Integrated –omics study of deep-sea microbial community and new *Pseudoalteromonas* isolate

by

Jieying Wu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the degree
Doctor of Philosophy

Approved September 2013 by the
Graduate Supervisory Committee:

Deirdre R. Meldrum, Chair
Weiwen Zhang
Susanne Neuer
Morteza Abbaszadegan

ARIZONA STATE UNIVERSITY

December 2013

ABSTRACT

This thesis research focuses on phylogenetic and functional studies of microbial communities in deep-sea water, an untapped reservoir of high metabolic and genetic diversity of microorganisms. The presence of photosynthetic cyanobacteria and diatoms is an interesting and unexpected discovery during a 16S ribosomal rRNA-based community structure analyses for microbial communities in the deep-sea water of the Pacific Ocean. Both RT-PCR and qRT-PCR approaches were employed to detect expression of the genes involved in photosynthesis of photoautotrophic organisms. Positive results were obtained and further proved the functional activity of these detected photosynthetic microbes in the deep-sea. Metagenomic and metatranscriptomic data was obtained, integrated, and analyzed from deep-sea microbial communities, including both prokaryotes and eukaryotes, from four different deep-sea sites ranging from the mesopelagic to the pelagic ocean. The RNA/DNA ratio was employed as an index to show the strength of metabolic activity of deep-sea microbes. These taxonomic and functional analyses of deep-sea microbial communities revealed a 'defensive' life style of microbial communities living in the deep-sea water. *Pseudoalteromonas* sp.WG07 was subjected to transcriptomic analysis by application of RNA-Seq technology through the transcriptomic annotation using the genomes of closely related surface-water strain *Pseudoalteromonas haloplanktis* TAC125 and sediment strain *Pseudoalteromonas* sp. SM9913. The transcriptome survey and related functional analysis of WG07 revealed unique features different from TAC125 and SM9913 and provided clues as to how it

adapted to its environmental niche. Also, a comparative transcriptomic analysis of WG07 revealed transcriptome changes between its exponential and stationary growing phases.

ACKNOWLEDGEMENTS

One of the pleasures of completion is to look over the journey past and remember all the friends and family who have helped and supported me along this long but fulfilling road.

I would like to express my heartfelt gratitude to my mentors Professor Deirdre Meldrum and Dr.Weiwen Zhang. I could not have asked for better role models, each inspirational, supportive, and patient. I could not be prouder of my academic roots and hope that I can in turn pass on the research values and the dreams that they have given to me.

I would also like to thank my examiners, Professor Susanne Neuer, Professor Morteza Abbaszadegan, and Dr.Weimin Gao, who provided encouraging and constructive feedback. It is not an easy task, reviewing a thesis, and I am grateful for their thoughtful and detailed comments. To the many anonymous reviewers at the various journals and conferences, thank you for helping to shape and guide the direction of the work with your careful and instructive comments.

I am grateful to Professor John Delaney and Professor Deborah Kelley of the University of Washington. They were the Chief Scientists for the INSITE08 National Science Foundation (NSF) Ocean Observatories Initiative (OOI) mapping cruise on the research vessel (R/V) Thompson TN 221 in the summer 2008. The samples used for this PhD research were obtained as part of this cruise in the NE Pacific Ocean (http://ooi.ocean.washington.edu/cruise/).

I would like to thank Arizona State University NEPTUNE fund for their generous support. As a member involved in NEPTUNE project in CBDA, Biodesign Institute, I

have been surrounded by wonderful colleagues and a rich and fertile environment to study and explore new ideas. To the staff and students at CBDA, I am grateful for the chance to join and be a part of the lab. Thank you for welcoming me as a friend and helping to develop the ideas in this thesis.

I would not have contemplated this road if not for my parents, who instilled within me a love of creative pursuits, science and language, all of which finds a place in this thesis, and my younger brother, to whom I would like to show my passion and my happiness toward the life. To my dear family, thank you. My best friend Mingji, with whom I've been brabbling and reconciling in the past 24 years, has also been the best companion along this journey. Thank you for being you.

To my dear comrades in the Biodesign Insititute, Xu, Jia, Bo, Saeed, and Xiaofang, thanks for being there to listen and console. Thank Dr.Yanqing Tian and Dr. Fengyu Su, and many other friends in the various households that have sheltered me over the years. You have been like surrogate families, bearing the brunt of the frustrations, and sharing in the joy of the successes.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figures Page

CHAPTER 1
BACKGROUND LITERATURE

The water body underlying the photic zone in the world's oceans (depths > 200 m), representing the largest water mass on earth (comprising 1.3×10^{18} m^3), is the largest aqueous habitat for microbial life (Orcutt et al. 2011). This fraction of the ocean is believed to harbor 6.5×10^{28} cells which accounts for about 55% of all the prokaryotes found in aquatic habitats (Whitman et al. 1998). Depth 200 m is an average division between deep and Surface Ocean. Sun light can hardly penetrate beyond this depth, which causes the decrease of the light-dependent primary production. The realm of the ocean below 200 m is also different distinctly from the photic zone, presenting low temperature (approximately 2~4^{o}C), higher pressure (pressure increases by roughly 1 atm with each 10 m of depth) and higher inorganic nutrient levels (Aristegui et al. 2009). Inorganic compounds (NO^{3+}, PO_4^{3-}, Fe^{3+}, etc.) rarely exist in the photic zone, while they are re-mineralized and released into the water by microbial activity in the aphotic zone which makes the concentrations of these compounds increase in the depths below 200 m (Orcutt et al. 2011). In the middle depths of the ocean, the oxygen concentration is typically lower than that in the surface water, mainly due to the decomposition of sinking organic material by aerobic respiration as well as the introduction of high-latitude, oxygen-rich surface water to deeper depths (Wyrtki 1962; Ulloa et al. 2012). However, oxygen is in general relatively abundant for the microbial population living in those depths, and the oxygen-direct respiration is often observed in the microorganisms around this depth (Orcutt et al. 2011).

Differences in physical geochemical parameters between the upper level of sea water and deep sea suggest that microbial communities in these environments are confronted by fundamentally different challenges. Our knowledge of microbial processes below the sunlit surface ocean has been enormously expanded in the most recent 30 years, especially owing to the stirring discoveries of hydrothermal vent (Prieur et al. 1995; Anantharaman et al. 2013), cold seeps (Fang et al. 2006; Orcutt et al. 2008), and whale falls (Deming et al. 1997; Smith 2003). These discoveries, as well as numerous studies being performed continuously, make great contributions to uncover the veil over the microorganisms living in the dark ocean where we cannot easily reach for, generate continuous and paradigm-shifting findings, and fundamentally change our thoughts toward deep-sea microbial communities (Orcutt et al. 2011). Study from the Global Ocean Sampling (GOS) expedition indicated that new microorganism families are being discovered at a rate that is linear or almost linear with the addition of new sequences, implying that we are still far from discovering all protein families in nature. It was also predicted that marine microorganisms will continue to be a source of novel enzymes in the near future (Rusch et al. 2007).

Many marine microorganisms can synthesize various metabolic compounds. In recent decades (since 1970s), a considerable amount of drug candidates were from marine natural products (Haefner 2003). These natural products are a rich source of new chemical diversity and also a vital component of pharmaceutical industry (Fenical & Jensen 2006). Hence, the development of tools to access deep-sea microorganisms and microbial community promises to provide insight into this significant new source of drug

discovery and developing. In this chapter, we introduce two leading approaches of studying microbial communities, metagenomics and metatranscriptomics, and their applications in exploring deep-sea microbial communities. Unique features and genetic adaptations of microbial communities living in this special environment are also discussed.

Metagenomics: a tool to catalog marine microbial communities

Biotechnology applications are fueling the exploration of microbial resources in the oceans. Although many studies have led to cultivation, characterization and application of various ocean microorganisms, it is now widely accepted that the overwhelming majority (i.e. 99%) of the microorganisms in the ocean are uncultivable and thus not accessible by traditional microbiological methodologies (Rappé et al. 2002). To address these issues, various culture-independent approaches have been proposed to explore microbial communities in the oceans (Schmidt et al. 1991; Poretsky et al. 2005; Parro et al. 2007). These techniques allow analysis of microbial communities without the necessity to cultivate the microorganisms, and have greatly enlarged the fraction of microorganisms that can be identified (Fuhrman 2002). The traditional culture-independent techniques include cloning and sequencing of taxonomically indexed genes (Pace 1997), denaturant gradient gel electrophoresis (DGGE) (Muyzer & Smalla 1998), fluorescent in situ hybridization (FISH) (Amann et al. 1995), qPCR (Head et al. 1998), and DNA microarray (Giovannoni et al. 1990). Recent progress led to some new techniques such as 16S ribosomal RNA (rRNA) target sequencing which has offered a glimpse into the phylogenetic diversity of uncultured organisms (Tringe et al. 2005), and

a new era based on next generation sequencing. Among the methods designed to approach the phylogenetics of uncultured organisms, metagenomics, the genomic analysis of a population of microorganisms, has emerged as a powerful centerpiece (Handelsman 2004).

Since the first proposal by Pace et al. that microorganisms could be studied by extracting the mixed microbial nucleic acid from environmental samples and then analyzing DNA sequences (Pace et al. 1986), the term 'metagenomics' as a game-changing concept opened a new era in environmental microbial exploration. Metagenomics is defined as the study of the direct-isolated genetic components of a given environmental sample (Kennedy et al. 2010). The following analysis of metagenomics can be either sequence-driven or function-driven (Handelsman 2004; Kennedy et al. 2010). Initially the technology involves environmental DNA isolation, DNA fragment cloning and transforming, and the resulting transformants screened and sequenced (Handelsman 2004). However, it is low in efficiency and very labor intensive and hard to achieve good coverage for genetic material isolated from environmental microbial communities. Recent developments in high-throughput and more affordable sequencing platforms have the advantages of high efficiency, labor savings, and high coverage. To date, several commercially available high-throughput sequencing platforms have been introduced, including Sanger sequencing using chain termination method (Orsi et al. 2011; Youssef et al. 2012), 454 pyrosequencing (Rusch et al. 2007; Eloe et al. 2011; Wu et al. 2011), Illumina sequencing using synthesis sequencing technology (Shakya et al. 2013), SOLiD (Sequencing by Oligonucleotide Ligation and Detection) sequencing by

ligation (Valouev et al. 2008), and Ion Torrent sequencing using ion semiconductor technology (Whiteley et al. 2012), etc. These technologies can generate 1~5 million reads in a single run, with read lengths ranging from 50 ~ 1,000 bps. In recent years, these high-throughput sequencing technologies are fueling a rapid increase in the number and scope of microbial community-targeted studies of marine environment using metagenomic approaches. Genes involved in photosynthesis, carbon fixation, and nitrogen metabolism were found to be highly represented in the cDNA pool of surface water microbial communities (Frias-Lopez et al. 2008). Comparative genomic analyses of stratified planktonic microbial communities in the ocean's interior revealed a depth-dependent trend of microbial community function in the North Pacific Subtropical Gyre. Genes involved in carbon and energy metabolism, attachment and motility, gene mobility, and host-viral interactions were found to be more enriched in the photic surface water (0~200 m) (DeLong et al. 2006). Comparative metagenomic analysis showed that deep-sea microbial communities cope with their dwelling environment by a collection of very subtle adaptations, instead of dramatic alterations of gene content or structure. The studies found that deep-sea prokaryotes own higher metabolic versatility and genomic plasticity, and they typically have larger genome sizes and higher content of transposases and prophages (Konstantinidis et al. 2009).

As mentioned before, there are two approaches for metagenomic analysis, sequence-based and function-based (Handelsman 2004). Sequence-based metagenomic approaches compare sequence data with previously constructed databases, while function-based metagenomics identify genes directly by their function. The sequence-

based analysis, by sequencing huge amount of random clones, can generate insight data for massive scale samples (e.g. deep-sea microbial community) and provide information about genomic component and features, distribution and redundancy of functions in the community, and horizontal gene transfer (Venter et al. 2004; Handelsman 2004). On the other hand, for function-based metagenomics, there is no need to recognize the targeted genes by sequences so that this kind of approach can play a role in novel function detection (Handelsman 2004; Kennedy et al. 2010). However, the significant limitation of function-based metagenomics is that function is always linked with the 'expressed' genes, which is a big burden to obtain complete information of the target. To date, DNA sequence based metagenomics that involves random community DNA sequencing, as well as taxonomic and functional gene diversity determination, has become a standard tool for the analysis of natural microbial communities in various natural environments including alkaline oil sands tailings ponds (Saidi-Mehrabad et al. 2012), acidic hot spring (Jiménez et al. 2012), wastewater treatment bioreactors (Ye et al. 2012), Antarctic soil (Pearce et al. 2012), and oxygen minimum zones in the ocean (Stewart et al. 2011; Cassman et al. 2012).

Metatranscriptome: insights into 'real-time' activity

Though metagenomic analysis is able to provide diversity information and functional potential of various samples including those of uncultured and unknown species (Ogura et al. 2011), it cannot provide informations regarding expression level of functional genes, hence not to mention the physiology of microbes under specific environmental conditions. To address this issue, metatranscriptomics was introduced into

this field. The major metatranscriptomics technologies include microarray (Parro et al. 2007; Abell et al. 2012) and mRNA-derived cDNA clone libraries (Poretsky et al. 2005). Microarray is a tool for the highly parallel hybridization of a single target to a multitude of probes in order to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome (Schena et al. 1995; Lashkari et al. 1997), while mRNA-derived cDNA clone library methods apply randomLy primed reverse transcription of enriched environmental mRNA to produce a cDNA template population and then to generate cDNA clone libraries (Poretsky et al. 2005). These technologies have been widely used and brought significant insights into the functionality of microbial communities. However, microarray analysis is limited to the known genes since it involves designing specific probes, while mRNA-derived cDNA clone library method may bring in additional biases, such as the relative abundance skewing of sequences because of the toxic products encoded by cloned transcripts (Gilbert et al. 2008). With the new next-generation sequencing technologies, current metatranscriptomics is similar to metagenomics except that it analyzes RNA samples. To date, it is possible to identify genes being transcribed in given conditions and further to depict gene expression patterns in natural microbial communities without any prior information (Gifford et al. 2011). Next-generation sequencing based metatranscriptomics has allowed assessment of microbial functionality in the natural environments (Shi et al. 2009). The technology can overcome various constraints inherent in the use of traditional qPCR and microarrays to monitor gene expression in complex microbial communities, such as a limited number of genes can be examined, a prior need to select gene targets,

and requirements for probes or primers (Gilbert et al. 2008). In addition, sequencing based metatranscriptomics has no significant bias towards any known sequences within microbial communities, and thus provides a new and powerful approach for identifying previously unknown genes and analyzing community-specific variants of functional genes in the nature (McGrath et al. 2008). In recent years, several metatranscriptomics analyses of marine and soil samples have been reported. The study made by Gilbert et al. (2008) provided a good example that when paired with metagenomic analysis, metatranscriptomic studies can offer unprecedented exploration of both structural and functional information of a surface-ocean microbial community. In another study of gene expression in ocean microbial communities, half of the genes detected were undetected in previous DNA-based metagenomic surveys (Frias-Lopez et al. 2008). The study also showed that genes involved in photosynthesis, carbon fixation, and nitrogen acquisition occupy the largest portion in the cDNA pool (Frias-Lopez et al. 2008). While most of these studies started from selective mRNA enrichment by depleting ribosomal RNA or enriching for polyA-tailed mRNA, in a recent study, a total RNA pool of microbial community, including both mRNA and rRNA, was analyzed to reveal both functionality and taxonomic diversity of the environments (Urich et al. 2008). This approach avoided extensive purification steps of mRNA, and enabled simple and quick preparation even from difficult samples. Together these studies have significantly enhanced our understanding of the diversity and functionality of microbial communities in the nature.

Defining the relationship between microbial community composition and function (metabolic characteristics) has been a major challenge in studying heterotrophic carbon

cycling in marine systems (Teske et al. 2011). The combination of metagenomic and metatranscriptomic approaches has proven to be a powerful tool in deciphering both the phylogenetic composition, and the metabolic activities of deep-sea microbial communities. In a recent study, coupled metagenomic and metatranscriptomic analyses were utilized for taxonomic and functional characterization of marine microbial communities living at depths between 25-500 m, and have found differential relative transcriptional activities per cell (Shi et al. 2011). Functional characterization of highly expressed genes in this study led to the finding of taxon-specific contributions to specific biogeochemical processes, such as *Roseobacter* sp. to aerobic anoxygenic phototrophy at 75 m, *Crenarchaea* to ammonia oxidation at 125 m, and transcriptionally active light-preferring *Prochlorococcus* population in the bottom of the photic zone (Shi et al. 2011). The results provided novel insights into not only microbial diversity but also specific metabolic processes transpiring in the ecosystems.

Technique challenges in applying metatranscriptomics

Limitations in sample quality and quantity have become big obstacles for applying metagenomic/metatranscriptomic technologies to explore deep-sea microbial communities (Wu et al. 2011). Compared with the biomass abundant samples, such as surface seawater, the intact nucleic acid that can be isolated from these environments is very limited in quantity, usually trace to several hundred nanograms level. Moreover, environmental samples, different from pure culture grown in the laboratory, often contain various minerals and humic acids, which may inhibit amplification reactions (Broemeling et al. 2008; Demeke & Jenkins 2010). Thus, in most of these cases, quality of the isolated

environmental nucleic acid is not comparable with that from laboratory grown pure culture either. Moreover, typical commercial next-generation sequencing platforms require 3-5 μg or more of input DNA/cDNA to produce reliable sequencing data, approximately equal to the total amount of DNA isolable from 10^9 Escherichia coli cells. In order to analyze these environmental samples, it is inevitable to perform amplification to fulfill amount requirement by the current pyrosequencing technologies. However, obtaining unbiased and evenly amplified whole metagenomic amplification products which are suitable for downstream sequencing is still difficult. Quality control is necessary before sending the products to final sequencing. Clone library construction combining randomLy picked-clone sequencing was proven to be a cost-effective method to examine the quality of the amplification products and to ensure the efficiency of next-generation sequencing (Wu et al. 2011). With more attention paid to community structure and functionality of microbial communities present in extreme environments, such as deep-sea, cold spring, and deep frozen earth where very low microbial cell density and high concentration of potential inhibitors are often expected, the protocol developed here could be very a useful tool in allowing direct next-generation based metatranscriptomic analysis. Related biotechnologies are still under development to enhance the whole metagenome amplification efficiency and to lower the starting mass of the next-generation sequencing platform to the nanogram or even hundred picogram level.

The key to a comprehensive study of the transcriptome is the identification of the full set of transcripts (Martin & Wang 2011). Despite the increasing ease of obtaining large quantities of DNA sequences during metatranscriptomic analysis, it is still a great

challenge to analyze the meta-data efficiently to obtain meaningful results, especially for those microorganisms without a high-quality reference genome (Singhal 2013). For the same reason, the metatranscriptomic analysis could be biased and partial since the analysis can only access DNA sequences with gene annotation or prediction (Martin & Wang 2011). Currently, the sequence reads obtained from next-generation sequencing platforms (except Sanger sequencing), including 454 pyrosequencing, Illumina, SOLiD, and Ion Torrent, are typically short (50 – 500 bps) (Metzker 2010), hence it is necessary to assemble short reads when reconstructing the total transcriptome, a complicated and challenging step that requires computing systems with large memories to run parallel algorithms. Although several de novo transcriptome assembly methods for single genomes have been developed (Zhao et al. 2011; Surget-Groba & Montoya-Burgos 2010), efforts are still needed for analysis of meta-data of complex microbial communities.

Deep-sea microbial community and its genetic adaptation

The biomass in the photic zone is remarkably higher than that in aphotic ocean since the photic surface ocean is the place where the majority of marine primary productions take place (Orcutt et al. 2011). On the other hand, the biomass concentration decreases by roughly 2 orders of magnitude in the aphotic ocean (Nagata et al. 2000). In the deep sea, microbial cell density can be as low as $10^3 – 10^4$ cells per mL of seawater (Kawagucci et al. 2012). Although cell abundances decrease with depth, the biodiversity of microbial communities in the deep-sea is surprisingly high. In a previous study, archaeal communities' diversity maintains a similar level or even higher level compared to that in upper level ocean (Winter et al. 2009). Deep-sea microbial community studies

indicated that the deep-sea lifestyle is composed of a series of very subtle adaptations to a cold, high-pressure environment, instead of dramatic alterations of gene content or structure. Lauro et al. (2007) compared several deep-sea psychrophilic and barophilic bacterial isolates including *Moritella* sp. PE36 (Delong et al. 1997), *Psychromonas* sp. CNPT3 (Delong et al. 1997), *Photobacterium profundum* SS9 (Vezzi et al. 2005), *Carnobacterium* sp. AT7 (Lauro et al. 2007), and *Shewanella* sp. KT99 (Lauro et al. 2007) with their corresponding surface-water strains, and have found that the deep-sea psychrophilic and barophilic strains typically possess higher rRNA operon copies, genome size or above-average intergenic regions, all of which are suggestive of a rapid response to environmental changes (Klappenbach et al. 2000; Lauro & Bartlett 2008). Other deep-sea specific features include lower membrane saturation to maintain cell plasticity (Allen et al. 1999; Allen & Bartlett 2002), starvation endurance to fight against oligotrophic dwelling conditions (Bidle & Bartlett 2001; Mahdhi et al. 2012), motility to quest for nutrients (Wirsen & Molyneaux 1999; Wang et al. 2008), and pressure-sensitive flagella assembly and motor systems (Shibata et al. 2005; Qin et al. 2011).

Using high-throughput sequencing data, Sogin et al (2006) provided a comprehensive survey of the vertical zonation of phylogenetic distribution of microbial communities in the North Pacific Ocean and found that the percentage of *Gammaproteobacteria* increases with depth. Similar findings were also obtained in the Mediterranean Sea and the Greenland Sea (Zaballos et al. 2006). In another study, bacteria *Prochlorococcus*, *Verrucomicrobiales*, *Flexibacteraceae*, *Gammaproteobacteria*, *Alphaproteobacteria*, and *Deltaproteobacteria* were found to be dominant in the ocean's

surface water, while *Deferribacteres*; *Planctomycetaceae*, *Acidobacteriales*, *Gemmatamonadaceae*, *Nitrospina*, *Alteromonadaeceae* were enriched in deep-sea water (DeLong et al. 2006). In this thesis (Wu et al. 2013), we found that chemolithoautotrophic and major nitrifier *Thaumarchaeota*, *Betaproteobacteria* and *Gammaproteobacteria*, which were recognized as ammonia oxidizing archaea (AOA) and ammonia oxidizing bacteria (AOB) that oxidize ammonia to nitrite (Pester et al. 2012), are enriched in the deep sea. Many of these AOA and AOB have a high ammonia affinity and can grow in extremely oligotrophic environments (Schleper & Nicol 2010). *Planctomycetes* are main members of the well-known anammox bacterial genus identified previously in the marine sub-oxic zone including *Rhodopirellula*, *Blastopirellula*, *Planctomyces*, *Pirellula*, *Candidatus Kuenenia*, *Gemmata* and *Isosphaera* (Kirkpatrick et al. 2006; Byrne et al. 2009). The presence of *Planctomycetes* is also an indicator that anaerobic ammonium oxidation (anammox), a globally important microbial process of the nitrogen cycle, may be another metabolic pathway supporting primary production in the low-oxygen, dark pelagic ocean. *Actinobacteria* and *Firmicutes*, two commonly seen deep-sea members, are believed to have adaptive advantage under low-nutrient conditions of the deep sea (Gärtner et al. 2011).

Except for the phylogenetic features of deep-sea microbial communities, microorganisms living in deep-sea environments also possess special metabolic adaptation to their environment. In previous studies, bacteria living in the deep-sea have been shown to cope with the high hydrostatic pressure environment mainly through membrane phospholipid fatty acid unsaturation (Allen et al. 1999; Nagata et al. 2000;

Allen & Bartlett 2002) and cell division (Bidle & Bartlett 2001; Ishii et al. 2005). Also, hydrostatic pressure was proven to be able to affect microorganisms' cellular processes including replication and translation (Welch et al. 1993; Ishii et al. 2005). By employing metagenomic and metatranscriptomic analysis, it has been found that in the deep-sea environments, genes involved in functional categories including 'energy production and conversion', 'transcription', 'replication, recombination and repair', 'cell motility', 'inorganic ion transport and metabolism', 'secondary metabolism', 'signal transduction mechanism', and 'defense mechanism' are more enriched in deep-sea microbial communities (Konstantinidis et al. 2009), while genes involved in functional categories including 'cell division, chromosome partitioning', 'amino acid transport and metabolism', 'nucleotide transport and metabolism', 'carbohydrate transport and metabolism', 'coenzyme transport and metabolism', 'lipid transport and metabolism', 'translation and biogenesis', 'cell wall/membrane/envelope', 'protein turnover, chaperones', and 'intracellular trafficking and secretion' are less enriched in deep-sea microbial communities than in surface-ocean communities. One recent study in this thesis, using integrated metagenomic and metatranscriptomic analysis of deep-sea microbial communities in the North Pacific Ocean, led to similar conclusions except that it showed genes related to 'cell wall/membrane/envelope' are more abundant in the deep-sea than in surface water (Wu et al. 2013). The results were consistent with previous metagenomic studies that genes from abundant planktonic microbial populations are distinct between upper and lower levels of sea water (Frias-Lopez et al. 2008; Konstantinidis & DeLong 2008; Konstantinidis et al. 2009; Shi et al. 2011).

Conclusions

Nowadays, more taxonomic and functional studies of deep-sea microbial communities are conducted to unveil deep-sea microbial ecology. However, the deep-sea environments, as a huge microbial reservoir, still remain mostly unknown. The development and broad utilization of next-generation sequencing based metagenomic and metatranscriptomic analysis, and also the increasingly sophisticated computer modeling technologies in this field make it possible to uncover the full taxonomical picture of microbial communities in the special dwelling environments. Recent studies found that prokaryotic microbes in deep-sea environments are well-adapted to the special dwelling environments after long evolution; carrying genetic features that enable them to live and reproduce in the extreme environmental conditions. Recent progress in sequencing technologies is fueling a rapid increase in the number and scope of deep-sea microbial community-targeted studies. While metagenomic analysis can provide information on the taxonomic composition and metabolic potential of microbial communities in deep sea, metatranscriptomics serves to unveil the actual metabolic activities of the communities at a specific time and location, and how those activities are changing in response to environmental and biotic parameters.

Ph.D. Research Contributions:

Phylogenetic analysis of deep-sea bacteria, archaea, and eukaryota communities using 16s rDNA genes; phylogenetic and gene expression analysis of photosynthetic cyanobacteria and diatoms in deep-sea;

Performance evaluation of 4 different commercially available whole-transcriptomic amplification kits and optimization of whole-transcriptome amplification from low cell density deep-sea microbial samples for metatranscriptomic analysis;

Phylogenetic and functional analysis of deep-sea microbial community using integrated metagenomic and metatranscriptomic approach; strength of metabolic activity evaluation of microbial community in species level by introducing the ratio of RNA/DNA;

Phenotypic and potentially functional characterization of deep-sea *Pseudoalteromonas* strains isolated from deep-sea by annotating and comparing its transcriptome with the genomes of reference strains *Pseudoalteromonas* haloplanktis TAC125 and *Pseudoalteromonas* sp. SM9913; linked in-depth observation of transcriptomic and proteomic changes of WG07 under different growing temperatures.

CHAPTER 2

SEARCHING FOR CYANOBACTERIA AND DIATOMS IN THE TWILIGHT WATERS OF THE TEMPERATE NORTHEAST PACIFIC OCEAN

Introduction

The ocean is a major regulator of climate and has been the only true net sink for anthropogenic CO_2 over the past 200 years (Sabine et al. 2004). Marine photosynthetic microbes (photosynthetic prokaryotes and unicellular eukaryotes) in the open ocean have been well known for decades, and their activities are responsible for a big portion of this drawdown of carbon by means of the biological carbon pump: the photosynthetic fixation of CO_2 (primary production) and subsequent sequestration of mainly particulate organic carbon into deep water (Falkowski et al. 1998). It has been evident that sunlight-driven aerobic photosynthesis is widespread in tropical surface waters of the eastern Pacific Ocean and in temperate coastal waters of the northwestern Atlantic (Kolber et al. 2000).

An ocean can be divided into several zones based on the depth and how much light can penetrate: i) the epipelagic zone is from the surface down to around depth 200 m where there is enough light for photosynthesis, and thus plants and animals are largely concentrated in this zone. Most of the ocean photosynthetic activity occurs in this zone and it accounts for about half of the global primary production (Falkowski et al. 1998; Kolber et al. 2001); ii) the mesopelagic zone, also called the twilight zone, is from depth 200 m down to around 1000 m where some sunlight can penetrate, but in general, is sufficient for photosynthesis; iii) the bathypelagic zone from depth 1000 m down to around 4000 m; and iv) the abyssopelagic zone from depth 4000 m down to above the ocean floor where it is almost entirely dark and no sunlight can penetrate to this depth

(Ikeda et al. 2007). By a narrower or stricter definition, photosynthesis is a biological process of converting light-energy to chemical-bond energy, accompanied by assimilation of inorganic carbon in conjunction with photolysis of water, which produces oxygen as a by-product (Falkowski & Raven, 2007; Zubkov 2009) and photoautotrophic microbes have been considered responsible for this narrowly defined photosynthesis activity. Due to their roles in global carbon cycles and oxygen generation, significant research has been on photoautotrophic microbes in the ocean and the results suggested that they are abundant in the epipelagic zone. In a recent study, Vilibić & Šantić (2008) described the finding of photoautotroph cyanobacteria *Synechococcus* in the 800 m deep Adriatic waters during the Spring of 2006 using a flow cytometer with detection according to the red (chlorophyll) and orange (phycoerytrin) emissions simultaneously (Vilibić & Šantić 2008). However, there is no evidence showing those photoautotrophic microbes are actively functioning in the deep-sea beyond the epipelagic zone rather than just as dormant cells carried down by seasonal water movement.

In recent decades, molecular biology techniques, such as cultivation-independent phylogenetic surveys using ribosomal RNA (rRNA) sequencing, have been applied to decipher the taxonomic and metabolic diversity of microbes in the ocean (Stahl et al. 1984; Olsen et al. 1986; Giovannoni et al. 1990; Schmidt et al. 1991; Tringe & Hugenholtz 2008). In this section of the study, we performed a 16S-rRNA based community structure analysis for microbial communities present in the sea water collected from the seafloor of 765-790 meters in depth in the Pacific Ocean. Although the depth belongs to the mesopelagic zone where it is believed there is not enough sunlight

for autotrophic photosynthesis, oxygenic photoautotrophic organisms, such as diatoms *Thalassiosira* spp. and cyanobacteria *Synechococcus* sp. were identified.

Materials and Methods

***Sampling of the Deep-sea Water*.** Samples were obtained as part of the INSITE08 National Science Foundation (NSF) Ocean Observatories Initiative (OOI) mapping cruise on the research vessel (R/V) Thompson TN 221 with Chief Scientists Professor John Delaney and Professor Deborah Kelley in the NE Pacific Ocean (http://ooi.ocean.washington.edu/cruise/). The Towed Digital Camera System and Multi-Rock Coring System ("TowCam") of the Woods Hole Oceanographic Institution was on-board the R/V Thompson and was used to collect the deep-sea water samples. TowCam is a specially designed digital camera system that photographs the seafloor as it is towed above the ocean bottom behind an oceanographic research vessel. It is also equipped with six Niskin water bottles; each can collect 5 liters of sea water from the seafloor. The water was taken for each of the bottles at sites about 102 miles from Newport, Oregon (Bottle 1 & Bottle 2: 768 m in depth, N44 34.1931, W125 09.8414; Bottle 3: 763 m in depth, N44 34.0170, W125 09.0652; Bottle 4: 789 m in depth, N44 33.9816, W125 09.0652; Bottle 5: 789 m in depth, N44 34.9448, W125 09. 0642; and Bottle 6: 785 m in depth, N44 33.8831, W125 09.0622) during the research vessel (R/V) Thompson TN 221 Research Cruise in the Pacific Ocean at 7:12 to 7: 31 GMT on July 27, 2008. Since the camera lighting was located in the bottom whereas all six Niskin bottles were located on the top of the TowCam, exposure to light during water sampling was avoided. Niskin water bottle triggering was done using analog ports on the SeaBird SBE25 CTD, which

fires a solenoid, thereby mechanically releasing the monofilament cord holding open the Niskin bottle. After surfacing, a total of 30 liters of sea water (0~4ºC) was combined and filtered immediately through vacuum-driven 0.22 μm Stericup filteration system (1,000 mL; Millipore, Billerica, MA). The whole process was performed in the microbiology lab on the deck. To avoid possible light interruption, the microbiology lab was kept as dark as possible and the whole filtration process was done quickly. The filters (designed as RT6) with microorganisms were then immediately stored in -80°C and shipped back to the laboratory on dry ice for genomics analysis.

Cell Recovery, Extraction of Chromosomal DNA and Total RNA. The cells were washed from the membrane of the RT6 sample using 3 x 5.0 mL of RNAlater (Ambion, Austin, TX). After centrifugation, 2.0 mL concentrated cells were obtained and they were then divided into two microcentrifuge tubes for storage at -20°C. Following a modified DNeasy method (QIAGEN, Valencia, CA), total DNA was extracted from 1.0 mL recovery cells. Following a modified method combining Trizol (Invitrogen, Grand Island, NY) and RNeasy method (QIAGEN, Valencia, CA), total RNA was extracted from 1 mL recovery cells. Briefly, 1.0 mL of recovery cells was transferred to a pre-chilled microcentrifuge tube and centrifuged at 12,000 x g for 1 min at 4°C. After centrifugation, the supernatant was decanted and the cell pellet was frozen in a -80°C freezer. Then, the cell pellet was resuspended in preheated 200 μL Max Bacterial Enhancement Reagent and incubated at 95°C for 4 min. 1.0 mL TRIzol® Reagent was added to the lysate, mixed well and incubated at room temperature for 5 min. For phase separation, 0.2 mL cold chloroform was added to the mixture and mixed by shaking the tube vigorously by

hand for 15 sec. After being incubated at room temperature for 2 – 3 min, the sample was centrifuged at 12,000 x g for 15 min at 4°C and the mixture in the tube separated into a lower red, phenol-chloroform phase, an interphase, and a colorless aqueous phase containing RNA. This aqueous phase was transferred into a new Eppendorf tube and proceeded to RNeasy Kit. Equal volume (700 µL) of 70% ethanol was added to the supernatant and mixed well. Then the mixture along with the precipitate was applied to RNeasy mini spin column sitting in a 2 mL collection tube and centrifuged at 11,000 rpm for 15 sec. The flow through was discarded. 350 µL of Buffer RW1 was pipetted on to the RNeasy column and the column was centrifuged at 11,000 rpm for 15 sec. After discarding the flow through and changing the collection tube, DNase treatment on the column using QIAGEN RNase-free Dnase Set was performed. Then another 350 µL of Buffer RW1 was added onto the spin column, and centrifuged at 11,000 for 15 sec followed by the flow through discarding and the collection tube change. 500 µL of Buffer RPE was applied twice onto the RNeasy column and centrifuged at 11,000 rpm for 15 sec and 2 min respectively. Up to this step, RNA was extracted and filtered on the silica-gel membrane of the column. At last, 30 µL of RNase-free water was directly applied onto the membrane and centrifuged at 11,000 rpm for 1 min. The RNA concentration is 6.3 ng/µL, and 260/280 ratio is more than 2.0, indicating a good RNA quality.

***PCR Amplification of 16S rRNA Gene of Both Archaea and Bacteria, and 23S rRNA Gene of Both Eukaryotic Algae and* Cyanobacteria**. Using the DNA isolated from one of the ocean samples, we successfully constructed the three following cloning libraries: 16S rRNA Bacteria (forward primer FD1: 5’ AGAGTTTGATCCTGGCTCAG

3’, reverse primer 1540R: 5’AAGGAGGTGATCCAGCC 3’), 16S rRNA Archaea (forward primer Ar20F: 5’ TTCCGGTTGATCCYGCCRG 3’, reverse Arch958R: 5’TCCGGCGTTGAMTCCAATT 3’), and 23S rRNA Cyanobacteria and Euryotic algal chloroplasts (forward primer p23SrV-f1: 5’GGACAGAAAGACCCTATGAA3’, and reverse primer p23SrV-r1: 5’TCAGCCTGTTATCCCTAGAG3’). The optimal conditions for PCR amplification were established as: i) bacterial library: 94°C for 2 min, 30 cycles of 94°C for 30 seconds, 58°C for 1 min and 72°C for 1 min, final cycle at 72°C for 7 min; ii) archaeal library: 94°C for 2 min, 30 cycles of 95°C for 1.5 min, 55°C for 1.5 min and 72°C for 1.5 min, final cycle at 72°C for 7 min; and iii) 94°C for 2 min, 30 cycles of 95°C for 1.5 min, 55°C for 1.5 min and 72°C for 1.5 min, final cycle at 72°C for 7 min . The expected PCR products were recovered by gel electrophoresis and extracted using Qiaquick Gel DNA Extraction Kit (QIAGEN, Valencia, CA).

Cloning Library Construction and Sequencing. The concentration of the PCR products was measured using the NanoDrop before it was cloned into a pGEM-T easy vector (Promega, Madison, WI). In general, a 2:1 ratio of insert vs. vector was used to achieve the optimal ligation. The total ligation reaction volume was 20 µL which contained 10 µL 2×Rapid Ligation Buffer, 1 µL pGEM-T easy vector which has a concentration of 50 ng/µL, and 2 µL T4 DNA ligase. After incubation at 4°C overnight, 2 µL of the ligation mixture was used for electroporation. The conditions of electroporation were 2 mm and 2.5 KV, and ~70 µL competent cells were used. After electroporation, cells were cultivated in Super Optimal broth with Catabolite repression (SOC) media at 37°C for 1 h, and then plated on a Lysogeny broth (LB) plate which contains 100 µg/mL

of ampicillin. Before spreading the sample, 100 µL of 100 mM IPTG and 20 µL 50 mg/mL X-Gal were spread over the surface of an LB-ampicillin plate. After spreading the sample onto the plates, the plates were cultured at 37°C overnight. RandomLy picked clones were checked by isolating plasmid and enzymatic digestion for the right size insert, and were cultivated in 96-well plates and used for plasmid isolation and sequencing by an ABI 373 Sequencer using PCR primers (Table 1). The resultant DNA sequences were subjected to manually editing using Sequence Scanner Software 2.0 (Applied Biosystems, Grand Island, NY).

