar to several group I fragments. End sequences similar to the *Deferribacteres*-like fosmid 4050020-J15 were also well represented and very similar to sequences derived from oxic through suboxic (but not anoxic) NESAP and SI libraries.

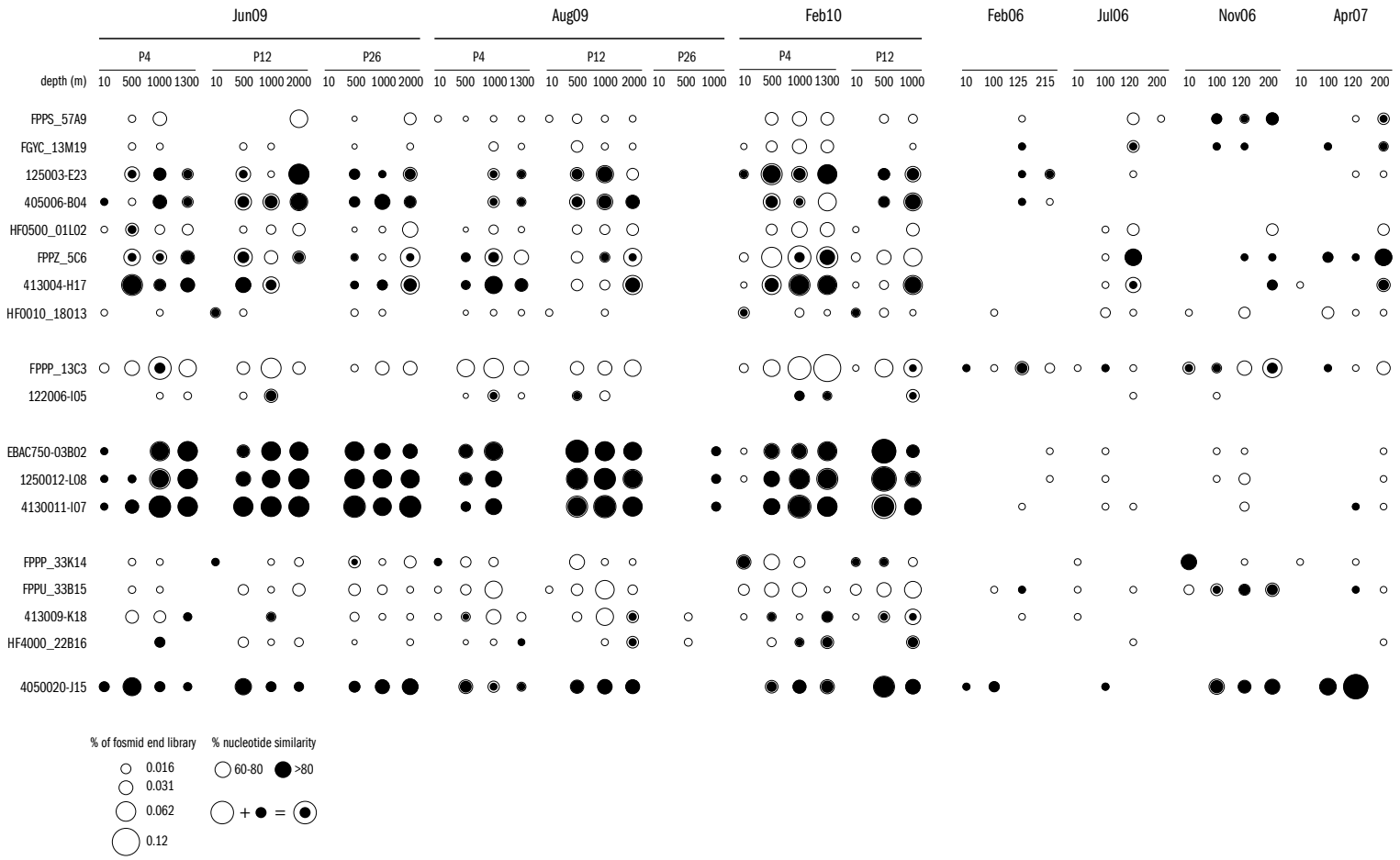
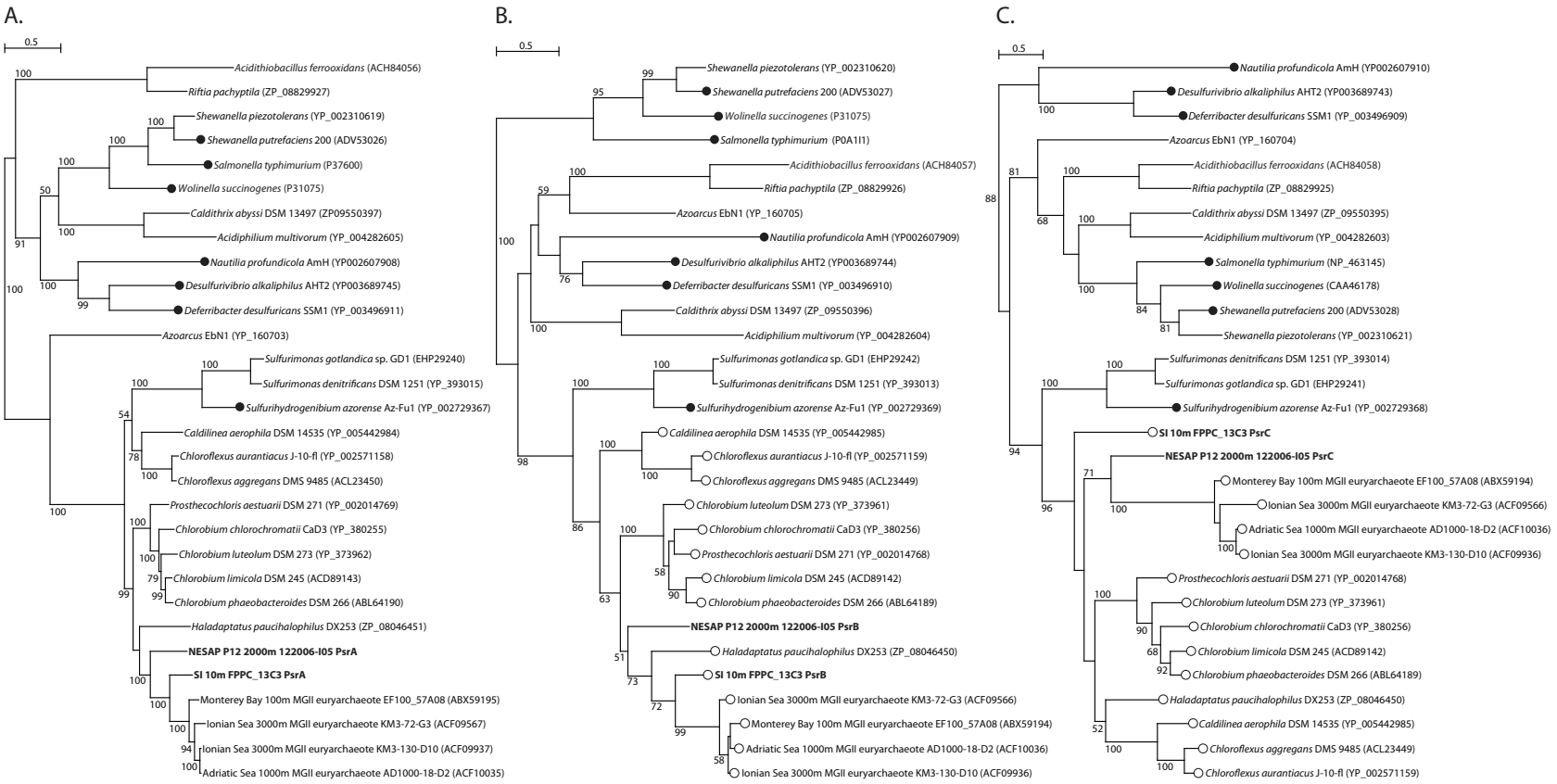

Jun09
Aug09
Feb10
Feb06
Jul06
Nov06
Apr07
P4
P12
P26
depth (m)
FPPS_57A9
FGYC_13M19
125003-E23
405006-B04
HF0500_01L02
FPPZ_5C6
413004-H17
HF0010_18O13
FPPP_13C3
122006-I05
EBAC750-03B02
1250012-L08
4130011-I07
FPPP_33K14
FPPU_33B15
413009-K18
HF4000_22B16
4050020-J15
% of fosmid end library
0.016
0.031
0.062
0.12
% nucleotide similarity
60-80
>80

Figure 4.9 Dot plot showing the proportion of fosmid end sequenced libraries recruiting to MGA large-insert DNA fragments at various sample points and depths

Depicted at NESAP stations P4, P12 and P26 and SI station S3. Hollow circles represent percent of fosmid end sequenced libraries recruiting to large-insert fragments with nucleotide similarity 60%-80%; solid circles >80%.

4.3.6 Phylogenetic analysis and distribution of polysulfide reductase

The identification of proteins homologous to polysulfide reductase (Psr) on two fosmids suggested that specific MGA subgroups have the capacity to use polysulfide as a terminal electron acceptor in the process of dissimilatory polysulfide reduction to hydrogren sulfide (H_2S) (Schröder et al., 1988; Klimmek et al., 1991; Krafft et al., 1992; Krafft et al., 1995; Jormakka et al., 2008). The Psr complex, which has only been thoroughly characterized in the anaerobic epsilonproteobacterium *Wolinella succinogenes*, is encoded by the *psrABC* genes and consists of two periplasmic subunits (a catalytic molybdopterin-containing PsrA subunit and a [4Fe-4S]-binding PsrB subunit) and a membrane-anchoring PsrC subunit (Krafft et al., 1992). The Psr holoenzyme is a membrane-associated protein that, in complex with its quinone cofactor, reduces polysulfide (S_n) to H_2S, an energy-yielding process that is common in extreme environments such as deep-sea vents and hot springs (Jormakka et al., 2008). In contrast to the organization of the *psrABC* operon originally described in *W. succinogenes* (Krafft et al., 1995), the ORFs encoding PsrB and PsrC homologues on both MGA fosmids were located upstream of ORFs encoding PsrA homologues (Figure 4.6). Also in contrast to the *W. succinogenes psrABC* operon, the genes encoding PsrB and PsrC on fosmid FPPP_13C3 appeared to form a gene fusion (*psrBC*), a feature also detected in several PSR-containing GSB (Frigaard and Bryant, 2008) and other PSR-containing bacteria and archaea.



● Capable of dissimilatory sulfur or polysulfide reduction

○ Encoded by *psrBC* gene fusion

Figure 4.10 Unrooted phylogenetic trees based on protein sequences with homology to Psr protein subunits (a) predicted polysulfide reductase molybdopterin-containing subunit (PsrA); (b) predicted [4Fe-4S]-binding subunit (PsrB); and (c) membrane anchor subunit (PsrC) identified on fosmids FPPP_13C3 and 122006-I05. The trees were inferred using maximum likelihood implemented in PhyML. Solid circle indicates proteins derived from organisms that have been demonstrated to grow by reducing elemental sulfur or polysulfide with concomitant H_2S production; hollow circle indicates presence of a *psrBC* gene fusion. The scale bar represents estimated number of amino acid substitutions per site. Bootstrap values below 50% are not shown.

To gain insight into the evolutionary history of *psrA*, *psrB* and *psrC* genes detected on MGA fosmids, phylogenetic trees of their predicted protein products were constructed (Figure 4.10). Phylogenetic analysis of the catalytic subunit, PsrA, confirmed that predicted PsrA homologues detected on MGA fosmids were most closely related to Psr and thiosulfate reductase (Phs) of the DMSO reductase family of molybdenum containing enzymes (Figure 4.11). Predicted PsrA homologues from MGA fosmids were ~63% similar to one another, and most closely related to proteins encoded on fosmids from the Mediterranean Sea and Monterey Bay derived from Marine Group II (MGII) euryarchaeota (Figure 4.10a). Predicted MGA proteins were less similar to canonical PsrA proteins originally characterized in *W. succinogenes* (Krafft et al., 1995). Phylogenetic trees of predicted PsrB with PsrB-like respiratory proteins containing [4Fe-4S]-binding-subunits and of predicted PsrC with PsrC-like membrane anchor subunits indicated similar phylogenetic relationships (Figures 4.10b, 4.10c). The *psrBCA* format of operon organization detected on MGA fosmids was also detected on MGII fosmids and several PSR-containing green sulfur bacteria (GSB) in addition to *Sulfurimonas denitrificans* DSM 1251, *Caldilinea aerophila* DSM 14535, *Chloroflexus aggregans* DSM 9485, and *Haladaptatus paucihalophilus* DX253 (Figures 4.10b, 4.10c). A third format of operon organization (*psrACB*) was detected in *Sulfurimonas gotlandica* GD1 and *Sulfurihydrogenibium azorense* Az-Fu1.

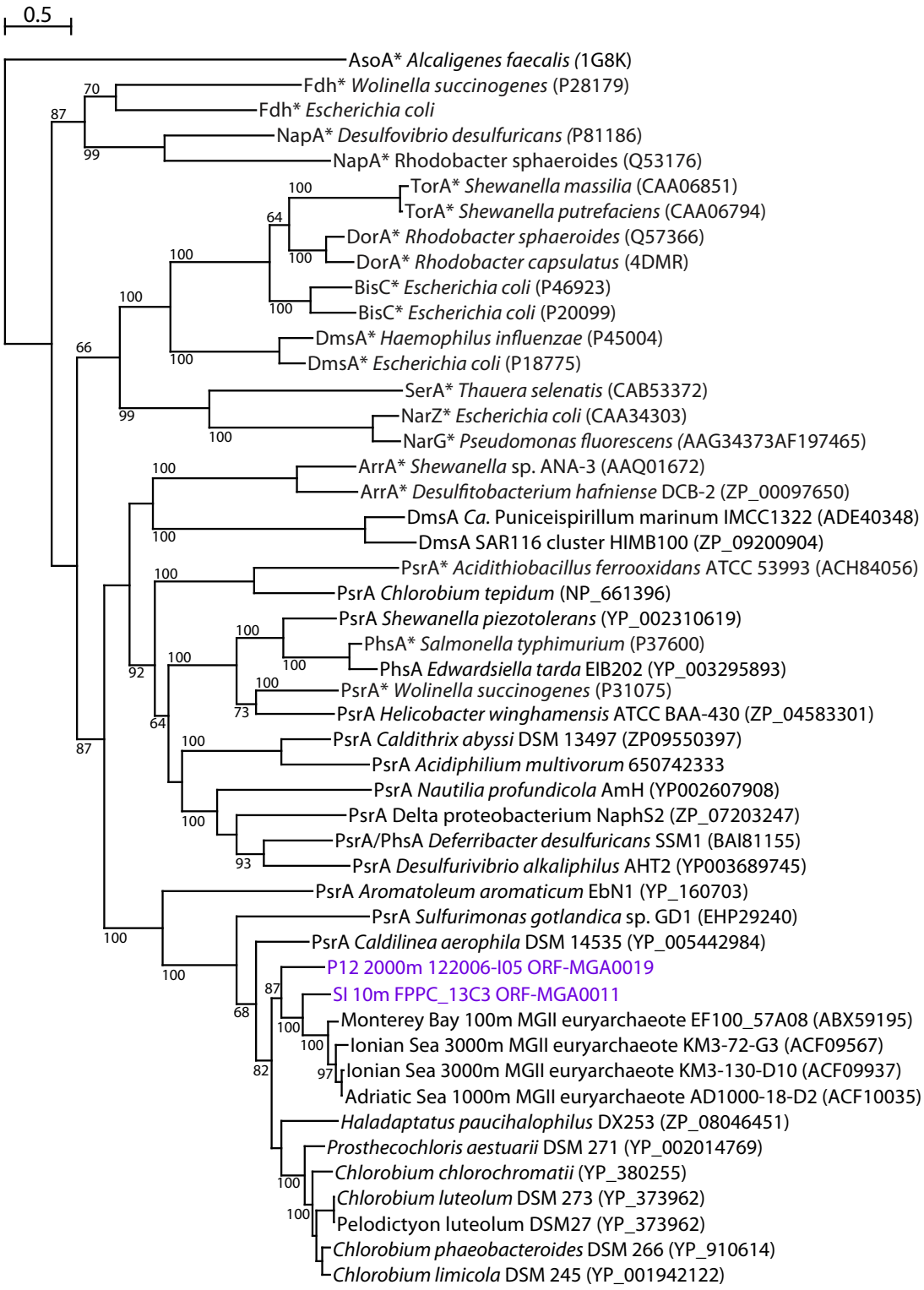

0.5
AsoA* Alcaligenes faecalis (1G8K)
Fdh* Wolinella succinogenes (P28179)
Fdh* Escherichia coli
NapA* Desulfovibrio desulfuricans (P81186)
NapA* Rhodobacter sphaeroides (Q53176)
TorA* Shewanella massilia (CAA06851)
TorA* Shewanella putrefaciens (CAA06794)
DorA* Rhodobacter sphaeroides (Q57366)
DorA* Rhodobacter capsulatus (4DMR)
BisC* Escherichia coli (P46923)
BisC* Escherichia coli (P20099)
DmsA* Haemophilus influenzae (P45004)
DmsA* Escherichia coli (P18775)
SerA* Thauera selenatis (CAB53372)
NarZ* Escherichia coli (CAA34303)
NarG* Pseudomonas fluorescens (AAG34373AF197465)
ArrA* Shewanella sp. ANA-3 (AAQ01672)
ArrA* Desulfitobacterium hafniense DCB-2 (ZP_00097650)
DmsA Ca. Puniceispirillum marinum IMCC1322 (ADE40348)
DmsA SAR116 cluster HIMB100 (ZP_09200904)
PsrA* Acidithiobacillus ferrooxidans ATCC 53993 (ACH84056)
PsrA Chlorobium tepidum (NP_661396)
PsrA Shewanella piezotolerans (YP_002310619)
PhsA* Salmonella typhimurium (P37600)
PhsA Edwardsiella tarda EIB202 (YP_003295893)
PsrA* Wolinella succinogenes (P31075)
PsrA Helicobacter winghamensis ATCC BAA-430 (ZP_04583301)
PsrA Caldithrix abyssi DSM 13497 (ZP09550397)
PsrA Acidiphilium multivorum 650742333
PsrA Nautilia profundicola AmH (YP002607908)
PsrA Delta proteobacterium NaphS2 (ZP_07203247)
PsrA/PhsA Deferribacter desulfuricans SSM1 (BAI81155)
PsrA Desulfurivibrio alkaliphilus AHT2 (YP003689745)
PsrA Aromatoleum aromaticum EbN1 (YP_160703)
PsrA Sulfurimonas gotlandica sp. GD1 (EHP29240)
PsrA Caldilinea aerophila DSM 14535 (YP_005442984)
P12 2000m 122006-I05 ORF-MGA0019
SI 10m FPPC_13C3 ORF-MGA0011
Monterey Bay 100m MGII euryarchaeote EF100_57A08 (ABX59195)
Ionian Sea 3000m MGII euryarchaeote KM3-72-G3 (ACF09567)
Ionian Sea 3000m MGII euryarchaeote KM3-130-D10 (ACF09937)
Adriatic Sea 1000m MGII euryarchaeote AD1000-18-D2 (ACF10035)
Haladaptatus paucihalophilus DX253 (ZP_08046451)
Prosthecochloris aestuarii DSM 271 (YP_002014769)
Chlorobium chlorochromatii (YP_380255)
Chlorobium luteolum DSM 273 (YP_373962)
Pelodictyon luteolum DSM27 (YP_373962)
Chlorobium phaeobacteroides DSM 266 (YP_910614)
Chlorobium limicola DSM 245 (YP_001942122)

Figure 4.11 Unrooted phylogenetic trees based on DMSO-reductase family protein sequences with homology to PsrA identified on fosmids FPPP_13C3 and 122006-I05

The trees were inferred using maximum likelihood implemented in PhyML. '*' indicates enzymes with experimentally supported activity. The bar represents estimated amino acid substitutions per site. Psr/Phs, polysulfide/thiosulfate reductase; Nrf, nitrite reductase; Tor, trimethylamineoxide reductase; Bis, biotinsulfoxide reductase; Dms/Dor, DMSO reductase; Ser, selenate reductase; Nar, membrane-associated nitrate reductase; Aso, arsenite oxidase; Fdh, formate dehydrogenase; Nap, periplasmic nitrate reductase; Arr, arsenate respiratory reductase; Frd, fumarate reductase.

To determine the prevalence of predicted MGA *psr* genes in NESAP and SI fosmid end sequenced libraries, the proportion of fosmid end sequences that recruited to the *psrBCA* operon on fosmids FPPP_13C3 and 122006-I05 was calculated for each end sequenced library (Figure 4.12). The majority of end sequences recruiting to *psr* genes were derived from ≥500 m depth in the NESAP and ≥100 m depth in SI, and *psr* homologues were most consistently present throughout O_2-deficient waters of the NESAP in August 2009 at station P4.

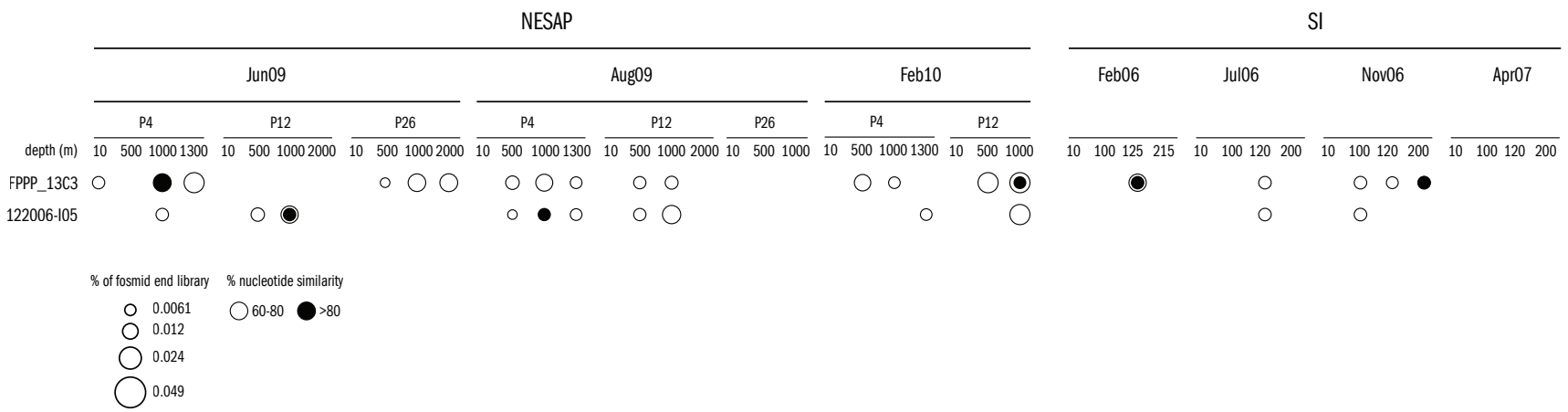


Figure 4.12 Dot plot showing the proportion of fosmid end sequenced libraries recruiting to psr genes on fosmids FPPP_13C3 and 122006-I05

Depicted at various sample points and depths in the NESAP at stations P4, P12 and P26 and SI station S3. Solid circles represent proportion of fosmid end sequenced libraries recruiting to large-insert fragments with nucleotide similarity 60%-80%; hollow circles >80%.

4.4 Discussion

Five out of 10 previously defined MGA subgroups were recovered in SI clone libraries, and 3 novel subgroups were identified (SHBH1141, SHBH391, and SHAN400) (Figure 4.2). The novel subgroups were exclusively found in SI and were most abundant in clone libraries generated from suboxic (1-20 μmol kg^{-1} O_2) or anoxic samples (Figure 4.3, Table 4.2). Shared subgroups ZA3312c, Arctic96B-7, and SAR406 were abundant in SI and NESAP clone libraries from dysoxic samples (20-90 μmol kg^{-1} O_2), while subgroups HF770D10, P262000D03, P41300E03, P262000N21 and Arctic95A-2 were solely detected in NESAP clone libraries generated from dysoxic and suboxic samples. Subgroup Arctic95A-2 was the most prevalent MGA subgroup detected in NESAP clone libraries, consistent with results obtained from pyrotag libraries generated from these waters (Allers et al., 2012). These observations are consistent with previous reports indicating that MGA subgroups partition along O_2 gradients within coastal and open ocean OMZs (Allers et al., 2012), and extends the range of MGA subgroup partitioning to include anoxic-sulfidic water column conditions.

The 17 large-insert DNA fragments containing MGA 16S rRNA genes derived from North Pacific Ocean metagenomic libraries were affiliated with 7 previously defined and 2 novel MGA subgroups, while the 16S rRNA gene on an 18th insert was more closely related to the phylum *Deferribacteres* (Figure 4.2). Although large-insert DNA fragments were obtained from multiple environments manifesting distinct oxyclines, fragments did not coalesce into coherent groups based on GC content, tetranucleotide frequency, or global nucleotide similarity. However, fragments did coalesce into 5 syntenic groups based on shared amino acid similarity of predicted ORFs. Group membership was not generally consistent with shared environmental origin, O_2 concentration, or 16S rRNA gene sequence identity (Table 4.1). These observations could be explained in several ways. MGA subgroups may contain multiple unlinked copies of the 16S rRNA operon (Acinas et al., 2004). Alternatively, large-insert fragments may be derived from flanking regions of the same 16S rRNA operon, as observed for syntenic groups I and II. It is also possible that subgroups ZA3312c through A714018 actually represent one subgroup of MGA, evidenced by a lack of bootstrap support for nodes encompassing these subgroups within the MGA 16S rRNA gene tree (Figure 4.2).

Recruitment of fosmid end sequences from NESAP and SI libraries to large-insert DNA fragments reflected 16S rRNA-based patterns of MGA distribution in that the proportion of MGA sequences was maximal in waters ≥500 m depth in the NESAP and ≥100 m depth in SI (Figure 4.9). MGA sequences comprised a much larger proportion of NESAP (open ocean) than SI (coastal basin) end libraries, a pattern also reflected in MGA 16S rRNA distribution. The proportion of SI end sequences recruiting to large-insert fragments was maximal in dysoxic and suboxic samples from Nov 2006 and April 2007, supporting the hypothesis that dominant MGA subgroups are adapted to O_2-deficiency in this location. The largest proportion of end sequences from NESAP libraries recruited to group III fragments affiliated with subgroup Arctic95A-2 supporting 16S rRNA-based observations that Arctic95A-2 is a dominant subgroup in the NESAP open-ocean. Group I fragments affiliated with subgroups Arctic96B-7 and SAR406 also recruited a relatively large proportion of NESAP end sequences. A reasonable proportion of end sequences from NESAP and SI libraries also recruited to *Deferribacteres*-like fosmid 4050020-J15, with a pattern of distribution suggesting adaptation to suboxic and dysoxic, but not anoxic, conditions.