Table 1. Primers used for PCR to detect archaea, bacteria, cyanobacteria and eukaryotic diatoms in the deep-sea microbial community.

Name	Primer type	Sequene (5'-3')	Specificity	Reference
FD1	16S rRNA, forward	AGAGTTTGATCCTGGCTCAG	Bacteria	(Weisburg et al. 1991)
1540R	16S rRNA, reverse	AAGGAGGTGATCCAGCC	Bacteria	(Weisburg et al. 1991)
Ar20F	16S rRNA, forward	TTCCGGTTGATCCYGCCRG	Archaea	(DeLong et al. 2006)
Arch958R	16S rRNA, reverse	TCCGGCGTTGAMTCCAATT	Archaea	(DeLong et al. 2006)
p23SrV-f1	23S rRNA, forward	GGACAGAAAGACCCTATGAA	Plastid of eukaryotic algae and *Cyanobacteria*	(Sherwood & Presting 2007)
p23SrV-r1	23S rRNA, reverse	TCAGCCTGTTATCCCTAGAG	Plastid of eukaryotic algae and *Cyanobacteria*	(Sherwood & Presting 2007)
S-psbA-F	*psbA*, forward	CAATGCACGGTTCTCTCGTA	*Synechococcus* sp. WH8102	This study
S-psbA-R	*psbA*, reverse	ATACCGACTACAGGCCATGC	*Synechococcus* sp. WH8102	This study
T-psaC-F	*psaC*, forward	TACGTGTATCGGTTGTACTCAATG	*Thalassiosira pseudonana*	This study
T-psaC-R	*psaC*, reverse	GGACATGCTGTTTCACAACG	*Thalassiosira pseudonana*	This study
T-psbA-F	*psbA*, forward	GCTGGTGTATTCGGTGGTTC	*Thalassiosira pseudonana*	This study
T-psbA-R	*psbA*, reverse	GTAACCGTGTGCAGCTACGA	*Thalassiosira pseudonana*	This study

Sequence Alignment and Phylogenetic Analysis. Sequences were compared against those in GenBank through the NCBI internet using BLAST 2.2.10 (Altschul et al. 1997). For bacteria clones, sequences were also compared against the 16S rDNA database from the Ribosomal Database Project (Cole et al. 2005). Alignment of sequences was done with Clustal W (Thompson et al. 1994). Construction of Neighbor-Joining phylogenic trees and bootstrapping were performed using MultiPhyl Online service (Keane et al. 2007). Operational taxonomic units (OTU) were determined based on a 3% divergence cutoff for individual "species" OTU. To estimate the extent of coverage of diversity in our clone libraries, rarefaction analysis was applied in this purpose using a RarefactWin shareware program. All DNA sequences were submitted to the NCBI GenBank database. Accession numbers GU317449–GU317689 were assigned to Archaea-related sequences, and numbers GU317690–GU317940 were assigned to bacteria-related sequences, and GU317367–GU317448 were assigned to photosynthetic microbe-related sequences.

Gene Expression Analysis. To determine the gene expression involved in protein synthesis in photoautotrophic microbes, a set of 23S based primers specific to eukaryotic algae and cyanobacteria was adopted from the literature (Sherwood & Presting 2007), or designed using Primer3 v. 0.4.0 (DNA Software, Totowa, NJ), and then synthesized by Invitrogen Corporation (Carlsbad, CA). The same primers were also used in the 23S rRNA gene cloning library construction as described above. Using genomic DNA *Synechocystis* sp PCC6803 as a PCR template, we amplified a 500-bp fragment of 23S rRNA gene for the purpose of generating a standard curve for quantitative RT-PCR. SYBR Green One–Step Quantitative RT-PCR with BioRad reagents was performed on

Rotor-Gene 6000 (Corbett Life Science). The PCR reaction in a total 20 µL volume set up was as follows: 10 µL 2 x SYBR Green Rxn mix (BioRad), 1.5 µL forward primer (4 µM), 1.5 µL reverse primer (4 µM), 4.5 µL nuclease free-water. 0.5 µL iScript RT enzyme for One-Step (BioRad), and 2 µL RNA/DNA template. The cDNA synthesis was performed at 50°C for 15 min followed by 95°C for 5 min for inactivation of reverse-transcriptase. The PCR cycling program was: 40 cycles of 95°C for 10 sec, 56°C for 56 sec, 72°C for 30 sec. Data analysis was carried out using the software provided by (QIAGEN, Valencia, CA).

One gene directly involved in photosynthesis in the *Synechococcus* sp., psbA gene in photosynthesis system II, was analyzed. Primers were designed based on the complete genome sequence of *Synechococcus* sp. WH8102 (Palenik et al. 2003) using Primer3 v. 0.4.0 (DNA Software, Totowa, NJ) and synthesized by Invitrogen Corporation (Carlsbad, CA) (Table 1). Similarly, two genes directly involved in photosynthesis in *T. pseudonana* were analyzed: the psaC gene involved in photosynthesis system I iron-sulfur center and the psbA gene involved in the photosynthesis system II. Primers were designed based on the complete genome sequence of *T. pseudonana* (Armbrust et al. 2004), using Primer3 v. 0.4.0 (DNA Software, Totowa, NJ) and synthesized by Invitrogen Corporation (Carlsbad, CA) (Table 1). RNA isolated from the deep-sea water samples was pretreated by heating under 65ºC for 5 min and chilling on the ice immediately. SuperScript VILO cDNA synthesis kit (Invitrogen) was used to reverse transcript RNA into cDNA according to the instructions from the manufacturer. To enhance the sample following PCR, the corresponding reverse primer was added into the reverse transcription mixture. Conventional PCR was carried out to characterize the

expression of the *Synechococcus* psbA gene and the *T. pseudonana* psaC gene in an Eppendorf Mastercycler® (Eppendorf, Hauppauge, NY) using PCR Master Mix (2X) and Taq DNA Polymerase (Fermentas Life Science, Pittsburgh, PA) with a reaction volume of 20 µL. The thermal profile for the *Synechococcus* psbA gene was 1 cycle at 94°C for 5 min; 35 cycles at 92°C for 1 min, 57°C for 1 min, and 68°C for 1 min; final extension at 72°C for 10 min and soak at 4°C; and for the *T. pseudonana* psaC gene was 1 cycle at 95°C for 10 min; 30 cycles at 94°C for 15 sec, 58°C for 20 sec, and 72°C for 30 sec; final extension at 72°C for 7 min and soak at 4°C. While we failed to amplify the T. pseudonana psbJ gene using conventional PCR, we applied touchdown PCR to detect the gene with the annealing temperature range from 50°C to 60°C. After PCR, 2 µl of the reaction mixture was analyzed by electrophoresis in a 2% agarose gel containing CYBR Green prior to being visualized under UV light with the Molecular Imager® Gel Doc™ XR System (Bio-Rad, Hercules, CA). A PCR band of the correct size was cut out, purified, and sequenced for verification. A similar real-time PCR was run as a negative control to prove the amplification products were from RNA and not DNA.

Results

DNA and RNA Isolation from the Deep-sea Water. Direct counting of the cell number was performed under light microcopy. The results showed that a total of 3-5 x 10^7 cells were recovered. Based on the number, it was estimated that the microbial cell density was around $1.0–1.7\times10^3$ cells per milliliter of the sea water, consistent with the cell number suggested for the similar depth in the tropical/subtropical Atlantic Ocean (Varela et al. 2008).

Following a modified DNeasy method (QIAGEN, Valencia, CA), we extracted total DNA from 1.0 mL (Cell No = ~10^7) of recovery cells. DNA was eluted three times, each time into 50 µL ddH_2O. The DNA concentration was about 24 ng/µL, 17 ng/µL and 11 ng/µL for the first, second and third elution, respectively. When combined, a total of 2600 ng of chromosomal DNA was obtained. Although the yield is relatively low, the DNA quality is good with 260/280 ratio or more than 1.8.

Using a modified method combining Trizol (Invitrogen, Carlsbad, CA) and RNeasy methods (QIAGEN, Valencia, CA), we also extracted total RNA from 1.0 mL (Cell No = 10^7) of recovery cells. The RNA was eluted into 30 µL ddH_2O. A total of 189 ng of RNA was obtained. Although recovery was low, probably due to the low metabolic activity of microbes at this depth, the quality of RNA was good, as suggested by a DNA quality of 260/280 ratio greater than 2.0.

Community Structure Analysis. DNA isolated from the deep-sea water was used to construct three clone libraries for microbial community structure analyses: i) 16S rRNA bacterial, ii) 16S rRNA archaeal and iii) algae and *Cyanobacteria* specific 23S rRNA libraries. To improve the PCR performance, we used PCRboost (Biomatrica, San Diego, CA) to replace water in the PCR mixture.

Archaea: Under the optimized condition, PCR amplification of 16S rRNA genes using archaea-specific primers generated a single band with the expected size of about 700-bp on the gel. Sequence analysis of 156 random clones revealed 21 different phylotypes based upon 97% sequence similarity, and they corresponded closely to the sequences of many groups of unknown microorganisms found in marine environments, such as EU199549 (Kalanetra et al. 2009) and FJ002876 (Agogué et al. 2008). The

majority of clones (150 in total) are phylogenetically similar to uncultured marine group I *Crenarchaote*, with 20 OTUs having been recognized using a 97% sequence identity as a cut-off value (Figure1). In addition, a small number of clones (6 in total) are phylogenetically affiliated to uncultured marine group II *Euryarchaeote*, and they represented two previously identified OTUs, sequenced clone UEU78206 from the Santa Barbara Channel (Massana et al. 1997) and DQ300553 from the North Pacific Subtropical Gyre (DeLong et al. 2006), respectively.

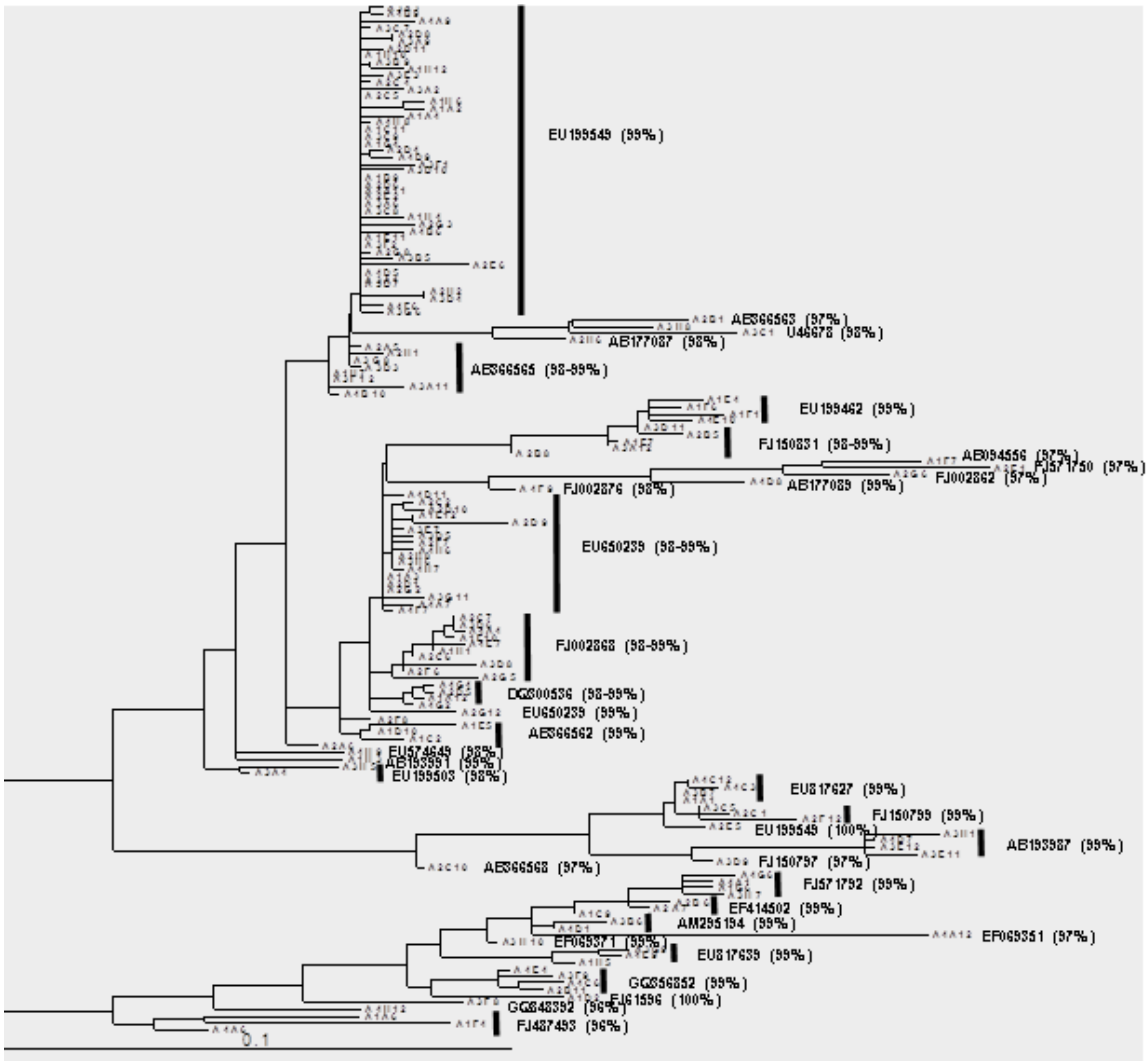

Figure 1. Reasonable Archaeal groups found in phylogenetic analysis of the microbial community.

Bacteria: Under the optimized condition, PCR amplification of bacterial 16S rRNA genes using bacteria-specific primers generated a single band with the expected size of about 1300-bp on the gel. Sequencing analysis of 251 random clones revealed significant bacterial diversity (Figure 2) present in this site after comparing to the 16S rDNA database from the Ribosomal Database Project (Cole et al. 2005). A majority of the clones (168 in total) from the clone library are phylogenetically associated with the phylum Proteobacteria, and they can be further grouped into classes of Alpha- (46 clones),

Beta- (4 clones), Gamma- (94 clones), Delta- (23 clones) and Epsilon-*proteobacteria* (1 clone). Among Alpha-*proteobacteria*, 15 total OTUs have been recognized based on a 97% sequence identity as a cut-off value. Based on sequence BLAST searches, 28 clones from 8 OTUs in Alpha-*proteobacteria* class are highly similar to marine SAR 11 clones with sequence identity more than 95%, such as EU361476 (Pham et al., 2008). Seven other OTUs belong to unknown species. The Beta- group contains 2 OTUs, one has a 98% sequence identity to AB239007, an uncultured *Nitrosospira* sp. (Arakawa et al., 2006), and another belongs to an unknown species. The Gamma-group contains 23 OTUs with most OTUs belonging to unknown species, and only five OTUs can be classified into uncultured members belonging to a genus including Colwellia (5 clones, 1 OTU), *Halomonas* (1 clone, 1 OTU), *Moritella* (5 clones, 1 OTU), *Psychromonas* (1 clone, 1 OUT), as well as *Shewanella* (1 clone, 1 OTU). Thirteen OTUs from the Delta-group are identified and most of them belong to unknown species. However, based on BLAST searches, they seem associated with sulfate-reducing bacteria, such as *Desulfobacterales* (8 clones, 6 OTUs), *Desulfovibrionales* (1 clones, 1 OTUs), *Desulfuromonales* (2 clones, 2 OTUs) and *Syntrophobacterales* (6 clones, 3 OTUs). In addition, 6 clones representing 2 OTUs seem to be *Moxococcals*-related. Finally, one clone (B3E5) sharing 99% sequence identity to an uncultured Epsilon-*proteobacterial* isolate AJ535216 has been recognized (Knittel et al. 2003).

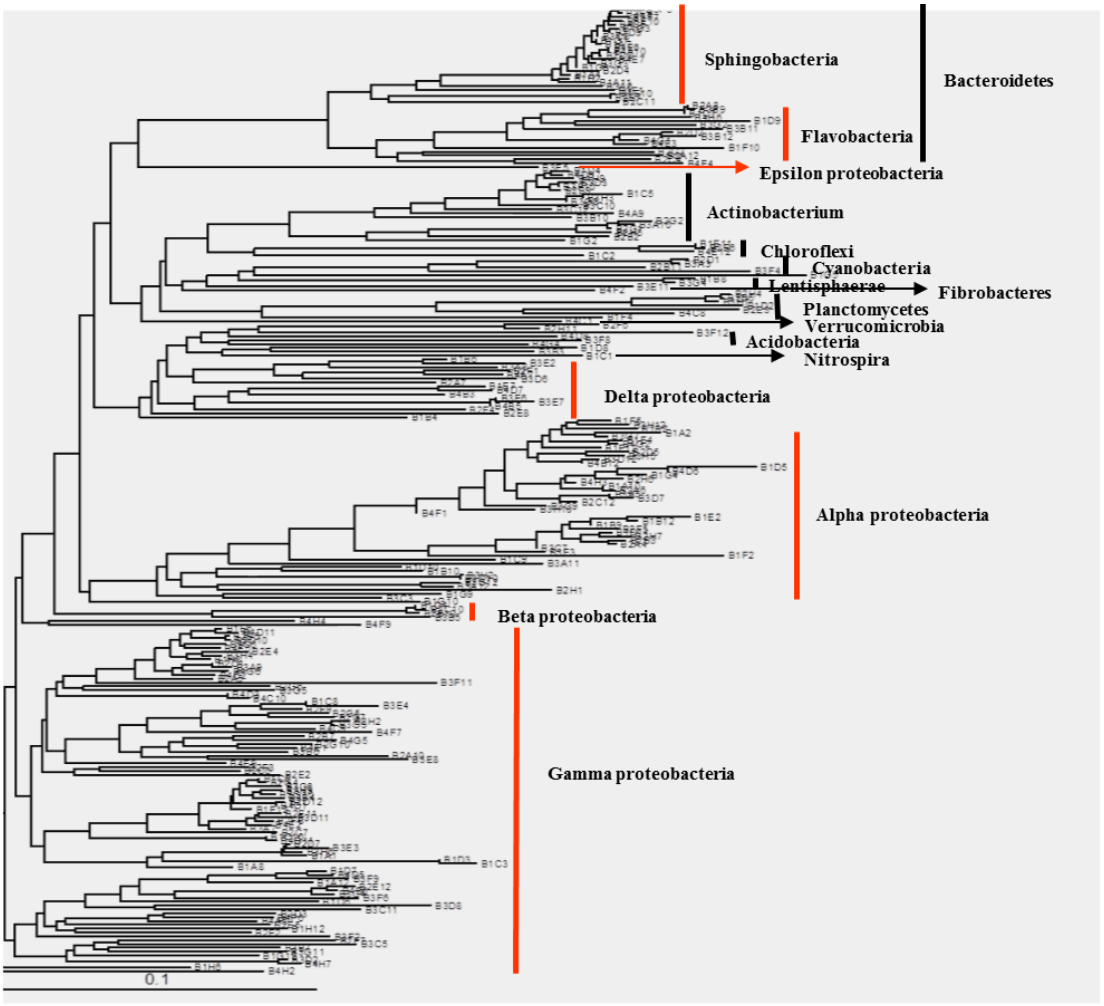

Figure 2. Reasonable Bacterial groups found in phylogenetic analysis of the microbial community.

Forty-two clones (13 OTUs) are phylogenetically affiliated with the phylum *Bacteroidetes*, and a majority (41 clones) of them can be further classified into orders *Flavobacteriales* (16 clones, 8 OTUs) and *Sphingobacteriales* (25 clones, 3 OTUs). *Flavobacterials* clones are divided into families including *Flavobacteriaceae* (13 clones, 8 OTUs) and *Crymorphaceaea* (3 clones, 2 OTUs). All *Sphingobacteriales* clones belong to the family *Flexibacteraceae* and they share sequence identify ranging from 88% to 98% with a previously isolated uncultured clone AB015263 (Li et al. 1999). Finally, B2A12 cannot be attributed to either one of two classes mentioned above. This clone shares 100%

sequence identity to GQ348796, an uncultured *Bacteroidetes* bacterium clone from an oxygen minimum zone (Walsh et al. 2009). Nineteen clones (4 OTUs) belong to *Actinobacteria*, among them, one OTU consisting of 10 clones is closely related to the previously isolated clone GQ349866 with a sequence similarity of over 98%. One OTU consisted of 3 clones (B1E11, B2E6 and B4E12) that are closely related to uncultured green non-sulfur bacteria *Chloroflexi* sp. In addition, clones phylogenetically affiliated to other phylum including *Planctomycete* (7 clones, 3 OTUs), *Cynaobacteria* (4 clones, 1 OTU), *Acidobacteria* (3 clones, 1 OTU), *Lentisphaerae* (2 clones, 1 OTU), *Nitrospira* (1 clone, 1 OTU), *Fibrobacteres* (1 clone, 1 OUT) and *Verrucomicrobia* (1 clone, 1 OTU) have been identified.

In regards to eukaryotic algae and cyanobacteria, it was intriguing that some clones identified in the bacterial 16S rRNA clone library of the deep-sea water belong to phototrophic bacteria, such as photoautotrophic *Cyanobacteria* and photoheterotrophic *Chloroflexi*. To further confirm the presence of the photoautotrophic microorganisms at this ocean depth with very little sunlight, we applied one pair of universal PCR primers aimed at the 23S rRNA plastid gene for the construction of a 23S rRNA clone library (Sherwood & Presting 2007). PCR amplification of 23S rRNA genes using eukaryotic algae and cyanobacteria-specific primers generated a single band around 500-bp in size. Sequencing of 94 random clones revealed that a majority of them are divided between diatoms (61 clones) and *Cyanobacteria* solely belonging to *Synechococcus* sp. (16 clones) (Figure 3). Among the diatom group, 45 clones (7 OTUs) are similar to *Thalassiosira* spp. Other closely related diatoms include: *Nitzsch* spp. (11 clones, 7 OTUs), *Odontella* spp. (2 clones, 2 OTUs), *Phaeodactylum* spp. (3 clones, 2 OTU). Also, a few clones

phylogenetically affiliated to other algae commonly found in the oceans such as Chlorella sp. and *Emiliania* sp. (Cattolico et al., 2008) were also identified based on a sequence similarity search. In terms of Syne-affiliated clones, 14 clones are mostly close to *Synechococcus* sp. CC9311 with 97-99% sequence identity (Palenik et al., 2006) and two to *Synechococcus* sp. CC9902 with 99% identity, respectively (Figure 3).

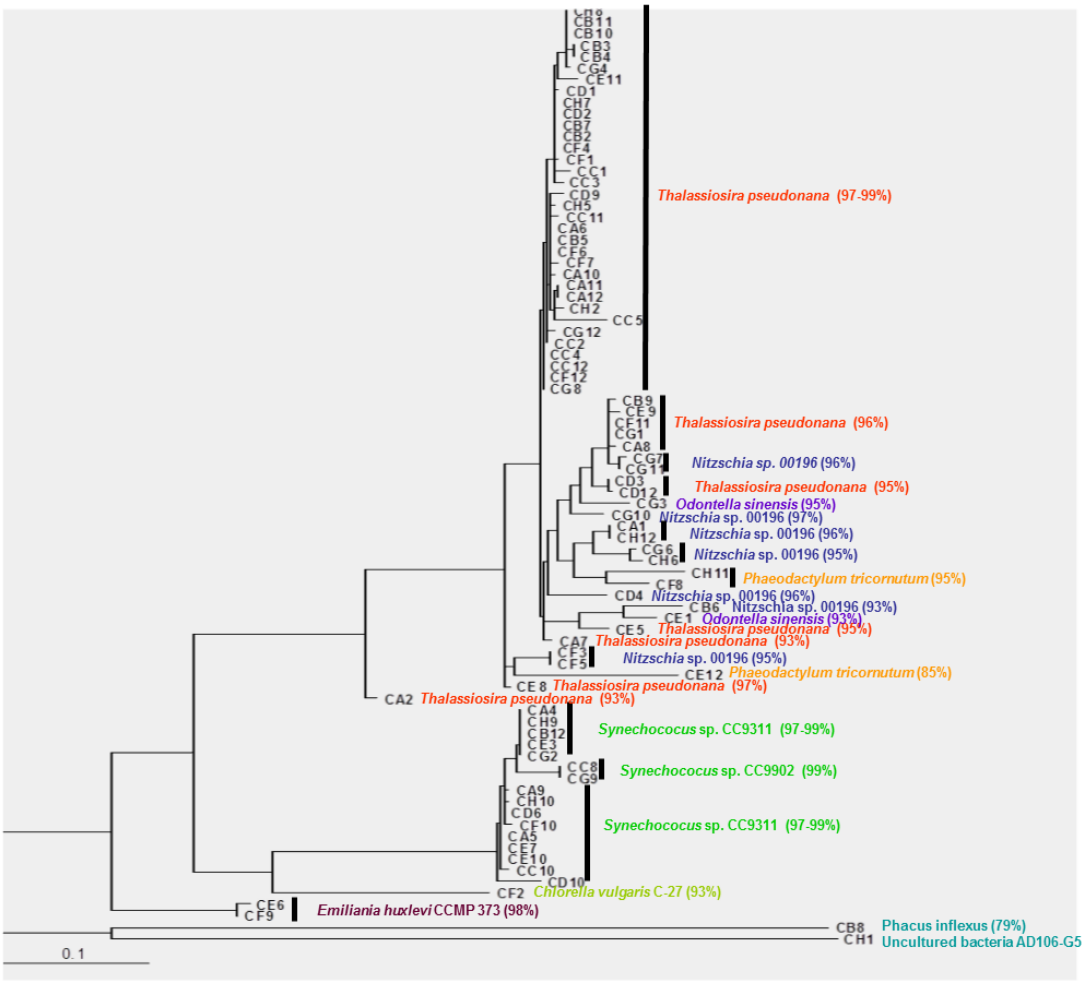

Figure 3. Phylogenetic analysis of the deep-sea photosynthetic eukaryotic algae and cyanobacteria.

Gene Expression Analysis. While the active presence of photoheterotrophic microbes in the deep sea has been demonstrated in some early studies (Karl et al. 1984; Zubkov 2009), the finding of photoautotrophic organisms similar to *Synechococcus* sp.

and *Thalassiosira* sp. in the deep-sea microbial samples raise an immediate question about whether these photoautotrophic microbes are in the status of actively metabolizing or are simply dormant cells. To seek an answer to this question, we performed further studies to determine the gene expression directly involved in protein synthesis and photosynthesis, using total RNA isolated from the deep-sea samples as a template. We first evaluated the expression of highly expressed rRNA genes involved in protein synthesis using a quantitative PCR method. Primers were designed based on the two complete genome sequences of *Synechococcus* sp. WH8102 (Palenik et al. 2003) and *Thalassiosira pseudonana* (Armbrust et al. 2004). Figure 4 shows the real-time PCR analysis of the copy number of 23S rRNA gene. The analysis was done separately using DNA or cDNA as a template. The calculated average copy number of the algae and *cyanobacteria* specific 23S rRNA gene is 5.63×10^4 and 2.04×10^5 per reaction for DNA and cDNA template, respectively. Considering the dilution factors of each template, the normalized copy number for 23S rRNA from cDNA (or RNA) and DNA is 2.44×10^7 and 4.25×10^6, respectively. Thus, the copy number of 23S rRNA is about 6 times higher than that of 23S rDNA. The results suggested that the photosynthetic algae and *cyanobacteria* are metabolically active even in the ~800 m-deep mesopelagic zone in the Pacific Ocean.

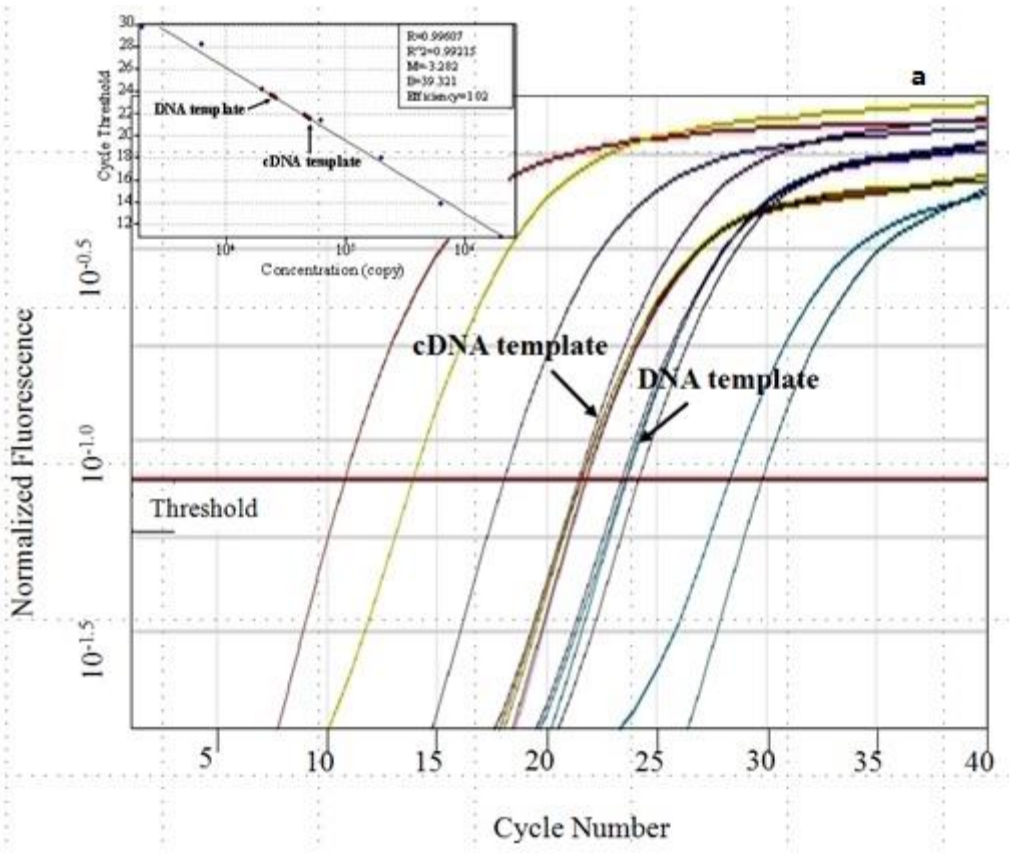

Figure 4. A quantitative RT-PCR analysis of algae and cyanobacteria specific 23S rRNA gene.

We then evaluated the expression of genes involved in photosynthesis in *Synechoccus* sp. and *Thalassiosira* spp. There are two types of photosynthetic systems existing in both *Synechoccus* sp. and *Thalassiosira* spp.: type I (PSI) and type II (PSII) (Palenik et al., 2003; Armbrust et al., 2004). Several genes from both photosynthesis systems were selected as targets for RT-PCR analysis: psbA gene (PSII) in *Synechococcus* sp., and psaC (PSI) and psbA (PSII) genes in *T. pseudonana*. Although initial attempts using a quantitative RT-PCR approach failed to detect the expression of those genes, probably due to the relative low copy number of their mRNA (cDNA) compared with 23S rRNA which we can measure successfully, we were able to amplify PCR bands of correct sizes of all three genes from the cDNA templates using optimized

conventional and touchdown PCR approaches. The PCR products were purified from the gel (Figure 5) and validated as the correct target genes by DNA sequencing. Using the default parameters, blast search of the PCR primer band based on the *Synechococcus* sp. psbA gene against NCBI Genbank gave two hits, 88% identity (E value 3e-34) to *Synechococcus* sp. CC9311 and 87% to *Synechococcus* sp. WH 8102 (E value 6e -31). Blast search of the PCR primer band based on the *T. pseudonana* psaC gene gave only one hit, 93% identity (E value 2e -26) to *T. pseudonana*. Blast search of the PCR primer band based on the *T. pseudonana* psbA gene showed 100% identity (E value 6e -32) to *T. pseudonana.*

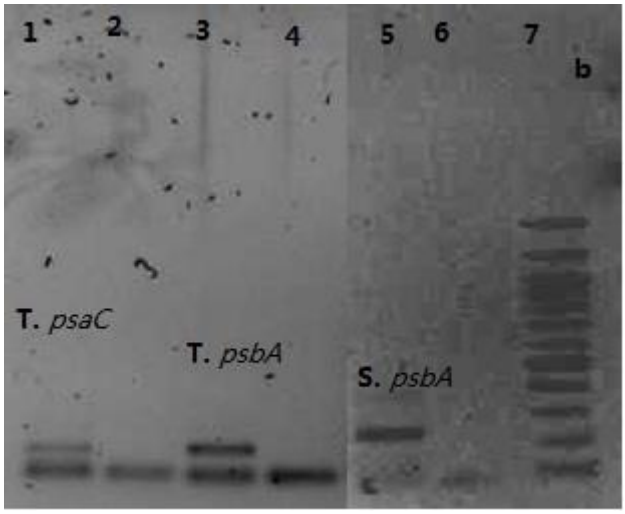

Figure 5. Electrophorisis gel image of three photosynthesis related genes identified in deep-sea microbial community RNA

To exclude the possibility that cDNA samples were contaminated with DNA during the RNA isolation step and to demonstrate that the expression was from DNA based templates, we set up negative controls for all genes using original RNA solution as template (without the cDNA synthesis step) for direct PCR amplification under the identical PCR conditions, and no amplification was observed. The results demonstrated

that the psbA gene of *Synechococcus* sp., psaC and psaA genes of *T. pseudonana* were actively expressed in the sea water samples.

Discussion

Photoautotrophic microbes such as the well-known genera *Prochlorococcus* and *Synechococcus* are abundant members of the microbial community in the euphotic zone and are responsible for much of the primary production in the ocean (Worden et al. 2000; DuRand et al. 2001; Herndl et al. 2008). Meanwhile, early studies have suggested that viable photoautotrophic microbes, such as *Synechococcus* sp. can be found from deep-sea water or sediments where there is supposed to little or no sunlight (Thiel et al. 1990; Reid et al. 1991; Patterson et al. 1993; Martín-Cuadrado et al. 2007; Vilibić & Šantić 2008; Zubkov 2009). In a recent study, Vilibić & Šantić (2008) described the discovery of the photoautotroph cyanobacteria *Synechococcus* in the 800 m deep Adriatic waters during the Spring of 2006. However, so far there is no evidence showing those photoautotrophic microbes are metabolically actively functioning rather than being dormant cells at this depth. To seek more insights on the functional boundary of photoautotrophic microbes in the ocean, in this study, we performed community and gene expression analyses for microbial communities in the sea water collected from the seafloor of 765-790 m in depth in the Pacific Ocean. This depth belongs to the deep mesopelagic zone where some light is normally insufficient for photosynthesis. In addition to clone library analysis and direct auto-fluorescence imaging, we improved the RNA isolation protocols so that total RNA of good quality was successfully obtained from a relatively small amount of biomass, and then was used to analyze genes directly related to photosynthesis. The presence of photoautotrophic microbes in the zone was demonstrated by a bacterial 16S rRNA based

clone library analysis, and the metabolic activity of the oxygenic photoautotrophic organisms *Thalassiosira* spp. and cyanobacteria *Synechococcus* sp. was demonstrated by expression of the 23S rRNA gene involved in protein synthesis in *Cyanobacteria*, and expression of photosynthesis-involved *psaC* and *psbA* genes in *Thalassiosira* spp., and expression of photosynthesis-involved *psbA* gene in *Synechococcus* sp., respectively. However, it still remains unclear how those active photoautotrophic microbes reach the deep mesopelagic zone. One less plausible explanation is that they sink to this depth in the water, although it could be argued that it is less possible for bacteria compared to larger phytoplankton due to their minute size (1-2 μm in diameter). For instance, it has been suggested that a cell 1.0 μm in diameter is predicted to sink at a rate of only 2.6 mm a day (Raven 1986). Another possibility is that they are going down as microbial assemblages associated with large sinking particles, as *cyanobacteria* was found associated with sinking particles sampled in sediment traps deployed at depths beyond 1000 m the Porcupine Abyssal Plain (PAP, NE Atlantic) (Vanucci et al. 2001). The third possibility is through mechanisms similar to what has been proposed for *cyanobacteria* found in deep Adriatic waters (Vilibić & Šantić 2008). Two processes involved were: 1) deep convection in the South Adriatic Pit (SAP); and 2) density current going downslope. The first process was responsible for bringing the cyanobacteria down to 600-m depth in the area of convection, and the second one triggered the downslope transport of the *Cyanobacteria* to the very bottom of the SAP. The depletion rate of *Synechoccocus*, *Cyanobacteria* in an extremely hostile environment has been computed to equal about 1 month (Vilibić & Šantić 2008). However, none of those possibilities make enough sense here when taking into consideration that RNA molecules (especially mRNA) are

generally only stable less than several hours after the transcription is stopped. For example, an early study in *Synechocystis* PCC 6803 by Gill et al. (2002) showed that genes involved in photosystem II such as structural subunit genes psbA2 and psbD2 can only maintain high transcript levels for 60 min after the lights were turned off (Gill et al. 2002). It is thus well expected that the RNA involved in photosynthesis will be fully degraded when they reach the deep mesopelagic zone from the surface ocean through any of three possible mechanisms. The forth possibility is that those photoautotrophic microbes are a natural part of activity at this depth. As one support to this hypothesis, it has been suggested that diatoms *Thalassiosira* spp. and cyanobacteria *Synechococcus* sp. are microbes with high adaptation ability and can function well in less favorable environments. In early studies, bacterial chemolithotrophy has been suggested as a key player for organic carbon generation at 700−900 m sea depth, and the energy for this process may be in part provided by detrital NH_4^+ derived from the downward flux of large particles (Karl et al. 1984). This possibility, if true, will imply that the functional boundary of photoautotrophic activity may be beyond the epipelagic zone, deeper than previously described, and their contribution to the oceanic carbon cycles may need further investigation. Finally, the facts that most bacteria and archaea identified from our clone libraries are similar to those previously found in the deep-sea environments, and deep-sea photoautotrophic community identified from our clone library is limited to several species (dominated by only *Synechococcus* sp. and *Thalassiosira* spp.) in comparison to that in the surface water, measured against the possibility of any surface water contamination in our sample, since no abundant surface-water species was observed in three of the clone libraries.