Although large-insert fragments did not clearly partition into ecologically distinct groups based on O_2 concentration, predicted protein-coding genes associated with adaptation to O_2-deficiency and sulfur-based energy metabolism were detected on multiple fosmids. With respect to adaptation to O_2-deficiency, a gene encoding *rpoE* RNA polymerase sigma-70 factor, known to have a role in oxidative stress response, was detected on SI fosmid FPPZ_5C6, obtained from an anoxic-sulfidic 200 m sample. A gene encoding rubrerythrin, also involved in oxidative stress response in some anaerobic prokaryotes, was detected on SI fosmid FPPP_13C3, obtained from an oxic 10 m sample. With respect to sulfur-based energy metabolism, a polysulfide reductase operon was detected on SI fosmid FPPP_13C3 and on NESAP fosmid 122006-I05, obtained from a dysoxic 2000 m sample. In *Wolinella succinogenes*, PSR and hydrogenase (HYD) or formate dehydrogenase (FDH) allows respiration on polysulfide (S_n) using H_2 or formate as an electron donor, with concomitant production of H_2S (Jankielewicz et al., 1995). The PSR complex isolated from *W. succinogenes* has also been documented to catalyze sulfide oxidation to polysulfide by dimethylnaphthoquinone, however with much lower efficiency (Hedderich et al., 1999).

Predicted PsrA proteins detected on MGA fosmids were only distantly related to isolated PsrA from *W. succinogenes*, but more closely related to PsrA homologues encoded on MGII euryarchaeotal fosmids derived from the Mediterranean Sea and Monterey Bay. PsrA proteins detected on MGA fosmids were also similar to PsrA homologues found in the green sulfur bacteria (GSB) *Prostheticochloris aestuarii* DSM 271, *Chlorobium chlorochromatii* CaD3*, Chlorobium luteolum* DSM273, *Chlorobium limicola* DSM 245, and *Chlorobium phaeobacteroides* DSM 266, the halophilic euryarchaeon *Haladaptatus paucihalophilus* DX253, the thermophilic *Chloroflexi* strain *Caldilinea aerophila* DSM 14535, the thermophilic *Aquificales* strain *Sulfurihydrogenibium azorense* Az-Fu1, and the sulfur oxidizing Epsilonproteobacteria *Sulfurimonas gotlandica* GD1 and *Sulfurimonas denitrificans* DSM 1251. Interestingly, in GSB, the phylogeny of PsrA homologues is congruent with a number of phylogenetic anchor genes, suggesting that PSR was present in the last common ancestor of PSR-containing GSB (Gregersen et al., 2011). Given the proximal phylogenetic relationship of MGA and GSB based on 16S rRNA gene sequences (Figure 4.2), it is possible that MGA inherited this operon from a common ancestor. The *psrBC* genes on MGA fosmid 122006-I05 were encoded by separate ORFs (*psrB* and *psrC*), while in fosmid FPPP_13C3, these genes were fused (*psrBC*). A *psrBC* gene fusion has been described previously in members of the PSR-containing GSB (including *P. aestuarii, C. chlorochromatii, and C. luteolum*; (Frigaard and Bryant, 2008)), and was detected in MGII fosmids from the Mediterranean Sea and Monterey Bay in addition to *H. paucihalophilus* and *C. aerophila*. The broad phylogenetic origins of *psrABC* genes similar to those detected on MGA fosmids are consistent with multiple lateral transfer events across phyla and domains.

Although direct evidence for the role of PSR in sulfur-based energy metabolism has only been obtained from *W. succinogenes*, many cultivated reference strains encoding PSR are capable of generating energy using sulfur compounds. The PSR sequences derived from several such reference strains, including *S. azorense* Az-Fu1 and the GSB, branched with predicted PSR homologues detected on MGA fosmids. *S. azorense* Az-Fu1 is capable of growth by coupling reduction of elemental sulfur (S°) to hydrogen oxidation, although polysulfide was not directly tested as an electron acceptor (Aguiar et al., 2004). *S. azorense* Az-Fu1 has also been documented to oxidize S° and sulfite (SO_2^{3-}) (Aguiar et al., 2004). Similarly, the PSR complex

found in many GSB (including *P. aestuarii, C. chlorochromatii, C. luteolum, C. limicola, and C. phaeobacteroides*) has been proposed to oxidize sulfite produced by the dissimilatory sulfate reduction (Dsr) system (Gregersen et al., 2011). While the actual substrate of PSR cannot be determined based on sequence similarity alone, the phylogenetic position of MGA PSR homologues provides a circumstantial link between MGA and sulfur cycling in the environment.

Oxygen-deficient marine systems, including OMZs and permanent or seasonally stratified anoxic basins, are known to harbor active sulfur cycles that have been linked to the activities of sulfur oxidizing gamma and epsilon-proteobacteria (Walsh et al., 2009; Canfield et al., 2010; Grote et al., 2012). The presence of PSR homologues on MGA affiliated genome fragments suggests a potential role for MGA in the cryptic sulfur cycle of O_2-deficient marine systems where the abundance of these bacteria seems to be concentrated. Process rate measurements linking sulfur chemistry with MGA activity are required to support this hypothesis (Milucka et al., 2012). Given the lack of cultivated representatives of MGA, the application of single-cell genomics could aid in providing the genome-wide information needed to fully describe the metabolic capacity of defined MGA subgroups residing in distinct locations (Stepanauskas, 2012; Swan et al., 2011; Woyke et al., 2009). Such high-resolution genomic data may provide additional clues as to the evolutionary history and biogeochemical roles of these widely distributed marine bacteria.

4.5 Conclusions

Phylogenetic analysis of small subunit ribosomal RNA (16S rRNA) gene clone libraries recovered five previously described MGA subgroups and defined three novel subgroups (SHBH1141, SHBH391, and SHAN400) in SI. Seventeen large-insert DNA fragments containing MGA 16S rRNA genes derived from North Pacific Ocean metagenomic libraries were affiliated with 9 out of 13 MGA subgroups. Large-insert DNA fragments did not partition into discrete groups based on similar GC content, tetranucleotide frequency, or global nucleotide similarity, however fragments did coalesce into 5 syntenic groups based on shared amino acid similarity of predicted open reading frames (ORFs). Predicted protein-coding genes associated with adaptation to O_2-deficiency and sulfur-based energy metabolism were detected on multiple fosmids. Of particular interest was an operon encoding polysulfide reductase (PSR), detected on

two fosmids derived from the NESAP and SI. The PSR complex has been implicated in dissimilatory polysulfide reduction to hydrogen sulfide and dissimilatory sulfur oxidation. These results posit a potential role for specific MGA subgroups in the marine sulfur cycle.

Chapter 5: Concluding Chapter

In this dissertation, I presented a thorough comparison of bacterial and archaeal community structure documented in oxygen minimum zones (OMZs) around the world, in addition to reviewing what is known about the involvement of specific taxa in relevant biogeochemical processes occurring within OMZs (Chapter 1). I described microbial community structures present within surface and mesopelagic regions of a previously understudied OMZ located in the Northeast subarctic Pacific Ocean (NESAP) using 454-pyrotag sequencing data, and I described co-occurrence patterns between microbial groups, highlighting patterns between Marine Group A (MGA) bacteria and other microorganisms that could represent important ecological interactions occurring in this region (Chapter 2). I focused on further characterizing the diversity, distribution, and population structure of the little-known MGA candidate phylum in the NESAP using a combination of complementary methods for assessing microbial diversity (SSU rRNA gene sequencing of clone libraries and pyrotags, and CARD-FISH) (Chapter 3). Finally, I aimed to uncover clues regarding the metabolic potential of MGA by examining large-insert genomic DNA fragments derived from MGA bacteria (Chapter 4).

5.1 Microbial community structure in the Northeast subarctic Pacific Ocean

Chapter 2 presented the first deep-sequencing survey of microbial community structure in the NESAP. From a taxonomic perspective, dominant microbial groups in surface waters included the SAR11 cluster of Alphaproteobacteria, Marine Group II Euryarchaeota of the class Thermoplasmata, Cyanobacteria of the genus *Synechococcus*, and Haptophytes of the genus *Phaeocystis*. Dominant microbial groups in mesopelagic waters included Marine Group I (MGI) Thaumarchaeota, the candidate phylum Marine Group A (MGA), Marine Group II (MGII) Euryarchaeota, and several additional Proteobacterial subgroups that have been documented to be abundant in other OMZs (SAR11, Deltaproteobacterial SAR324 and *Nitrospina*, and Gammaproteobacterial ZD0405 and ZD0417). Hierarchical cluster analysis of microbial community profiles indicated that surface communities were highly distinct between August 2007 and February 2010, a pattern that was reflected albeit to a lesser extent in mesopelagic communities sampled at these time points. In surface and mesopelagic waters, rare microbial

groups (present at frequencies <0.1% across all samples) displayed a more distinct biogeography than abundant (present at frequencies >1%) microbial groups, reinforcing the conclusion that rare microbes are not cosmopolitan in distribution and are subject to selective mechanisms. Further supporting the conclusion that rare microbes are biogeographically partitioned, the majority of operational taxonomic units (OTUs) determined to be indicative of specific clusters of samples (based on Indicator Species Analysis) belonged to the rare frequency class. The number of significant indicator OTUs was greater in February 2010 than in August 2007, which may be a result of increased depth of sequencing achieved for February vs. August samples. Indicator OTUs affiliated with MGA and Alpha- and Gammaproteobacteria were identified in all indicator groups of samples, suggesting that these taxa contain a broad diversity of subgroups that are adapted to different water column conditions.

The microbial co-occurrence network generated in Chapter 2 appeared to be highly modular, indicating the presence of distinct clusters of co-occurring OTUs at distinct locations and times within the NESAP, supporting results of the hierarchical cluster and indicator species analyses. The network also followed a scale-free distribution with a relatively low degree exponent, indicating that the network topology was shifted towards the presence of several highly connected hubs with many low-degree nodes. The most prevalent microbial group detected in the network was MGA (22% of all nodes), and MGA nodes were involved in 37% of all correlations depicted in the network. MGA nodes were most frequently connected to other MGA nodes, suggesting that intra-phylum interactions may play a role in governing microbial processes in the NESAP.

5.1.1 Limitations

The original goal of this chapter was to assess the spatiotemporal dynamics in microbial community structure in the NESAP across a 4-year times-series of data (August 2007 – February 2010), including 110 total samples taken at 8 sampling time points. Unfortunately, there were problems with the pyrotag primers used by the sequencing facility (the Department of Energy's Joint Genome Institute) to amplify the 2008 and 2009 samples, whereby the primers were heavily biased against archaeal sequences, making these datasets inappropriate to use for assessing whole community structure across all 3 domains. As such, the utility of this study in

assessing spatiotemporal dynamics over a consecutive set of time points was diminished, and I was only able to compare August 2007 and February 2010 as two discrete points in time. In addition, the fact that February 2010 sampling took place during a strong El Niño event likely decreases the value of these data as being representative of an average winter state along the Line P transect. It is also important to note that pyrotag data is only semi-quantitative, and proportions of tags affiliated with different groups of organisms should most likely not be directly compared across independent studies or interpreted as truly quantitative assessments of microbial community composition.

At present, there is no single taxonomic database that performs optimally for assignment of sequences affiliated from all 3 domains of life. The Greengenes database (DeSantis et al., 2006a) provides relatively high resolution taxonomic information for classification of bacterial and archaeal sequences, however, does not include eukaryotic sequences. The SILVA database (Pruesse et al., 2007) used to taxonomically assign pyrotag sequences in this chapter was chosen because it is the largest public database that includes reference sequence from all 3 domains, however, the level of resolution for taxonomic assignment is low, particularly for environmentally-derived sequences (e.g. most eukaryotic sequences captured in this study were assigned as "environmental eukaryote" or "uncultured eukaryote"). Future studies will be aided by the development of more detailed taxonomic databases that will enable more accurate assessments of community structure and interpretation of the ecological relevance of microbial co-occurrence networks.

At present, we are in the early stages of learning to quantify, compare, and interpret microbial co-occurrence networks. As such, there is currently a lack of bioinformatic tools available to analyse these networks. For example, a tool that could be used to quantify connections between nodes at shifting levels of taxonomy would be incredibly useful in helping to narrow in on the level of taxonomy that might be relevant for assessing patterns in correlations (for example, quantifying all connections with all *Proteobacteria*, then all *Alphaproteobacteria*, then all SAR11 and being able to easily zoom in and out on connections determined at these varying levels of taxonomy to identify patterns). In the absence of such tools that would enable identification of co-occurrence patterns among microbial groups at all levels of taxonomy, I focused on defining patterns involving the dominant bacterial group MGA, with the

understanding that this is certainly not the only important group of microbes in the NESAP and that further characterization of co-occurrence patterns involving other groups is necessary.

As discussed in Chapter 2, it is important to note that correlations observed in co-occurrence networks do not distinguish between true ecological interactions (e.g. syntrophy) and other non-random processes (e.g. niche overlap). However, the identification of strong correlation patterns can provide a source of new hypotheses regarding interactions that may be important in particular environments.

5.1.2 Future directions

In order to obtain a more detailed understanding of microbial community dynamics in the NESAP, community structure studies including more consecutive time points are needed. This will help determine whether patterns in community structure are stable throughout a given year or other time step, seasonally reoccurring, or continuously morphing from one time point to the next. Combining assessments of microbial community structure with measurements of export flux (for example, using sediment traps) would be very helpful in testing the hypothesis that quantity and quality of export production affects microbial community structure in the NESAP. In addition, analyzing samples from a higher resolution of depths throughout the euphotic zone into the oxycline and OMZ would assist in developing a more detailed understanding of vertical partitioning of microbial communities in the water column. Application of algorithms to mathematically derive clusters of nodes in the network will assist in determining whether fine-scale clusters of nodes mapped onto the network using the results of Indicator Species Analysis actually represent cohesive modules in the network (Newman and Banfield, 2002; Newman, 2004). Once derived, global properties and node properties of these modules can be calculated and compared to obtain a more thorough understanding of the similarities and differences in network architecture at different times and depths in the NESAP (Guimera and Amaral, 2005). Analysis and comparison of metabolic genes and pathways present in metagenomic datasets derived from different times and depths in the NESAP water column will aid in determining whether variability in microbial community structure is associated with distinct patterns of community metabolism (functionally unique), or whether metabolic capacity of communities is the same over time (functionally redundant).