CHAPTER 3
OPTIMIZATION OF WHOLE-TRANSCRIPTOME AMPLIFICATION FROM LOW CELL DENSITY DEEP-SEA MICROBIAL SAMPLES FOR METATRANSCRIPTOMIC ANALYSIS

Introduction

Next-generation sequencing based metatranscriptomics can analyze gene expression directly from natural microbial assemblages, and has allowed assessment of microbial functionality in the natural environments (Shi et al. 2009). The technology can overcome various constraints inherent in the use of traditional qPCR and microarrays to monitor gene expression in complex microbial communities, such as limited number of genes can be examined, a prior need to select gene targets, and requirements for probes or primers. In addition, sequencing based metatranscriptomics has no significant bias towards any known sequences within microbial communities, and thus provides a new and powerful approach for identifying previously unknown genes and analyzing community-specific variants of functional genes in nature (McGrath et al. 2008). In recent years, several metatranscriptomics analyses of marine and soil samples have been reported (Poretsky et al. 2009; Gilbert et al. 2008; Vila-Costa et al. 2010). Most of these studies started from selective mRNA enrichment by depleting ribosomal RNA or enriching for polyA-tailed mRNA, respectively. In a recent study, a total RNA pool of microbial community, including mRNA and rRNA, was analyzed to reveal both functionality and taxonomic diversity of the environments (Urich et al. 2008). This approach avoided extensive purification steps of mRNA, and enabled simple and quick preparation even from difficult samples. Together these studies have significantly

enhanced our understanding of the diversity and functionality of microbial communities in the nature.

Several commercial next-generation sequencing platforms are currently available for metatranscriptomic analysis, such as Roche's 454 FLX Genome Sequencer, Illumina's Genome Analyzer, and Applied Biosystems' SOLiD. Typically these sequencing platforms require at least 2 – 5 ug of input DNA/cDNA in order to achieve good sequencing data, which is equal to the total amount of DNA isolated from approximate 10^9 Escherichia coli cells. Unfortunately, in many natural environments (e.g., the deep sea), microbial cell density can be as low as 10^3 – 10^4 cells per mL of seawater (Takai et al. 2004), and total RNA amount in the cells is also low due to their relatively slow metabolism. Compared with biomass abundant samples, such as surface seawater, the intact RNA that can be isolated from these environments is very limited in quantity; usually low to several hundred nanograms level. Moreover, environmental samples, different from pure culture grown in the laboratory, often contain various minerals and humic acids, which may inhibit amplification reactions (Broemeling et al. 2008; Demeke and Jenkins 2010). Thus, in most of these cases, quality of the isolated environmental RNA is not comparable with that from laboratory grown pure culture either. In order to analyze these samples, it is inevitable to perform amplification to fulfill amount requirement by the current pyrosequencing technologies. However, to our knowledge, few protocols are currently available for whole-transcriptome amplification from small amounts of microbial samples for pyrosequencing.

Our center is working on in situ sensor development in order to provide real-time oceanographic observations (Qin et al. 2010; Zhou et al. 2012; Zhu et al. 2012). As part

of this effort, we are working on developing genomics-based biosensors to monitor microbial communities and their dynamic changes in the deep ocean. In this respect, obtaining baseline understanding of the diversity and functionality of microbial communities through metagenomic and metatranscriptomic analyses becomes essential. In this thesis, we report a new molecular protocol to isolate total RNA from a very small amount of deep-sea microbial samples, and then to amplify them to the amount required for pyrosequencing. The effectiveness of this protocol was proven through analyzing sequence data from both a random cloning library and pyrosequencing of amplified metatranscripts. The results demonstrated that the protocol was capable of reliably amplifying whole-transcriptome from small amount of environmental microbial samples for pyrosequencing.

Materials and Methods

Sample collection and cell recovery. Microbial samples used in this study were collected from seven different sites, at depths of 784 - 2,319 m in the Northeast Pacific Ocean. For each sample, a total of 30 liters of deep-sea water was taken at 2-3 m above the sea floor by six water bottles carried on the Towed Camera System ('TowCam') of the Woods Hole Oceanographic Institution. Deep sea water was then filtered immediately through 0.22 μm filters which were then stored in -80°C until used. Since our goal was to isolate RNA for community structure and gene expression analyses, we washed cells off of the filter membrane gently so that the RNA profiles could be persevered with 5.0 mL of the chilled RNAlater solution (Ambion, Austin, TX) for each wash. A total of five washes were performed for the filter, cells recovered from each wash were combined into a single pool, and stored at -80°C immediately for subsequent RNA isolation.

Isolation of total RNA. Using a modified protocol combining Trizol (Invitrogen, Carlsbad, CA) and RNeasy methods (QIAGEN, Valencia, CA), we extracted total RNA from the collected microbial cells. Briefly, 1.0 mL of collected cells (i.e. 3×10^8) was transferred to a pre-chilled microcentrifuge tube and centrifuged at 12,000 x g for 1 min at 4°C. After centrifugation, the supernatant was decanted and the cell pellet was re-suspended in preheated 200 µL Max Bacterial Enhancement Reagent provided by the Trizol kit and incubated at 95°C for 4 min. 1.0 mL TRIzol® Reagent was added to the lysate, mixed well and incubated at room temperature for 5 min. For phase separation, 0.2 mL cold chloroform was added to the mixture and mixed by shaking the tube vigorously by hand for 15 sec. After incubation at room temperature for 2 – 3 min, the sample was centrifuged at 12,000 x g for 15 min at 4°C and the mixture in the tube separated into a lower red, phenol-chloroform phase, an interphase, and a colorless aqueous phase containing RNA. This aqueous phase was transferred into a new Eppendorf tube and the RNeasy method applied. Equal volume (700 µL) of 70% ethanol was added to the supernatant and mixed well. Then the mixture along with the precipitate was applied to an RNeasy mini spin column sitting in a 2 mL collection tube and centrifuged at 11,000 rpm for 15 sec. The flow through was discarded. 350 µL of Buffer RW1 was pipetted on to the RNeasy column and the column was centrifuged at 11,000 rpm for 15 sec. After discarding the flow through and changing the collection tube, DNase treatment on the column using QIAGEN RNase-free DNase Set was performed for 15 min at room temperature. Then another 350 µL of Buffer RW1 was added into the spin column, and centrifuged at 11,000 for 15 sec followed by the flow through discarding and the collection tube change. 500 µL of Buffer PE was applied twice onto the RNeasy column

and centrifuged at 11,000 rpm for 15 sec and 2 min respectively. At last, 30 µL of RNase-free water was directly applied onto the membrane and centrifuged at 11,000 rpm for 1 min. After 1 min centrifugation, total RNA was eluted from the membrane and kept at - 80ºC unil used.

Whole transcriptome amplification. Total RNA isolated from the deep-sea microbial samples was directly applied to a whole transcriptome amplification process. Three different amplification methods were evaluated in this study: i) Illustra GenomiPhi V2 DNA Amplification Method (GE Healthcare, NJ): This method is designed for whole-genome DNA amplification. To modify it for whole-transcriptome amplification, SuperScript® VILO™ cDNA Synthesis Method (Invitrogen, Carlsbad, CA) was employed to synthesize cDNA. The cDNA synthesis condition was: 25°C for 10 min, 42°C for 60 min, and 85°C for 5 min. Amplification was then performed using purified cDNA product with the GE Healthcare method following the instructions from the manufacturer: denaturation at 95°C for 3 min, incubation at 30°C for 90 min, and inactivation at 65°C for 10 min. After amplification, the product was purified with QIAquick® PCR Purification method (QIAGEN, Valencia, CA) and the concentration was measured using NanoDrop; ii) QuantiTect® Whole Transcriptome method (QIAGEN, Valencia, CA): This method is designed for preparing cDNA from total RNA by whole transcriptome amplification. This method was different from the GE method since before final amplification, there was a ligation step in the QuantiTect method, which proved to be a key factor influencing amplification efficiency in our following experiment. We used 12 hrs instead of the maximum 8 hrs suggested in the manufacture's manual. Hence, the thermal cycling condition was: RNA reverse transcription at 37°C for 60 min and

inactivation at 95°C for 5 min; ligation at 22°C for 6.5/12 hrs; cDNA amplification at 30oC for 3.5 hrs and inactivation at 95°C for 5 min. After amplification, the product was purified with QIAquick® PCR Purification method (QIAGEN, Valencia, CA) and the concentration was measured using NanoDrop; and iii) WT-Ovation™ Pico RNA Amplification System (NuGEN, San Carlos, CA): This method is designed for preparing amplified cDNA from a small amount of total RNA (as low as 500 pg) for gene expression analysis. As described in the protocol provided by the manufacturer, three steps including first strand cDNA synthesis, second strand cDNA synthesis and SPIA amplification were performed. For the first strand cDNA synthesis, primer annealing was at 65°C for 2 min, then 4°C forever; first strand synthesis was 4°C for 1 min, 25°C for 10 minutes, 42°C for 10 min, 70°C for 15 min, then 4°C forever. In the second strand cDNA synthesis, the thermal cycling condition was 4°C for 1 min, 25°C for 10 minutes, 50°C for 30 min, 70°C for 5 min, then 4°C forever. In the final SPIA amplification step, the heating condition was 4°C for 1 min, 47°C for 60 min, 95°C for 5 min, then 4°C forever. Before SPIA amplification step, Beckman Coulter's Agencourt® RNAClean® purification beads were added into reaction tube to purify products. After amplification, the product was purified with QIAquick® PCR Purification method (QIAGEN, Valencia, CA), and the concentration was determined using NanoDrop.

Products from whole transcriptome amplification were single strand DNA (ssDNA), which needs to be converted to double strand DNA (dsDNA) for sequencing and clone library construction. Hence, WT-Ovation™ Exon Module (NuGEN, San Carlos, CA) was employed to synthesize the second strand cDNA. As the protocol illustrated, reaction tubes containing template and reagents were cultivated in the

following thermal cycling condition: 95°C for 5 min for primer annealing; 4°C for 1 min, 30°C for 10 min, 42°C for 60 min, 75°C for 10 min, then 4°C forever for second strand cDNA generation. The double strand cDNA (ds cDNA) obtained was then purified with QIAquick® PCR Purification method (QIAGEN, Valencia, CA), determined using NanoDrop, and then cloned and sequenced.

Clone library analysis. Products from the whole-transcriptome amplification were sheared into pieces (100 ~ 300 bp) by sonication, and then end-repaired with End-It™ DNA End-Repair Method (EPICENTRE Biotechnologies, Madison, WI). Fragments after end repair were ligated to pSMART® vector and used to transform *E. coli* 10G Competent Cells (Lucigen, Middleton, WI). Transformed cells were spread on nutrient agar plates containing the antibiotic kanamycin (25 µg/mL). Twenty of the grown transformed bacterial colonies were randomLy chosen with toothpicks and used in colony PCR with the primers provided by the cloning method to check the presence of cloned inserts and their lengths. The amplification was carried out in a total volume of 20 µL, containing 4 µM of each primer, 0.2 µL of Taq DNA polymerase (Fermentas Life Science, Pittsburgh, PA), 10 uL of PCR Master Mix (2X) (Fermentas Life Science, Pittsburgh, PA) and 7.8 µL of DEPC-treated water (Ambion, Austin, TX). Amplification involved an initial denaturation at 95°C for 2 min, then 30 cycles of 94°C for 15 sec, 50°C for 15 sec and 72°C for 30 sec, followed by a final 7 min extension at 72°C. Results were observed on an agarose gel by electrophoresis. Confirmed clones were used for plasmid isolation and sequencing using an ABI 373 Sequencer (Applied Biosystem Inc., Carlsbad, CA). Sequencing reads were edited using Sequence Scanner V1.0 (Applied Biosystem Inc., Carlsbad, CA). Vector sequences were identified and removed using

VecScreen program available in the NCBI online service. The DNA sequences were annotated using BLAST programs available from the NCBI database.

Metatranscriptomic analysis. Roche GS FLX Titanium chemistry pyrosequencing (Roche 454 Life Sciences, CT) was employed to sequence the amplified cDNA products. Approximate 2-3 µg of the amplified double-strand DNA from each sample was used as templates. The whole sequencing process included eight steps: product fragmentation, library preparation, binding single strand fragments to their own unique beads, emulsion PCR, clonally amplified bead purification, loading beads onto a plate, sequencing, and final pyrosequencing including signal processing of the images. Annotation, phylogenetic classification, and comparison of data obtained from each sample were performed using MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology) online service (Meyer et al. 2008).

Results and Discussion

RNA isolation. Out of 30 liters of deep-sea water, approximately 10^6-10^8 cells were collected from each sample, based on the estimated cell density of collected samples being about 10^2 to 10^3 cells per liter of deep-sea water, even lower than the reference data (10^3 – 10^4 cells per mL) reported before (Takai et al. 2004). Cells from four washes were combined together and used for RNA isolation. RNA was successfully isolated using our modified protocol combining Trizol (Invitrogen, Carslbad, CA) and RNeasy methods (QIAGEN, Valencia, CA) (Figure 6). In order to get rid of possible inhibitors carried in the samples, the purified RNA samples were washed and precipitated additionally 1-2 times before they were used for cDNA synthesis and amplification.

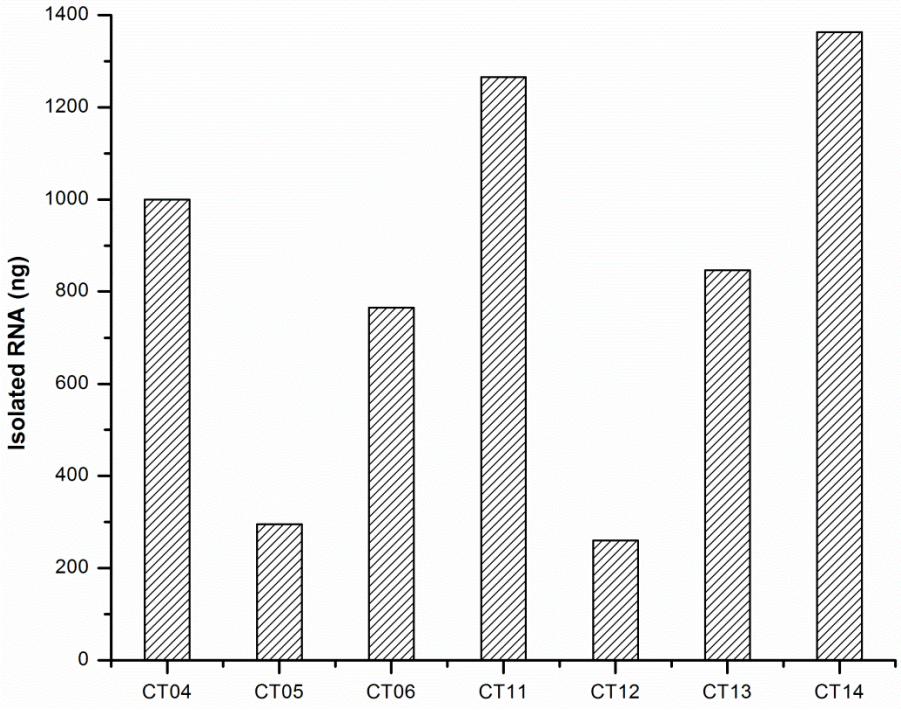

Figure 6. RNA isolation amount from different samples

Whole-transcriptome amplification using three different methods. We chose sample CT14-057 as the first trial target to develop the whole-transcriptome amplification protocol. This sample was collected from a site closest to the hydrothermal vent and a relatively higher abundant RNA was recovered. It has been reported previously that the microbial samples from hydrothermal vent fields were often contaminated with various inorganic inhibitors and thus were difficult for molecular biology manipulation (Sorensen et al. 2004). Hence, the protocol developed from this challenging sample was expected to have a good chance to be successful with other samples from similar environments.

We first evaluated the application of GE Illustra GenomiPhi V2 DNA Amplification method to our RNA sample. In an early study, 500 ng RNA of bacterial Tropheryma whipplei was used to synthesize cDNA with reverse transcriptase, and then

amplified successfully using the Genomiphi DNA amplification method to reach µg level DNA for microarray analysis (Van La et al. 2007). However, when this method was applied to our sample as described in the protocol of the method, no reasonable amplification result was achieved. Little increase of DNA amount was observed based on a NanoDrop reading (Table 2). One of the possible reasons for the low amplification efficiency could be that the overall poor quality of the RNA isolated from natural environments is poor; the sample may carry inhibitors from the environments.

Table 2. Comparison of whole transcriptomic amplification efficiencies of three different methods

Samples	Methods	Template (ng)	Product (ng)	Negative control (ng)	Amplification times
	GE Healthcare	9.6*	10.0	-	1
CT14-057	QIAGEN (12 h ligation)	10.9	2,060	100	189
	Nugen	5.0	16,260	600	3,252
CT04-020	Nugen	20	6,425	600	321
CT05-025	Nugen	5.9	14,025	600	2,377
CT05-030	Nugen	15.3	6,905	600	451
CT11-042	Nugen	3.5	4,857	-	1,388
CT12-047	Nugen	5.2	6,060	600	1,165
CT13-052	Nugen	4.5	15,759	-	3,502

* indicates cDNA as template in the corresponding experiment.

We then evaluated the application of the QIAGEN QuantiTect Whole Transcriptome Amplification (WTA) method to our RNA samples. The method has previously been used for amplifying 40 – 80 µg viral RNA as an alternative to random RT-PCR (Berthet et al. 2008). Amplifying viral RNA by WTA provides considerably better sensitivity and accuracy of detection compared to random RT-PCR. This method is different from the Illustra Genomiphi method in that it has an extra 'ligation' step before amplification. In general, the longer the amplification time, or when amplification time is fixed, the longer the ligation time is, the higher the amplification will be. However, longer ligation or amplification time can also result in higher amplification in negative controls. We started this method by optimizing the ligation and amplification time using the RNA purified from *E. coli*, and the results showed that 12 hrs of ligation time and 3 hrs of amplification time gave the best amplification for *E. coli* RNA with only minimal amplification in the negative control (data not shown). When the optimized method was applied to our deep-sea sample, we produced ~2,060 ng of amplified products from ~10.9 ng of template RNA for sample CT14-057, whereas only ~100 ng DNA in the negative control (Table 2).

We finally evaluated the application of the WT-Ovation™ Pico RNA Amplification System (NuGEN, CA) on our samples. Previously, this method has been compared to three other commercial protocols, i) Arcturus RiboAmp™ system (Applied Biosystem, Carlsbad, CA), ii) Ambion MessageAmp™ (Ambion, Austin, TX), and iii) Epicentre TargetAmp™ (EPICENTRE Biotechnology, Madison, WI), for amplification of picogram amounts of input RNA for microarray expression profiling (Clement-Ziza et al. 2009). In the study, the WT-Ovation pico system proposed by Nugen appears to be the

most suitable for RNA amplification. This method applied a 'cleaning-bead' step to purify the template before the final amplification step, which has been found to be very useful in minimizing amplification in the negative control (Boehme et al. 2009). When applied to our samples, the method achieved 16,260 ng of the final amplification products from ~5.0 ng RNA template, with only ~600 ng DNA in the negative control (Table 2). Compared with two other methods we evaluated, the NuGEN method demonstrated better amplification and relatively higher reproducibility on whole-transcriptome amplification. When applied to the rest of our deep-sea samples, the method generated a similar level of amplification (i.e. CT04-020, CT05-025, CT05-030, CT11-042 and CT12-047), suggesting a good reproducibility of the method (Table 2).

The DNA product obtained directly from the NuGEN method was in single-strand form. In order to perform pyrosequencing and other molecular biology operation, single-strand DNA needs to be converted to double strands. For this step, we used a NuGEN WT-Ovation™ Exon Module to synthesize the second strand cDNA. From the NanoDrop reading, the products almost exactly doubled after synthesis of the second strands, suggesting the overall good quality of the single-strand templates (Figure 7). The λ260/λ280 readings of double-stranded DNA products were around 1.90, indicating relatively good product quality.

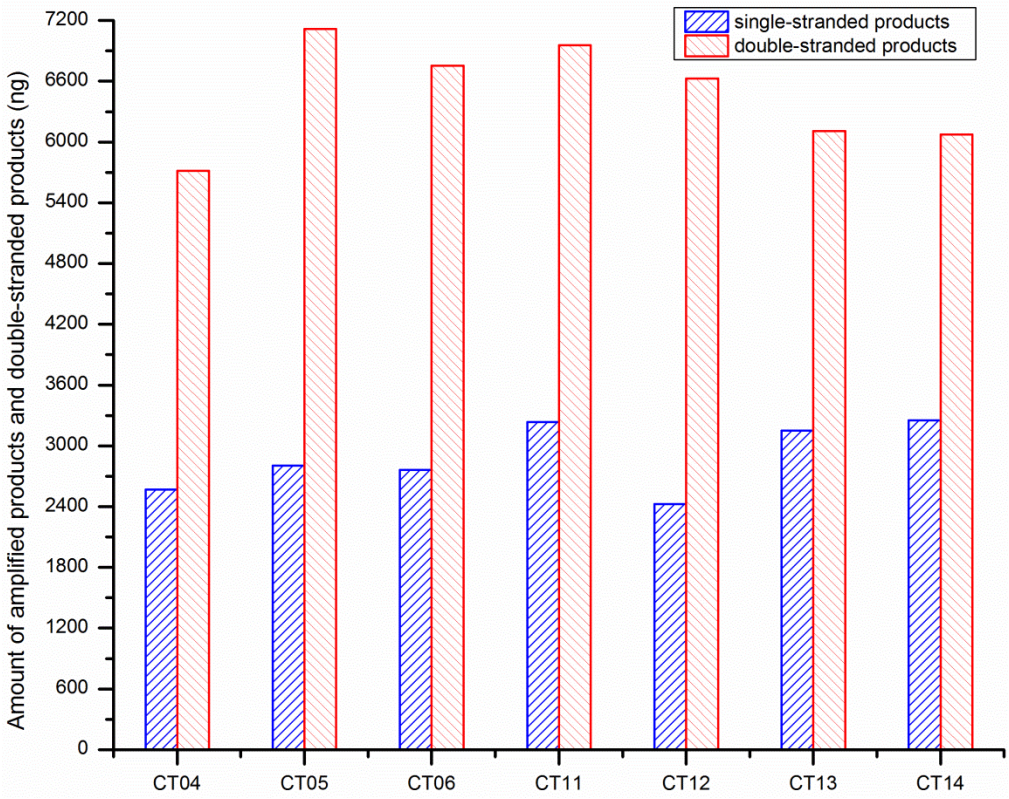

Figure 7. Products' amount before and after 2nd strand synthesis.

Evaluation of amplification product 'quality' by clone library analysis. The quality of the amplification product is crucial to pyrosequencing results. To confirm the amplified DNA products were indeed from our deep-sea samples, we used the whole-transcriptome amplification products for clone library construction and analysis. Approximately 20 colonies were randomLy picked from the clone libraries constructed for each of the samples, and plasmids were extracted for sequencing. The sequences obtained were then annotated by BLAST search against the NCBI database. The results showed that almost all BLAST hits (i.e. the organisms) were marine-based. Some of the top BLAST hits are summarized in Table 3. Their domination in the libraries was consistent with the source of these samples. The marine microorganisms identified from the clone libraries included *Calanus marshallae*, which is a dominant marine zooplankton

in boreal and arctic waters (Frost 1974), *Pareucalanus attenuates* which was previously found in northern South China Sea (Tseng et al. 2008), and *Paraeuchaeta* sp. which was a kind of copepoda found in the Norwegian Deep (northern North Sea) (Klimpel et al. 2004). Other identified marine microorganisms included: *Warnowia* sp. (Greuet 1978), *Nitrosopumilus maritimus* (Mincer et al. 2007), *Euchirella rostromagna* (Boysen-Ennen et al. 1991), *Silicibacter* sp. (González et al. 2003), *Pontella fera* (Jeong et al. 2008), *Scolecithricella longispinosa* (Hwang 2006), *Scaphocalanus brevicornis* (Kosobokova et al. 1997), *Pareucalanus attenuates* (Madhupratap & Haridas 1990), *Nematodinium* sp. (Gailhard et al. 2002), *Chlorobium luteolum* (Imhoff 2001), *Cibicidoides* sp. (Katz et al. 2003), *Epistominella exigua* (Schnitker 1980), *Ceratium horridum* (Burns 1982), and *Vorticella* sp. (Kimor et al. 1992). They have been studied or mentioned in many previous studies on marine ecosystems and inhabitants. The green algae, *Ceratium horridum*, which was normally found in freshwater was also identified in the clone libraries (Wei & Yu 2005).

Table 3. Typical microorganisms in different samples and corresponding e-values.

Name	e-value
Calanus simillimus	4.00E-07
Thermoanaerobacter mathranii subsp. mathranii str. A3	0.031
Anatoma euglypta isolate S34	9.00E-20
Uncultured marine bacterium clone SJC1.9	5.00E-21
Paraeuchaeta pseudotonsa	6.00E-13
Warnowia sp.	3.00E-82
Nitrosopumilus maritimus SCM1	2.00E-17
Euchirella rostromagna	9.00E-10
Uncultured marine picoeukaryote clone	2.00E-14
Silicibacter sp. TM1040	9.00E-09
Pontella fera	3.00E-11
Uncultured Myxococcales bacterium clone	2.00E-37
Scolecithricella longispinosa	9.00E-28
Euchaeta indica	3.00E-37
Bathypolypus arcticus actin gene	7.00E-31
Scaphocalanus brevicornis	2.00E-23
Calanus marshallae	2.00E-17
Pareucalanus attenuates	4.00E-07
Nematodinium sp.	7.00E-84
Bacillus anthracis str. A0248	8.00E-06
Chlorobium luteolum DSM 273	0.2
Uncultured gamma proteobacterium clone	7.00E-15
Uncultured Mycobacteriaceae bacterium	1.00E-19
Cibicidoides sp.	6.00E-17
Epistominella exigua	2.00E-09
Staurastrum punctulatum chloroplast	3.00E-27
Ceratium horridum	8.00E-37
Vorticella sp.	5.00E-15

Metatranscriptomic analysis by Roche GS FLX Titanium pyrosequencing. We employed Roche GS FLX Titanium chemistry pyrosequencing for metatranscriptomics analysis. The typical sequencing plate contains 11 sections, of which we used only 7 for metatranscriptomic analysis in this study. Preliminary analysis of the pyrosequencing results and its comparison with other recent metatranscriptomic studies are summarized in Table 4.. In our study, we obtained a total reads number of 120,000. Considering the reading number was generated from only 7/11 of the plate, a total reading number of 200,000 could be expected if a whole plate was used. The average read length of a transcriptome is 186 bps, which was in accordance with the given size distribution of amplified cDNA products in the Roche GS FLX protocol (mostly in the range of 100 – 300 bps). In several recent metatranscriptomic studies, microbial samples from ocean surface water and soil were collected for pyrosequencing (Frias-Lopez et al. 2008; Gilbert et al. 2008; Urich et al. 2008; Vila-Costa et al. 2010). However, since the microbial abundance is high in surface water, no whole transcriptome amplification was involved in these studies. The comparison of our metatranscriptomic analysis with these studies is provided in Table 5. The results show that a comparable pyrosequencing quality has been achieved by using the DNA amplified with our protocol.

Table 4. Pyrosequencing data summary and the comparison with some recent studies.

Source	Total number of reads (K)	Total basepairs (Mbp)	Ave. reading length (bp)	Reference
Amplified total RNA	120	22.3	186	This thesis
Amplified mRNA	128	14.7	114	Frias-Lopez et al. 2007
Genomic DNA	414	45.4	110	Frias-Lopez et al. 2007
Amplified RNA	-	117.4	215	Gilbert et al. 2008
Amplified mRNA	303	-	209	Vila-Costa et al. 2010
Total RNA	258	-	96	Urich et al. 2008

Since the major goal of this thesis is to introduce the method to amplify whole-transcriptome from deep-sea samples. A preliminary overview of the metatranscriptomic data is provided in Table 5. The 16S based analysis showed that bacteria are the dominant microbes in the environments, consistent with mRNA based analysis although the mRNA based sequence readings have significant hits for small eukaryotes as well. In general, 8.70-42.76% of 16S sequences and 17.71-63.85% of mRNAbased sequences have no significant hits in the databases, suggesting that many novel microbial species may be present in these deep-sea environments. For most of the samples, the number of the mRNA-based sequence readings was much higher than the number of the 16S based

sequence readings (except sample CT11-042). The results showed the method has better coverage on protein-based (mRNA-based) sequences than 16S rRNA-based sequences, consistent with the manufacturer's claim that NuGEN WT-Ovation™ Pico RNA Amplification System was mRNA favorable, which was preferable for generating more mRNA-based reads for functional analysis of microbial communities. As a comparison, in a recent metatranscriptomic analysis of total RNA isolated from soil samples without any amplification steps, mRNA-based analysis yielded 21,133 reads, and rRNA-based analysis generated 193,219 reads (Urich et al. 2008).

Table 5. Overview of metatranscriptomics analysis.

Name	mRNA-based readings (%)				16s-based readings (%)			
	Archae	Bacteria	Eukaryota	Other*	Archae	Bacteria	Eukaryota	Other*
CT04-020	0.23	12.66	40.45	46.66	0.00	85.82	0.00	14.18
CT05-025	0.00	3.21	69.08	17.71	0.00	82.05	0.00	17.95
CT05-030	1.59	16.79	22.07	59.55	0.00	78.14	0.00	21.86
CT11-042	1.38	14.83	32.41	51.38	0.00	92.50	0.00	7.50
CT12-047	0.25	5.54	30.36	63.85	0.00	57.24	0.00	42.76
CT13-052	0.47	0.47	76.42	22.64	0.00	83.33	0.00	16.67
CT14-057	0.00	3.45	70.00	26.55	0.00	91.30	0.00	8.70

* Other indicates viral sequences and sequences without blast hits

Conclusions

We reported a new protocol to isolate and amplify RNA from small amounts (nano gram level) of deep-sea microbial samples. The protocol included three key steps: total RNA isolation by a modified method combining Trizol (Invitrogen, Carlsbad, CA) and RNeasy methods, amplification with WT-Ovation™ Pico RNA Amplification System (NuGEN, San Carlos, CA), and then conversion of single-stranded cDNA to double-stranded DNA with the NuGEN WT-Ovation™ Exon Module. Based on the results from the clone library sequencing and the preliminary analysis of the Roche GS FLX Titanium chemistry pyrosequencing, we conclude that the method is suitable for metatranscriptomic study of microbial samples of impurity and low cell density. With more attention paid to community structure and functionality of microbial communities present in extreme environments, such as the deep-sea, cold springs, and deep frozen earth where very low microbial cell density and high concentration of potential inhibitors are often expected, the protocol developed here could be a very useful tool in applying direct next-generation based metatranscriptomic analysis.

CHAPTER 4
INTEGRATED METAGENOMIC AND METATRANSCRIPTOMIC ANALYSES OF MICROBIAL COMMUNITIES IN THE MESO- AND BATHYPELAGIC REALM OF THE NORTH PACIFIC OCEAN

Introduction

The water body underlying the photic zone in the world's oceans, representing the largest water mass on earth (comprising 1.3×10^{18} m^3), is the largest aqueous habitat for microbial life (Orcutt et al. 2011). This realm differs distinctly from the photic zone, presenting low temperature (approximately 2~4^{o}C), high pressure (250 atmospheres in the depth of 2,500 m)and high inorganic nutrient levels (Arístegui et al. 2009). Differences in physical geochemical parameters between the upper level of sea water and deep sea suggest that microbial communities in these environments are confronted by fundamentally different challenges. Although emerging evidence indicates that deep-sea water contains an untapped reservoir of high metabolic and genetic diversity, and microbial communities in deep-sea water play an important role in ocean biogeochemistry (Arístegui et al. 2002, 2005, 2009; Teske et al. 2011), this realm has not been studied well when compared with surface sea water. This is probably due to the difficulties in sampling and the significant costs incurred (Arístegui et al. 2009). Recent studies have found that prokaryotic plankton is one of the main drivers of biogeochemical cycles over large ocean expanses (Schattenhofer et al. 2009), and that eukaryotic microbes account for a significant fraction of the biomass and activity of marine microbial communities (Massana & Pedrós-Alió 2008; Díez et al. 2001). To better understand the influence of microbes on ocean geochemistry, we set about exploring the

structure and metabolic characteristics of microbial communities in deep-sea environments.

Progress in next-generation sequencing is fueling a rapid increase in the number and scope of microbial community-targeted studies (Frias-Lopez et al. 2008; Gilbert et al. 2008; Poretsky et al. 2009; Vila-Costa et al. 2010; Stewart et al. 2011; Rinta-Kanto et al. 2011). While metagenomics provides information on the taxonomic composition and metabolic potential of a microbial community, metatranscriptomics serves to unveil the actual metabolic activities of the community at a specific time and place, and how those activities change in response to environmental forces or biotic interactions (Mitra et al. 2011). Defining the relationship between microbial community composition and function (metabolic characteristics) has been a major challenge in studying heterotrophic carbon cycling in marine systems (Teske et al. 2011). The combination of metagenomic and metatranscriptomic approaches has proven efficacious in deciphering the phylogenetic composition, metabolic potential and pathways of deep-sea microbial communities. For example, in a recent study, coupled metagenomic and metatranscriptomic analyses were utilized for taxonomic and functional characterization of marine microbial communities living at depths between 25-500 m (Shi et al. 2011). The results provided novel insight into not only microbial diversity but also specific metabolic processes transpiring in the ecosystems. Moreover, the different relative abundance of taxonomic groups identified in the metagenomic and metatranscriptomic libraries arising from the study revealed differential relative transcriptional activities per cell (Shi et al. 2011).

In this chapter, we present a study focused on simultaneous metagenomic and metatranscriptomic analysis of deep-sea microbial communities, including both

prokaryotes and eukaryotes, from four different deep-sea sites ranging from the mesopelagic to pelagic ocean (depth of 784 m ~ 1,937 m, 2 ~ 4 m above the sea floor). Many marine microbial community studies have focused on either prokaryotes or eukaryotes (Martín-Cuadrado et al. 2007; Frias-Lopez et al. 2008; Gilbert et al. 2008; Rinta-Kanto et al. 2011; Roussel et al. 2011; Eloe et al. 2011; Stewart et al. 2011). Whole-genome and whole-transcriptome amplifications were performed to meet the entry requirements of nucleic acids for pyrosequencing. Integrated metagenomic and metatranscriptomic results revealed a defensive life style instead of active growing/metabolic life style of both prokaryotic and eukaryotic communities living in the deep-sea water. Microbial community structures and their metabolic characteristics in the environments are presented and discussed.

Materials and Methods

Microbial sample collection and DNA/RNA isolation. Microbial samples for this study were collected from waters in four deep-sea Northeast Pacific Ocean sites during the R/V Thompson TN 221 Research Cruise between July 28th and August 2nd, 2008, as described in Table 6. Among these sampling sites, CT04, CT05 and CT06 are 164.2 km east of Newport, Oregon, while CT12 is much further away from the coast, lying approximately 20 km southwest of a deep-sea hydrothermal vent in the Juan de Fuca Ridge. At each site a total of 30 liters of deep-sea water was collected in six rigid, tightly-sealed 5-liter bottles carried on the Woods Hole Oceanographic Institute's Towed Camera System ('TowCam'). The entire TowCam sampling excursion into the deep sea and return to the surface took about 1 h. Deep sea water was immediately filtered through 0.22-μm filters which were then stored at -80° C until used. The filtration time for each

single filter (sample) was about 0.5 h. To preserve the RNA profiles' integrity, cells were gently washed from the filter membranes with chilled RNAlater solution (Ambion, Austin, TX). A total of five 5.0-mL washes were performed for each filter, and cells recovered from the five washes were combined into a single pool and stored immediately at -80° C for later DNA/RNA isolation. Genomic DNA was isolated with the DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA) following the bacterial DNA isolation protocol provided by the manufacturer. Total RNA was extracted from the collected microbial cells using a modified protocol combining Trizol (Invitrogen, Grand Island, NY) and the prescribed RNeasy methods (QIAGEN, Valencia, CA). The detailed procedures for sampling, storage and nucleic acid isolation have been described in Chapter 3.

DNA/RNA amplification, second strand cDNA synthesis, and quality examination. Genomic DNA and total RNA (rRNA and mRNA) were isolated, purified and amplified for pyrosequencing. Briefly, the Ovation WGA system (NuGEN, San Carlos, CA) was employed to amplify isolated DNA following the manufacturer's protocol. After amplification, the products were purified with the QIAquick® PCR Purification kit (QIAGEN, Valencia, CA). The concentration and purity were determined using a NanoDrop (Thermo Scientific, West Palm Beach, FL). Total RNA amplification was performed using the WT-Ovation™ Pico RNA Amplification System (NuGEN, San Carlos, CA). The products from whole-transcriptome amplification were purified and measured in the same way as DNA amplification products. DNA and RNA amplification products, both single-stranded, were converted into double-stranded products before conducting clone library construction and sequencing using the WT-Ovation™ Exon

Module (NuGEN, San Carlos, CA) method described in Chapter 3. To check the quality of amplification products and preclude the possibility of external contamination, we constructed clone libraries for the amplified products of each sample. The isolated plasmid DNA/cDNA from 20 random clones from each library were sequenced using the ABI 373 Sequencer (Applied Biosystem Inc., Carlsbad, CA) and BLAST annotated against the NCBI databases.

Data analysis. We used the MG-RAST metagenomics analysis server provided by Argonne National Laboratory to perform phylogenetic and functional analysis on the metagenomic and metatranscriptomic data (Meyer et al. 2008). In MG-RAST version 3, reads are considered replicates if the first 50 bp are identical. De-replication was processed when uploading data. The M5NR, a searchable database integrated from several existing sequence databases, including KEGG (Kyoto Encyclopedia of Genes and Genomes), NCBI (National Center for Biotechnology Information), SEED (The SEED Project), and COG (Clusters of Orthologous Groups of proteins), was used to perform organism hierarchical classification and functional hierarchical classification searches in the pyrosequencing data (Meyer et al. 2008). All searches were performed using the default parameters suggested by the MG-RAST server.

Results and Discussion

Overview of data generation and analysis. Among the four sampling sites in this study, CT04, CT05, and CT06 are 164.2 km west of the coast, while CT12 is much further out to sea, approximately 20 km southwest of a deep-sea hydrothermal vent. Detailed information about the sampling sites is provided in Table 6. Although all sampling sites are close to the sea floor (2-3 m above the sea floor), sites CT05 and CT06

are in the mesopelagic realm (200-1,000 m), while CT04 and CT12 are within the bathypelagic realm (1,000-4,000 m). Environmental parameters, such as pressure and nutrient level, varied according to depth and location of the sites. By studying samples from four disparate sites we were better able to reveal a more complete picture of microbial communities' structures and their metabolic characteristics in the deep-sea water of the Northeast Pacific Ocean.

In natural environments such as the deep sea, microbial cell density can be as low as 10^3 – 10^4 cells per mL of seawater (Kawagucci et al. 2012). In addition, RNA abundance decreases with depth due to a relatively slow metabolism (Danovaro et al. 1999). Compared to biomass-abundant samples such as surface seawater, the intact RNA that can be isolated from these environments is very limited in quantity: usually in the low to several hundred nanogram level. On the other hand, typical commercial next-generation sequencing platforms require 3-5 µg or more of input DNA/cDNA to produce reliable sequencing data, approximately equal to the total amount of DNA isolable from 10^9 *Escherichia coli* cells. Due to the limited samples collected from the deep-sea water, an amplification process was employed in this study for both genomic DNA and total RNA samples to provide sufficient DNA and cDNA for metagenomic and metatranscriptomic analysis. Total RNA was used for metatranscriptomic analysis, allowing simultaneous assessment of rRNA and mRNA to produce both taxonomic and metabolic information on the studied microbial communities (Urich et al. 2008; Shi et al. 2011).

Table 6. Description of deep-sea biomass sampling.

Sampling sites	Position	Depth (m)	Sampling time
CT04	44˚29′34.80″N 125˚8′49.61″W	1181 ~ 1194	7/23/2008 21:00 ~ 21:24 GMT
CT05	44˚34′1.02″N 125˚9′3.75″W	785 ~ 790	7/27/2008 3:05 ~ 7:21 GMT
CT06	44˚33′52.99″N 125˚9′3.73″W	763 ~ 789	7/27/2008 3:12 ~ 7:31 GMT
CT12	45˚51′57.01″N 129˚47′19.47″W	1840 ~ 1913	8/2/2008 00:01 ~ 00:24 GMT

Using the amplification method introduced in Chapter 3, 4.9 ~ 16 μg of the amplification products after purification was generated from 5.0 ~ 20.0 ng of DNA/RNA template, with only ~600 ng DNA detected in negative controls. The statistical summary of the sequenced data determined by MG-RAST is shown in Table 7. The Nugen Ovation WGA system (NuGEN, San Carlos, CA) used for whole genome amplification has been evaluated by the manufacturer and demonstrated the ability to faithfully replicate genomic DNA (https://gs107.genesifter.net/users/?mod=ngd&action=view_result&analysis_id=1583&ras_analysis_file_id=2717197&kegg=1). The performance of WT-Ovation™ Pico RNA Amplification System (NuGEN) used for total RNA amplification in this study has been evaluated in two other studies and shown to be reproducible with minimal bias (Morse et al. 2010; Skalitzky 2010). Pyrosequencing of community DNA and RNA across four deep-sea sampling sites generated 160,072 and 64,928 sequencing reads (after quality control and de-replication), with mean lengths of 225.5 bp and 182.2 bp, respectively.

Table 7 shows the microbial community compositions of different samples as revealed by metagenomic and metatranscriptomic analysis. The results demonstrate that the taxonomic compositions of microbial communities as revealed by metagenomic analysis differ markedly from those obtained by metatranscriptomic analysis, especially for archaeal and bacterial groups, suggesting the importance of using both approaches to avoid possible methodological bias. In metagenomic and metatranscriptomic data, the proportion of eukaryotic reads was much higher than that of archaeal and bacterial reads (Table 7). The proportion of eukaryotic reads (58.73-71.92% based on metagenomic analysis, and 73.45-87.54% based on metatranscriptomic analysis) in this study without pre-filtration were almost an order of magnitude higher than those in several previous studies using pre-filtered sea water, which were typically under 5% (DeLong et al. 2006; Martín-Cuadrado et al. 2007; Rusch et al. 2007), but was similar to one recent study which reported higher proportions of eukaryotes at different depths (10 m – 25.60%; 800 m – 48.08%; 4,400 m – 37.27%) in the North Pacific Ocean (Brown et al. 2009).

Table 7. Statistical summary of the sequencing results of amplified deep-sea microbial communities' nucleotide acids.

	DNA				RNA			
	CT04	CT05	CT06	CT12	CT04	CT05	CT06	CT12
Total reads	61,650	46,012	50,491	40,049	12,804	13,722	18,367	33,509
Ave. length (bp)	217 ± 143	228 ± 145	252 ± 156	236 ± 144	170 ± 113	184 ± 120	177 ± 124	202 ± 121
Ave. GC%	49 ± 8	50 ± 8	48 ± 8	48 ± 8	49 ± 6	36 ± 6	49 ± 7	48 ± 6
Failed QC	15,938	11,537	11,133	0	3,745	4,421	5,443	0
Annotated protein	8,545	7,328	9,070	12,720	201	0	0	10,300
Unknown protein	18,581	14,359	17,011	15,723	4,602	5,790	6,981	13,569
Ribosomal RNA	1,505	1,042	999	2,811	1,932	1,009	3,663	6,175
Unknown	17,081	11,746	12,278	8,795	2,324	2,502	2,280	3,465

To better define the characteristics of deep-sea microbial communities, we also employed, as a reference for the data analysis, metagenomic data from the microbial community isolated from surface waters (sampling depth: 1 m) of Browns Bank, Gulf of Maine, in the Global Ocean Sampling (GOS) Expedition project (Rusch et al. 2007). We chose this data set as our reference in part because the timing of the GOS project's sample collection was similar to that in our study (GOS: Aug. 21, 2003); and in part because its sampling location, especially the latitude, was also relatively close to that of our study (GOS: +43° 39' 53.95", -65° 33' 50.78").

Metagenomic analysis of the deep-sea prokaryotic communities. When compared with archaea, bacterial reads represented more than 90% of the prokaryotic sequencing reads from both metagenomic and metatranscriptomic data for all sampling sites, suggesting that bacteria are absolutely dominant in the prokaryotic communities in the deep-sea water samples. This is similar to previous studies conducted on soil, surface water, deep sea, and marine sediment (DeLong et al. 2006; Rusch et al. 2007; Urich et al. 2008; Brown et al. 2009; Liao et al. 2011). However, archaea were typically found at higher levels in the marine sediment than in sea water (Schippers et al. 2012). Figure 8 shows the compositions of prokaryotic communities in metagenomic and metatranscriptomic data from four sampling sites and the reference surface water community revealed by metagenomic data (Rusch et al. 2007). The results revealed remarkably high microbial diversity even though the cell densities were much lower in the deep sea. Prokaryotic communities of the GOS surface water and the four deep-sea sampling sites diverged significantly in terms of phylogenetic composition at broad levels of phyla and classes (Figure 8). In the surface water, the proportion of archaea was less

than 1% of the prokaryotic community, while in the deep sea, it was increased to as high as 13.54% (CT06). This result is in accord with a general trend observed in multiple ocean basins: the proportion of archaea increases with depth (Karner et al. 2001; Herndl et al. 2005). The changes from GOS surface water to deep-sea water were primarily manifest by the emergence of archaeal phyla *Crenarchaeota*, *Euryarchaeota* and *Thaumarchaeota*, bacterial phyla *Actinobacteria* and *Firmicutes*, sub-phyla *Betaproteobacteria*, *Deltaproteobacteria* and *Gammaproteobacteria*, and the decreasing abundances of bacterial phyla *Bacteroidetes* and *Alphaproteobacteria*. Among them, phylogenetic lineages within chemolithoautotrophic *Thaumarchaeota*, *Betaproteobacteria* and *Gammaproteobacteria* were recognized as ammonia oxidizing archaea (AOA) and ammonia oxidizing bacteria (AOB) that can oxidize ammonia to nitrite (Pester et al. 2012). This process is the first and rate-limiting step in nitrification and is also a vital component of the global biogeochemical nitrogen cycle (Gruber & Galloway 2008). The AOB *Betaproteobacteria* and *Gammaproteobacteria* are considered a major mediator of ammonia oxidation processes (Purkhold et al. 2000). A previous study has confirmed that at least some AOA (e.g. *Nitrosopumilus*, one of the main members identified in the *Thaumarchaeota* group in this study) have a high ammonia affinity and can grow in extremely oligotrophic environments (Schleper & Nicol 2010). Considering the trace substrate concentration in the deep sea, AOA *Thaumarchaeota* may be the major nitrifier in that environment (Pester et al. 2011). Figure 8 also reveals a slight increase of *Planctomycetes* in the deep sea, all of which are members of the well-known anammox bacterial genus identified previously in the marine sub-oxic zone including *Rhodopirellula*, *Blastopirellula*, *Planctomyces*, *Pirellula*,

Candidatus Kuenenia, *Gemmata* and *Isosphaera* (Kirkpatrick et al. 2006; Byrne et al. 2009). The presence of *Planctomycetales* in the low-oxygen, dark pelagic ocean is an indicator that anaerobic ammonium oxidation (anammox), a globally important microbial process of the nitrogen cycle, may be another metabolic pathway supporting primary production in this environment (Kuypers et al. 2005). More information on environmental parameters, especially oxygen concentration, will be required before a definitive conclusion can be reached. *Actinobacteria* and *Firmicutes* had higher representation in the deep-sea prokaryotic community compared to surface water, probably due to the adaptive advantage of *Actinobacteria* and *Firmicutes* under low-nutrient conditions of the deep sea (Gärtner et al. 2011). Moreover, it has been proposed that the ability of *Actinobacteria* to survive in cold and dystrophic environments might be due to its adaptive ability to go into resting states with low metabolic activity (Simon et al. 2009).

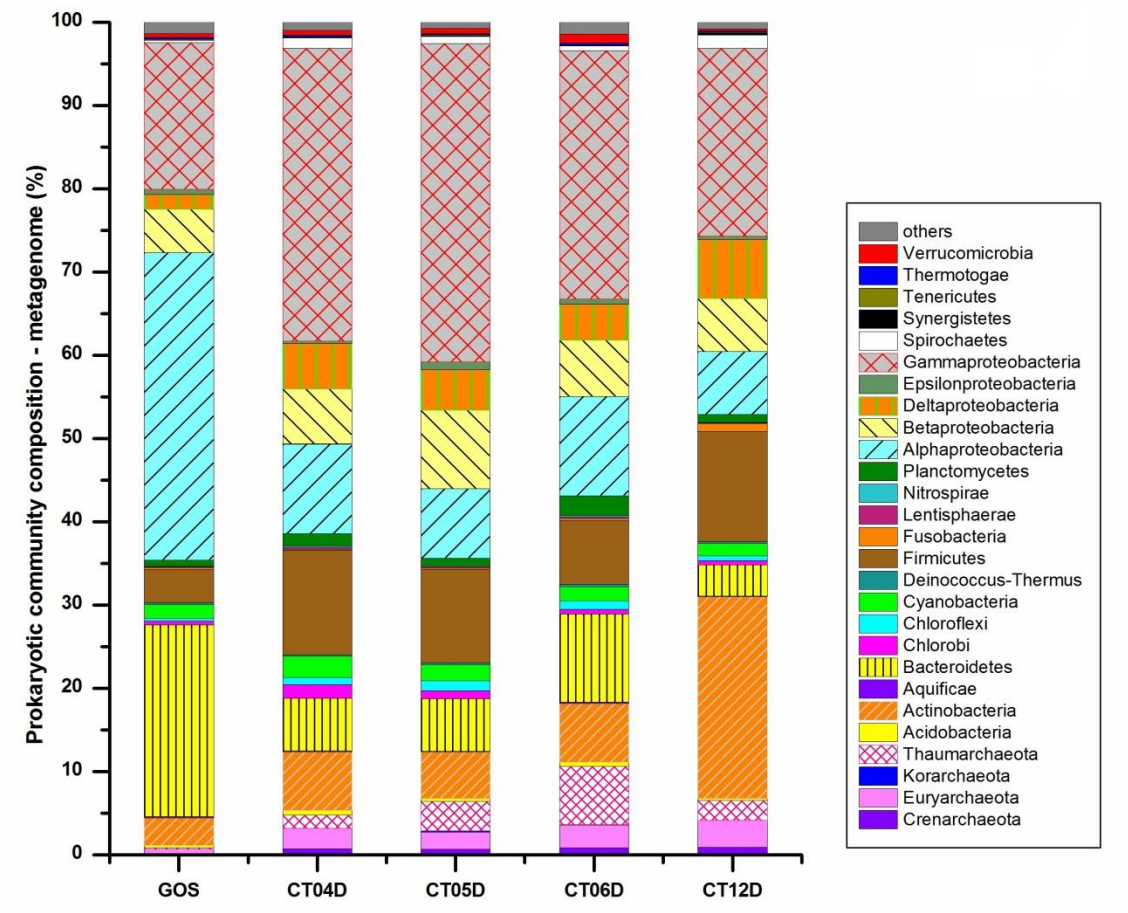

Figure 8. Comparison of prokaryotic microbial communities' composition in four deep-sea sampling sites and in GOS surface water.

Our results demonstrated a dramatic decrease of *Alphaproteobacteria* in the deep sea (Figure 8), consistent with a previous study at the Hawaii Ocean Time-series (HOT) station ALOHA (DeLong et al. 2006). In the *Alphaproteobacteria* group, *Candidatus pelagibacter*, an abundant member of the SAR11 clade (Sun et al. 2011) and one of the most abundant groups of bacteria in the upper surface waters of the oceans (Rusch et al. 2007; Viklund et al. 2012), was found to be the most conspicuously decreased genus in the deep-sea samples. The SAR11 clade was found to contribute to the variability in utilization of nutritional compounds (glucose, ATP, a combination of amino acids, and organic compounds, the sources of C, N, and P) by the bacterial community, and its

activity can be linked to the bacterial community's activity as a whole because of its ability to adapt to nutrient limitation (Laghdass et al. 2012). Hence, the decreased representation of *Candidatus pelagibacter* in the deep-sea prokaryotic communities indicated lower levels of nutrient utilization and bacterial community activity in the deep sea relative to surface water. In the surface water, class *Flavobacteria* dominated within the *Bacteroidetes* phylum, while in the deep sea the taxonomic composition within *Bacteroidetes* changed to a mixed assemblage of *Bacteroidia* and *Prevotella*. This represented a shift from aerobic to anaerobic phylotype, that is, aerobic *Flavobacteria* in surface waters giving way to most anaerobic *Bacteroidia* and *Prebotella* spp. in the deep-sea water, consistent with the decrease in dissolved oxygen (DO) in deep sea relative to surface environments.

Metatranscriptomic analysis of the deep-sea prokaryotic communities. Of the 38 total prokaryotic phyla identified in metagenomic libraries from the four samples, there were 20 phyla in CT04, 10 phyla in CT05, 17 phyla in CT06, and 26 phyla in CT12 also detected in the metatranscriptomic libraries (Figure 9). Of the four deep-sea sampling sites, CT12 harbored the most diverse metabolically active prokaryotes. In the metatranscriptomic libraries of all four samples, Gammaproteobacteria constituted the highest proportion, similar to the metagenomic libraries. Compared to prokaryotic community composition as explored from metagenomic data, proportions of *Crenarchaeota*, *Thaumarchaeota*, *Alphaproteobacteria*, *Bacteroidetes* and *Planctomycetes* were decreased markedly, while *Euryarchaeota*, *Deltaproteobacteria*, *Firmicutes* and *Actinobacteria* were increased in the metatrascriptome libraries (Figure 9). The archaeal phyla *Korarchaeota* and *Nanoarchaeota*, and bacterial phyla *Poribacateria*,

Chrysiogenetes, *Deferribacteres*, *Elusimicrobia*, *Fibrobacteres*, *Gemmatimonadetes*, *Lentisphaerae* and *Zetaproteobacteria*, which made up a minor portion of the metagenomic libraries, were completely absent in the metatranscriptome libraries from all four deep-sea sites. As discussed in more detail below, the relative abundances of these taxonomic groups (phylum level) were very different in the metagenomic and metatranscriptomic libraries, indicating differential relative transcriptional activities per cell (Shi et al. 2011).

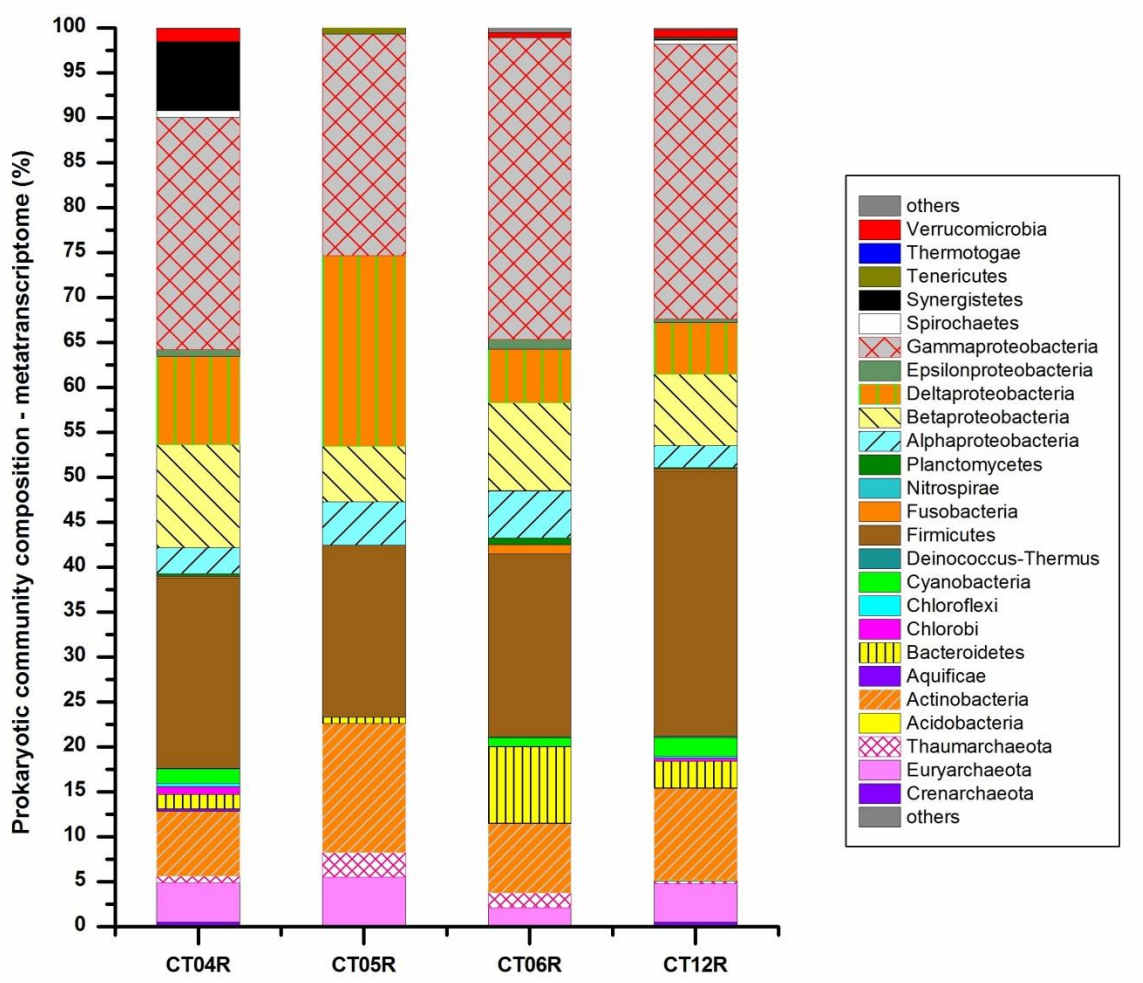

Figure 9. Distribution of metabolically active prokaryotes in different deep-sea microbial communities.

Integrated metagenomic and metatranscriptomic analysis of prokaryotic communities in deep-sea water. From the metagenomic analysis, a total of 453, 462, 598 and 564 different prokaryotic genera were identified for site CT04, CT05, CT06 and CT12, respectively; while from the metatranscriptome analysis, the numbers of detected genera were 113, 34, 138 and 332. The number of species identified and overall community diversity revealed by metatranscriptomic data was relatively low compared to the metagenomic data. Metagenomic analysis provides better coverage of microbial species but lacks the information afforded by metatranscriptomic analysis on the metabolic activities of the communities described. Given the differences observed between metagenomic and metatranscriptomic analysis, it is clear that neither metagenome-based nor metatranscriptome-based analysis is sufficiently comprehensive to fully characterize a microbial community. In this study we defined a parameter to measure the strength of metabolic activity at the genus level as the ratio of transcript abundance in the RNA pool to gene abundance in the DNA pool. The abundances calculated and provided by MG-RAST are counts of taxon (each abundance represents the number of times a particular taxon is detected) or function (each count represents the number of times a particular functional role is detected) (Meyer et al. 2008). The calculation was based on the assumption that both the DNA and RNA used for sequencing analysis were isolated from equal numbers of cells (Values calculated by MG-RAST). Thus, we did not standardize the abundance to dataset size. Prokaryotic genus with very low abundance values (as low as 1), or lack of either DNA or RNA data, was excluded from the analysis. Figure 10 shows the RNA/DNA ratio of the major prokaryotic classes identified in the deep-sea samples. We found that the RNA/DNA

ratio of single phylum/sub-phylum varied considerably among the deep-sea sites, likely reflective of the different metabolic conditions the prokaryotic communities were experiencing in the varied environments (Figure 10). Overall, the prokaryotic community in site CT12 (20.0 km away from a hydrothermal vent on the Juan de Fuca Ridge (Wang et al. 2009) showed the highest RNA/DNA ratios among the four sites for almost all detected phyla except *Synergistetes*, consistent with the relatively higher temperature and mineral-rich environment which support highly diverse and more active communities of microbes around hydrothermal vents (Schrenk et al. 2003; Xie et al. 2010). Due to the small taxonomic coverage from the RNA pool, RNA/DNA ratios were missing for most of these phyla in CT05, except for phyla *Euryarchaeota*, *Thaumarchaeota*, *Actionbacteria*, *Firmicutes* and *Proteobacteria*.

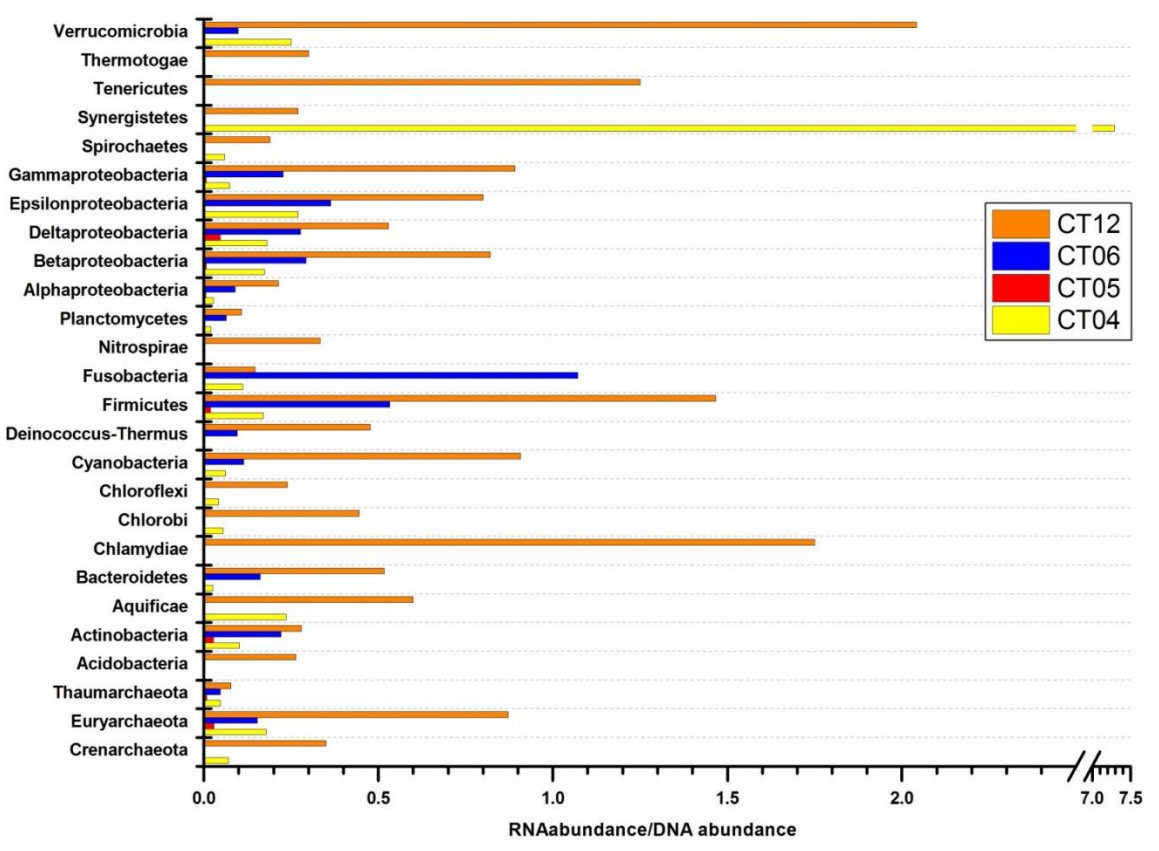

Figure 10. Metabolic activity strength comparison using RNA/DNA ratios of prokaryotic communities in the deep sea.

Figure 11 shows the RNA/DNA ratios of some interesting prokaryotic genera identified in the studied samples. The RNA/DNA ratio was employed as an index to show strength of metabolic activity (SMA) of prokaryotes in deep-sea environments. Piezophilic bacteria *Shewanella*, thermophilic archaea/bacteria, sulfur/sulfate reducing bacteria, methanotrophic archaea/bacteria, genera within the green sulfur bacteria *Chlorobia* group, and *Cyanobacteria* genera detected in the deep-sea samples were chosen as targets. Of all the sites, overall, deep-sea site CT12 harbored the largest number of genera showing high potential SMA (Figure 11), even for the photoautotrophic *Cyanobacteria* group. Considering the metabolic characteristics of these targeted genera (piezophilic, thermophilic, methanotrophic and sulfur/sulfate

reducing), along with the unique physico-chemical properties of hydrothermal vents or proximal sites (Roussel et al. 2011), we infer that though CT12 is 20 km away from the hydrothermal vent, it is still under its environmental influence.

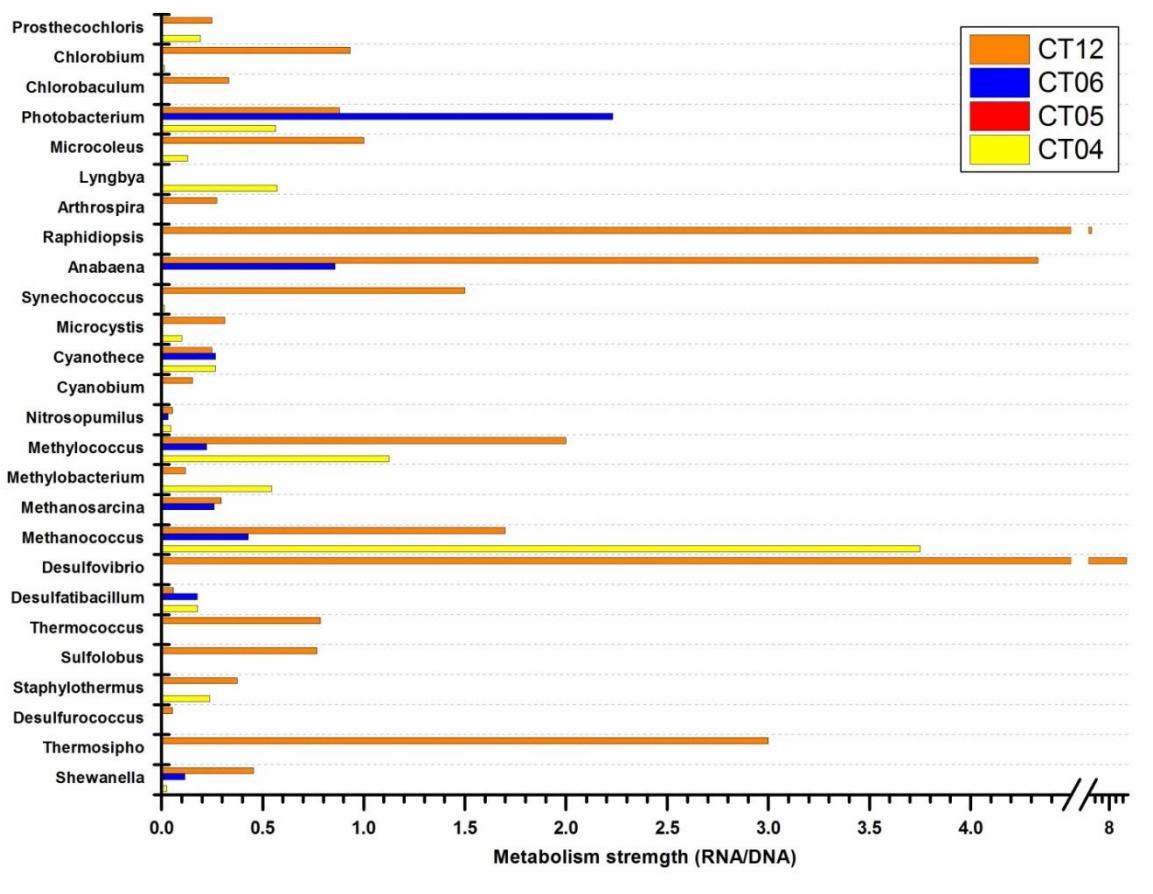

Figure 11. Metabolic activity strength comparison using RNA/DNA ratios of selected prokaryotic genera.

Functionality analysis of deep-sea prokaryotic communities. To functionally annotate the sequences collected from each site, the metagenome and metatranscriptome reads obtained were searched against Clusters of Orthologous Groups (COG) of the National Center for Biotechnology Information (NCBI). The COG function clusters revealed in the DNA and RNA data showed huge differences in their diversity and the number of reads assigned to each category (Table 8). Unique functions identified in

metatranscriptomic data were less diverse than those in the metagenomic data, probably due to the fact that only a fraction of prokaryotic organisms were metabolically active and expressing their functional genes or because, even though many genes were expressed, their expression levels were lower than the built-in detection threshold of MG-RAST. Assuming the density of sea water to be 1,025 kg/m^3, pressure increases by 1 atm with each 10 m of depth (www.calctool.org). Based on this premise, we estimated that the hydrostatic pressure in our sampling sites ranged from 77.29 ~ 190.78 atm, with the highest pressure present at the deepest site CT12. Hence, in contrast to the microbial communities near the surface, genes involved in COG clusters detected in deep-sea metagenomic libraries showed characteristics strongly associated with the high hydrostatic pressures including 'cell wall/membrane/envelope biogenesis', 'cytoskeleton', 'defense mechanisms', 'signal transduction mechanisms', 'replication, recombination and repair', and 'inorganic ion transport and metabolism' (Figure 12). The transcriptional potential of protein-coding genes varied remarkably among the sampling sites. In particular, genes from GOS surface water had a higher representation in COG clusters including 'cell cycle control, cell division, chromosome partitioning', 'amino acid transport and metabolism', 'carbohydrate transport and metabolism', 'lipid transport metabolism', 'nucleotide transport and metabolism', 'secondary metabolites biosynthesis', and 'transport and catabolism' (Figure 12).

Table 8. Taxonomic compositions of metagenomic and metatranscriptomic libraries.

	Metagenome (%)				Metatranscriptome (%)			
	CT04	CT05	CT06	CT12	CT04	CT05	CT06	CT12
Archaea	1.31	1.97	4.23	1.60	1.24	0.80	0.64	1.05
Bacteria	26.56	30.27	34.30	24.16	23.06	9.42	17.14	19.98
Eukaryotes	69.94	65.73	58.73	71.92	73.45	87.54	81.40	78.37
Viruses	0.41	0.29	0.47	0.44	0.30	0.50	0.23	0.25
Others*	1.78	1.74	2.26	1.88	1.96	1.77	0.59	0.35

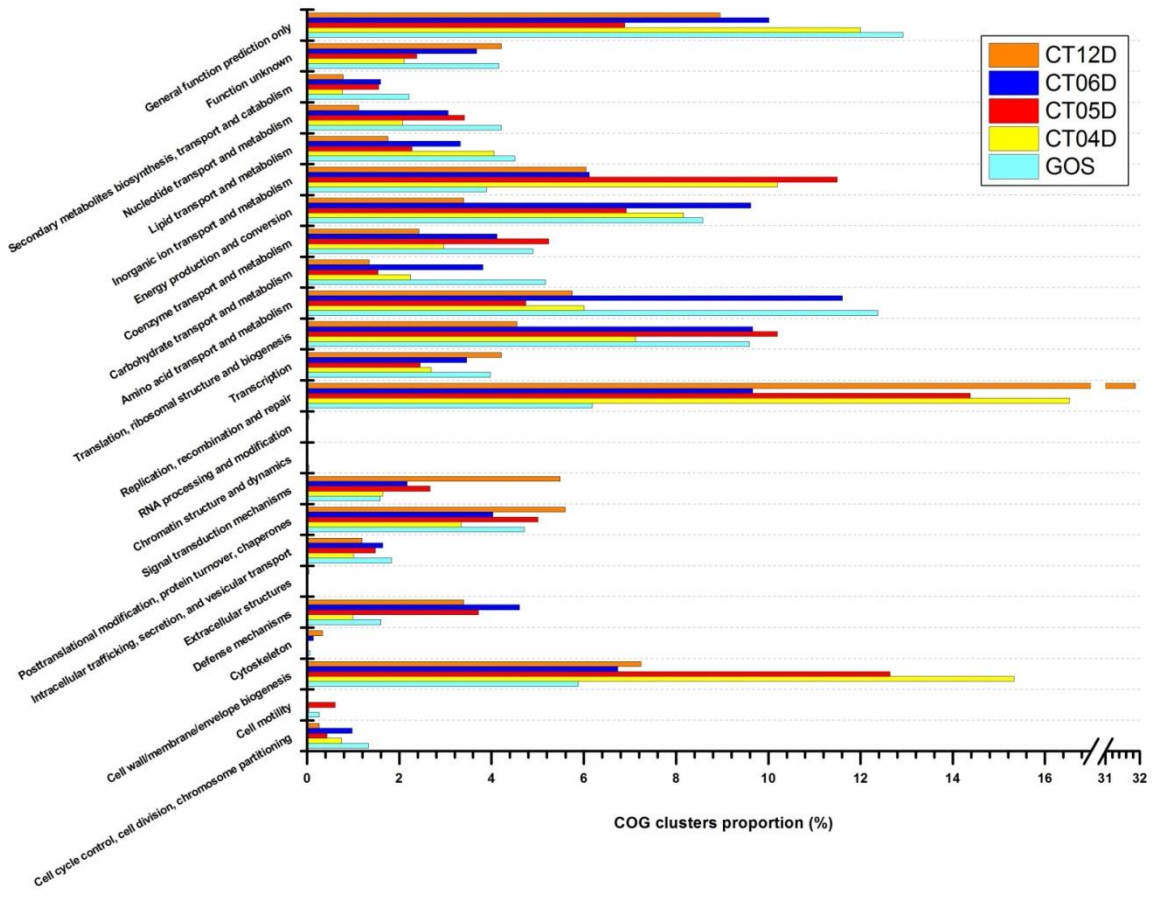

Figure 12. COG distributions of prokaryotic metagenomic libraries.

Cell wall/membrane/envelope biogenesis. Several functional groups related to lipopolysaccharide biosynthesis were found, including LPS: glycosyltransferases (COG1442), Dihydrodipicolinate synthase/N-acetylneuraminate lyase (COG0329), Glycosyltransferases involved in cell wall biogenesis (COG0463), and a predicted sugar phosphate isomerase involved in capsule formation (COG0794). Lipopolysaccharide (LPS) constitutes the outermost leaflet of the outer membrane of gram-negative bacteria (El-Hajj et al. 2009), while dihydrodipicolinate synthase (DHDPS) catalyses the first step in the biosynthetic pathway producing meso-diaminopimelate (DAP) and (S)-lysine, required components of the cell wall (Domigan et al. 2009). The higher proportion of

‘Cell wall/membrane/envelope biogenesis’ related genes in deep-sea samples compared to surface water is probably due to the requirement for cell wall integrity under high pressure (Shi et al. 2011).

Signal transduction mechanisms. COG cluster ‘signal transduction mechanisms’ including Signal transduction histidine kinase (COG 0642), FOG: PAS/PAC domain (COG 2202), FOG: CheY-like receiver (COG0784), a predicted membrane GTPase involved in stress response (COG1217), and FOG: GGDEF, and GAF domain (COG2199, COG 2203) were identified in the metagenomic data of deep-sea prokaryotic communities. Signal transduction histidine kinase (COG 0642) and FOG: PAS/PAC domains (COG 2202) are two groups that were particularly enriched in deep-sea protein coding genes. Genes encoding histidine kinase are important for chemotaxis and quorum sensing (Wolanin et al. 2002). The PAS domain is integral to proteins that sense environmental stimuli such as oxygen and redox potential (Taylor & Zhulin 1999). The proportion of signal transduction pathways evident in deep-sea prokaryotic communities was higher than in GOS surface water, a manifestation of the need for deep-sea prokaryotic communities to sense and adapt to dystrophic deep sea environments, and consistent with previous findings from a study on microbial communities at 6,000 m depth in the Puerto Rico Trench (Eloe et al. 2011).