The development of bioinformatic tools for quantifying and interpreting microbial co-occurrence networks will greatly improve the ability of these analyses to inform our understanding of microbial community interactions and to generate testable hypotheses regarding specific interactions between or among groups. Hypotheses regarding metabolic interactions could be tested by querying metagenomic datasets for genes affiliated with correlated microbial groups, and searching for complementary steps in putatively shared metabolic pathways.

5.2 Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean

The first step in a study of microbial ecology is often to define the phylogeny and distribution of the breadth of organismal types in question. In this chapter, I used small subunit ribosomal RNA (SSU rRNA) gene survey data to classify evolutionary relationships among MGA sequence types (OTUs) detected in the NESAP, and to estimate the abundance and distribution of these OTUs throughout the water column. To begin, I used three complementary approaches to measure total MGA abundance in the NESAP: (1) CARD-FISH, using probe SAR406-97 (2) SSU rRNA gene clone library sequencing, and (3) 454-pyrotag sequencing of SSU rRNA genes. MGA abundance estimates across methods ranged between 0% and 1.3% of total bacteria in surface waters, and between 3.3% and 11.6% of total bacteria in OMZ waters. In surveying bacterial SSU rRNA gene clone libraries generated from the NESAP, I identified 290 sequences affiliated with MGA comprising distinct 121 OTUs when clustered at 97% identify, and placed these OTUs into phylogenetic context with relevant reference sequences from other environments. I defined 5 novel subgroups of MGA and recovered 5 previously defined subgroups. To explore the population structure of MGA subroups with increased resolution, I recruited all pyrotags to all SSU rRNA gene clone library sequences affiliated with MGA and obtained direct matches to 78 out of 121 OTUs. OTUs affiliated with the MGA subgroup Arctic95A-2 were consistently the most abundant within NESAP OMZ waters.

To explore the hypothesis of O_2 and other environmental factors as drivers of habitat selection for MGA subgroups in the NESAP water column I calculated Spearman's rank correlation coefficients between pyrotag sequence abundance estimates for specific OTUs and environmental parameters. The relative abundance of 4 OTUs identified in 454-pyrotag datasets

showed significant correlations with decreasing O_2 after a Bonferroni correction was applied ($p<0.000079$). These OTUs were affiliated with 2 subgroups of MGA (Arctic95A-2 and A714018). An additional 13 OTUs were weakly correlated with decreasing O_2.

This is the first reported analysis of MGA diversity and population structure, and the first study to establish the phylogeny of MGA subgroups. Results indicate that MGA is a broadly diverse candidate phylum and that many distinct OTUs of MGA are present in the NESAP at varying relative abundances. The negative correlation between presence of certain MGA OTUs and O_2 concentration is consistent with habitat selection in suboxic waters, and suggests the potential for adaptation to O_2-deficiency.

5.2.1 Limitations

MGA total abundance estimates were highly correlated between CARD-FISH and SSU rRNA gene clone libraries, but not between CARD-FISH and pyrotag sequences, while SSU rRNA gene clone library and pyrotag sequences were correlated based on Spearman's rank correlations. In addition, CARD-FISH-based estimates were consistently lower than SSU rRNA gene clone library or pyrotag sequence estimates for the same samples, suggesting an under or overestimation of MGA abundance by some or all of the methods applied. These incongruities could be based on limited probe access to target cells during the application of CARD-FISH, or to variability in efficiency of binding specific sequence types by primers and probes. Alternatively, MGA subgroups could contain multiple copies of the SSU rRNA gene, which would inflate PCR-based estimates of abundance.

The ability to assign pyrotags to only 78 out of 121 OTUs defined by SSU rRNA gene clone library sequences may have resulted from the stringent nature of the chosen approach, having required full-length pyrotag sequences to match a cognate SSU rRNA gene clone library sequence with no mismatches. It is also possible that there are temporal patterns in the abundance and distribution of MGA OTUs, which prevented assignment of pyrotags to all SSU rRNA-derived OTUs because pyrotag datasets were generated from June 2009 samples while SSU rRNA gene clone libraries were derived from February 2009 samples.

5.2.2 Future directions

More extensive studies of MGA diversity are required to ascertain the biogeographic range and complete phylogenetic diversity of this candidate phylum of bacteria, including surveys of terrestrial, freshwater, and other non-marine environments. Improvements in phylogenetic resolution of MGA subgroups in public taxonomic databases (e.g. Greengenes (DeSantis et al., 2006a)) using the taxonomy defined in this study would greatly simplify future studies of MGA population structure. Quantitative studies (for example, using qPCR) documenting the temporal dynamics of MGA subgroups across multiple marine environments are needed to understand the uniformity and stability of MGA populations in the ocean. Genome-scale sequence data are needed to determine whether observed SSU rRNA-based patterns of distribution across the oxycline are indeed associated with metabolic adaptation to O_2-deficiency and ecotype differentiation among MGA subgroups.

5.3 Metabolic capacity of Marine Group A bacteria

Despite the prevalence of MGA bacteria in marine environments and particularly within O_2-deficient waters, the metabolic capacity and ecological role of these organisms in OMZs or in the ocean at large has never before been studied. In this chapter, I first extended the range of known MGA diversity to include 3 additional subgroups detected in suboxic and anoxic-sulfidic waters of the seasonally anoxic basin, Saanich Inlet (SI). I then performed phylogenetic anchor screening on 23 large-insert metagenomic libraries generated from NESAP and SI samples and identified and sequenced to completion 14 fosmid inserts derived from MGA bacteria as a route to studying MGA function. I obtained 4 additional large-insert DNA fragments derived from MGA from public databases and included these inserts in comparative analyses. Phylogenetic analysis of SSU rRNA genes located on large-insert DNA fragments found the fragments to be affiliated with 9 discrete MGA subgroups, while one fragment grouped outside of MGA was more closely affiliated with the phylum *Deferribacteres*.

Large-insert DNA fragments did not partition into discrete groups based on similar GC content, tetranucleotide frequency, or global nucleotide similarity, however, fragments did coalesce into 5 syntenic groups based on shared amino acid similarity of predicted open reading frames (ORFs). Syntenic groups did not consistently reflect subgroup partitioning patterns

observed in SSU rRNA gene clone libraries or patterns of distribution along the oxycline. However, predicted protein-coding genes associated with adaptation to O_2 deficiency and sulfur-based energy metabolism were detected on multiple fosmids. Of particular interest was an operon encoding polysulfide reductase (PSR), of the dimethylsulfoxide (DMSO) family of oxidoreductases, detected on two fosmids derived from the NESAP and SI. In the anaerobic Epsilonproteobacterium *Wolinella succinogenes* (where this protein complex has been characterized), PSR in combination with hydrogenase (HYD) or formate dehydrogenase (FDH) allows respiration on polysulfide (S_n) using H_2 or formate as an electron donor, and results in the production of H_2S. The PSR complex has also been documented to catalyze sulfide oxidation to polysulfide. Analysis of the predicted protein sequences of MGA-encoded PSR indicated that these proteins were most closely related to Psr proteins encoded on euryarchaeotal fosmids derived from the Mediterranean Sea and Monterey Bay, and to Psr proteins found in several green sulfur bacteria (GSB), suggesting that this operon has been involved in multiple lateral transfer events or that it arose in an ancient microbial lineage before the bacterial – archaeal split. While direct evidence for PSR activity cannot be inferred from protein sequence alone, the detection of genes encoding PSR homologues suggests a potential role for MGA bacteria in the cryptic sulfur cycle recently discovered to play a central role in the microbial ecology and biogeochemical cycling of OMZs.

5.3.1 Limitations

Although 255.3 Mb of metagenomic DNA sequence data was screened to identify MGA derived sequences, only ~540 kb were able to be linked to MGA based on phylogenetic anchor screening for MGA SSU rRNA sequences. Given the abundance and diversity of MGA in the environment, this is not a large enough sample size of genomic DNA to obtain complete metabolic characterization of the capacity within the MGA candidate phylum. In addition, of the 496 total predicted ORFs linked to MGA in this analysis, 73% were annotated as 'hypothetical proteins', further reducing any ability to confer metabolic properties onto distinct MGA subgroups. It is possible that this inefficiency in informative annotation is a result of inadequate coverage of the protein universe in public databases, but could also be due to highly divergent proteins encoded

by this deeply-branching bacterial group that do not show similarity to existing protein databases.

With respect to inferring PSR activity, we cannot infer protein substrate or activity from protein sequence alone. As this study lacks experimental evidence of PSR activity, the role of MGA in marine sulfur cycling remains a hypothesis.

5.3.2 Future directions

As obtaining pure cultures of environmentally relevant microoganisms is notoriously difficult, more extensive insights into MGA metabolic capacity and potential ecophysiology might be best obtained through the application of single-cell genomics. Analysis of single-cell amplified genomes (SAGs) affiliated with various subgroups of MGA could aid in providing the genome-wide information needed to more adequately describe the metabolic capacity of defined subgroups and to explore the possibility of ecotype differentiation and niche partitioning among MGA subgroups. In addition, comparison of SAGs derived from the same MGA populations could aid in the search for potential patterns of distributed metabolism within discrete MGA populations, as suggested by MGA co-occurrence patterns described in Chapter 2. This high-resolution genomic data may also provide clues as to the evolutionary history of MGA bacteria. Finally, experimental verification of the activity of putative MGA polysulfide reductase (PSR), for example by quantifying expression of the *psr* operon detected on MGA fosmids in an *E. coli* host, is needed to conclusively link sulfur transformations to MGA activity in the ocean. Alternatively, environmental ecophysiology techniques such as Halogen In Situ Hybridization-Secondary Ion Mass Spectroscopy (HISH-SIMS) could be used to perform simultaneous phylogenetic identification and quantification of sulfur-related metabolic activities of single MGA cells in the environment (Musat et al., 2008). Finally, it will be of interest to explore the diversity of microorganisms expressing *psr* operons, and to quantify the extent and utility of this metabolic strategy in the ocean.

5.4 Significance of this research

Oxygen minimum zones are sites of intensive biogeochemical cycling with important feedbacks on ocean ecology and climate. Given that OMZs are expanding and intensifying, primarily as a

result of global climate change, it is of increasing importance to define the diversity and ecosystem function of dominant microorganisms within these systems in order to predict the systemic impacts on ocean ecology and biogeochemistry. The work presented in this dissertation characterized, for the first time, the diversity and distribution of one of the most abundant groups of microorganisms detected in the world's largest OMZ: Marine Group A. In addition, this dissertation applied novel techniques to assess the connectedness of MGA within the microbial ecosystem of the Northeast subarctic Pacific Ocean, and provided first insights into a potential role of MGA bacteria in the marine sulfur cycle.

Bibliography

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. Nature *430*, 551-54.

Agrawal, H. (2002). Extreme self-organization in networks constructed from gene expression data. Phys Rev Lett *89*, 268702.

Aguiar, P., Beveridge, T., and Reysenbach, A. (2004). Sulfurihydrogenibium azorense, sp. nov., a thermophilic hydrogen-oxidizing microaerophile from terrestrial hot springs in the Azores. International Journal Of Systematic And Evolutionary Microbiology *54*, 33-39.

Albert, R., Jeong, H., and Barabási, A.-L. (1999). Internet: Diameter of the world-wide web. Nature *401*, 130-31.

Alldredge, A.L., and Cohen, Y. (1987). Can microscale chemical patches persist in the sea? Microelectrode study of marine snow, fecal pellets. Science *235*, 689-691.

Alldredge, A.L., and Silver, M.W. (1988). Characteristics, dynamics and significance of marine snow. Progress in oceanography *20*, 41-82.

Allers, E., Wright, J.J., Konwar, K.M., Howes, C.G., Beneze, E., Hallam, S.J., and Sullivan, M.B. (2012). Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. ISME J *7*, 256-268.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol *56*, 1919-925.

Anderson, J.J., and Devol, A.H. (1973). Deep water renewal in Saanich Inlet, an intermittently anoxic basin. Estuarine and Coastal Marine Science *1*, 1-10.

Assenov, Y., Ramirez, F., Schelhorn, S.E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. Bioinformatics *24*, 282-84.

Bakker, J.D. (2008). Increasing the utility of indicator species analysis. Journal of Applied Ecology *45*, 1829-835.

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

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science *286*, 509-512.

Barabasi, A.L., and Bonabeau, E. (2003). Scale-free networks. Sci Am *288*, 60-69.

Barabási, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nature Reviews Genetics *5*, 101-113.

Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J *6*, 343-351.

Barwell-Clarke, J., and Whitney, F. (1996). Institute of Ocean Sciences nutrient methods and analysis. Can. Tech. Rep. Hydrogr. Ocean Sci., 49.

Belmar, L., Molina, V., and Ulloa, O. (2011). Abundance and phylogenetic identity of archaeoplankton in the permanent oxygen minimum zone of the eastern tropical South Pacific. FEMS Microbiol Ecol *78*, 314-326.

Beman, J.M., Popp, B.N., and Francis, C.A. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. ISME J *2*, 429-441.

Biegala, I.C., Cuttle, M., Mary, I., and Zubkov, M. (2005). Hybridisation of picoeukaryotes by eubacterial probes is widespread in the marine environment. Aquatic microbial ecology *41*, 293-97

Bograd, S.J., Castro, C.G., Di Lorenzo, E., Palacios, D.M., Bailey, H., Gilly, W., and Chavez, F.P. (2008). Oxygen declines and the shoaling of the hypoxic boundary in the California Current. Geophysical Research Letters *35*

*B*ooth, B.C., Lewin, J., and Postel, J.R. (1993). Temporal variation in the structure of autotrophic and heterotrophic communities in the subarctic Pacific. Progress in Oceanography *32*, 57-99.

Boyd, P., and Harrison, P.J. (1999). Phytoplankton dynamics in the NE subarctic Pacific. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2405-432.