Replication, recombination and repair. In this category, Recombinational DNA repair protein (RecF pathway) (COG0353) and Adenine specific DNA methylase Mod (COG2189) are major functions among the functions array detected in the metagenomic data from the deep-sea water samples. Methylation of DNA by the DNA adenine methylase plays an important role in DNA mismatch repair and replication regulation

(Heusipp et al. 2007). The enrichment of DNA repair protein coding genes probably indicated the strong DNA repair capacity of prokaryotes in deep-sea water to protect themselves from DNA degradation in the harsh environments (Nakai et al. 2011).

Defense mechanisms. ABC-type multidrug transport systems, ATPase and permease components (COG1132) and Cation/multidrug efflux pump (COG0841) are two main function groups detected in this category. Multidrug efflux systems contribute significantly to intrinsic and acquired resistance to antimicrobials in some gram-negative bacteria (Poole 2004). This efflux system may be a defensive mechanism for deep-sea microorganisms against external aggression and toxins (Nakai et al. 2011), increasing their likelihood of survival (DeLong et al. 2006).

Inorganic ion transport and metabolism. Outer membrane receptor proteins, mostly Fe transport (COG1629), is the most dominant function detected in this category for all four deep-sea sampling sites. Iron is an important component of most redox enzymes (Braun & Braun 2002). We also detected the sequences of functional genes associated with membrane transport of heavy metals such as Co, Cu, Zn, Mn and Cd in site CT12. It has been reported that some microorganisms in hydrothermally active areas can remove heavy metals from hydrothermal fluid (Jannasch 1995). The presence of these functional genes supported our hypothesis that the microbial community in CT12 was under the influence of the hydrothermal vent 20 km away.

In metatranscriptomic libraries of deep-sea prokaryotic communities, far more limited COG clusters were identified compared to the metagenomic libraries. The transcriptional activity of protein-coding genes varied distinctly among the four sampling sites. In total, 24 functional categories were identified in the metagenomic libraries; 6 in

CT04, 9 in CT06 and 15 categories in CT12 were characterized. Only two COG groups, Transcriptional regulator (COG0583) and Ribosomal protein S3 (COG0092), were identified in CT05. Under the cell wall/membrane/envelope biogenesis category, the two bathypelagic sites CT04 and CT12 showed significantly higher percentages than the other two sites. In the metagenomic pool, this category accounted for 15.33% in CT04 and 7.24% in CT12; while in the metatranscriptomic pool, the percentages were increased to 60.87 and 40.53, respectively. While the increase may be due to the smaller number of COG categories detected in the metatranscriptomic pool, it still shows that mRNA related to cell wall/membrane/envelope biogenesis was more highly expressed in the two deeper sites compared to the shallower CT05 and CT06 sites. 'Replication, recombination and repair' was a category occupying a high percentage in all deep-sea sites except CT05. Serine/threonine protein kinase (COG0515) and Adenine specific DNA methylase Mod (COG2189) enriched this category. Sites CT12 and CT06 showed relatively higher percentages of the category 'Inorganic ion transport and metabolism', which was enriched by Outer membrane receptor proteins (COG1629) in CT06, and Outer membrane receptor proteins (COG1629) and ABC-type phosphate/phosphonate transport system ATPase component (COG3638) in CT12. Hence, the defensive life style of prokaryotes in the deep sea could be inferred from metatranscriptomic analysis (Martín-Cuadrado et al. 2007; Konstantinidis et al. 2009) since the detected mRNA COG categories were mostly related to cell wall/membrane and capsule formation for high pressure resistance, nucleotide repair, and membrane transporter for virulence.

Active presence of photosynthetic bacteria in deep-sea water. Photoautotrophic microbes such as genera *Prochlorococcus* and *Synechococcus* are abundant members of

the microbial community in the euphotic zone and responsible for much of the primary production in the ocean (Herndl et al. 2008; Walsh et al. 2009). Meanwhile, early studies have suggested that viable photoautotrophic microbes such as *Synechococcus* sp. can be found in deep sea water or sediments exposed to little or no sunlight (Martín-Cuadrado et al. 2007; Vilibić & Šantić 2008). However, these studies provided no evidence as to whether these photoautotrophic microbes were metabolically active in the deep sea, or merely present as dormant cells. In Chapter 2, in addition to direct auto-fluorescence imaging, the presence of photoautotrophic microbes in the deep mesopelagic zone (765-790 m) was demonstrated by both bacterial 16S and 23S rRNA-based clone library analysis, and the metabolic activity of the oxygenic photoautotrophic bacteria *Cyanobacteria Synechococcus* sp. was demonstrated by expression level quantification of the 23S rRNA gene involved in protein synthesis and photosynthesis-involved psbA gene using RT-qPCR (Gao et al. 2011). Here we employed metagenomic and metatranscriptomic analysis to further demonstrate the existence of active photosynthetic microbes in samples from four deep-sea sites. Photoautotrophic bacterial phylum *Cyanobacteria* (including genera *Cyanobium*, *Cyanothece, Synechococcus*, *Anabaena*, *Cylindrospermopsis*, *Nodularia*, *Nostoc*, *Arthrospira*, *Lyngbya*, *Microcoleus* and *Oscillatoria*) were found by both metagenomic and metatranscriptomic analysis in samples from all four sites, further suggesting the presence of metabolically active photosynthetic bacteria in the pelagic ocean realm. SEED subsystems identified in both metagenomic and metatranscriptomic pools are listed in Table 9. In contrast to the diverse *Cyanobacteria* subsystems identified in metagenomic data, only a few were identified from the metatranscriptomic data. This might be due to a combination of the

low quality of the RNA samples and the low metabolic activity of deep-sea *Cyanobacteria*. In site CT05, no *Cyanobacteria* SEED subsystems were detected. The Photosystem II protein D1 encoding gene (*psbA*) was detected in the metatranscriptomic pool of CT06, which belongs to *Nostoc* sp. (strain PCC 7120 / UTEX 2576), suggesting that a 'photosynthesis' pathway may be still maintained when *Cyanobacteria* find themselves in deep-sea environments. Counterintuitively, the most diverse cyanobacteria subsystems were identified in samples from the deepest site, CT12, the site also most close to the hydrothermal vent.

Table 9. Cyanobacteria identified in metagenomic and metatranscriptomic libraries.

Samples	Level 1	Function	Closest Species
CT04D	Amino Acids and Derivatives	Proline iminopeptidase (EC 3.4.11.5)	*Nostoc punctiforme* (strain ATCC 29133 / PCC 73102)
		Gluconolactonase (EC 3.1.1.17)	*Synechococcus* sp. CC9311
	Carbohydrates	Acetate permease ActP (cation/acetate symporter)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
		Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22)	*Synechococcus* sp. (strain WH7803)
	Clustering-based subsystems	Phytoene desaturase (EC 1.14.99.-)	*Nostoc punctiforme* (strain ATCC 29133 / PCC 73102)
		Adenylate cyclase	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
	Cofactors, Vitamins, Prosthetic Groups, Pigments	5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193)	*Synechococcus elongatus* PCC 7942
		Diaminohydroxyphosphoribosylaminopyrimidine deaminase (EC3.5.4.26)	*Synechococcus elongatus* PCC 7942
	Fatty Acids, Lipids, and Isoprenoids	Phytoene desaturase (EC 1.14.99.-)	*Nostoc punctiforme* (strain ATCC 29133 / PCC 73102)
	Miscellaneous	Adenylate cyclase (EC 4.6.1.1)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
		Serine peptidase (Alpha/beta hydrolase superfamily) fused to N- terminal uncharacterized domain specific to cyanobacteria	*Synechococcus* sp. CC9311
		Na+/H+ antiporter	*Nostoc punctiforme* PCC 73102
	Potassium metabolism	Potassium voltage-gated channel subfamily KQT	*Synechococcus* sp. RCC307
	Protein Metabolism	Acetate permease ActP (cation/acetate symporter)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)

		Selenide,water dikinase (EC 2.7.9.3)	*Microcystis aeruginosa* (strain NIES-843)
	Regulation and Cell signaling	Adenylate cyclase (EC 4.6.1.1)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
		ATP synthase alpha chain (EC 3.6.3.14)	*Crocosphaera watsonii* WH 8501
	Respiration	Cytochrome c oxidase polypeptide I (EC 1.9.3.1)	*Prochlorococcus marinus* NATL1A
		soluble [2Fe-2S] ferredoxin	*Cyanothece* sp. (strain PCC 7425 / ATCC 29141)
	Virulence, Disease and Defense	Negative regulator of beta-lactamase expression	*Synechococcus* sp. WH 7805
CT04R	Monosaccharides	Ribokinase (EC 2.7.1.15)	*Synechococcus* sp. RS9917
	Sugar utilization in Thermotogales	Ribokinase (EC 2.7.1.15)	*Synechococcus* sp. RS9917
CT05D	Amino Acids and Derivatives	5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)	*Synechococcus* sp. CC9605
	Carbohydrates	5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)	*Synechococcus* sp. CC9605
	Carbohydrates	decarboxylase	*Prochlorococcus marinus* MIT 9313
	Clustering-based subsystems	Adenylosuccinate synthetase (EC 6.3.4.4)	*Cyanothece* sp. PCC 7425
	Cofactors, Vitamins, Prosthetic Groups, Pigments	5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)	*Synechococcus* sp. CC9605
	Miscellaneous	Scaffold protein for [4Fe-4S] cluster assembly, MRP-like, similar to chloroplast-targeted plant protein HCF101	*Synechocystis* sp. (strain ATCC 27184 / PCC 6803 / N-1)
	Nucleosides and Nucleotides	Adenylosuccinate synthetase (EC 6.3.4.4)	*Cyanothece* sp. PCC 7425
	Photosynthesis	photosystem II protein D1 (PsbA)	*Synechococcus* sp. RCC307

	Protein Metabolism	Arginyl-tRNA synthetase (EC 6.1.1.19)	*Synechococcus* sp. CC9605
	Protein Metabolism	Chaperone protein HtpG	*Synechococcus* sp. (strain WH8102)
	Stress Response	Glutamate decarboxylase (EC 4.1.1.15)	*Prochlorococcus marinus* (strain MIT 9313)
CT06D	Amino Acids and Derivatives	Threonine dehydratase, catabolic (EC 4.3.1.19)	*Anabaena variabilis* (strain ATCC 29413 / PCC 7937)
		Urease accessory protein UreG	*Cyanobacteria* bacterium Yellowstone B-Prime (*Synechococcus* sp. JA- 2-3B'a(2-13))
		Threonine dehydratase, catabolic (EC 4.3.1.19)	*Anabaena variabilis* (strain ATCC 29413 / PCC 7937)
		Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	*Cyanothece* sp. PCC 7424
		Threonine dehydratase, catabolic (EC 4.3.1.19)	*Anabaena variabilis* (strain ATCC 29413 / PCC 7937)
		Transketolase (EC 2.2.1.1)	*Nostoc punctiforme* PCC 73102
		L-alanine:glyoxylate aminotransferase (EC 2.6.1.44)	*Cyanothece* sp. (strain PCC 7424) (*Synechococcus* sp. (strain ATCC 29155))
	Carbohydrates	Transketolase (EC 2.2.1.1)	*Nostoc punctiforme* PCC 73102
		Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22)	*Synechococcus* sp. CC9902
		Serine--pyruvate aminotransferase (EC 2.6.1.51)	*Cyanothece* sp. (strain PCC 7424) (*Synechococcus* sp. (strain ATCC 29155))
		Transketolase (EC 2.2.1.1)	*Nostoc punctiforme* PCC 73102
	Cell Division and Cell Cycle	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	*Prochlorococcus marinus* MIT 9303

Cell Wall and Capsule	Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22)	*Synechococcus* sp. CC9902
Clustering-based subsystems	Peptidyl-tRNA hydrolase (EC 3.1.1.29)	*Synechococcus elongatus* PCC 7942
	Molybdenum cofactor biosynthesis protein MoaD	*Thermosynechococcus elongatus* (strain BP-1)
Cofactors, Vitamins, Prosthetic Groups, Pigments	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	*Cyanothece* sp. PCC 7424
	Flavodoxin	*Synechococcus* sp. CC9311
	Flavodoxin 1	Synechococcus sp. CC9311
	Sulfur carrier protein adenylyltransferase ThiF	*Cyanobacteria bacterium* Yellowstone B-Prime (*Synechococcus* sp. JA- 2-3B'a(2-13))
Dormancy and Sporulation	Peptidyl-tRNA hydrolase (EC 3.1.1.29)	*Synechococcus elongatus* PCC 7942
Membrane Transport	Dipeptide transport ATP-binding protein DppD (TC 3.A.1.5.2)	*Cyanothece* sp. ATCC 51142
Nitrogen Metabolism	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	*Cyanothece* sp. PCC 7424
Nucleosides and Nucleotides	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	*Prochlorococcus marinus* MIT 9303
Photosynthesis	Octaprenyl-diphosphate synthase (EC 2.5.1.-)	-
Protein Metabolism	SSU ribosomal protein S6p	*Prochlorococcus marinus* SS120 (subsp. *marinus* CCMP1375)
	Peptidyl-tRNA hydrolase (EC 3.1.1.29)	*Synechococcus elongatus* PCC 7942
	Prolyl-tRNA synthetase (EC 6.1.1.15)	*Synechococcus* sp. CC9902
	Chaperone protein DnaJ	*Synechococcus* sp. WH 7803
	Urease accessory protein UreG	*Thermosynechococcus elongatus* BP-1

	Stress Response	Chaperone protein DnaJ	*Cyanobacteria bacterium* Yellowstone B-Prime (*Synechococcus* sp. JA- 2-3B'a(2-13)) *Synechococcus* sp. WH 7803
		5-oxoprolinase (EC 3.5.2.9)	*Cyanothece* sp. (strain ATCC 51142)
CT06R	Photosynthesis	photosystem II protein D1 (PsbA)	*Nostoc* sp. (strain PCC 7120 / UTEX 2576)
CT12D	Amino Acids and Derivatives	Glutamate N-acetyltransferase (EC 2.3.1.35)	*Synechococcus* sp. JA-2-3B'a(2-13)
		N-acetylglutamate synthase (EC 2.3.1.1)	*Synechococcus* sp. JA-2-3B'a(2-13)
	Carbohydrates	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
	Cell Wall and Capsule	Bacillosamine/Legionaminic acid biosynthesis aminotransferase PglE	*Prochlorococcus marinus* (strain MIT 9303)
		ADP-heptose synthase (EC 2.7.-.-)	*Cyanothece* sp. (strain PCC 7424) (*Synechococcus* sp. (strain ATCC 29155)
		D-glycero-beta-D-manno-heptose 7-phosphate kinase	*Cyanothece* sp. (strain PCC 7424) (*Synechococcus* sp. (strain ATCC 29155))
	Clustering-based subsystems	Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)	*Microcystis aeruginosa* NIES-843
		Glycosyltransferase	*Anabaena variabilis* ATCC 29413
		Glycosyltransferase (EC 2.4.1.-)	*Anabaena variabilis* ATCC 29413
		Translation elongation factor Ts	*Prochlorococcus marinus* (strain MIT 9303)
		Alanyl-tRNA synthetase (EC 6.1.1.7)	*Anabaena variabilis* ATCC 29413
		Amino acid permease in hypothetical Actinobacterial gene cluster	*Cyanothece* sp. (strain PCC 7425 / ATCC 29141)
		Helicase PriA essential for oriC/DnaA-independent DNA replication	*Cyanothece* sp. PCC 7425
		Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)	*Microcystis aeruginosa* NIES-843

	Competence/damage-inducible protein CinA	*Synechococcus* sp. (strain JA-3-3Ab) (*Cyanobacteria bacterium* Yellowstone A-Prime)
	Phosphomethylpyrimidine kinase (EC 2.7.4.7)	*Prochlorococcus marinus* (strain MIT 9215)
Cofactors, Vitamins, Prosthetic Groups, Pigments	Orotate phosphoribosyltransferase (EC 2.4.2.10)	*Synechococcus* sp. (strain RCC307)
	Phosphomethylpyrimidine kinase (EC 2.7.4.7)	*Prochlorococcus marinus* (strain MIT 9215)
	Sulfur carrier protein adenylyltransferase ThiF	*Cyanothece* sp. PCC 8801
DNA Metabolism	Helicase PriA essential for oriC/DnaA-independent DNA replication	*Cyanothece* sp. PCC 7425
	Membrane protein PxcA, involved in light-induced proton extrusion	*Synechocystis* sp. (strain ATCC 27184 / PCC 6803 / N-1)27184 / PCC 6803 / N-1)
	Glutathione S-transferase family protein	*Prochlorococcus marinus* (strain MIT 9313)
	Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)	*Microcystis aeruginosa* NIES-843
Miscellaneous	Glycosyltransferase	*Anabaena variabilis* ATCC 29413
	Phosphomethylpyrimidine kinase (EC 2.7.4.7)	*Prochlorococcus marinus* (strain MIT 9215)
	Competence/damage-inducible protein CinA	*Synechococcus* sp. (strain JA-3-3Ab) (*Cyanobacteria bacterium* Yellowstone A-Prime)
Nucleosides and Nucleotides	Orotate phosphoribosyltransferase (EC 2.4.2.10)	*Synechococcus* sp. (strain RCC307)
Phages, Prophages, Transposable elements, Plasmids	Heat shock protein 60 family chaperone GroEL	*Prochlorococcus marinus* MED4 (subsp. pastoris str. CCMP1378)
Photosynthesis	photosystem II protein D1 (PsbA)	*Cyanothece* sp. (strain PCC 8802) (*Synechococcus* sp. (strain RF-2))
	Translation elongation factor Ts	*Prochlorococcus marinus* (strain MIT 9303)
Protein Metabolism	Translation elongation factor Tu	*Cyanobacteria bacterium* Yellowstone A-Prime (*Synechococcus* sp. JA- 3-3Ab)
	Alanyl-tRNA synthetase (EC 6.1.1.7)	*Anabaena variabilis* ATCC 29413

		Valyl-tRNA synthetase (EC 6.1.1.9)	*Trichodesmium erythraeum* (strain IMS101)
		Prolyl endopeptidase (EC 3.4.21.26)	*Gloeobacter violaceus* PCC7421
		Heat shock protein 60 family chaperone GroEL	*Prochlorococcus marinus* MED4 (subsp. pastoris str. CCMP1378)
		4-keto-6-deoxy-N-Acetyl-D-hexosaminyl-(Lipid carrier) aminotransferase	*Prochlorococcus marinus* (strain MIT 9303)
	Respiration	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	*Anabaena* sp. PCC 7120 (*Nostoc* sp. PCC 7120)
	Stress Response	Glutathione S-transferase family protein	*Prochlorococcus marinus* (strain MIT 9313)
	Carbohydrates	Ribokinase (EC 2.7.1.15)	*Synechococcus* sp. RS9917
CT12R	Fatty Acids, Lipids, and Isoprenoids	Phytoene desaturase, pro-zeta-carotene producing (EC 1.-.-.-)	*Synechococcus elongatus* (strain PCC 7942) (*Anacystis nidulans* R2)
	Phosphorus Metabolism	Alkaline phosphatase (EC 3.1.3.1)	*Cyanothece* sp. CCY0110

The reason for the presence of metabolically active *Cyanobacteria* may be the sinking mechanism of marine particles. Cell aggregation phenomena are frequently observed in many *cyanobacterial* species (Koblížek et al. 2001). A study on particle interceptor traps at the Bermuda Atlantic Time-series Study found that the cyanobacteria *Prochlorococcus* and *Synechococcus* were consistently detected in the water column, and that they trap samples at different depths in the euphotic zone, showing that *Cyanobacteria* can contribute to downward particle flux (Amacher et al.). Further study is necessary to provide more information about the active presence of *Cyanobacteria* in dark, deep-sea environments.

Taxonomic and functional study of the deep-sea eukaryotic community. In line with our observations regarding the deep-sea prokaryotes, deep-sea eukaryotic communities displayed remarkably different compositions from those in the GOS surface water (Figure 13). Though eukaryotes accounted for a far higher proportion in the sampled deep-sea microbial communities (58.73 ~ 71.92% in metagenomic data) than in the GOS surface water (3.31% in metagenomic data) (Rusch et al. 2007), the diversity of the deep-sea eukaryotic community was very limited: In the GOS surface water, 25 different eukaryotic phyla were identified, while 17, 17, 18 and 20 eukaryotic phyla were identified at the deep-sea sampling sites CT04, CT05, CT06 and CT12, respectively. In contrast, fungi *Chytridiomycota*, flagellate *Euglenida*, parasite *Rhombozoa*, animal *Tardigrada*, and yellow-green algae *Xanthophyceae* were absent in all four deep-sea sites. Besides, the community changes from the GOS surface water to the deep-sea water were primarily manifest in the emergence of phyla *Chordata* and *Nematoda*, and the decrease of fungi *Ascomycota*, *Basidiomycota*, green algae *Chlorophyta*, plant

Streptophyta and *Cnidaria*. Fungus *Chytridiomycota*, *Glomeromycota*, *Microsporidia*, *Neocallimastigomycota*, algae *Phaeophyceae*, *Xanthophyceae*, diatom *Bacillariophyta* and other animal phyla were also identified in the deep-sea metagenomic data, although making up only a small proportion of the total. Because we studied the microbial diversity using non-size-fractionated deep-sea samples, animalia *Chordata* and Nematode, whose sizes are typically around 2.50 mm, turned out to be dominant groups here. The proportion and diversity of fungi was also found to be decreased in the deep-sea water. In the GOS surface water, 100 fungus species were identified; while in the deep-sea water 63, 66, 68 and 75 different fungus species were identified in sites CT04, CT05, CT06 and CT12, respectively. The decrease in fungus was also reported in a previous study which reported that, compared to surface water, fungi were rare and less diverse in high-pressure, deep-sea environments (Bass et al. 2007). Our results also identified three phyla belonging to eukaryotic parasites, including *Apicomplexa*, *Annelida* and *Nematode*. *Apicomplexa* were commonly found in deep-sea environments including hydrothermal vents (Gestal et al. 2010) and methane cold seeps (Takishita et al. 2007). *Annelida*, which we identified with very small abundance in site CT12, is one of the few parasites found in hydrothermal vents (de Buron & Morand 2004). Another detected eukaryotic parasite, *Nematodes*, is among the most abundant metazoan taxa in deep-sea ecosystems in general (Lambshead & Boucher 2003; Merckx et al. 2009; Vanreusel et al. 2010).

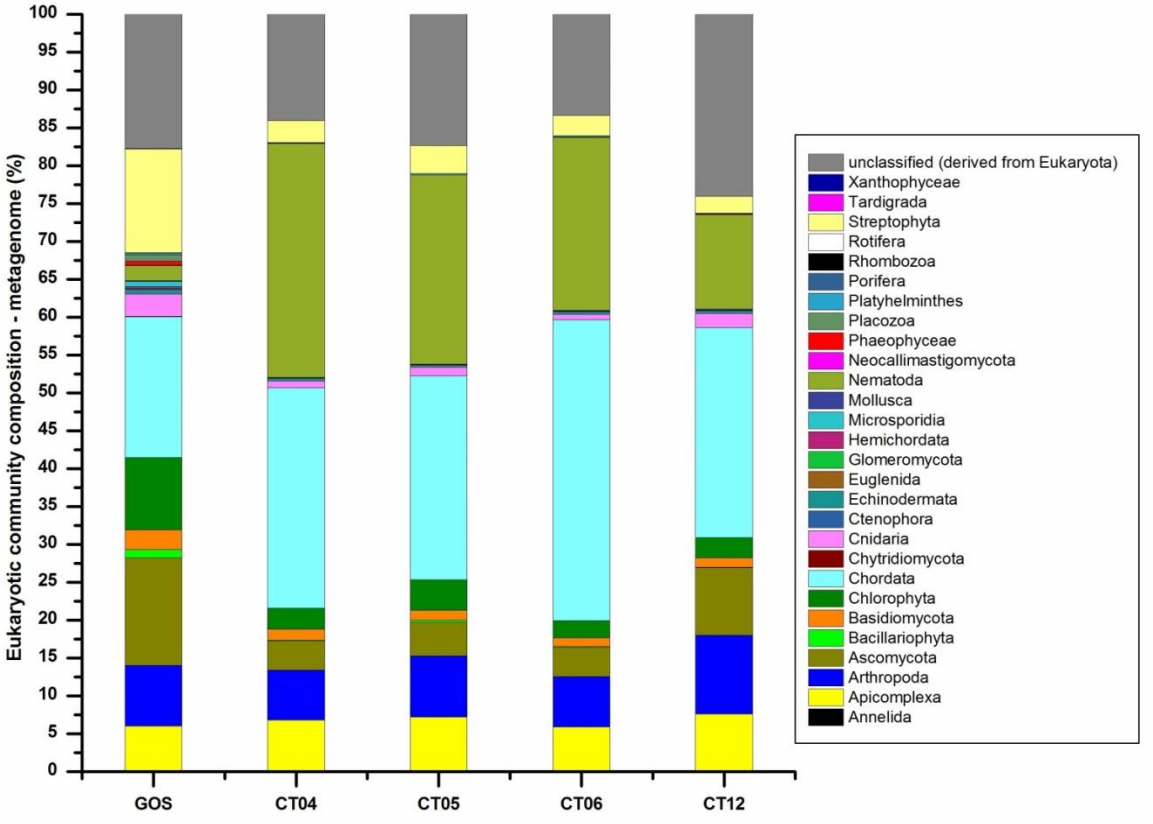

Figure 13. Eukaryotic community compositions in deep sea.

Searching against the COG database, we found a similar trend in functional categories for deep-sea eukaryotes and prokaryotes in the metagenomic libraries (Figure 14). In contrast to GOS surface water, deep-sea eukaryotic protein coding genes are more involved in functions such as ‘defense mechanisms’, ‘intracellular trafficking, secretion, and vesicular transport’, ‘signal transduction mechanisms’, ‘replication, recombination and repair’ and ‘inorganic ion transport and metabolism’, which are useful for leading a defensive life style as discussed in the section on prokaryotes. For deep-sea eukaryotic communities relatively limited COG clusters were identified in the metatranscriptomic libraries compared to the metagenomic libraries. The transcriptional activity of protein-coding genes varied remarkably among the four sampling sites, and it was difficult to identify any obvious trends analogous to those identified in the metagenomic libraries

(Figure 15). Of the 23 functional categories identified in the metagenomic libraries, 17 categories in CT04, 8 in CT05, 15 in CT06 and 20 categories in CT12 were identified in their metatranscriptomic pool: better coverage compared to the prokaryotic metatranscriptomic libraries, possibly due to the larger number of sequencing reads obtained (Table 10). Among the four deep-sea metatranscriptomic libraries, site CT12 showed the highest percentage of genes related to the functional category 'inorganic ion transport and metabolism'. This category from CT12 was mainly enriched with clusters including 'Outer membrane receptor proteins, mostly Fe transport' COG0629, 'Predicted divalent heavy-metal cations transporter' COG0428, and 'Cation transport ATPase' COG0474. For the categories 'signal transduction mechanisms' and 'replication, recombination and repair', site CT12 also harbored a significantly higher percentage of genes than other three sites. 'Diadenosine tetraphosphatase and related serine/threonine protein phosphatases' COG0639 and 'Universal stress protein UspA and related nucleotide-binding proteins' COG0589 are two main components included in this category. Site CT05 showed discernibly lower metabolic strength in most eukaryotic phyla, while CT06, very close and at similar depth to CT05, showed the highest metabolic strength in almost all eukaryotic phyla, similar to our findings in the deep-sea prokaryotic study.

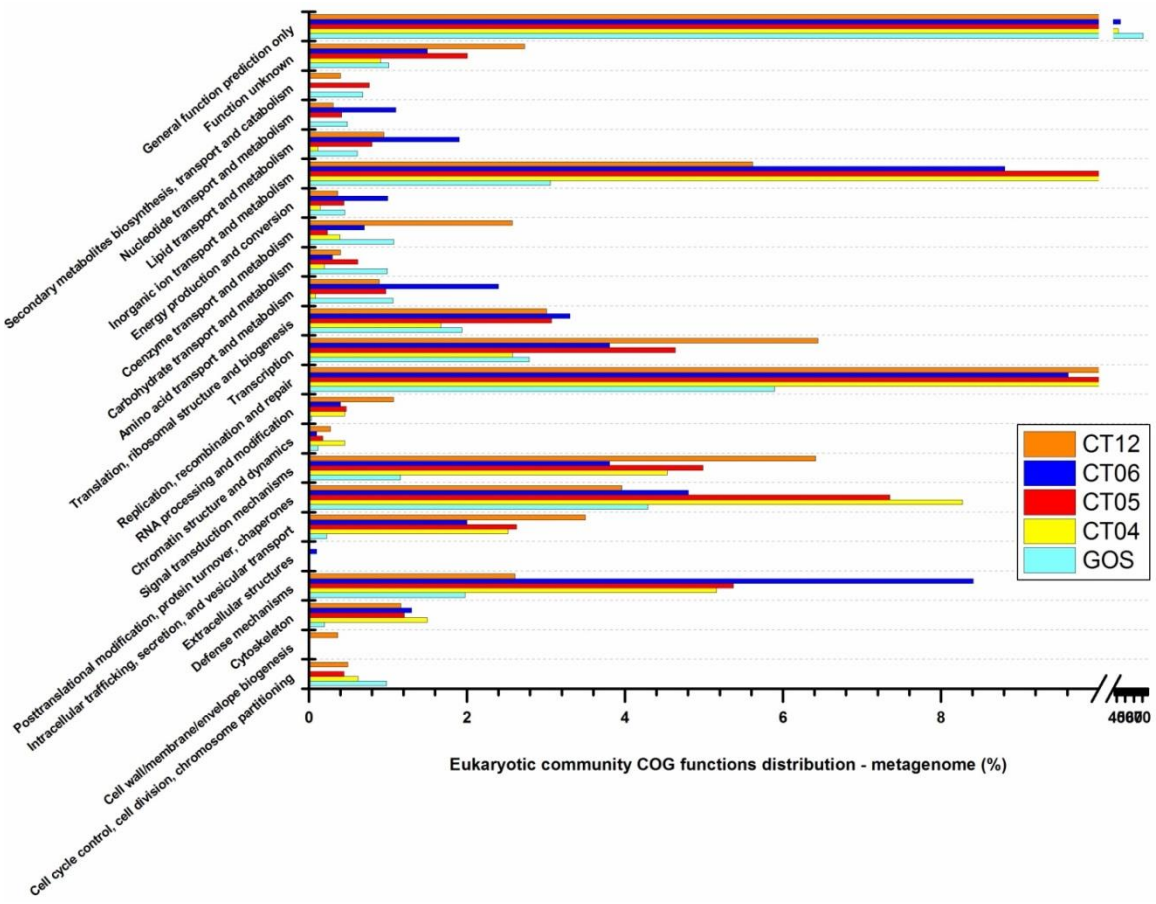

Figure 14. COG distributions of eukaryotic metagenomic libraries.

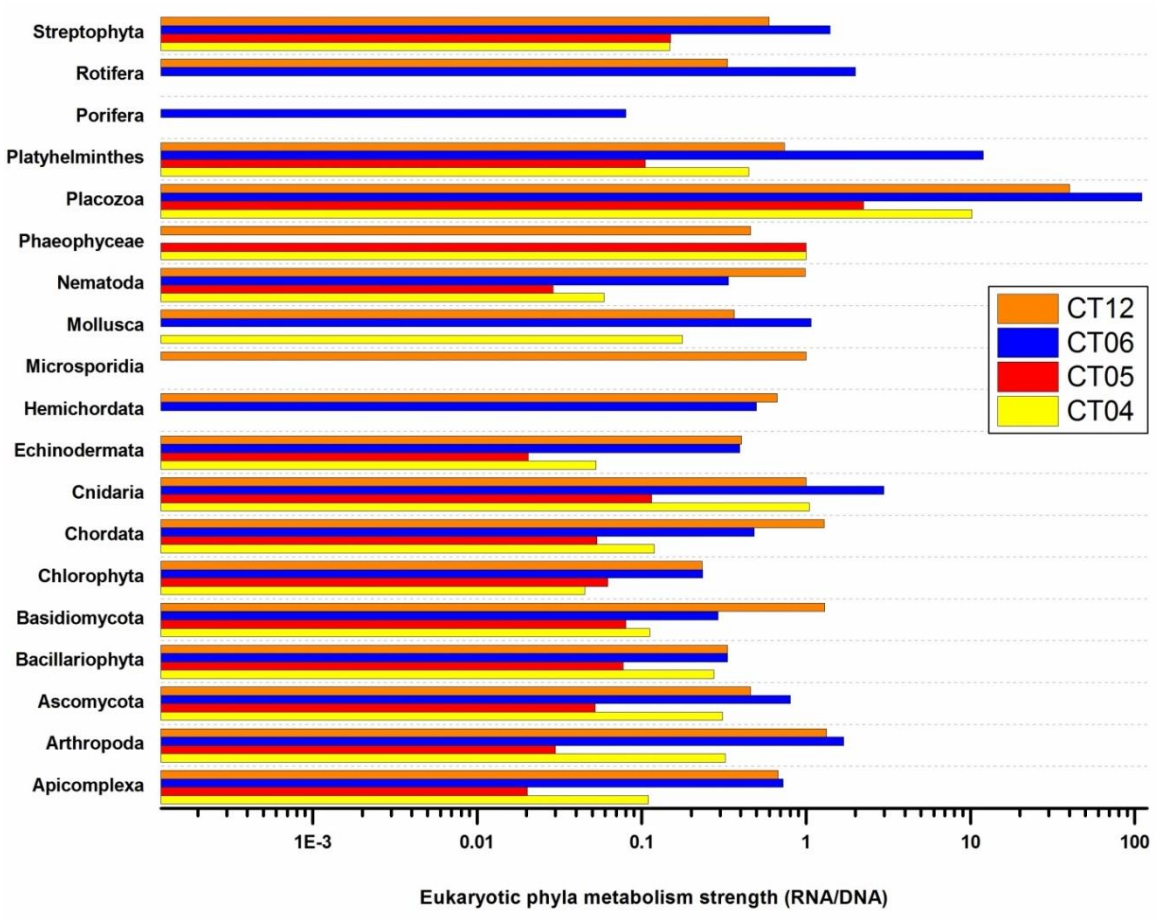

Figure 15. Metabolic activity strength comparison using RNA/DNA ratios of eukaryotic communities in deep sea.

Table 10. COG clusters identified in metatranscriptomic libraries of deep-sea prokaryotic communities.

Level 1	Level 2	CT04	CT05	CT06	CT12
Cellular processes and signaling	Cell cycle control, cell division, chromosome partitioning	0	0	0	0.12
	Cell wall/membrane/envelope biogenesis	60.87	0	19.46	40.53
	Defense mechanisms	0	0	0	0.12
	Intracellular trafficking, secretion, and vesicular transport	0	0	0	0.23
	Posttranslational modification, protein turnover, chaperones	12.17	0	4.70	1.51
	Signal transduction mechanisms	0	0	0	1.51
Information storage and processing	Replication, recombination and repair	21.74	0	27.52	26.13
	Transcription	0	33.33	14.09	3.83
	Translation, ribosomal structure and biogenesis	0	66.67	0	0.35
Metabolism	Amino acid transport and metabolism	0.87	0	4.70	0.58
	Carbohydrate transport and metabolism	0	0	0	0.46
	Energy production and conversion	0.87	0	7.38	0
	Inorganic ion transport and metabolism	0	0	9.06	12.89
	Lipid transport and metabolism	0	0	0	0.12
	Secondary metabolites biosynthesis, transport and catabolism	0	0	0	0.46
Poorly characterized	Function unknown	3.48	0	9.73	3.95
	General function prediction only	0	0	3.36	7.20

The RNA/DNA ratio was again employed as an index to describe the metabolic strength of deep-sea eukaryotes. Placozoa, considered the simplest organized metazoan model system (Schierwater et al. 2009; Eitel et al. 2010), showed the highest SMA value in all four deep-sea sampling sites (Figure 16). The highly active metabolism of *Placozoa* in the cold, dark, deep sea is quite surprising, because it has been suggested that the growth rate and vegetative reproduction of *Placozoa* may be positively correlated to increasing temperature (Maruyama 2004). Deep-sea autotrophic ecosystems, such as hydrothermal vents or cold seeps, are normally not considered conducive living environments for fungi that are abundant in terrestrial ecosystems because of their ability to degrade organic matter (Le Calvez et al. 2009; Bhadury et al. 2011). However, in our study the fungi *Basidiomycota* and *Microsporidia* showed the highest SMA value in site CT12, suggesting their affinity for the hydrothermal vent-influenced habitat. These results are consistent with the unexpected diversity of fungal species, such as *Basidiomycota*, detected in hydrothermal areas in previous studies (Takishita et al. 2007; Le Calvez et al. 2009).

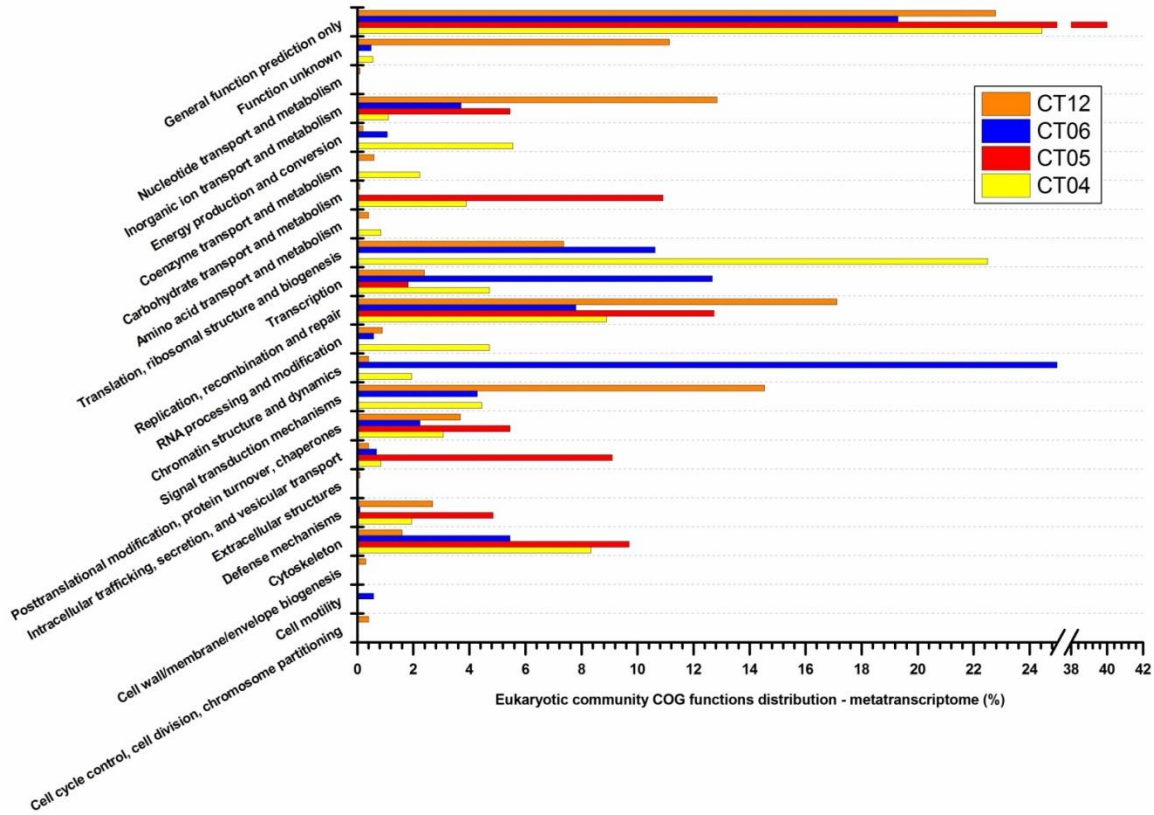

Figure 16. COG distributions of eukaryotic metatranscriptomic libraries.

Conclusions

Microbial communities in seawater at four deep-sea sites were isolated and amplified to perform taxonomic and functional analyses by using an integrated metagenomic and metatranscriptomic approach. The results showed that within the prokaryote community bacteria is absolutely dominant over archaea (~90%) in both metagenomic and metatranscriptomic pools in the deep-sea prokaryotic samples. When compared with the microbial communities of the GOS surface water, the proportion of archaea in the prokaryotic community was increased in the deep-sea water. The emergence of archaeal phyla *Crenarchaeota*, *Euryarchaeota* and *Thaumarchaeota*, bacterial phyla *Actinobacteria*, *Firmicutes*, sub-phyla *Betaproteobacteria*, *Deltaproteobacteria*, and *Gammaproteobacteria*, and the decrease of bacterial phyla

Bacteroidetes and *Alphaproteobacteria* were the main differences observed in the prokaryotic community compositions present in the deep-sea water. *Cyanobacteria* were identified in samples from all four deep-sea sites by metagenomic and metatranscriptomic analysis, suggesting their active functionality in deep-sea environments in spite of very little sunlight. Employing the RNA/DNA ratio as a metric indicative of the metabolic strength of microbes, we found that the metabolic strength of single phylum/sub-phylum varied remarkably across deep-sea sites, suggesting that the prokaryotic communities are experiencing distinctly different metabolic conditions at the different sites. In contrast to the GOS surface water communities, functional groups related to cell wall/membrane and capsule forming for high pressure resistance; signal sensing and transduction for adapting to the low-nutrient deep-sea environment; multidrug efflux for intrinsic and acquired resistance to antimicrobials; and defense mechanisms for self-protection were enhanced in the deep-sea water, indicative of a defensive life style rather than an active growing/metabolic style on the part of the prokaryotic community living in the deep sea. Taxonomic and functional analysis of the CT12 site, located 20 km away from the Juan de Fuca hydrothermal vent, harbored higher diversity than other deep-sea sites. In addition, decreases in abundance of fungi and algae in the deep sea were detected in our eukaryote study. Similar to prokaryotes, COG distribution analysis revealed that eukaryotes adapted a more defensive life style in the harsh deep-sea environments. This study provides the first integrated genomic and transcriptomic view of the microbial communities in deep-sea water of the North Pacific Ocean.

CHAPTER 5
PHENOTYPIC AND OMIC CHARACTERIZATION OF PSEUDOALTEROMONAS STRAINS ISOLATED FROM THE DEEP NORTH PACIFIC OCEAN

Introduction

The deep-sea water body, underlying the photic zone in the world's oceans, comprises 1.3×10^{18} m^3 of water and is the largest aqueous habitat for microbial life (Orcutt et al. 2011). It differs distinctly from the photic zone, exhibiting low temperature (approximately 2~4^{o}C), high pressure, and high inorganic nutrient levels (Aristegui et al. 2009). Compared to deep-sea sediment hosting rich microbial communities (Schippers et al. 2005; Jørgensen & Boetius 2007), deep-sea water which serves as a filter between the ocean bottom and the sunlit surface water, shows different characteristics. The pressure there is lower than in sediment, while the temperature is higher as ocean sediment is close to freezing point. Bacteria in deep-sea water are not primary producers (Orcutt et al. 2011), as they utilize sinking organic matter produced by phototroph in the ocean-photic zone via chemical redox reactions as a primary metabolic strategy. A huge amount of aged and buried sedimentary organic materials exist, although it remains unclear whether this serves as the main carbon source and is directly utilized by seafloor bacteria (Jørgensen & Boetius 2007). Many marine bacteria, especially those living in the surface water (Kolber et al. 2000; Morris et al. 2002; Thomas et al. 2008; Frias-Lopez et al. 2008; Biers et al. 2009) or in sediment (Schippers et al. 2005; Do et al. 1990; Braker et al. 2000; Toffin et al. 2004; Liao et al. 2011), have been subjected to extensive investigation. However, only limited studies on psychrotrophic, free-floating, deep-sea bacteria (Médigue et al. 2005a) have been conducted so far.

Pseudoalteromonas, belonging to one of the largest genera within the class gammaproteobacteria (Matsuyama et al. 2012) are Gram negative bacteria commonly found in the ocean environment from surface water to deep-sea sediment (Qin et al. 2011). *Pseudoalteromonas* are also found in association with marine eukaryotes and possess anti-bacterial, bacteriolytic, agarolytic and algicidal characteristics (Holmström & Kjelleberg 1999), and release many biologically active agents including extracellular enzymes (Cui et al. 2008; Zhou et al. 2009), toxins, and polysaccharides (Holmström & Kjelleberg 1999), indicating their role in important ecological functions in the marine environment. Production of anti-fouling agents from one strain of *Pseudoalteromonas* (Holmström et al. 1998) also demonstrates a potential commercial use of these bacteria. To date, the surface ocean strain *Pseudoalteromonas haloplanktis* TAC125 and marine sediment strain *Pseudoalteromonas* sp. SM9913 have been studied in depth through genome mapping and comparative genomics (Médigue et al. 2005a; Qin et al. 2011). However, to enrich and complete a full picture of the life of marine *Pseudoalteromonas*, more strains, especially those representatives from other geological environments are absolutely needed.

RNA-seq, the transcriptome revealed by next generation sequencing (NGS) of cDNA (Nagalakshmi et al. 2008), provides us with a highly sensitive novel approach to be utilized for characterizing the whole transcriptome of an organism (Pinto et al. 2011). The employment of this transcriptomic approach has proven efficacious in deciphering the metabolic potential and pathways of microorganisms. Normally, the prerequisite for transcriptomic analysis of an organism requires the availability of the whole genome of this organism for the purpose of RNA sequence alignment and annotation. While

genome-wide comparative studies based on NGS platforms are widely accepted to determine inter-species differences (Thomas et al. 2008; Moran et al. 2004; Hou et al. 2004; Vezzi et al. 2005; Nakagawa et al. 2007; Hijum et al. 2008; Hu et al. 2011; Eloe et al. 2011; Park et al. 2012), whole genome sequencing remains a pricey and time consuming technique. Hence it is difficult to obtain whole genome sequences for every single bacterial strain of interest. As proximity, it becomes an apparent choice to utilize the genomes most phylogenically related for RNA sequence alignment and annotation.

In this study, we initially phenotypically characterized properties of nine deep-sea *Pseudoalteromonas* isolates from the Pacific Ocean. One of them, WG07, was chosen for further in-depth transcriptomic study by application of RNA-Seq technology. The RNA sequence alignment and annotation of WG07 was realized through the completed genomes of two closed related strains of WG07, including the marine surface strain *Pseudoalteromonas haloplanktis* TAC125 (TAC125) and the marine sediment strain *Pseudoalteromonas* sp. SM9913 (SM9913). The transcriptome sequencing and related functional analysis of WG07 reveal unique features of this deep-sea isolate and provide clues as to how it adapts to its dwelling environment. Further comparative transcriptomic analysis of WG07 provides us with novel information on transcriptome changes between exponential and stationary growing phases, as well as between its grown in 4^{o}C and 25^{o}C . A comparative proteomic analysis between WG07 cells grown at both 4^{o}C and 25^{o}C has been conducted and the resultant proteomic results didn't concord the trends revealed by transcriptomic analysis. The possible reason and underlying mechanism causing these differences between transcriptomic and proteomic analysis have been discussed in the end of this chapter.

Materials and methods

***Microbial sample collection and phylogeny affiliation determination*.** All strains were isolated from deep-sea water samples collected at depths ranging from 1,504 ~ 2,320 m in the Northeast Pacific Ocean during the Thompson TN 221 Research Cruise from July to August of 2008. Samples were supplemented with 15% (V/V) glycerol and then stored in -80ºC before being shipped back to the lab on dry ice. Upon receipt, samples were stored at -80ºC again until use. For colony formation, samples were thawed at room temperature and 0.1 mL of seawater was homogenized and spread on LB agar plates containing 3% NaCl. After dark incubation for approximately one week at room temperature, colonies were observed and subcultured until pure cultures were obtained. For phylogeny affiliation determination, each isolate was transferred to 0.04 M NaOH (50 µl) and heated to 95ºC for 10 min. A 1-µL aliquot of lysed cells was used for the PCR template. The 16S rRNA genes were amplified from isolates with the PCR primers FD1 (5_-AGA GTTTGATCCTGGCTCAG-3_) and 1540R (5_-AAGGAGGTGATCCAGCC-3_). PCR conditions were as follows: 95ºC for 5 min, 35 cycles of 95ºC for 15 sec, 56ºC for 30 sec and 72ºC for 1 min, followed by 72ºC for 7 min. PCR products were purified with the QIAquick PCR Purification Kit (Cat No. 78104) according to manufaturer's instructions. DNA sequences were determined with a BigDye Terminator v3.1 Cycle Sequencing 175 Kit (Applied Biosystems) using a 3700 DNA Analyzer (Perkin-Elmer) according to the manufacturer's instructions using the primer 529R (5_-CGC GGC TGC TGG CAC-3_). Sequences were compared against those in GenBank using BLAST 2.2.10.

Phenotypical characterization of bacterial isolates.

Gram staining. Fifty microliters of culture grown at room temperature for 24-48 hours in marine broth 2216 (Becton, Dickinson and Company, Sparks, MD) was dried and heat fixed on a glass slide (VWR, Radnor, Pennsylvania) before gram staining. Gram staining was performed with gram crystal violet, gram iodine, gram decolorizer, and then counterstained with gram safranin (Becton, Dickinson and Company, Franklin Lakes, NJ) according to manufacturer's directions. Slides were air dried and examined with a light microscope.

Determination of pH range of growth. Marine broth 2216 (Becton, Dickinson and Company, Sparks, MD) was adjusted to the appropriate pH with HCl or NaOH and then inoculated with a single colony. Presence or absence of growth was determined following the four-day incubation at room temperature.

DNA extraction. DNA was extracted with the DNeasy blood and tissue kit (QIAGEN, Hilden Germany) without addition of RNase A. Extraction of nucleic acids was performed using the spin column protocol for animal tissue following a pretreatment for Gram-negative bacteria consisting of resuspending the bacterial pellets with buffer ATL as per manufacturer's directions. DNA was eluted with DEPC-treated water (Ambion, Austin, TX) and stored at -20°C until further analysis.

G+C mol%. The DNA G+C mol% was determined by thermal denaturation temperature according to Gonzalez & Saiz-Jimenez (2002) on a Step One PCR instrument (Applied Biosystems, Foster City, CA). *Pseudoalteromonas tetradonis* IAM 14160, *Pseudoalteromonas elyakovii* KMM 162, and Escherichia coli K12 standards (ATCC, Boulevard Manassas, VA) with known G+C mol% of 41.5%, 38.5%, and 50.8%

respectively (Blattner et al. 1997; Sawabe et al. 2000; Ivanova et al. 2001) were used to determine the optimal concentration of SSC (Invitrogen, Carlsbad, CA). The default melting curve analysis program was used to generate the melting curve and calculate the Tm for each isolate. Melting curve analysis for each isolate was performed in triplicate in 10 µL volume reactions. A standard curve of known G+C mol% versus Tm was generated thereafter. A final concentration of 0.1× SSC, 1× SYBR Green I (Invitrogen, Carlsbad, CA), and 34 ng to 1.4 µg of DNA per reaction yielded G+C mol % comparable to known values. *Pseudoalteromonas Tetradonis* IAM 14160, *Pseudoalteromonas elyakovii* KMM 162, and Escherichia coli K12 standards G+C mol% was experimentally determined to be 42.2 ± 0.26, 38.6 ± 0.56, and 50.31 ± 0.87, respectively. The slope equation generated from the standards was used to determine G+C mol% of unknown isolates after performing triplicate reactions using the same conditions outlined above.

Substrate utilization. Enzymatic activity was determined using API 20E, API 50CH, and API ZYM strips (Biomérieux Marcy I'Etoile, France) according to manufacturer's directions. Freshly cultured isolates were resuspended in sterile 0.85% NaCl before inoculation of the API 20E and API ZYM strips. Isolates for API 50 CH were resuspended in API 50 CHB/E medium prior to inoculation. API ZYM suspensions were prepared with turbidity of approximately 6 McFarland. All reactions were incubated at 37°C for 24-48 hours before examination. API ZYM strips were additionally exposed to light to render negative reactions colorless before interpretation of results.

Antibiotic susceptibility. Sulfamethoxazole, chloramphenicol, cefadroxil, trimethoprim, erythromycin, and tetracycline (Sigma-Aldrich, St. Louis, MO) were dissolved in 1 mL of acetone, ethanol, methanol, 50:50 mixtures of chloroform and

isopropanol, and ethanol respectively. Antibiotics were added to cooled LB agar (Difco Franklin Lakes, NJ) at final concentrations of 25 µg/mL sulfamethoxazole, 39 µg/mL chloramphenicol, 21 µg/mL cefadroxil, 30 µg/mL trimethoprim, 20 µg/mL erythromycin, 6.5 µg/mL and tetracycline, respectively. Isolates were subcultured onto antibiotic plates and presence or absence of growth was determined after 24-48 hour incubation at room temperature. Ampicillin, kanamycin, bacitracin, and vancomycin were dissolved in 1 mL of sterile distilled water to a final concentration of 50 µg/mL, 25 µg/mL, 30µg/mL, and 38 µg/mL, respectively. Disk diffusion antibiotic sensitivity testing was performed using the Kirby-Bauer method (Bauer et al. 1966). Briefly, 200 µL of freshly inoculated LB broth was spread uniformLy onto a LB agar plate and allowed to dry. Antibiotic disks were placed with sterile forceps and presence or absence of growth in the area surrounding the disks was determined after incubation for 24-48 hours at room temperature.

Bacteria growth assay and RNA isolation. The marine broth was made by dissolving 37.4g DifcoTM marine broth (Becton, Dickinson and Company, Sparks, MD) into 1 L distilled water. Cultivation of WG07 was performed in dark at room temperature (25^{o}C) in three 125 mL flasks with 50 mL marine broth. The OD600 was measured at various times during growth, using the average value of 5 reads as the final read. Microbial samples were collected at time points 24 hours (exponential growth phase, OD600 $\approx$ 0.42) and 168 hours (stationary growth phase, OD600 $\approx$ 0.92). Sampled bacteria were directly centrifuged and total RNA isolation was obtained with the ZR Fungal/Bacterial RNA MicroPrep™ (Zymo Research, Irvine, CA) following the manufacturer's instructions. mRNA was then further enriched by using a RiboMinus™

Transcriptome Isolation Kit (Bacteria) (Invitrogen, Grand Island, NY) following the manufacturer's instruction. BioAnalyzer (serial number: DE20901519) (Agilant, Santa Clara, CA) was employed to check the integrity and purity of mRNA obtained to ensure the good performance of further RNA whole-transcriptome sequencing.

Transcriptome sequencing and sequence alignment and annotation. RNA-seq was performed by using Ion Torrent RNA whole-transcriptome sequencing. The whole sequencing process included eight steps: product fragmentation, library preparation, binding single strand fragments to their own unique beads, emulsion PCR, clonally amplified bead purification, loading beads onto a plate, sequencing, and final pyrosequencing (signal processing of images). The De Novo transcriptome assembler Trinity was then employed to assemble the sequences. Gene expression profiling provided by GeneSifter® analysis edition (Geospiza, Seattle, WA) was utilized for next generation sequencing data interpretation and pairwise comparative analysis; and the MG-RAST metagenomics analysis server provided by Argonne National Laboratory (Meyer et al. 2008) was employed for functional subsystems categorization. For the pairwise comparative analysis, the parameters are as follows: normalization is mapped (NGS), statistics is performed with the likelihood ratio test, quality is 10, the quality bar is that one group must pass, minimum fold change is 1.5, correction is Benjamini and Hochberg correction, and no data transformation is allowed. In MG-RAST analysis, the maximum e-value cutoff is 1e-5, minimum percent of identity cutoff is 60%, and minimum alignment length cutoff is 15.

Sequence data of *Pseudoalteromonas haloplanktis T*AC125 (TAC125) and *Pseudoalteromonas* sp. SM9913 were obtained from the National Center for Biotechnology Information (NCBI) databank.

Electron microscopy (EM) imaging. A JEOL JSM-6300 SEM (North Billerica, MA) fitted with an IXRF digital imaging system (Austin, TX) and Philips CM12S Scanning Transmission EM with EDAX 9800 (Mahwah, NJ) were employed to perform imaging analysis. For samples of both systems, cells were fixed by the following procedures: cells were pelleted at 3,000×g for 5 min, re-suspended in 2.5% glutaraldehyde and incubated in 0.1M NaPO4 solution (adjusted pH 7 with 2% NaCl (PBS)) for 1 hour at room temperature then pelleted again at 3,000×g for 1 min and washed 2 times in suspension with PBS. For TEM, the pellet was encased in 0.8% agarose, washed three times more in PBS, and incubated in 1% osmium tetroxide in PBS for 1 hour at room temperature. After washing 3-4 times in distilled H_2O, the pellet was then incubated in 0.5% aqueous uranyl acetate for 1 hour at room temperature, followed by washing another 3-4 times. Dehydration was performed three times (7-8 min/time) in graded series of acetone with 20% increments with 100% acetone. The product was infiltrated three times in graded series of epoxy resin (Spurr's formula) with 25% increments with 100% resin, followed by polymerization in BEEM capsules at 60ºC for 16 hours. The capsules were then cut into 60 nm sections with Leica Ultracut-R microtome and post-stained with uranuyl acetate and Sato's lead citrate. The observations were performed in Philips CM12 TEM operated at 80 kV, and images were acquired with Gatan model 791 CCD camera (Pleasanton, CA). For SEM, the pellet was fixed in suspension with 1% osmium tetroxide for 1 hour at room temperature, and then washed

2-3 times in suspension with purified water. The cells were then adhered to poly-L-lysine coated coverslips and dehydrated with a series of acetone as illustrated in the TEM section. The product was dried at the critical point with CO_2 (Balzers CPD-020 unit), a coverglass was mounted on the SEM stubs, and coated with approximate 15 nm gold-palladium (Technics Hummer-II sputter-coater). Observations were performed in JEOL JSM-6300 SEM operated at 15 kV. The images were acquired with a IXRF Systems digital processing unit. The EM imaging service was provided by The School of Life Sciences Electron Microscopy Lab at Arizona State University.

Protein isolation and proteomic analysis. Bacteria used for protein isolation is sampled together with bacteria used for RNA isolation. Minute™ Bacterial Total Protein Extraction Kits (InventBiotech, Eden Prairie, MN) were used to extract denatured proteins from WG07 following the manufacturer's manual. LC-MS/MS using solution protein digestion that can provide absolute quantitation was used to obtain protein expression information. Micro Lowry Peterson's Modification Total Protein Kit (Sigma, St. Louis, MO) was used for protein quantitation. *Pseudoalteromonas* sp. proteins were downloaded from Uniprot_Sprot, and T33V Rhodobacter cytochrome c standard primary sequence was employed as a standard to perform the quantitation. The proteome assay service was provided by Arizona Proteomics Consortium, The University of Arizona.

Results and Discussion

***Phenotypic characterization of isolates*.** Table 11 summarizes the assayed phenotypic properties of all isolates. A phylogenetic tree of the marine isolates and TAC125 and SM9913 is shown in Figure 17. All strains were very similar. WG07, isolated from the deepest depths (2,300 m), was selected as the target to perform the

transcriptomic analysis since it is neither too far away nor too close to the references to provide an adequate comparison. Note that WG07 total DNA G+C mol% determined by thermal denaturation was 40.31 ± 0.99 which is similar to SM9913 and TAC125 chromosome I 40.4 and 41% and chromosome II 39.8 and 39.3%, respectively (Médigue et al. 2005a; Qin et al. 2011). Experimental results revealed that WG07, SM9913, and TAC125 were able to utilize gelatinase and arginine dihydrolase but only WG07 was capable of using β-glucosidase which is confirmed in the literature (Qin et al. 2011). WG07 was unable to use trypsin and alpha-glucosidase while SM9913 and TAC125 utilized these substrates (Qin et al. 2011). Experimentally, WG07 did not use glycogen but instead, may have been accumulating it as an energy reserve if other food sources were depleted. According to the Qin study in 2007, WG07 utilized N-acetyl-glucosamine and maltose like SM9913 while in comparison did not utilize gentiobiose, D-trehalose, and D adonitol. WG07 did not utilize L-rhamnose, D- turanose, D-sorbitol, inositol, D-arabitol, and L-arabinose while Qin in 2007 reported that SM9913 utilization was also not determined for L-rhamnose, turanose, sorbitol, meso-inositol, and adonitol but was determined for D-arabitol and L-arabinose. D-mannose was used by both WG07 and SM9913. Unlike SM9913, WG07 was not able to use the sugar D-lactose or the alcohol glycerol. Our experimental findings also showed that WG07 utilized the amino acids L-arginine and L-lysine but did not use L-ornithine and L-tryptophane.

Table 11. Phenotypic characteristic of marine *Pseudoalteromonas* isolates

	WG02	WG06	WG07	WG12	WG13	WG16	WG25	WG28	WG32
G/C mol (%)	41.34± 0.76	41.96± 0.60	40.31± 0.99	40.32± 0.62					
Antibody resistance									
Tetracycline	--/--	+/+	+/+	+/+	--/--	+/+	--/--	+/+	--/--
Sulfmethoxalate	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
Chloraphenical	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--
Cefadroxl	--/--	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
Trimehoprim	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
Erthomycin	--/--	+/+	+/+	+/+	--/--	+/+	--/--	+/+	--/--
Ampicillin	--/--	--/--	--/--	+/+	--/--	+/+	--/--	--/--	--/--
Kanamycin	--/--	+/+	--/--	+/+	--/--	+/+	--/--	--/--	--/--
Vancomycin	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--
Bacitracin	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--	--/--

Tetracycline	--/--	+/+	+/+	+/+	--/--	+/+	--/--	+/+	--/--
Substrate utilization									
2-nitrophenyl-betaD-galactopyranoside	+	+	+	+	-	+	-	+	-
L-arginine	+	+	+	+	-	+	-	+	+
L-lysine	-	+	+	+	-	+	-	+	-
L-ornithine	-	-	-	-	-	-	-	-	-
trisodium citrate	+	+	+	+	-	+	-	+	+
sodium thiosulfate	-	-	-	-	-	-	-	-	-
urea	-	-	-	-	-	-	-	-	-
L-tryptophane	+	-	-	-	-	+	-	+	+
L-trytophane	-	-	-	-	-	-	-	-	-
sodium pyruvate	-	-	-	-	-	-	-	-	-
Gelative	+	+	+	+	+	+	-	+	+
D-glucose	+	+	+	+	-	+	-	+	+

d-mannitol	-	-	-	-	-	-	-	-	-
inositol	-	-	-	-	-	-	-	-	-
D-sorbitol	-	-	-	-	-	-	-	-	-
L-rhamnose	-	-	-	-	+	-	-	-	-
D-sucrose	-	-	-	-	-	-	-	-	-
D-melibiose	-	-	-	-	-	-	-	-	-
amygdalin	-	-	-	-	-	-	-	-	-
L-arabinose	-	-	-	-	-	-	-	-	-

WG07 was susceptible to kanamycin (25µg/mL) and ampicillin (50µg/mL) but grew in the presence of low concentration (6.5 µg/mL) of tetracycline whereas SM9913 and TAC125 did not grow in the presence of higher concentration (30µg per disc) of tetracycline (Qin et al. 2011). WG07 also grew in the presence of sulfamethoxazole (25 µg/mL), cefadroxil (21 µg/mL), trimethoprim (30 µg/mL), and erythromycin (20 µg/mL) but was susceptible to chloramphenicol (39 µg/mL), vancomycin (38 µg/mL), and bacitracin (30 µg/mL).

Scanning electron micrographs of WG07 cells revealed the rod shape of the cells (Figure 17). Nearly spherical cells were also observed, which is highly possible due to

increased stress (Santos et al. 2005). Lengths of the cells ranged from 0.8 μm to 2.5 μm, while the diameter was around 0.6 μm. The transmission electron micrograph clearly shows the cell membrane and inner structure of the cell (Figure 17).

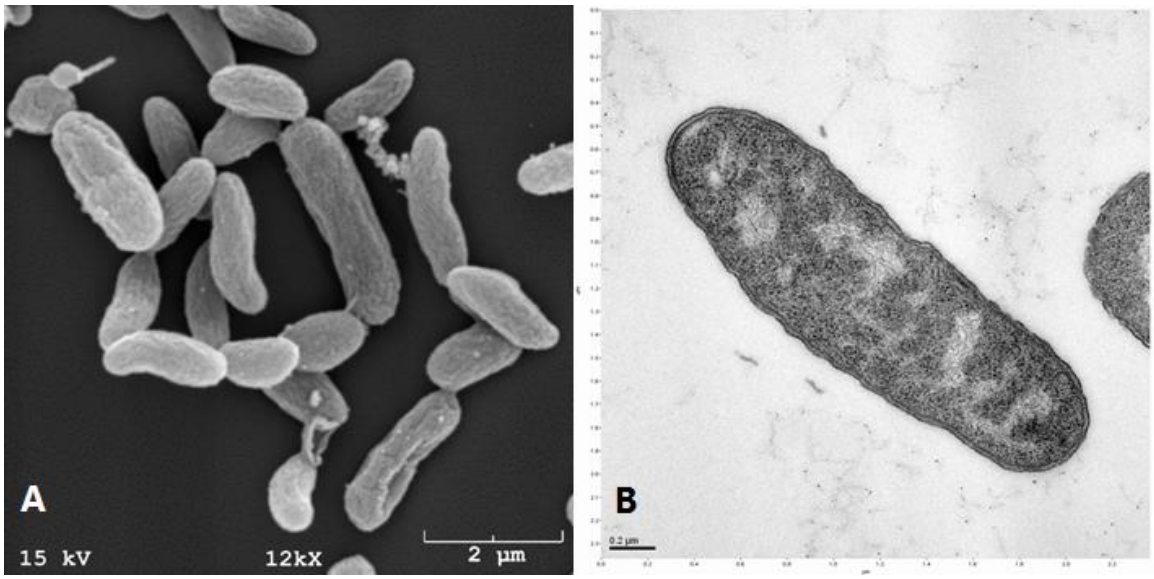

Figure 17. Electron Microscopy images of the *Pseudoalteromonas* isolate WG07. A) Scanning Electron Microscopy; B) Transmission Electron Microscopy.

General information of WG07's transcriptome. Like many other marine γ-*Proteobacteria*, the genome of *Pseudoalteromonas* isolate WG07 consists of two chromosomes. In this study, partial 16S rRNA of WG07 was cloned and sequenced. WG07 is highly similar to the surface-water strain *Pseudoalteromonas haloplanktis* TAC125 and ocean-sediment strain *Pseudoalteromonas* sp. SM9913, with 16S rRNA gene maximum identity of both alignments of 99%. Here we use symbols $WG07_{TAC125}$ and $WG07_{SM9913}$ to represent WG07's transcriptome using TAC125 and SM9913 as the reference, respectively. WG07 was found to be relatively phylogenetically closer to SM9913 (Figure 18). Within the mapped reads, 3,408 protein-coding genes were found in

$WG07_{SM9913}$, while 3,288 were found in $WG07_{TAC125}$. As a reference, there are 3,712 and 3,485 protein coding genes in SM9913 and TAC125, respectively. A total of 2,747 out of 2,847 COG clusters of SM9913 were found in $WG07_{SM9913}$, while 2,650 out of 2,726 COG clusters in TAC125 were found in $WG07_{TAC125}$. In Table 12, it is clear that in COG clusters 'Amino acid transport and metabolism', 'Carbohydrate transport and metabolism', 'Cell motility', 'Cell wall/membrane/envelope biogenesis', 'Defense mechanisms', 'Signal transduction mechanisms', and 'Translation, ribosomal structure and biogenesis', WG07 and SM9913's corresponding genes are more than those of TAC125; while in COG clusters 'Cell cycle control, mitosis and meiosis', 'Coenzyme transport and metabolism', 'Energy production and conversion', 'Intracellular trafficking, secretion, and vesicular transport', 'Lipid transport and metabolism', 'Posttranslational modification, protein turnover, chaperones', and 'Secondary metabolites biosynthesis, transport and catabolism', WG07 and TAC125's related genes are more than those of SM9913. This part will be discussed in detail below.

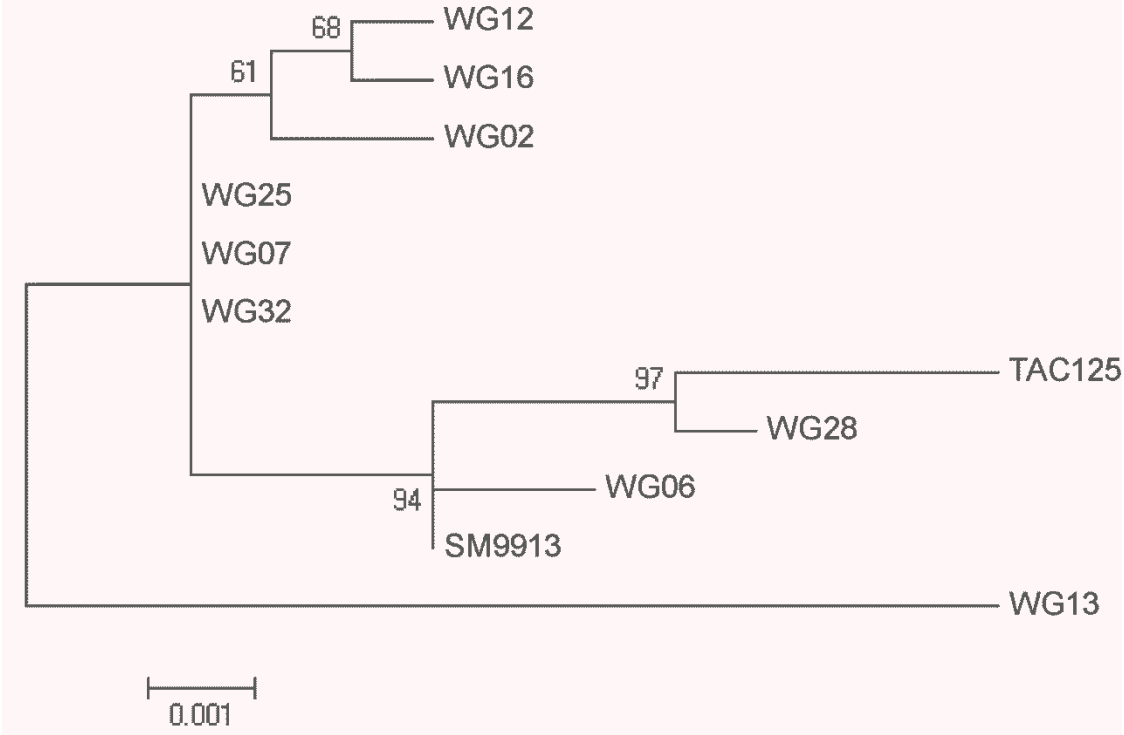

Figure 18. Phylogenetic tree from neighbor-joining method based on the 16S rRNA gene sequences of deep-sea Pseudoalteromonas isolates, SM9913 and TAC125.

Table 12. COG matches of WG07 in SM9913 and TAC125.

COG clusters	SM9913	$WG07_{SM9913}$	TAC125	$WG07_{TAC125}$
Amino acid transport and metabolism	218	212	211	207
Carbohydrate transport and metabolism	100	99	95	90
Cell cycle control, mitosis and meiosis	32	32	36	35
Cell motility	61	59	23	22
Cell wall/membrane/envelope biogenesis	188	180	157	155
Coenzyme transport and metabolism	103	102	108	107
Defense mechanisms	53	51	42	42
Energy production and conversion	151	149	162	159
Function unknown	284	271	269	248
General function prediction only	321	305	323	319
Inorganic ion transport and metabolism	165	161	162	159
Intracellular trafficking, secretion, and vesicular transport	95	93	113	108
Lipid transport and metabolism	74	73	98	98
Nucleotide transport and metabolism	58	58	59	57
Posttranslational modification, protein turnover, chaperones	131	126	145	141

Replication, recombination and repair	163	147	148	144
RNA processing and modification	1	1	1	1
Secondary metabolites biosynthesis, transport and catabolism	41	40	46	44
Signal transduction mechanisms	242	235	164	159
Transcription	175	164	178	170
Translation, ribosomal structure and biogenesis	192	189	186	185

Metabolic features. The growth curve shows that strain WG07 is a fast growing strain, suggesting that it may live in a relatively nutrient-rich environment. In the survey of WG07's transcriptome, genes involved in glycolysis/gluconeogenesis, citric acid cycle (TCA cycle), and the pentose phosphate metabolism pathways, including 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, 6-phosphogluconolactonase, enolase, glucokinase, glucose-6-phosphate 1-dehydrogenase, NAD-dependent glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, pyruvate kinase, and 6-phosphogluconate dehydratase, were identified, implying that WG07 utilizes glucose to obtain energy, similar to SM9913 (Qin et al. 2011). WG07 possesses a phosphoenolpyruvate-dependent phosphotransferase system (expressed gene PtsP identified) for the transportation and initiation of carbohydrate degradation. This phosphotransferase system phosphorylates glucose by glucokinase PSHAa1364. Experimental results also confirmed that WG07 can metabolize glucose (Table 11). On the other hand, TAC125 was unable to grow in the presence of glucose unless tyrosine

was supplemented to the medium (Médigue et al. 2005). No genes related to sucrose were found expressed in WG07's transcriptome. It is consistent with the experimental result that WG07 cannot metabolize sucrose (Table 11). In comparison, SM9913 was able to utilize sucrose as substrate, while neither TAC125 nor WGO7 could. Additionally, mannose can be utilized by TAC125 and WG07 but not by SM9913. As a proof, the transcriptome of WG07 harbored 4 mannose metabolism related genes (Mannose-1-phosphate guanylyltransferase EC 2.7.7.22, Mannose-6-phosphate isomerase EC 5.3.1.8, Phosphomannomutase EC 5.4.2.8, and Mannosides-regulated TonB-dependent outer membrane receptor). It is highly possible that WG07 is able to utilize mannose as a carbon source. For nitrogen metabolism, no genes involved in nitrite or nitrate reduction were identified, suggesting that WG07 may not be able to utilize nitrite or nitrate, which is similar to SM9913.

The distribution of carbon, nitrogen, and phosphorous in the ocean is generally unbalanced (Moran et al. 2004). Bacterial glycogen production is considered to be a reservoir that can provide both energy and food (carbon) during starvation periods (Strange 1968). The glycogen production operon is made up of glucose-1-phosphate adenylyltransferase, glycogen synthase and glycogen branching enzyme and is present in chromosome II from PSM_B0508 to PSM_B0514 in SM9913. Glycogen production operon was also identified in WG07 (Table 13). The operon is conserved in *Alteromonadales* sp. TW-7 and *Pseudoalteromonas tunicate* but absent in TAC125 (Thomas et al. 2008). Several studies have illustrated that a large fraction of the annual nutrient load can arrive at deep-sea in pulses within a few days (Witte et al. 2003; Billett et al. 1983; Lampitt 1985). Therefore, deep-sea bacteria WG07 and SM9913 have to

accumulate glycogen using the glycogen production operon when the nutrients are present, and consume the stored glycogen when food is limited or absent.

Table 13. Glycogen production related genes using SM9913 as reference.

Gene id	Gene name	Info.	Description
10035097	PSM_B0508	chrII:555589:555891:302:3685	1,4-alpha-glucan branching enzyme
10035098	PSM_B0509	chrII:555901:557889:1988:3686	4-alpha-glucanotransferase
10035099	PSM_B0510	chrII:557890:560112:2222:3687	glycogen branching enzyme
10035100	PSM_B0511	chrII:560105:562168:2063:3688	glycogen operon protein
10035101	PSM_B0512	chrII:562161:564686:2525:3689	Phosphorylase
10035102	PSM_B0513	chrII:564707:565981:1274:3690	glucose-1-phosphate adenylyltransferase
10035103	PSM_B0514	chrII:566040:567572:1532:3691	Starch (glycogen) synthase

Expression of genes involved in signal transduction and transposable elements.