Boyd, P.W., Law, C.S., Wong, C.S., Nojiri, Y., Tsuda, A., Levasseur, M., Takeda, S., Rivkin, R., Harrison, P.J., and Strzepek, R. (2004). The decline and fate of an iron-induced subarctic phytoplankton bloom. Nature *428*, 549-553.

Boyd, P.W., Muggli, D.L., Varela, D.E., Goldblatt, R.H., Chretien, R., Orians, K.J., and Harrison, P.J. (1996). In vitro iron enrichment experiments in the NE subarctic Pacific. Marine ecology progress series. Oldendorf *136*, 179-193.

Breitburg, D.L. (2002). Effects of hypoxia, and the balance between bypoxia and enrichment, on coastal fishes and fisheries. Estuaries *25*, 767-781.

Breitburg, D.L., Hondorp, D.W., Davias, L.A., and Diaz, R.J. (2009). Hypoxia, nitrogen, and fisheries: integrating effects across local and global landscapes. Ann Rev Mar Sci *1*, 329-349.

Brown, M.V., and Donachie, S.P. (2007). Evidence for tropical endemicity in the Deltaproteobacteria Marine Group B/SAR324 bacterioplankton clade. Aquat. Microb. Ecol. *46*, 107-115.

Brune, A., Frenzel, P., and Cypionka, H. (2000). Life at the oxic--anoxic interface: microbial activities and adaptations. FEMS Microbiology Reviews *24*, 691-710.

Brüchert, V., Jørgensen, B.B., Neumann, K., Riechmann, D., Schlösser, M., and Schulz, H. (2003). Regulation of bacterial sulfate reduction and hydrogen sulfide fluxes in the central Namibian coastal upwelling zone. Geochimica et Cosmochimica Acta *67*, 4505-518.

Canfield, D.E. (2006). Models of oxic respiration, denitrification and sulfate reduction in zones of coastal upwelling *70*, 5753-765.

Canfield, D.E., Stewart, F.J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E.F., Revsbech, N.P., and Ulloa, O. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *330*, 1375-78.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., and Gordon, J.I. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods *7*, 335-36.

Carlson, C.A., and Ducklow, H.W. (1996). Growth of bacterioplankton and consumption of dissolved organic carbon in the Sargasso Sea. Aquatic Microbial Ecology *10*, 69-85.

Chaffron, S., Rehrauer, H., Pernthaler, J., and von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res *20*, 947-959.

Chisholm, S.W., Falkowski, P.G., and Cullen, J.J. (2001). Oceans. Dis-crediting ocean fertilization. Science *294*, 309-310.

Cipollone, R., Ascenzi, P., and Visca, P. (2007). Common themes and variations in the rhodanese superfamily. IUBMB life *59*, 51-59.

Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. Australian journal of ecology *18*, 117-143.

Clarke, K.R., and Gorley, R.N. (2006). PRIMER v6. User manual/tutorial. Plymouth routine in mulitvariate ecological research. Plymouth Marine Laboratory

Codispoti, L.A., Brandes, J.A., Christensen, J.P., Devol, A.H., Naqvi, S.W.A., Paerl, H.W., and Yoshinari, T. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? Scientia Marina *65*, 85-105.

Codispoti, L.A.A.J.C. (1985). Nitrification, denitrification and nitrous oxide cycling in the easter tropical South Pacific Ocean. Marine Chemistry *16*, 277-300.

Cole, J.A., and Brown, C.M. (1980). Nitrite reduction to ammonia by fermentative bacteria: a short circuit in the biological nitrogen cycle. FEMS Microbiology Letters *7*, 65-72.

Coleman, M.L., and Chisholm, S.W. (2007). Code and context: Prochlorococcus as a model for cross-scale biology. Trends Microbiol *15*, 398-407.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006). Genomic islands and the ecology and evolution of Prochlorococcus. Science *311*, 1768-770.

Conley, D.J., Humborg, C., Rahm, L., Savchuk, O.P., and Wulff, F. (2002). Hypoxia in the Baltic Sea and basin-scale changes in phosphorus biogeochemistry. Environmental science & technology *36*, 5315-320.

Coolen, M.J., Abbas, B., van Bleijswijk, J., Hopmans, E.C., Kuypers, M.M., Wakeham, S.G., and Sinninghe Damste, J.S. (2007). Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. Environ Microbiol *9*, 1001-016.

Cottrell, M.T., and Kirchman, D.L. (2003). Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. Limnology and Oceanography *48*, 168-178.

Cuttelod, A., and Claustre, H. (2010). ALMOFRONT 2 cruise in Alboran sea: Chlorophyll fluorescence calibration. Journal of Oceanography, Research and Data *3*

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H., and Wagner, M. (1999). The Domain-specific Probe EUB338 is Insufficient for the Detection of all< i> Bacteria: Development and Evaluation of a more Comprehensive Probe Set. Systematic and Applied Microbiology *22*, 434-444.

Damsté, J.S.S., Rijpstra, W.I.C., Hopmans, E.C., Prahl, F.G., Wakeham, S.G., and Schouten, S. (2002). Distribution of membrane lipids of planktonic Crenarchaeota in the Arabian Sea. Appl Environ Microbiol *68*, 2997-3002.

del Giorgio, P.A., and Duarte, C.M. (2002). Respiration in the open ocean. Nature *420*, 379-384.

DeLong, E.F. (1992). Archaea in coastal marine environments. Proc Natl Acad Sci U S A *89*, 5685-89.

DeLong, E.F. (1998). Everything in moderation: archaea as 'non-extremophiles'. Curr Opin Genet Dev *8*, 649-654.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. Science *311*, 496-503.

deMaré, F., Kurtz, D.M., and Nordlund, P. (1996). The structure of Desulfovibrio vulgaris rubrerythrin reveals a unique combination of rubredoxin-like FeS4 and ferritin-like diiron domains. Nature Structural & Molecular Biology *3*, 539-546.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006a). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol *72*, 5069-072.

DeSantis, T.Z.J., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., Phan, R., and Andersen, G.L. (2006b). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res *34*, W394-99.

Deutsch, C., Sarmiento, J.L., Sigman, D.M., Gruber, N., and Dunne, J.P. (2007). Spatial coupling of nitrogen inputs and losses in the ocean. Nature *445*, 163-67.

Devol, A.H., and Hartnett, H.E. (2001). Role of the oxygen-deficient zone in transfer of organic carbon to the deep ocean. Limnology and oceanography , 1684-690.

DFO (2011). State of the Pacific Ocean 2010 (DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2011/032).

Diaz, R.J., and Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. Science *321*, 926-29.

Doherty, S. (1995). The abundance and distribution of heterotrophic and autotrophic nanofagellates in the NE subarctic Pacific.

Doney, S.C. (2010). The growing human footprint on coastal and open-ocean biogeochemistry. Science *328*, 1512-16.

Dubilier, N., Mülders, C., Ferdelman, T., de Beer, D., Pernthaler, A., Klein, M., Wagner, M., Erséus, C., Thiermann, F., and Krieger, J. (2001). Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. Nature *411*, 298-302.

Dufrêne, M., and Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological monographs *67*, 345-366.

Dugdale, R., Goering, J., Barber, R., Smith, R., and Packard, T. (1977). Denitrification and hydrogen sulfide in the Peru upwelling region during 1976 *24*, 601-08.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-97.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-61.

Edgcomb, V., Orsi, W., Taylor, G.T., Vdacny, P., Taylor, C., Suarez, P., and Epstein, S. (2011). Accessing marine protists from the anoxic Cariaco Basin. ISME J *5*, 1237-241.

Ekau, W., Auel, H., Portner, H.O., and Gilbert, D. (2010). Impacts of hypoxia on the structure and processes in pelagic communities (zooplankton, macro-invertebrates and fish). Biogeosciences *7*, 1669-699.

Emerson, S., Watanabe, Y.W., Ono, T., and Mecking, S. (2004). Temporal trends in apparent oxygen utilization in the upper pycnocline of the North Pacific: 1980-2000. Journal of Oceanography *60*, 139-147.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., and Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J *4*, 642-47.

Estrada, M., and Marrasé, C. (1987). Phytoplankton biomass and productivity off the Namibian coast. South African Journal of Marine Science *5*, 347-356.

Falkowski, P.G., Algeo, T., Codispoti, L., Deutsch, C., Emerson, S., Hales, B., Huey, R.B., Jenkins, W.J., Kump, L.R., and Levin, L.A. (2011). Ocean deoxygenation: Past, present, and future. Eos, Transactions American Geophysical Union *92*, 409.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science *320*, 1034-39.

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. Nature Reviews Microbiology *10*, 538-550.

Featherstone, D.E., and Broadie, K. (2002). Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. Bioessays *24*, 267-274.

Field, K.G., Gordon, D., Wright, T., Rappe, M., Urback, E., Vergin, K., and Giovannoni, S.J. (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. Appl Environ Microbiol *63*, 63-70.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005). Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. P Natl Acad Sci Usa *102*, 14683-88.

Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., and Feldman, M.W. (2002). Evolutionary rate in the protein interaction network. Science Signaling *296*, 750.

Freeland, H. (2007). A short history of ocean station papa and Line P. Progress in Oceanography *75*, 120-25.

Freeland, H., Denman, K., Wong, C.S., Whitney, F., and Jacques, R. (1997). Evidence of change in the winter mixed layer in the Northeast Pacific Ocean. Deep-Sea Res Pt I *44*, 2117-129.

Frigaard, N.U., and Bryant, D.A. (2008). Genomic insights into the sulfur metabolism of phototrophic green sulfur bacteria. Sulfur Metabolism in Phototrophic Organisms , 337-355.

Frost, B.W. (1991). The role of grazing in nutrient-rich areas of the open sea. Limnology and Oceanography *36*, 1616-630.

Fuchs, B.M., Woebken, D., Zubkov, M.V., Burkill, P., and Amann, R. (2005). Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. Aquat. Microb. Ecol. *39*, 145-157.

Fuenzalida, R., Schneider, W., Garcés-Vargas, J., Bravo, L., and Lange, C. (2009). Vertical and horizontal extension of the oxygen minimum zone in the eastern South Pacific Ocean *56*, 992-1003.

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

Fuhrman, J.A., and Davis, A.A. (1997). Widespread Archaea and novel Bacteria from the deep sea as shown by 16S rRNA gene sequences. Marine Ecology Progress Series *150*, 275-285.

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

Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences *106*, 22427-432.

Galand, P.E., Lovejoy, C., Pouliot, J., and Vincent, W.F. (2008). Heterogeneous archaeal communities in the particle-rich environment of an arctic shelf ecosystem. Journal of Marine Systems *74*, 774-782.

Galán, A., Molina, V., Thamdrup, B., Woebken, D., Lavik, G., Kuypers, M.M.M., and Ulloa, O. (2009). Anammox bacteria and the anaerobic oxidation of ammonium in the oxygen minimum zone off northern Chile. Deep Sea Research Part II: Topical Studies in Oceanography *56*, 1021-031.

Garcia, H.E., Locarnini, R.A., Boyer, T.P., and Antonov, J.I. (2010). World Ocean Atlas 2009, vol. 4, Nutrients (Phosphate, Nitrate, Silicate), NOAA Atlas NESDIS, vol. 71. US Gov. Print. Off., Washington, DC

Gihring, T.M., Green, S.J., and Schadt, C.W. (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environ Microbiol *14*, 285-290.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., et al. (2012). Defining seasonal marine microbial community dynamics. ISME J *6*, 298-308.

Goericke, R., Olson, R., and Shalapyonok, A. (2000). A novel niche for *Prochlorococcus* sp. in low-light suboxic environments in the Arabian Sea and the Eastern Tropical North Pacific. Deep Sea Research Part I *47*, 1183-1205.

González, J.M., Covert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B., Schmitt, R., Buchan, A., Fuhrman, J.A., and Kiene, R.P. (2003). Silicibacter pomeroyi sp. nov. and Roseovarius nubinhibens sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. International journal of systematic and evolutionary microbiology *53*, 1261-69.

Gordon, D.A., and Giovannoni, S.J. (1996). Detection of stratified microbial populations related to Chlorobium and Fibrobacter species in the Atlantic and Pacific oceans. Appl Environ Microbiol *62*, 1171-77.

Grantham, B.A., Chan, F., Nielsen, K.J., Fox, D.S., Barth, J.A., Huyer, A., Lubchenco, J., and Menge, B.A. (2004). Upwelling-driven nearshore hypoxia signals ecosystem and oceanographic changes in the northeast Pacific. Nature *429*, 749-754.

Gregersen, L.H., Bryant, D.A., and Frigaard, N.U. (2011). Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Frontiers in microbiology *2*

Grimaud, R., Ezraty, B., Mitchell, J.K., Lafitte, D., Briand, C., Derrick, P.J., and Barras, F. (2001). Repair of oxidized proteins: Identification of a new methionine sulfoxide reductase. Journal of Biological Chemistry *276*, 48915-920.

Grote, J., Jost, G., Labrenz, M., Herndl, G.J., and Jürgens, K. (2008). Epsilonproteobacteria represent the major portion of chemoautotrophic bacteria in sulfidic waters of pelagic redoxclines of the Baltic and Black Seas. Appl Environ Microbiol *74*, 7546-551.

Grote, J., Schott, T., Bruckner, C.G., Glöckner, F.O., Jost, G., Teeling, H., Labrenz, M., and Jürgens, K. (2012). Genome and physiology of a model Epsilonproteobacterium responsible for sulfide detoxification in marine oxygen depletion zones. Proc Natl Acad Sci U S A *109*, 506-510.

Gruber, N., and Galloway, J.N. (2008). An Earth-system perspective of the global nitrogen cycle. Nature *451*, 293-96.

Grzymski, J.J., Riesenfeld, C.S., Williams, T.J., Dussaq, A.M., Ducklow, H., Erickson, M., Cavicchioli, R., and Murray, A.E. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. ISME J

Guimera, R., and Amaral, L.A.N. (2005). Functional cartography of complex metabolic networks. Nature *433*, 895-900.

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res *33*, W557-59.