Signal transduction genes play a vital role in the adaptation of bacteria to deep-sea environments (Vezzi et al. 2005). For prokaryote, two-component system is a basic stimulus-response coupling mechanism to enable them to sense and respond to many different stimuli (Stock et al. 2000). Under normal conditions, two component system consists of a histidine kinase that senses the specific environmental stimulus and a

cognate response regulator that can regulate the cellular response through the control of target genes' expression (Mascher et al. 2006). The GGDEF domain-contained protein is involved in cyclic diguanosine monophosphate synthesis (RömLing et al. 2005). Based on the gene database of TAC125 and SM9913 obtained from NCBI, there are 20 histidine kinase related genes and 15 GGDEF/diguanylate cyclase related genes in TAC125's genome, while 33 histidine kinase related genes and 30 GGDEF/diguanylate cyclase related genes exist in SM9913's genome. All 20 histidine kinase related genes and 15 GGDEF/diguanylate cyclase related genes were identified in $WG07_{TAC125}$, while 33 histidine kinases related genes and 28 GGDEF/diguanylate cyclase related genes were identified in $WG07_{SM9913}$. Moreover, in WG07, two genes were found to be mutated compared to corresponding genes' sequences in TAC125 and SM9913. BarA in WG07, the membrane-located sensor kinase, an important component of the BarA/UvrY two-component system that is well conserved in species of the γ-*proteobacteria* (Binnenkade et al. 2011), has one insertion and one single nucleotide variation compared to TAC125 but is the same as that in SM9913. On the other hand, the transmembrane histidine kinase PhoQ, that can limit concentrations of extracellular Mg^{2+} through increasing the net phosphorylation of the PhoP transcriptional response regulator (Marina et al. 2001), was also found to be mutated by having two single nucleotide variants (SNV) when compared to SM9913 but maintained homology with TAC125.

Transposable elements transposases and integrases categorized into COG cluster 'Replication, recombination and repair', control the displacement of nucleic acid sequences from one place to another within and between genomes (Polard & Chandler 1995). In the comparative study of metagenomes collected from different depths

(Konstantinidis et al. 2009), it was reported that transposable elements are more enriched in genomes of bacteria living in deep-sea rather than in surface water, indicating that more gene-translocations happen in the deep-sea. As a proof, there are 28 transposases and 12 integrases in SM9913, while only 13 transposases and 7 integrases in TAC125. For WG07, 17 transposases and 9 integrases were all identified in the transcriptome of $WG07_{SM9913}$, while 12 transposases and 6 integrases in $WG07_{TAC125}$. When living in sub-optimal environments, bacteria can respond by increasing the frequency of mutants created by base substitution, frame-shift and transposition mutations (Higgins 1992). Under the high pressure of the deep-sea, mutation, which is also an adaptation of bacteria to the environment, is important to help bacteria survive. Remarkably larger groups of transposable elements serve as a promoter for these adaptive mutations in deep-sea bacteria.

Expression of genes related to reactive oxygen species (ROS). Protein oxidative damages usually disturb cellular processes and metabolism (Cabiscol et al. 2000; Shacter 2000). Microbes need to build up mechanisms to fight against oxidative stress with enzymes including catalase and superoxide dismutase, and small proteins such as thioredoxin (Cabiscol et al. 2000). As reported in a previous study about TAC125 and SM9913, the numbers of proteins involved in cold/salt adaptation in the two strains are almost the same (Qin et al. 2007; Médigue et al. 2005), indicating similar cold/salt adaptation strategies of these two strains. In TAC125's genome there are 12 dioxygenase genes (PSHAa0187, PSHAa0570, tesB, PSHAa0904, PSHAa2137, PSHAa2147, melA, PSHAa2449, PSHAb0029, PSHAb0041, PSHAb0115, and hmgA), while there are only 4 (PSM_A0972, PSM_A1718, PSM_A1738, and PSM_B0404) in SM9913's genome. In

WG07's transcriptome, we identified all 12 and 4 dioxygenase genes in $WG07_{TAC125}$ and $WG07_{SM9913}$, respectively, which indicates that surface water strain TAC125 and deep-sea strain WG07 fight against ROS in ocean better than marine sediment strain SM9913. Gene melA was found mutated in $WG07_{SM9913}$ with 2 deletions and 2 SNV. No genes were found mutated in $WG07_{TAC125}$. A previous study explored that the concentration of oxygen in marine sediment is even lower than that in deep-sea water at a depth of 1,800 m (Glud 2008). This is in accordance with the finding that sediment strain SM9913 has the smallest number of dioxygenase genes among these three isolates, which indicates that SM9913's dwelling environment might be with lower oxygen concentration. WG07's surrounding deep-sea water (~2,320 m depth) is highly possible with a relatively high oxygen concentration due to the increased solubility of oxygen at low temperature, since WG07 has a similar strategy with surface water strain TAC125 against ROS. The fatty acid metabolism gene cluster from PSHAa0894 to PSHAa0910 that is related to ROS in TAC125 was also detected in $WG07_{TAC125}$ while absent in SM9913. All these findings imply that WG07 is well protected against ROS at cold temperatures, and has similar ROS resistance as TAC125 but better than SM9913.

***Genes involved in defense mechanism*.** Genes encoding two AcrB/AcrD/AcrF multidrug resistance proteins (COG0841), ABC transporter permease (COG0842), two ABC transporter ATP-binding proteins (COG1131), inner membrane lipoproteins ABC transporter ATP-binding protein (COG1136), restriction endonuclease (COG1403), 5,10-methylenetetrahydrofolate reductase and beta-lactamase (COG1680), restriction endonuclease (COG1715), toxin secretion ATP-binding protein (COG2274), and two restriction endonucleases (COG3183) in SM9913 are absent in TAC125. These genes are

related to multidrug resistance (Nishino & Yamaguchi 2001; Nikaido 1996). On the other hand, genes encoding ABC transporter ATP-binding protein/permease (COG1132) and undecaprenyl-diphosphatase (COG1968) in TAC125 are absent in SM9913. There are more genes encoding multidrug resistance proteins in SM9913 than in TAC125, indicating a better antibiotic resistance of SM9913 than TAC125. In fact, this feature has been experimentally proven to show that SM9913 can survive with some antibiotics including ampicillin, penicillin, and amoxicillin, whereas TAC125 is susceptible to them (Qin et al. 2011). For strain WG07, as Table 2 shows, all 42 defense mechanism related gene clusters were identified in $WG07_{TAC125}$, while 51 out of 53 genes involved in the defense mechanism were identified in $WG07_{SM9913}$. The two absent genes in $WG07_{SM9913}$ are encoding multidrug resistance efflux pump (COG1566) and Mrr restriction system (COG1715). Among tested anbibodies (Table 11), WG07 is susceptible to Kanamycin and resistant to Tetracycline, the same as SM9913 and TAC125. But WG07 is susceptible to Ampicillin, which TAC125 is also susceptible to, but SM9913 is resistant to Ampicillin. For mutations in $WG07_{SM9913}$, 2 insertions and 2 SNV in the RND family multidrug efflux protein coding gene acrB, 1 SNV in inner membrane lipoproteins ABC transporter ATP-binding protein coding gene lolD, and 1 SNV and 2 insertions in beta-lactamase coding gene ampH, were identified. For mutations in $WG07_{TAC125}$, 1 insertion in the type I restriction-modification system M involving gene hsdS and 1 deletion in the peptide ABC transporter membrane protein coding gene sapB were identified.

On the basis of the transcriptomic survey, WG07 is expected to be resistant to a series of heavy metals including cadmium (CzcA, CzcB), cobalt (CzcA, CzcB, CorA, CorC), Zinc (CzcA, CzcB, Zur), magnesium (CorA, CorC, PSM_A2359, PSM_A0395),

copper (CopAB, CopC, PSM_A2644, PSM_A2645, PSM_A2647), and mercury (McrC, PSM_B0282). Here we use SM9913 as the reference which has a better heavy metal resistant potential than TAC125 (Médigue et al. 2005; Qin et al. 2011). These heavy metal resistance and efflux-involved genes make WG07 have a similar strategy against heavy metal with SM9913 but better than TAC125. Deep-sea water usually mixes with hydrothermal fluid and hence contains high levels of heavy metals (Hourdez & Weber 2005). This finding may be a possible explanation for why most deep-sea isolates are more resistant to heavy metals than surface bacteria (Qin et al. 2011).

***Genes involved in motility and flagella*.** Motility is an important mechanism for bacteria to acquire food, form biofilms, and avoid harmful chemicals and predators. Flagellar motility is one of the most pressure-sensitive cellular processes in mesophilic bacteria (Eloe et al. 2008). Usually, there are two gene clusters, lateral flagellum (LF) gene cluster and polar flagellum (PF) gene cluster, involved in the synthesis of the flagella. PF and LF are two typical properties of deep-sea bacteria for swimming and swarming, respectively (Eloe et al. 2008; Wang et al. 2008). The sodium motive force drives PF rotation, and the proton motive force powers LF translocation (McCarter 2004). PF and LF gene clusters have been detected in the genome of SM9913 and enable it to swim in open waters and also swarm in deep-sea sediments (Qin et al. 2011), whereas only the PF gene cluster is identified in TAC125 (Médigue et al. 2005). Both LF and PF gene clusters were found in $WG07_{SM9913}$; hence the existence of dual flagella system in WG07 was confirmed. All genes involved in the PF gene cluster were observed in the transcriptome of $WG07_{TAC125}$ except FliQ and FliO; whereas in $WG07_{SM9913}$, more genes involved in PF are absent including FliK, FlrB, FliS, FliD, FlgK, and FlgA. Many of

these observed genes are mutated when compared to the corresponding genes in TAC125 and SM9913 (Table 14). Overall, PF genes in WG07 are more similar to TAC125 rather than SM9913 since there are 10 mutated genes when compared to TAC125 and 20 mutated genes when SM9913 is the reference. All genes encoding LF related proteins are identified in $WG07_{SM9913}$ except LfgA, LfgK, and LafK. Four genes including motA, motB, motX, and motY responsible for sodium-driven complexes and PF rotation were observed in WG07's transcriptome, and further indicated the presence and functioning of PF equipment in WG07 (Eloe et al. 2008).

Table 14. Polar flagellum genes' mutations in WG07.

		SNV	Deletion	Insertion
TAC125	FlgL	1	-	1
	FlgC	4	-	-
	FlgB	2	-	-
	FliD	1	-	-
	ParA	2	-	-
	MotB	-	-	1
	CheY	4	-	-
	FliA	6	-	-
	FlhG	1	-	-
	FliM	3	-	-
SM9913	CheB	4	1	2
	MotA	2	-	-
	ParA	-	-	1
	CheW	3	-	2
	FlhG	-	2	3
	FliM	5	-	1
	FliG	1	-	2
	FliF	-	-	2
	FliE	4	-	3
	FlaM	2	-	-
	FlgJ	4	-	-
	FlgG	1	-	-
	FlgF	8	1	1
	FlgD	4	1	1
	FlgC	5	1	2
	FlgB	2	-	5
	CheR	1	-	1
	CheV	2	1	4
	FlgN	2	-	-
	FlgM	-	-	1

3 SNV and 1 deletion in polar flagellar protein coding gene motX, 2 SNV in flagellar motor protein PomA coding gene *motA*, 1 SNV and 2 insertions in flagellar motor switch protein G coding gene *fliG*, 5 SNV, 2 insertions and 1 deletion in flagellar biosynthesis gene *fligC*, 2 SNV and 5 insertions in flagellar biosynthesis gene *fligB*, 4 SNV, 1 insertion, and 1 deletion in flagellar basal body rod modification protein coding gene *flgD*, 5 SNV and 1 insertion in flagellar motor switch protein coding gene *fliM*, 1 insertion in negative regulator of flagellin synthesis related gene *flgM*, 2 SNV in flagellar biosynthesis chaperone *flgN*, 1 SNV in flagellar basal body rod protein coding gene *flgG*, and 8 SNV, 1 deletion, and 1 insertion in flagellar biosynthesis related gene *flgF* were detected in $WG07_{SM9913}$. In $WG07_{TAC125}$, 1 SNV in flagellar hook-associated protein coding gene *fliD*, 1 insertion in flagellar motor protein coding gene *motB*, and 2 SNV in flagellar biosynthesis related gene *flgB* were identified.

Genes involved in exopolysaccharide (EPS) biosynthesis, biofilm formation, and membranes. Strain WG07 was found to adapt to fast growth through growth experiment (Figure 19), suggesting that WG07 may live in a relatively nutrient-rich environment. Since the deep sea is often nutrient-poor, the formation of biofilms may aid in WG07's survival by maximizing its ability to capture nutrients. Also, when grown in marine broth media, WG07 quickly coagulate into amorphous particles. The formation of dense biofilms on the air-liquid interface was observed after one week post seeding. Biofilm formation usually happens in nutrient-sufficient ecosystems (Costerton et al. 1995), and biofilm formation on the air-liquid interface is a prevalent survival strategy of surface-associated bacteria such as TAC125 which can rapidly occupy this interface with a dense layer of compact cells (Médigue et al. 2005). It is speculated that biofilm

formation by WG07 may be due to the nutrient-richness of marine broth and the large air-liquid interface present in the cultivation flask. The rapid formation of particles falls in line with the life of WG07 pursuing food in the deep sea. These aggregations are mainly from extracellular release (Iriberri & Herndl 1995). EPS genes identified in TAC125 and SM9913 were observed in WG07's transcriptome as well. 16 mannose-sensitive agglutinin (MSHA) biogenesis genes were identified in $WG07_{TAC125}$. These genes have been revealed to be advantageous in *Vibrio cholerae* when adhering to a chitin surface (Meibom et al. 2004). EPS biosynthesis is a common protective mechanism for bacteria in extreme environments. Nutrient uptake by bacteria in aggregates was found to be higher than for free-living cells in low nutrient systems. Also, EPS can concentrate dissolved organic matter in the deep-sea (Nichols et al. 2005). It helps deep-sea bacteria endure extremes of temperature, salinity, and nutrient flux by changing the physicochemical environment around the bacterial cell (Nichols et al. 2005).

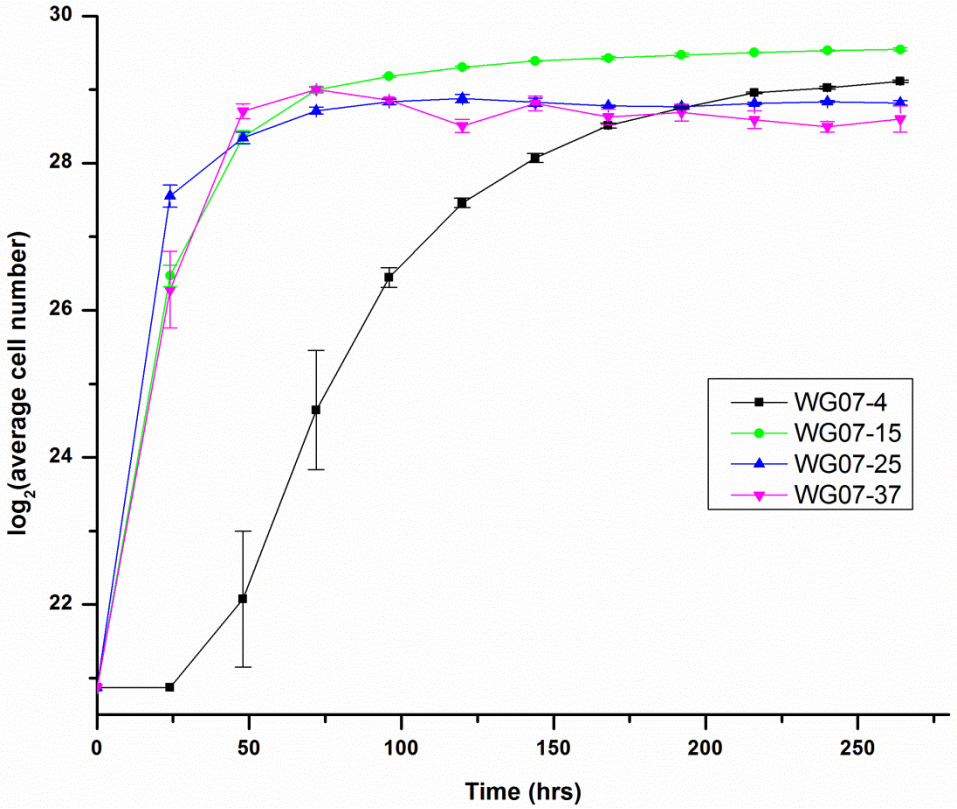

Figure 19. Growth curve of marine strain WG07 under different temperatures.

In the deep sea, WG07 faces low temperatures, hence ensuring membrane fluidity using lipid desaturases is necessary in order to survive. Lipid desaturases can also protect against dioxygen simultaneously (Médigue et al. 2005). The numbers of fatty acid desaturation genes in TAC125, SM9913, and WG07 are the same. In WG07, type II secretion is functional since general secretion pathway (GSP) proteins are present. The twin-arginine translocation (TAT) system involved in protein secretion covering most functions of the Sec pathway (Voulhoux et al. 2001) were also identified in WG07's transcriptome. 13 genes located in chromosome I responsible for the biogenesis of type IV pili in both TAC125 and SM9913 were also observed in WG07. No genes involved in type III secretion were identified in TAC125, SM9913, or WG07.

Changes of WG07's transcriptome in exponential and stationary growing phases. Microorganisms' growth is partly controlled by the availability of nutrients (Voulhoux et al. 2001). Gram-negative bacteria can enter and exit the stationary phase through complicated physiological and morphological changes depending on starvation conditions (Kolter et al. 1993). To explore how transcriptional control influences different cell growth phases in WG07, we performed a comparison of cells isolated in exponential growing phase (25E) versus stationary growing phase (25S). During the experimental process, sampled biomass for these two phases was kept as similar as possible. In total, there were 4,622 genes present and 1,748 genes were categorized into 20 COG categories in 25E and 25S, respectively, indicating a reduction in cell vitality. Previous studies suggest the reorganization of the nucleoid in the cell's stationary phase is accompanied by an overall repression of gene expression (Ramírez Santos et al. 2005). Genes expressed in different growth phases distributed differently in COG functional categories (Figure 20). In both phases, the largest portion of genes was involved in 'translation, ribosomal structure and biogenesis'. In 25E, expressed genes are enriched in COG categories 'amino acid transport and metabolism' (9.35%), 'energy production and conversion' (7.70%), and 'cell wall/membrane/envelope biogenesis' (7.40%). While in 25S, expressed genes are enriched in 'energy production and conversion' (12.53%), 'posttranslational modification, protein turnover, chaperones' (9.67%), and 'cell wall/membrane/envelope biogenesis' (5.95%). Genes involved in COG categories decline differently in 25S compared to 25E, especially those involved in 'defense mechanisms' and 'coenzyme transport and metabolism'. Using the pairwise comparison in GeneSifter with the parameters' illustrated in the methodology section, we compared genes' copy

number change in 25E and 25S and obtained the scatter plot with r^2=0.975 (Figure 21). Through this comparison, there are apparently more genes down regulated (25E > 25S) than genes up regulated (25S > 25E). The distribution of COG clusters these genes categorized in is shown in Figure 6. In COG categories 'translation, ribosomal structure and biogenesis', 'cell cycle control, cell division, chromosome partitioning', 'energy production and conversion', 'lipid transport and metabolism', 'secondary metabolites biosynthesis, transport and catabolism', 'nucleotide transport and metabolism' and 'transcription', the proportion of down regulated genes was remarkably higher than that of up regulated genes, while in COG categories 'defense mechanisms', 'cell motility', and 'cell wall/membrane/envelope biogenesis', the proportion of up regulated genes was higher. This is consistent with previous findings that show when proliferating cells exhaust available nutrients and enter the stationary phase, they develop a multiple stress-resistance state characterized by morphological changes which include thickening of the cell wall, rounded shape, and loss of flagella (Ramírez Santos et al. 2005). The COG distribution observed here is not out of the realm of expectations, further indicating that the comparative transcriptomic approach is a reliable method to study the transcriptomic variation of gram-negative bacteria subject to different conditions. Among down regulated genes, ribosomal replication protein coding gene priB in COG category 'replication, recombination and repair' shows the highest change fold (over 13), followed by ribosomal protein coding genes rpsH (change fold ~11) and ATP synthase gene *atpC* (change fold ~ 10). While among down regulated genes, the twitching motility protein coding gene *pilT* in category 'intracellular trafficking, secretion, and vesicular transport'

shows the highest change fold of 9, followed by glutamy kinase gene *proB* (change fold ~8.5) and mechano sensitive channel protein coding gene PSHAa2520 (change fold ~7).

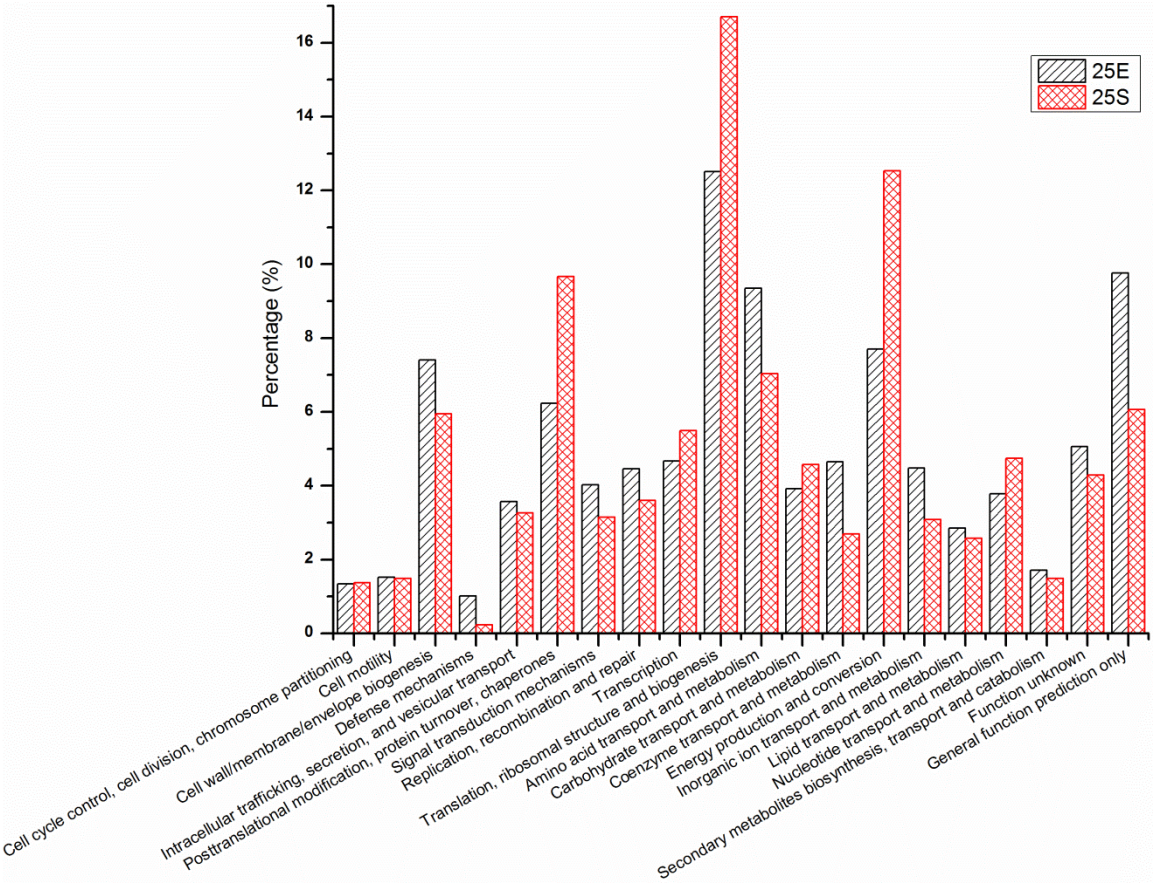

Figure 20. COG distribution of WG07's transcriptome in exponential and stationary growing phases.

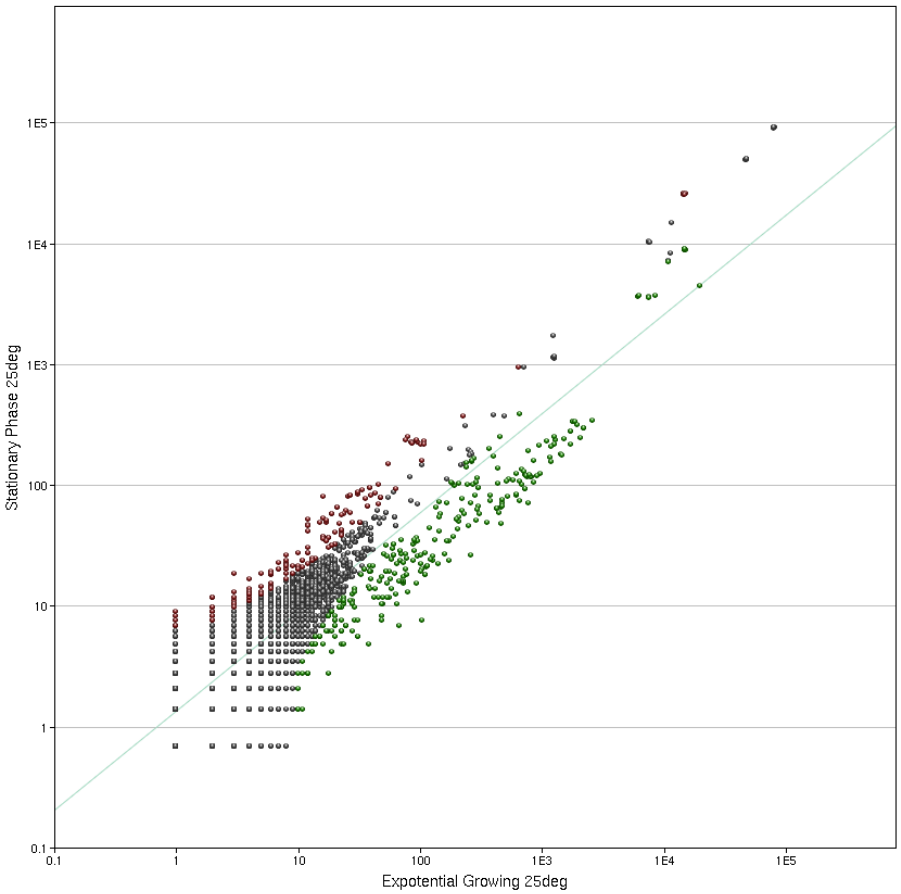

Figure 21. Scatter plot of up and down regulated genes identified in the transcriptome of WG07 in its exponential and stationary phases.

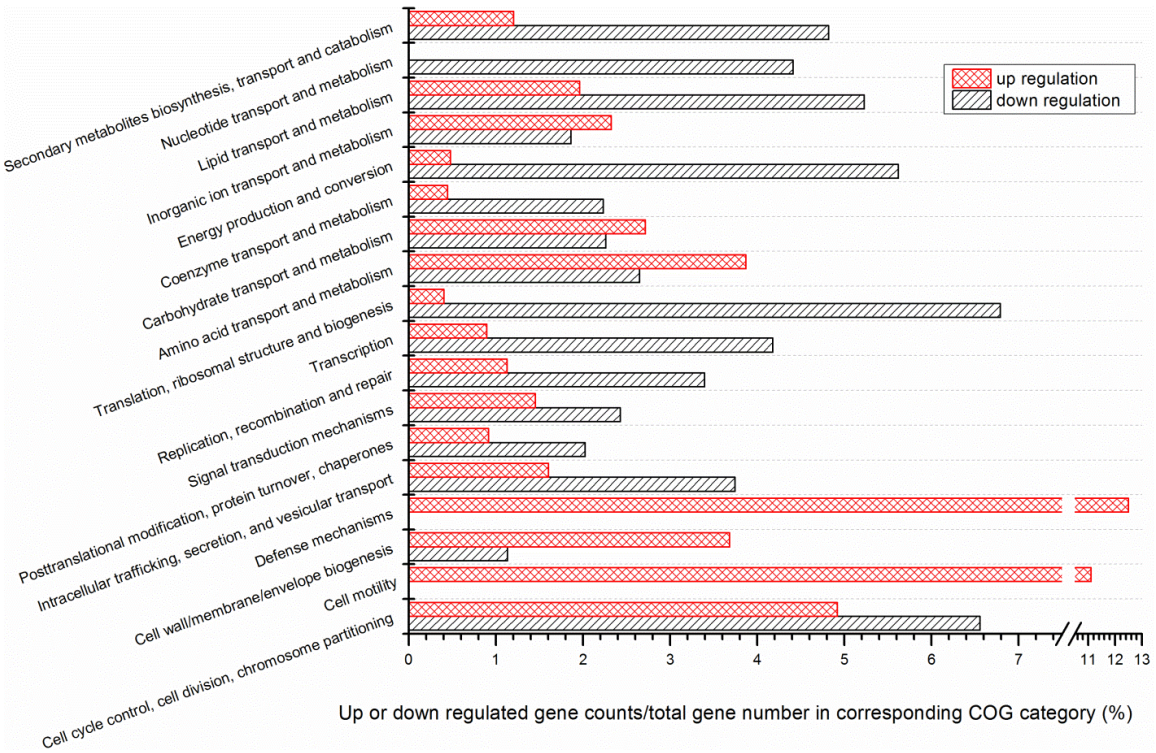

Figure 22. Distribution of COG clusters differently expressed genes categorized into in WG07's exponential and stationary phases.

Transcriptomic and proteomic change of WG07 grown under different temperatures. To further describe the new deep-sea isolate WG07, its growth under four different temperatures, 4°C, 15°C, 25°C, and 37°C, was examined (Figure 19). Overall, 25°C turned out to be the optimal temperature for WG07 to grow with the shortest generation time (3.62 hours) compared to the generation times of 13.00 hours (4°C), 4.32 hours (15°C), and 6.17 hours (37°C), respectively. The same strategy was used to look through the gene regulations as was done for the different growing-phase experiment. 4°C and 25°C were chosen to see the following transcriptome and proteome change. Surprisingly, genes involved in almost all COG categories of WG07 grown under lower temperature 4°C showed much higher expression activity than under higher temperature 25°C, except those in categories 'Intracellular trafficking and secretion' and 'Posttranslational

modification, protein turnover, chaperones' (Figure 23). Most genes that are more active in 25ºC than in 4ºC in 'Intracellular trafficking and secretion' are involved in general secretion pathway and translocation, such as general secretion pathway protein coding genes PSM_A0732, *exeD*, and *exeE*, and preprotein translocase subunit gene *secY*, etc. This result is further supported by the proteomic data, in which remarkably less protein was identified in 4ºC than in 25ºC (data not shown), which is in contrast to proteomic data. Among these, only 2 proteins were expressed in an obviously higher level in 4ºC than in 25ºC, CspX (change fold ≈ 3) and CspC (change fold ≈ 2), both of which are cold shock proteins (Figure 24).

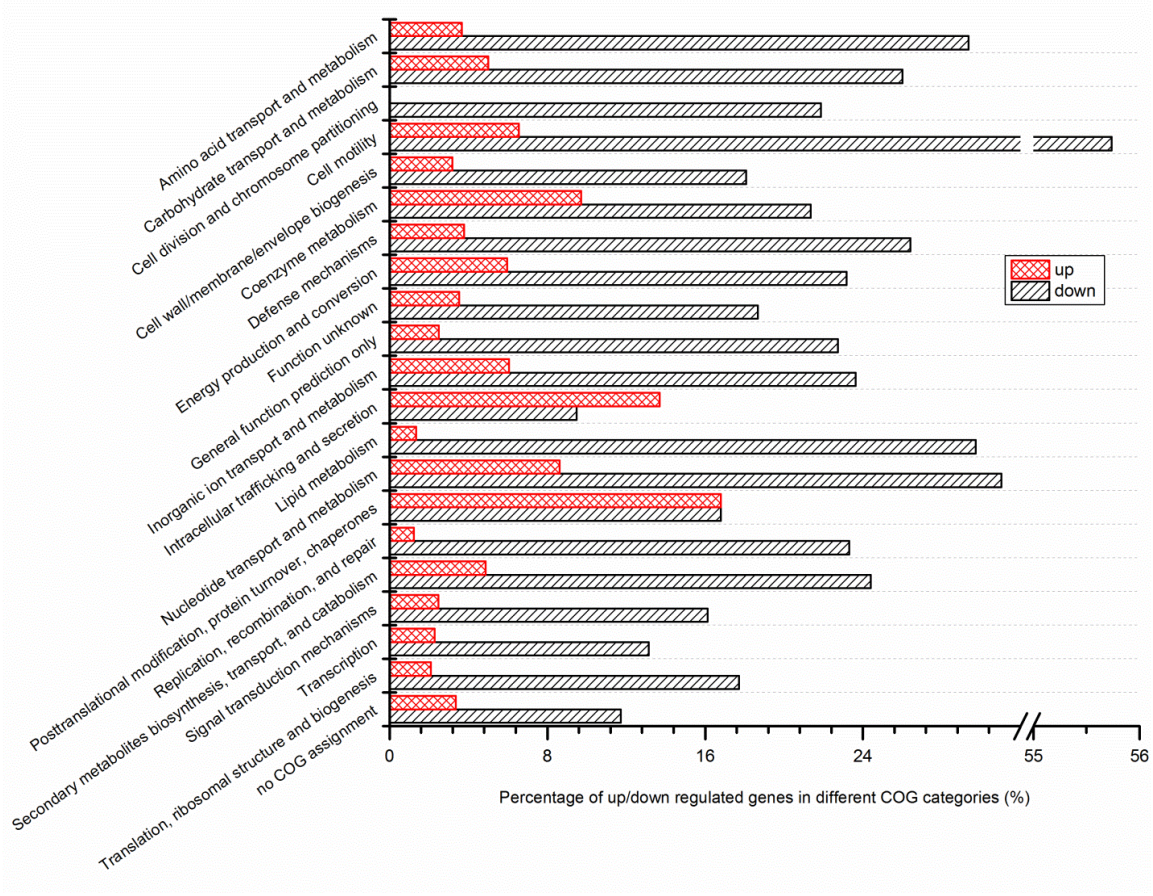

Figure 23. Distribution of COG clusters differently expressed genes categorized into in transcriptomes of WG07 grown in 25ºC and 4ºC.

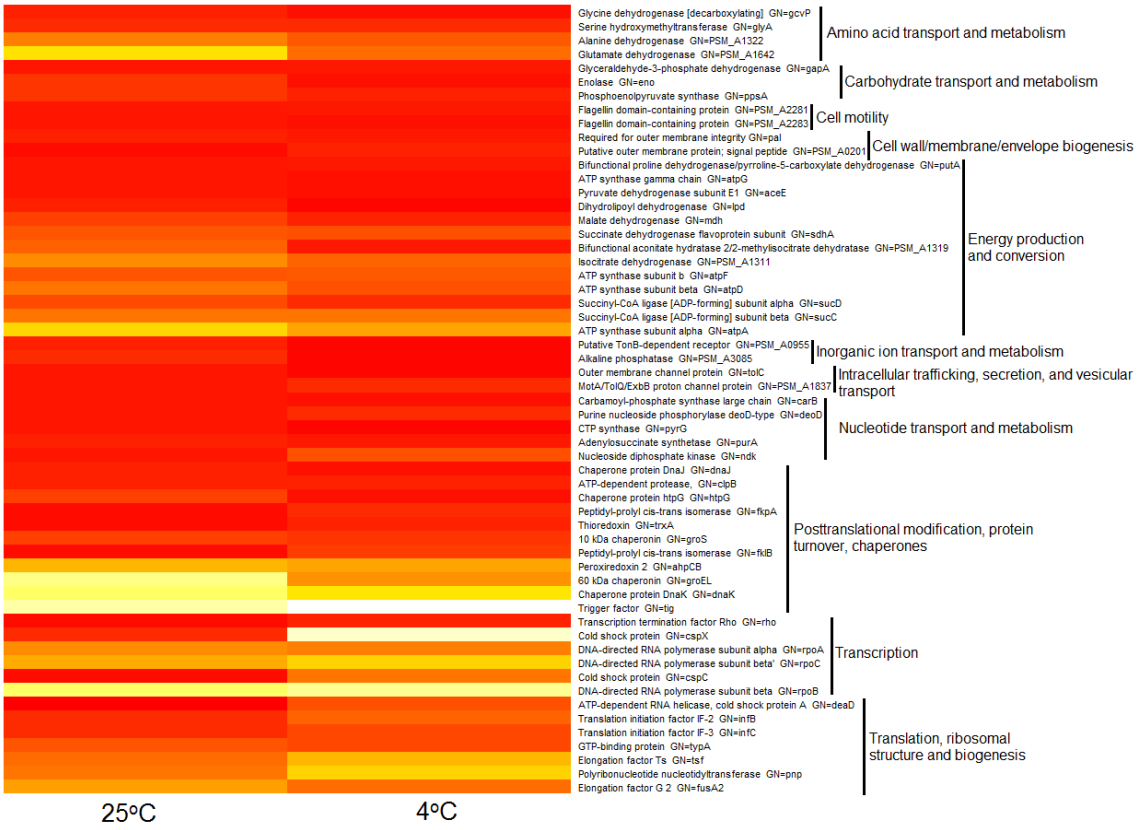

Figure 24. Heatmap of some differently expressed proteins in WG07 grown in 25ºC and 4ºC.

The transcriptomic data shows that, in contrast to slower cellular growth, transcriptomic level except protein secretion and translocation related genes was enhanced in the WG07 strain exposed to low temperature compared to the strain in optimal temperature (25ºC). Considering the proteomic data revealing down-regulation in overall expression at 4°C, this result further suggests that during growth under these conditions, the deep-sea marine isolate WG07, although able to transcribe the required mRNA, may lack the cellular resources required for translation. Similar phenomena were observed in previous studies in *E.coli* in cold (Kocharunchitt et al. 2011), and *Shewanella oneidensis* MR-1in hypo-osmotic stress (Gao et al. 2005).