Guinotte, J.M., and Fabry, V.J. (2008). Ocean acidification and its potential effects on marine ecosystems. Ann N Y Acad Sci *1134*, 320-342.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., and DeLong, E.F. (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. Proc Natl Acad Sci U S A *103*, 18296-8301.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. Science *305*, 1457-462.

Harris, S.L., Varela, D.E., Whitney, F.W., and Harrison, P.J. (2009). Nutrient and phytoplankton dynamics off the west coast of Vancouver Island during the 1997/98 ENSO event. Deep Sea Research Part II: Topical Studies in Oceanography *56*, 2487-2502.

Harrison, P.J., Boyda, P.W., Varela, D.E., Takeda, S., Shiomoto, A., and Odate, T. (1999). Comparison of factors controlling phytoplankton productivity in the NE and NW subarctic Pacific gyres. Progress in Oceanography *43*, 205-234.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47-C52.

Hedderich, R., Klimmek, O., Kröger, A., Dirmeier, R., Keller, M., and Stetter, K.O. (1999). Anaerobic respiration with elemental sulfur and with disulfides. FEMS microbiology reviews *22*, 353-381.

Helly, J.J., and Levin, L.A. (2004). Global distribution of naturally occurring marine hypoxia on continental margins. Deep-Sea Res Pt I *51*, 1159-168.

Helm, K.P., Bindoff, N.L., and Church, J.A. (2011). Observed decreases in oxygen content of the global ocean. Geophysical Research Letters *38*, 1-6.

Hild, E., Takayama, K., Olsson, R.M., and Kjelleberg, S. (2000). Evidence for a role of rpoE in stressed and unstressed cells of marine Vibrio angustum strain S14. J Bacteriol *182*, 6964-974.

Holm-Hansen, O., Lorenzen, C.J., Holmes, R.W., and Strickland, J.D. (1965). Fluorometric determination of chlorophyll. Journal du Conseil *30*, 3-15.

Hoshino, T., Yilmaz, L.S., Noguera, D.R., Daims, H., and Wagner, M. (2008). Quantification of target molecules needed to detect microorganisms by fluorescence in situ hybridization (FISH) and catalyzed reporter deposition-FISH. Appl Environ Microbiol *74*, 5068-077.

Howard, E.C., Henriksen, J.R., Buchan, A., Reisch, C.R., Burgmann, H., Welsh, R., Ye, W., Gonzalez, J.M., Mace, K., et al. (2006). Bacterial Taxa That Limit Sulfur Flux from the Ocean. Science *314*, 649-652.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics *20*, 2317-19.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., Salter, I., Agogué, H., Mary, I., and Galand, P.E. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci U S A *110*, 6004-09.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119.

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K., Nealson, K.H., and Horikoshi, K. (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. Appl Environ Microbiol *69*, 7224-235.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences *98*, 4569-574.

Jankielewicz, A., Klimmek, O., and Kröger, A. (1995). The electron transfer from hydrogenase and formate dehydrogenase to polysulfide reductase in the membrane of Wolinella succinogenes. Biochimica et Biophysica Acta (BBA)-Bioenergetics *1231*, 157-162.

Jeon, S.-O., Ahn, T.-S., and Hong, S.-H. (2008). A novel archaeal group in the phylum Crenarchaeota found unexpectedly in an eukaryotic survey in the Cariaco Basin. The Journal of Microbiology *46*, 34-39

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41-42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. Nature *407*, 651-54.

Johnson, Z., Landry, M., Bidigare, R., Brown, S., Campbell, L., Gunderson, J., Marra, J., and Trees, C. (1999). Energetics and growth kinetics of a deep Prochlorococcus spp. population in the Arabian Sea *46*, 1719-743.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M., and Chisholm, S.W. (2006). Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science *311*, 1737-740.

Jormakka, M., Yokoyama, K., Yano, T., Tamakoshi, M., Akimoto, S., Shimamura, T., Curmi, P., and Iwata, S. (2008). Molecular mechanism of energy conservation in polysulfide respiration. Nat Struct Mol Biol *15*, 730-37.

Jørgensen, B.B. (1982). Ecology of the bacteria of the sulphur cycle with special reference to anoxic-oxic interface environments. Philos Trans R Soc Lond B Biol Sci *298*, 543-561.

Jürgens, G., Glöckner, F.-O., Amann, R., Saano, A., Montonen, L., Likolammi, M., and Münster, U. (2000). Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridization1. FEMS Microbiol Ecol *34*, 45-56.

Jürgens, G., Lindström, K., and Saano, A. (1997). Novel group within the kingdom Crenarchaeota from boreal forest soil. Appl Environ Microbiol *63*, 803-05.

Kamiya, E., Izumiyama, S., Nishimura, M., Mitchell, J.G., and Kogure, K. (2007). Effects of fixation and storage on flow cytometric analysis of marine bacteria. Journal of oceanography *63*, 101-112

Kanehisa, M., and Goto, S. (1999). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res *27*, 29-34.

Karl, D.M., and Tilbrook, B.D. (1994). Production and transport of methane in oceanic particulate organic matter. Nature *368*, 732-34.

Karl, D.M., Knauer, G.A., Martin, J.H., and Ward, B.B. (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. Nature *309*, 54-56.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S.M., and Pellegrini-Toole, A. (2000). The ecocyc and metacyc databases. Nucleic Acids Res *28*, 56-59.

Karstensen, J., Stramma, L., and Visbeck, M. (2008). Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans. Progress in Oceanography *77*, 331-350.

Kartal, B., Kuypers, M.M.M., Lavik, G., Schalk, J., Op den Camp, H.J.M., Jetten, M.S.M., and Strous, M. (2007). Anammox bacteria disguised as denitrifiers: nitrate reduction to dinitrogen gas via nitrite and ammonium. Environ Microbiol *9*, 635-642.

Kasting, J.F., and Siefert, J.L. (2002). Life and the evolution of Earth's atmosphere. Science *296*, 1066-68.

Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-066.

Keeling, R.F., Körtzinger, A., and Gruber, N. (2010). Ocean deoxygenation in a warming world. Ann Rev Mar Sci *2*, 199-229.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., et al. (2007). Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. PLoS Genet *3*, e231.

Kirchman, D.L., Keel, R.G., Simon, M., and Welschmeyer, N.A. (1993). Biomass and production of heterotrophic bacterioplankton in the oceanic subarctic Pacific. Deep Sea Research Part I: Oceanographic Research Papers *40*, 967-988.

Kirchman, D.L., Rich, J.H., and Barber, R.T. (1995). Biomass and biomass production of heterotrophic bacteria along 140 W in the equatorial Pacific: Effect of temperature on the microbial loop. Deep Sea Research Part II: Topical Studies in Oceanography *42*, 603-619.

Klimmek, O., Kröger, A., Steudel, R., and Holdt, G. (1991). Growth of Wolinella succinogenes with polysulphide as terminal acceptor of phosphorylative electron transport. Arch Microbiol *155*, 177-182.

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, A., Ward, D.M., Rooney, A.P., Brambilla, E., Connor, N., et al. (2008). Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. Proceedings of the National Academy of Sciences *105*, 2504-09.

Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature *437*, 543-46.

Krafft, T., Bokranz, M., Klimmek, O., Schröder, I., Fahrenholz, F., Kojro, E., and Kröger, A. (1992). Cloning and nucleotide sequence of the pst A gene of Wolinella succinogenes polysulphide reductase. European Journal of Biochemistry *206*, 503-510.

Krafft, T., Gross, R., and Kröger, A. (1995). The function of Wolinella succinogenes psr genes in electron transport with polysulphide as the terminal electron acceptor. European Journal of Biochemistry *230*, 601-06.

Krzywinski, M., Schein, J., Birol, \.I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res *19*, 1639-645.

Kumar, S. (2007). Fourth assessment report of the Intergovernmental Panel on Climate Change: Important observations and conclusions. Current Science *92*, 1034-4.

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

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol *5*, R12.

Labrenz, M., Jost, G., and Jürgens, K. (2007). Distribution of abundant prokaryotic organisms in the water column of the central Baltic Sea with an oxic-anoxic interface. Aquat. Microb. Ecol. *46*, 177.

Labrenz, M., Sintes, E., Toetzke, F., Zumsteg, A., Herndl, G.J., Seidler, M., and Jürgens, K. (2010). Relevance of a crenarchaeotal subcluster related to Candidatus Nitrosopumilus maritimus to ammonia oxidation in the suboxic zone of the central Baltic Sea. ISME J

Lam, P., Jensen, M.M., Lavik, G., McGinnis, D.F., Mueller, B., Schubert, C.J., Amann, R., Thamdrup, B., and Kuypers, M.M.M. (2007). Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. P Natl Acad Sci Usa *104*, 7104-09.

Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D., Gutierrez, D., Amann, R., Jetten, M.S.M., and Kuypers, M.M.M. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. P Natl Acad Sci Usa *106*, 4752-57.

Lavik, G., Stührmann, T., Brüchert, V., Plas, A.V.D., Mohrholz, V., Lam, P., Mußmann, M., Fuchs, B.M., Amann, R., et al. (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. Nature *457*, 581-85.

Lees, H., and Simpson, J.R. (1957). The biochemistry of the nitrifying organisms. 5. Nitrite oxidation by Nitrobacter. Biochemical Journal *65*, 297.

Lilley, M.D., Baross, J.A., and Gordon, L.I. (1982). Dissolved hydrogen and methane in Saanich Inlet, British Columbia. Deep Sea Res. *29*, 1471-484.

Lin, X., Scranton, M.I., Chistoserdov, A.Y., Varela, R., and Taylor, G.T. (2008). Spatiotemporal dynamics of bacterial populations in the anoxic Cariaco Basin. Limnology and Oceanography , 37-51.

Lin, X., Wakeham, S.G., Putnam, I.F., Astor, Y.M., Scranton, M.I., Chistoserdov, A.Y., and Taylor, G.T. (2006). Comparison of vertical distributions of prokaryotic assemblages in the anoxic Cariaco Basin and Black Sea by use of fluorescence in situ hybridization. Appl Environ Microbiol *72*, 2679-690.

Lipschultz, F., Wofsy, S.C., WARD, B.B., Codispoti, L.A., Friedrich, G., and Elkins, J.W. (1990). Bacterial transformations of inorganic nitrogen in the oxygen-deficient waters of the eastern tropical South Pacific Ocean. Deep Sea Research Part A. Oceanographic Research Papers *37*, 1513-541.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res *32*, 1363-371.

Madrid, V.M., Taylor, G.T., Scranton, M.I., and Chistoserdov, A.Y. (2001). Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. Appl Environ Microb *67*, 1663-674.

Maldonado, M.T., Boyd, P.W., Harrison, P.J., and Price, N.M. (1999). Co-limitation of phytoplankton growth by light and Fe during winter in the NE subarctic Pacific Ocean. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2475-485.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the american Meteorological Society *78*, 1069-079.

Martin, J.H., and Fitzwater, S.E. (1988). Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. Nature *331*, 341-43.

Martinez-Garcia, M., Brazel, D., Poulton, N.J., Swan, B.K., Gomez, M.L., Masland, D., Sieracki, M.E., and Stepanauskas, R. (2011). Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. ISME J *6*, 703-07.

Massana, R., DeLong, E.F., and Pedrós-Alió, C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. Appl Environ Microbiol *66*, 1777-787.

Matear, R.J., and Hirst, A.C. (2003). Long-term changes in dissolved oxygen concentrations in the ocean caused by protracted global warming. Global Biogeochemical Cycles *17*, 1125.

Miller, C.B., Frost, B.W., Wheeler, P.A., Landry, M.R., Welschmeyer, N., and Powell, T.M. (1991). Ecological dynamics in the subarctic Pacific, a possibly iron-limited ecosystem. Limnol. Oceanogr *36*, 1600-615.

Milucka, J., Ferdelman, T.G., Polerecky, L., Franzke, D., Wegener, G., Schmid, M., Lieberwirth, I., Wagner, M., Widdel, F., and Kuypers, M.M. (2012). Zero-valent sulphur is a key intermediate in marine methane oxidation. Nature *491*, 541-46.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environ Microbiol *9*, 1162-175.

Miroshnichenko, M.L., Kolganova, T.V., Spring, S., Chernyh, N., and Bonch-Osmolovskaya, E.A. (2010). Caldithrix palaeochoryensis sp. nov., a thermophilic, anaerobic, chemo-organotrophic bacterium from a geothermally heated sediment, and emended description of the genus Caldithrix. International journal of systematic and evolutionary microbiology *60*, 2120-23.

Miroshnichenko, M.L., Kostrikina, N.A., Chernyh, N.A., Pimenov, N.V., Tourova, T.P., Antipov, A.N., Spring, S., Stackebrandt, E., and Bonch-Osmolovskaya, E.A. (2003). Caldithrix abyssi gen. nov., sp. nov., a nitrate-reducing, thermophilic, anaerobic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent, represents a novel bacterial lineage. International journal of systematic and evolutionary microbiology *53*, 323-29.

Molina, V., Belmar, L., and Ulloa, O. (2010). High diversity of ammonia-oxidizing archaea in permanent and seasonal oxygen-deficient waters of the eastern South Pacific. Environ Microbiol *12*, 2450-465.

Monteiro, P., Vanderplas, A., Melice, J., and Florenchie, P. (2008). Interannual hypoxia variability in a coastal upwelling system: Ocean–shelf exchange, climate and ecosystem-state implications. Deep Sea Research Part I: Oceanographic Research Papers *55*, 435-450.

Morris, R. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature *420*, 803-06.

Mulder, A., Graaf, A.A., Robertson, L.A., and Kuenen, J.G. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiol Ecol *16*, 177-184.

Murray, A.E., Preston, C.M., Massana, R., Taylor, L.T., Blakis, A., Wu, K., and DeLong, E.F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. Appl Environ Microbiol *64*, 2585-595.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., Horreard, F., Amann, R., Jørgensen, B.B., and Kuypers, M.M. (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. Proc Natl Acad Sci U S A *105*, 17861-66.

Naqvi, S.W.A., Bange, H.W., Farías, L., Monteiro, P.M.S., Scranton, M.I., and Zhang, J. (2010). Marine hypoxia/anoxia as a source of CH_4 and N_2O. Biogeosciences *7*, 2159-190.

Naqvi, S.W.A., Yoshinari, T., Jayakumar, D.A., Altabet, M.A., Narvekar, P.V., Devol, A.H., Brandes, J.A., and Codispoti, L.A. (1998). Budgetary and biogeochemical implications of N2O isotope signatures in the Arabian Sea *394*, 462-64.

Newman, D.K., and Banfield, J.F. (2002). Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. Science *296*, 1071-77.

Newman, M. (2010). Networks: an introduction (Oxford University Press, Inc.).

Newman, M.E.J. (2004). Analysis of weighted networks. Physical Review E *70*, 056131.

Norris, T.B., Wraith, J.M., Castenholz, R.W., and McDermott, T.R. (2002). Soil microbial community structure across a thermal gradient following a geothermal heating event. Appl Environ Microbiol *68*, 6300-09.

Orsi, W., Edgcomb, V., Jeon, S., Leslin, C., Bunge, J., Taylor, G.T., Varela, R., and Epstein, S. (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. II. Habitat specialization. ISME J *5*, 1357-373.

Orsi, W., Song, Y.C., Hallam, S., and Edgcomb, V. (2012). Effect of oxygen minimum zone formation on communities of marine protists. ISME J *6*, 1586-1601.

Paerl, H.W., and Pinckney, J.L. (1996). A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling. Microb Ecol *31*, 225-247.

Park, J., and Barabási, A.L. (2007). Distribution of node characteristics in complex networks. Proceedings of the National Academy of Sciences *104*, 17916-920.

Parsons, T.R., and Lalli, C.M. (1988). Comparative oceanic ecology of the plankton communities of the subarctic Atlantic and Pacific oceans. Oceanogr. Mar. Biol. Ann. Rev *26*, 317-359.

Paulmier, A., and Ruiz-Pino, D. (2009). Oxygen minimum zones (OMZs) in the modern ocean. Progress In Oceanography *80*, 113-128.

Paulmier, A., Ruiz-Pino, D., and Garcon, V. (2008). The oxygen minimum zone (OMZ) off Chile as intense source of CO2 and N2O. Continental Shelf Research *28*, 2746-756.

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

Peña, M.A., and Bograd, S.J. (2007). Time series of the northeast Pacific. Progress in Oceanography *75*, 115-19.

Peña, M.A., and Varela, D.E. (2007). Seasonal and interannual variability in phytoplankton and nutrient dynamics along Line P in the NE subarctic Pacific. Progress in Oceanography *75*, 200-222.

Pernthaler, A., Pernthaler, J., Amann, R., Kowalchuk, G.A., de Bruijn, F.J., Head, I.M., Akkermans, A.D.L., and Elsas, J.D.V. (2004). Sensitive multi-color fluorescence in situ hybridization for the identification of environmental microorganisms. Molecular microbial ecology manual. Volumes 1 and 2 , 711-725.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.

Prosser, J.I., and Nicol, G.W. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. Environ Microbiol *10*, 2931-941.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., and Glockner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res *35*, 7188-196.

Pruitt, K.D., and Maglott, D.R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res *29*, 137-140.

Rabalais, N.N., Diaz, R.J., Levin, L.A., Turner, R.E., Gilbert, D., and Zhang, J. (2010). Dynamics and distribution of natural and human-caused hypoxia. Biogeosciences *7*, 585-619.

Rabalais, N.N., Turner, R.E., and Wiseman, W.J. (2001). Hypoxia in the Gulf of Mexico. Journal of Environmental Quality *30*, 320-29.

Raes, J., and Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. Nature Reviews Microbiology *6*, 693-99.

Ramírez-Flandes, S., and Ulloa, O. (2008). Bosque: integrated phylogenetic analysis software. Bioinformatics *24*, 2539-541.

Rappé, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annual Reviews in Microbiology *57*, 369-394.

Rasko, D.A., Myers, G.S., and Ravel, J. (2005). Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics *6*, 2.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551-55.

Reinthaler, T., van Aken, H.M., and Herndl, G.J. (2010). Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. Deep Sea Research Part II: Topical Studies in Oceanography *57*, 1572-580.

Ribalet, F., Marchetti, A., Hubbard, K.A., Brown, K., Durkin, C.A., Morales, R., Robert, M., Swalwell, J.E., Tortell, P.D., and Armbrust, E.V. (2010). Unveiling a phytoplankton hotspot at a narrow boundary between coastal and offshore waters. Proceedings of the National Academy of Sciences *107*, 16571-76.

Ribalet, F., Marchetti, A., Hubbard, K.A., Brown, K., Durkin, C.A., Morales, R., Robert, M., Swalwell, J.E., Tortell, P.D., and Armbrust, E.V. (2010). Unveiling a phytoplankton hotspot at a narrow boundary between coastal and offshore waters. Proceedings of the National Academy of Sciences *107*, 16571-76.

Rich, V.I., Pham, V.D., Eppley, J., Shi, Y., and DeLong, E.F. (2011). Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray. Environ Microbiol *13*, 116-134.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., et al. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature *424*, 1042-47.

Royer, S.-J., Levasseur, M., Lizotte, M., Arychuk, M., Scarratt, M.G., Wong, C.S., Lovejoy, C., Robert, M., Johnson, K., et al. (2010). Microbial dimethylsulfoniopropionate (DMSP) dynamics along a natural iron gradient in the northeast subarctic Pacific. Limnology and Oceanography *55*, 1614-626.

Ruan, Q., Dutta, D., Schwalbach, M.S., Steele, J.A., Fuhrman, J.A., and Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics *22*, 2532-38.

Samuelsson, M.-O., and Rönner, U. (1982). Ammonium production by dissimilatory nitrate reducers isolated from Baltic sea water, as indicated by 15N study. Appl Environ Microbiol *44*, 1241-43.

Santoro, A.E., Buchwald, C., McIlvin, M.R., and Casciotti, K.L. (2011). Isotopic signature of N(2)O produced by marine ammonia-oxidizing archaea. Science *333*, 1282-85.

Sarmiento, J.L., Slater, R., Barber, R., Bopp, L., Doney, S.C., Hirst, A.C., Kleypas, J., Matear, R., Mikolajewicz, U., and Monfray, P. (2004). Response of ocean ecosystems to climate warming. Global Biogeochemical Cycles *18*, 3001-023.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ Microbiol *11*, 2078-093.

Schink, B. (2002). Synergistic interactions in the microbial world. Antonie Van Leeuwenhoek *81*, 257-261.

Schloss, P.D., and Westcott, S.L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol *77*, 3219-226.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol *75*, 7537-541.

Schmittner, A., Oschlies, A., Matthews, H.D., and Galbraith, E.D. (2008). Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual CO2 emission scenario until year 4000 AD. Global Biogeochemical Cycles *22*, GB1013.

Schönhuber, W., Fuchs, B., Juretschko, S., and Amann, R. (1997). Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. Appl Environ Microbiol *63*, 3268-273.

Schröder, I., Kröger, A., and Macy, J.M. (1988). Isolation of the sulphur reductase and reconstitution of the sulphur respiration of Wolinella succinogenes. Arch Microbiol *149*, 572-79.

Schwalbach, M.S., Tripp, H.J., Steindler, L., Smith, D.P., and Giovannoni, S.J. (2010). The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. Environ Microbiol *12*, 490-500.

Scranton, M.I., Astor, Y., Bohrer, R., Ho, T.Y., and Muller-Karger, F. (2001). Controls on temporal variability of the geochemistry of the deep Cariaco Basin. Deep Sea Research Part I: Oceanographic Research Papers *48*, 1605-625.

Shaffer, G., Olsen, S.M., and Pedersen, J.O.P. (2009). Long-term ocean oxygen depletion in response to carbon dioxide emissions from fossil fuels. Nature Geoscience *2*, 105-09.

Shanks, A.L., and Reeder, M.L. (1993). Reducing microzones and sulfide production in marine snow. Marine Ecology Progress Series *96*, 43-47.

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

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., Polz, M.F., and Alm, E.J. (2012). Population genomics of early events in the ecological differentiation of bacteria. Science *336*, 48-51.

Sherry, N.D., Boyd, P.W., Sugimoto, K., and Harrison, P.J. (1999). Seasonal and spatial patterns of heterotrophic bacterial production, respiration, and biomass in the subarctic NE Pacific. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2557-578.

Siegenthaler, U., and Sarmiento, J.L. (1993). Atmospheric carbon dioxide and the ocean. Nature *365*, 119-125.

Simon, J. (2002). Enzymology and bioenergetics of respiratory nitrite ammonification. FEMS microbiology reviews *26*, 285-309.

Smethie Jr, W.M. (1987). Nutrient regeneration and denitrification in low oxygen fjords. Deep Sea Research Part A. Oceanographic Research Papers *34*, 983-1006.

Sogin, M., Morrison, H., Huber, J., Welch, D., Huse, S., Neal, P., Arrieta, J., and Herndl, G. (2006). Microbial Diversity in the Deep Sea and the Underexplored "Rare Biosphere". Proc Natl Acad Sci U S A *103*, 12115-120.

Stams, A.J., and Plugge, C.M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. Nature Reviews Microbiology *7*, 568-577.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E., Sachdeva, R., Jones, A.C., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J *5*, 1414-425.

Stepanauskas, R. (2012). Single cell genomics: an individual look at microbes. Curr Opin Microbiol *15*, 613-620.

Stevens, H., and Ulloa, O. (2008). Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. Environ Microbiol *10*, 1244-259.

Stewart, F.J., Newton, I.L., and Cavanaugh, C.M. (2005). Chemosynthetic endosymbioses: adaptations to oxic--anoxic interfaces. Trends Microbiol *13*, 439-448.

Stewart, F.J., Ulloa, O., and Delong, E.F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environ Microbiol *14*, 23-40.

Stramma, L., Johnson, G.C., Sprintall, J., and Mohrholz, V. (2008). Expanding Oxygen-Minimum Zones in the Tropical Oceans. Science *320*, 655-58.

Stramma, L., Schmidtko, S., Levin, L.A., and Johnson, G.C. (2010). Ocean oxygen minima expansions and their biological impacts. Deep Sea Research Part I: Oceanographic Research Papers *57*, 587-595.

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., et al. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature *440*, 790-94.

Sunamura, M., Higashi, Y., Miyako, C., Ishibashi, J.-I., and Maruyama, A. (2004). Two bacteria phylotypes are predominant in the Suiyo Seamount hydrothermal plume. Appl Environ Microbiol *70*, 1190-98.

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

Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N.J., Masland, E.D.P., et al. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. Science *333*, 1296-1300.

Sztukowska, M., Bugno, M., Potempa, J., Travis, J., and Kurtz, D.M. (2002). Role of rubrerythrin in the oxidative stress response of Porphyromonas gingivalis. Mol Microbiol *44*, 479-488.

Sørensen, J. (1987). Nitrate reduction in marine sediment: pathways and interactions with iron and sulfur cycling. Geomicrobiology Journal *5*, 401-421.

Sørensen, K.B., and Teske, A. (2006). Stratified communities of active archaea in deep marine subsurface sediments. Appl Environ Microbiol *72*, 4596-4603.

Takai, K., and Horikoshi, K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. Genetics *152*, 1285-297.

Takai, K., Komatsu, T., Inagaki, F., and Horikoshi, K. (2001). Distribution of archaea in a black smoker chimney structure. Appl Environ Microbiol *67*, 3618-629.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res *29*, 22-28.

Taupp, M., Lee, S., Hawley, A., Yang, J., and Hallam, S.J. (2009). Large insert environmental genomic library production. J Vis Exp

Taylor, F.J.R., and Haigh, R. (1996). Spatial and temporal distributions of microplankton during the summers of 1992 1993 in Barkley Sound, British Columbia, with emphasis on harmful species. Canadian Journal of Fisheries and Aquatic Sciences *53*, 2310-322.

Tedersoo, L., Nilsson, R.H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G., and Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytologist *188*, 291-301.

Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., and Glockner, F.O. (2004a). Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ Microbiol *6*, 938-947.

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., and Glockner, F.O. (2004b). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. BMC Bioinformatics *5*, 163.

Tortell, P.D., Maldonado, M.T., and Price, N.M. (1996). The role of heterotrophic bacteria in iron-limited ocean ecosystems. Nature *383*, 330-32.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. Environ Microbiol *7*, 1985-995.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W.H., Wilhelm, L.J., and Giovannoni, S.J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature *452*, 741-44.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., and Pochart, P. (2000). A comprehensive analysis of protein--protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-27.

Ulloa, O., and Pantoja, S. (2009). The oxygen minimum zone of the eastern South Pacific. Deep-Sea Res Pt Ii *56*, 987-991.

Ulloa, O., Canfield, D.E., Delong, E.F., Letelier, R.M., and Stewart, F.J. (2012). Microbial oceanography of anoxic oxygen minimum zones. Proc Natl Acad Sci U S A *109*, 15996-16003.

Ulloa, O., Wright, J.J., Belmar, L., and Hallam, S.J. (2013). Pelagic Oxygen Minimum Zone Microbial Communities. In The Prokaryotes (Berlin, Heidelberg: Springer Berlin Heidelberg).

Urbach, E., and Chisholm, S.W. (1998). Genetic diversity in Prochlorococcus populations flow cytometrically sorted from the Sargasso Sea and Gulf Stream. Limnology and oceanography *43*, 1615-630.

Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (1998). Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of Prochlorococcus and Synechococcus (Cyanobacteria). J Mol Evol *46*, 188-201.