Conclusion

This chapter presents an in-depth comparison of the newly isolated deep-sea strain *Pseudoaltermonas* sp.WG07's transcriptome with the genomes of the surface-water strain *Pseudoalteromonas haloplanktis* TAC125 and the marine-sediment strain *Pseudoalteromonas* sp. SM9913. Transcriptome survey and related functional analysis of WG07 reveal unique features of this deep-sea isolate allowing it to adapt to its niche between surface water and ocean sediment. Genes encoding proteins involved in 'Amino acid transport and metabolism', 'Carbohydrate transport and metabolism', 'Cell motility', 'Cell wall/membrane/envelope biogenesis', 'Defense mechanisms', 'Signal transduction mechanisms', and 'Translation, ribosomal structure and biogenesis' are more enriched in WG07 and SM9913 than in TAC125; while for 'Cell cycle control, mitosis and meiosis', 'Coenzyme transport and metabolism', 'Energy production and conversion', 'Intracellular trafficking, secretion, and vesicular transport', 'Lipid transport and metabolism', 'Posttranslational modification, protein turnover, chaperones', and 'Secondary metabolites biosynthesis, transport and catabolism', WG07 and TAC125's related genes are more than those of SM9913. WG07 was found to be phylogenetically closer to SM9913, a reasonable conclusion considering their dwelling depths. Comparative transcriptomic analysis of WG07 in exponential growing phase and stationary growing phase provided us with information on WG07's transcriptome change during growth. Genes involved in 'translation, ribosomal structure and biogenesis', 'cell cycle control, cell division, chromosome partitioning', 'energy production and conversion', 'lipid transport and metabolism', 'transcription', 'transcription' etc. are actively expressed in exponential phase, while those involved in 'defense mechanisms',

'cell motility', and 'cell wall/membrane/envelope biogenesis' were up regulated in stationary phase. This is consistent with previous findings in gram-negative bacteria *E.coli* and yeast *Saccharomyces cerevisiae* and further supports the use of comparative transcriptomic analysis as are liable tool to study transcriptome-wise variation in gram-negative bacteria under varying conditions. Transcriptomic and proteomic comparison of WG07 grown in 4^{o}C and 25^{o}C revealed that during growth in low temperature, the deep-sea marine isolate WG07, although able to transcribe the required mRNA, may lack the cellular resources required for translation.

CHAPTER 6
THESIS CONTRIBUTION SUMMARY AND FUTURE WORK

Summary of research contributions

Ph.D. Research Contributions:

Phylogenetic analysis of deep-sea bacteria, archaea, and eukaryota communities using 16s rDNA genes; phylogenetic and gene expression analysis of photosynthetic *Cyanobacteria* and diatoms in deep-sea;

Performance evaluation of 4 different commercially available whole-transcriptomic amplification kits and optimization of whole-transcriptome amplification from low cell density deep-sea microbial samples for metatranscriptomic analysis;

Phylogenetic and functional analysis of deep-sea microbial community using integrated metagenomic and metatranscriptomic approach; strength of metabolic activity evaluation of microbial community in species level by introducing the ratio of RNA/DNA;

Phenotypic and potentially functional characterization of deep-sea *Pseudoalteromonas* strains isolated from deep-sea by annotating and comparing its transcriptome with the genomes of reference strains *Pseudoalteromonas* haloplanktis TAC125 and *Pseudoalteromonas* sp. SM9913; linked in-depth observation of transcriptomic and proteomic changes of WG07 under different growing temperatures.

Future work

Nowadays, more taxonomic and functional studies of deep-sea microbial communities are conducted to unveil deep-sea microbial ecology. However, the deep-sea environments, huge microbial reservoirs, still remain mostly unknown. The development

and broad utilization of next-generation sequencing based metagenomic and metatranscriptomic analysis, and also the increasingly sophisticated computer modeling technologies in this field make it possible to uncover the full taxonomical picture of microbial communities in the special dwelling environments. Recent studies found that prokaryotic microbes in deep-sea environments are well-adapted to the special dwelling environments after long evolution; carrying genetic features that enable them to live and reproduce in the extreme environmental conditions. Recent progress in sequencing technologies is fueling a rapid increase in the number and scope of deep-sea microbial community-targeted studies. While metagenomic analysis can provide information on the taxonomic composition and metabolic potential of microbial communities in deep sea, metatranscriptomics serves to unveil the actual metabolic activities of the communities at a specific time and location, and how those activities are changing in response to environmental and biotic parameters.

For the study of searching for photosynthetic cyanobacteria and diatoms in the twilight waters, one of several possibilities is that those photoautotrophic microbes are a natural part of activity at this depth. As one support to this hypothesis, it has been suggested that diatoms *Thalassiosira* spp. and cyanobacteria *Synechococcus* sp. are microbes with high adaptation ability and can function well in less favorable environments. In early studies, bacterial chemolithotrophy has been suggested as a key player for organic carbon generation at 700−900 m sea depth, and the energy for this process may be in part provided by detrital NH_4^+ derived from the downward flux of large particles (Karl et al. 1984). This possibility, if true, will imply that the functional boundary of photoautotrophic activity may be beyond the epipelagic zone, deeper than

previously described, and their contribution to the oceanic carbon cycles may need further investigation.

In the study of integrated metagenomic and metatranscriptomic analysis of microbial communities in the deep-sea, it is quite surprising to observe high abundance of plocozoa down at this depth. Further study is definitely worthwhile to try to compare with other deep-sea microbial communities in different locations. Also, *Cyanobacteria* were identified in samples from all four deep-sea sites by metagenomic and metatranscriptomic analysis, suggesting their active functionality in deep-sea environments in spite of very little sunlight. RNA degradation process under dark environment and low temperature (~4^{o}C) would be an interesting and meaningful topic to study to further prove the hypothesis we suggested in Chpater 2.

During the study of –omic characterization of newly isolated deep-sea Pseudoalteromonas strain, transcriptomic and proteomic comparison of WG07 grown in 4^{o}C and 25^{o}C revealed that during growth in low temperature, the deep-sea marine isolate WG07, although able to transcribe the required mRNA, may lack the cellular resources required for translation. It is another point that conflicts with previous expectations. This is consistent with previous findings in gram-negative bacteria *E.coli* and yeast *Saccharomyces cerevisiae* and further supports the use of comparative transcriptomic analysis as a liable tool to study transcriptome-wise variation in gram-negative bacteria under varying conditions. However, more case studies and deeper mechanism research are required to further explore this phenomenon.

REFERENCES

D. A. Burns JSM. (1982). Further examples of the dinoflagellate genus Ceratium from New Zealand coastal waters. New Zealand Journal of Marine and Freshwater Research 16:57–67.

Agogué H, Brink M, Dinasquet J, Herndl GJ. (2008). Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. Nature 456:788–791.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Amacher J, Neuer S, Lomas M. DNA-based molecular fingerprinting of eukaryotic protists and cyanobacteria contributing to sinking particle flux at the bermuda atlantic time-series study. Deep-Sea Research II.

Arístegui J, Duarte CM, Agustí S, Doval M, Alvarez-Salgado XA, Hansell DA. (2002). Dissolved organic carbon support of respiration in the dark ocean. Science 298:1967.

Arístegui J, Duarte CM, Gasol JM, Alonso-Sáez L. (2005). Active mesopelagic prokaryotes support high respiration in the subtropical northeast Atlantic Ocean. Geophys. Res. Lett. 32:4 PP.

Aristegui J, GASOL JM, DUARTE CM, HERNDL GJ. (2009). Microbial oceanography of the dark ocean's pelagic realm. Limnol. Oceanogr. 54:1501–1529.

Arístegui J, GASOL JM, DUARTE CM, Herndl GJ. (2009). Microbial oceanography of the dark ocean's pelagic realm. Limnol. Oceanogr. 54. http://www.aslo.org/lo/toc/vol_54/issue_5/1501.html (Accessed November 29, 2011).

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science 306:79–86.

Bass D, Howe A, Brown N, Barton H, Demidova M, Michelle H, et al. (2007). Yeast forms dominate fungal diversity in the deep oceans. Proc. Biol. Sci. 274:3069–3077.

Bauer AW, Kirby WMM, Sherris JC, Turck M. (1966). Antibiotic susceptibility testing by a standardized single disk method. American Journal of Clinical Pathology 45:493–496.

Bhadury P, Bik H, Lambshead JD, Austen MC, Smerdon GR, Rogers AD. (2011). Molecular diversity of fungal phylotypes co-amplified alongside nematodes from coastal and deep-sea marine environments. PLoS ONE 6:e26445.

Biers EJ, Sun S, Howard EC. (2009). Prokaryotic Genomes and Diversity in Surface Ocean Waters: Interrogating the Global Ocean Sampling Metagenome. Appl. Environ. Microbiol. 75:2221–2229.

Billett DSM, Lampitt RS, Rice AL, Mantoura RFC. (1983). Seasonal sedimentation of phytoplankton to the deep-sea benthos. , Published online: 07 April 1983.

Binnenkade L, Lassak J, Thormann KM. (2011). Analysis of the BarA/UvrY Two-Component System in Shewanella oneidensis MR-1. PLoS ONE 6:e23440.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. (1997). The complete genome sequence of Escherichia coli K-12. Science 277:1453–1462.

Boysen-Ennen E, Hagen W, Hubold G, Piatkowski U. (1991). Zooplankton biomass in the ice-covered Weddell Sea, Antarctica. Mar. Biol. 111:227–235.

Braker G, Zhou J, Wu L, Devol AH, Tiedje JM. (2000). Nitrite Reductase Genes (nirK andnirS) as Functional Markers To Investigate Diversity of Denitrifying Bacteria in Pacific Northwest Marine Sediment Communities. Appl. Environ. Microbiol. 66:2096–2104.

Braun V, Braun M. (2002). Active transport of iron and siderophore antibiotics. Current Opinion in Microbiology 5:194–201.

Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM, et al. (2009). Microbial community structure in the North Pacific ocean. ISME J 3:1374–1386.

De Buron I, Morand S. (2004). Deep-sea hydrothermal vent parasites: why do we not find more? Parasitology 128:1–6.

Byrne N, Strous M, Crépeau V, Kartal B, Birrien J-L, Schmid M, et al. (2009). Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. The ISME Journal 3:117–123.

Cabiscol E, Tamarit J, Ros J. (2000). Oxidative stress in bacteria and protein damage by reactive oxygen species. Int. Microbiol. 3:3–8.

Le Calvez T, Burgaud G, Mahé S, Barbier G, Vandenkoornhuyse P. (2009). Fungal diversity in deep-sea hydrothermal ecosystems. Appl. Environ. Microbiol. 75:6415–6421.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, et al. (2005). The ribosomal database project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. 33:D294–296.

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappin-Scott HM. (1995). Microbial biofilms. Annual Review of Microbiology 49:711–745.

Cui Z, Lai Q, Dong C, Shao Z. (2008). Biodiversity of polycyclic aromatic hydrocarbon-degrading bacteria from deep sea sediments of the Middle Atlantic Ridge. Environmental Microbiology 10:2138–2149.

Danovaro R, Dell'anno A, Pusceddu A, Fabiano M. (1999). Nucleic acid concentrations (DNA, RNA) in the continental and deep-sea sediments of the eastern Mediterranean: relationships with seasonally varying organic inputs and bacterial dynamics. Deep Sea Research Part I: Oceanographic Research Papers 46:1077–1094.

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

Díez B, Pedrós-Alió C, Massana R. (2001). Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl. Environ. Microbiol. 67:2932–2941.

Do HK, Kogure K, Simidu U. (1990). Identification of deep-sea-sediment bacteria which produce tetrodotoxin. Appl. Environ. Microbiol. 56:1162–1163.

Domigan LJ, Scally SW, Fogg MJ, Hutton CA, Perugini MA, Dobson RCJ, et al. (2009). Characterisation of dihydrodipicolinate synthase (DHDPS) from Bacillus anthracis. Biochim. Biophys. Acta 1794:1510–1516.

DuRand MD, Olson RJ, Chisholm SW. (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. Deep Sea Research Part II: Topical Studies in Oceanography 48:1983–2003.

Eitel M, Schierwater B, Eitel M, Schierwater B. (2010). The phylogeography of the Placozoa suggests a taxon-rich phylum in tropical and subtropical waters, The

phylogeography of the Placozoa suggests a taxon-rich phylum in tropical and subtropical waters. Molecular Ecology, Molecular Ecology 19, 19:2315, 2315–2327, 2327.

El-Hajj ZW, Tryfona T, Allcock DJ, Hasan F, Lauro FM, Sawyer L, et al. (2009). Importance of proteins controlling initiation of DNA replication in the growth of the high-pressure-loving bacterium photobacterium profundum SS9. J. Bacteriol. 191:6383–6393.

Eloe EA, Fadrosh DW, Novotny M, Zeigler Allen L, Kim M, Lombardo M-J, et al. (2011). Going Deeper: Metagenome of a hadopelagic microbial community. PLoS ONE 6:e20388.

Eloe EA, Lauro FM, Vogel RF, Bartlett DH. (2008). The deep-sea bacterium Photobacterium profundum SS9 utilizes separate flagellar systems for swimming and swarming under high-pressure conditions. Appl. Environ. Microbiol. 74:6298–6305.

Falkowski PG, Barber RT, Smetacek V. (1998). Biogeochemical controls and feedbacks on ocean primary production. Science 281:200–206.

Falkowski PG, Raven JA. Aquatic photosynthesis: (second edition). http://press.princeton.edu/titles/8337.html (Accessed February 12, 2013).

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

Frost BW. (1974). Calanus marshallae, a new species of calanoid copepod closely allied to the sibling species C. finmarchicus and C. glacialis. Mar. Biol. 26:77–99.

Gailhard I, Gros P, Durbec JP, Beliaeff B, Belin C, Nzan E, et al. (2002). Variability patterns of microphytoplankton communities along the French coasts. Mar Ecol Prog Ser 242:39–50.

Gao W, Liu Y, Wu L, Liu X, Yan T, Zhou J. (2005). A transcriptome perspective of shewanella oneidensis MR-1 cells adapted to hypoosmotic stress. In:Atlanta, GA.

Gao W, Shi X, Wu J, Jin Y, Zhang W, Meldrum DR. (2011). Phylogenetic and gene expression analysis of cyanobacteria and diatoms in the twilight waters of the temperate northeast Pacific Ocean. Microb. Ecol. 62:765–775.

Gärtner A, Blümel M, Wiese J, Imhoff JF. (2011). Isolation and characterisation of bacteria from the Eastern Mediterranean deep sea. Antonie Van Leeuwenhoek 100:421–435.

Gestal C, Pascual S, Hochberg FG. (2010). Aggregata bathytherma sp. nov. (Apicomplexa:Aggregatidae), a new coccidian parasite associated with a deep-sea hydrothermal vent octopus. Dis. Aquat. Org. 91:237–242.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, et al. (2008). Detection of large numbers of novel sequences in the Metatranscriptomes of complex marine microbial communities. PLoS ONE 3:e3042.

Gill RT, Katsoulakis E, Schmitt W, Taroncher-Oldenburg G, Misra J, Stephanopoulos G. (2002). Genome-wide dynamic transcriptional profiling of the light-to-dark transition in Synechocystis sp. strain PCC 6803. J. Bacteriol. 184:3671–3681.

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

Glud RN. (2008). Oxygen dynamics of marine sediments. Marine Biology Research 4:243–289.

González JM, Covert JS, Whitman WB, Henriksen JR, Mayer F, Scharf B, et al. (2003). Silicibacter pomeroyi sp. nov. and Roseovarius nubinhibens sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. Int. J. Syst. Evol. Microbiol. 53:1261–1269.

Greuet C. (1978). Ultrastructural organization of the ocelloide of Nematodinium. Phylogenetic aspect of the evolution of Warnowiidae Lindemann dinoflagellates photoreceptor. Cytobiologie 17:114–136.

Gruber N, Galloway JN. (2008). An Earth-system perspective of the global nitrogen cycle. Nature 451:293–296.

Herndl GJ, Agogu H, Baltar F, Reinthaler T, Sintes E, Varela MM. (2008). Regulation of aquatic microbial processes: the 'microbial loop' of the sunlit surface waters and the dark ocean dissected. Aquat Microb Ecol 53:59–68.

Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, et al. (2005). Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. Appl. Environ. Microbiol. 71:2303–2309.

Heusipp G, Fälker S, Schmidt MA. (2007). DNA adenine methylation and bacterial pathogenesis. Int. J. Med. Microbiol. 297:1–7.

Higgins NP. (1992). Death and transfiguration among bacteria. Trends in Biochemical Sciences 17:207–211.

Hijum S van, Baerends R, Zomer A, Karsens H, Martin-Requena V, Trelles O, et al. (2008). Supervised Lowess normalization of comparative genome hybridization data – application to lactococcal strain comparisons. BMC Bioinformatics 9:93.

Holmström C, James S, Neilan BA, White DC, Kjelleberg S. (1998). Pseudoalteromonas tunicata sp. nov., a bacterium that produces antifouling agents. Int J Syst Bacteriol 48:1205–1212.

Holmström C, Kjelleberg S. (1999). Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. FEMS Microbiology Ecology 30:285–293.

Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, Kawarabayasi Y, et al. (2004). Genome sequence of the deep-sea gamma-proteobacterium Idiomarina loihiensis reveals amino acid fermentation as a source of carbon and energy. Proc. Natl. Acad. Sci. U.S.A. 101:18036–18041.

Hourdez S, Weber RE. (2005). Molecular and functional adaptations in deep-sea hemoglobins. Journal of Inorganic Biochemistry 99:130–141.

Hu S, Zheng H, Gu Y, Zhao J, Zhang W, Yang Y, et al. (2011). Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in Clostridium acetobutylicum EA 2018. BMC Genomics 12:93.

Hwang J-S. (2006). A 5-year study of the influence of the northeast and southwest monsoons on copepod assemblages in the boundary coastal waters between the East China Sea and the Taiwan Strait. Journal of Plankton Research 28:943–958.

Ikeda T, Sano F, Yamaguchi A, Matsuishi T. (2007). RNA:DNA ratios of calanoid copepods from the epipelagic through abyssopelagic zones of the North Pacific Ocean. Aquat Biol 1:99–108.

Imhoff JF. (2001). True marine and halophilic anoxygenic phototrophic bacteria. Arch Microbiol 176:243–254.

Iriberri J, Herndl GJ. (1995). Formation and microbial utilization of amorphous aggregates in the sea: ecological significance. Microbiologia 11:309–322.

Ivanova EP, Romanenko LA, Matté MH, Matté GR, Lysenko AM, Simidu U, et al. (2001). Retrieval of the species Alteromonas tetraodonis Simidu et al. 1990 as Pseudoalteromonas tetraodonis comb. nov. and emendation of description. Int J Syst Evol Microbiol 51:1071–1078.

Jannasch H. (1995). Microbial Interactions With Hydrothermal Fluids. In seafloor hydrothermal systems: physical, chemical, biological, and geological interactions, American Geophysical Union, pp. 273–296.

Jeong HG, Suh H-L, Yoon YH, Soh HY. (2008). Redescription of Pontella species (Calanoida, Pontellidae) from Korean waters, with notes on their spatio-temporal distribution. Ocean Sci. J. 43:209–222.

Jørgensen BB, Boetius A. (2007). Feast and famine — microbial life in the deep-sea bed. Nature Reviews Microbiology 5:770–781.

Kalanetra KM, Bano N, Hollibaugh JT. (2009). Ammonia-oxidizing Archaea in the Arctic Ocean and Antarctic coastal waters. Environ. Microbiol. 11:2434–2445.

Karl DM, Knauer GA, Martin JH, Ward BB. (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. Nature 309:54–56.

Karner MB, DeLong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature 409:507–510.

Katz ME, Tjalsma RC, Miller KG. (2003). Oligocene bathyal to abyssal benthic foraminifera of the Atlantic Ocean. Micropaleontology 49:1–45.

Kawagucci S, Yoshida YT, Noguchi T, Honda MC, Uchida H, Ishibashi H, et al. (2012). Disturbance of deep-sea environments induced by the M9.0 Tohoku Earthquake. Sci Rep 2.

Keane TM, Naughton TJ, McInerney JO. (2007). MultiPhyl: a high-throughput phylogenomics webserver using distributed computing. Nucleic Acids Res. 35:W33–37.

Kimor B, Gordon N, Neori A. (1992). Symbiotic associations among the microplankton in oligotrophic marine environments, with special reference to the Gulf of Aqaba, Red Sea. J. Plankton Res. 14:1217–1231.

Kirkpatrick J, Oakley B, Fuchsman C, Srinivasan S, Staley JT, Murray JW. (2006). Diversity and distribution of Planctomycetes and related bacteria in the suboxic zone of the Black Sea. Appl Environ Microbiol 72:3079–3083.

Klimpel S, Palm HW, Rückert S, Piatkowski U. (2004). The life cycle of Anisakis simplex in the Norwegian Deep (northern North Sea). Parasitol. Res. 94:1–9.

Knittel K, Boetius A, Lemke A, Eilers H, Lochte K, Pfannkuche O, et al. (2003). Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia Margin, Oregon). Geomicrobiology Journal 20:269–294.

Koblížek M, Komenda J, Masojídek J, Pechar L. (2001). Cell aggregation of the cyanobacterium synechococcus elongatus: role of the electron transport chain. Journal of Phycology 36:662–668.

Kocharunchitt C, King T, Gobius K, Bowman JP, Ross T. (2011). Integrated transcriptomic and proteomic analysis of the physiological response of Escherichia coli O157:H7 Sakai to steady-state conditions of cold and water activity stress. Molecular & Cellular Proteomics: MCP. (Accessed October 29, 2011).

Kolber ZS, Dover CLV, Niederman RA, Falkowski PG. (2000). Bacterial photosynthesis in surface waters of the open ocean. Nature 407:177–179.

Kolber ZS, Gerald F, Plumley, Lang AS, Beatty JT, Blankenship RE, et al. (2001). Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. Science 292:2492–2495.

Kolter R, Siegele DA, Tormo A. (1993). The stationary phase of the bacterial life cycle. Annu. Rev. Microbiol. 47:855–874.

Konstantinidis KT, Braff J, Karl DM, DeLong EF. (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. Appl. Environ. Microbiol. 75:5345–5355.

Kosobokova KN, Hanssen H, Hirche H-J, Knickmeier K. (1997). Composition and distribution of zooplankton in the Laptev Sea and adjacent Nansen Basin during summer, 1993. Polar Biol 19:63–76.

Kuypers MMM, Lavik G, Woebken D, Schmid M, Fuchs BM, Amann R, et al. (2005). Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. Proc. Natl. Acad. Sci. U.S.A. 102:6478–6483.

Laghdass M, Catala P, Caparros J, Oriol L, Lebaron P, Obernosterer I. (2012). High contribution of sar11 to microbial activity in the north west Mediterranean Sea. Microb. Ecol. 63:324–333.

Lambshead PJD, Boucher G. (2003). Marine nematode deep-sea biodiversity – hyperdiverse or hype? Journal of Biogeography 30:475–485.

Lampitt RS. (1985). Evidence for the seasonal deposition of detritus to the deep-sea floor and its subsequent resuspension. Deep Sea Research Part A. Oceanographic Research Papers 32:885–897.

Li, Kato, Horikoshi. (1999). Microbial diversity in sediments collected from the deepest cold-seep area, the Japan Trench. Mar. Biotechnol. 1:391–400.

Liao L, Xu X-W, Jiang X-W, Wang C-S, Zhang D-S, Ni J-Y, et al. (2011). Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the Pacific Ocean. FEMS Microbiol. Ecol. 78:565–585.

Madhupratap M, Haridas P. (1990). Zooplankton, especially calanoid copepods, in the upper 1000 m of the south-east Arabian Sea. J. Plankton Res. 12:305–321.

Marina A, Mott C, Auyzenberg A, Hendrickson WA, Waldburger CD. (2001). Structural and mutational analysis of the phoq histidine kinase catalytic domain insight into the reaction mechanism. J. Biol. Chem. 276:41182–41190.

Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A, et al. (2007). Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. PLoS ONE 2:e914.

Maruyama YK. (2004). Occurrence in the field of a long-term, year-round, stable population of placozoans. Biol. Bull. 206:55–60.

Mascher T, Helmann JD, Unden G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol. Mol. Biol. Rev. 70:910–938.

Massana R, Murray AE, Preston CM, DeLong EF. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. Appl. Environ. Microbiol. 63:50–56.

Massana R, Pedrós-Alió C. (2008). Unveiling new microbial eukaryotes in the surface ocean. Curr. Opin. Microbiol. 11:213–218.

Matsuyama H, Minami H, Kasahara H, Kato Y, Murayama M, Yumoto I. (2012). Pseudoalteromonas arabiensis sp. nov., a novel marine polysaccharide-producing bacterium. Int J Syst Evol Microbiol. (Accessed December 1, 2012).

McCarter LL. (2004). Dual flagellar systems enable motility under different circumstances. J. Mol. Microbiol. Biotechnol. 7:18–29.

Médigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, et al. (2005a). Coping with cold: the genome of the versatile marine Antarctica bacterium Pseudoalteromonas haloplanktis TAC125. Genome Res. 15:1325–1335.

Médigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, et al. (2005b). Coping with cold: the genome of the versatile marine Antarctica bacterium Pseudoalteromonas haloplanktis TAC125. Genome Res. 15:1325–1335.

Meibom KL, Li XB, Nielsen AT, Wu C-Y, Roseman S, Schoolnik GK. (2004). The Vibrio cholerae chitin utilization program. Proc Natl Acad Sci U S A 101:2524–2529.

Merckx B, Goethals P, Steyaert M, Vanreusel A, Vincx M, Vanaverbeke J. (2009). Predictability of marine nematode biodiversity. Ecological Modelling 220:1449–1458.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386.

Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environ. Microbiol. 9:1162–1175.

Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, et al. (2011). Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinformatics 12:S21.

Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, et al. (2004). Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment. Nature 432:910–913.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. (2008). The transcriptional landscape of the yeast genome defined by rna sequencing. Science 320:1344–1349.

Nakagawa S, Takaki Y, Shimamura S, Reysenbach A-L, Takai K, Horikoshi K. (2007). Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. Proc. Natl. Acad. Sci. U.S.A. 104:12146–12150.

Nakai R, Abe T, Takeyama H, Naganuma T. (2011). Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. Mar. Biotechnol. 13:900–908.

Nichols CAM, Guezennec J, Bowman JP. (2005). Bacterial Exopolysaccharides from Extreme Marine Environments with Special Consideration of the Southern Ocean, Sea Ice, and Deep-Sea Hydrothermal Vents: A Review. Marine Biotechnology 7:253–271.

Nikaido H. (1996). Multidrug efflux pumps of gram-negative bacteria. J. Bacteriol. 178:5853–5859.

Nishino K, Yamaguchi A. (2001). Analysis of a complete library of putative drug transporter genes in Escherichia coli. J. Bacteriol. 183:5803–5812.

Olsen GJ, Lane DJ, Giovannoni SJ, Pace NR, Stahl DA. (1986). Microbial ecology and evolution: a ribosomal RNA approach. Annu. Rev. Microbiol. 40:337–365.

Orcutt Beth N, Sylvan JB, Knab NJ, Edwards KJ. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. Microbiol. Mol. Biol. Rev. 75:361–422.

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, et al. (2003). The genome of a motile marine Synechococcus. Nature 424:1037–1042.

Park H, Myung S, Kim W. (2012). Comparative transcriptomic analysis of streptococcus pseudopneumoniae with viridans group streptococci. BMC Microbiology 12:77.

Patterson DJ, Nygaard K, Steinberg G, Turley CM. (1993). Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid North Atlantic. Journal of the Marine Biological Association of the United Kingdom 73:67–95.

Pester M, Rattei T, Flechl S, Gröngröft A, Richter A, Overmann J, et al. (2012). amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. Environmental Microbiology 14:525–539.

Pester M, Schleper C, Wagner M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr. Opin. Microbiol. 14:300–306.

Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. (2011). Application of RNA-seq to reveal the transcript profile in bacteria. Genet. Mol. Res. 10:1707–1718.

Polard P, Chandler M. (1995). Bacterial transposases and retroviral integrases. Molecular Microbiology 15:13–23.

Poole K. (2004). Efflux-mediated multiresistance in gram-negative bacteria. Clin. Microbiol. Infect. 10:12–26.

Poretsky RS, Gifford S, Rinta-Kanto J, Vila-Costa M, Moran MA. (2009). Analyzing gene expression from marine microbial communities using environmental transcriptomics. J Vis Exp. (Accessed December 26, 2011).

Purkhold U, Pommerening-Röser A, Juretschko S, Schmid MC, Koops HP, Wagner M. (2000). Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: implications for molecular diversity surveys. Appl. Environ. Microbiol. 66:5368–5382.

Qin G, Zhu L, Chen X, Wang PG, Zhang Y. (2007). Structural characterization and ecological roles of a novel exopolysaccharide from the deep-sea psychrotolerant bacterium Pseudoalteromonas sp. SM9913. Microbiology (Reading, Engl.) 153:1566–1572.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65.

Qin Q-L, Li Y, Zhang Y-J, Zhou Z-M, Zhang W-X, Chen X-L, et al. (2011). Comparative genomics reveals a deep-sea sediment-adapted life style of Pseudoalteromonas sp. SM9913. ISME J 5:274–284.

Ramírez Santos J, Contreras Ferrat G, Gómez Eichelmann MC. (2005). Stationary phase in Escherichia coli. Rev. Latinoam. Microbiol. 47:92–101.

Raven JA. (1986). Physiological consequences of extremely small size for autotrophic organisms in the sea. Canadian Bulletin of Fisheries and Aquatic Sciences 1–70.

Reid PC, Turley CM, Burkill PH. (1991). Protozoan association with marine 'snow' and 'fluff'. (Accessed February 25, 2013).

Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. (2011). Bacterial community transcription patterns during a marine phytoplankton bloom. Environmental Microbiology. (Accessed October 23, 2011).

Römling U, Gomelsky M, Galperin MY. (2005). C-di-GMP: the dawning of a novel bacterial signalling system. Mol. Microbiol. 57:629–639.

Roussel EG, Konn C, Charlou J-L, Donval J-P, Fouquet Y, Querellou J, et al. (2011). Comparison of microbial communities associated with three Atlantic ultramafic hydrothermal systems. FEMS Microbiol. Ecol. 77:647–665.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. 5:e77.

Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL, et al. (2004). The oceanic sink for anthropogenic CO2. Science 305:367–371.

Sawabe T, Tanaka R, Iqbal MM, Tajima K, Ezura Y, Ivanova EP, et al. (2000). Assignment of Alteromonas elyakovii KMM 162T and five strains isolated from spot-wounded fronds of Laminaria japonica to Pseudoalteromonas elyakovii comb. nov. and the extended description of the species. Int. J. Syst. Evol. Microbiol. 50 Pt 1:265–271.

Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ. Microbiol. 11:2078–2093.

Schierwater B, de Jong D, Desalle R. (2009). Placozoa and the evolution of Metazoa and intrasomatic cell differentiation. Int. J. Biochem. Cell Biol. 41:370–379.

Schippers A, Kock D, Höft C, Köweker G, Siegert M. (2012). Quantification of Microbial Communities in Subsurface Marine Sediments of the Black Sea and off Namibia. Front Microbiol 3:16.

Schippers A, Neretin LN, Kallmeyer J, Ferdelman TG, Cragg BA, Parkes RJ, et al. (2005). Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. Nature 433:861–864.

Schleper C, Nicol GW. (2010). Ammonia-oxidising archaea--physiology, ecology and evolution. Adv. Microb. Physiol. 57:1–41.

Schmidt TM, DeLong EF, Pace NR. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 173:4371–4378.

Schnitker D. (1980). Quaternary deep-sea benthic foraminifers and bottom water masses. Annual Review of Earth and Planetary Sciences 8:343–370.

Schrenk MO, Kelley DS, Delaney JR, Baross JA. (2003). Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. Appl. Environ. Microbiol. 69:3580–3592.

Shacter E. (2000). Quantification and significance of protein oxidation in biological samples. Drug Metab. Rev. 32:307–326.

Sherwood AR, Presting GG. (2007). Universal primers amplify a 23S rDNA plastid marker in Eukaryotic algae and Cyanobacteria. Journal of Phycology 43:605–608.

Shi Y, McCarren J, Delong EF. (2011). Transcriptional responses of surface water marine microbial assemblages to deep-sea water amendment. Environmental Microbiology. (Accessed October 14, 2011).

Simon C, Wiezer A, Strittmatter AW, Daniel R. (2009). Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. Appl. Environ. Microbiol. 75:7519–7526.

Stahl DA, Lane DJ, Olsen GJ, Pace NR. (1984). Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. Science 224:409–411.

Stewart FJ, Ulloa O, Delong EF. (2011). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environmental Microbiology. (Accessed December 1, 2011).

Stock AM, Robinson VL, Goudreau PN. (2000). Two-component signal transduction. Annual Review of Biochemistry 69:183–215.

Strange RE. (1968). Bacterial ‘Glycogen’ and Survival. , Published online: 09 November 1968.

Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. (2011). One carbon metabolism in SAR11 pelagic marine bacteria. PLoS ONE 6:e23973.

Takishita K, Yubuki N, Kakizoe N, Inagaki Y, Maruyama T. (2007). Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. Extremophiles 11:563–576.

Taylor BL, Zhulin IB. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol. Mol. Biol. Rev. 63:479–506.

Teske A, Durbin A, Ziervogel K, Cox C, Arnosti C. (2011). Microbial community composition and function in permanently cold seawater and sediments from an arctic fjord of svalbard. Appl. Environ. Microbiol. 77:2008–2018.

Thiel H, Pfannkuche O, Schrieber G, Lochte K, Gooday AJ, Hemleben C, et al. (1990). Phytodetritus on the deep-sea floor in a central oceanic region of the northeast Atlantic. Biol Oceanogr. (Accessed February 25, 2013).

Thomas T, Evans FF, Schleheck D, Mai-Prochnow A, Burke C, Penesyan A, et al. (2008). Analysis of the Pseudoalteromonas tunicata genome reveals properties of a surface-associated life style in the marine environment. PLoS ONE 3:e3252.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Toffin L, Webster G, Weightman AJ, Fry JC, Prieur D. (2004). Molecular monitoring of culturable bacteria from deep-sea sediment of the Nankai Trough, Leg 190 Ocean Drilling Program. FEMS Microbiology Ecology 48:357–367.

Tringe SG, Hugenholtz P. (2008). A renaissance for the pioneering 16S rRNA gene. Curr. Opin. Microbiol. 11:442–446.

Tseng L-C, Dahms H-U, Chen Q-C, Hwang J-S. (2008). Copepod assemblages of the Northern South China Sea. Crustaceana 81:1–22.

Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS ONE 3:e2527.

Vanreusel A, De Groote A, Gollner S, Bright M. (2010). Ecology and biogeography of free-living nematodes associated with chemosynthetic environments in the deep sea: a review. PLoS ONE 5:e12449.

Vanucci S, Dell'Anno A, Pusceddu A, Fabiano M, Lampitt RS, Danovaro R. (2001). Microbial assemblages associated with sinking particles in the Porcupine Abyssal Plain (NE Atlantic Ocean). Progress in Oceanography 50:105–121.

Varela MM, van Aken HM, Herndl GJ. (2008). Abundance and activity of Chloroflexi-type SAR202 bacterioplankton in the meso- and bathypelagic waters of the (sub)tropical Atlantic. Environ. Microbiol. 10:1903–1911.

Vezzi A, Campanaro S, D'Angelo M, Simonato F, Vitulo N, Lauro FM, et al. (2005). Life at depth: Photobacterium profundum genome sequence and expression analysis. Science 307:1459–1461.

Viklund J, Ettema TJG, Andersson SGE. (2012). Independent Genome Reduction and Phylogenetic Reclassification of the Oceanic SAR11 Clade. Mol. Biol. Evol. 29:599–615.

Vila-Costa M, Rinta-Kanto JM, Sun S, Sharma S, Poretsky R, Moran MA. (2010). Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J 4:1410–1420.

Vilibić I, Šantić D. (2008). Deep water ventilation traced by Synechococcus cyanobacteria. Ocean Dynamics 58:119–125.

Voulhoux R, Ball G, Ize B, Vasil ML, Lazdunski A, Wu L-F, et al. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. The EMBO Journal 20:6735–6741.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, et al. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science 326:578–582.

Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, et al. (2009). GeoChip-based analysis of metabolic diversity of microbial communities at the Juan De Fuca Ridge hydrothermal vent. PNAS 106:4840–4845.

Wang Fengping, Wang J, Jian H, Zhang B, Li S, Wang Feng, et al. (2008). Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium Shewanella piezotolerans WP3. PLoS ONE 3:e1937.

Wei Y, Yu M. (2005). Phytoplanktonic desmids community in Donghu Lake, Wuhan, China. Chin. J. Ocean. Limnol. 23:91–97.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. (1991). 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173:697–703.

Witte U, Wenzhöfer F, Sommer S, Boetius A, Heinz P, Aberle N, et al. (2003). In situ experimental evidence of the fate of a phytodetritus pulse at the abyssal sea floor. Nature 424:763–766.

Wolanin PM, Thomason PA, Stock JB. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. Genome Biology 3:reviews3013.

Worden AZ, Chisholm SW, Binder BJ. (2000). In situ hybridization of Prochlorococcus and Synechococcus (marine cyanobacteria) spp. with rRNA-targeted peptide nucleic acid probes. Appl. Environ. Microbiol. 66:284–289.

Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, et al. (2010). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. The ISME Journal 5:414–426.

Zhou M-Y, Chen X-L, Zhao H-L, Dang H-Y, Luan X-W, Zhang X-Y, et al. (2009). Diversity of both the cultivable protease-producing bacteria and their extracellular proteases in the sediments of the South China Sea. Microb Ecol 58:582–590.

Zhou X, Su F, Lu H, Senechal-Willis P, Tian Y, Johnson RH, et al. (2012). An FRET-based ratiometric chemosensor for in vitro cellular fluorescence analyses of pH. Biomaterials 33:171–180.

Zhu H, Tian Y, Bhushan S, Su F, Meldrum DR. (2012). High throughput micropatterning of optical oxygen sensor for single cell analysis. IEEE Sens J 12:1668–1672.

Zubkov MV. (2009). Photoheterotrophy in marine prokaryotes. J. Plankton Res. 31:933–938.