Vanoni, M.A., and Curti, B. (2008). Structure--function studies of glutamate synthases: A class of self-regulated iron-sulfur flavoenzymes essential for nitrogen assimilation. IUBMB life *60*, 287-300.

Vaquer-Sunyer, R., and Duarte, C.M. (2008). Thresholds of hypoxia for marine biodiversity. Proc Natl Acad Sci U S A *105*, 15452-57.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66-74.

Vetriani, C., Jannasch, H.W., MacGregor, B.J., Stahl, D.A., and Reysenbach, A.-L. (1999). Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. Appl Environ Microbiol *65*, 4375-384.

Vetriani, C., Tran, H.V., and Kerkhof, L.J. (2003). Fingerprinting microbial assemblages from the oxic/anoxic chemocline of the Black Sea. Appl Environ Microbiol *69*, 6481-88.

Wagner, A., and Fell, D.A. (2001). The small world inside large metabolic networks. Proceedings of the Royal Society of London. Series B: Biological Sciences *268*, 1803-810.

Walker, C.B., De La Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., Brochier-Armanet, C., Chain, P.S.G., Chan, P.P., et al. (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proceedings of the National Academy of Sciences *107*, 8818-823.

Wallner, G., Amann, R., and Beisker, W. (1993). Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry *14*, 136-143.

Walsh, D.A., and Hallam, S.J. (2011). Bacterial Community Structure and Dynamics in a Seasonally Anoxic Fjord: Saanich Inlet, British Columbia. In Molecular Microbial Ecology II: Metagenomics in Different Habitats, F.J. de Bruijn, eds. (Hoboken, New Jersey: John Wiley & Sons, Inc.).

Walsh, D.A., Zaikova, E., and Hallam, S.J. (2009). Large volume (20L+) filtration of coastal seawater samples. J Vis Exp

Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., Tortell, P.D., and Hallam, S.J. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science *326*, 578-582.

Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G., Zhang, C., Song, Z., Zhang, Y., and Ren, H. (2013). Control of Temperature on Microbial Community Structure in Hot Springs of the Tibetan Plateau. PLoS One *8*, e62901.

Wang, W., Duan, D., Liu, L., Yang, Y., Gu, G., and Mu, M. (2012). Molecular analysis of the microbial community structures in water-flooding petroleum reservoirs with different temperatures. Biogeosciences Discussions *9*, 5177-5203.

Ward, B.B., and Zafiriou, O.C. (1988). Nitrification and nitric oxide in the oxygen minimum of the eastern tropical North Pacific. Deep Sea Research Part A. Oceanographic Research Papers *35*, 1127-142.

Ward, B.B., Devol, A.H., Rich, J.J., Chang, B.X., Bulow, S.E., Naik, H., Pratihary, A., and Jayakumar, A. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. Nature *461*, 78-U77.

Ward, B.B., Kilpatrick, K.A., Wopat, A.E., Minnich, E.C., and Lindstrom, M.E. (1989). Methane oxidation in Saanich Inlet. Cont. Shelf Res. *9*, 65-75.

Wasserman, S., and Galaskiewicz, J. (1994). Advances in social network analysis: Research in the social and behavioral sciences (SAGE Publications, Incorporated).

Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature *393*, 440-42.

Whitney, F.A., and Freeland, H.J. (1999). Variability in upper-ocean water properties in the NE Pacific Ocean. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2351-370.

Whitney, F.A., Freeland, H.J., and Robert, M. (2007). Persistently declining oxygen levels in the interior waters of the eastern subarctic Pacific. Progress in Oceanography *75*, 179-199.

Whitney, F.A., Wong, C.S., and Boyd, P.W. (1998). Interannual variability in nitrate supply to surface waters of the Northeast Pacific Ocean. Marine Ecology Progress Series *170*, 15-23.

Wilhelm, L.J., Tripp, H.J., Givan, S.A., Smith, D.P., and Giovannoni, S.J. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. Biol Direct *2*, 27.

Woebken, D., Fuchs, B.M., Kuypers, M.M., and Amann, R. (2007). Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. Appl Environ Microbiol *73*, 4648-657.

Woebken, D., Lam, P., Kuypers, M.M., Naqvi, S.W., Kartal, B., Strous, M., Jetten, M.S., Fuchs, B.M., and Amann, R. (2008). A microdiversity study of anammox bacteria reveals a novel Candidatus Scalindua phylotype in marine oxygen minimum zones. Environ Microbiol *10*, 3106-119.

Wong, C.S., and Matear, R.J. (1999). Sporadic silicate limitation of phytoplankton productivity in the subarctic NE Pacific. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2539-555.

Wong, C.S., Whitney, F.A., Crawford, D.W., Iseki, K., Matear, R.J., Johnson, W.K., Page, J.S., and Timothy, D. (1999). Seasonal and interannual variability in particle fluxes of carbon, nitrogen and silicon from time series of sediment traps at Ocean Station P, 1982--1993: relationship to changes in subarctic primary productivity. Deep Sea Research Part II: Topical Studies in Oceanography *46*, 2735-760.

Wong, C.S., Whitney, F.A., Iseki, K., Page, J.S., and Zeng, J. (1995). Analysis of trends in primary productivity and chlorophyll-a over two decades at Ocean Station P (50 N, 145 W) in the subarctic northeast Pacific Ocean. Canadian Special Publication of Fisheries and Aquatic Sciences , 107-117.

Wong, C.-S., Wong, S.-K.E., Peña, A., And Levasseur, M. (2006). Climatic effect on DMS producers in the NE sub-Arctic Pacific: ENSO on the upper ocean. Tellus B *58*, 319-326.

Woyke, T., Xie, G., Copeland, A., Gonzalez, J.M., Han, C., Kiss, H., Saw, J.H., Senin, P., Yang, C., et al. (2009). Assembling the marine metagenome, one cell at a time. PLoS One *4*, e5299.

Wright, J.J., Konwar, K.M., and Hallam, S.J. (2012). Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol *10*, 381-394.

Wright, J.J., Lee, S., Zaikova, E., Walsh, D.A., and Hallam, S.J. (2009). DNA extraction from 0.22 microM Sterivex filters and cesium chloride density gradient centrifugation. J Vis Exp

Wright, T.D., Vergin, K.L., Boyd, P.W., and Giovannoni, S.J. (1997). A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. Appl Environ Microbiol *63*, 1441-48.

Wuchter, C., Abbas, B., Coolen, M.J., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G.J., et al. (2006). Archaeal nitrification in the ocean. Proc Natl Acad Sci U S A *103*, 12317-322.

Yook, S.H., Oltvai, Z.N., and Barabási, A.L. (2004). Functional and topological characterization of protein interaction networks. Proteomics *4*, 928-942.

Zaikova, E., Hawley, A., Walsh, D.A., and Hallam, S.J. (2009). Seawater sampling and collection. J Vis Exp

Zaikova, E., Walsh, D.A., Stilwell, C.P., Mohn, W.W., Tortell, P.D., and Hallam, S.J. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Environ Microbiol *12*, 172-191.

Zehnder, A.J., and Stumm, W. (1988). Geochemistry and biogeochemistry of anaerobic habitats. In Biology of anaerobic microorganisms, A.J. Zehnder, eds. (New York: John Wiley & Sons).

Zhao, F., and Qin, S. (2007). Comparative molecular population genetics of phycoerythrin locus in Prochlorococcus. Genetica *129*, 291-99.

Zhou, J., Deng, Y., Luo, F., He, Z., and Yang, Y. (2011). Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. MBio *2*

*Z*umft, W.G. (1997). Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev *61*, 533.

Appendices

Appendix A In silico binding efficiency of probe SAR406-97 with full-length MGA 16S rRNA gene clone sequences from Line P (Chapter 3).

Appendix A *In silico* efficiency of probe SAR406-97† at binding MGA sequences from Line P

OTU	E-4	E-3	E-2	E-1	E0	# of sequences in OTU
MGA_03	97.92			2.08		48
MGA_09	100.00					26
MGA_08	100.00					22
MGA_07	95.24			4.76		21
MGA_13	100.00					14
MGA_15	8.33	91.67				12
MGA_19	100.00					6
MGA_20		20.00	60.00		20.00	5
MGA_21	100.00					5
MGA_22	100.00					5
MGA_23	100.00					3
MGA_06		100.00				2
MGA_24			100.00			2
MGA_25	100.00					2
MGA_26	100.00					2
MGA_27		100.00				2
MGA_28	100.00					2
MGA_29		100.00				2
MGA_30	100.00					2
MGA_31	100.00					2
MGA_32			100.00			2
MGA_33	100.00					2
MGA_34	100.00					2
MGA_35	100.00					2
MGA_17	100.00					1
MGA_36		100				1
MGA_37	100.00					1
MGA_38				100.00		1
MGA_39				100.00		1
MGA_40					100.00	1
MGA_41					100.00	1
MGA_42				100.00		1
MGA_43	100.00					1
MGA_44	100.00					1
MGA_45					100.00	1
MGA_46	100.00					1
MGA_47					100.00	1
MGA_48	100.00					1
MGA_49	100.00					1
MGA_50					100.00	1
MGA_51				100.00		1
MGA_52	100.00					1
MGA_53				100.00		1
MGA_54	100.00					1
MGA_55	100.00					1
MGA_56					100.00	1
MGA_57				100.00		1
MGA_58			100.00			1
MGA_59		100.00				1
MGA_60	100.00					1
MGA_61	100.00					1
MGA_62	100.00					1
MGA_63					100.00	1
MGA_64		100.00				1
MGA_65				100.00		1
MGA_66	100.00					1
MGA_67				100.00		1
MGA_68				100.00		1
MGA_69	100.00					1
(continued...)						
% of MGA sequences hit at E value	76.21	9.31	3.79	6.21	4.14	290

†Fuchs et al. 2005
*E value categories defined as follows:
E-4: Up to 1 missing 3' base, no mismatches
E-3: Up to 3 missing 5' bases or 1 mismatch
E-2: Up to 4 missing 5' or 3' bases and 1 mismatch
E-1: Up to 5 missing 5' or 3' bases or 2 mismatches
E0: Up to 5 missing 5' or 3' bases and 1 mismatch

Appendix A: *In silico* efficiency of probe SAR406-97† at binding MGA sequences from Line P

OTU	E-4	E-3	E-2	E-1	E0	# of sequences in OTU
(continued from Table S1a)						
MGA_70	100.00					1
MGA_71	100.00					1
MGA_72				100.00		1
MGA_73			100.00			1
MGA_74			100.00			1
MGA_75	100.00					1
MGA_76	100.00					1
MGA_77	100.00					1
MGA_78	100.00					1
MGA_79	100.00					1
MGA_80	100.00					1
MGA_81	100.00					1
MGA_82					100.00	1
MGA_83	100.00					1
MGA_84	100.00					1
MGA_85	100.00					1
MGA_86	100.00				Table	1
MGA_87	100.00					1
MGA_88	100.00					1
MGA_89						1
MGA_90	100.00					1
MGA_91		100.00				1
MGA_92		100.00				1
MGA_93	100.00					1
MGA_94	100.00					1
MGA_95	100.00					1
MGA_96	100.00					1
MGA_97		100.00				1
MGA_98	100.00					1
MGA_99	100.00					1
MGA_100	100.00					1
MGA_101	100.00					1
MGA_102	100.00					1
MGA_103				100.00		1
MGA_104			100.00			1
MGA_105				100.00		1
MGA_106	100.00					1
MGA_107	100.00					1
MGA_108		100.00				1
MGA_109	100.00					1
MGA_110	100.00					1
MGA_111	100.00					1
MGA_112	100.00					1
MGA_113	100.00					1
MGA_114	100.00					1
MGA_115	100.00					1
MGA_116		100.00				1
MGA_117					100.00	1
MGA_118	100.00					1
MGA_119	100.00					1
MGA_120		100.00				1
MGA_121				100.00		1
MGA_122	100.00					1
MGA_123	100.00					1
MGA_124	100.00					1
MGA_125					100.00	1
MGA_126	100.00					1
MGA_127	100.00					1
MGA_128				100.00		1
MGA_129				100.00		1
MGA_130				100.00		1
MGA_131					100.00	1
% of MGA sequences hit at E value	76.21	9.31	3.79	6.21	4.14	290

†Fuchs et al. 2005
*E value categories defined as follows:
E-4: Up to 1 missing 3' base, no mismatches
E-3: Up to 3 missing 5' bases or 1 mismatch
E-2: Up to 4 missing 5' or 3' bases and 1 mismatch
E-1: Up to 5 missing 5' or 3' bases or 2 mismatches
E0: Up to 5 missing 5' or 3' bases and 1 mismatch

Appendix B Primers for verification of IonTorrent sequencing errors on select fosmids (Chapter 4)

Appendix B: Primers for verification of IonTorrent sequencing errors on select fosmids

Primer name	Forward	Reverse	Fosmid Target
MGA_1	ggactccatccatacccaca	cagcagctgtccgttcatta	4130011-I07
MGA_2	acgttctatccgcagcaagt	ccatgctgattaaggggcta	125003-E23
MGA_3	gaaaggcagttttcaacatgg	gcaacagcaatggcatctaa	413004-H17
MGA_4	aggccaatttggatgtgaaa	gcggggaaattagatcgttt	413004-H17
MGA_5	accggggatctaaaggagaa	caatgcagaaacgcaatgtt	413004-H17
MGA_6	gccaactgcaacaccctatt	cttccatggtcgctggttat	405006-B04
MGA_7	tttggccggaacttgaatac	tcagcgtgtttcctgtgaac	405006-B04
MGA_8	gggctagggagaagccatac	aaatggtggtcgcaatgatg	405006-B04
MGA_9	ccgatgagccagataccataa	gtctgcaataccgccaagat	405006-B04
MGA_10	tgaaattggcgttgcatcta	gatatgaccacggggtgttt	405006-B04
MGA_11	ggaacgagactgcctactgg	gcccctgtaagaccaggaat	122006-I05